

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP**

JOSÉ RONALDO ALVES ARAÚJO

**CONTRIBUTOS DE RECURSOS DIGITAIS PARA ANÁLISE GRÁFICA DE DADOS
NA DISCIPLINA ESTATÍSTICA DE UMA FORMAÇÃO INICIAL DE
PROFESSORES**

DOUTORADO EM EDUCAÇÃO MATEMÁTICA

**São Paulo
2023**

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP**

JOSÉ RONALDO ALVES ARAÚJO

**CONTRIBUTOS DE RECURSOS DIGITAIS PARA ANÁLISE GRÁFICA DE DADOS
NA DISCIPLINA ESTATÍSTICA DE UMA FORMAÇÃO INICIAL DE
PROFESSORES**

DOUTORADO EM EDUCAÇÃO MATEMÁTICA

*Tese apresentada à Banca Examinadora do Programa de
Estudos Pós-Graduados em Educação Matemática da
PUC-SP, como exigência parcial para obtenção do título
de **Doutor em Educação Matemática**, sob a orientação da
Prof.^a Dr.^a Celina Aparecida Almeida Pereira Abar.*

**São Paulo
2023**

Araújo , José Ronaldo Alves
Contributos de recursos digitais para análise
gráfica de dados na disciplina de estatística de uma
formação inicial de professores . / José Ronaldo Alves
Araújo . -- São Paulo: [s.n.], 2023.
196p. il. ; cm.

Orientador: Celina Aparecida Almeida Pereira Abar.
Tese (Doutorado)-- Pontifícia Universidade Católica de
São Paulo, Programa de Estudos Pós-Graduados em Educação
matemática.

1. Recursos digitais . 2. Transnumeração. 3.
Visualização de dados. 4. Análise gráfica de dados . I.
Abar, Celina Aparecida Almeida Pereira. II. Pontifícia
Universidade Católica de São Paulo, Programa de Estudos
Pós-Graduados em Educação matemática. III. Título.

CDD

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta tese por processos de fotocopiadora ou eletrônicos.

Assinatura: _____ *São Paulo, ___/___/___*

JOSÉ RONALDO ALVES ARAÚJO

**CONTRIBUTOS DE RECURSOS DIGITAIS PARA ANÁLISE GRÁFICA DE DADOS
NA DISCIPLINA ESTATÍSTICA DE UMA FORMAÇÃO INICIAL DE
PROFESSORES**

Banca Examinadora

Prof.^a Dr.^a Celina Aparecida Almeida Pereira Abar – Orientadora
PUC-SP

Prof. Dr. Daniel Couto Gatti

Prof. Dr. Celso Ribeiro Campos

Prof. Dra. Rosa Monteiro Paulo

Prof. Dr. Lori Viali

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, sob número do processo: 88887.370467/2019-00

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, process number: 88887.370467/2019-00

Aos meus pais, Fátima e Luiz.
Aos meus irmãos: Gilson, Fernanda, Uilton, Marcos e Elizangela.
A todo o povo do Estado do Ceará.

AGRADECIMENTOS

A Deus, por prover meios para a realização deste trabalho. A Ele, toda honra e toda glória.

À Prof.^a Dr.^a Celina Abar, orientadora deste trabalho, por seu profissionalismo, orientação e estímulo para a pesquisa. Obrigado por sua paciência, por sua dedicação, por acreditar em mim e em meu projeto.

Ao Prof. Dr. Zsolt Lavicza, por sua colaboração ao longo do desenvolvimento desta pesquisa. Por me receber junto ao seu grupo de pesquisa e oportunizar a aplicação desta investigação.

Aos futuros professores, por terem aceitado participar deste estudo.

Aos professores, Dr.^a Rosa Monteiro Paulo, Dr. Lori Viali, Dr. Celso Ribeiro Campos e Dr. Daniel Couto Gatti, por lerem e tecerem assertivas e valiosas sugestões no exame de qualificação desta pesquisa.

Aos professores do Programa de Estudos Pós-Graduados em Educação Matemática da Pontifícia Universidade Católica de São Paulo.

Aos meus colegas do Grupo de pesquisa, Adriana Dias, Alexandre Russo, Amabile Mesquita, Daniel de Souza, David Correia, Érika Vieira e Thiago Silva.

Aos colegas de turma, Ana Maria Campos, Chrystian Bastos, Dosilia Barreto, João Carlos Leal, José Roberto Campos, Lilian Trindade, Simone Guedes e Sofia Seixas, por tantas discussões e apoio mútuo ao longo dessa jornada.

À Pontifícia Universidade Católica de São Paulo, por promover meios para minha formação pessoal e profissional. Por sua celeridade em se adaptar e manter sua reconhecida qualidade, mesmo em adversidades como a pandemia de Covid-19.

À *Johannes Kepler Universität* que, por meio da *Linz School of Education*, me recebeu e permitiu que eu aprimorasse minha formação ao longo do estágio doutoral.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, por financiar esta pesquisa. É inestimável a sua contribuição para o avanço da ciência brasileira.

The most fundamental idea in a statistical approach to learning is that of forming and changing data representations of aspects of a system to arrive at a better understanding of that system.

(WILD; PFANNKUCH, 1999, p. 227)

RESUMO

Esta pesquisa se concentra na discussão sobre recursos digitais na educação estatística, especificamente o *Gapminder*, o GeoGebra e o SPSS, e seus possíveis benefícios na construção e análise gráfica de dados. O estudo foi realizado durante o estágio de doutorado do pesquisador na *Linz School of Education* da *Johannes Kepler Universität*, Áustria, de outubro de 2021 a janeiro de 2022, durante a participação nas aulas da disciplina de Estatística ministrada pelo professor Dr. Zsolt Lavicza aos alunos de geografia, educação física e outros cursos de formação inicial de professores. Durante a disciplina, foram propostas atividades nas quais os futuros professores utilizaram recursos digitais para analisar dados de problemas reais diversos. Pesquisas preliminares indicaram a escassez de estudos sobre o uso de tecnologias digitais no ensino de Estatística. Como referenciais teóricos, foram adotados o conceito de visualização de dados, que permite explorar e comunicar informações relevantes sobre grandes volumes de dados por meio de representações gráficas, e o conceito de transnumeração, que se refere à formação e alteração de representações de dados para um melhor entendimento. As três atividades propostas, baseadas no ciclo investigativo PPDAC (problema, plano, dados, análises e conclusões), foram realizadas com cada um dos recursos digitais, permitindo que os futuros professores delimitassem problemas relevantes, realizassem busca, limpeza e organização dos dados relacionados aos problemas escolhidos. A análise dos relatórios produzidos por cinco duplas de futuros professores possibilitou a seleção dos materiais para esta pesquisa. O *corpus* de análise consiste nos resultados dos relatórios de dois dos grupos, analisados à luz dos referenciais teóricos da transnumeração e da visualização de dados. Na primeira atividade, utilizando o *Gapminder*, a organização e a análise gráfica dos dados foram realizadas para contribuir na compreensão das variáveis exploradas. Na segunda atividade, com o GeoGebra, os participantes importaram, organizaram e construíram representações de dados, além de realizar análises relacionadas. Na terceira atividade, com a utilização do SPSS, os participantes executaram análises de correlação entre pelo menos três pares de variáveis e interpretaram os resultados, seguindo as fases do ciclo investigativo. Os resultados obtidos demonstram os benefícios de cada um dos recursos digitais e indicam que sua utilização pode complementar significativamente a aprendizagem de conceitos estatísticos.

Palavras-chave: *Gapminder*, GeoGebra, SPSS, Transnumeração, Visualização de dados, Análise gráfica de dados.

ABSTRACT

This research focuses on the discussion about digital resources in Statistical Education about the possible contributions of Gapminder, GeoGebra and SPSS in the constructions and graphical analysis of data. The study was carried out during the author's doctoral internship, from October 2021 to January 2022, at the Linz School of Education of Johannes Kepler Universität, Austria, during the researcher's participation in the statistics classes organized by Professor Dr. Zsolt Lavicza to students of Geography, Physical Education, among other courses, for the initial training of teachers. Throughout the course, activities were proposed in which future teachers used digital technological resources to analyze data of various real problems. Results of preliminary research conducted indicate the scarcity of studies for the use of digital technologies in the teaching of statistics. It is part of the theoretical references the proposition about Data Visualization because it allows to explore and effectively communicate relevant information about large volumes of data through graphical representations. In addition, the concept of transnumeration was also adopted and refers to the idea of forming and changing data representations of a system to reach a better understanding of them. The three proposed activities, built with reference to the PPDAC investigative cycle (problem, plan, data, analyses and conclusions), with each of the digital resources, allowed the future teachers to delimit problems relevant to them, perform searches, cleaning and organization of the data related to the chosen problems. The study of the reports produced by five pairs of future teachers, identified which productions would constitute as analysis material for this study and, thus, the corpus of analysis of this research presents the results of the reports of two of the groups that, in the light of the theoretical references of transnumeration and data visualization, were analyzed. In the first activity with Gapminder the organization and graphical analysis of data was a process carried out to contribute to the understanding of the variables explored. The second activity had as objectives that the participants could import, organize, build data representations mediated by GeoGebra and their respective analyses. In the third activity, with the integration of the SPSS, the participants performed correlation analyses of at least three pairs of variables and interpreted the results, seeking to meet the phases of the investigative cycle. In the results obtained, the contributions of each of the digital technological resources can be verified and indicate that their integration can significantly complement the learning of statistical objects.

Keywords: *Gapminder*, GeoGebra, SPSS, Transnumeration, Data visualization, Graphical analysis of data

LISTA DE FIGURAS

Figura 1 – Framework dimensional para o pensamento estatístico em investigação empírica.....	64
Figura 2 – Atividade proposta com o <i>Gapminder</i>	90
Figura 3 – Atividade proposta com o GeoGebra.....	91
Figura 4 – Atividade proposta com o SPSS	92
Figura 5 – Tela inicial do <i>Gapminder</i>	95
Figura 6 – Porcentagem da população humana vivendo em extrema pobreza em 1800 e 2019	99
Figura 7 – Porcentagem da população mundial por idade em 1950 (esquerda) e em 2021 (direita) ..	101
Figura 8 – Porcentagem da população mundial por idade em 2100.....	102
Figura 9 – Cores representando diferentes regiões do planeta	106
Figura 10 – Representação gráfica por bolhas	107
Figura 11 – Distribuição de renda em percentual da população (1983).....	110
Figura 12 – Expectativa de vida em função do <i>score</i> de felicidade	114
Figura 13 – Número de filhos por mulher em função do <i>score</i> de felicidade	115
Figura 14 – Renda em função do <i>score</i> de felicidade	118
Figura 15 – <i>Score</i> de felicidade em função da média de anos escolares frequentados	119
Figura 16 – Emissões de CO2 per capita em função do <i>score</i> de felicidade.....	120
Figura 17 – Tela do GeoGebra para exploração de ferramentas estatísticas.....	123
Figura 18 – Altura média mundial por sexo de 134 países	126
Figura 19 – <i>Box-plot</i> das alturas médias da população, por sexo, de 134 países	128
Figura 20 – Peso médio da população mundial por sexo	130
Figura 21 – Relação entre a altura média mundial e o peso por sexo	133
Figura 22 – Planilha com dados estatístico sobre a taxa média de vacinação e o número relativo de mortes acumuladas	140
Figura 23 – Representação gráfica dos dados da taxa média de vacinação pelo número relativo de mortes	141
Figura 24 – Representação gráfica da curva do total acumulado de pessoas vacinadas ao longo do tempo na Áustria	145
Figura 25 – Representação da curva e dos setores dinâmicos para o total de vacinados contra Covid-19 ao longo do tempo na Áustria	148
Figura 26 – Tela de apresentação dos recursos estatísticos do SPSS.....	150
Figura 27 – Estatísticas das distribuições dos dados estudados	152
Figura 28 – Distribuição dos dados relativos à emissão de CO ₂	153
Figura 29 – Distribuição dos dados relativos à renda <i>per capita</i>	154
Figura 30 – Distribuição dos dados relativos à população dos 192 países.....	155

Figura 31 – Valores estatísticos relativos ao estudo de correlação das variáveis estudadas	156
Figura 32 – Representação gráfica da razão aluno-professor na educação primária no mundo desde 1970	160
Figura 33 – Representação gráfica da dispersão entre razão aluno-professor e o letramento de jovens (15-24 anos)	163
Figura 34 – Representação gráfica da dispersão entre razão aluno-professor no ensino primário e percentual de matrícula no ensino secundário	164
Figura 35 – Representação gráfica da dispersão entre a razão aluno-professor no ensino primário e a razão aluno-professor no ensino secundário	165
Figura 36 – Valores estatísticos relativos ao estudo de correlação entre razão aluno-professor e letramento entre jovens de 15 a 24 anos	166
Figura 37 - Valores estatísticos relativos ao estudo de correlação entre razão aluno-professor no ensino primário e percentual de matrícula no ensino secundário	168
Figura 38 - Valores estatísticos relativos ao estudo de correlação entre razão aluno-professor no ensino primário e razão aluno-professor no ensino secundário	169

LISTA DE QUADROS

Quadro 1 – Teses e dissertações brasileiras relativas a problemática da pesquisa.....	35
Quadro 2 – Publicações da ICOTS10 e da ICOTS11	49
Quadro 3 – Elementos do letramento estatístico na era <i>OPEN Data</i>	75
Quadro 4 – Tradução da atividade proposta com o <i>Gapminder</i>	90
Quadro 5 – Tradução da atividade proposta com o GeoGebra	92
Quadro 6 – Tradução da atividade proposta com SPSS.....	93

SUMÁRIO

1 INTRODUÇÃO.....	19
2 PERCURSO, PROBLEMÁTICA E PROBLEMA DA PESQUISA	25
3 ESTUDOS PRELIMINARES	31
3.1 Trabalhos selecionados no catálogo de teses e dissertações da CAPES.....	34
3.2 Trabalhos disponíveis nos anais de 2018 e 2022 da ICOTS.....	48
4 REFERENCIAIS TEÓRICOS.....	63
4.1 Pensamento estatístico – transnumeração.....	63
4.2 Visualização de dados	69
5.2 Momento histórico e restrições para a aplicação do estudo.....	84
5.2 As instituições envolvidas e os recursos subjacentes à realização do estudo.....	85
5.3 Contexto de implementação da pesquisa.....	86
5.4 Os recursos digitais utilizados	87
5.5 A proposição das atividades	89
5.6 <i>Corpus</i> do material de análise da pesquisa	94
6 DESENVOLVIMENTO, RESULTADOS E ANÁLISES DAS ATIVIDADES	95
6.1 Atividades desenvolvidas com o <i>Gapminder</i>	95
Dupla 1 - <i>Gapminder</i>.....	95
Dupla 2 - <i>Gapminder</i>.....	105
6.2 Atividades desenvolvidas com o GeoGebra	122
Dupla 1 - GeoGebra	124
Dupla 2 – GeoGebra.....	137
6.3 Atividades desenvolvidas com o SPSS.....	149
Dupla 1 - SPSS	151

Dupla 2 – SPSS	159
7 CONSIDERAÇÕES	177
REFERÊNCIAS.....	187
ANEXO	196

1 INTRODUÇÃO

O levantamento da produção de teses no campo da educação estatística, realizado por Viali e Ody (2020) entre os anos de 1994 e 2018, revela que, embora tenha havido um crescimento ao longo do tempo, esse crescimento ainda era bastante irregular. A partir de 2010, entretanto, parece haver uma tendência de aumento significativo na produção de teses nessa área, quando se considera desde a primeira defesa de tese em 1994: “com exceção dos anos de 1995, 1997 e 1998 quando não ocorreram defesas, todos os demais anos apresentaram pelo menos uma produção na área” (VIALI; ODY, 2020, p. 74).

Os autores consideram que:

A educação estatística está evoluindo, contudo, ocorre de maneira lenta e irregular. Por enquanto não é possível um desmembramento da educação matemática, pois a maioria dos orientadores e pesquisadores da área são licenciados ou bacharéis em matemática. É perceptível a ausência dos formados em estatística na área (VIALI; ODY, 2020, p. 83).

Os pesquisadores observam que o número de doutores na área de educação estatística tem evoluído, talvez pelo crescimento de cursos de pós-graduação na área de ensino de ciências ou de educação matemática.

As motivações para o estudo desta tese tiveram início já em 2014, durante a realização do estágio supervisionado na graduação em matemática, e continuaram no curso de pós-graduação *lato sensu* em estatística aplicada, concluído em 2016.

Os resultados da dissertação de mestrado do autor, defendida em 2018, que envolvia recursos tecnológicos, especialmente o GeoGebra, inspiraram o projeto de pesquisa para o doutorado, aprovado no primeiro semestre de 2019 no programa de Estudos Pós-Graduados em Educação Matemática da Pontifícia Universidade Católica de São Paulo cujo objetivo era investigar as possíveis contribuições do GeoGebra e suas representações gráficas em atividades de análise gráfica de dados, a ser desenvolvido com alunos do ensino médio.

Estudos como Votto (2018) e Santos et al. (2022) revelam que, no contexto brasileiro, há carência de pesquisas que se dedicam a investigar a utilização de recursos tecnológicos no campo da educação estatística. Numa perspectiva estrangeira, Biehler (2018) e McNmara (2018) salientam a necessidade de que os recursos tecnológicos se aproximem das problemáticas latentes à educação estatística, sobretudo porque muitos dos recursos não dispõem de atributos que favoreçam essa aproximação.

Autores como Oliveira (2016) e Sera (2016) indicam que a construção e a análise gráfica ainda são desafios no processo de aprendizagem estatística. Nesse sentido, esta pesquisa foi desenvolvida no contexto das discussões sobre a utilização de recursos digitais na educação estatística, buscando contribuições desses recursos para o campo de pesquisa. Especificamente, o foco desta tese é o estudo da análise gráfica de dados como um processo operacional.

Nesse contexto, devido à importância da utilização de recursos digitais na educação estatística, no início do ano letivo de 2020, foi decidido organizar os meios para implementar o projeto de pesquisa em uma escola pública do estado de São Paulo. Com as autorizações da instituição de ensino e as solicitações éticas legais para o desenvolvimento da pesquisa, deu-se início à mobilização dos potenciais participantes para o estudo.

Entretanto, em meados de março de 2020, a infecção respiratória causada pelo SARS-CoV-2, um vírus da família dos coronavírus que, ao infectar humanos, causa uma doença chamada Covid-19¹, desencadeou uma pandemia e as escolas, assim como toda a sociedade, foram afetadas. Como medida de prevenção ao aumento de infecções, as escolas foram fechadas e todo o processo de ensino passou a ser realizado de forma remota, em caráter emergencial. Devido ao contexto das aulas remotas, a pesquisa sofreu alterações e não foi possível investigar as possíveis contribuições da utilização do GeoGebra em atividades de análise de dados com alunos do ensino médio, devido à perda de contato com os estudantes e ao desligamento institucional do pesquisador da Escola Pública do Estado de São Paulo.

Desse modo, o desenvolvimento deste estudo ocorreu durante o estágio de doutorado realizado entre os meses de outubro de 2021 e janeiro de 2022, sob a supervisão do professor doutor Zsolt Lavicza, na *Linz School of Education* da *Johannes Kepler Universität*, na Áustria, ao longo da disciplina de Estatística oferecida para uma turma de futuros professores, graduandos em matemática, geografia, educação Física, entre outros cursos relacionados à formação de professores.

A viabilidade para o desenvolvimento do pesquisa se concentrou na utilização dos recursos digitais *Gapminder* (<https://www.gapminder.org>), *GeoGebra* (<https://www.geogebra.org/>) e *Statistical Package for the Social Sciences* (SPSS)

¹ <https://butantan.gov.br/covid/butantan-tira-duvida/tira-duvida-noticias/qual-a-diferenca-entre-sars-cov2-e-covid-19-prevalencia-e-incidencia-sao-a-mesma-coisa-e-mortalidade-e-letalidade>.

(<https://www.ibm.com/products/spss-statistics>), levando-se em consideração as condições de desenvolvimento desta tese de doutorado.

Os estudos preliminares realizados elucidaram a utilização de recursos tecnológicos, como *software* e *websites*, como meios que possibilitam construções, explorações e interações no estudo da Estatística. Nesta tese, o foco na análise gráfica de dados pode advir de representações de dados oriundos dos mais diversos contextos.

Tais estudos iniciais, na direção da problemática, foram levantados a partir de dois repositórios: o Catálogo de Teses e Dissertações da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e o repositório de publicações da *International Association for Statistical Education*, especificamente os anais das duas últimas edições da *International Conference on Teaching Statistics (ICOTS)*, realizadas em 2018 e 2022.

A seleção desses trabalhos seguiu pressupostos da pesquisa bibliográfica, de acordo com Soares, Picolli e Casagrande (2018), e os resultados da aproximação às pesquisas disponíveis no Catálogo de Teses e Dissertações da Capes levaram a conceber algumas possibilidades e justificativas para a promoção da educação estatística, sob o ponto de vista da utilização de recursos digitais.

Ao longo da pesquisa, tornou-se possível investigar as possibilidades de contribuições dos recursos tecnológicos *Gapminder*, GeoGebra e SPSS em atividades de análise de dados. Dessa forma, a tese abarca recursos digitais que podem contribuir para a compreensão de análise gráfica de dados. Para conduzir a pesquisa, com base na tese estabelecida, o estudo tem como questão:

Como a utilização de recursos digitais pode contribuir para análise gráfica de dados na disciplina estatística em uma formação inicial de professores?

A partir da questão foram delimitados os seguintes objetivos para o estudo desta tese, oriundos nos relatos dos futuros professores participantes:

- 1) Identificar e analisar possibilidades do *Gapminder*, GeoGebra e SPSS que se configuram como favoráveis para análise gráfica de dados;
- 2) Identificar recursos presentes nessas ferramentas que caracterizam tais possibilidades;
- 3) Identificar nesses recursos possibilidades que permitam aprimoramentos na análise gráfica de dados.

Ao delimitar o problema de pesquisa e os objetivos deste estudo, considerou-se como aporte teórico as contribuições dos autores Wild e Pfannkuch (1999, p. 227, tradução nossa²) sobre a transnumeração, “[...] um processo dinâmico de mudanças de representações para gerar compreensão”, mais especificamente nos seus estágios (WILD; PFANNKUCH, 1999; PFANNKUCH, RUBICK, 2002, CHICK, 2004). Além da transnumeração, utiliza-se o aporte teórico da visualização de dados, que “[...] pode ser visto como incluindo análise exploratória de dados assistida por computador de conjuntos de dados volumosos e complexos” (PRODROMOU; DUNNE, 2017, p. 3, tradução nossa³) e contribuiu para a análise dos relatórios das atividades desenvolvidas.

O ciclo investigativo – PPDAC (Problema, Planejamento, Desenvolvimento, Análise e Conclusão) é uma dimensão que descreve o processo de investigação estatística, conforme Wild e Pfannkuch (1999) e foi o referencial metodológico para a implementação da pesquisa com a proposição de atividades, denominadas *Homework*, desenvolvidas em duas duplas.

A metodologia para análise dos resultados obtidos tomou como referência uma abordagem qualitativa. O acesso aos relatórios dos participantes, em formato *Portable Document Format* (PDF), permitiu investigar os possíveis contributos dos recursos digitais utilizados e analisados sob os aportes da transnumeração e da visualização de dados.

Na *Homework 1*, o resultado da utilização do *website Gapminder* para a realização de explorações dos dados por ele mediada possibilitou aos participantes identificar e selecionar temas envolvendo as variáveis relevantes.

A utilização do GeoGebra na *Homework 2*, com a importação de dados, as possíveis construções de representações deles e suas respectivas análises, possibilitou a compreensão dos participantes sobre os dados estudados.

A *Homework 3*, após os futuros professores escolherem bancos de dados e realizarem a importação deles ao SPSS, evidencia que a análise dos dados, a partir das representações gráficas e tabular visualizadas, se mostraram contributos relevantes para a utilização do recurso tecnológico digital.

Ao introduzir esta tese (Capítulo 1), segue a sua organização: no Capítulo 2, estão apresentados o percurso, a problemática e o problema de pesquisa; no Capítulo 3, estão os estudos preliminares, delimitados pelos interesses revelados na problemática da tese; no Capítulo 4, são abordados os referenciais teóricos e metodológicos; no Capítulo 5, são

² A dynamic process of changing representations to engender understanding.

³ [...] can be viewed as including computer-assisted exploratory data analysis of voluminous complex data sets.

apresentados os procedimentos adotados; no Capítulo 6, encontram-se o desenvolvimento, os resultados obtidos e as análises das atividades de duas duplas de futuros professores; e, ao final, são tecidas considerações.

2 PERCURSO, PROBLEMÁTICA E PROBLEMA DA PESQUISA

A apresentação desta pesquisa esclarece o percurso, a problemática e o problema abordados pela investigação realizada. Inicialmente, concentra-se na discussão sobre a implementação de recursos digitais na Educação Estatística e está relacionada aos resultados obtidos na pesquisa de mestrado (ARAÚJO, 2018), que mostraram as potencialidades do GeoGebra para o estudo da estatística, especificamente as Medidas de Tendência Central, por meio de uma sequência de atividades com alunos ingressantes do ensino médio.

Observou-se que o GeoGebra, por meio das representações dos dados, dos valores encontrados para os objetos de estudo e das propriedades implícitas nas representações construídas, contribuiu para a compreensão dos participantes em relação às Medidas de Tendência Central e seus significados no contexto das atividades.

As múltiplas representações, como tabelas, listas, gráficos e valores de medidas, por exemplo, auxiliaram os participantes a compreenderem aspectos dos dados das atividades e os significados relacionados às Medidas de Tendência Central no contexto dos problemas.

De fato, Abar e Lavicza (2019, p. 50) apontam que “algumas ferramentas tecnológicas oferecem condições que favorecem professores e alunos no processo de ensino e aprendizagem”, como foi o caso na pesquisa de mestrado (ARAÚJO, 2018). Nessas condições, “os professores precisam propor situações em que os alunos possam construir conhecimentos com o auxílio das ferramentas e que tais ferramentas permitam identificar os caminhos percorridos pelos alunos e possibilitar respostas às suas atividades” (ABAR; LAVICZA, 2019, p. 50).

Bortolossi (2016) havia destacado as possíveis potencialidades do GeoGebra no contexto da Estatística, considerando as múltiplas representações de dados que o *software* permite como uma vantagem didática. Identificou-se que essas representações contribuem para o poder de argumentação sobre o objeto de estudo e sua relação com os dados originais (ARAÚJO; ABAR, 2019).

A partir de considerações dessa natureza, com o projeto de pesquisa aprovado no primeiro semestre de 2019 para o curso de doutorado do programa de Estudos Pós-Graduados em Educação Matemática da Pontifícia Universidade Católica de São Paulo, foi proposta uma investigação acerca das possíveis contribuições do GeoGebra, de suas representações gráficas em atividades relativas à análise gráfica de dados.

No âmbito do pensamento estatístico, especificamente da Transnumeração, viu-se ser um processo imbricado com possíveis argumentos mobilizados na análise gráfica de dados.

A proposta do projeto era desenvolver uma pesquisa com alunos do Ensino Médio e buscar, por meio de atividades de análise de dados, especialmente análise gráfica, quais seriam os possíveis contributos do GeoGebra nesse processo. Destaca-se que a análise gráfica de dados é concebida de acordo como salienta Patsiomitou (2019), com base nos estudos de Sfard (1989, 1991, 1992):

[...] um objeto matemático, ou um objeto abstrato em geral, pode ser concebido ou interpretado tanto operacionalmente, quando é considerado como um processo realizado ou a ser realizado, quanto estruturalmente quando é interpretado como um objeto permanente com propriedades concretas (PATSIOMITOU, 2019, p. 33, tradução do autor).

Ao propor um estudo a partir de atividades que envolvam o acesso a dados provenientes de contextos específicos, admite-se, conforme Neto (2002), que esses dados (nesta tese entendidos como dados estatísticos), são aqueles observados ou obtidos a partir de experimentos.

Com base nos resultados preliminares dos estudos realizados por outros autores, no Congresso Brasileiro de Informática na Educação de 2019 e na Revista de Produção Discente em Educação Matemática foram destacadas as primeiras percepções sobre o tema da pesquisa (ABAR; ARAÚJO, 2019; ARAÚJO; ABAR, 2020), tomando como referência o estudo de Silva (2007), que enfatiza que o pensamento estatístico é compreendido como as estratégias mentais utilizadas pelo indivíduo para tomar decisões em todas as etapas de um ciclo investigativo. Além disso, de acordo com Lee et al. (2014), o uso de tecnologias oferece oportunidades para a proposição de atividades estatísticas, pois os *softwares* podem facilitar a criação adequada de representações de dados e ações transnumerativas.

Dessas percepções, no Boletim Cearense de Educação e História da Matemática foi publicado um relato sobre o percurso em desenvolvimento (ARAÚJO, 2020), que buscou investigar o GeoGebra como recurso para o estudo da estatística e questionar quais processos de pensamento estatístico permitiriam compreender as representações dos dados, bem como identificar os possíveis recursos do GeoGebra que o tornariam um meio adequado para estudar tais objetos.

No início do ano letivo de 2020, foi decidido organizar os meios para implementar o projeto de pesquisa em uma escola pública do Estado de São Paulo. Com autorizações da instituição de ensino e as solicitações éticas legais para desenvolvimento da pesquisa, iniciou-se a mobilização dos potenciais participantes para o estudo.

Em decorrência da pandemia de Covid-19, em meados de 2020, toda a sociedade foi afetada. Como medida de prevenção ao aumento de infecções, as escolas foram fechadas e todo o processo de ensino passou a ser realizado, em caráter emergencial, de forma remota.

Dado o contexto das aulas remotas, a pesquisa sofreu alterações, e investigar as possíveis contribuições do GeoGebra em atividades de análise de dados com alunos do ensino médio só seria possível se houvesse ajustes metodológicos. Assim, com o aporte metodológico do Ciclo Investigativo proposto por Wild e Pfannkuch (1999), como citou Silva (2007), seria possível investigar as possíveis contribuições do GeoGebra nas estratégias mentais utilizadas pelos participantes em atividades que resultariam em análises gráficas de dados para tomadas de decisões.

Em abril de 2021, foi dado início ao processo de organização do ambiente remoto para a proposição das atividades de pesquisa e os recursos do *Google for Education*, o *Google Classroom* e o *Google Meet* foram adotados. Por meio dessas ferramentas, foram realizadas interações e conduzidas diversas discussões cujo objetivo era estabelecer contextos relevantes para os participantes, a fim de propor a implementação de ciclos investigativos, nos quais seriam identificadas e analisadas as possíveis contribuições do GeoGebra.

As leituras e reflexões sobre o desenvolvimento do GeoGebra e a relevância de estudar estatística mediada por essa ferramenta aproximaram a problemática do estudo de outros centros de pesquisa. Nesse sentido, ao longo de 2020 e 2021, houve aproximação e participação nas discussões realizadas na *Linz School of Education* da *Johannes Kepler Universität*, na Áustria, por meio do Professor Zsolt Lavicza, que formalizou um convite para a realização de um estágio de doutorado junto ao seu grupo de pesquisa.

O estágio proporcionou a oportunidade de participar de seminários e congressos organizados pelo professor e seu grupo de pesquisa, como, por exemplo:

- *Innovations in STEAM Education and Action Research Conference* (<https://www.geogebra.org/m/tsnr5dpi>), realizada em janeiro de 2020;
- *Doctoral Student's Conference: Tradition, Development & Innovation in Didactics-Early Career Researchers Conference*

(<http://sddid.psiedu.ubbcluj.ro/tdidecr2021.htm>), realizada em dezembro de 2021;

- *Arts in STEAM 2022 Conference* (<https://www.geogebra.org/m/txxgevnd>), realizada em janeiro de 2022.

Essas participações também possibilitaram a conexão da pesquisa com o grupo de Lavicza e destacaram a relevância da estadia na *Linz School of Education*, entre outubro de 2021 e janeiro de 2022, com algum suporte financeiro do Plano de Incentivo à Pesquisa (PIPeq – PUC-SP). Em relação ao contexto pandêmico na Áustria, naquela época, uma vez que estava controlado, foi possível realizar o estágio de doutorado. Todavia, a ida para a Áustria implicou no desligamento institucional do pesquisador da Escola Pública do Estado de São Paulo, o que comprometeu o andamento da pesquisa e exigiu um redirecionamento.

Durante o estágio doutoral, foi possível a participação nas aulas da disciplina *Research Methods* (Métodos de Pesquisa) e nos seminários conduzidos pelo Professor Zsolt Lavicza aos alunos dos cursos de mestrado e doutorado da *Linz School of Education*. Além disso, foi possível acompanhar e participar das aulas da disciplina de estatística lecionadas por Lavicza aos alunos de graduação em Matemática, Geografia, Educação Física, entre outros cursos para a formação inicial de professores.

O redirecionamento da pesquisa contou com o apoio do Professor Zsolt Lavicza, responsável pela disciplina de estatística oferecida, o qual viu a oportunidade de ajustar a pesquisa para esse público. Ao longo da disciplina, foram propostas atividades nas quais os futuros professores utilizariam recursos digitais para analisar dados de problemas reais diversos. Isso foi relevante, pois, de acordo com Lopes (2008), problemas próximos à realidade dos participantes se mostram como atividades de natureza que podem complementar de forma significativa a aprendizagem da estatística.

Ao considerar o planejamento do professor para a disciplina de estatística, viu-se que a proposta da pesquisa poderia ser ampliada, pois havia a previsão da implementação dos recursos *Gapminder*, *GeoGebra* e *SPSS*.

Com a implementação da pesquisa, foram publicados e/ou apresentados resultados da utilização dos recursos tecnológicos em atividades de análise gráfica de dados em congressos, como:

- *International Conference on Teaching Statistics* (<https://icots.info/11/>), no qual foi publicado um artigo sobre possíveis contribuições do *Gapminder* (ARAÚJO; ABAR; LAVICZA, 2022);

- *Latin American Conference of STEAM Education (2023)* (<https://www.geogebra.org/m/hhqqq8g6>), na qual foram apresentados resultados quanto à utilização do GeoGebra;
- XV Encontro Paulista de Educação Matemática (<https://www.even3.com.br/xv-epem-encontro-paulista-de-educacao-matematica-294768/>), no qual foram publicados resultados da utilização do SPSS.

Na elaboração dessas publicações, bem como no desenvolvimento da presente tese, a formulação de um problema de pesquisa possibilitou a delimitação de uma potencial contribuição para o campo da Educação Estatística, pelo menos no que diz respeito às lacunas aparentes que surgiram ao longo da revisão da literatura, apresentada no próximo capítulo.

Vinculada às oportunidades de investigar a utilização do *Gapminder*, do GeoGebra e do SPSS como recursos em atividades, a possibilidade de identificar como esses recursos podem contribuir para a análise gráfica de dados é um objetivo a ser alcançado nas discussões sobre a viabilidade da incorporação de recursos tecnológicos para promover a Educação Estatística.

Como Fiorentini e Lorenzato (2012) salientam, ao delimitar objetivos do estudo, esses nortearão as escolhas metodológicas. Nesse sentido, embora a aplicação do estudo tenha passado por redirecionamentos, seja pelo contexto da pandemia de COVID-19, seja pela oportunidade de implementação em contexto diferente do Brasil, considera-se que ao investigar as possíveis contribuições desses recursos tecnológicos implementados em atividades, pode-se evidenciar contribuições que atendam à problemática da Educação Estatística, inclusive em outros contextos, como o brasileiro.

A essa justificativa, acrescenta-se a relevância da demanda curricular que, embora até aqui não tenha sido mencionada, há que se considerar oportuno. Nesse caso, ao contexto do Brasil, as normativas da Base Nacional Comum Curricular (BNCC) (BRASIL, 2018) contribuem para justificar esta pesquisa, na ótica da implementação de recursos tecnológicos no ambiente educacional.

O documento sugere, por exemplo, que a construção de conhecimentos em álgebra, números, geometria e probabilidade e estatística pode contribuir para o desenvolvimento do pensamento matemático dos alunos: “[...] eles precisam ser capazes de traduzir uma situação dada em outras linguagens, como transformar situações-problema, apresentadas em língua materna, em fórmulas, tabelas e gráficos e vice-versa” (BRASIL, 2018, p. 271).

A BNCC, ao se referir à área de conhecimento da probabilidade e da estatística, propõe “a abordagem de conceitos, fatos e procedimentos presentes em muitas situações-problema da vida cotidiana, das ciências e da tecnologia” (BRASIL, 2018, p. 274).

[...] os cidadãos precisam desenvolver habilidades para coletar, organizar, representar, interpretar e analisar dados em uma variedade de contextos, de maneira a fazer julgamentos bem fundamentados e tomar as decisões adequadas. Isso inclui raciocinar e utilizar conceitos, representações e índices estatísticos para descrever, explicar e prever fenômenos (BRASIL, 2018, p. 274).

Essa indicação salienta que há a necessidade de os alunos desenvolverem habilidades que vão ao encontro da problemática desta pesquisa, o que justifica identificar como os recursos tecnológicos contribuem para viabilizar processos para o desenvolvimento delas.

Diretamente relacionada à implementação de uma tecnologia digital, a BNCC indica como foco identificar potencialidades de recursos dessa natureza para a realização de atividades relativas a todas as áreas do conhecimento, inclusive considerando as diversas práticas sociais e anseios do mundo do trabalho (BRASIL, 2018).

Essas considerações permitem justificar a pesquisa que esta tese apresenta, sobretudo sob o ponto de vista da demanda curricular do contexto brasileiro e de pesquisas significativas em Educação Matemática, que, no Brasil, tem sido a área maior em que são observados estudos relativos à Educação Estatística.

Cai et al. (2019a) sugerem que pesquisas na área da Educação Matemática devem ter uma conexão próxima com a sala de aula e impactar diretamente a prática. Ao investigar como o *Gapminder*, o GeoGebra e o SPSS podem contribuir para a análise gráfica de dados, os resultados obtidos podem fornecer subsídios para práticas a serem implementadas em sala de aula.

Ao descrever o percurso, a problemática e o problema abordados nesta tese, juntamente com os objetivos derivados da questão de pesquisa, é importante ressaltar que essas delimitações foram direcionadas e se tornaram plausíveis por meio de uma revisão da literatura existente, abordada no próximo capítulo, no qual se busca identificar as contribuições dos estudos que envolvem o uso de recursos tecnológicos no ensino e aprendizagem da estatística. Somente após essa revisão, no Capítulo 4, são apresentados os fundamentos teóricos para a implementação do estudo.

No próximo capítulo, são apresentados estudos preliminares sobre o tema e suas respectivas contribuições para esta pesquisa.

3 ESTUDOS PRELIMINARES

Neste capítulo, são apresentados estudos preliminares realizados sobre Educação Estatística, com ou sem a mediação de ferramentas tecnológicas, e suas respectivas contribuições para este trabalho.

Chernoff (1978) observou que o computador é um recurso capaz de produzir respostas relevantes, por exemplo, para a organização de grandes quantidades de dados. Além disso, o potencial dos recursos computacionais para apoiar o estudo de bases de dados, especialmente pelo uso adequado de representações gráficas, direcionou a potencialidade computacional.

Fey (1989) salientou o acesso aos computadores como um meio que permitisse aos alunos trabalharem com coleções realistas de dados numéricos, e observou que a comunidade de Educação Estatística se mostrou ativa na exploração dessa oportunidade.

Ao se referir às simulações que os recursos tecnológicos podem permitir, Lèvy (1993) destacou que as visualizações por meio de imagens em uma tela (representações de dados), bem como as simulações, desempenhariam um papel de comunicação ou de persuasão importante. Nesta tese, as simulações apresentadas podem contribuir para o uso de recursos tecnológicos e oferecer uma possibilidade de tornar compreensível o que os conjuntos de dados, por meio de representações gráficas, teriam a dizer.

Ben-Zvi e Friedlander (1997) defendem que tecnologia oferece a oportunidade de criar um ambiente de aprendizado totalmente novo, inclusive como ferramentas na solução de problemas. O poder de comunicar, de persuadir, de tornar compreensível ou mesmo de resolver problemas pode ser considerado como indicações de que os avanços tecnológicos podem favorecer seus usuários quanto a essas ações (BEN-ZVI; FRIEDLANDER, 1997).

Ao se voltar para um aspecto relativo à estatística, a Análise Exploratória de Dados, desenvolvida conforme Tukey (1977), Andre (2020) pontua que os pacotes de *software* permitem explorações de conjuntos de dados, muitas vezes, por métodos de visualização, de modo que a conexão entre estatística, as visualizações e tecnologia computacional tem sido indissociável.

Viali e Ody (2020), ao realizarem um levantamento de pesquisas de mestrado e doutorado no âmbito da educação estatística no Brasil, sublinham que desde a década de 1990 tem havido o desenvolvimento de estudos dessa natureza, mas a produção, embora crescente, tem se mostrado irregular ao longo dos anos.

Diante dessas considerações preliminares, este capítulo contempla estudos que permitem problematizar a implementação de recursos tecnológicos no âmbito da educação estatística. Com base em Martins e Theóphilo (2016), o levantamento de bibliografias possibilita conhecer, analisar e explicar contribuições sobre determinado assunto, tema ou problema. Para Fonseca (2002, p. 31), “permite ao pesquisador conhecer o que já se estudou sobre o assunto”. Soares, Picolli e Casagrande (2018) enfatizam que pesquisa bibliográfica é a essência do estudo exploratório e como uma referência metodológica – adotada para conduzir o processo de revisão desta pesquisa – tem as seguintes etapas:

- I. levantamento de referências sobre o tema de pesquisa;
- II. pré-leitura;
- III. leitura seletiva;
- IV. leitura crítica.

A primeira etapa indica atentar-se à cobertura, qualidade e relevância das referências; a segunda etapa consiste, essencialmente, em buscar eliminar obras que não condizem com o objetivo de pesquisa; a terceira etapa é destinada a eliminar referências que não contenham informação útil para a pesquisa, e, a quarta etapa, realizada de forma aprofundada, tem como objetivo separar o indispensável do complementar ou desnecessário em relação ao objeto de pesquisa.

Na primeira etapa, o levantamento dos estudos foi realizado nos âmbitos nacional e internacional. Quanto à cobertura das pesquisas no âmbito nacional, foi realizado um levantamento de teses e dissertações no Catálogo do Banco de Teses da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. Essa é uma plataforma que disponibiliza estudos realizados por pesquisadores de programas de todo Brasil. Os aspectos relativos à qualidade e à relevância podem ser justificados pelo fato de esses estudos serem realizados por pesquisadores de programas credenciados pela CAPES. Esses programas passam, periodicamente, por avaliações que atestam a qualidade de cada um deles.

No catálogo da CAPES, encontram-se trabalhos das mais diferentes áreas de pesquisa, e, para identificar trabalhos relacionados ao assunto do interesse desse estudo, foram inseridas palavras-chave que os localizassem no base de dados.

As palavras-chave foram “ensino de estatística”, “tecnologias digitais e ensino de estatística”, “educação estatística”, “tecnologias digitais e educação estatística”, “tecnologias e ensino de estatística”, “tecnologias e educação estatística”, “aprendizagem de estatística”, “ensino e aprendizagem de estatística” e “pensamento estatístico”.

No âmbito internacional, como a problemática está ligada ao uso de recursos tecnológicos para o ensino e à aprendizagem da estatística, a busca inicial foi direcionada para organizações que se atentam a essas questões. A *Red Latinoamericana de Investigación en Educación Estadística*⁴ – RELIEE, frequentemente aborda discussões e publicações produzidas por pesquisadores estrangeiros, por exemplo, publicações e eventos que a *International Association for Statistical Education*⁵– IASE organiza e que, geralmente, envolvem pesquisadores dos mais diversos países.

Ao acessar os anais desses eventos, foram identificadas contribuições para esta pesquisa. Mas, para delimitar a busca nos anais de eventos, foi selecionado um evento suficientemente abrangente, por conta da amplitude que a revisão bibliográfica exige. Dadas as questões de qualidade e relevância, a revisão ficou concentrada nos anais da *International Conference on Teaching Statistics* – ICOTS.

A razão para essa escolha se deve ao fato de que a própria IASE, em seu *website*, destaca esse evento como o mais relevante do calendário internacional da Educação Estatística, que, a cada quatro anos, congrega as contribuições e discussões da comunidade internacional.

Desde sua primeira edição, ocorrida em 1982, o evento totaliza onze edições até o momento. Para a revisão realizada, foram consideradas discussões mais atuais, de modo que nesta pesquisa utilizam-se os anais das duas últimas edições do evento ocorridas em 2018 e 2022.

As discussões empreendidas na ICOTS abordam diferentes aspectos relativos ao ensino de estatística: para diferentes idades escolares, para não estatísticos, sobre o desenvolvimento de materiais didáticos, a formação de professores, o uso de recursos tecnológicos no ensino, entre outros. Esse levantamento se debruça sobre aspectos relativos ao uso de recursos digitais para o ensino da estatística.

Esclarecido o percurso realizado para esse levantamento de estudos preliminares relacionados ao tema desta pesquisa, a seguir estão os resultados obtidos a partir do Banco de Teses e Dissertações (BTD) da CAPES e, em seguida, a revisão a partir dos anais da ICOTS.

⁴ Disponível em: <https://reliee.weebly.com/>. Acesso em: jun. 2023.

⁵ Disponível em: <https://iase-web.org/>. Acesso em: jun. 2023.

3.1 Trabalhos selecionados no catálogo de teses e dissertações da CAPES

Para o levantamento nesta base de dados, além das palavras-chave, foram utilizados os filtros:

- espaço temporal: de 2016 a 2023;
- tipo de trabalho: dissertações de mestrado e teses de doutorado;
- áreas de conhecimento: Ensino e Ensino de Ciências e Matemática;
- área de avaliação: Ensino;
- área de concentração: Educação Científica e Tecnológica, Educação em Ciências, Educação em Ciências e Matemática, Educação Matemática, Ensino de Ciências e Matemática, Ensino de Ciências, Ensino de Ciências e Educação Matemática.

As palavras consideradas para o levantamento foram as que indicavam objetos de estudo da estatística para o ensino e/ou a aprendizagem dela, ou para o uso de recurso tecnológico aplicado para essas finalidades. Palavras como gráficos, variabilidade, dados, letramento estatístico, educação estatística, pensamento estatístico, entre outras, foram consideradas. Como resultados da leitura dos títulos, conclui-se a primeira fase do levantamento no Catálogo de Teses e Dissertações da CAPES com trinta e cinco (35) trabalhos.

Com o objetivo de eliminar trabalhos que se afastassem da problemática, foi realizada a segunda fase da pesquisa bibliográfica em relação a essa base de dados, a partir da leitura dos resumos de cada um deles cujo resultado da leitura dos resumos foi a exclusão de três (3) trabalhos que não atendiam ao contexto desta pesquisa. Então, para a fase de leitura seletiva, foram selecionados os trinta e dois (32) trabalhos restantes.

Inicialmente, para essa terceira fase, foi possível identificar, a partir dos resumos, que as pesquisas estavam voltadas para a formação de professores e à aplicação com alunos da educação básica. Como nessa etapa requeria uma seleção aprimorada, foi necessária a leitura integral dos trabalhos, a fim de identificar as contribuições das pesquisas para o ensino e para a aprendizagem da estatística, especialmente no que se refere ao uso de recursos tecnológicos para essas finalidades.

O resultado dessa leitura seletiva delimitou quatorze (13) estudos, como estabelecido na quarta fase da pesquisa bibliográfica. No Quadro 1, constam os estudos brasileiros, em ordem alfabética dos autores.

Quadro 1 – Teses e dissertações brasileiras relativas a problemática da pesquisa

AUTOR, ANO	TÍTULO	NÍVEL	INSTITUIÇÃO
ALBUQUERQUE, 2018	<i>Escala Apresentada em Gráficos: Conhecimentos Matemáticos para o Ensino dos anos iniciais do Ensino Fundamental (crianças e EJA)</i>	Doutorado em Educação Matemática e Tecnológica	Universidade Federal do Pernambuco
ARAÚJO, 2018	<i>Atividades para o estudo das Medidas de Tendência Central: uma proposta com o apoio do GeoGebra</i>	Mestrado em Educação Matemática	Pontifícia Universidade Católica de São Paulo
FRANCISCO, 2016	<i>Interpretação de dados estatísticos: um estudo com alunos do ensino médio da educação de jovens e adultos</i>	Mestrado em Educação Matemática e Tecnológica	Universidade Federal do Pernambuco
GIORDANO, 2017	<i>Panorama da competência estatística no Ensino Médio brasileiro: das ideias e práticas dos professores ao desempenho dos alunos no ENEM</i>	Doutorado em Multiunidades em Ensino de Ciências e Matemática	Universidade Estadual de Campinas
LUCAS, 2019	<i>Estudo dos Níveis de letramento estatístico e dos Estágios de Desenvolvimento Cognitivo no Programa LeME</i>	Mestrado em Educação em Ciências Química da Vida e Saúde	Fundação Universidade do Rio Grande
MELO, 2017	<i>O pensamento estatístico no Ensino Fundamental: uma experiência articulando o desenvolvimento de projetos de pesquisa com os conceitos básicos da estatística implementados em uma sequência didática eletrônica</i>	Doutorado em Ensino de Ciências e Matemática	Universidade Luterana do Brasil
OLIVEIRA, 2016	<i>Educação Estatística em escolas do povo Xukuru do Ororubá</i>	Mestrado em Educação Matemática e Tecnológica	Universidade Federal do Pernambuco
PONTES, 2020	<i>Aprendizagem de gráficos com e sem uso do Excel por alunos do 5º ano Ensino Fundamental</i>	Mestrado em Educação Matemática e Tecnológica	Universidade Federal do Pernambuco
RENAUX, 2017	<i>O uso de Objetos de Aprendizagem de Estatística em um curso de Pedagogia: algumas possibilidades e potencialidades</i>	Mestrado em Educação em Ciências e em Matemática	Universidade Federal do Paraná
SANTOS, 2019	<i>A construção do letramento estatístico em estratégias com o uso de tecnologias digitais em aulas de estatística de cursos de graduação</i>	Mestrado em Educação Matemática	Pontifícia Universidade Católica de São Paulo
SERA, 2016	<i>Conhecimentos de professores para o ensino da leitura e construção de gráficos estatísticos na educação básica</i>	Mestrado em Educação Matemática	Universidade Anhanguera de São Paulo
SOUZA, 2019	<i>Interpretação de Gráficos: explorando o letramento estatístico dos professores de escolas</i>	Mestrado em Educação	Universidade Federal do Pernambuco

	<i>públicas no campo nos espaços de oficinas de formação continuada</i>	Matemática e Tecnológica	
VOTTO, 2018	<i>As potencialidades lúdicas nas estratégias para o ensino e a aprendizagem estatística nos anos iniciais do ensino fundamental</i>	Mestrado em Educação em Ciências Química da Vida e Saúde	Fundação Universidade do Rio Grande

Fonte: Dados da pesquisa.

A partir das pesquisas levantadas, conforme o Quadro 1, foram quatorze (13) trabalhos, os quais passaram por uma leitura crítica com o objetivo de encontrar contribuições relevantes para a implementação de recursos tecnológicos no ensino e na aprendizagem da Estatística, especialmente no que diz respeito à análise gráfica de dados.

A discussão emergente do levantamento bibliográfico teve início a partir da problemática das escalas gráficas, presente na tese de Albuquerque (2018), que destaca a importância desse elemento para o letramento estatístico. Para Albuquerque (2018), as mensagens estatísticas geralmente são apresentadas em representações gráficas, e ao observar essas representações, “é possível constatar que os eixos e a escala são elementos extremamente importantes para a estrutura do gráfico” (ALBUQUERQUE, 2018, 48), pois são eles que fornecem informações quantitativas sobre os dados representados que influenciam diretamente o aspecto visual da representação dos dados.

Albuquerque (2018) salienta que a proporcionalidade é um aspecto relevante no que diz respeito à escala apresentada em gráficos. Assim, nas etapas de construção de escalas, é a proporcionalidade o aspecto garantidor da fidedignidade dos dados representados em seus diferentes tipos de intervalo. Portanto, uma dificuldade é “a compreensão dos valores contínuos apresentados na escala, uma vez que é necessário que os alunos estabeleçam a proporcionalidade entre os pontos explicitados na escala adotada” (ALBUQUERQUE, 2018, 55).

Nesse sentido, observa que “alunos de diferentes níveis de escolaridade e idade, além de docentes em formação, apresentam dificuldades tanto com a interpretação como com a construção de escalas” (ALBUQUERQUE, 2018, 61). Sob essa conjectura, com referência ao Conhecimento Matemático para o Ensino – MKT, proposto por Ball, Thames e Phelps (2008), o estudo da pesquisadora teve como objetivo definir Conhecimentos Matemáticos para o Ensino de escala apresentada em gráficos estatísticos (ALBUQUERQUE, 2018, p. 86) e se ateve ao contexto da formação continuada no intuito de permitir a apropriação, pelos professores, do ensino de escalas apresentadas em gráficos.

Os resultados relevantes obtidos a partir das pesquisas preliminares, de acordo com Albuquerque (2018), são os seguintes:

1. Conhecimento Comum de Escala - saber localizar corretamente um valor na escala e saber construir corretamente a escala do gráfico; 2. Conhecimento Especializado de Escala - considerar as especificidades dos diferentes intervalos da escala, considerar a diferença de dificuldade em localizar um valor implícito ou explícito na escala, considerar que os diferentes conjuntos numéricos influenciam no intervalo da escala e considerar que a proporcionalidade é um atributo da escala; 3. Conhecimento de Escala e do Aluno - compreender as dificuldades e facilidades dos alunos para interpretar e construir a escala; 4. Conhecimento de Escala e Ensino - explorar as habilidades representar, localizar, analisar, comparar e construir escalas em diversas situações e com diferentes unidades escalares e variar as situações de ensino e uso do conceito de escala envolvendo atividades de medida de comprimento, de reta numérica e de mapas; 5. Conhecimento de Escala e Currículo - conhecer o que os programas curriculares oficiais (BNCC e currículos locais) prescrevem para o trabalho com escala e conhecer o que o livro didático propõe para o trabalho com escala” (ALBUQUERQUE, 2018, p. 280-281).

Com base nos resultados de Albuquerque (2018), fica evidente que o conhecimento sobre o estudo de escalas apresentadas em gráficos se mostra relevante, à medida que oferece parâmetros para análises de algumas representações gráficas. À luz da problemática desta tese, a escala em representações gráficas de dados emerge como um objeto de estudo, que deve ser considerado na proposição de atividades, especialmente com o aporte de recursos tecnológicos.

Sobre a utilização de recursos tecnológicos, a pesquisa de Araújo (2018) buscou identificar potencialidades do GeoGebra para o estudo das medidas de tendência central. Com referência nas dialéticas de ação, formulação e validação e na institucionalização, presentes na Teoria das Situações Didáticas do pesquisador francês Guy Brousseau, o estudo traz a aplicação de uma sequência de atividades com alunos ingressantes no ensino médio, sob o auxílio metodológico de pressupostos da engenharia didática, para investigar as possíveis potencialidades do GeoGebra.

A escolha do GeoGebra considerou a vantagem didática de oferecer diferentes representações de um mesmo objeto que interagem entre si, verificando, nesse processo, que o recurso contribuiu com a dialética de ação dos participantes, ao possibilitar a organização dos dados dos problemas. O *software* possibilitou encontrar respostas para alguns questionamentos relacionados a esses dados (ARAÚJO, 2018, p. 133).

Na dialética de formulação, por parte dos alunos, Araújo (2018, p. 133) observou que organizar os dados em tabelas, listas, listas ordenadas, representação gráfica, ou os valores

para as medidas de tendência central, permitiu que eles discutissem sobre a representação dos valores em relação ao conjunto de dados. A representação gráfica auxiliou a expressarem ou descreverem significados para as medidas de tendência central e a representação gráfica contribuiu para argumentação de uma validação de significados para os objetos de estudo.

Quanto à aprendizagem estatística, na dialética de formulação, os valores encontrados para as medidas de tendência central e as múltiplas representações dos dados nas janelas do GeoGebra favoreceram aos participantes conjecturarem sobre a relação dos valores encontrados com os dados de sua origem (ARAÚJO, 2018).

No estudo de Araújo, as múltiplas representações de um mesmo conjunto de dados proporcionaram questionamentos e discussões adicionais sobre as propriedades dos objetos estudados, contribuindo, no contexto estudado, para a institucionalização do saber envolvido. Isso evidenciou as múltiplas representações de dados possíveis com o GeoGebra para a proposta de atividades relacionadas às medidas de tendência central.

A pesquisa de Francisco (2016) investigou o raciocínio da variação de dados em representações gráficas, constatando que há uma carência dessa discussão na disciplina de matemática na educação básica, o que o instigou a propor um estudo sobre interpretação de representações gráficas de dados com alunos do ensino médio da Educação de Jovens e Adultos (EJA).

Para Francisco (2016, p. 43), “faz-se necessário que os leitores desenvolvam as capacidades de pensar, raciocinar, a fim de se letrarem estatisticamente, para que possam interpretar corretamente as informações que recebem”, por exemplo, na forma de gráficos de barras e gráficos de linhas.

O autor considera o raciocínio estatístico “como o desenvolvimento de um sistema de conhecimento, para realizar algo que não foi dito, claramente, como fazê-lo” (FRANCISCO, 2016, p. 43). Sobre essa competência estatística, o pesquisador ressalta “ser difícil de ser alcançado pelos alunos, uma vez que requer uma grande vivência em conhecimentos estatísticos, seja na escola ou fora dela” (p. 47).

Para ele, os processos fundamentais na análise de dados são a descrição, a organização, a redução, a análise e a representação desses dados, componentes do pensamento estatístico, que, no contexto de sua pesquisa, se voltam a identificar tendências, fazer inferências ou previsões a partir de representações gráficas.

Os resultados obtidos por Francisco (2016) emergiram de uma análise exploratória a partir de uma proposta de atividades aplicada a vinte e três (23) alunos de uma turma do III

módulo do ensino médio da EJA. O autor observou lacunas de aprendizagem referentes à leitura e interpretação de gráficos, especialmente em relação à análise da variação. Além disso, observou que os participantes não tinham conhecimento ou não se lembravam do significado de variação e, sobretudo, do conceito de amplitude, fatores que contribuíram para o baixo desempenho nas atividades (FRANCISCO, 2016).

Tais resultados apontam a existência de lacunas de aprendizagem sobre a análise de variação, mesmo por meio de representações gráficas de dados, revelando a necessidade de fomentar meios para promover a superação de lacunas dessa natureza, de modo que a proposta desta tese pode contribuir nessa direção.

Giordano (2017) realizou seu estudo com o objetivo de compreender o *status* atual do ensino de Estatística do ensino médio brasileiro. Sobre as origens do ensino de estatística, a pesquisadora aponta que “o desenvolvimento de recursos estatísticos para resolver problemas aplicados a diversas áreas do conhecimento no início do século XX despertou a necessidade de ensinar estatística a futuros profissionais e pesquisadores” (GIORDANO, 2017, p. 35).

Nesse sentido, a estudiosa assevera que, além do domínio sobre conceitos e aplicações de modelos estatísticos para a resolução de situações-problema ao se referir ao ensino, “espera-se que os alunos desenvolvam uma postura crítica e reflexiva diante das informações estatísticas que, a nosso ver, se assemelha à postura característica de pesquisadores científicos” (GIORDANO, 2017, p. 66). Dentre os elementos que compõem essa postura crítica e reflexiva, a autora destaca que “é necessário elaborar conjecturas, buscar regularidades, generalização de padrões, ser capaz de argumentar, dentre outras ações que estão intimamente relacionadas à habilidade de resolver problemas” (p. 97).

No *status* atual apresentado por Giordano (2017), uma contribuição à problemática desta tese revela-se no anseio de uma abordagem de ensino de estatística que promova oportunidades de desenvolvimento de posturas críticas e reflexivas. Como esta tese busca analisar como a inserção de recursos digitais pode contribuir para análise gráfica de dados, é possível que nesse processo haja contribuições para argumentos que favoreçam o desenvolvimento desses tipos de posturas.

Ao analisar a dissertação de Lucas (2019) do ponto de vista da ação pedagógica, nota-se que seu estudo investigou os níveis de letramento estatístico em atividades planejadas por professores. O recurso teórico utilizado para identificar esses níveis se baseou nas proposições de Watson e Callingham (2003).

Com auxílio da metodologia de análise de conteúdo, Lucas (2019) identificou, a partir dos planejamentos e registros de um grupo de professores, possibilidades de cada atividade fomentar a mobilização de habilidades acerca do letramento estatístico. Nessa direção, os resultados obtidos em sua pesquisa apontam que os professores planejaram atividades que objetivavam “estimular, de forma dinâmica, os estudantes a descobrirem estatísticas presentes em materiais impressos e digitais [...] uma visão crítica sobre a estatística nas mídias” (LUCAS, 2019, p. 102).

Sobre os objetos de estudo da estatística, os planejamentos dos professores envolveram conceitos relativos à proporcionalidade, unidades de medidas, medidas de tendência central e gráficos. Lucas (2019) destaca outras intenções nos planejamentos, como a de fomentar a compreensão da variação, probabilidades, hipóteses e deduções, vislumbradas pelos professores. O estudo aponta que as atividades presentes nos planejamentos podem sugerir a possibilidade de promover “uma compreensão a um nível crítico de letramento estatístico” (p. 102).

A partir da pesquisa de Lucas (2019), nesses estudos preliminares, foi possível refletir sobre o planejamento de atividades e a elaboração de um esquema para promover a formação e o ensino da estatística, de modo que, no âmbito da problemática desta pesquisa, reflexões dessa natureza podem ser consideradas potenciais contribuições das representações de dados, mediadas por algum recurso tecnológico, para esse propósito.

No contexto da análise da uma representação da curva normal, a formação promovida por Macedo (2016) permitiu reflexões sobre dados, de modo que, para a situação avaliada, “a média deixou de ser apenas um valor que representa um valor central para o grupo, e passou a compor um conjunto de dados a serem considerados, juntamente com outras medidas como mediana e moda” (MACEDO, 2016, p. 146).

Melo (2017) avaliou as contribuições para o desenvolvimento do pensamento estatístico, a partir da implementação de uma sequência didática eletrônica: “um conjunto de atividades pedagógicas organizadas e implementadas com uso das tecnologias” (MELO, 2017, p. 130), prevista para alunos do ensino fundamental, por meio de uma articulação da estratégia metodológica de projetos.

Sobre a implementação de recursos tecnológicos para a Educação Estatística, Melo ressalta a utilização da resolução de problemas para a realização de experimentos e simulações. A autora salienta que, “no processo de ensino e aprendizagem da Estatística, o uso da tecnologia como estratégia metodológica permite auxiliar a interação com diferentes

formas de representação simbólica, como gráficos, planilhas, textos e tabelas, por meio de *software*” (MELO, 2017, p. 127). Ainda:

As tecnologias que têm função educativa são instrumentos essenciais para elaboração e desenvolvimento de atividades que estimulem o questionamento e o espírito de investigação dos alunos. Assim, ao utilizar a tecnologia o professor poderá diversificar e inovar a forma de trabalho, utilizando métodos e técnicas pedagógicas, indispensáveis no preparo do cidadão (MELO, 2017, p. 128).

Para Melo (2017), cabe aos educadores buscarem por recursos tecnológicos que se adequem às propostas de ensino, de modo a atingir os objetivos educacionais e a formação dos alunos. Com suporte metodológico da plataforma de ensino SIENA⁶, o estudo de Melo (2017) teve a participação de 54 alunos de nono ano do ensino fundamental que, durante as aulas de matemática, de forma não-presencial, acessaram materiais de estudos disponibilizados no sistema SIENA.

Os resultados da aplicação da sequência didática eletrônica apontaram para uma apropriação dos conceitos estatísticos como “amostra, população, frequência absoluta e frequência relativa, dados brutos, rol e elementos essenciais às representações tabulares e gráficas” (MELO, 2017, p. 393), que foram mobilizados a partir das atividades propostas no desenvolvimento de projetos.

As atividades buscaram promover a investigação e a aplicação dos conceitos estatísticos estudados sob aporte de situações reais e concretas. Como sugerido na proposta de Melo (2017), os participantes utilizaram o Excel como recurso tecnológico para construir as representações tabulares e gráficas, o que possibilitou aos estudantes condições necessárias à realização de procedimentos para organizar e representar os dados coletados, “favorecendo assim: a transnumeração, evidenciada na passagem dos dados brutos para o rol e do rol para a representação tabular e/ou gráfica” (MELO, 2017, p. 399).

A pesquisadora identificou a utilização correta da terminologia estatística para analisar, interpretar e comunicar os resultados obtidos:

[...] a implementação (desenvolvimento, aplicação e avaliação) de uma sequência didática eletrônica com a estratégia metodológica de projetos de pesquisa possibilitou [...] condições necessárias para a transformação do objeto teórico de

⁶ Um sistema inteligente que é capaz de comunicar informações sobre o conhecimento dos alunos em determinado tema, tem o objetivo de auxiliar no processo de recuperação de conteúdos matemáticos, utilizando a combinação de mapas conceituais e testes adaptativos (GROENWALD, C. L. O.; RUIZ, L. M. Formação de professores de matemática: uma proposta de ensino com novas tecnologias. *Acta Scientiae*, Canoas, v. 8, n. 2, jul./dez. 2006).

estudo (conceitos estatísticos) em instrumentos cognitivos aplicados a situações práticas e reais, favorecendo o desenvolvimento de habilidades estatísticas articuladas ao desenvolvimento de componentes responsáveis pela formação do pensamento estatístico (MELO, 2017, p. 402).

Segundo destaca Melo (2017), a transnumeração esteve presente em todas as etapas de desenvolvimento da sequência didática eletrônica com a contribuição do recurso tecnológico adotado, o Excel.

As contribuições da pesquisa de Melo (2017) a esta tese residem no fato de apontar o que um recurso tecnológico pode trazer para o ensino da estatística, colaborando para as estratégias transnumerativas. Ao implementar uma sequência didática eletrônica, Melo (2017) encoraja a busca por recursos dessa natureza para o ensino de Estatística.

No caso de Oliveira (2016), seu estudo envolveu a formação de professores no planejamento e realização de atividades relacionadas à Educação Estatística em sala de aula. No que diz respeito ao planejamento das atividades, Oliveira (2016) observou a proposição de etapas do ciclo investigativo para a Educação Estatística e uma idealização metodológica abrangente voltada para a constituição de um ambiente de grupo colaborativo.

Os resultados mostraram algumas fragilidades nas proposições dos professores participantes, por exemplo, a construção e a interpretação de representações de gráficos de colunas. “Suas produções apresentaram ausência de alguns elementos fundamentais na estrutura da representação gráfica, tais como os eixos, valores percentuais e legendas” (OLIVEIRA, 2016, p. 140).

Sob o aporte do ciclo investigativo, seus passos motivaram os professores a desenvolverem sua prática. Oliveira (2016, p. 142) destaca que as atividades planejadas para o ensino buscaram “problematizar criticamente alguns aspectos do contexto de vida da comunidade indígena – contexto em que se desenvolveu a pesquisa – se configurando uma característica potencializadora deste trabalho”.

A proposição de atividades embasadas no ciclo investigativo foi um fator relevante no trabalho, e, embora o pesquisador não tenha vislumbrado essa problemática, as possíveis contribuições que o uso de recursos tecnológicos teria nesse ciclo investigativo revelam potenciais fragilidades a serem superadas na construção de atividades relativas à análise gráfica de dados. A conjectura dessa possibilidade de superação instiga a problemática desta tese em relação à inserção de recursos tecnológicos na Educação Estatística.

Sobre as práticas que utilizam a tecnologia para o ensino da Estatística, o estudo de Pontes (2019) ressalta que ainda existem lacunas, visto que “ensinar, nos dias de hoje, requer práticas diferenciadas e metodologias que considerem os estudantes do contexto atual, que estão imersos em um mundo permeado pelas tecnologias digitais” (PONTES, 2019, p. 20). Dessa problemática, Pontes propôs investigar a aprendizagem de alunos do ensino fundamental em relação à compreensão da construção de gráficos de barras, a partir de tabelas mediadas pelo Excel, pois, de acordo com o autor, “os gráficos são constantemente utilizados para diversos fins e em variados contextos sociais, como forma de comunicação no cotidiano das pessoas” (PONTES, 2019, p. 38).

Embora tenha destacado a utilização do Excel em seu experimento, Pontes ressalta que essa ferramenta não é um aplicativo criado para fins educativos, portanto, sua inserção na prática educativa requer uma reflexão cuidadosa sobre sua aplicação. Os recursos oferecidos pelo Excel, por exemplo, a construção de diferentes formas de representação, como gráficos, planilhas, banco de dados e tabelas, podem proporcionar ao aluno a oportunidade de interagir com diferentes recursos, indo além do uso tradicional de papel, régua, lápis e borracha (PONTES, 2019).

Em seu estudo, Pontes (2019) propôs uma intervenção com duplas de alunos de quinto ano do ensino fundamental, envolvendo o estudo de gráficos estatísticos e observou que os alunos tinham conhecimentos referentes às representações gráficas, mas não conseguiam extrair delas, por exemplo, informações que permitissem organizar uma representação tabular, ou seja, os resultados revelaram que os alunos participantes apresentavam fragilidades em relação à organização de representações dos dados.

O resultado de Pontes (2019) evidencia que a utilização de recursos digitais pode contribuir para facilitar diferentes formas de representação de dados. Além disso, ressalta a existência de lacunas de pesquisas sobre práticas que utilizam a tecnologia para o ensino da Estatística, de modo que, ao se adotar determinado recurso tecnológico para essa prática, há que se considerar se ele efetivamente é destinado para promoção do ensino ou não. Para esta tese, essas considerações revelam a necessidade de refletir e realizar possíveis adaptações para que os objetivos possam ser alcançados.

Renaux (2017), ao investigar objetos de aprendizagem de estatística, sugere que as tecnologias digitais estão presentes no cotidiano das pessoas e que elas podem criar conexões com o contexto do ensino. As tecnologias digitais “podem promover interação entre professores, alunos e conteúdos” (RENAUX, 2017, p. 13), de modo que, diante dessa

realidade, posturas e metodologias de ensino devem ser objeto de reflexão, para que haja uma adequação a essa nova realidade dos alunos.

A partir dessa problemática, Renaux (2017) define objetos de aprendizagem como “quaisquer recursos virtuais multimídia, que podem ser usados e reutilizados com o intuito de dar suporte à aprendizagem de um conteúdo específico por meio de atividade interativa, apresentada na forma de animação ou simulação” (p. 13).

Nos resultados, a percepção de um participante revelou que “os conteúdos no OA (objeto de aprendizagem) são trabalhados de forma dinâmica e aplicados de maneira que podemos interagir com eles e ter uma maior compreensão” (RENAUX, 2017, p. 101). Além dessa percepção, Renaux (2017) aponta que os objetos de aprendizagem “permitiram a interatividade dos alunos com o conteúdo explorado e puderam trazer, na forma de simulações e atividades interativas, uma maneira com que os conteúdos tivessem mais significado do que a obtida pelos meios tradicionais de ensino” (p. 101).

Para esta tese, ao considerar o OA como definido por Renaux (2017), é possível observar que a inserção de recursos digitais pode fomentar outras formas de compreender os objetos estatísticos. Ao explicitar um caso de um participante que encontrou dificuldades para entender a construção de gráficos somente com aulas expositivas, a autora destaca que, ao interagir com o objeto de aprendizagem, o participante “teve mais facilidade em compreender o que os gráficos significavam e como fazê-los” (RENAUX, 2017, p. 101).

Diante do exposto, verifica-se não só a importância dos OA na implementação de recursos tecnológicos para o estudo e a aprendizagem de objetos estatísticos, mas também sua relevância na problemática deste estudo.

Uma investigação sobre possíveis contribuições de uma estratégia didática com uso do *software* R⁷, para o ensino de estatística, foi desenvolvida por Santos (2019), partindo da conjectura de que esse recurso permitiria a mobilização de conhecimentos para a resolução de problemas envolvendo dados.

Santos (2019) recorreu a uma dimensão de níveis de mobilização de conhecimentos para a resolução de atividade: de conhecimentos técnicos sobre a resolução de um problema vinculado a uma técnica; de conhecimentos mobilizáveis nos quais existe um princípio de justaposição de saberes em dado domínio; de conhecimentos disponíveis, adquiridos em outras tarefas e mobilizáveis para a realização de novas tarefas.

⁷ R é um ambiente de *software* livre para computação estatística e gráficos. Disponível em: <https://www.r-project.org>. Acesso em: jun. 2023.

Em seu experimento, Santos utilizou os pressupostos da Engenharia Didática e os participantes foram alunos de graduação em medicina, enfermagem e matemática. O pesquisador esperava que os participantes, para mobilizar conhecimentos matemáticos e estatísticos, com a implementação do *software* R “contribuísse para a manipulação do banco de dados, por meio da visualização, dinamismo e experimentação” (SANTOS, 2019, p. 83).

Os resultados revelaram que os participantes apresentaram dificuldades em interpretar medidas descritivas, suas interpretações gráficas se mostraram parciais, da mesma forma que as representações tabulares, de modo que os conhecimentos mobilizados, essencialmente, foram técnicos.

A partir dos resultados de Santos (2019), embora a estratégia didática com a implementação do R tenha sido bem aceita pelos participantes, ele destaca: “não conseguimos concluir se contribuiu para a construção do letramento estatístico dos alunos, dado que tivemos indícios que os alunos atingiram apenas o nível técnico de conhecimento” (SANTOS, 2019, p. 118). Esse resultado revela que é possível propor uma atividade mediada por recursos tecnológicos, embora haja limitações, revelando a necessidade de reflexões e ajustes na proposição, de modo que questionamentos sobre os objetos estatísticos envolvidos possam ser compreendidos pelos participantes.

A pesquisa de Sera (2016) investigou o processo de ensino e aprendizagem da leitura e construção de gráficos estatísticos, com foco na formação de professores, e os resultados incluíram considerações relevantes sobre o letramento estatístico.

Com o aporte em *Design Experiments*, Sera (2016), inicialmente, constatou uma fragilidade dos participantes em relação à leitura e construção de gráficos estatísticos, durante a formação, propondo, portanto, a inserção de uma tecnologia digital, o Excel⁸, para as atividades realizadas na formação.

Além disso, verificou que os participantes não consideravam a possibilidade de utilizar um *software*, em sala de aula, para o ensino da leitura e construção de gráficos estatísticos, aspecto que se deve, segundo Sera (2016), à falta de habilidade em manipulá-los. Sera (2016, p. 175) observou que eles “não sabiam determinar tipos de gráficos adequados para representar variáveis específicas e, muito menos, se preocupavam com os

⁸ O Excel é um dos programas do pacote *Office* da empresa Microsoft, que permite a criação de planilhas, a organização de dados e a utilização de comandos (fórmulas) para realizar cálculos. Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/excel>. Acesso em: jun. 2023.

elementos que compõem um gráfico estatístico, tais como título, rótulo dos eixos vertical e horizontal ou linhas de grade”.

Em seu experimento, a introdução do uso de tecnologia digital “permitiu familiarização dos sujeitos, eles observaram que o uso de tecnologias digitais proporciona uma aula mais dinâmica e reflexiva, como “testar” diversas construções gráficas para um conjunto de dados” (SERA, 2016, p. 176). A autora verificou que os participantes passaram a discutir a respeito da melhor representação ou manipulação dos elementos da representação gráfica. Nesse sentido, considerou os avanços dos participantes, pois eles compreenderam as informações e discutiram sobre as conclusões e possíveis consequências em cada contexto abordado.

A implementação de um recurso tecnológico, como o Excel, pode oferecer meios para compreender e discutir resultados obtidos em uma atividade, o que comunga com as discussões propostas nesta tese.

O estudo de Souza (2019) concentrou-se na investigação da interpretação de representações gráficas de dados, em que Souza ressalta que o ensino de Estatística deve ser conduzido “a partir da realidade social e da valorização da cultura do estudante [...] para que haja o desenvolvimento das habilidades necessárias e seja favorecida a tomada de decisões diante das questões que se apresentam no contexto de vida dos sujeitos” (p. 15).

A autora afirma que “o letramento estatístico é considerado como uma competência de conotação abrangente, mantendo o raciocínio e o pensamento estatístico em seu termo, vistos como posteriores no desenvolvimento do cidadão estatisticamente competente” (SOUZA, 2019, p. 33). Na interpretação de gráficos, pontua que “há um ponto importante a considerar, que diz respeito à familiaridade do leitor com determinado tema e com a representação do gráfico e influencia as interpretações dele com relação aos dados” (SOUZA, 2019, p. 48).

Com uma organização metodológica qualitativa, a pesquisadora conduziu um estudo de natureza exploratória, no qual explorou as características dos participantes em cenários específicos. Suas análises eram realizadas a partir de materiais oriundos de oficinas realizadas com um grupo de professores, e um dos principais objetivos do estudo foi potencializar as reflexões para o desenvolvimento do letramento estatístico por meio das oficinas.

Souza (2019) identificou algumas fragilidades dos participantes, como incompreensões ao interpretar os gráficos estatísticos: “dificuldades e inseguranças na

mobilização dos elementos do conhecimento necessários ao letramento estatístico para a compreensão das mensagens estatísticas expostas nos gráficos” (SOUZA, 2019, p. 125). Além disso, observou que o ambiente de interação e trocas de experiências contribuiu para avanços em relação às fragilidades constatadas, inicialmente, sobre a interpretação de gráficos.

Os contributos da pesquisa de Souza (2019) à problemática desta tese situam nas fragilidades em relação à interpretação de gráficos e no ambiente de interação. Como o foco da pesquisadora esteve em professores, é possível que essas fragilidades estejam presentes neste estudo, de modo que é essencial pensar um ambiente apropriado como forma de minimizar tais fragilidades.

O estudo de Votto (2018) observou as atitudes e crenças na Educação Estatística e como elas podem refletir na resolução de problemas. Em relação ao escopo desta pesquisa, no levantamento de Votto (2018), apenas três trabalhos tiveram como objetivo desenvolver um estudo baseado nas tecnologias da informação e comunicação, dentro do contexto da Educação Estatística. “Esse fato mostra o quanto ainda precisamos evoluir em produções de conhecimentos sobre as TIC na Educação Estatística” (VOTTO, 2018, p. 36). TIC lê-se Tecnologias da Informação e Comunicação.

Em sua pesquisa, Votto identificou, a partir de um grupo de noventa e dois (92) professores que ensinam matemática, que 90% deles desenvolvem atividades relacionadas à Educação Estatística e que as estratégias utilizadas por eles “envolvem habilidades referentes a estatística, principalmente por meio de levantamentos e pesquisas de opinião em sala de aula, o que, por sua vez, possibilita a coleta e organização de dados por meio de construção de tabelas e gráficos” (VOTTO, 2018, p. 153).

A existência de poucos trabalhos que abordam a utilização das TIC na Educação Estatística, conforme Votto (2018), contribui e justifica a pertinência do estudo proposto por esta tese, que se dedica à inserção dos recursos digitais nesse contexto. Além disso, como a autora observa, a maior parte dos professores desenvolve, em suas aulas, atividades relativas à Educação Estatística.

Ao realizar uma leitura crítica dos trabalhos de pesquisadores de programas oriundos ou afins ao campo da educação matemática, no contexto brasileiro, a quarta fase da pesquisa bibliográfica (SOARES; PICOLLI; CASAGRANDE, 2018) apresentou estudos preliminares no contexto nacional para esta tese.

Os resultados dos estudos preliminares indicam a escassez de estudos para o uso de tecnologias no ensino de estatística, como destaca Votto (2018). Alguns estudos, como Sera (2016) e Renaux (2017), em elaboração de planejamentos de aulas, apresentam possíveis potencialidades do Excel e de objetos de aprendizagem, respectivamente, para promover a aprendizagem da Estatística, como o fez Santos (2019) ao agregar o R em aplicações com alunos. Ou, como apontou Pontes (2019), ao inserir recursos dessa natureza na prática educativa, é necessário realizar reflexões e possíveis adaptações da prática, levando em consideração os objetivos desejados, porque algumas dessas tecnologias não foram originalmente desenvolvidas com finalidade educativa.

As contribuições dos recursos tecnológicos para múltiplas representações de um conjunto de dados, permitidas pela inserção do GeoGebra, foram identificadas como importantes para o estudo das medidas de tendência central (ARAÚJO, 2018).

Os trabalhos levantados a partir do Catálogo de Teses e Dissertações da CAPES revelaram, em diferentes níveis, fragilidades em relação aos conhecimentos necessários para a construção e para a análise gráfica de dados, como nas pesquisas de Oliveira (2016) e Albuquerque (2018), nas quais problemas relacionados à análise gráfica de dados foram vistos como percalços da aprendizagem estatística.

No tópico a seguir, seguindo os procedimentos propostos por Soares, Picolli e Casagrande (2018) em um levantamento bibliográfico, estão estudos preliminares no contexto internacional e selecionados a partir dos anais da ICOTS.

3.2 Trabalhos disponíveis nos anais de 2018 e 2022 da ICOTS

As conferências denominadas ICOTS são promovidas pela *International Association for Statistical Education – IASE* e ocorrem a cada quatro anos. De acordo com o *website* IASE, esse é o evento mais relevante do calendário internacional da educação estatística e totaliza onze (11) edições.

Os estudos preliminares aqui apresentados se aportam nos anais das duas últimas edições do evento, delimitados aos trabalhos com o uso de recursos digitais para o ensino da Estatística, pois a conferência tem essa categorização para algumas pesquisas publicadas.

Uma versão preliminar do autor desta tese sobre o levantamento realizado na décima edição da ICOTS foi publicada no XXIV Encontro Brasileiro de Estudantes de Pós-Graduação em Educação Matemática, organizado de forma *online* em novembro de 2020.

As contribuições advindas das críticas recebidas durante o evento colaboraram para as análises dos estudos preliminares apresentados adiante. Além disso, nos anais da décima primeira edição, consta um trabalho resultante dos avanços realizados.

No levantamento dos trabalhos, ao iniciar a primeira fase da pesquisa bibliográfica, ao acessar o *site* da IASE, na página inicial é possível encontrar na aba de conferências⁹ a ICOTS e os *hiperlinks* para todas as suas edições, os quais permitiram acessar os anais das duas últimas edições, a ICOTS10 de 2018 e a ICOTS11 de 2022, realizadas em Kyoto, Japão, e em Rosário, Argentina, respectivamente.

Na base de dados da ICOTS, não foram utilizadas palavras-chave como descritores, pois há apenas publicações no âmbito da educação estatística. A estratégia foi realizar a leitura de todos os títulos das publicações, de modo a identificar as eventuais contribuições para a problemática desta tese.

Decorrente do levantamento nos anais das duas últimas edições da ICOTS (2018 e 2022), foram identificadas vinte e nove (29) publicações, e o critério de seleção adotado foi a inclusão de estudos que mencionassem a utilização de recursos tecnológicos, especialmente *softwares* específicos, para promover o ensino ou a aprendizagem da Estatística sob diferentes abordagens, finalizando a primeira fase da pesquisa bibliográfica.

Na segunda fase, procedeu-se à leitura dos resumos das vinte e nove publicações identificadas, resultando na seleção de quatorze delas para uma leitura seletiva mais aprofundada, a terceira fase da pesquisa bibliográfica. A leitura na íntegra delimitou o conjunto de doze publicações.

A quarta fase da pesquisa bibliográfica compreendeu a análise dos trabalhos apresentados no Quadro 2. Nos anais da ICOTS, foram encontrados alguns resultados de aplicações e proposições teóricas para atender à problemática desta tese.

Quadro 2 – Publicações da ICOTS10 e da ICOTS11

AUTOR, ANO	TÍTULO	TRADUÇÃO DO TÍTULO
BIEHLER, 2018	<i>Design principles, realizations and uses of software supporting the learning and the doing of statistics – A reflection on developments since the late 1990s</i>	Princípios de projetos, realizações e usos de <i>software</i> que apoiam o aprendizado e o fazer estatístico – Uma reflexão sobre os desenvolvimentos desde o final dos anos 1990

⁹Disponível em: <https://iase-web.org/Conferences.php>. Acesso em: jun. 2023.

MCNAMARA, 2018	<i>Imagining the future of statistical education software</i>	Imaginando o futuro de <i>software</i> de educação estatística
HABIBULLAH, 2018	<i>A Workshop in Pakistan on Teaching Statistical Inference Through a Freely Available Online Software</i>	Um <i>workshop</i> no Paquistão sobre o ensino de inferência estatística por meio de um <i>software</i> online disponível gratuitamente
MANTILLA, 2018	<i>Using ICT's for to Introduce to Statistic</i>	Usando TIC para introduzir a estatística
KRAFT; CARVER, 2018	<i>Keeping it Real with Data Visualization</i>	Observando a realidade com a visualização de dados
BOLCH; JACOBBE, 2018	<i>Students' Understanding of Data Visualizations</i>	Compreensão dos alunos sobre visualizações de dados
WOODARD, 2018	<i>Defining the relationship between statistical thinking and statistical computing</i>	Definindo a relação entre pensamento estatístico e computação estatística
FUJII, 2018	<i>Current situation and issue for effective use of statistical software in Japan</i>	Situação atual e questão para uso efetivo de <i>software</i> estatístico no Japão
ECKERT, 2018	<i>Implementing research results in technology induced teaching of statistics - A literature review</i>	Implementação de resultados de pesquisa em tecnologia induzida ao ensino de estatística - Uma revisão da literatura
CABRERA; TAUBER; GILI; ROMERO, 2022	<i>La Transnumeración como insumo del pensamiento crítico</i>	A transnumeração como insumo do pensamento crítico
OGBONNAYA, 2022	<i>Teaching statistics with MATLAB: how and why?</i>	Ensinar estatística com o MATLAB: como e por quê?
SANTOS; OLIVEIRA JÚNIOR; SANTOS; CARVALHO; SAITO, 2022	<i>The use of technological resources in teacher education as a contribution to the teaching of statistics and probability in Brazil</i>	O uso de recursos tecnológicos na formação de professores como contribuição ao ensino de estatística e probabilidade no Brasil

Fonte: Dados da pesquisa.

As primeiras considerações são baseadas no texto de Biehler (2018), que discute o surgimento de *software* para o fazer e para o ensino da estatística. De acordo com Biehler (2018), se quisermos implementar novas práticas de ensino, devemos considerar a introdução de ferramentas computacionais, visto que a tecnologia computacional tem potencial para auxiliar a aprendizagem ativa por visualizações interativas, por ambientes para experimentação com métodos e pelo uso de simulações.

O autor ressalta como uma ferramenta tecnológica utilizada para o fazer estatístico também poderia auxiliar na aprendizagem. Que recursos uma ferramenta computacional

deveria ter para apoiar a aprendizagem e a implementação de novas práticas em estatística e probabilidade? (BIEHLER, 2018).

Sobre essa problemática, destaca que, desde a década de 1980, diferentes recursos tecnológicos foram desenvolvidos para auxiliar a prática estatística, como o JMP¹⁰ e o R, que recentemente têm sido utilizados com finalidades educacionais. No entanto, sob aporte de um estudo anterior de sua autoria, assevera que essas ferramentas não cumpriam vários requisitos para serem úteis para a Educação Estatística, como: suporte à experimentação com métodos estatísticos, modelagem com modelos de probabilidade simples e discussões necessárias na introdução do estudo da estatística. Biehler (2018) observou que um dos principais motivos para isso é que “a visão atual de estatística e probabilidade e seu ensino, que está subjacente a essas ferramentas, ainda não se tornou um padrão no ensino médio e superior em qualquer país” (p. 2, tradução do autor¹¹).

Biehler (2018) destaca que ferramentas gerais como GeoGebra são frequentemente preferidas, pois não restringem seus usos apenas à estatística e probabilidade. No entanto, para o autor, essas ferramentas nem sempre possibilitam uma visão suficiente da estatística e probabilidade. Além disso, há a necessidade de domínio no uso desses recursos tecnológicos e “o tempo de aprendizagem dedicado à probabilidade e estatística é muito pequeno para tal abordagem” (BIEHLER, 2018, p. 2, tradução do autor¹²).

Uma contribuição de Biehler (2018) para a problemática desta tese é que algumas ferramentas tecnológicas possuem limitações que nem sempre permitem uma visão abrangente da estatística, exigindo, assim, a criatividade do pesquisador para explorar o tema de forma mais aprimorada.

No estudo de McNamara (2018) e no questionamento sobre o que deveria estar presente em recursos tecnológicos para que avançassem nessa direção, há caminhos para essa discussão e a autora pontua alguns atributos que seriam pertinentes em um recurso tecnológico e que se submetam à promoção da Educação Estatística, tais como:

- Acessibilidade
- Entrada fácil para novos usuários
- Dados como um objeto persistente de primeira ordem
- Suporte para um ciclo de análises exploratórias e confirmatórias
- Criação de plotagem flexível
- Suporte para randomização completa

¹⁰ *Software* de análise de dados usado por exploradores de dados em todo o mundo. Interativo e visual, o JMP revela perspectivas que tabelas numéricas primárias ou gráficos estáticos tendem a ocultar”

¹¹ “[...] the new vision of statistics and probability and its teaching that is underlying these tools has not yet become a mainstream at high school and college level in any country”.

¹² “[...] that the learning time devoted to probability and statistics is too small for such an approach”.

- Interatividade em todos os níveis
- Documentação inerente
- Suporte simples para narrativa, publicação e reprodutibilidade
- Flexibilidade para construir extensões (McNAMARA, 2018, p. 1, tradução do autor¹³).

Ao questionar para onde estamos indo com *software* de Educação Estatística, McNamara (2018) sugere que os atributos podem servir como um guia, uma vez que não existem atualmente pacotes de *software* que possuam todos os atributos delineados, de modo que imaginar uma ferramenta que preencha todos os atributos é um exercício a ser considerado. Por exemplo, aspectos de linguagens de programação podem ser combinados com ferramentas visualização e análise de dados (McNAMARA, 2018).

McNamara (2018) salienta que, ao longo de todas essas explorações, é necessário ter em mente as pessoas, como elas visualizam os dados e como analisam dados em suas mentes. Sob o ponto de vista da viabilidade da tecnologia a ser utilizada nesse processo, afirma que essas ferramentas tecnológicas para educação em estatística podem fazer mais do que ampliar do autores habilidades, o que contribui para a problemática desta tese.

O estudo de Habibullah (2018) problematizou que “a capacidade de gerar números aleatórios através de computadores levou a simulações em larga escala que eram *impensáveis*” (HABIBULLAH, 2018, p. 1, tradução do autor¹⁴), abrindo caminhos para outras abordagens no ensino da estatística. Sob essa problemática, Habibullah (2018) propôs, durante um *workshop* com alunos e professores universitários, dezesseis participantes no total, de áreas diversas, a realização de simulações com o StatKey¹⁵, destacando que “[...] um *software* capaz de gerar simulações é uma ferramenta excelente para elucidar resultados matemáticos” (HABIBULLAH, 2018, p. 4, tradução do autor¹⁶), de modo que a exposição

¹³ • Accessibility
 • Easy entry for novice users
 • Data as a first-order persistent object
 • Support for a cycle of exploratory and confirmatory analysis
 • Flexible plot creation
 • Support for randomization throughout
 • Interactivity at every level
 • Inherent documentation
 • Simple support for narrative, publishing, and reproducibility
 • Flexibility to build extensions

¹⁴ “[...] the ability to generate random numbers through computers has led to large-scale simulations that were *unthinkable*”.

¹⁵ Disponível em: <http://www.lock5stat.com/statkey>. Acesso em: jun 2023.

¹⁶ “[...] that software capable of generating simulations is a wonderful tool for elucidating mathematical results”.

a essas simulações pode possibilitar “um mundo totalmente novo” em relação à maneira de aprender e ensinar alguns conceitos fundamentais, por exemplo, de inferência estatística.

Juntamente com os participantes de sua pesquisa, Habibullah observou que a possibilidade de realizar simulações pode ser um aspecto favorável no uso de recursos tecnológicos para o ensino de estatística, o que contribui para a problemática desta pesquisa.

Mantilla (2018) considera que as TIC desempenham um papel significativo para o ensino da estatística, especialmente, porque “[...] permitem ao aluno explorar, conhecer e analisar dados de forma dinâmica e colaborativa” (p. 1, tradução do autor¹⁷).

Essa perspectiva emergiu para Mantilla (2018) ao desenvolver um estudo piloto de introdução à estatística com estudantes de Ciências Humanas, Engenharias e Economia de uma universidade colombiana, sob a conjectura de que adventos das TIC, os canais de informação, como fóruns e debates online, trabalho colaborativo e os aplicativos interativos facilitam a interação dos atores do processo.

Mantilla (2018) se referiu a *softwares* que permitem visualizações especiais, como o Power BI¹⁸, no estudo de conceitos estatísticos e nas construções de representações interativas, forma dinâmica e interativa que pode favorecer o ensino da Estatística, como explorado nesta tese.

Ao abordar as vantagens da visualização de dados, Carver e Kraf (2018) propõem uma discussão baseada nos estágios de um cenário de solução de problemas, começando por: integrar dados reais em um contexto da realidade e propósito; focar na compreensão conceitual e usar tecnologia para explorar conceitos e analisar dados.

Em relação ao último estágio, usar tecnologia para explorar conceitos e analisar dados, Carver e Kraf (2018) salientam que a tecnologia deve integrar a visualização com procedimentos analíticos. “Combinar tecnologia adequada com dados interessantes e problemas do mundo real também pode fornecer um “roteiro para a felicidade” no ensino do pensamento estatístico, tanto para o professor quanto para os alunos” (p. 4, tradução do autor¹⁹), o que vai ao encontro e contribui com a problemática desta tese.

¹⁷ “[...] it allows the student to explore, know and analyze data in a dynamic and collaborative way”.

¹⁸ Ferramenta que unifica dados de várias fontes para criar painéis e relatórios interativos e envolventes que fornecem percepções e geram resultados. Disponível em: <https://powerbi.microsoft.com/es-es/>. Acesso em: jun 2023.

¹⁹ “Combining adequate technology with exciting data and real-world problems can also provide a “roadmap to happiness” in teaching statistical thinking, both for the teacher and the students”.

No estudo publicado por Bolch e Jacobbe (2018), os autores destacam que os alunos podem passar por um curso inteiro de estatística e serem ensinados a fazer cálculos estatísticos, mas ainda assim não desenvolverem o pensamento estatístico necessário para abordar uma questão ou um problema. Nesse contexto, ressaltam que essa competência pode advir de proposições de organizações didáticas que propiciem esse desenvolvimento. Por exemplo “[...] para entender e desenvolver essas habilidades de visualização de dados, os alunos precisam primeiro desenvolver uma base de habilidades de letramento estatístico” (BOLCH; JACOBBE, 2018, p. 1, tradução do autor²⁰).

Bolch e Jacobbe (2018) propõem um estudo cujo objetivo é investigar o impacto das experiências dos alunos na aprendizagem em relação à leitura, interpretação e criação de visualizações de dados com o uso de tecnologia. Nesse caso, consideram a inserção de um *software* estatístico dinâmico e as habilidades dos participantes de compreensão de gráficos e atitudes em relação à Estatística.

Os problemas propostos por Bolch e Jacobbe (2018) aos participantes se ancoram em dados do mundo real, abordagem considerada propícia aos seus objetivos, pois “[...] requer que os alunos apliquem habilidades de pensamento crítico para responder questões de pesquisa usando esses dados” (BOLCH; JACOBBE, 2018, p. 2, tradução do autor²¹).

Os resultados da implementação de *software* estatístico dinâmico consideraram cinco pontos essenciais: a computação, a análise, a síntese, a apresentação visual e a comunicação escrita, perspectiva que trouxe avanços, como escolha e exibição gráfica corretas, identificação dos principais recursos, comunicação de suas descobertas com clareza e de forma concisa, por meio de apresentação visual e comunicação escrita (BOLCH; JACOBBE, 2018).

Para esta tese, os pontos destacados por Bolch e Jacobbe (2018) podem ser vistos como relevantes, ao proporem a inserção de tecnologias para atividades no campo da educação estatística, em atividades relativas à análise gráfica de dados.

O estudo de Woodard (2018) salienta que há poucas pesquisas avaliando como os recursos computacionais para a Estatística afetam as habilidades dos alunos para pensar e raciocinar estatisticamente; “[...] como os alunos usam as ferramentas para resolver problemas estatísticos” (p. 1, tradução do autor²²). Sob esse prisma, o objetivo de Woodard

²⁰ “[...] To understand and develop these data visualization skills, students need to first develop a foundation of statistical literacy skills”.

²¹ “[...] and required students to apply critical thinking skills to answer research questions using this data”.

²² “[...] how students use the tools to solve statistical problems”.

(2018) foi buscar introduzir definições para as relações entre as habilidades de pensar e raciocinar estatisticamente com o uso dos recursos computacionais, baseando-se em investigações disponíveis a respeito do pensamento estatístico e *frameworks* de computação estatística²³.

Os resultados da pesquisa da autora foram identificados a partir de registros escritos dos participantes em que ela destacou cinco aspectos:

- A produção escrita dos alunos nem sempre fornece evidências de computação estatística;
- Os argumentos dos alunos usam computação estatística para fornecer fatos, não implicações;
- Os argumentos escritos dos alunos podem demonstrar pensamento estatístico e computação estatística corretamente;
- Os alunos às vezes aplicam métodos ou têm pensamento computacional incorreto em suas evidências para argumentos;
- A aplicação correta da computação estatística nem sempre leva ao pensamento estatístico devido a fragilidades do trabalho feito pelos alunos (WOODARD, 2018).

Woodard (2018) destaca que esses aspectos permitem uma introdução das ideias de como o pensamento estatístico e os recursos tecnológicos utilizados pela estatística funcionam juntos, mas foi nas entrevistas com os alunos que a pesquisadora identificou uma perspectiva diferente sobre como eles utilizaram os recursos tecnológicos para resolver problemas envolvendo estatística:

- usaram o *software* para automatizar procedimentos computacionais;
- criaram uma estratégia de resolução de problemas quando não tinham certeza de como programar;
- passaram muito tempo pesquisando como resolver um problema.

Diante dos resultados, Woodard (2018) salienta que nos registros escritos relacionados à computação estatística a automação para procedimentos computacionais foi o recurso mais evidente, seguido do pensamento computacional. Mas, ao avaliar os resultados da entrevista, o pensamento computacional foi o recurso mais acessado em relação à computação estatística, enquanto a automação foi a menos citada.

²³ *Frameworks* de computação estatística são aqueles que se submetem ao estudo da inserção dos recursos computacionais na própria estatística.

Em suas considerações, Woodard (2018) observa que os alunos demonstram bom domínio do pensamento estatístico e das tecnologias voltadas para a Estatística, apresentando argumentos sólidos para problemas estatísticos. Além disso, as falhas relacionadas à utilização das tecnologias não implicam necessariamente que os alunos sejam incapazes de pensar estatisticamente um dado problema, visto que “mesmo com pequenas fragilidades em seu planejamento ou implementação, eles ainda podem pensar estatisticamente sobre o problema e sobre os resultados que encontram” (WOODARD, 2018, p. 5, tradução do autor²⁴).

Ao implementar uma abordagem em que se espera que os alunos aprendam estatística e técnicas computacionais, simultaneamente, para resolver problemas, é possível que eles enfrentem dificuldades. “Os estudantes que tinham experiência com o uso das tecnologias ou algum domínio estatístico, foram capazes de demonstrar um método para resolver os problemas estatísticos utilizando a tecnologia, depois da formação” (WOODARD, 2018, p. 6, tradução do autor²⁵).

As proposições de Woodard (2018) apontam que há uma relação entre o uso de recursos tecnológicos e o pensamento estatístico necessária para a resolução de problemas, possibilitando identificar quais contribuições advêm de seus usos, um dos objetivos desta tese.

Diante da importância de empreender pesquisas que investigam as contribuições dos recursos tecnológicos para a Educação Estatística, Fujii (2018) observou que no Japão a inserção de tais recursos para o ensino da Estatística na educação básica tem se mostrado carente. Para Fujii (2018), ao menos três razões justificam esse descompasso:

- 1) O uso de recursos tecnológicos não é um objetivo da Educação Estatística, mas uma ferramenta para a educação. Nesse sentido, os professores japoneses consideram importante ter controle sobre como os alunos usam esses recursos nas aulas.
- 2) São necessárias algumas evidências de que o uso dos recursos tecnológicos traga resultados positivos para a educação, e os professores devem ter objetivos claros para utilizá-los.
- 3) Os professores precisam ter tecnologias adequadas que correspondam à intenção do trabalho docente (FUJII, 2018).

²⁴ “Even with minor flaws in their planning, or implementing, then they can still think statistically about the problem and the results they found”.

²⁵ “Students that had strong statistical or computational backgrounds prior to the course were able to demonstrate a method to solve the statistical problems while using technology after the course was over”.

Fujii (2018) contribui para esta tese ao apontar as dificuldades identificadas em seu estudo para a implementação de recursos tecnológicos em aulas de Estatística. Além dessas constatações, o autor salienta o fato de que a promoção de ambientes favoráveis à implementação de tecnologias depende do currículo das escolas, observando a questão do tempo que os professores não dispõem para o ensino de Estatística.

Eckert (2018), em seu estudo, parte da premissa de que, “à medida que a tecnologia digital se torna mais ou menos uma parte natural da vida cotidiana, a pesquisa em educação continuada fornece *insights* sobre como isso pode ser utilizado em sala de aula para aprimorar o aprendizado dos alunos” (ECKERT, 2018, p. 1, tradução do autor²⁶), e, buscando na literatura, aponta que há uma notável coerência entre o papel do professor e o papel da tecnologia. Pelo menos dois aspectos são destacados:

Quando a tecnologia estava conceituada como uma ferramenta de provocação à reflexão, o professor foi evocado de uma forma mais retraída, com papel de um promotor ou organizador;

Quando a tecnologia foi usada como uma ferramenta de simulação para os alunos ganharem experiências de processos e conceitos estatísticos, o professor foi visto como um ator de interação com os alunos (ECKERT, 2018, p. 3, tradução do autor²⁷).

Sobre os aspectos identificados por Eckert (2018), acerca da tecnologia digital tornar-se, em maior ou menor grau, uma parte natural da vida cotidiana, a depender de como ela é inserida, este estudo revela diferentes oportunidades para o professor atuar na Educação Estatística, o que se pretende implementar nesta tese.

O estudo de Cabrera et al. (2022) teve como foco a compreensão e a interpretação de gráficos por alunos de um curso de graduação em medicina de uma universidade argentina, a fim de investigar a implicação da transnumeração no desenvolvimento do pensamento crítico, e, a partir da inserção do *website* do *Observatório de la Deuda Social de Argentina* (ODSA) (<https://observatorio.unr.edu.ar/observatorio-de-la-deuda-social-argentina-uca/>), os autores propuseram uma atividade sobre dados relativos à pobreza na Argentina.

Sobre o acesso às representações gráficas dos dados, “os alunos possuíam o relatório ODSA em formato digital, ou seja, podiam acessar as informações completas quantas vezes

²⁶ “As digital technology becomes more or less a natural part of everyday life ongoing educational research provides insights into how it can be utilized in the classroom to enhance students' learning”.

²⁷ Where the technology was conceptualized as a tool for provoking reflection, the teacher was thought of in a more withdrawn role of a promotor or organizer.

When technology was used as a simulation tool for students to gain experiences of statistical processes and concepts, the teacher was thought of as an actor in interaction with the students.

precisassem para responder aos questionamentos propostos pela tarefa” (CABRERA et al., 2022, p. 3, tradução do autor²⁸). Os autores analisaram, por meio de uma metodologia interpretativa baseada no método comparativo constante, as respostas de quarenta e três alunos, sob os seguintes critérios: identificar e definir variáveis contendo os dados do gráfico proposto; tirar conclusões lógicas a partir dos dados apresentados no gráfico proposto; discutir e comunicar seus pontos de vista sobre as implicações das informações estatísticas extraídas no gráfico ou suas preocupações sobre a aceitabilidade de certas conclusões.

Uma das formas que os dados são representados pelo ODSA, conforme os autores, é a “matriz multidimensional de pobreza”, que representa dados relativos a recursos econômicos básicos, como alimentação e saúde, moradia digna, acesso à educação, acesso a serviços públicos, acesso a um ambiente saudável, emprego e segurança social. Para Cabrera et al. (2022, p. 4), essa forma de representação “dá conta da ocorrência de um processo de transnumeração em que se altera de uma representação para outra, com a intenção de gerar maior compreensão dos dados e comunicá-los com clareza”. Sob o ponto de vista da análise, “para aqueles que interpretam e buscam compreender o significado dessas informações, o caminho da transnumeração ocorre na direção oposta, ou seja, a mudança de representações requer ser decodificada e recolocada no contexto dos dados” (CABRERA et al., 2022, p. 4, tradução do autor²⁹).

Os resultados observados pelos pesquisadores apontaram que 60% dos participantes tiveram resultados insuficientes em relação à identificação e definição das variáveis que contêm os dados. A elaboração de conclusões lógicas, a partir dos dados representados graficamente, foi observada em 37% dos participantes, que conseguiram apresentar resultados aceitáveis, enquanto 58% das respostas se mostraram insuficientes em relação à discussão e à comunicação de pontos de vista sobre as implicações das informações estatísticas extraídas no gráfico.

Sobre essas fragilidades, os autores ressaltam que:

Vale notar aqui que esses alunos apresentaram dificuldades em decodificar, interpretar e ressignificar a complexa mensuração do construto – pobreza multidimensional – que não é diretamente observável. Ou seja, essas mudanças de representações nos dados, dadas na operacionalização dessa variável, resultaram

²⁸ Los estudiantes disponían del informe de ODSA en formato digital, esto es, podían acceder a la información completa las veces que lo necesitaran para responder a las cuestiones e interrogantes que les proponía la tarea.

²⁹ “para quien interpreta y busca comprender el sentido de dicha información, el camino de la transnumeración se dá en sentido inverso, esto es, el cambio de representaciones requiere ser decodificado y puesto nuevamente en el contexto de los datos”.

em um obstáculo cognitivo para esses estudantes (CABRERA et al., 2022, p. 4, tradução do autor³⁰).

Como forma de lidar com essas fragilidades dos alunos, os pesquisadores destacam a relevância de propor uma atividade emergencial, na qual, por meio de um processo de reflexão-ação-reflexão, com base na tarefa inicial, a atividade busca recontextualizar, identificar e definir as variáveis representadas graficamente, analisando a correspondência de informações significativas, a partir de um olhar crítico sobre essa informação estatística (CABRERA et al., 2022).

Com relação à proposta do estudo, os autores consideram que:

A transnumeração, um dos tipos fundamentais do pensamento estatístico, quando interfere em contextos sociopolíticos complexos, como o problema da mensuração da pobreza e, em particular, o construto, pobreza multidimensional, possibilita um cenário educacional no qual os alunos desenvolvem primeiro a capacidade de identificar informações relevantes, estabelece a intenção e a razoabilidade dos argumentos baseados em dados e, assim, estejam cientes de como eles afetam a objetividade dos julgamentos (CABRERA et al., 2022, p. 4, tradução do autor³¹).

Ressaltam, ainda, que “é preciso entender que não basta desenvolver estatísticas aplicadas no contexto, se esse contexto não entrar na interpelação do trabalho profissional e, ao mesmo tempo, nas crises, nos conflitos do mundo em que vivemos” (CABRERA et al., 2022, p. 4, tradução do autor³²).

O percurso realizado por Cabrera et al. (2022), ao utilizar um *website* para acessar representações gráficas de dados, contribui para esta tese e pode ser visto como um fator que viabiliza o processo de proposição de uma atividade. As fragilidades dos participantes quanto à interpretação de dados representados por meio de gráficos estatísticos no *website* da ODSA revelam a necessidade de o aprendiz conhecer o contexto, a origem dos dados que

³⁰ Vale señalar aquí, que estos estudiantes evidenciaron dificultades para decodificar, interpretar y resignificar la compleja medición del constructo—pobreza multidimensional—que no es observable de manera directa. Dicho de otro modo, estos cambios de representaciones en los datos, dado en la operacionalización de esta variable, resultó un obstáculo cognoscitivo para estos estudiantes que fue abordado a través de la tarea emergente (Figura 4) en la secuencia de clases.

³¹ La transnumeración, uno de los tipos fundamentales de pensamiento estadístico, cuando se inmiscuye en contextos socio-políticos complejos, como lo es la problemática de la medición de la pobreza y en particular del constructo, pobreza multidimensional, posibilita un escenario educativo en el que los estudiantes desarrollen habilidad para identificar información relevante, establecer la intención y razonabilidad de los argumentos basados en datos y con ello, ser conscientes de cómo éstos afectan la objetividad de los juicios.

³² “se precisa comprender que no alcanza con el desarrollo de una estadística aplicada en el contexto, si ese contexto no se adentra en la interpelación del quehacer profesional y a la vez, en las crisis, los conflictos del mundo en el que vivimos; sólo así será posible el desarrollo de una AEC para leer y escribir el mundo con la lente de la Estadística Crítica”.

pretende estudar. Como os autores consideram, a transnumeração pode ser observada como pertinente nas análises do contexto dos problemas abordados.

O estudo de Ogbonnaya (2022) discute recursos e potencialidades do MATLAB para o ensino de Estatística e análise de dados. “O MATLAB é uma linguagem de programação orientada a objetos que é comumente usada na academia, bem como na indústria para computação científica, simulação e análise de dados (OGBONNAYA, 2022, p. 1).

Ogbonnaya (2022) considera que o MATLAB permite a análise de dados, mas essa potencialidade é frequentemente ignorada. Diante disso, propõe a implementação introdutória do MATLAB para análise de dados e destaca algumas potencialidades, como:

Flexibilidade. O MATLAB oferece a flexibilidade de escrever comandos; portanto, um usuário pode facilmente criar suas próprias funções para executar uma análise que não está disponível.

Capacidade trabalhar álgebra linear. Devido aos seus operadores matemáticos intuitivos, o MATLAB é ótimo para álgebra linear e permite operações em vetores e matrizes, o que é útil no ajuste de modelos complexos.

Interface gráfica do usuário (GUI). Em comparação com linguagens de programação semelhantes, o MATLAB tem uma interface gráfica amigável e pode ser bastante fácil de se locomover (OGBONNAYA, 2022, p. 2, tradução do autor³³).

Além de apontar recursos presentes no MATLAB, Ogbonnaya (2022) oferece algumas comparações desse *software* com outros, como o SPSS e o R, de modo a justificar a viabilidade da análise de dados mediada pelo MATLAB. Ao tratar do ensino de Estatística, Ogbonnaya (2022) destaca algumas ferramentas dessa linguagem de programação:

Demonstração de dados reais. O uso de dados reais no ensino permite que os alunos resolvam problemas reais e entendam o motivo de cada análise [...]; *Compartilhamento de código.* Ensinar ativamente enquanto compartilha código permite que os alunos se atualizem com os materiais do curso e participem da aula [...]; *Exercícios ao vivo.* Ter exercícios para trabalhar ajuda os alunos a praticarem o que aprenderam e fornece a garantia de que eles estão confortáveis escrevendo seus próprios comandos para responder a perguntas específicas [...]; *Avaliador MATLAB.* Esta é uma ferramenta útil para exercícios assíncronos[...] (p. 3-4, tradução do autor³⁴).

³³ *Flexibility.* MATLAB gives the flexibility of writing commands; therefore, a user can easily create their own functions to perform an analysis that is unavailable.

Linear algebra capabilities. Due to its intuitive mathematical operators, MATLAB is great for linear algebra and allows for operations on vectors and matrices, which is useful when fitting complex models. *Graphical user interface (GUI).* Compared to similar programming languages, MATLAB has a friendly graphical user interface and can be quite easy to get Around.

³⁴*Real data demonstration.* Using real data in teaching enables learners to solve real problems and understand the reason for each analysis [...]; *Code sharing.* Actively teaching while sharing code allows learners catch up with course materials and participate in the classroom; [...]; *Live exercises.* Having exercises to work through

Sobre essas potencialidades do MATLAB, as considerações de Ogbonnaya (2022) apontam para sua capacidade de mediar o ensino de Estatística, abrangendo desde análises estatísticas básicas até análises mais avançadas, como aprendizagem de máquina. O uso do MATLAB como uma ferramenta de avaliação também é destacado, possibilitando a criação de atividades assíncronas e facilitando as atividades não presenciais.

À medida que Ogbonnaya (2022) discute as possíveis contribuições da implementação do MATLAB em atividades de análise de dados, seus apontamentos favorecem a problemática desta tese, ao ressaltar que há a possibilidade de ensino mediado pelo uso de um recurso tecnológico.

O estudo de Santos et al. (2022) considera que há uma abundância de ferramentas tecnológicas educacionais para o ensino da Estatística e faz uma revisão sobre os trabalhos publicados nos anais do Encontro Nacional de Educação Matemática – ENEM. Conforme destacam, no evento há um tema para discutir o uso de recursos tecnológicos na formação inicial e continuada de professores no âmbito da Educação Estatística. O foco dos autores, nesse estudo, foi identificar eventuais tendências na evolução da pesquisa e as fragilidades da produção científica.

Ao analisarem os anais das edições do evento, entre 1987 e 2019, Santos et al. (2022) identificaram quinze trabalhos e destacaram que eles tratam de pesquisas que envolvem quatro tipos de tecnologias: objetos de aprendizagem, *software*, ambientes virtuais de aprendizagem e linguagem de programação.

Quanto à formação para o ensino de Estatística, os resultados apontam:

O foco são conteúdos estatísticos desde a educação infantil e os anos iniciais do ensino fundamental (licenciatura em pedagogia), passando pelos anos finais do ensino fundamental e médio (licenciatura em matemática), até o ensino superior (formação inicial de professores), com apenas um trabalho focado na análise combinatória (SANTOS et al., 2022, p. 4, tradução do autor³⁵).

helps learners practice what they have learned and provides assurance that they are comfortable writing their own commands to answer specific questions. *MATLAB grader*. This is a useful tool for asynchronous exercises.
³⁵ The focus is on statistics content from early childhood education and the early years of elementary school (pedagogy degree), through the final years of elementary and high school (mathematics degree), to higher education (initial teacher training), with only one work focusing on combinatorial analysis. Thus, we realize that there is a need to develop more works aimed at teaching probability and combinatorics or aspects aimed at stochastics by considering approaches that associate statistical and/or probabilistic and/or combinatorial contents.

Como os autores salientam, há, nos diversos níveis de formação de professores, resultados de pesquisas que envolvem objetos estatísticos, de modo que, “sem dúvida, a formação de professores precisa ser atenta para que os professores tenham a oportunidade de adquirir esses conhecimentos” (SANTOS et al., 2022, p. 4).

O estudo de Santos et al. (2022, p. 4) revela que “a apresentação de trabalhos desenvolvidos com suporte tecnológico no ENEM ainda é incipiente, tendo como suporte os softwares que podem ser utilizados e/ou classificados como educacionais (Excel, BrOffice Calc, R, SPSS, etc.)³⁶”. Sobre como esses recursos poderiam ser implementados, Santos et al. (2022, p. 5) sugerem que “sejam utilizados nas aulas como alternativas para priorizar o raciocínio, a compreensão dos processos de análise de dados e dos conceitos que os permeiam”. Para os autores, esse poderia ser o diferencial dos recursos tecnológicos quando comparados com alternativas didáticas.

Na direção da problemática desta tese, os resultados do estudo de Santos et al. (2022) apresentam contribuições relevantes, ao revisar as pesquisas publicadas no ENEM que tratam da inserção de recursos tecnológicos na formação de professores e mostrarem que há, ao longo da formação, oportunidades de estudo da Estatística. Além disso, a carência de estudos que tratam dos recursos tecnológicos no âmbito dessa formação justifica a pertinência da problemática desta tese.

Como o objetivo deste capítulo foi apresentar uma discussão relacionada à problemática desta tese, os estudos que emergiram da pesquisa bibliográfica na última fase permitem uma compreensão das possibilidades de inserção de recursos tecnológicos para a Educação Estatística. Os resultados obtidos a partir dos estudos preliminares mostram que ainda existe a necessidade de que tais recursos estejam alinhados com as problemáticas latentes à Educação Estatística.

Esta tese, ao problematizar a inserção de recursos digitais em atividades de análise gráfica de dados, considera contribuir nessa direção. No próximo capítulo, são apresentados os referenciais teóricos.

³⁶ “the presentation of works developed with technological support in ENEM is still incipient, having as support the software that can be used and/or classified as educational (Excel, BrOffice Calc, R, SPSS, etc.)”.

4 REFERENCIAIS TEÓRICOS

Neste capítulo, são apresentadas as bases teóricas e metodológicas que subsidiaram a aplicação e as análises dos resultados de um experimento que investigou como a utilização dos recursos *Gapminder*, GeoGebra e SPSS contribuem para a análise gráfica de dados. A proposta desta tese concebe esses recursos digitais como potenciais promissores para fomentar a Educação Estatística.

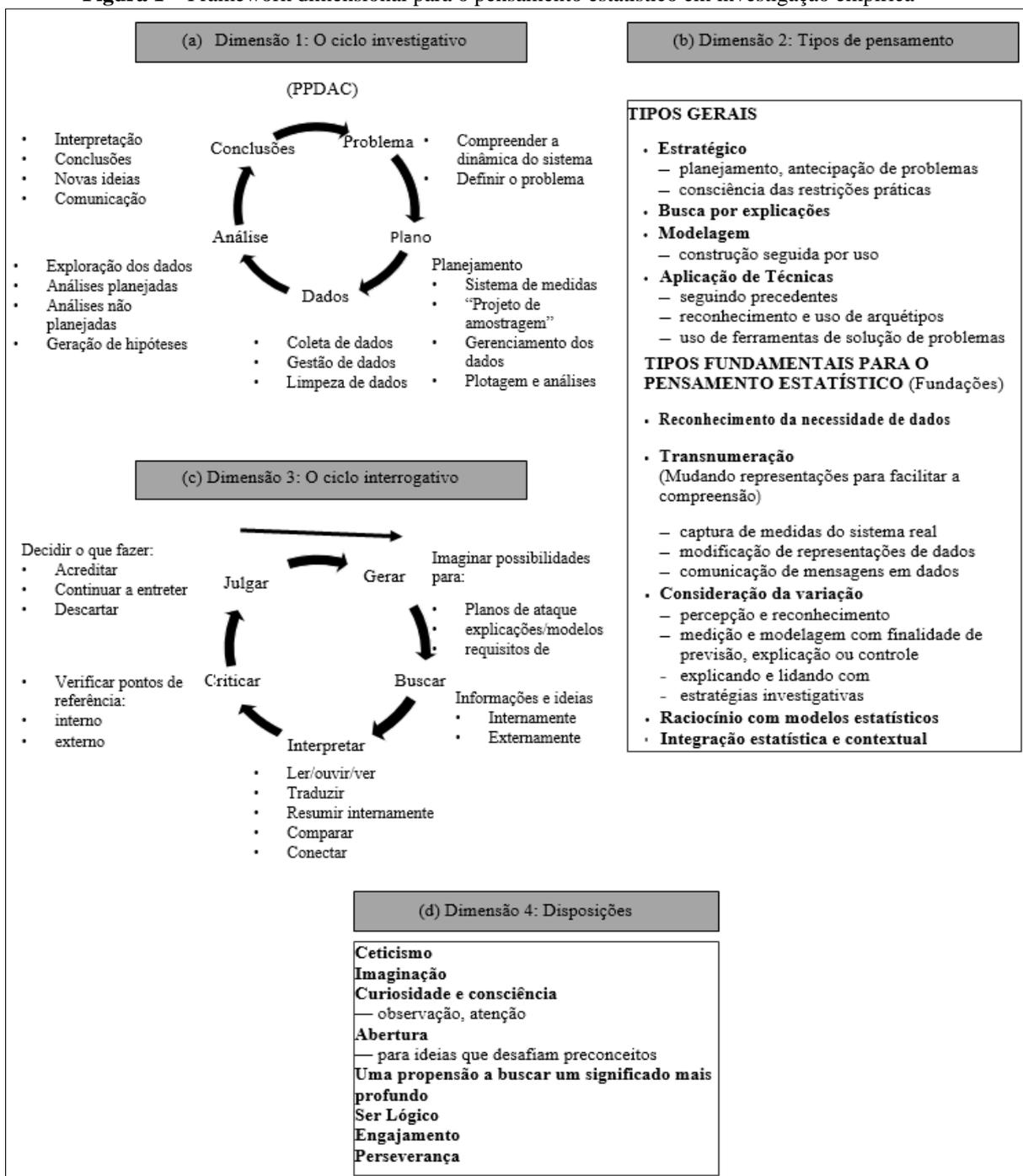
Fazem parte dos referenciais teóricos a proposição de Prodromou e Dunne (2017) sobre a Visualização de Dados, visto que esses autores consideram que ela permite explorar e comunicar efetivamente informações relevantes sobre grandes volumes de dados por meio de representações gráficas. Além disso, o conceito de transnumeração, discutido por Wild e Pfannkuch (1999), também foi adotado, e é definido por eles como “um processo dinâmico de mudança de representações para gerar compreensão” (WILD; PFANNKUCH, 1999, p. 227). A transnumeração refere-se à ideia de formar e alterar representações de dados de um sistema para chegar a um melhor entendimento deles.

4.1 Pensamento estatístico – transnumeração

Especificamente, o uso de *software* e os avanços tecnológicos na direção de manipulação e exibição de dados tornaram a pesquisa sobre transnumeração relevante (DAIGA, 2018). Lee et al. (2014) observaram que as ferramentas de *software* para criar representações é um aspecto proveitoso em muitas ações transnumerativas.

Segundo Wild e Pfannkuch (1999), a transnumeração é “um processo dinâmico de mudança de representações para gerar compreensão” (p. 227). A discussão sobre a transnumeração foi apresentada como parte do interesse dos autores no desenvolvimento de uma estrutura para padrões de pensamento, estratégias e utilização de elementos estatísticos na resolução de problemas, como mostra a Figura 1. Na sequência, o *framework* proposta por Wild e Pfannkuch (1999) é apresentada, permitindo explorar as dimensões desse referencial teórico e delinear a transnumeração como aporte teórico para esta tese.

Figura 1 – Framework dimensional para o pensamento estatístico em investigação empírica



Fonte: Pfannkuch e Wild (1999, p. 226, tradução nossa).

Na Figura 1, é apresentado o modelo dimensional proposto por Wild e Pfannkuch (1999), que descreve quatro dimensões para o pensamento estatístico.

A dimensão um, *o ciclo investigativo*, aborda a resolução de um problema real e propõe o modelo PPDAC (problema, plano, dados, análises e conclusões), objetivando, geralmente, mudar o sistema para melhorar algo. Wild e Pfannkuch (1999) ressaltam que

esse ciclo é iniciado para promover a aprendizagem, de modo que certos objetivos devem ser alcançados para que seja possível chegar a um nível desejado de compreensão do problema. Os autores destacam que:

O conhecimento adquirido e as necessidades identificadas dentro desses ciclos podem iniciar novos ciclos investigativos. As conclusões das investigações alimentam uma base de conhecimento de contexto ampliada que pode, em seguida, informar quaisquer ações (WILD; PFANNKUCH, 1999, p. 225, tradução nossa³⁷).

Com referência aos apontamentos desses autores, esta tese considera o ciclo investigativo como recurso metodológico para propor a implementação das atividades da pesquisa. No capítulo de metodologia e procedimentos, o ciclo PPDAC é retomado.

Na dimensão dois, representada na Figura 1, são apresentados os *tipos de pensamento* estatístico, incluindo os tipos geral e fundamental. Embora os autores citem os do tipo geral, nesta tese o referencial se concentra nos tipos fundamentais de pensamento estatístico presentes no *framework* dos autores, entre os quais os pressupostos relativos às representações de dados, que podem ser objetos de análise.

Um tipo de pensamento estatístico fundamental trata-se do *reconhecimento da necessidade de dados*, ou seja, “o reconhecimento das inadequações das experiências pessoais e evidências pontuais que levam ao desejo de basear decisões em dados deliberadamente coletados é um impulso estatístico” (WILD; PFANNKUCH, 1999, p. 227, tradução nossa³⁸).

A *transnumeração* é um tipo de pensamento estatístico voltado à construção e alteração das representações de dados de um sistema, buscando melhor compreendê-lo, e é uma ideia fundamental para a aprendizagem estatística (WILD; PFANNKUCH, 1999). Essa mudança de representações de dados e a compreensão deles é o que se pode observar ao realizar a análise gráfica de dados.

A análise gráfica de dados, viabilizada por meio de mudanças de representações desses dados, é um aspecto central abordado na problemática e objeto de estudo desta tese, abordagem que segue os pressupostos de Wild e Pfannkuch (1999) e é observada como resultados da transnumeração, que serve como referencial neste estudo. Na sequência, após

³⁷ Knowledge gained and needs identified within these cycles may initiate further investigative cycles. The conclusions from the investigations feed into an expanded context-knowledge base which can then inform any actions.

³⁸ The recognition of the inadequacies of personal experiences and anecdotal evidence leading to a desire to base decisions on deliberately collected data is a statistical impulse.

situar os demais tipos de pensamento fundamentais e as demais dimensões do *framework*, conforme Figura 1, procede-se a um aprofundamento deste tipo de pensamento estatístico.

A *consideração da variação*, outro tipo de pensamento apresentado por Wild e Pfannkuch (1999), coloca em pauta o pensamento estatístico no contexto da aprendizagem e da tomada de decisão em um ambiente de incerteza, situando a variação como um fenômeno onipresente. Os autores ressaltam a necessidade de examinar o papel da variação, porque, além da onipresença, há a possibilidade de consequências sob o ponto de vista prático. Ainda, enfatizam o fato de a estatística permitir a compreensão de um mundo cercado de variações.

Quanto ao *raciocínio com modelos estatísticos*, Wild e Pfannkuch (1999, p. 227) definem: “todo pensamento usa modelos”. Para os autores, a principal contribuição da disciplina de Estatística ao pensamento é “o seu próprio conjunto distinto de modelos, ou *Frameworks*, para pensar sobre certos aspectos da investigação de forma genérica” (WILD; PFANNKUCH, 1999, p. 227, tradução nossa³⁹).

Sobre a *integração estatística e contextual*, nesse tipo de pensamento se considera a habilidade de integrar o conhecimento estatístico com o conhecimento do contexto e da informação contida nos dados. De acordo com Wild e Pfannkuch (1999, p. 228), “o próprio pensamento é a síntese desses elementos para produzir implicações, *insights* e conjecturas. Um não pode se relacionar ao pensamento estatístico sem algum conhecimento de contexto⁴⁰”. Nesse sentido, ressaltam que é necessário, ao ensino de estatística, considerar os diferentes conhecimentos e que a integração deles permite conexões que podem gerar sentido.

Após apresentar os tipos fundamentais de pensamento estatístico da dimensão dois, na dimensão seguinte do *framework* de Wild e Pfannkuch (1999), têm-se o ciclo interrogativo, “um processo de pensamento genérico em uso constante na resolução de problemas estatísticos⁴¹” (p. 231). Para os autores, o aprendiz está sempre em um dos estados interrogativos durante a resolução de problemas, seja para gerar ideias, buscar informações, interpretar resultados de buscas, criticar a consistência de ideias e informações recebidas e julgar com base nas críticas realizadas.

³⁹ [...] its own distinctive set of models, or frameworks, for thinking about certain aspects of investigation in a generic way.

⁴⁰ The thinking itself is the synthesis of these elements to produce implications, insights and conjectures. One cannot indulge in statistical thinking without some context knowledge.

⁴¹ [...] is a generic thinking process in constant use in statistical problem solving.

Na dimensão quatro, as *disposições* se referem ao compromisso com um problema que se submete a resolvê-lo. Conforme Wild e Pfannkuch (1999), nessa dimensão discutem-se as qualidades pessoais que afetam ou mesmo iniciam a entrada em um modo de pensamento no contexto da resolução de problemas estatísticos. A curiosidade e a consciência, a imaginação, o ceticismo, o ser lógico, a perseverança, uma propensão a buscar por um significado mais profundo e o engajamento são qualidades consideradas relevantes para uma pessoa que se coloca a resolver um problema estatístico.

As dimensões apresentadas e propostas por Wild e Pfannkuch (1999) permitem situar que a transnumeração é parte de um marco teórico mais amplo, o qual, como um tipo fundamental de pensamento estatístico, é adotado como referencial para esta tese.

A transnumeração refere-se às “transformações de numeracia feitas para facilitar a compreensão” (WILD; PFANNKUCH, 1999, p. 227, tradução nossa⁴²).

A transnumeração ocorre quando encontramos formas de obter dados (por meio de medição ou classificação) que captam elementos significativos do sistema real. Ela permeia toda a análise estatística de dados, ocorrendo toda vez que mudamos nossa maneira de ver os dados, na esperança de que isso nos transmita um novo significado (WILD; PFANNKUCH, 1999, p. 227, tradução nossa⁴³).

Como pontuam os autores, a transnumeração é um tipo fundamental de pensamento estatístico que, sob o ponto de vista de análises na resolução de um problema, pode ser considerada como:

Ideias sobre possíveis causas e outros fatores que podem ser preditores importantes do comportamento da resposta, são traduzidas em um conjunto de variáveis para medir (transnumeração) e os dados são coletados para facilitar a investigação das relações entre variáveis medidas e as respostas de interesse (WILD; PFANNKUCH, 1999, p. 227, tradução nossa⁴⁴).

De acordo com os autores, as causas esperadas são aquelas que têm alguma uniformidade ou regularidade e que só é possível detectá-las se pensarmos em alguma forma de olhar para a situação que revele essa uniformidade ou regularidade, essencialmente, se conseguirmos “transnumerar” sobre os dados do problema (WILD; PFANNKUCH, 1999).

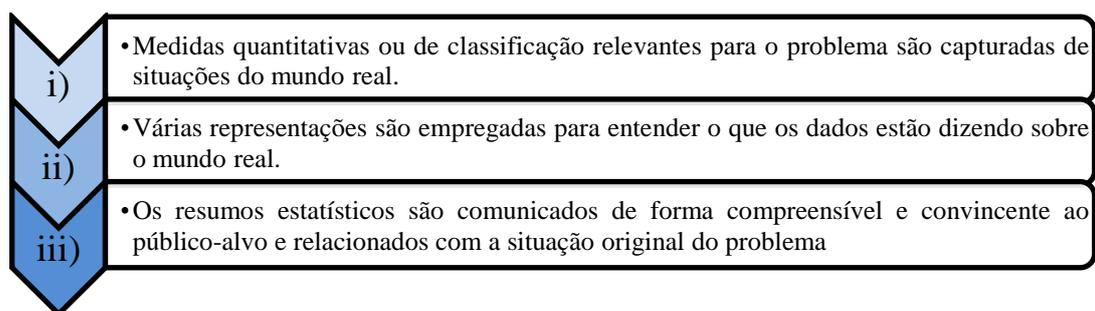
⁴² [...] "numeracy transformations made to facilitate understanding".

⁴³ Transnumeration occurs when we find ways of obtaining data (through measurement or classification) that capture meaningful elements of the real system. It pervades all statistical data analysis, occurring every time we change our way of looking at the data in the hope that this will convey new meaning to us.

⁴⁴ Ideas about possible causes and other factors that might be important predictors of the behaviour of the response are translated into a set of variables to measure (transnumeration) and data is collected to facilitate investigation of relationships between measured variables and the responses of interest.

De forma a delinear como a transnumeração se insere como aporte teórico nesta tese, são consideradas as propostas de Pfannkuch, Rubick e Yoon (2002) e Pfannkuch, Rubick (2002), que admitem a transnumeração como um processo em três estágios, como mostra a Figura 2:

Figura 2 – Estágios de transnumeração



Fonte: Adaptado de Pfannkuch, Rubick e Yoon (2002).

Conforme Pfannkuch, Rubick e Yoon (2002), como mostra a Figura 2, no primeiro estágio o foco está na captura de qualidades ou características da situação real; no segundo estágio, para que sejam apresentadas representações de dados, o foco está na necessidade de os dados serem coletados, transformados de brutos em múltiplas representações gráficas, resumos estatísticos, de modo a permitir a obtenção de significado da situação real a partir deles; no terceiro estágio, dada a obtenção de significados, o foco reside na comunicação a partir dos dados e julgamentos de forma a serem compreendidos, em termos das situações reais, por outros. Nesta tese, os estágios de transnumeração permitiram analisar o percurso dos participantes em cada atividade proposta, de modo a elucidar as contribuições dos recursos tecnológicos implementados que são apresentadas no Capítulo 6.

Pfannkuch e Rubick (2002) observam que os conhecimentos prévios do contexto e da estatística, o pensamento em um nível elevado em relação às representações construídas, a representação e interpretação ativa, o entrelaçamento do pensamento local e do pensamento global e a mudança de diálogo do pensamento estatístico entre as representações são pontos relevantes a serem considerados na análise da construção de significados pautados em dados reais, pontos relevantes nas análises dos resultados obtidos a partir das atividades apresentadas nesta tese.

Os estágios delineados por Pfannkuch, Rubick e Yoon (2002) permitem compreender como a transnumeração ocorre ao longo de uma atividade que envolve dados, e os pontos

levantados por Pfannkuch e Rubick (2002) realçam a oportunidade de analisar conhecimentos mobilizados pelos participantes durante o processo de transnumeração.

A problemática desta tese está centrada na implementação de recursos tecnológicos em atividades envolvendo a análise gráfica de dados e visa investigar as possíveis contribuições da utilização dos recursos *Gapminder*, GeoGebra e SPSS. As atividades são analisadas com base na definição de transnumeração, entendida como um processo dinâmico de mudança de representações para gerar compreensão, conforme proposto por Wild e Pfannkuch (1999), levando em consideração os estágios delineados por Pfannkuch, Rubick e Yoon (2002) e os pontos relevantes destacados por Pfannkuch, Rubick (2002).

A problematização da implementação de recursos digitais em atividades que envolvem a análise gráfica de dados direciona para uma investigação de possíveis potencialidades desses recursos, pois, como destacam Prodromou e Dunne (2017), o desenvolvimento no poder de computação tem melhorado significativamente as representações gráficas. Essa afirmação emerge em uma proposição sobre a visualização de dados como um conjunto de competências para examinar criticamente declarações baseadas em dados e representações de dados.

4.2 Visualização de dados

Ao introduzir a visualização de dados como referencial para esta tese, considera-se, de acordo com Uwin (2020), que ela envolve a construção de representações (exposições) gráficas de dados incluindo exibições transformadas que têm como principal objetivo facilitar a interpretação e obtenção de informações.

Keim et al. (2006) ressaltam que a ideia básica da análise visual é representar visualmente a informação, de modo que o ser humano interaja diretamente com ela, obtenha *insights*, tire conclusões e, em última instância, melhore as decisões. A vantagem específica da análise visual é que “aqueles que tomam decisões podem concentrar suas capacidades cognitivas e perceptuais completas no processo analítico, ao mesmo tempo que permite a aplicação de recursos computacionais para aumentar o processo de descoberta” (KEIM; MANSMANN; SCHNEIDEWIND; ZIEGLER, 2006, p. 9).

Para Prodromou e Dunne (2017), a visualização de dados, no campo da Educação Estatística, engloba a análise exploratória de dados assistida por computador, o que

possibilita explorar e comunicar efetivamente informações relevantes por meio de representações gráficas.

Como esta tese investiga as possíveis contribuições de recursos tecnológicos implementados para a análise gráfica de dados, a visualização de dados é um referencial para as análises das atividades propostas, pressupondo esse processo assistido pelos recursos *Gapminder*, *GeoGebra* e *SPSS*. Essa escolha se justifica porque a proposição de Prodromou e Dunne (2017) para a visualização de dados enfatiza as possibilidades de aprendizagem que podem ocorrer a partir do desenvolvimento tecnológico.

[...] os avanços da computação também fortaleceram as representações gráficas exploratórias que, por sua vez, forneceram suporte para explorar ainda mais os dados. A qualidade e a quantidade de representações gráficas foram melhoradas. Muitas exibições distintas dos mesmos dados podem ser esboçadas para lançar luz sobre aspectos ocultos da informação (PRODROMOU; DUNNE, 2017, p. 2, tradução nossa⁴⁵).

Embora Prodromou e Dunne ressaltem que a revolução não reside nos recursos que permitem a manipulação diversificada dos dados estatísticos, eles consideram que a revolução está nas formas de utilização dos dados e afirmam que “a grande importância da disponibilidade de *software* e da popularidade associada, na determinação de quais análises são realizadas e como elas são apresentadas, é um tópico de grande relevância” (PRODROMOU; DUNNE, 2017, p. 2, tradução nossa⁴⁶).

A proposta de Prodromou e Dunne (2017) emerge do contexto *OPEN Data*⁴⁷ em que vivemos e busca apreciar a relevância de recursos de visualização de dados para a educação estatística. Os autores discutem questões relacionadas ao papel da visualização, sob o ponto de vista da *compreensão*, de suas *finalidades* e da *verbalização*. O resultado dessa discussão leva às questões relativas ao letramento estatístico, do qual, ao considerarem o contexto *OPEN Data*.

⁴⁵ [...] computing advances have also strengthened exploratory graphical representations that in turn provided support to exploring data further. The quality and quantity of graphical representations has been improved. Many distinct displays of the same data can be drawn to shed light on hidden aspects of the information.

⁴⁶ The great importance of computer software availability and associated popularity, in determining what analyses are carried out and how they are presented, is a topic of major relevance. In the world of business, the spreadsheet Excel has long been in common use for creating graphical.

⁴⁷ *OPEN Data* são dados que podem ser utilizados, reutilizados e redistribuídos livremente por qualquer pessoa, e que estão sujeitos, quando muito, à exigência de atribuição e a serem compartilhados da mesma forma que aparecem. Este artigo descreve o conceito de dados abertos e dados de pesquisa e por que eles precisam ser compartilhados (ALEIXANDRE-BENAVENT; SAPENA; PESET, 2021, p. 1). Disponível em: <https://www.sciencedirect.com/science/article/pii/S1575181319301287>. Acesso em: jun. 2023.

Nesta tese, o letramento estatístico, considerado na literatura e em pesquisas por autores como Gal e Garfield (1997) e Gal (2002), será concebido na perspectiva de Prodromou e Dunne (2017), os quais propõem uma outra organização do letramento estatístico como referencial para esta tese.

Na sequência, estão apresentados elementos da discussão dos autores sobre as questões relacionadas à visualização, seguidas pela organização teórica proposta por eles.

Sobre a *compreensão* no contexto de visualização, Prodromou e Dunne (2017) incluem descrições estáticas e dinâmicas de características multidimensionais e definem que “a visualização envolve a produção de uma representação visual de um ou mais aspectos de um conjunto de dados” (PRODROMOU; DUNNE, 2017, p. 4, tradução nossa⁴⁸).

Quanto aos objetivos da visualização, os autores indicam que um deles é permitir ciclos de repetição e acúmulo de experiências com diferentes ferramentas e representações, cada uma das quais constitui um envolvimento com os dados e suas possíveis características subjacentes, aspectos considerados nesta tese.

Sob o ponto de vista da *compreensão* mediada por recursos de visualização:

[...] o usuário acessa processos cognitivos rápidos que permitem o reconhecimento de características estruturais insuspeitas ou latentes de um corpo detalhado de números, cadeias ou símbolos. Em consequência, a visualização faz parte da compreensão e integra-se na sua comunicação (PRODROMOU; DUNNE, 2017, p. 4, tradução nossa⁴⁹).

Os autores definem que as *finalidades* da visualização podem oferecer suporte para a produção de significados do contexto e a interpretação por meio de estratégias cognitivas disponíveis. Nesse sentido, apresentam alguns pontos a serem considerados:

- Descrever e comparar condições ou estados observados em um contexto (em um instante).
- Descrever e avaliar associações entre categorias, contagens e medidas (muitas vezes com aspectos de tempo ignorados).
- Descrever e comparar mudanças ou processos atuais em um contexto (ao longo de um período, às vezes com intervalos de interobservação iguais).

⁴⁸ Visualisation involves the production of a visual representation of one or more aspects of a data set.

⁴⁹ [...] the user accesses rapid cognitive processes that permit recognition of unsuspected or latent structural features of a detailed body of numbers, strings, or symbols. In consequence visualisation is part of comprehension and integrates into its communication.

- Descrever e avaliar associações entre mudanças nas variáveis observadas (ao longo de alguns comprimentos de intervalo de tempo implícitos ou especificados) (PRODROMOU; DUNNE, 2017, p. 4, tradução nossa⁵⁰).

Os pontos destacados pelos autores, também considerados nesta tese, podem ser estratégias úteis para professores ao implementarem atividades relacionadas à visualização de dados, de modo que, para o aluno, a visualização de dados pode fornecer um resumo dos dados associados. Quanto às condições para análise de dados, “a fluência com a visualização implica uma consciência das possíveis características dos dados que são marginalizadas por qualquer visualização em particular” (PRODROMOU; DUNNE, 2017, p. 4, tradução nossa⁵¹).

De acordo com os autores, dada a oportunidade de acessar essa fluência, os alunos podem desenvolver competências para reformular crenças, de modo a acomodar novas percepções. “As crenças, inferências e visualizações também podem informar ou permitir *insights* retrospectivos sobre a causalidade associada aos fenômenos que deram origem aos dados e à(s) representação(ões) produzida(s) pelo aluno” (PRODROMOU; DUNNE, 2017, p. 5, tradução nossa⁵²).

Sob a perspectiva das finalidades da visualização, os pesquisadores destacam que favorecer a tomada de sentido e de decisões é o seu propósito, e, no contexto do letramento estatístico, permite ao aluno o controle consciente de benefícios e a consciência de suas limitações.

Ao se referirem à *verbalização*, Prodromou e Dunne (2017) a observam como um complemento e resultado da visualização. Para os autores, em um contexto de aprendizagem no qual se considera a visualização, é esperado que emergjam comentários, *insights* ou perguntas. Isso é o que se espera nas análises dos resultados desta tese.

Nesse contexto, o papel do professor é “estar atento às potencialidades e fragilidades desses *insights* espontâneos e possibilitar intercâmbios que permitam aos alunos formular e

⁵⁰ Describe and compare observed conditions or states in a context (at an instant). Describe and assess associations between categories, counts and measures (often with time aspects ignored). Describe and compare current changes or processes in a context (over a period, sometimes with equal inter-observation intervals). Describe and assess associations between changes in observed variables (over some implicit or specified time-interval lengths).

⁵¹ Fluency with visualisation implies an awareness of possible features of the data that are marginalised by any particular visualisation.

⁵² The beliefs, inferences and visualisations may also inform or allow retrospective insights into causation associated within the phenomena that gave rise to the data and the representation(s) as produced by the student.

revisar iterativamente suas impressões e raciocínios em formatos apropriados” (PRODRUMOU; DUNNE, 2017, p. 5, tradução nossa⁵³).

Segundo os estudiosos, as *verbalizações* precipitadas, no contexto de estudo de dados, devem atender aos critérios de explicitações, coerência e precisão de natureza como descrições de variabilidade e pressupostos articulados, necessários para a extrapolação. Prodromou e Dunne (2017) também destacam que, para a visualização, é necessário um conhecimento contextual, de modo que é relevante, antes de explorar padrões e tendências, explicitar o contexto de interesse incorporado nos dados. “Essa especificação de contexto auxilia na compreensão de como os processos de localização de padrões podem envolver níveis ocultos ou implícitos [identificação/designação] de agregação ou partição” (PRODRUMOU; DUNNE, 2017, p. 6, tradução nossa⁵⁴).

Na era *OPEN Data*, Prodromou e Dunne (2017) consideram que, ao lidar com dados, é conveniente utilizar uma abordagem orientada pelos próprios dados, permitindo escolhas exploratórias sobre quais variáveis podem ser examinadas. Nesse sentido, significa que “usamos comparações de grupos e associações entre variáveis (correlações dentro dos casos) para obter informações sobre as relações entre variáveis importantes em fenômenos de interesse” (PRODRUMOU; DUNNE, 2017, p. 6, tradução nossa⁵⁵). O que é almejado é a formação de múltiplas hipóteses provisórias.

Sob o ponto de vista do ensino da Estatística, de acordo com os autores, é fundamental proporcionar aos alunos oportunidades de interagir com diferentes formas de representação de dados, buscando promover compreensões mais profundas de construções e métodos complexos e detalhados.

A finalidade dessas fluências mais flexíveis orquestradas entre as formas é permitir e apoiar explorações da natureza e complexidade das vastas quantidades de dados disponíveis no mundo moderno em quase todos os tipos de tomada de decisão são baseadas no conhecimento (PRODRUMOU; DUNNE, 2017, p. 7, tradução nossa⁵⁶).

⁵³ [...] to be alert to the potential and the fragilities of these spontaneous insights and enable interchanges that allow the students to formulate and iteratively revise their impressions and reasoning in appropriate formats.

⁵⁴ This specification of context assists in understanding how processes of finding pattern may involve hidden or implicit [identification/designation] levels of aggregation or partition. Knowledge of context may give rise to a hypothesis-driven approach, but also more often to various aggregation strategies.

⁵⁵ We use comparisons of groups and associations between variables (correlations within cases) to gain insight into relationships between important variables in phenomena of interest.

⁵⁶ The purpose of these orchestrated looser fluencies between modes is to permit and support explorations of the nature and complexity of the vast quantities of data available in the modern world in almost every type of knowledge-based decision-making.

Para Prodromou e Dunne (2017, p. 7, tradução nossa⁵⁷), “os esforços na visualização de dados, tanto como uma atividade figural ou gráfica construtiva quanto como uma ferramenta de comunicação, serão mais eficazes se estiverem ligados às representações textuais e tabulares explícitas dos mesmos *insights*”.

Prodromou e Dunne (2017) abordam a importância da visualização no ensino da Estatística em um contexto de *OPEN Data*. A implementação de recursos tecnológicos para análise gráfica de dados proposta por esta tese busca investigar, dentro das possibilidades, contribuições de opções disponíveis nesses recursos para mediação de atividades de visualização, considerando o *framework* proposta pelos autores em relação ao letramento estatístico.

O *framework* de Prodromou e Dunne (2017) leva em conta toda a complexidade dos dados e argumenta pela necessidade de uma mudança na forma como se abordam o controle e a manipulação de dados.

Embora nos baseemos nos valores, elementos de conhecimento e elementos disposicionais usados para descrever os construtos de letramento estatístico, reconhecemos a necessidade urgente de novos princípios e disposições que se tornarão os blocos de construção de um novo modelo de letramento que aborda a era *OPEN Data* (PRODROMOU; DUNNE, 2017, p. 7, tradução nossa⁵⁸).

A proposição dos autores pondera que “a era *OPEN Data* desafia as pessoas a se tornarem mais bem informadas sobre as maneiras pelas quais poderiam aproveitar vastos corpos de dados, em vez de pequenos conjuntos de dados e, simultaneamente, aproveitar a tecnologia que facilita sua exploração” (PRODROMOU; DUNNE, 2017, p. 6, tradução nossa⁵⁹). No Quadro 3 estão os elementos do framework para o letramento estatístico na era *OPEN Data*.

⁵⁷ [...] efforts in data visualisation, both as a constructive figural or graphical activity and as a tool of communication, will be more effective if they are linked to explicit textual and tabular representations of the same insights.

⁵⁸ Although we will build upon the values, knowledge elements, and dispositional elements used to describe the constructs of statistical literacy, we recognise the urgent need for new principles and dispositions that will become the building blocks of a novel model of literacy addressing the *OPEN Data* era.

⁵⁹ The *OPEN Data* era challenges people to become better informed about the ways in which they could harness vast bodies of data rather than small datasets, and simultaneously harness the technology that facilitates exploration of data.

Quadro 3 – Elementos do letramento estatístico na era *OPEN Data*

1) Ver as inferências plausíveis dentro dos dados
2) Linguagem
3) Contexto, correlação, causalidade
4) Avaliar dados <ul style="list-style-type: none"> • Avaliar a qualidade da evidência • Conceber estatísticas como modelos • Estatísticas orientadas para a ação

Fonte: Adaptado de Prodromou e Dunne (2017, p. 17)

O primeiro elemento do *framework*, ver inferências plausíveis dentro dos dados, tem a ver com a familiaridade com várias ideias fundamentais, especialmente em *OPEN Data*, “conjuntos de dados multivariados, visualização e proveniência, e qualidade dos dados, fundamentam a capacidade dos alunos de entender a exploração e análise de conjuntos de dados multivariados em larga escala” (PRODROMOU; DUNNE, 2017, p. 17, tradução nossa⁶⁰)

Para os autores, cabe ao professor selecionar algumas ideias-chave estatísticas, representadas por símbolos matemáticos ou termos estatísticos, mas não só notações técnicas que ajudem os alunos a descreverem os dados de forma eficiente. Prodromou e Dunne (2017) afirmam que tanto as investigações causais quanto a análise associativa são recursos para que uma hipótese sob teste seja rejeitada ou adotada e, em seguida, descrita. De maneira geral, é necessário que o aluno acesse alguma orientação para reunir evidências de relações causais de associações, o que foi considerado nas propostas de atividades desta tese.

Ao se referirem à *linguagem*, Prodromou e Dunne (2017) destacam as dissonâncias vivenciadas pelos alunos advindas de novos construtos. Essas dissonâncias “decorrem dos múltiplos significados contextuais anteriores de termos adotados com nova interpretação em uma nova disciplina inserida para um aluno” (PRODROMOU; DUNNE, 2017, p. 17, tradução nossa⁶¹). Para eles, termos como possibilidade, probabilidade, razão de chances, entre outros, precisam ser recontextualizados para admitir a entrada de novos entendimentos.

Sob a perspectiva da *linguagem*, elementos como texto, tabela, figura e gráfico são considerados pelos autores como formas de engajamento visual que podem favorecer

⁶⁰ [...] multivariate data sets, visualisation, and data provenance and quality underlies students’ ability to understand the exploration and analysis of large-scale multivariate data sets.

⁶¹ [...] arise from the earlier multiple contextual meanings of terms adopted with fresh interpretation in a new discipline being entered by a student.

estratégias na busca de compreensões mais aprofundadas de termos complexos no contexto estatístico. Para Prodromou e Dunne (2017), a introdução de tais elementos linguísticos numa visualização de dados não é tão improvável como pode parecer pelo seu valor nominal.

Ao se referirem ao aspecto de *contexto, mudança e causa*, Prodromou e Dunne (2017, p. 17, tradução nossa⁶²) consideram que “a prática de observar ou articular explicitamente um contexto permite tanto a canalização da exploração quanto uma atenção crítica mais profunda”. Para os autores, o conhecimento de contexto leva à apreciação de pré-condições latentes na estrutura e composição dos dados e pode ser considerado pelo professor na proposição de atividades.

Sob o ponto de vista da análise dos dados, Prodromou e Dunne (2017) observam que o ato de articular a análise com o contexto fundamenta a descrição de relações e associações entre variáveis.

Quanto às mudanças, Prodromou e Dunne (2017) explicam que, em *OPEN Data*:

As configurações de dados com uma ordem de tempo inerente de quaisquer observações múltiplas podem levar a um foco sobre variáveis derivadas de mudança em categoria ou mudança em contagem ou mudança em medida sobre sequência especificada ou intervalos de tempo de interesse (p. 18, tradução nossa⁶³).

Prodromou e Dunne (2017) definem que os níveis mais fortes de evidência envolvem necessariamente a associação de pelo menos uma mudança em uma variável explicativa com uma mudança em uma variável de resposta. A percepção desses níveis de evidências ocorre pela observância do desenvolvimento de efeitos esperados ao longo do tempo. Para Prodromou e Dunne (2017), elementos contextuais permitem uma apreciação da complexidade do estabelecimento de relações de causa e efeito.

A causa e o efeito são vistos pelos autores como possíveis de serem observados em um ciclo investigativo, uma das dimensões do pensamento estatístico definida por Wild e Pfannkuch (1999). Prodromou e Dunne (2017) salientam que, necessariamente, a organização do ciclo pode conduzir os alunos a contextos nos quais inferências e decisões envolvendo noções de causa e efeito devem ser realizadas.

⁶² The practice of explicitly noting or articulating a context permits both channeling of exploration and deeper critical attention.

⁶³ OPEN Data settings with an inherent time ordering of any multiple observations may lead to a focus on the derived variables of change in category or change in count or change in measure over specified sequence lengths or time intervals of interest.

Ao considerarem o contexto, a mudança e a causa como elementos necessários ao letramento em *OPEN Data*, Prodromou e Dunne (2017) destacam o conhecimento do contexto como um fator relevante para a seleção e redução de dados, de modo que os estudos realizados possam considerar as associações significativas. “Se o aluno não está ciente do conhecimento do contexto dentro do qual os dados são gerados, é muito difícil identificar partes apropriadas e buscar relações entre essas partes” (PRODROMOU; DUNNE, 2017, p. 18, tradução nossa⁶⁴). Essa situação se manifesta na proposição das atividades desta tese.

Prodromou e Dunne (2017) observam que a visualização de dados em *OPEN Data* proporciona uma perspectiva de inferências válidas no mundo real, especialmente quando se espera encontrar evidências de efeitos causais em determinados conjuntos de mudanças. Para eles, se desejamos promover explorações de causa e efeito em atividades de ensino de Estatística, é necessário apresentar aos alunos as condições ou suposições para justificar qualquer inferência causal, ou seja, o aluno deve ser apresentado a pelo menos um objeto de estudo que ofereça o rigor necessário para inferências causais formais. Ainda asseveram que, antes de considerar as evidências gráficas disponíveis como adequadas para uma análise, é necessário observar os possíveis efeitos causais.

Prodromou e Dunne (2017, p. 20, tradução nossa⁶⁵) apontam como último elemento do framework a *avaliação dos dados* e consideram alguns pontos necessários para a evolução do letramento estatístico:

- Avaliar a qualidade da evidência: Envolve esperar, buscar e fornecer informações sobre fontes de dados e inserir especificações para acessar os metadados em bancos de dados que apresentem dados multivariados e discussões que chamem a atenção para a fonte de geração de dados e qualidade associada;
- Conceber a estatística como modelo: É a capacidade de usar modelos estatísticos para descrever uma ampla gama de fenômenos em nosso mundo, e quantificar aspectos desses fenômenos é algo importante;

⁶⁴ If the student is not aware of the context knowledge within which data are generated, it is very difficult to identify appropriate parts and seek relationships amongst those parts.

⁶⁵ **Evaluating the Quality of Evidence:** This element involves expecting, seeking and providing information about data sources and inserting specifications to access the metadata into displays that present multivariate data, and discussions that draw attention to the source of data generation and associated quality. **Conceiving Statistics as Forming Models:** The ability to use statistical models to describe a wide range of phenomena in our world and quantify an aspect of these phenomena is important. **Action-Oriented Statistics:** According to Ridgway et al. (2013), it is the purposive nature of investigation and consequent actions that will engage and enrich student learning.

- Estatísticas orientadas para a ação: É a natureza proposital da investigação e consequentes ações que envolverão e enriquecerão a aprendizagem dos alunos.

Diante desses pontos, considerados como aspectos relevantes no letramento estatístico na era *OPEN Data*, Prodromou e Dunne (2017) explicam que a Estatística é vista no contexto de um ciclo investigativo, em que o ponto final é um relato teórico e alguma ação destinada a mudar a situação atual, processo facilitado pelo uso de ferramentas de visualização. Tais aspectos são considerados na problemática desta tese, ou seja, investigar as possibilidades de análise gráfica de dados mediados por recursos digitais.

Prodromou e Dunne (2017, p. 23, tradução nossa⁶⁶) argumentam sobre a incorporação de novos elementos e disposições para o letramento estatístico na era *OPEN Data*:

Em particular, as inovações tecnológicas na arquitetura de computadores não só permitem o armazenamento de grandes volumes de dados, mas também permitem que os usuários obtenham representações gráficas de maior qualidade que contribuem para uma proeminência da visualização de dados e, de forma mais geral, aos dados.

Além de destacar a importância dos recursos emergentes do desenvolvimento tecnológico, na direção da problemática desta tese, Prodromou e Dunne (2017) enfatizam que, para promover o letramento estatístico, é necessária a organização de muitas formas de visualização, que devem ser complementadas com uma linguagem explícita, de modo a explorar a capacidade humana de perceber o espaço e o tempo tridimensionais e invocar esses elementos relacionando-os com os dados.

A transnumeração, um tipo fundamental de pensamento estatístico, é observada a partir das dimensões propostas no framework de Wild e Pfannkuch (1999) e dos estágios apresentados por Pfannkuch, Rubick e Yoon (2002). Dada a problemática de investigar as possibilidades de análise gráfica de dados mediados por recursos digitais, questões relacionadas à *compreensão*, às *finalidades* e à *verbalização*, no âmbito da visualização de dados, foram apresentadas, conforme Prodromou e Dunne (2017). Esses autores consideram

⁶⁶ In particular, technological innovations in computer architecture not only allow the storage of large volumes of data but also enable users to obtain higher-quality graphical representations that have contributed to giving a prominent role to data visualisation and more generally to data.

o letramento estatístico na era *OPEN Data* e apresentam um referencial com elementos necessários ao nesse contexto.

Como evidenciado por Wild e Pfannkuch (1999) e apontado por Prodromou e Dunne (2017), como uma organização pertinente para o estudo da estatística, o ciclo investigativo é exposto no próximo capítulo, o qual apresenta o referencial metodológico.

Neste capítulo foi apresentada a transnumeração como um tipo fundamental de pensamento estatístico, observada a partir das dimensões propostas no framework de Wild e Pfannkuch (1999), bem como foram considerados os estágios apresentados por Pfannkuch, Rubick e Yoon (2002). As possibilidades de análise gráfica de dados mediada por recursos digitais estão relacionadas à *compreensão*, às *finalidades* e à *verbalização* discutidas no âmbito da visualização de dados, apresentadas conforme Prodromou e Dunne (2017) propuseram na era *OPEN Data*, um referencial com elementos necessários ao letramento estatístico.

No próximo capítulo, são expostos os procedimentos para a implementação do experimento, levando em consideração os aportes teórico e metodológico apresentados anteriormente.

5 METODOLOGIA E PROCEDIMENTOS

Esse capítulo descreve a metodologia adotada para a implementação da parte experimental desta tese, o ciclo investigativo PPDAC (Problema, Planejamento, Desenvolvimento, Análise e Conclusão), conforme indicado no *framework* dimensional para o pensamento estatístico em investigação empírica Wild e Pfannkuch (1999).

Como citado no Capítulo 2, sobre o percurso desta pesquisa, foram explicitadas as condições que viabilizaram a aplicação do estudo e, com base nessas condições, adotou-se uma abordagem qualitativa. Nos termos de Godoy (1995), algumas características dessa abordagem considera o ambiente natural como fonte de dados e o pesquisador é visto como instrumento fundamental; observa-se um caráter descritivo e um enfoque indutivo.

Nesse sentido, são destacados neste capítulo informações relativas ao momento histórico e às restrições para a aplicação do estudo, as instituições envolvidas e os recursos subjacentes à sua realização, contexto de sua implementação, os recursos digitais utilizados, os participantes da pesquisa e as atividades propostas. Todos esses elementos condicionaram o desenvolvimento desta investigação.

5.1 Referencial metodológico - ciclo investigativo - PPDAC

Os procedimentos metodológicos desta pesquisa permitiram a aplicação de uma proposta que investigou as possíveis contribuições de recursos tecnológicos em atividades relativas à análise gráfica de dados.

Como propõe o *framework* de Wild e Pfannkuch (1999), em atividades estatísticas, o ciclo investigativo permite a organização para promover a aprendizagem da Estatística. Denominado como PPDAC, o ciclo descreve o processo de investigação estatística, que, eventualmente, pode ser uma possibilidade para questionar, sob alguma medida, algo no mundo.

A parte experimental deste estudo é uma proposta de atividades que envolve o estudo de situações reais, sob ciclos investigativos, buscando identificar contribuições dos recursos *Gapminder*, GeoGebra e SPSS em análise gráfica de dados.

O ciclo PPDAC é a descrição de uma abordagem sistemática de investigação (WILD; PFANNKUCH, 1999), e considerado como um recurso metodológico desta tese, logo, para

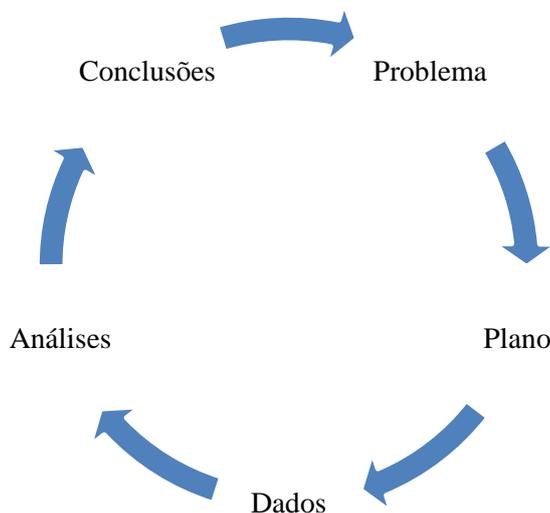
que possa ser compreendido, são expostos os elementos que os autores concebem e que compõem esse ciclo.

Um ciclo PPDAC está preocupado em abstrair e resolver um problema estatístico fundamentado em um problema "real" maior. A maioria dos problemas está inserida no desejo de mudar um "sistema" para melhorar algo. [...] Uma solução baseada no conhecimento para o problema real requer uma melhor compreensão de como um sistema funciona e talvez também como ele reagirá às mudanças nos fluxos de entrada, configurações ou ambiente (WILD; PFANNKUCH, 1999, p. 225, tradução nossa⁶⁷).

Wild e Pfannkuch (1999) apontam que um ciclo investigativo pode ser definido para atingir objetivos de aprendizagem. O conhecimento adquirido e as necessidades identificadas nesses ciclos podem iniciar novos ciclos investigativos, de modo que as conclusões das investigações alimentam uma base de conhecimento sobre o contexto e ao ser expandida pode informar quais ações tomar, indicando uma tomada de decisão.

Como os autores o modelam, o PPDAC é um ciclo composto por fases, que estão dispostas na Figura 3:

Figura 3 – Ciclo Investigativo - PPDAC



Fonte: Adaptado de Wild e Pfannkuch (1999, p. 226).

⁶⁷ A PPDAC cycle is concerned with abstracting and solving a statistical problem grounded in a larger "real" problem. Most problems are embedded in a desire to change a "system" to improve something. [...] A knowledge-based solution to the real problem requires better understanding of how a system works and perhaps also how it will react to changes to input streams, settings or environment.

Com a disposição das fases, é possível identificar o que supõe cada uma delas. O comportamento das setas na figura indicam que o conhecimento adquirido e as necessidades identificadas no ciclo podem iniciar novos ciclos investigativos. Com referência em Wild e Pfannkuch (1999), a seguir está descrito o que prevê cada fase do ciclo PPDAC.

Problema

Ao iniciar um ciclo, a definição dessa fase está ancorada em *um sistema de apreensão dinâmico*, no estabelecimento de *um problema* a ser estudado. A apreensão está ligada ao fato de buscarmos perceber o contexto em que estamos inseridos e a proposição de um problema. Um questionamento ao mundo que vemos certamente emerge dessa apreensão. Nesta tese, o problema é identificado pelos respectivos participantes em cada atividade proposta.

Plano

Refere-se a estabelecer um problema, a delimitar *um sistema de medição* para compreendê-lo, delimitar *uma amostra* de uma população que seja suficiente para representar o seu contexto em estudo, à *gestão dos dados* para permitir que eles atendam à demanda do problema e à *plotagem*, ou seja, a inserção de dados numa planilha ou a construção de uma representação para os dados com auxílio de um recurso, e à *análise* desses dados. Identificado o problema pelos participantes, um plano de ação é estabelecido para cada atividade.

Dados

Essa fase inclui *o recolhimento de dados*, reconhecendo que eles são essenciais para estudar o problema, *a gestão de dados* que decorrem do recolhimento a partir de uma amostra selecionada e *a limpeza de dados* que tem a ver com a seleção daqueles que são relevantes para o problema estabelecido. Nessa fase, os dados de cada atividade são gerenciados pelos participantes, de modo a permitir as respectivas análises.

Análises

Para essa fase, estão previstas *a exploração de dados* que pode permitir a observação de aspectos relevantes em relação ao problema, *as análises planejadas* que tem a ver com o que foi previsto para solucionar o problema, *as análises não planejadas* advindas da necessidade observada a partir do que os dados mostram e que não haviam sido planejadas previamente e *a geração de hipóteses* que indicam possíveis soluções para o problema inicial, considerando o que as análises dos dados permitem conjecturar. Na exploração dos dados de cada atividade, os participantes podem construir representações distintas para os dados e identificar os respectivos elementos relevantes.

Conclusões

A fase de conclusões prevê *a interpretação*, o que implica olhar para os dados como eles foram representados, mensurados, identificando as relações com o problema inicial. A partir dessa interpretação, são tiradas *conclusões* sobre o que emerge dos dados, surgindo *as novas ideias* que apontam para possíveis novos ciclos investigativos, baseados no conhecimento adquirido e nas necessidades identificadas. A *comunicação* é uma parte crucial dessa fase, pois permite apresentar uma possível solução para o problema, destacando aspectos do estudo que levaram à solução. Nessa última fase, os participantes têm a oportunidade de explicitar, respectivamente, a compreensão do problema abordado em cada atividade.

Como referencial metodológico, foram expostas as fases que compõem o ciclo investigativo cujo “objetivo final da investigação estatística é aprender no domínio do contexto de um problema real” (WILD; PFANNKUCH, 1999, p. 244, tradução nossa⁶⁸).

5.2 Momento histórico e restrições para a aplicação do estudo

O projeto para a realização deste estudo foi concebido no âmbito da linha de pesquisa “Tecnologias Digitais e Educação Matemática” do Programa de Estudo Pós-Graduados em Educação Matemática da Pontifícia Universidade Católica de São Paulo, no início do ano de 2019. Nas possíveis contribuições de recursos digitais para a análise gráfica de dados, estava

⁶⁸ The ultimate aim of statistical investigation is learning in the context domain of a real problem. Learning consists of the construction of mental models of the system under study.

prevista uma aplicação em aulas presenciais para o ensino médio, mas, devido à pandemia, foram necessárias adequações.

As adaptações decorreram do momento histórico vivido mundialmente. Desde o final do ano de 2019, a pandemia de COVID-19, causada pelo coronavírus do tipo SARS-CoV-2, espalhou-se por todos os continentes, resultando em medidas restritivas e necessárias, inclusive no território brasileiro, que impactaram o andamento desta pesquisa.

No intuito de conter e evitar um desastre sanitário – ainda maior – foram implantadas medidas, como a publicação da Portaria nº 188⁶⁹, em 3 de fevereiro de 2020, que declarou a emergência da pandemia e estabeleceu diretrizes de prevenção, controle e contenção de riscos, danos e agravos à saúde pública. Diante da situação de restrição de circulação da população, incluindo o fechamento de empresas, fábricas, comércios e instituições escolares, medidas foram tomadas para dificultar a propagação do vírus.

À medida que a situação continuou a se agravar, em 28 de abril de 2020, o Conselho Nacional de Educação – CNE aprovou um documento⁷⁰ com diretrizes para as escolas durante a pandemia, incluindo estratégias de ensino não presencial, o que comprometeu a proposta desta pesquisa.

As condições de trabalho para a implementação do estudo só permitiriam a aplicação em um contexto de ensino remoto. Desse modo, por meio de interações de pesquisadores do grupo de pesquisa e com incentivos institucionais de aprovação de projetos, ao longo do processo de desenvolvimento desta tese, foi viabilizada a aplicação da pesquisa em formato presencial em outro país, cujas condições da pandemia já estavam sob controle e eram permitidas aulas presenciais, obedecendo às normas de segurança sanitária e, também, com a comprovação de vacinação do pesquisador para viabilizar a viagem necessária.

5.2 As instituições envolvidas e os recursos subjacentes à realização do estudo

O desenvolvimento desta tese foi possível graças à inserção do pesquisador no Grupo de Pesquisa Tecnologias e Mídias em Educação Matemática - TecMEM, liderado pela orientadora deste estudo. A participação no TecMEM permitiu interagir com outros

⁶⁹ Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-188-de-3-de-fevereiro-de-2020-241408388>. Acesso em: 12 mar. 2020.

⁷⁰ Disponível em: <https://www.gov.br/pt-br/noticias/educacao-e-pesquisa/2020/06/mec-orienta-instituicoes-sobre-ensino-durante-pandemia>. Acesso em: mar. 2022.

pesquisadores e grupos de pesquisa internacionais, o que contribuiu para viabilizar a implementação do estudo.

Essa interação possibilitou o contato com pesquisadores da *Linz School of Education da Johannes Kepler Universität*, Áustria, por meio do professor, pesquisador e doutor Zsolt Lavicza, que colaborou fornecendo sugestões de leituras, referenciais teóricos e orientações para adequações desta tese. A partir dessa aproximação, foi realizado um estágio de doutorado junto ao grupo de pesquisa do professor Lavicza, entre outubro de 2021 e janeiro de 2022, com a utilização do idioma inglês, embora na Áustria algumas atividades e cursos fossem ministrados em alemão. Esse estágio contou com algum apoio financeiro do Plano de Incentivo à Pesquisa, por meio da orientadora, aprovados pela PUC-SP.

No projeto de estágio, além da colaboração e produção de pesquisas, foi definida a participação nas aulas de *Research Methods* (Métodos de Pesquisa) e *Statistics* (Estatística) oferecidas pelo professor Zsolt Lavicza, respectivamente, em níveis de pós-graduação e graduação. As aulas de Métodos de Pesquisa são oferecidas aos alunos dos cursos de mestrado e de doutorado da *Linz School of Education* e as aulas de Estatística são oferecidas para alunos de diferentes cursos de graduação para formação de professores, como matemática, geografia, educação física, etc.

A realização do estágio de doutorado promoveu a colaboração e produção de pesquisas, por meio da participação em congressos e publicação de trabalhos, como alguns citados no Capítulo 2. As contribuições das aulas de Métodos de Pesquisa permitiram uma aproximação com os estudos de outros pesquisadores, como, por exemplo, sobre a visualização de dados, conforme Prodromou e Dunne (2017) para pesquisas no campo da educação estatística. Sob a orientação do professor Zsolt Lavicza, ao participar das aulas de Estatística, foi possível organizar atividades, solicitadas pelo professor, que viabilizaram a pesquisa proposta nesta tese.

5.3 Contexto de implementação da pesquisa

Ao planejar a aplicação do estudo durante as aulas da disciplina de Estatística, foi necessário considerar a programação da disciplina, de modo a não afetar o andamento das aulas e garantir que as atividades propostas contribuíssem para a formação dos participantes.

A disciplina de Estatística, sendo uma introdução à matéria, teve uma carga horária de três horas semanais, divididas entre aulas teóricas ministradas em salas equipadas com

retroprojetores, lousas e espaço para toda a turma, e três sessões de 1 hora e 30 minutos dedicadas a atividades em grupos, permitindo que o professor dividisse a turma em equipes e organizasse apresentações e discussões relacionadas aos tópicos estudados. As sessões em grupos foram realizadas em salas menores, proporcionando aos participantes a oportunidade para realizar apresentações.

As aulas ocorreram às quartas-feiras, com duas sessões de aula teórica, a primeira iniciando às 8h30 e a segunda às 13h45, ambas com duração de 1 hora e 30 minutos. As sessões destinadas às atividades em grupos foram realizadas em três momentos, com início às 10h15, às 15h30 e às 17h15, respectivamente, todas com duração de 1 hora e 30 minutos.

Com o objetivo de promover a compreensão dos tópicos estudados, nos três encontros de atividades, o professor propunha uma abordagem que envolvia o estudo e a aplicação desses tópicos em diferentes contextos, o que viabilizou a implementação do projeto desta tese.

Conforme mencionado anteriormente, os estudantes de diferentes cursos de graduação estavam matriculados na disciplina de Estatística. Para a realização das atividades propostas ao longo da disciplina, os alunos foram organizados em duplas ou trios. Essa mesma organização, autorizada pelo professor Zsolt Lavicza (ANEXO) e pelos participantes, foi mantida durante a implementação das atividades propostas nesta pesquisa.

Em relação à problemática abordada nessa tese, ao longo da disciplina, o professor incentivou o uso de recursos digitais para representar os conceitos estatísticos nas aulas teóricas e durante os encontros. Nesse contexto, foram utilizados os recursos *Gapminder*, *GeoGebra* e *SPSS*.

5.4 Os recursos digitais utilizados

O projeto inicial desta tese visava explorar as possibilidades de análise gráfica de dados por meio do *software* *GeoGebra* (<https://www.geogebra.org/>), mas, conforme descrito no tópico anterior, durante as aulas e encontros da disciplina, foi observada a implementação de outros recursos digitais, como o *website* do *Gapminder* e o *software* *SPSS*, os quais foram considerados para aprimorar este estudo. Neste tópico, fornecemos informações sobre cada um desses recursos.

O *Gapminder* (<https://www.gapminder.org>) é uma fundação sueca independente, sem afiliações políticas, religiosas ou econômicas. Sua missão é promover uma visão de

mundo baseada em fatos, tornando-os acessíveis a todos. Como descrito em seu *website*, o *Gapminder* identifica equívocos sistemáticos sobre tendências e proporções globais importantes, e utiliza dados confiáveis para desenvolver materiais de ensino acessíveis. Seus recursos e materiais didáticos estão disponíveis gratuitamente.

No *website* do *Gapminder*, há uma variedade de materiais que apresentam dados sobre diversos temas, como educação, renda, pobreza, entre outros, fornecendo representações gráficas desses dados. Estudos como os de Andre, Lavicza e Prodromou (2019), Andre e Lavicza (2019) e Andre, Schneider e Frischemeier (2022) demonstram a viabilidade de utilizar o *Gapminder* como recurso em aulas de Estatística. Assim, considerando a organização da disciplina de Estatística oferecida pelo professor, esta pesquisa buscou analisar as possibilidades do *Gapminder* na análise gráfica de dados.

Em relação à implementação do GeoGebra, o *software* é descrito em seu *site* oficial como uma ferramenta dinâmica de matemática que abrange tópicos como geometria, álgebra, estatística, probabilidade em uma única plataforma, permitindo trabalhar com planilhas e construir representações gráficas de dados. O *software* é de acesso gratuito e possui versões que podem ser utilizadas *online* ou *offline*, sendo acessível também em dispositivos móveis. Além disso, o GeoGebra oferece uma plataforma *online* com mais de 1 milhão de recursos gratuitos criados pela comunidade em vários idiomas.

Estudos como os de Gonzales e Lopes (2017), Araújo (2018) e Souza e Calejon (2019) evidenciam implementações do GeoGebra em aulas de Estatística. Ao considerar o contexto da aplicação da pesquisa e a problemática inicial do projeto desta tese, a implementação do GeoGebra em atividades foi considerada para esta pesquisa.

O SPSS (<https://www.ibm.com/products/spss-statistics>) é um *software* estatístico desenvolvido pela IBM® que oferece ferramentas para análise estatística e é acessível a usuários de todos os níveis de habilidade. O uso do SPSS requer uma licença e não é um recurso tecnológico digital de acesso livre, todavia, a IBM® oferece um período experimental gratuito de um mês para utilização do *software*. Conforme destaca em seu site oficial, o SPSS é de fácil utilização e possui uma ampla biblioteca de algoritmos de *machine learning*, análise de texto, extensibilidade de *software* livre e integração com *big data*, permitindo abordar problemas em diferentes níveis.

O SPSS tem sido utilizado em pesquisas para análise de dados. No contexto do ensino de Estatística, estudos como Strelhow e Câmara (2011) e Gomes (2016) apontam o uso do

SPSS como uma possibilidade. Nesta tese, são investigadas as contribuições da inserção desse recurso para a análise gráfica de dados.

Neste tópico, foram apresentados os recursos digitais utilizados na disciplina de Estatística oferecida pelo professor Zsolt Lavicza cuja organização permitiu que as atividades propostas para o experimento desta tese fossem viabilizadas, seguindo o referencial metodológico da dimensão do ciclo investigativo - PPDAC, conforme Wild e Pfannkuch (1999).

5.5 A proposição das atividades

Como parte integrante da disciplina de Estatística ministrada pelo professor Zsolt Lavicza, o experimento desta tese foi conduzido em inglês, com a respectiva tradução para o português neste trabalho. O experimento consistiu na proposição de três atividades, cada uma utilizando um dos recursos: *Gapminder*, GeoGebra e SPSS, e todos os participantes tiveram a oportunidade de realizar as atividades.

O número total de alunos matriculados na disciplina foi de 72, denominados nesta tese como “futuros professores”, pois se tratava de estudantes de cursos de formação inicial de professores.

As três atividades propostas, chamadas de “*Homework*”, são apresentadas neste trabalho em inglês, seguidas pela tradução para o português. Tais atividades, de caráter investigativo e exploratório, permitiam que os futuros professores, organizados em duplas ou trios, abordassem problemas relevantes para eles, realizassem pesquisas e fizessem a limpeza e organização dos dados relacionados aos problemas escolhidos. A proposição das atividades seguiu o ciclo investigativo PPDAC, considerando a importância de abordar problemas significativos para os participantes, e fornecia sugestões de acesso e utilização dos recursos.

A primeira atividade proposta solicitava que os participantes acessassem o *website* do *Gapminder* e realizassem uma atividade de exploração de dados, utilizando as representações disponíveis dos dados relacionados aos problemas escolhidos pelas duplas ou trios. A Figura 2 apresenta a solicitação da primeira atividade em inglês.

Figura 2 – Atividade proposta com o *Gapminder*

Homework 1

Gapminder Statistical Explorations

1. Watch some videos on the Gapminder website, for example, these two: <https://www.gapminder.org/fw/world-health-chart/> <https://www.gapminder.org/answers/how-did-the-world-population-change/> with the explanation of Hans Rosling, to get an idea how Gapminder visualization works.

2. Explore some of the Gapminder tools: <https://www.gapminder.org/tools/> and try to select some interesting topics for explorations (can be countries, different data, different tools, etc.)

3. Write up two exploration topics that you found interesting and give explanation of the charts that you selected. Use some screenshots to illustrate your ideas, write some key ideas and explanations, but you can explain more in detail in your presentation.

Fonte: Dados da pesquisa.

O Quadro 4 apresenta uma tradução da atividade proposta com a inserção do *Gapminder*. Essa atividade foi implementada na parte inicial da disciplina e seu objetivo era permitir que os futuros professores, ao interagirem com as ferramentas presentes no recurso tecnológico digital, fizessem a escolha de dois tópicos de exploração e as respectivas análises gráficas dos dados selecionados, seguindo as fases metodológicas do ciclo investigativo – PPDAC, conforme Wild e Pfannkuch (1999).

Quadro 4 – Tradução da atividade proposta com o *Gapminder*

Homework 1

1. Assista a alguns vídeos no site da *Gapminder*, por exemplo, estes dois: <https://www.gapminder.org/fw/world-health-chart/> <https://www.gapminder.org/answers/how-did-the-world-population-change/> com a explicação de Hans Rosling, para ter uma ideia de como funciona a visualização do *Gapminder*.

2. Explore algumas das ferramentas do *Gapminder*: <https://www.gapminder.org/tools/> e tente selecionar alguns tópicos interessantes para explorações (podem ser países, dados diferentes, ferramentas diferentes etc.)

3. Escreva dois tópicos de exploração que você achou interessante e dê explicações das representações gráficas selecionadas. Use algumas capturas de tela para ilustrar suas ideias, escreva algumas ideias e explicações importantes, mas você pode explicar mais detalhadamente em sua apresentação.

Fonte: Dados da pesquisa.

As sugestões para assistir a vídeos e explorar as ferramentas tiveram como objetivo permitir a familiarização dos participantes com o *Gapminder*. Ao considerar a possibilidade de identificar variáveis e estudos baseados em dados relacionados a elas, a atividade proposta buscava promover explorações, seleção de tópicos (temas envolvendo as variáveis estudadas) e dissertações sobre as representações gráficas, por meio da interação com as ferramentas do Gapminder. A proposta visava incentivar uma análise exploratória, em que a análise gráfica dos dados seria um processo a ser realizado para contribuir para a compreensão das variáveis investigadas.

A Figura 3 apresenta a atividade proposta a partir da implementação do GeoGebra para o estudo de dados. Como citado, ao longo das aulas foram apresentadas algumas ferramentas do GeoGebra para que os futuros professores pudessem observar algumas possibilidades de utilizar o recurso tecnológico digital com objetos estatísticos. A atividade tinha como objetivo que os participantes pudessem importar, organizar, construir representações de dados mediados pelo GeoGebra e suas respectivas análises, novamente, seguindo as fases metodológicas do ciclo investigativo – PPDAC, conforme Wild e Pfannkuch (1999).

Figura 3 – Atividade proposta com o GeoGebra

Homework 2

GeoGebra Statistical Explorations

1. Explore some GeoGebra statistics tools, similar that were presented by Martin and Ronaldo in the class:

<https://www.geogebra.org/t/statistics?lang=en>

some very good examples

<https://www.geogebra.org/m/g3vLRuB8>

2. Find some datasets online that are interesting for you, here are some useful websites:

<https://www.gapminder.org/tools/>

<https://covid19.who.int/>

<https://www.statista.com/statistics/1109188/coronavirus-confirmed-cases-by-age-group-austria/>

<https://data.humdata.org/event/covid-19?>

<https://www.data.gv.at/covid-19/>

3. Show some visualization with GeoGebra and offer some explanations, similarly as you did with Gapminder

Fonte: Dados da pesquisa.

A proposta da atividade (Quadro 5) consistia em que os futuros professores explorassem as ferramentas e se apropriassem de procedimentos e funcionalidade do GeoGebra para análise gráfica de dados.

Quadro 5 – Tradução da atividade proposta com o GeoGebra

Homework 2

Explorações Estatísticas com GeoGebra

1. Explore algumas ferramentas estatísticas do GeoGebra, semelhantes às apresentadas por Martin e Ronaldo nas aulas: <https://www.geogebra.org/t/statistics?lang=en>. Alguns exemplos muito bons <https://www.geogebra.org/m/g3vLRuB8>
2. Encontre alguns conjuntos de dados *online* que são interessantes para você, aqui estão alguns sites úteis: <https://www.gapminder.org/tools/>
<https://covid19.who.int/>
<https://www.statista.com/statistics/1109188/coronavirus-confirmed-cases-by-age-group-austria/>
<https://data.humdata.org/event/covid-19?>
<https://www.data.gv.at/covid-19/>
3. Mostre algumas visualizações com o GeoGebra e apresente algumas explicações, do mesmo modo que você fez com o *Gapminder*.

Fonte: Dados da pesquisa.

Os *links* apresentados aos participantes objetivaram favorecer essa aproximação com o *software* e com possíveis fontes de dados para realização de estudos estatísticos de forma a possibilitar as respectivas análises.

Figura 4 – Atividade proposta com o SPSS

Homework 3

SPSS Statistical Explorations – data preparation – correlations

1. Find some datasets online that are suitable for exploring correlations (Parametric: continuous, normal), you can find this data anywhere, but these are some source websites:
<https://www.gapminder.org/tools/>
<https://data.worldbank.org/>
<https://www.data.gv.at;>
2. Import at least 3 variables (suitable to explore correlations) from excel to SPSS and set data description in the Data View;
3. Run correlation analyses (0-ask some questions why these correlations can be interesting; 1- check assumptions; 2- state hypotheses; 3 – calculate suitable statistics numbers to check strength of relationships, significance, effect size;
4. interpret results, be careful correlations do not mean causation) – you should have correlations of at least 3 pairs of variables.

Fonte: Dados da pesquisa.

A Figura 4 apresenta a terceira atividade, na qual era proposta a implementação do SPSS. Ao considerar o momento dos futuros professores na disciplina, o estudo de correlações foi uma das possibilidades de utilizar o *software*. Assim, o objetivo dessa atividade era executar análises de correlação de pelo menos 3 pares de variáveis e interpretar os resultados, procurando atender as fases do ciclo investigativo – PPDAC, conforme Wild e Pfankuch (1999).

Assim como nos casos de utilização do *Gapminder* e do GeoGebra, ao longo das aulas teóricas desenvolvidas com os participantes, foram apresentados recursos do SPSS. A atividade três (Quadro 6) considerou esse *software* como outra ferramenta para o estudo de objetos estatísticos por meio de problemas envolvendo correlações.

Quadro 6 – Tradução da atividade proposta com SPSS

Homework 3	
Explorações estatísticas com o SPSS – preparação de dados – correlações	
1.	Encontre alguns conjuntos de dados <i>online</i> que são adequados para explorar correlações (Paramétrico: contínuo, normal), você pode encontrar esses dados em qualquer lugar, mas estes são alguns sites de origem: https://www.gapminder.org/tools/ https://data.worldbank.org/ https://www.data.gv.at;
2.	Importar pelo menos três variáveis (adequadas para explorar correlações) do Excel para o SPSS e definir a descrição desses dados na janela de visualização de dados;
3.	Execute análises de correlação (0- faça algumas perguntas porque essas correlações podem ser interessantes; 1- verifique suposições; 2- estabeleça hipóteses; 3- calcule valores estatísticos adequados para verificar a força das relações, significância, tamanho do efeito);
4.	Interprete os resultados e tenha cuidado, pois correlações não significam causalidade – você deve ter correlações de pelo menos 3 pares de variáveis.

Fonte: Dados da pesquisa.

Foram sugeridos alguns *links* que permitissem o acesso a banco de dados, orientações quanto à seleção de variáveis e procedimentos para os respectivos estudos de correlação, a partir dos dados importados. Como indicado na atividade, a planilha do *Excel* foi utilizada como intermédio no processo de importação de dados.

Neste tópico, foram apresentados os procedimentos que viabilizaram a aplicação das atividades da pesquisa, utilizando os recursos digitais *Gapminder*, GeoGebra e SPSS. As atividades foram realizadas no contexto da disciplina de Estatística, voltada para futuros professores, e possibilitaram a investigação de possibilidades de análise gráfica de dados,

objetivo desta tese. É relevante salientar que a disciplina estatística é obrigatória e pode ser frequentada em qualquer momento da formação.

5.6 *Corpus* do material de análise da pesquisa

Ao descrever o contexto em que as atividades foram aplicadas, destaca-se que a turma era composta por 72 futuros professores, que trabalharam em duplas ou trios para realizar as atividades propostas nesta pesquisa. As atividades, chamadas de *Homework*, foram desenvolvidas por 28 grupos, que exploraram os recursos digitais e escolheram problemáticas a serem estudadas com base em dados disponíveis.

Durante as aulas, os grupos apresentaram seus relatórios, revelando o percurso investigativo realizado (ciclo investigativo). Entretanto, para fins de análise, foram selecionados alguns desses relatórios como material de estudo. Os relatórios foram organizados em pastas individuais para cada grupo e, em seguida, foram lidos e avaliados. A seleção do material levou em consideração a adequada utilização dos recursos digitais e a realização das análises solicitadas.

O recorte metodológico inicial definiu que relatórios de três grupos poderiam ser incluídos no *corpus* do material de análise dessa tese, visto que esses grupos mantiveram a composição de participantes ao longo das atividades e apresentaram resultados relevantes para os objetivos deste estudo.

No entanto, o *corpus* do material apresentado nessa pesquisa consiste nos resultados dos relatórios de dois grupos, visto que os participantes se adequaram melhor aos objetivos desta tese que aborda as possibilidades de análise gráfica de dados mediadas por recursos digitais. A escolha por analisar os relatórios das duas duplas atende aos objetivos da tese, pois permaneceram juntas em todo o trajeto da pesquisa. Os trabalhos desenvolvidos pelos demais grupos serão objetos de investigações futuras.

Apresentamos neste capítulo os procedimentos metodológicos para a implementação da parte experimental desta tese, decorrente do referencial metodológico adotado, ciclo investigativo – PPDAC. No próximo capítulo, são apresentados o desenvolvimento das atividades de dois grupos e os respectivos resultados, analisados à luz dos referenciais teóricos da transnumeração e da visualização de dados.

6 DESENVOLVIMENTO, RESULTADOS E ANÁLISES DAS ATIVIDADES

Neste capítulo, apresentamos o desenvolvimento das atividades, tomando como referência as fases do ciclo investigativo - PPDAC (WILD; PFANNKUCH, 1999). Os resultados obtidos consideram a implementação dos recursos tecnológicos *Gapminder*, GeoGebra e SPSS, bem como as análises realizadas sob o apoio dos estágios de transnumeração (PFANNKUCH; RUBICK; YOUN, 2002) e da visualização de dados, especialmente elementos do letramento estatístico na era *OPEN Data*, conforme definido por Prodromou e Dunne (2017). A organização do capítulo segue a sequência de aplicação das respectivas atividades, com o suporte do *Gapminder*, GeoGebra e SPSS, utilizando diferentes versões dos *softwares* em cada dupla.

6.1 Atividades desenvolvidas com o *Gapminder*

Dupla 1 - *Gapminder*

O relatório da **Dupla 1** traz uma contextualização da exploração no *website* do *Gapminder*. Seguindo as orientações da *Homework 1* e utilizando os *links* sugeridos, os futuros professores realizaram uma primeira análise exploratória sobre temáticas que o *Gapminder* aborda, a partir de dados. Com base nas representações gráficas e nas explicações de Hans Roseling⁷¹, os dois trechos seguintes apresentam uma análise descritiva realizada pela **Dupla 1**:

The first video can be found under this link “<https://www.gapminder.org/fw/world-health-chart/>”, in this video one can realize that the income, the world population and the life expectancy increased with the passing of years. Apart from that one can also observe that China and India are the countries with the highest population in 2019. Besides to this fact, Japan and Singapore have the highest life expectancy of an average of 84,8 years in 2019.

⁷¹ Fundador da Fundação Gapminder. Disponível em: https://www.ipea.gov.br/desafios/index.php?option=com_content&view=article&id=3041&catid=30&Itemid=41. Acesso em: mar. 2023.

Tradução: *O primeiro vídeo pode ser encontrado no link "<https://www.gapminder.org/fw/world-health-chart/>", no qual se verifica que a renda, a população mundial e a expectativa de vida aumentaram com o passar dos anos. Além disso, pode-se também observar que a China e a Índia são os países com maior população em 2019. Além desse fato, Japão e Cingapura têm a maior expectativa de vida de uma média de 84,8 anos em 2019.*

As observações da **Dupla 1** sobre as variáveis renda, população e expectativa de vida em nível mundial são resultado de uma primeira aproximação com o *website*. As análises são baseadas em bancos de dados de escala relativamente grande, considerando que se trata de dados em nível global. Como observado por Prodromou e Dunne (2017), essa é a natureza dos dados na era *OPEN Data*.

Ao indicar a elevação dos valores das variáveis ao longo do tempo e destacar dados como os países com maior população e expectativa de vida mais elevada, no Japão e em Singapura, é possível perceber que a **Dupla 1** foi capaz de realizar uma análise exploratória inicial dos dados representados e discutidos, utilizando as representações gráficas disponíveis

The second video, which can be found under this link "<https://www.gapminder.org/answers/how-did-the-world-population-change/>" shows, how the size of the world population has changed over time. The Professor Hans Rosling affirms that in year -8.000, there were 10 million people on earth. From the beginning of agriculture to the industrial revolution the population growth was very slow. Then the population started to grow rapidly because fewer died young. Today, most women have fewer children and therefore the rapid growth will soon be slowing down. World population will most likely stabilize around 11 billion towards the end of the century.

Tradução: *O segundo vídeo, que pode ser encontrado sob este link "<https://www.gapminder.org/answers/how-did-the-world-population-change/>" mostra como o tamanho da população mundial mudou ao longo do tempo. O professor Hans Rosling afirma que no ano 8.000 A. C. havia 10 milhões de*

peças na Terra. Desde o início da agricultura até a revolução industrial, o crescimento populacional foi muito lento. Então a população começou a crescer rapidamente porque menos morreram jovens. Hoje, a maioria das mulheres tem menos filhos e, portanto, o rápido crescimento em breve desacelerará. A população mundial provavelmente se estabilizará em torno de 11 bilhões no final do século.

O segundo trecho da análise exploratória inicial da **Dupla 1** mostra considerações sobre o aumento da população mundial ao longo do tempo. Como pontuado em Hans Rosling, no vídeo, ao fazer uso de uma representação gráfica, a **Dupla 1** destaca alguns eventos históricos que explicam o crescimento da população mundial.

Sob a perspectiva da visualização de dados, de acordo com Prodromou e Dunne (2017), essa *compreensão* inicial mobilizada pelos futuros professores, considerando as representações gráficas apresentadas nos vídeos, também evidencia a *finalidade* das representações, permitindo a *verbalização* de ideias relacionadas aos temas abordados.

Com base na definição de transnumeração de Wild e Pfannkuch (1999) e nas representações gráficas apresentadas nos vídeos, é possível observar como resultado das mudanças nas representações de grandes conjuntos de dados uma compreensão do contexto em que as variáveis são discutidas pela **Dupla 1**.

No que diz respeito à organização metodológica da atividade, ainda é possível observar que a análise exploratória inicial foi um momento que precedeu a delimitação dos problemas a serem abordados, de acordo com a primeira fase do ciclo investigativo – PPDAC, problemas definidos a partir da segunda solicitação da atividade *Homework 1*.

A exploração das ferramentas do *Gapminder* (<https://www.gapminder.org/tools/>), na segunda solicitação da atividade, tinha como objetivo permitir a seleção de tópicos a serem explorados à luz das fases do ciclo investigativo - PPDAC, conforme apresentado na descrição e análise dos resultados obtidos ao acessar o relatório da **Dupla 1**, resultado da terceira solicitação da atividade.

First topic: Number of people by income and percentage of the human population living in extreme poverty. The first topic, we have chosen handles about the number of people by income and about the percentage of the human population living in extreme poverty. There is a video about this topic on the Gapminder

website, which can be found under this link "[https://www.gapminder.org/tools/#\\$chart-type=mountain&url=v1](https://www.gapminder.org/tools/#$chart-type=mountain&url=v1)".

Tradução: *Primeiro tópico: Número de pessoas por rendimento e porcentagem da população humana que vive em extrema pobreza. O primeiro tópico, que escolhemos, trata sobre o número de pessoas por renda e sobre a porcentagem da população humana que vive em extrema pobreza. Há um vídeo sobre este tópico no site da Gapminder, que pode ser encontrado neste link "[https://www.gapminder.org/tools/#\\$chart-type=mountain&url=v1](https://www.gapminder.org/tools/#$chart-type=mountain&url=v1)".*

A exploração das ferramentas do *Gapminder* pela **Dupla 1** permitiu estudar o número de pessoas, por extrato de renda e a taxa populacional em situação de extrema pobreza. Como a transnumeração é um tipo de pensamento estatístico que ocorre durante todo o ciclo investigativo, a delimitação do problema a ser estudado é que definirá medidas quantitativas ou de classificação relevantes dessa situação real, que, conforme Pfannkuch, Rubick e Yoon (2002), necessárias no primeiro estágio da transnumeração.

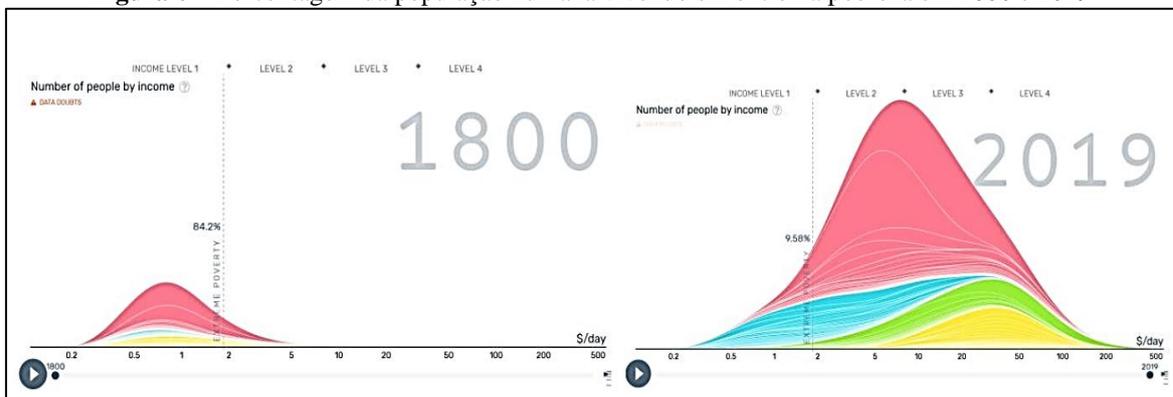
A definição de medidas quantitativas está relacionada às possibilidades de inferências plausíveis presentes nos dados, conforme destacado por Prodromou e Dunne (2017), ao relacionar essas inferências com a familiaridade de várias ideias fundamentais, como conjuntos de dados multivariados, visualização, procedência e qualidade dos dados, especialmente *OPEN Data*.

It shows the development of the percentage of the human population living in extreme poverty from 1,800 to 2,019. In 1,800 84.2% of the human population was living in extreme poverty and in 2,019 there were only 9.58%, therefore a reduction of 74.62% took place. "[https://www.gapminder.org/tools/#\\$chart-type=mountain&url=v1](https://www.gapminder.org/tools/#$chart-type=mountain&url=v1)) (see Diagram 1). (Figura 6).

Tradução: *Esse (tópico) mostra a progressão da porcentagem da população humana que vive em extrema pobreza de 1800 para 2019. Em 1800, 84,2% da população humana vivia em extrema pobreza e em 2019 havia apenas 9,58%, portanto ocorreu uma redução de 74,62%.*

“[https://www.gapminder.org/tools/#\\$chart-type=mountain&url=v1](https://www.gapminder.org/tools/#$chart-type=mountain&url=v1)”
(ver Diagrama 1). (Figura 6).

Figura 6 – Porcentagem da população humana vivendo em extrema pobreza em 1800 e 2019



Fonte: Dados da pesquisa.

Ao utilizar o Gapminder, como citado na apresentação da atividade, a fase de planejamento e levantamento de dados para estudar o problema escolhido (segunda fase do ciclo investigativo – PPDAC) poderia acontecer de forma interna no *website*. Portanto, ocorreram de forma implícita a delimitação do problema, a seleção, a limpeza e a gestão dos dados, conforme a interação da **Dupla 1** com o *Gapminder*, o que viabilizou o trabalho em relação ao recolhimento de dados, necessário para a terceira fase do ciclo investigativo – PPDAC. Essas operações implícitas, realizadas pelo *Gapminder*, podem ser exploradas à luz do letramento estatístico na era *OPEN Data*, considerando aspectos como a veracidade dos dados e as possibilidades de acesso a eles, que podem ser consultados no *Gapminder*.

Ao selecionar as variáveis abordadas, número de pessoas por extrato de renda e taxa populacional em situação de extrema pobreza, na escolha da forma de representação, o *Gapminder* possibilitou aos dois futuros professores acessarem a representação gráfica dos dados pertinentes para uma análise (quarta fase do ciclo investigativo – PPDAC).

É possível observar que a representação gráfica utilizada pela Dupla 1 tem potencial interativo por meio do recurso tecnológico digital, embora não esteja evidente na Figura 6, para estabelecer datas de referência e conduzir as análises. É possível observar a variação percentual da população em situação de extrema pobreza ao longo de toda a série temporal, considerando diferentes regiões ou países ao redor do mundo.

A escolha de uma métrica e uma representação gráfica dos dados adotadas pela **Dupla 1** para discutir o problema (porcentagem da população em situação de extrema pobreza) entre 1800 e 2019 possibilitou a construção de uma análise e o entendimento do

que os dados estão mostrando sobre o problema abordado, caracterizando, de acordo com Pfannkuch, Rubick e Yoon (2002), o segundo estágio da transnumeração.

O aspecto da linguagem, apontado por Prodromou e Dunne (2017) como relevante para o letramento estatístico na era OPEN Data, pode ser observado por meio da representação gráfica dos dados relacionados ao problema. É por meio dessa forma de apresentação dos dados, juntamente com a identificação de datas específicas conforme o objetivo da dupla, que surge o discurso escrito: “Em 1800, 84,2% da população humana vivia em extrema pobreza e em 2019 havia apenas 9,58%, portanto ocorreu uma redução de 74,62%”. No relatório dessa dupla o discurso adequado seria substituir 74,62% por uma redução percentual de 74,62 pontos.

Esse discurso escrito pode ser visto como uma comunicação compreensível e persuasiva sobre a situação original do problema abordado, correspondendo à terceira etapa da transnumeração proposta por Pfannkuch, Rubick e Yoon (2002). É nesse discurso que se observa uma conclusão (quinta fase do ciclo investigativo - PPDAC) sobre o que a **Dupla 1** propôs investigar no primeiro tópico.

No segundo tópico, a **Dupla 1** abordou a porcentagem da população mundial por idade:

The second topic, we have chosen handles about the development of the percentage of the world population by age from 1950 till 2100, as we have found a video about it on the Gapminder website. We have chosen three representative years, namely 1950, 2021 and 2100 in order to compare the distribution of the percentage of the world population by age.

Tradução: Para o segundo tópico, escolhemos tratar sobre a progressão da porcentagem da população mundial por idade de 1950 a 2100, pois encontramos um vídeo sobre isso no site do Gapminder. Escolhemos três anos representativos, a saber, 1950, 2021 e 2100, para comparar a distribuição do percentual da população mundial por idade.

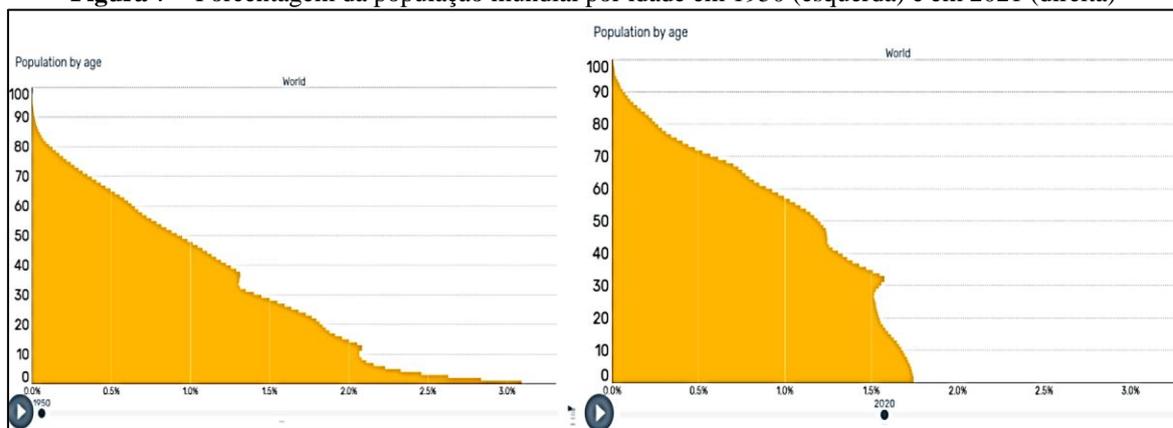
No trecho apresentado, é possível perceber que a **Dupla 1** demonstra interesse em estudar como a taxa percentual da população por faixa etária tem se comportado ao longo do tempo. Essa abordagem reflete a primeira fase do ciclo investigativo mencionado neste tópico. Assim como no primeiro tópico, também é observado o interesse da **Dupla 1** em

estudar dados de uma variável em nível mundial, o que está alinhado com as considerações de Prodromou e Dunne (2017) sobre a importância de abordar a estatística considerando o contexto de *OPEN Data*.

Ao limitar o contexto do estudo à idade da população, com base na variação da taxa percentual, a **Dupla 1** explicita a métrica adotada para compreender o problema escolhido, o que corresponde ao primeiro estágio da transnumeração, conforme Rubick, Pfannkuch e Yoon (2002).

Assim como no primeiro tópico, as fases de planejamento e trabalho com os dados, relacionadas à segunda e terceira fase do ciclo investigativo - PPDAC, respectivamente, envolvem operações implícitas realizadas pelo *Gapminder*⁷², dependendo das escolhas e interações dos usuários com a ferramenta tecnológica. A Figura 7 apresenta capturas de tela das representações gráficas dos dados possibilitadas pelo *Gapminder*, resultantes das interações da **Dupla 1**.

Figura 7 – Porcentagem da população mundial por idade em 1950 (esquerda) e em 2021 (direita)

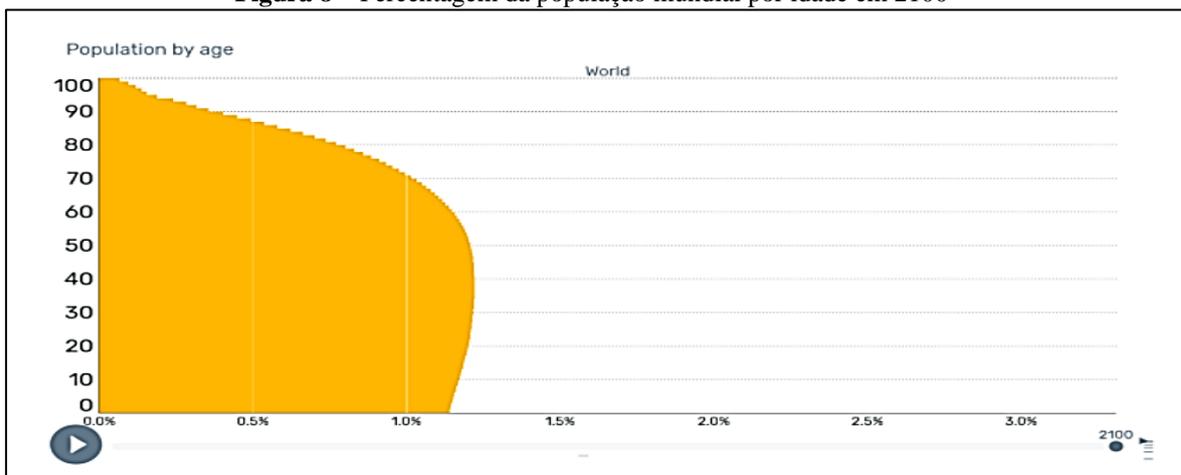


Fonte: Dados da pesquisa.

⁷² Disponível em:

[https://www.gapminder.org/tools/#\\$model\\$markers\\$pyramid\\$encoding\\$frame\\$value=2100&playbackSteps:1;;;&chart-type=popbyage&url=v1](https://www.gapminder.org/tools/#$model$markers$pyramid$encoding$frame$value=2100&playbackSteps:1;;;&chart-type=popbyage&url=v1). Acesso em: fev. 2022.

Figura 8 – Percentagem da população mundial por idade em 2100



Fonte: Dados da pesquisa.

As Figuras 7 e 8 apresentadas são resultados da fase de trabalho com os dados. É possível observar que a **Dupla 1** optou por estabelecer anos de referência para investigar a variação da taxa percentual da população por faixa etária. No relatório, a **Dupla 1** disponibiliza o *link* de acesso às representações gráficas dos dados delimitados por suas escolhas no ambiente do *website* do *Gapminder*.

As representações gráficas obtidas permitem compreender o que os dados disponíveis na plataforma do *Gapminder* revelam sobre o problema abordado, conforme o referencial de Pfannkuch, Rubick e Yoon (2002), etapa que corresponde ao segundo estágio de transnumeração. Assim, com base nessas representações gráficas dos dados, a **Dupla 1** apresenta uma análise:

By comparing these three diagrams one can come to the conclusion that the percentage of the old people of the world have increased with the pass of the time and according to the Gapminder website it will continue raising. Besides one can also realize that the percentage of the babies of the world have decreased with the passing of time and according to the Gapminder website it will continue reducing (see Diagram 2 and Diagram 3).

Tradução: *Ao comparar esses três diagramas, pode-se chegar à conclusão de que a porcentagem dos idosos do mundo aumentou com o passar do tempo e, de acordo com o site Gapminder, continuará aumentando. Além disso, pode-se também perceber que a porcentagem dos bebês do mundo diminuiu com o passar*

do tempo e, de acordo com o site Gapminder, continuará reduzindo (ver Diagrama 2 e Diagrama 3). (Figura 7 e Figura 8).

A análise apresentada pela **Dupla 1** evidencia uma *compreensão*, no contexto de visualização de dados, em conformidade com Prodromou e Dunne (2017). A **Dupla 1** se apoiou na produção de uma representação visual dos dados para tecer considerações sobre o problema, na perspectiva dos autores supracitados, a qual revela uma *finalidade* atendida quanto à visualização disponibilizada pelo *Gapminder*. As representações gráficas dos dados, apresentadas nas Figuras 7 e 8, oferecem a possibilidade para os futuros professores, na fase de análise do ciclo investigativo – PPDAC, interagirem com a representação visual de dados. À luz da visualização de dados, essa fase de análise elucida o aspecto da *verbalização* no discurso do relatório.

Ao concluir “*que a porcentagem dos idosos do mundo aumentou com o passar do tempo e, de acordo com o site Gapminder, continuará aumentando*” e “*que a porcentagem dos bebês do mundo diminuiu com o passar do tempo e, de acordo com o site Gapminder, continuará reduzindo*”, o relatório da **Dupla 1** revela que as representações gráficas dos dados, visualizadas pelos dois futuros professores, foram relevantes.

Na fase de conclusão do ciclo investigativo - PPDAC do segundo tópico, a **Dupla 1** buscou realizar uma comparação da variação da taxa percentual da população mundial por faixa etária ao longo do tempo. Para isso, as três representações gráficas dos dados forneceram possibilidades para essa análise comparativa.

É relevante observar que a **Dupla 1** identifica e compara tendências em determinadas faixas etárias e se utiliza do parâmetro de porcentagem estabelecido, o qual, à luz da transnumeração proposta por Pfannkuch, Rubick e Yoon (2002), no seu terceiro estágio, contribuiu para a comunicação de forma compreensível e convincente sobre os achados do problema abordado.

Sob o ponto de vista da visualização de dados, as *verbalizações* apresentadas no relatório da **Dupla 1** demonstram clareza e coerência em relação aos dados do problema abordado. A utilização da taxa percentual como forma de fornecer precisão à análise realizada, como mencionado pelos próprios futuros professores, juntamente com a representação gráfica mostrada na Figura 8, possibilitam extrapolar e apresentar algumas previsões sobre o problema estudado.

Com referência aos elementos do letramento estatístico na era *OPEN Data*, a escolha por estabelecer três momentos diferentes para observar o comportamento da variável em estudo pode ter permitido aos futuros professores visualizarem *inferências plausíveis dentro dos dados*. Um exemplo é a inferência do aumento porcentual da população idosa e da redução da porcentual da população de recém-nascidos, ao longo do tempo.

Quanto à *linguagem*, as representações gráficas dos dados, acessadas pela **Dupla 1**, no segundo tópico, assim como no primeiro, foram relevantes para o engajamento visual e favoreceram as estratégias na busca de compreensão dos dados representados.

Considerações sobre *o contexto, as mudanças e a causalidade* observadas no relatório da **Dupla 1** têm a ver com o fato de os participantes da dupla delimitarem espaços temporais, datas específicas para compararem e apontarem fatores que revelam as mudanças nas variáveis envolvidas no problema.

Do ponto de vista dos elementos do letramento estatístico na era *OPEN Data*, não há um questionamento sobre os dados que o *Gapminder* utiliza, mas o conhecimento do contexto poderia permitir aos futuros professores uma apreciação de pré-condições latentes na estrutura e na composição dos dados.

Ao permitir observar tendências sobre o comportamento das variáveis envolvidas no problema, as representações gráficas dos dados podem ser vistas como representações de *modelos estatísticos*, como é o caso da previsão da taxa porcentual por faixa de idade, previstas para 2100, apresentada no relatório da **Dupla 1**.

Sob o ponto de vista da questão de pesquisa desta tese, a utilização do *Gapminder* contribuiu para a realização de análises exploratórias a partir das representações gráficas. Um exemplo dessa visualização foi a análise inicial realizada pela **Dupla 1** sobre o aumento da expectativa de vida, renda e população humana ao longo do tempo.

A possibilidade de interagir com o *website*, selecionando as variáveis a serem representadas graficamente e escolhendo as formas de representação, é um dos benefícios oferecidos pelo *Gapminder*. Apesar de não ser evidente nas telas apresentadas pela dupla de futuros professores, o potencial interativo do recurso tecnológico digital permite visualizar o comportamento dos dados em relação à variável estudada ao longo do tempo.

Os resultados de análises realizadas pela **Dupla 1**, ao problematizarem sobre o porcentual da população humana vivendo em extrema pobreza, representados pelas construções gráficas do *Gapminder* (Figura 6), permitem observar os resultados da escolha de variáveis e das representações dos dados sobre elas. Além disso, as diferentes cores nas

representações visualizadas, que indicam as regiões do mundo, podem favorecer a compreensão sobre onde há maior concentração de pessoas na condição estudada.

No estudo sobre a progressão da população humana, estudado pela **Dupla 1**, as representações gráficas (Figuras 7 e 8) são exemplos de como, a partir do acesso aos dados, o *Gapminder* possibilitou a realização de previsões, além da observação do comportamento desses dados ao longo do tempo. São essas ferramentas e o potencial interativo, observados ao analisar o relatório da dupla de futuros professores, que revelam as possibilidades de aprimoramentos nas análises gráficas dos dados, permitidas pela utilização do *Gapminder* na *Homework 1*.

Dupla 2 - Gapminder

Ao acessar o relatório da **Dupla 2**, os resultados da exploração inicial no *website* do *Gapminder* trazem detalhes sobre a percepção dos futuros professores, ao interagirem com o recurso tecnológico digital integrado na atividade.

In general, the idea of the website Gapminder is to show statistical data of countries all over the world. All the 195 countries are represented in the different charts (that the website provides) as dots (in different sizes).

Tradução: *Em geral, a ideia do site Gapminder é mostrar dados de países de todo o mundo. Todos os 195 países estão representados nos diferentes gráficos (que o site fornece) como pontos (em diferentes tamanhos).*

Partindo dessa compreensão geral, a **Dupla 2** explicita ferramentas que podem auxiliar a análise de dados com a utilização do *Gapminder*.

Figura 9 – Cores representando diferentes regiões do planeta



Fonte: Dados da pesquisa.

Ao considerar cores diferentes para as regiões do planeta (Figura 9), para uma análise gráfica de dados sobre algum tópico, é possível identificar concentrações, tendências ou disparidades, a depender da região, que pode ser um aspecto relevante das ferramentas visuais do *Gapminder*. A **Dupla 2** explica:

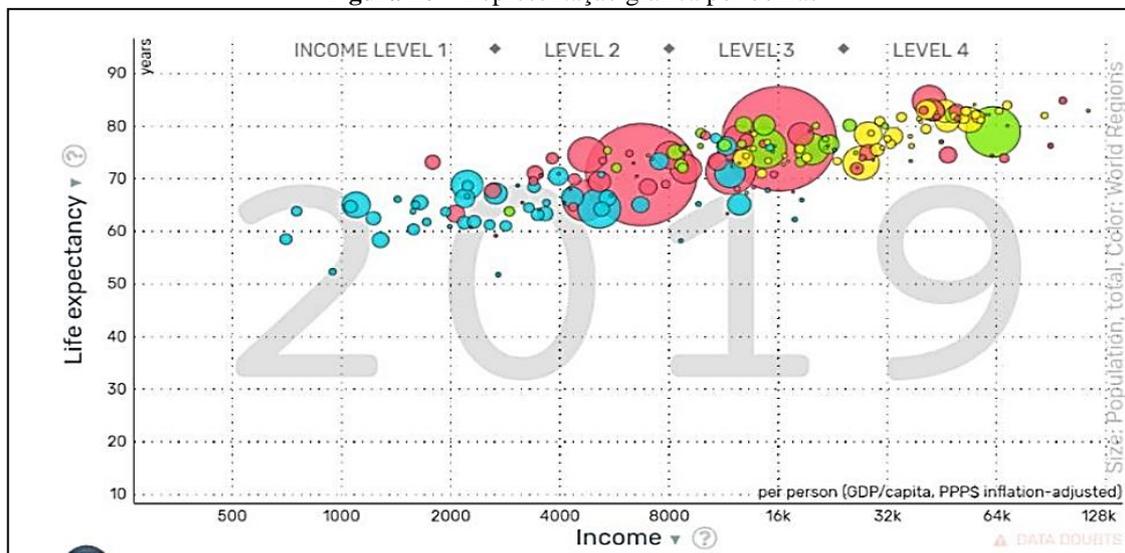
The colour of the dots represents the continents of the country they are representing (green for North and South America; yellow for Europe, Iceland and Russia; red for Asia, and Australia; blue for Africa. Although one must mention that one could customize the meaning of colour. In the standardized setting, the different sizes of the dots represent the population of each country (therefore the dots representing India and China e.g., are huge).

Tradução: A cor dos pontos representa os continentes dos países que estão representando (verde para a América do Norte e do Sul; amarelo para a Europa, Islândia e Rússia; vermelho para a Ásia e Austrália; azul para a África. Embora se deva mencionar que se poderia personalizar o significado da cor. No cenário padronizado, os diferentes tamanhos dos pontos representam a população de cada país (portanto, os pontos que representam a Índia e a China, por exemplo, são enormes).

Os apontamentos apresentados pela dupla de futuros professores são resultados de uma exploração sobre os aspectos visuais. Além disso, ao indicar as diferenças em relação à

área ocupada por cada ponto, a **Dupla 2** salienta uma variável implícita da representação do gráfico de bolhas, a proporção da população de cada país.

Figura 10 – Representação gráfica por bolhas



Fonte: Dados da pesquisa.

A Figura 10 se refere a um exemplo de representação gráfica de dados que o *Gapminder* disponibiliza, e a **Dupla 2** descreve sua exploração inicial sobre ela:

The pre-setting shows Income on the x-Axis and Life expectancy on the y-Axis, though you are able to choose from different variables as child-mortality, CO₂-emission per capita, Cell-phones per capita and many more. Every circle represents one country in the year 2019. The colours show where the countries come from (geographical aspect) and the size of the bubbles represents the population.

Tradução: A pré-configuração mostra Renda no Eixo x e Expectativa de vida no Eixo y, embora você possa escolher entre diferentes variáveis, como mortalidade infantil, emissão de CO₂ per capita, telefones celulares per capita e muito mais. Cada círculo representa um país no ano de 2019. As cores mostram a região dos países (aspecto geográfico) e o tamanho das bolhas representa a população.

Além de explorarem as ferramentas disponíveis para analisar os dados que a plataforma *Gapminder* possui, os futuros professores da **Dupla 2** perceberam que, a depender das variáveis de interesse, há a possibilidade de analisá-las com registros de um

espaço temporal relativamente extenso, entre os anos de 1800 a 2019, e, ao consultar o *website*, há bancos que contêm dados mais recentes, ou seja, o *Gapminder* segue atualizando a sua base de dados.

The collected data reaches all the way back to the year 1800 and contains all the data between 1800 and 2019 (although some special data like the human-development-index has fewer recorded years). Therefore, the chart can be animated to show the development of all the countries in this timeframe. Although this is already a huge feature of Gapminder, it really gets interesting, when you change the x and y-Axis of the chart.

Tradução: *Os dados coletados remontam ao ano de 1800 e contêm todos os dados entre 1800 e 2019 (embora alguns dados especiais, como o índice de desenvolvimento humano, tenham menos anos registrados). Portanto, o gráfico pode ser animado para mostrar o desenvolvimento de todos os países neste período. Embora isso já seja uma grande característica do Gapminder, realmente fica interessante, quando você altera o eixo x e y do gráfico.*

No relato da exploração realizada pela **Dupla 2** é possível identificar que os futuros professores interagiram com algumas ferramentas disponíveis no *Gapminder*, pois verificase ao animar as representações gráficas, alterar condições de dependência e independência de variáveis nessas representações, no momento que a dupla acessa o *link*: <https://www.gapminder.org/answers/how-did-the-world-population-change/>:

[...] with the explanation of Hans Rosling, to get an idea how Gapminder visualization works.

Tradução: *[...] com a explicação de Hans Rosling, consegue-se ter uma ideia de como funciona a visualização do Gapminder.*

O percurso de exploração inicial permitiu que a **Dupla 2** interagisse com o recurso tecnológico digital integrado na atividade, aproximando-a de temas relevantes para a sociedade a partir de representações de dados. Para a segunda solicitação da atividade, ao

explorar as ferramentas do *Gapminder* (<https://www.gapminder.org/tools/>), a **Dupla 2** apresenta dois exemplos de resultados dessa exploração.

Dollar-Street: The organization behind Gapminder travelled the world and took photos of different volunteer-families, that stated their weekly income. The publicized all those pictures and the data to illustrate, which weekly-income can provide which level of lifestyle for a family. You are even able to select which part of the families house you want to see in particular (e.g., the bedroom).

Tradução: *Dollar-Street: A organização por trás do Gapminder viajou pelo mundo e tirou fotos de diferentes famílias-voluntárias, que declararam sua renda semanal. Divulgamos todas essas fotos e os dados para ilustrar, qual renda semanal pode fornecer qual nível de estilo de vida para uma família. Você ainda é capaz de selecionar qual parte da casa da família deseja ver em particular (por exemplo, o quarto).*

A primeira ferramenta explorada explicita parte do trabalho realizado pela Fundação *Gapminder* para acessar e problematizar questões relevantes a partir da produção de dados. No exemplo apresentado pela **Dupla 2**, a ferramenta permite acessar dados sobre renda familiar em diferentes lugares do mundo, acompanhados de fotografias que aludem para as condições sociais dessas famílias. Isso pode ser visto como uma ferramenta que, no sentido do letramento estatístico na era *OPEN Data*, contribui para a compreensão da dimensão e do contexto de origem de dados.

Outra ferramenta explorada foi a da representação gráfica de “renda acumulada”:

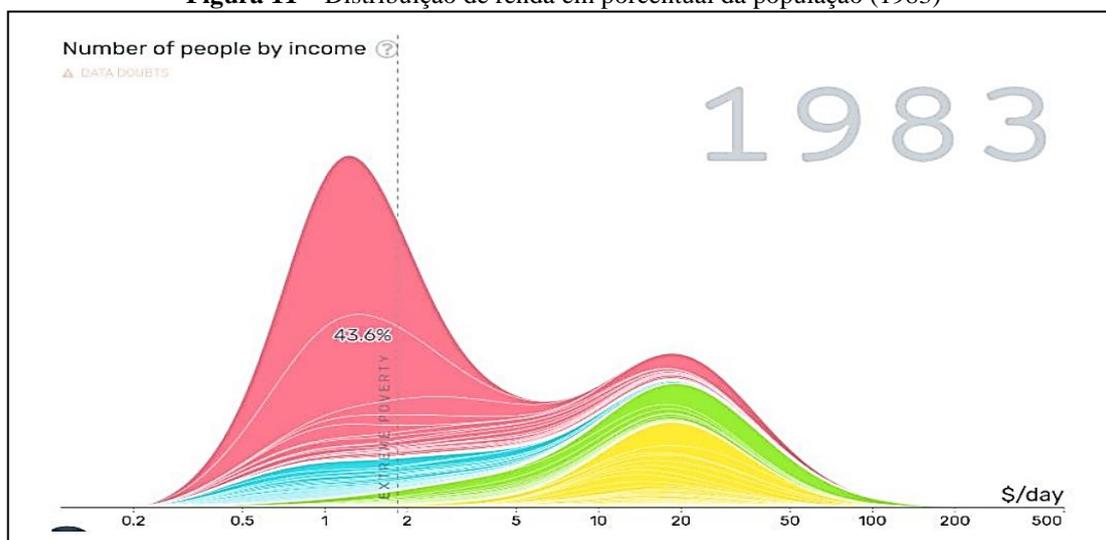
The second tool, that we found to be particularly interesting is the (as we would call it) the “accumulated income” chart, in which the worlds income today is shown to be normally distributed (which in general makes a lot of sense, when we think about the economy of our society), where each layer of the curve is depicting the distribution of income for a specific country. This illustrates the difference between the western countries and the development-countries quite well (many people having low income in development-countries vs. Many people in western countries having high-income). Especially in the 80s one can see a huge difference between Asia and Europe/America.

Tradução: A segunda ferramenta, que achamos particularmente interessante, é o (como chamaríamos) o gráfico de "renda acumulada", no qual a renda mundial hoje é mostrada como sendo normalmente distribuída (o que em geral faz muito sentido, quando pensamos na economia de nossa sociedade), onde cada camada da curva está representando a distribuição de renda para um país específico. Isso ilustra muito bem a diferença entre os países ocidentais e os países em desenvolvimento (muitas pessoas com baixa renda nos países em desenvolvimento versus muitas pessoas nos países ocidentais com alta renda). Especialmente nos anos 80 pode-se ver uma enorme diferença entre a Ásia e a Europa / América.

A ferramenta de representação gráfica que a dupla de futuros professores apresenta em seu relatório considera a distribuição de renda nos diferentes países. Como as cores permitem identificar as regiões do planeta, as análises a serem realizadas podem revelar tendências, concentrações e dispersões em relação às diferentes regiões, ao longo do tempo (Figura 11). Além disso, a taxa percentual, uma métrica evidente na representação gráfica disponibilizada, pode possibilitar ao usuário quantificar o comportamento desses dados.

Os exemplos de ferramentas acessadas pela **Dupla 2** revelam potenciais contribuições da utilização do *Gapminder* em atividades de análise gráfica de dados, visto que o discurso escrito pela dupla de futuros professores explicita a possibilidade de análise gráfica de dados mediada pela representação visualizada e pelos parâmetros que nela constam.

Figura 11 – Distribuição de renda em porcentual da população (1983)



Fonte: Dados da pesquisa.

A exploração e a interação com as ferramentas do *Gapminder* foram relevantes para a dupla de futuros professores realizarem a terceira solicitação da atividade, a seleção de problemas a serem estudados, a fase inicial do ciclo investigativo – PPDAC. No relatório, antes de apresentarem os problemas, a **Dupla 2** considerou:

In general, we thought that the original bubble-tool of gap-minder has the most potential to illustrate interesting exploration topics, because of the versatility of the variables that our yourself can pick for the x- and y-Axis respectively.

Tradução: *Em geral, pensamos que a ferramenta de bolha original do Gapminder tem o maior potencial para ilustrar tópicos de exploração interessantes, devido à versatilidade das variáveis que nós mesmos podemos escolher para os eixos x e y, respectivamente.*

Ao escolherem a forma de representar os dados pelo gráfico de bolhas, uma ação imbricada com um planejamento, é possível identificar que a dupla de futuros professores considera o potencial dessa representação gráfica para promover a compreensão das variáveis envolvidas nos problemas escolhidos.

Em relação aos referências teóricos, para a **Dupla 2**, a representação escolhida revela um potencial transnumerativo para os dados estudados. O potencial ilustrativo, salientado pela dupla, pode ser visto como resultado do desenvolvimento de ferramentas do recurso tecnológico digital para visualização de dados.

Com base nas explorações iniciais e nas interações com as ferramentas do *Gapminder*, a **Dupla 2** abordou em seu relatório o seguinte tópico:

The first interesting exploration topic that we thought of, was the impact of the “Happiness-Score” which determines (or at least is expected to determine) the average happiness in a specific country. Furthermore, we were not only interested in finding different interesting correlations between a variable and the happiness-score, but also in finding some “Wannabe-Correlations”⁷³, that just

⁷³ A dupla de futuros professores atribui a esse tipo de correlação aquele que, embora apresente resultados plausíveis, as análises podem chegar em conclusões que não fazem sentido, por exemplo, o estudo das variáveis *score* de felicidade e emissão de CO₂, apresentado pela dupla em seu relatório.

do not make any sense at all (to illustrate the problem with correlation, especially the problem with a third variable that causes this “Wannabe-Correlation”).

Tradução: O primeiro tópico de exploração interessante em que pensamos, foi o impacto do “score de felicidade”, que determina (ou pelo menos espera-se que determine) a felicidade média em um país específico. Além disso, não estávamos apenas interessados em encontrar diferentes correlações interessantes entre uma variável e o score de felicidade, mas também em encontrar algumas “Correlações Wannabe”, que simplesmente não fazem nenhum sentido (para ilustrar o problema com correlação, especialmente o problema com uma terceira variável que causa essa “Correlação Wannabe”).

A problemática estudada pela **Dupla 2** é um tema de relevância social e elucida a primeira fase do ciclo investigativo – PPDAC. A dupla de futuros professores, ao acessar o *Gapminder*, com base nos dados hospedados, considera os *scores* apresentados para investigar o problema. Os dados disponíveis, ao tratarem do assunto em uma magnitude global, se revelam como uma característica da era *OPEN Data*, salientada por Prodromou e Dunne (2017).

Os resultados (*scores*) observáveis a partir das representações gráficas dos dados, como forma de medir o nível de felicidade, mostram uma escolha necessária para a segunda fase do ciclo investigativo. Como a transnumeração ocorre durante todo o ciclo, ao identificar formas de medida e de quantificar resultados, revelam características de seu estágio inicial, conforme Rubick, Pfannkuch e Yoon (2002).

Do ponto de vista do *contexto*, no sentido do letramento estatístico na era *OPEN Data*, esses *scores* de felicidade de cada país podem ter relação com uma diversidade de variáveis. Na segunda fase do ciclo investigativo – PPDAC, realizada pela **Dupla 2**, a fase de planejamento induziu a dupla de futuros professores para as escolhas que permitiriam a *gestão dos dados*, a *plotagem* de representações para eles e as suas respectivas *análises*.

O trabalho relativo a essa fase do ciclo investigativo, a partir de escolhas e comandos realizados no *website*, ocorre de forma implícita, restando ao usuário a visualização dos dados por meio ferramentas de representações gráficas. Da mesma forma, tendo em conta a diversidade de variáveis que podem ter relação com o problema abordado, o *recolhimento*, a *gestão*, a *limpeza* dos dados selecionados, relevantes para o problema estabelecido, (aspectos delimitados na terceira fase do ciclo investigativo) ocorreram de forma implícita,

seguindo os comandos da dupla de futuros professores aplicados durante a interação com o *Gapminder*.

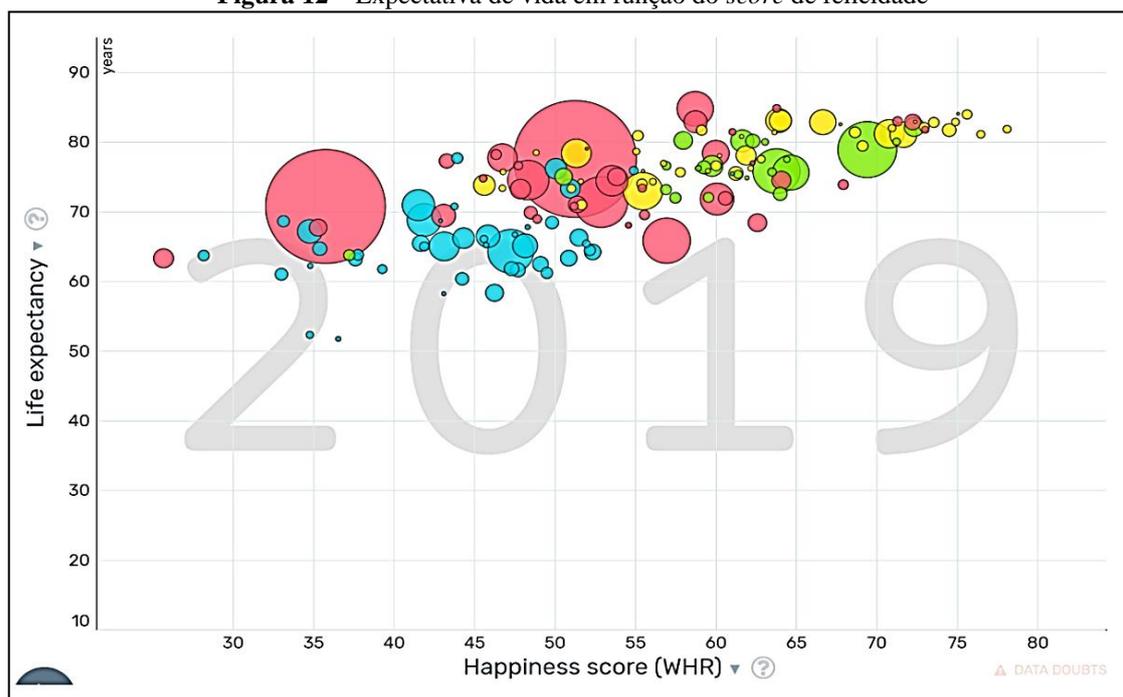
Nessa fase, em que os dados são gerenciados pelos participantes, as escolhas de objetos estatísticos, como correlação e as formas de representação gráfica, permitem identificar que os resultados apresentados no relatório da **Dupla 2**, sobre o *score* de felicidade, caracterizam o segundo estágio de transnumeração, de acordo com Rubick, Pfannkuch e Yoon (2002).

A **Dupla 2** correlacionou a variável *score* de felicidade com outras variáveis, como expectativa de vida, número de filhos por mulher, renda etc. e, em cada correlação realizada, teceu análises e conclusões sobre o problema abordado. Nesse sentido, foi identificado que, admitidas as escolhas e as delimitações anteriores, as fases de análise e conclusão do ciclo investigativo realizado pela **Dupla 2** oferecem, em cada correlação, resultados relevantes para esta pesquisa. Assim, a cada correlação realizada, por meio do *Gapminder* e discutida pela dupla de futuros professores, são apresentadas análises para esta tese, sob aporte dos referencias teóricos e destacando as contribuições da utilização do recurso tecnológico digital *Gapminder*.

One trivial example of the impact of the happiness-score is its (seeming) correlation with life expectancy shown in the chart below.

Tradução: *Um exemplo trivial do impacto do score de felicidade é sua (aparente) correlação com a expectativa de vida mostrada no gráfico abaixo.*

Figura 12 – Expectativa de vida em função do *score* de felicidade



Fonte: Dados da pesquisa.

De forma a exemplificar os resultados emergentes das escolhas e comandos realizados para estudar o *score* de felicidade no mundo, a **Dupla 2**, em uma análise geral, observa a correlação existente dessa variável com a variável expectativa de vida. A Figura 12 é uma representação gráfica resultante do trabalho com os dados (terceira fase do ciclo investigativo), que permitiu a dupla de futuros professores identificar a existência trivial de uma correlação entre *score* de felicidade e expectativa de vida.

Ao considerar o terceiro estágio de transnumeração de Rubick, Pfannkuch e Yoon (2002), o próprio trecho apresentando “*Um exemplo trivial do impacto do score de felicidade e sua (aparente) correlação com a expectativa de vida*” é uma análise conclusiva, considerando as duas últimas fases do ciclo investigativo, em que se observa a comunicação convincente da correlação existente entre o *score* de felicidade e a expectativa de vida.

A **Dupla 2** evidencia o potencial visual da ferramenta do gráfico de bolhas do *Gapminder* que, sob a ótica da visualização de dados, definida por Prodromou e Dunne (2017), colabora para a *compreensão* sobre a correlação existente entre as variáveis e explicita a sua *finalidade*, fundamental à *verbalização* desse aspecto correlacional.

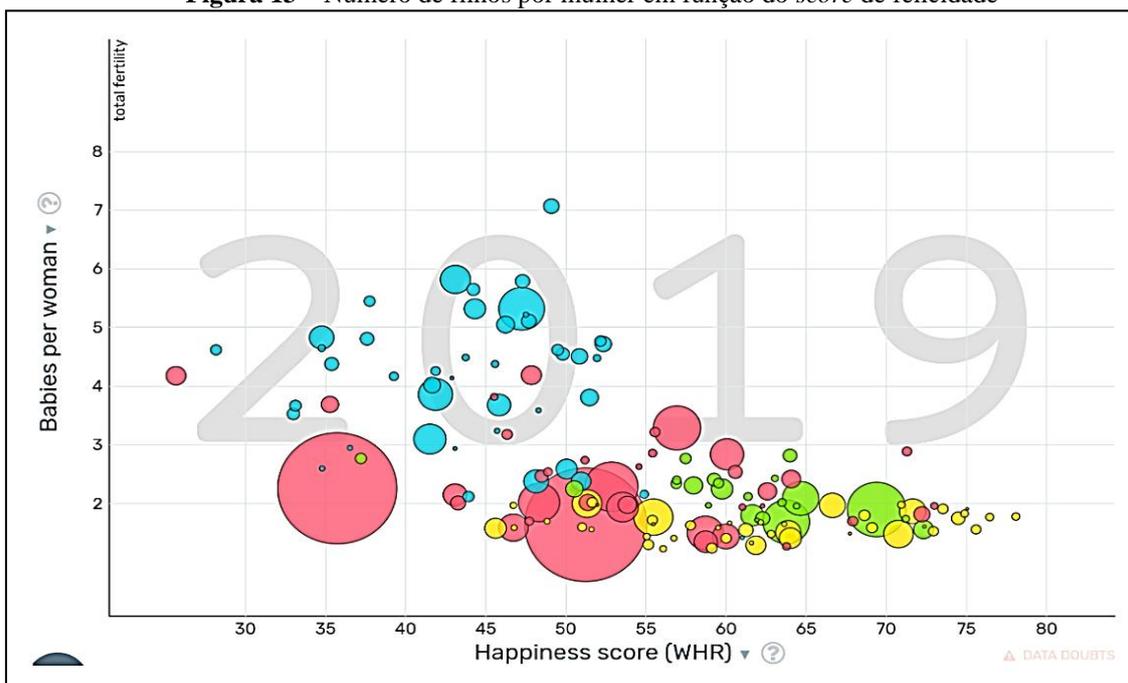
Nas correlações apresentadas a seguir, ao possibilitarem a escolha de variáveis pertencentes ao problema, o *Gapminder* contribuiu para *inferências plausíveis em relação aos dados*. O referencial no discurso escrito sobre a representação gráfica dos dados

favoreceu o aspecto da *linguagem* e, da mesma forma, o discurso apresentou uma compreensão da dupla de futuros professores sobre o *contexto*, admitidas as variáveis expectativa de vida e *score* de felicidade.

A correlação e a representação gráfica dos dados podem ser observadas como uma possível visualização de um *modelo* ao explicar como essas variáveis se correlacionam e podem indicar possíveis *ações* a serem tomadas, havendo alguma *evidência*. Todos esses aspectos estão alinhados com elementos do letramento estatístico na era *OPEN Data*, como definido por Prodromou e Dunne (2017).

Admitidas as três primeiras fases do ciclo investigativo (delimitação do problema, planejamento e trabalho com os dados), uma outra análise foi construída com base nas interações com o *Gapminder*, em outro tópico estudado pela **Dupla 2**: o *score* de felicidade em relação ao número de filhos por mulher.

Figura 13 – Número de filhos por mulher em função do *score* de felicidade



Fonte: dados da pesquisa.

A representação gráfica (Figura 13) é uma visualização dos dados referentes ao ano de 2019 sobre o *score* de felicidade e número de filhos por mulheres. No momento de aplicação da atividade, esses eram os dados mais recentes sobre essas variáveis.

Another interesting interaction is the connection between happiness-scale and children per women. As seen in the chart, it seems, as if the variable “Happiness-score” and the variable “babies per women” are correlating negatively, meaning that the happier a country is, the less children get born per women. If this really illustrates a negative correlation, one could ask the question why this is the case. The most probable explanation could be, that the happier a country is, the more developed and wealthier it is, meaning that sex-education and contraception is available for a bigger percentage of population, resulting in more planned and less accidental births.

Tradução: Outra interação interessante é a conexão entre a score de felicidade e os filhos por mulher. Como visto no gráfico, parece que a variável “score de Felicidade” e a variável “filhos por mulher” estão se correlacionando negativamente, o que significa que quanto mais feliz é um país, menos crianças nascem por mulheres. Se isso realmente ilustra uma correlação negativa, pode-se fazer a pergunta porque esse é o caso. A explicação mais provável poderia ser que quanto mais feliz é um país, mais desenvolvido e mais rico ele é, o que significa que a educação sexual e a contracepção estão disponíveis para uma porcentagem maior da população, resultando em nascimentos mais planejados e menos acidentais.

A discussão apresentada pela **Dupla 2** permite identificar que a representação gráfica dos dados relativos ao número de filhos por mulher em função do *score* de felicidade foi concebida como o recurso para visualizar a correlação existente entre essas variáveis. No sentido do ciclo investigativo, em sua quarta fase, essa discussão é resultado da *exploração dos dados*, que permitiu a observação de aspectos relevantes entre as duas variáveis, necessários para as *análises* que contribuíram para a *geração de hipóteses* sobre o problema inicial, nesse caso, o *score* de felicidade nos diferentes países.

Aos estágios de transnumeração, ao indicarem “*como visto no gráfico, parece que a variável “score de Felicidade” e a variável “filhos por mulher” estão se correlacionando negativamente, o que significa que quanto mais feliz é um país, mais desenvolvido e mais rico ele é*”, observa-se que a dupla de futuros professores utilizou a representação gráfica dos dados (Figura 13) para entender o que eles estavam indicando sobre o problema. O potencial de visualização desse tipo de representação gráfica já havia sido ponderado pela **Dupla 2** como viável para estudar o problema abordado.

A viabilidade está no sentido de que a representação do gráfico de bolhas do *Gapminder*, sob a ótica da visualização de dados, conforme Prodromou e Dunne (2017), pode oferecer suporte para *compreender* os dados do problema, explicitando uma *finalidade* relevante para a *verbalização* presente no discurso do relatório da dupla de futuros professores.

A correlação negativa entre dados representados graficamente e observada pela **Dupla 2** permite, como caracteriza a fase de conclusão do ciclo investigativo – PPDAC, uma *interpretação* sobre a correlação que há entre o *score* de felicidade e o número de filhos por mulher no mundo. A explicação “*quanto mais feliz é um país, mais desenvolvido e mais rico ele é, o que significa que a educação sexual e a contracepção estão disponíveis para uma porcentagem maior da população, resultando em nascimentos mais planejados e menos acidentais*” é uma *conclusão* que oferece *novas ideias* decorrentes dos resultados da correlação dos dados, observada graficamente.

A explicação apresentada pela **Dupla 2** evidencia que a representação gráfica de bolhas para os dados relativos às variáveis *score* de felicidade e número de filhos por mulher contribuiu para uma compreensão sobre como essas variáveis se correlacionam, ou seja, foi uma mudança de representação que promoveu a compreensão do problema, como definido por Wild e Pfannkuch (1999).

Quanto aos estágios de transnumeração de Rubick, Pfannkuch e Yoon (2002), a representação gráfica de bolhas dos dados subsidiou a explicação da **Dupla 2** e contribuiu para uma comunicação compreensível e convincente sobre a situação original do problema, ao considerar a correlação do *score* de felicidade com o número de filhos por mulher, caracterizando o terceiro estágio de transnumeração.

Da mesma forma, com base na explicação da **Dupla 2**, os elementos do letramento estatístico na era *OPEN Data* se evidenciam. As *Inferências plausíveis em relação aos dados* como a indicação de que países mais desenvolvidos têm maior *score* de felicidade são possíveis com o auxílio das ferramentas de representações de dados disponíveis, por exemplo, as cores para diferentes regiões do mundo ou a possibilidade de identificar por cada bolha um país e sua proporção populacional.

O conhecimento do *contexto* dos dados permitiu à dupla de futuros professores observar uma possível relação existente entre a educação sexual e a contracepção acessível em países desenvolvidos com nascimentos mais planejados e menos acidentais, *causando* o

comportamento correlacional negativo dos dados, ou seja, para a **Dupla 2**, o maior *score* de felicidade tem a ver com o que cada país oferece para seus habitantes.

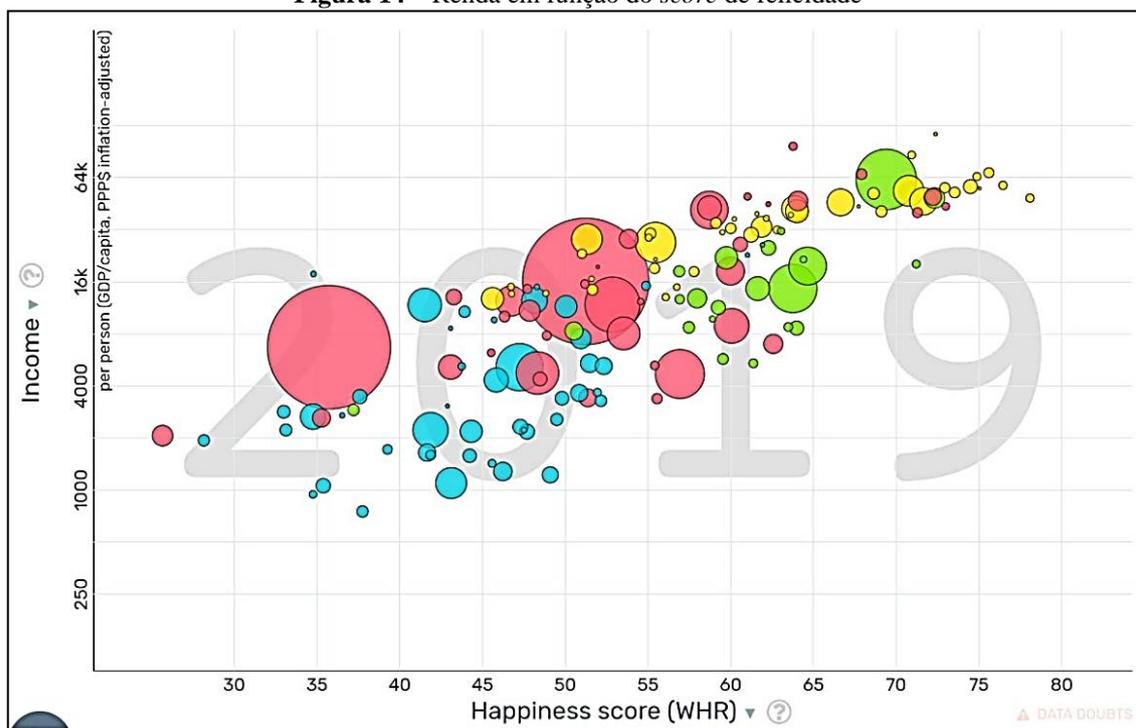
Indicar que “a variável “*score de felicidade*” e a variável “*filhos por mulher*”” estão se correlacionando negativamente” é um resultado que evidencia uma possibilidade da representação gráfica para os dados como auxílio à *linguagem*. A indicação também é um indicativo possível de concepção da representação gráfica dos dados visualizada como um *modelo* que permite compreender e agir sobre o problema.

Emergente na análise sobre a “*score de felicidade*” e a variável “*filhos por mulher*”, o fator desenvolvimento dos países pareceu ser influente. Nesse sentido, a **Dupla 2** ampliou seu relatório, ao apresentar outra representação gráfica dos dados, na qual correlaciona *score* de felicidade com a variável renda.

This is further highlighted by the chart depicting the interaction between the “happiness-score” and “Income”, showing that the intuitive explanation, that more happiness correlates positively with higher GDP.

Tradução: *Isso é ainda destacado pelo gráfico que descreve a interação entre o “score de felicidade” e a “renda”, mostrando que a explicação intuitiva de que mais felicidade se correlaciona positivamente com maior PIB.*

Figura 14 – Renda em função do *score* de felicidade



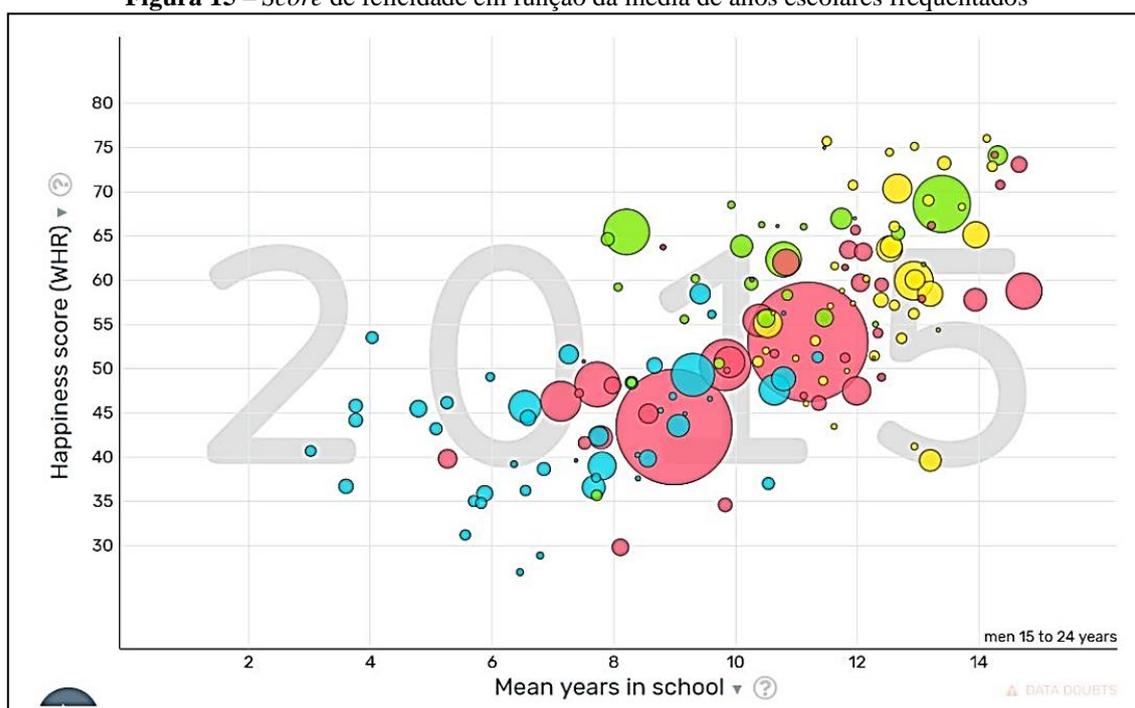
Fonte: dados da pesquisa.

Além da investigar a renda em função do *score* de felicidade (Figura 14) que permitiu conjecturar a correlação existente entre o *score* de felicidade e a renda, a **Dupla 2** explorou a possível influência que há da variável média de anos escolares frequentados com o *score* de felicidade (Figura 15).

As teachers we were also interested in the interaction between “happiness-score” and the mean of the attended school years. This could show, that after all, school isn't all that bad.

Tradução: *Como professores, também nos interessa a interação entre o “score de felicidade” e a média dos anos escolares frequentados. Isso pode mostrar que, afinal, a escola não é tão ruim assim.*

Figura 15 – *Score* de felicidade em função da média de anos escolares frequentados



Fonte: Dados da pesquisa.

No relatório da **Dupla 2** foi identificada a intenção de explorar essas representações dos dados relativos ao *score* de felicidade com dados de outras variáveis para mostrar que, mesmo intuitivamente, a partir da representação visualizada, seria possível processar uma análise gráfica desses dados.

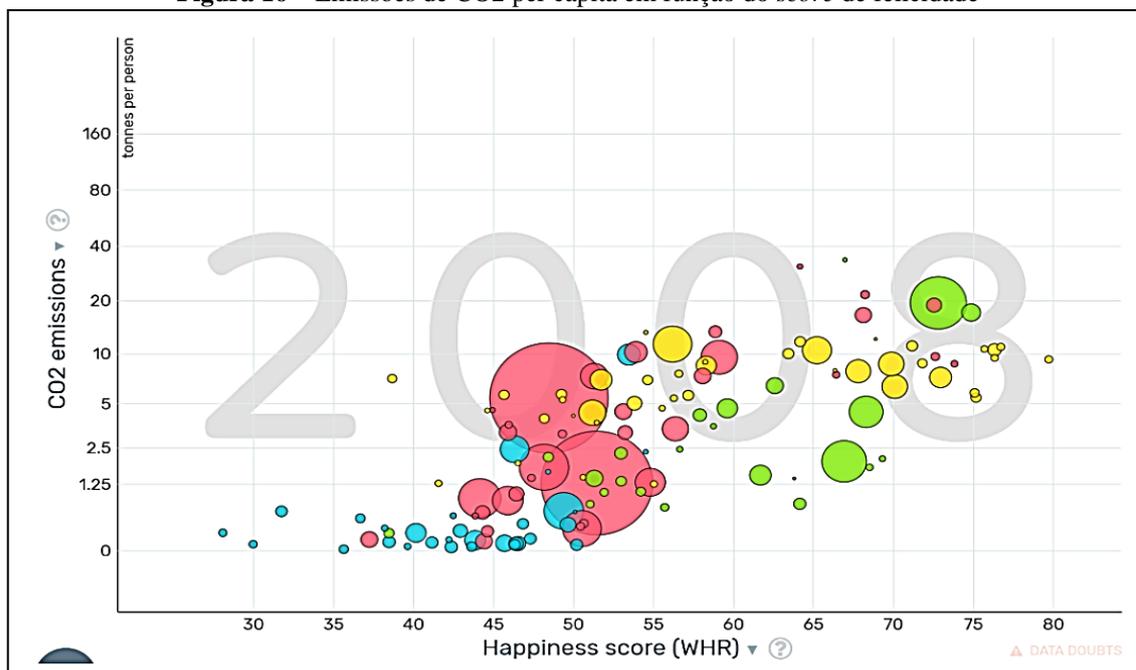
Outro potencial que essas explorações adicionais revelaram foi que os dados disponíveis no *Gapminder* permitiram à dupla de futuros professores ampliar sua compreensão sobre o problema. Como indicado inicialmente, o *score* de felicidade influencia e pode ser influenciado por outras variáveis.

Ainda no sentido de ampliar a percepção sobre o *score* de felicidade nos diferentes países, a **Dupla 2** realizou outra exploração (Figura16).

For the kind of “Wannabe -correlation” we found the interaction between Happiness and CO₂ emission per capita. This could be interpreted as a positive connection, meaning that we are happier the more CO₂ we emit, which is just ridiculous.

Tradução: *Para o tipo de “correlação inicial” encontramos a interação entre Felicidade e emissão de CO₂ per capita. Isso poderia ser interpretado como uma correlação positiva, o que significa que somos mais felizes quanto mais CO₂ emitimos, o que é simplesmente ridículo.*

Figura 16 – Emissões de CO₂ per capita em função do *score* de felicidade



Fonte: dados da pesquisa.

A exploração que a **Dupla 2** realiza no *Gapminder* e as possibilidades de selecionar dados e estudá-los evidenciam um aspecto relevante que vai ao encontro das proposições do letramento estatístico na era *OPEN Data*, a necessidade de considerar o *contexto* dos dados. No caso explorado, embora a representação visualizada apresente elementos que permitam

uma análise das variáveis emissões de CO₂ *per capita* e *score* de felicidade, a própria dupla de futuros professores salienta que não é possível conceber que alguém possa se sentir mais feliz ao emitir maior quantidade de CO₂ na atmosfera.

Sob o ponto de vista da questão de pesquisa desta tese, a utilização do *Gapminder* contribuiu para a realização de análises gráficas de dados a partir de explorações e interações com as representações gráficas. Exemplos dessas explorações foram as análises iniciais sobre o aumento da expectativa de vida, renda e população humana ao longo do tempo e sobre a distribuição de renda nos diferentes países, realizadas pelas **Dupla 1** e **Dupla 2**, respectivamente. As diferentes cores nas representações gráficas, permitem identificar as regiões do planeta e as possíveis análises revelam tendências, concentrações e dispersões em relação aos dados de determinadas variáveis nessas diferentes regiões.

A possibilidade de interagir com o *website*, selecionando as variáveis a serem representadas graficamente e escolhendo as formas de representação, é um dos benefícios oferecidos pelo *Gapminder*. Apesar de não ser evidente nas telas apresentadas pelas duplas de futuros professores, o potencial interativo do recurso tecnológico digital permite visualizar o comportamento dos dados em relação à variável estudada ao longo do tempo. Esses aspectos se mostram como possibilidades da utilização do *Gapminder* favoráveis para análise gráfica de dados.

Ao longo da análise sobre os resultados emergentes nos relatórios das duas duplas, algumas ferramentas se mostraram relevantes como possibilidades para análises gráficas de dados. Nesse sentido, o gráfico da curva de distribuição foi relevante ao ser utilizado para estudar o percentual da população mundial em situação de extrema pobreza ao longo do tempo (**Dupla 1**) ou para investigar a distribuição de renda em percentual da população (**Dupla 2**). No caso da ferramenta de representação gráfica de distribuição dos dados, as diferentes cores visualizadas, que indicam as regiões do mundo, podem favorecer a compreensão sobre onde há maior concentração de pessoas na condição estudada.

A representação gráfica para o estudo de faixas etárias se mostrou como uma ferramenta relevante. Nesse caso, o *Gapminder* possibilitou a realização de previsões, além de observações sobre o comportamento da variável ao longo do tempo.

O gráfico de bolhas é outra ferramenta que se mostrou relevante para análise de dados e, especialmente a **Dupla 2**, se ateu a essa possibilidade ao longo de seu relatório. A proporção das bolhas, como observado pela dupla de futuros professores, permite fazer

referência à população de cada país, além das cores diferentes para identificar as regiões do planeta, nas quais cada bolha está associada.

Ao longo da análise do relatório da **Dupla 2**, as disposições das bolhas possibilitaram discussões sobre as possíveis correlações entre variáveis, ao serem objetos de estudo. As demais características contribuíram para a argumentação sobre possíveis resultados representados, foi o caso, por exemplo, do estudo da renda em função do *score* de felicidade.

Ao considerar os referenciais teóricos delimitados para esta tese, ao longo da análise dos resultados presentes nos relatórios das duas duplas de futuros professores, foi possível identificar que as representações gráficas, de distribuição acumulada, de faixas etárias, de bolhas, foram possibilitadores para os avanços nos processos de transnumeração.

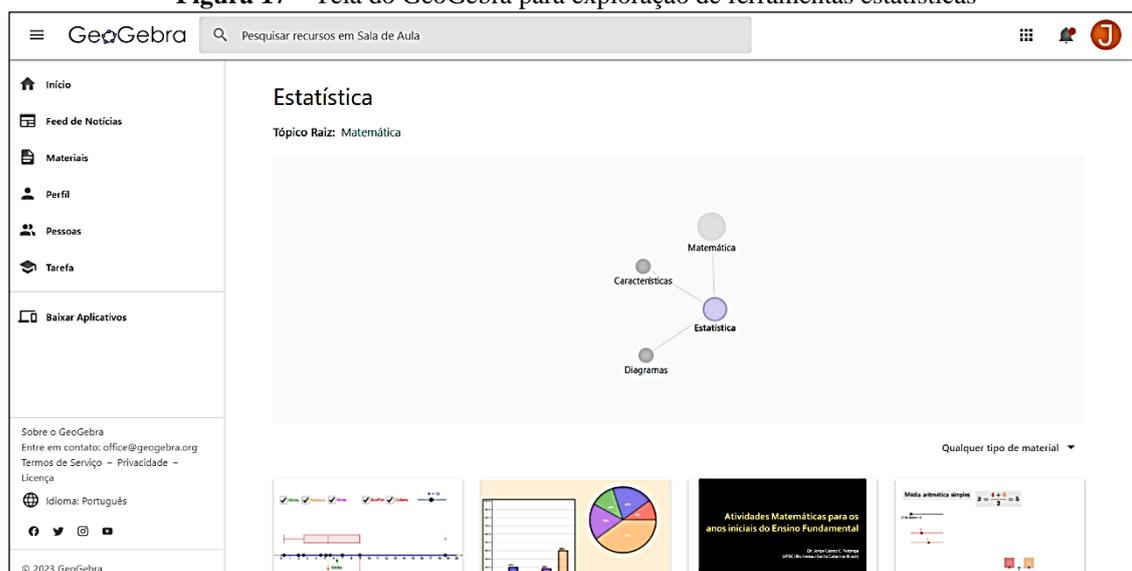
Especialmente, observou-se que, por vezes, discursos conclusivos sobre os problemas abordados tomaram como referência os valores percentuais, as tendências de comportamento dos dados, que foram representados graficamente. Nesse sentido, as análises gráficas dos dados ocorreram por transnumerações fomentadas pelas possibilidades de representações visualizadas pelos participantes. Os resultados que revelam a viabilidade visual dos dados mediados pelo *Gapminder* se evidenciam ao longo dos discursos das duplas.

6.2 Atividades desenvolvidas com o GeoGebra

Neste tópico, estão os resultados da atividade proposta com a utilização do GeoGebra. Na mesma sequência do tópico anterior, são apresentadas análises baseadas nos aportes teóricos desta tese, de modo a explicitar como o *software* contribuiu no desenvolvimento das atividades realizadas pelos participantes.

De acordo com as fases do ciclo investigativo – PPDAC, a solicitação de explorar as ferramentas estatísticas do GeoGebra (<https://www.geogebra.org/t/statistics?lang=en>) teve como objetivo permitir que os futuros professores se familiarizassem com o recurso tecnológico digital para trabalhar com dados (Figura 17).

Figura 17 – Tela do GeoGebra para exploração de ferramentas estatísticas



Fonte: GeoGebra, 2023.

Cabe ressaltar que, antes de propor a atividade aos participantes, o pesquisador apresentou uma possível aplicação do GeoGebra para análise de dados. Além disso, em uma discussão conduzida pelo professor Zsolt Lavicza, outro colega pesquisador, o professor doutor Martin Andre⁷⁴, abordou a relevância da análise de dados no contexto da sustentabilidade.

Após a exploração inicial das ferramentas estatísticas disponíveis no GeoGebra, a segunda solicitação da atividade era que os futuros professores pesquisassem bases de dados disponíveis *online* e estudassem problemas relevantes para eles. Embora alguns *links* de *sites* tenham sido apresentados na atividade, os participantes tinham a liberdade de buscar bases de dados em outras fontes disponíveis na rede.

Sob o ponto de vista metodológico, a escolha de bancos de dados e sua exploração, por meio da utilização do GeoGebra, fazia sentido se houvesse um problema a ser abordado com base nesses dados. A intenção era colocar os participantes em uma situação de investigação, seguindo o ciclo investigativo - PPDAC.

Diferentemente do *Gapminder*, para analisar dados no GeoGebra, era necessário importar os dados. Assim, era esperado que os grupos explorassem diferentes fontes de

⁷⁴ Foi orientando de doutorado do professor Zsolt Lavicza no Departamento de STEAM Education da Linz School of Education. Atua como professor e pesquisador na University College of Teacher Education Tyrol, Áustria, cooperando com o Departamento de Matemática da University of Innsbruck. Disponível em: <https://www.researchgate.net/profile/Martin-Andre>. Acesso em: jul. 2023.

dados sobre diversas problemáticas e escolhessem uma delas para abordar, o que correspondia à primeira fase do ciclo investigativo.

A fase de planejamento, a segunda do ciclo investigativo - PPDAC, exigia que os participantes delimitassem amostras de dados, gerenciassem esses dados, definissem unidades de medida e realizassem análises para compreender o problema. O processo de gerenciamento e limpeza dos dados e sua importação para o GeoGebra caracterizava a terceira fase do ciclo investigativo, o trabalho com os dados.

Esse trabalho, já com os dados importados na planilha de cálculo, envolvia a escolha de ferramentas de análise univariada, bivariada ou multivariada, representações dos dados por meio dessas ferramentas de análise e possíveis alterações na representação dos dados no GeoGebra.

O resultado esperado desses procedimentos era uma tela do GeoGebra com janelas que permitissem a realização de análises dos dados, correspondendo à quarta fase do ciclo investigativo. Além da planilha de cálculo, o GeoGebra oferecia recursos para visualizar estatísticas e representações gráficas desses dados. As conclusões, caracterizando a quinta fase do ciclo investigativo, eram esperadas a partir das análises proporcionadas pelas representações e estatísticas.

Apresentados os delineamentos metodológicos que possibilitaram o desenvolvimento da atividade mediada pelo GeoGebra, *Homework 2*, seguem os resultados observados nos relatórios das duplas, **Dupla 1 e Dupla 2**, e suas respectivas análises, embasadas na transnumeração de Wild e Pfannkuch (1999), Rubick, Pfannkuch e Yoon (2002), e na visualização de dados, conforme Prodromou e Dunne (2017). O desenvolvimento dessa atividade, assim como das demais realizadas pelas duplas, foi apresentado originalmente em inglês e traduzido posteriormente para o português.

Dupla 1 - GeoGebra

O relatório da dupla não apresenta detalhes sobre as primeiras explorações realizadas no GeoGebra ou sobre a experiência com esse recurso tecnológico digital, de modo que a dupla de futuros professores inicia o relatório apresentando cada tópico abordado.

A **Dupla 1** realizou um estudo envolvendo a altura média e peso médio da população mundial por sexo e as possíveis relações entre essas duas variáveis, delimitando o problema a ser abordado.

A escolha dessas variáveis evidencia a primeira fase do ciclo investigativo, e, sob o ponto de vista do letramento estatístico na era *OPEN Data*, os valores médios atribuídos às alturas e aos pesos das populações são dados que sintetizam o contexto de cada país. Certamente, há muito o que investigar sobre a composição e a diversidade de dados que resultam nesses valores médios atribuídos para essas variáveis.

No primeiro tópico do relatório, a **Dupla 1** aborda a altura média da população mundial por sexo.

The first topic, we have chosen handles about the average height by gender of the human population. The dates used for our visualizations were found under this link "<https://www.laenderdaten.info/durchschnittliche-koerpergroessen.php>" These are the average values of the countries.

Tradução: Para o primeiro tópico, escolhemos tratar sobre a altura média por sexo da população humana. As datas utilizadas para nossas visualizações foram encontradas neste link "<https://www.laenderdaten.info/durchschnittliche-koerpergroessen.php>" Estes são os valores médios dos países.

Sobre a origem dos dados e suas datas de publicação, ao acessar o *link* disponibilizado pela dupla de futuros professores, é possível considerar a possibilidade de consultar e questionar a qualidade e veracidade dos dados, o que é uma característica relevante no letramento estatístico na era *OPEN Data*.

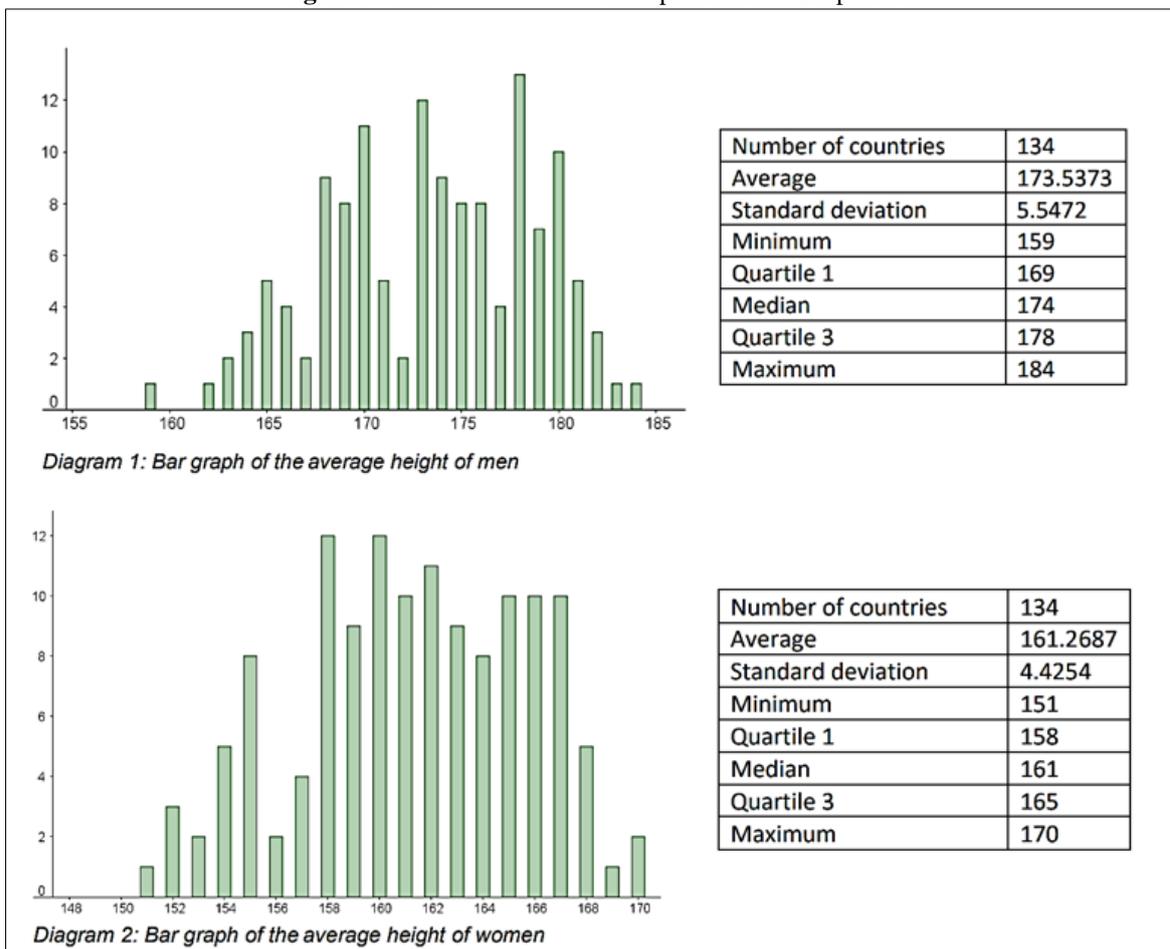
No relatório, a **Dupla 1** apresenta os resultados das manipulações realizadas com os dados, as quais envolvem procedimentos de escolha de amostra, observação das unidades de medida e análises, que são parte da gestão dos dados, segunda fase do ciclo investigativo e necessárias no primeiro estágio de transnumeração. Embora esses procedimentos não estejam explicitamente evidenciados no relatório, é importante reconhecer que foram realizados.

Da mesma forma, os procedimentos de importação dos dados para a planilha de cálculo do GeoGebra, limpeza e organização dos dados são essenciais para o trabalho com os dados, terceira fase do ciclo investigativo. Como resultado desses procedimentos, as possibilidades de representações gráficas foram utilizadas para compreender os dados e suas

relações com o problema em questão, o que corresponde ao segundo estágio de transnumeração, de acordo com Rubick, Pfannkuch e Yoon (2002). Essas representações gráficas, sob o ponto de vista da visualização de dados segundo Prodromou e Dunne (2017), destacam as contribuições do GeoGebra para a análise gráfica dos dados sobre a altura média da população mundial por sexo.

As primeiras representações gráficas da **Dupla 1** para o estudo da altura média da população mundial por sexo são apresentadas na Figura 18.

Figura 18 – Altura média mundial por sexo de 134 países



Fonte: Dados da pesquisa.

Os gráficos de colunas representados e as estatísticas na Figura 18 permitem à **Dupla 1** introduzir algumas análises:

These two diagrams show the average height of women and men of 134 countries. For men, the smallest value is 159 cm, only few countries have this average height. In the case of the women the smallest value is 151 cm. The average height

of men is 173.54 cm, this is shown in diagram 1. This value is higher than the average height of women, which is 161.27 cm and can be seen in diagram 2. The average height of the female and male population has been determined by calculating an average of the average values of the males and females of 134 countries. The median of the height of the men is 174 cm and of the women is 161 cm.

Tradução: Estes dois diagramas mostram a altura média de mulheres e homens de 134 países. Para os homens, o menor valor é de 159 cm, apenas alguns países têm essa altura média. No caso das mulheres, o menor valor é de 151 cm. A altura média dos homens é de 173,54 cm, isso é mostrado no diagrama 1. Este valor é maior do que a altura média das mulheres, que é de 161,27 cm e pode ser visto no diagrama 2. A altura média da população feminina e masculina foi determinada calculando-se uma média dos valores médios dos homens e mulheres de 134 países. A mediana da altura dos homens é de 174 cm e das mulheres é de 161 cm.

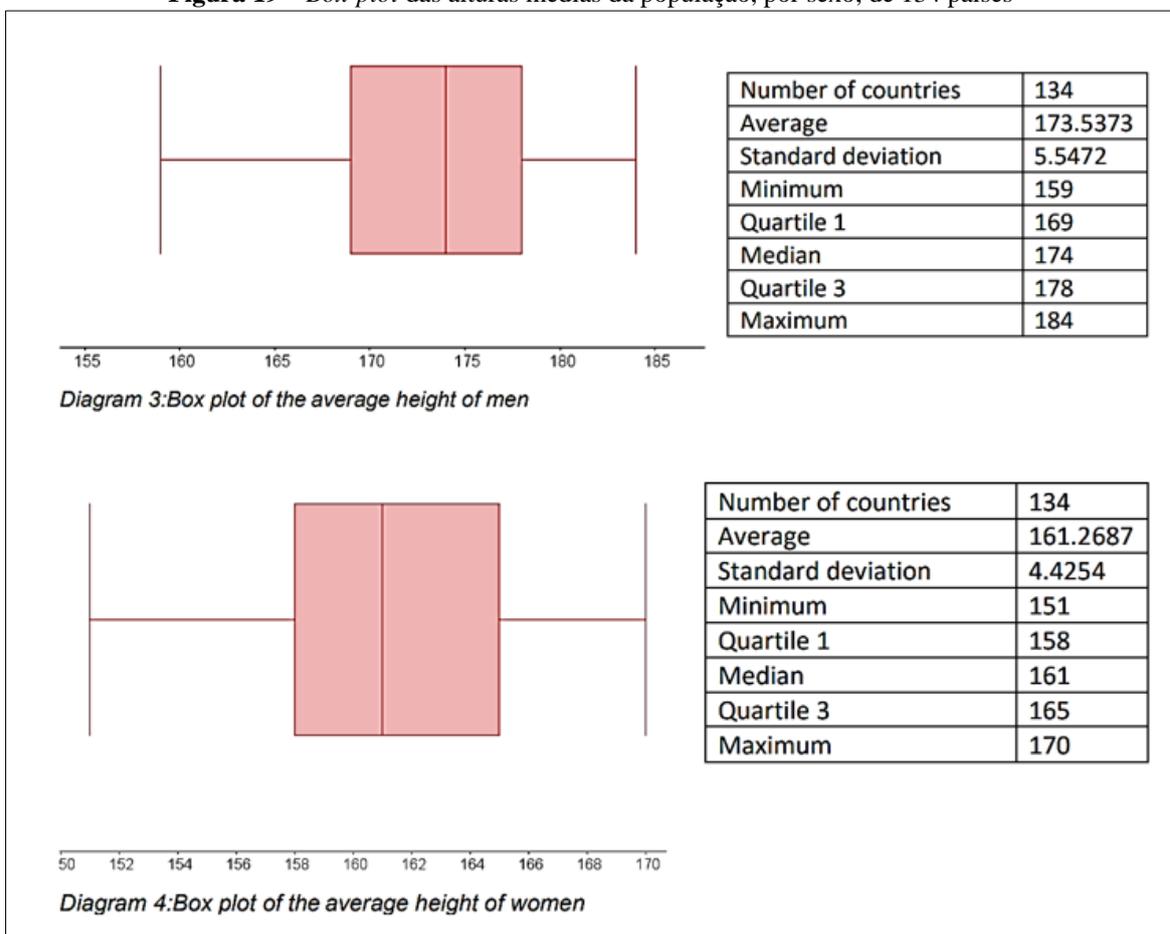
As representações gráficas (gráfico de barras) presentes na Figura 18 mostram as alturas médias de acordo com o sexo e a frequência de países, visualizadas nos eixos horizontal e vertical, respectivamente. Como a variável altura média é contínua, consideramos que a representação gráfica adequada no estudo realizado pela dupla seria o histograma.

É relevante apontar que, embora seja possível observar a frequência de países relativa a cada valor de média das alturas, não se pode, pela representação gráfica dos dados, visualizar quais países estão relacionados em cada barra de frequência. Todavia, a representação pode ser relevante para compreender como essas alturas médias da população humana varia a depender do sexo e de cada país.

O discurso presente no relatório da **Dupla 1**, ao destacar valores como mínimos, médias e medianas, permite que os futuros professores apresentem comparações entre esses valores em relação aos sexos. Essa descrição é resultado do que é possível visualizar na janela gráfica e nas estatísticas apresentadas pelo GeoGebra, que caracteriza uma contribuição do recurso tecnológico digital para a realização dessa análise inicial da **Dupla 1**, a quarta fase do ciclo investigativo.

Outra representação gráfica explorada pela *Dupla 1* foi o *box plot* (Figura 19), que utiliza os dados dos 134 países em quartis. Por meio dessa representação, é possível observar a dispersão das médias de altura.

Figura 19 – *Box-plot* das alturas médias da população, por sexo, de 134 países



Fonte: Dados da pesquisa.

A exploração da representação gráfica *box-plot* para os dados dos 134 países permitiu à **Dupla 1** ampliar a análise descritiva apresentada a partir da representação do gráfico de colunas.

In the boxplot we can see that the median for men is 174 cm and for women 161cm. The value of the first quartile is 158 cm and of the third quartile 165 cm.

Tradução: No box-plot podemos ver que a mediana para os homens é de 174 cm e para as mulheres de 161cm. O valor do primeiro quartil é de 158 cm e do terceiro quartil de 165 cm.

É possível identificar, por meio da representação gráfica do *box-plot*, as diferenças entre as médias de altura entre os sexos, representação que contribuiu para que a dupla de futuros professores verificasse como os dados, divididos em quartis, estão distribuídos. Além da representação gráfica, a exposição das estatísticas das alturas médias foi relevante para possibilitar a compreensão dos dados explorados sobre essa variável.

Sob o ponto de vista da transnumeração, conforme Wild e Pfannkuch (1999), essas representações e suas métricas, disponibilizadas pela utilização do GeoGebra, permitem gerar uma compreensão de como a altura média da população humana pode variar de acordo com o sexo e o país. Nesse sentido, as representações gráficas e as métricas observadas são contribuições do GeoGebra para a análise dos dados.

Ao considerar a visualização de dados, a representação do gráfico de colunas sobre a frequência de países com diferentes médias de altura em ambos os sexos e a representação do *box plot* em quartis, é possível compreender como essas médias de altura da população humana se distribuem, o que leva a identificar tais representações como contributos para a *compreensão* dos dados. O GeoGebra, nesse sentido, permitiu essa *compreensão* e as representações gráficas, com suas *finalidades* explicitadas, presentes na *verbalização* do discurso da **Dupla 1** em seu relatório.

A percepção da dupla de futuros professores, ao identificar e comparar as médias de altura por sexo considerando diferentes países, está alinhada ao letramento estatístico na era do *OPEN Data*, no qual é importante acessar, analisar e realizar interpretações sobre dados dessa natureza. Nesse caso específico, por meio de uma amostra estudada, é possível obter uma compreensão sobre a variável em nível mundial.

O segundo tópico de interesse da **Dupla 1** foi o peso médio da população mundial por sexo, conforme apontam dados⁷⁵. Como no primeiro tópico, as fases de planejamento e de trabalho com os dados, respectivamente, segunda e terceira fases do ciclo investigativo – PPDAC - ocorreram.

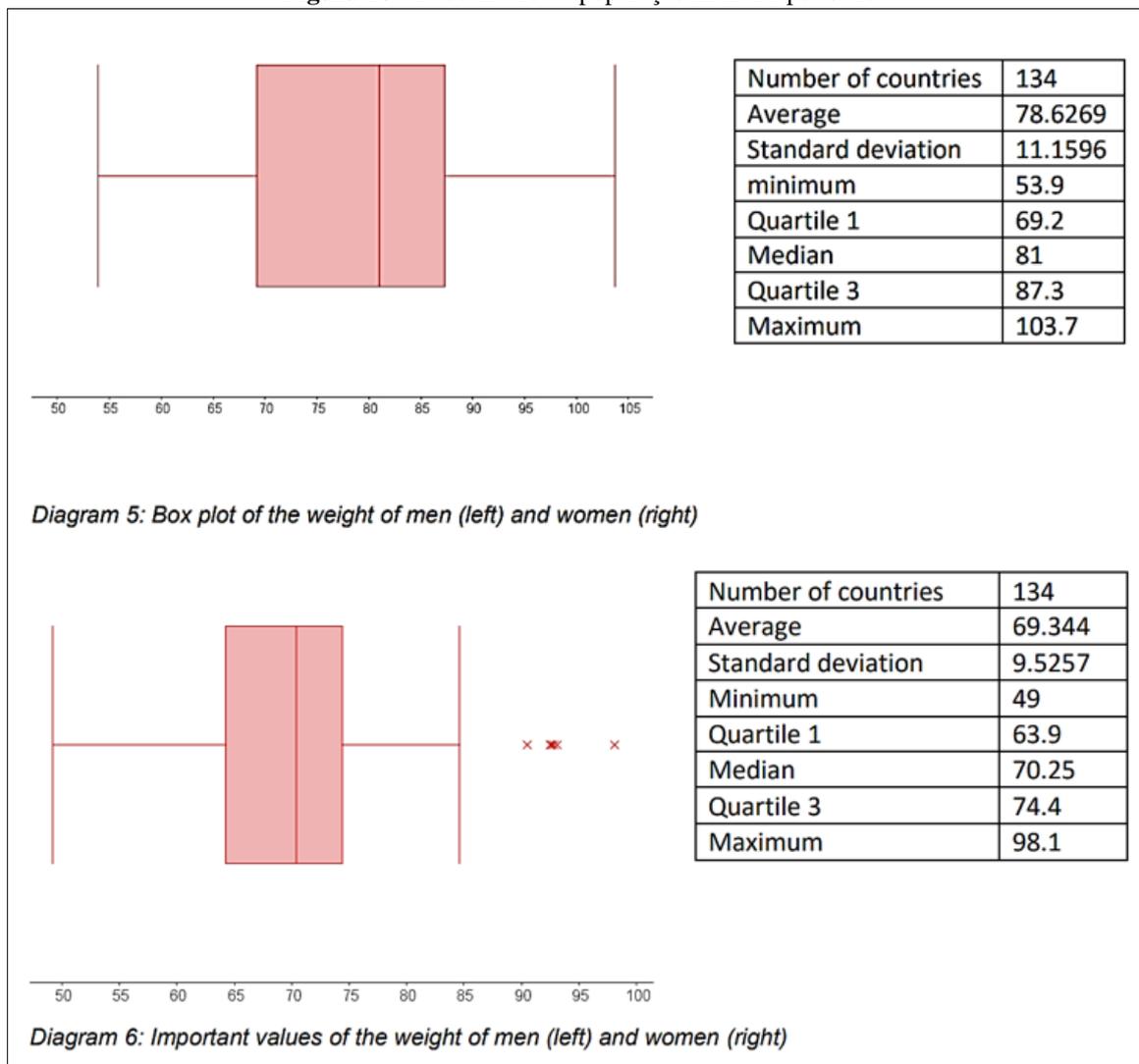
Sob a perspectiva dos estágios de transnumeração, a seleção de amostras e a definição das unidades de medida são etapas relevantes que certamente foram consideradas pela dupla de futuros professores. Além disso, ao lidar com as alturas médias da população mundial e ao reconhecer a validade dos dados disponíveis no *site* utilizado para acessar as médias de

⁷⁵ Disponível em: <https://www.laenderdaten.info/durchschnittliche-koerpergroessen.php>. Acesso em: maio. 2023.

peso, a dupla demonstra um posicionamento condizente com o letramento estatístico na era *OPEN Data*, o que implica *avaliar os dados e considerar a qualidade da evidência* fornecida por esses dados.

A figura 20 mostra a representação gráfica visualizada pela **Dupla 1**, resultado da utilização do GeoGebra na atividade.

Figura 20 – Peso médio da população mundial por sexo



Fonte: Dados da pesquisa.

Diante da visualização gráfica apresentada na Figura 20, a **Dupla 1** realiza uma análise que descreve os resultados obtidos em relação à média de peso da população por sexo nos 134 países, remetendo à quarta fase do ciclo investigativo.

Thanks to these two box plots one can see the average worldwide weight of the women and men. The average weight of the men is 78.63 kg and of the women is 69.34 kg. The standard deviation of the box plot showing the distribution of the average weight of men is 11.16 and of the women is 9.53. One can also realize that there are some outliers in diagram 6.

Tradução: Graças aos dois box-plot pode-se ver o peso médio mundial das mulheres e dos homens. O peso médio dos homens é de 78,63 kg e das mulheres é de 69,34 kg. O desvio padrão do box-plot mostrando a distribuição do peso médio dos homens é de 11,16 e das mulheres é de 9,53. Pode-se também perceber que existem alguns valores atípicos no diagrama 6).

Como é possível observar, a análise realizada pela **Dupla 1** sobre o peso médio da população dos 134 países é descritiva e se enquadra na análise univariada, assim como no tópico anterior. Ao atribuir um valor médio para o peso de cada sexo, a dupla de futuros professores está apresentando valores representativos para toda a população desses 134 países.

A análise também considera os valores do desvio padrão para cada sexo, o que permite à **Dupla 1** explicitar a variação do peso médio observada nos diferentes países em relação aos valores médios calculados na atividade. A representação gráfica do *box-plot* possibilita aos futuros professores identificar possíveis valores atípicos presentes nos dados.

É importante destacar que, neste tópico, a dupla não utilizou a representação gráfica de barras, apenas a do *box-plot*, que, de acordo com os estágios de transnumeração propostos por Rubick, Pfannkuch e Yoon (2002), possibilita compreender o que os dados mostram sobre o peso médio da população humana.

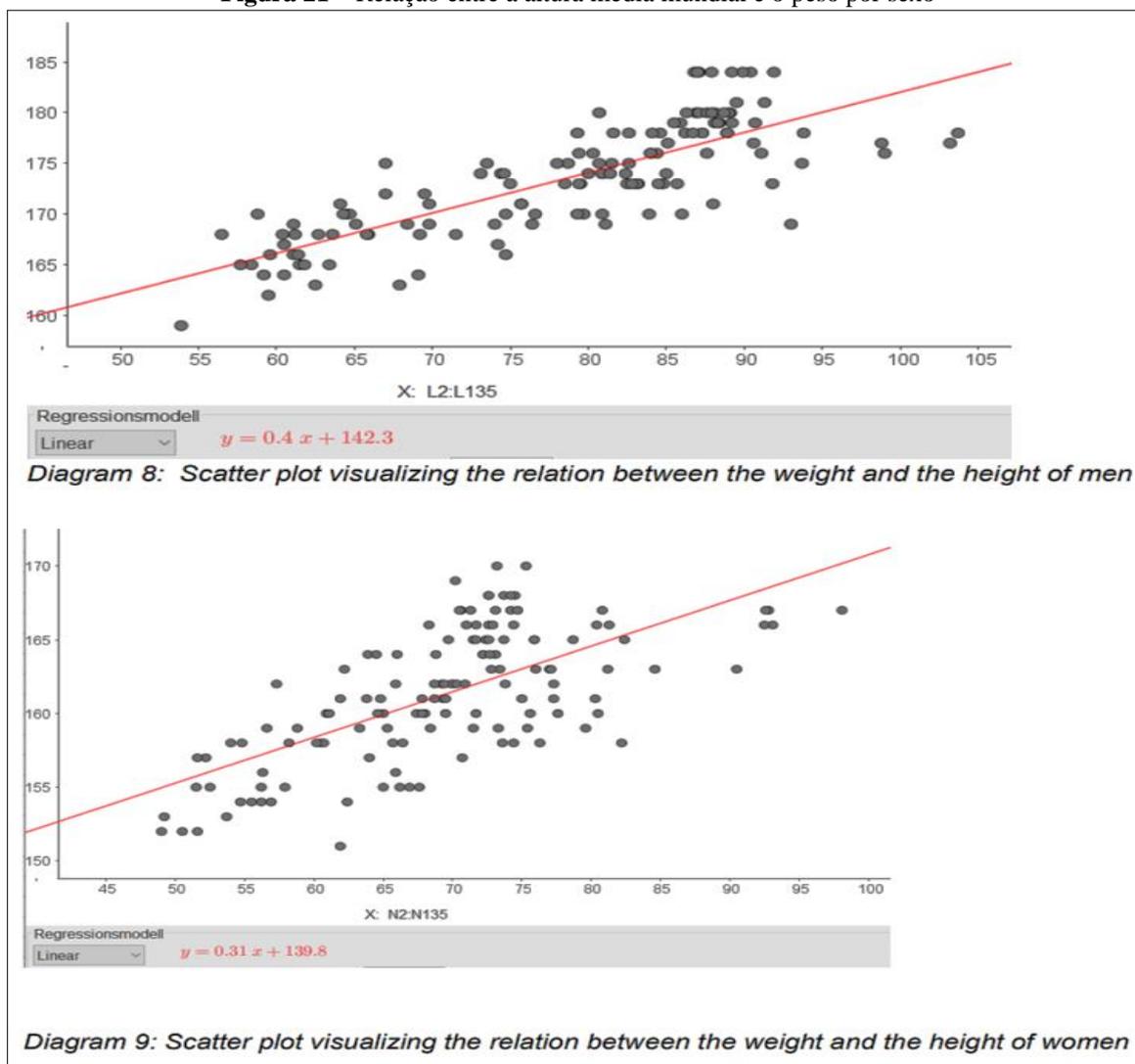
As estatísticas apresentadas e as representações gráficas do *box-plot*, referenciadas no relatório, indicam a compreensão da dupla de futuros professores sobre como os dados sobre o peso médio se distribuem e como os valores encontrados resumem a situação estudada, cumprindo a finalidade da representação gráfica.

Ao visualizar como os dados se distribuem em quartis e ao identificar pontos que representam valores atípicos no conjunto de dados estudado, assim como ao considerar a descrição presente na análise desses valores, é possível reconhecer outra contribuição da representação gráfica do *box-plot*, resultante da utilização do GeoGebra na atividade.

No contexto do letramento estatístico na era *OPEN Data*, de acordo com Prodromou e Dunne (2017), a combinação da representação gráfica com a linguagem utilizada na análise descritiva revela uma potencialidade resultante da utilização do GeoGebra para a análise gráfica de dados. Além disso, as medidas de média e desvio padrão, consideradas pela dupla de futuros professores, foram relevantes para questionar seus significados em relação à complexidade subjacente, mesmo antes do cálculo das médias dos pesos das populações por sexo de cada país. Ou seja, houve uma reflexão sobre o conhecimento do contexto de origem dos dados.

Ao analisar separadamente as variáveis altura média e peso médio por sexo da população dos 134 países, a **Dupla 1** realizou análises descritivas sobre cada uma delas, obtendo suas estatísticas e representações gráficas possibilitadas pela utilização do GeoGebra. Esse processo de análise, característico da quarta fase do ciclo investigativo, foi ampliado pela percepção da dupla de futuros professores, que considerou a possibilidade de existir alguma relação entre a altura média e o peso médio por sexo observados nos diferentes países.

Figura 21 – Relação entre a altura média mundial e o peso por sexo



Fonte: Dados da pesquisa.

Com a utilização da ferramenta de análise bivariada do GeoGebra, representada na Figura 21, a **Dupla 1** obteve o resultado que relaciona as duas variáveis: altura média e peso médio por sexo. É importante ressaltar que a representação gráfica pode ser interpretada como uma representação da correlação entre as variáveis, e a reta exibida, obtida a partir do modelo linear disponível no GeoGebra, representa um estudo de regressão linear para essas duas variáveis. Diante dos resultados, a dupla de futuros professores destaca:

The diagrams 8 visualizes the relation between the weight and the height of men and the diagram 9 of women. In both diagrams the values of the 134 countries were approximated by a linear equation. Both diagrams confirm that generally can be affirmed, that the taller a person is, the more they weigh. The variance of

the regression line shown in diagram 8 is 0.64 and in diagram 9 is 0.44, which means that the regression line approximating the values of the 134 countries of the men is better than the one of the women, as it has a higher variance.

Tradução: O diagrama 8 mostra a relação entre o peso e a altura dos homens e o diagrama 9 das mulheres. Em ambos os diagramas, os valores dos 134 países foram aproximados por uma equação linear. Ambos os diagramas confirmam que geralmente pode ser afirmado, que quanto mais alta uma pessoa é, mais ela pesa. A variância da reta de regressão mostrada no diagrama 8 é de 0,64 e no diagrama 9 é de 0,44, o que significa que a linha de regressão que se aproxima dos valores dos 134 países dos homens é melhor que a das mulheres, pois apresenta uma variância maior.

Os apontamentos da **Dupla 1**, baseados nos resultados da análise bivariada representada na janela gráfica do GeoGebra, trazem considerações gerais sobre a relação entre altura média e peso médio da população dos 134 países. A dupla de futuros professores identifica que a reta de regressão linear é um modelo que aproxima a altura média do peso médio, por sexo, da população dos 134 países. Como é possível observar na representação da reta, em ambos os sexos, faz sentido afirmar que, geralmente, quanto mais alta uma pessoa é, mais ela pesa.

Os valores 0,64 para o sexo masculino e 0,44 para o sexo feminino, concebidos pela **Dupla 1** e apresentados como variância nas estatísticas do GeoGebra, deveriam ser considerados como coeficientes de determinação da reta de regressão sobre o peso médio e a altura média.

A interpretação desses resultados é evidente quando a **Dupla 1** afirma: *a reta de regressão que se aproxima dos valores dos 134 países dos homens é melhor que a das mulheres, pois apresenta uma variância maior.*

De acordo com os referenciais teóricos desta tese, a representação gráfica resultante da análise bivariada realizada pela dupla de futuros professores, utilizando os dados dos 134 países em relação às variáveis altura média e peso médio por sexo, evidencia outra possibilidade de contribuição para a análise gráfica de dados.

Como mencionado pela **Dupla 1**, a transnumeração, conforme Wild e Pfannkuch (1999), ocorre a partir dos resultados obtidos pela utilização do GeoGebra na análise bivariada, permitindo compreender que há diferenças na variabilidade dos dados, dependendo do sexo, e que, embora possam ser modelados por uma regressão linear, um modelo é mais adequado do que o outro devido a essa diferença na variabilidade dos dados.

Ao considerar o referencial de Rubick, Pfannkuch e Yoon (2002), a escolha de representar a relação entre peso médio e altura média da população dos 134 países como uma regressão linear significa ações correspondentes ao segundo estágio de transnumeração, no qual se observa a contribuição da utilização do GeoGebra para que a dupla de professores faça considerações sobre como os dados das variáveis estão relacionados.

Ao tecerem considerações sobre o modelo de regressão linear visualizado, se observa que, para a *compreensão* da dupla, quando afirma que *quanto mais alta uma pessoa é, mais ela pesa*, evidencia-se uma *finalidade* atendida da representação gráfica visualizada. A afirmativa sobre a representação do modelo de regressão linear em relação aos dados é, explicitamente, a *verbalização* daquilo que a **Dupla 1** visualizou diante do resultado da análise bivariada mediada pelo GeoGebra.

No sentido de apontar para a fase de conclusão do ciclo investigativo – PPDAC no *Homework 2*, as análises apresentadas nos três tópicos colaboraram para elucidar como há uma variação em relação à altura média e o peso médio da população em 134 países. Em ambos os tópicos, a percepção sobre essas diferenças foi mediada pela utilização do GeoGebra, o qual possibilitou a construção de representações gráficas de dados relativos às variáveis estudadas pela **Dupla 1**.

Para concluir a quinta fase do ciclo investigativo, além dos resultados e como esses emergiram em suas análises interpretativas das representações gráficas dos dados, novas ideias surgiram, de modo que a dupla de futuros professores buscou encontrar explicações para as diferentes alturas médias e pesos médios.

Deviations can result from poor nutrition. Medical care is therefore also important. Body size can be influenced by diseases and allergies. In poor countries, the height is lower than in richer countries. Genes determine what height you can achieve. Families where people are taller will pass these genes on to the next generation. People are getting bigger and bigger. In the last 100 years people have grown by around 15 cm.

Tradução: *Os desvios podem resultar de má nutrição. A assistência médica é, portanto, também importante. O tamanho do corpo pode ser influenciado por doenças e alergias. Nos países pobres, a altura é menor do que nos países mais ricos. Os genes determinam a altura que você pode alcançar. As famílias em que as pessoas são mais altas transmitirão estes genes à próxima geração. As pessoas estão ficando cada vez maiores. Nos últimos 100 anos, as pessoas cresceram cerca de 15 cm.*

Na tentativa de explicar as diferenças observadas nos dados, a **Dupla 1**, em consonância com os elementos necessários para o letramento estatístico na era *OPEN Data*, conforme Prodromou e Dunne (2017), apresenta novas ideias que levam em consideração aspectos do *contexto* de origem dos dados.

No sentido dos estágios de transnumeração propostos por Rubick, Pfannkuch e Yoon (2002), a comunicação sobre os resultados observados pela dupla de futuros professores ocorreu de forma que, em cada tópico, ao analisar os dados, as afirmações se mostraram como resultados convincentes sobre os dados estudados. Cabe destacar, como um apontamento final do estudo realizado pela **Dupla 1** ao dissertar sobre possíveis causas de haver diferenças relevantes em relação aos dados de cada país, que isso se mostra como uma tentativa de, para além de apresentar resultados estatísticos, problematizar o porquê de eles serem como são. O sentido do problema abordado pela dupla revela, possivelmente, problemas sociais e econômicos ao redor do mundo.

Ao considerar a questão de pesquisa desta tese, observa-se que a utilização do GeoGebra contribuiu para que a **Dupla 1** estudasse os dados relacionados às variáveis altura média e peso médio da população humana em referência a 134 países.

A contribuição da utilização do GeoGebra pode ser observada ao identificar que, para compreender como as alturas médias e pesos médios nesses países variam, são utilizadas as representações gráficas de gráfico de colunas e *box-plot* resultantes das análises univariadas. Da mesma forma, para investigar a possível relação entre essas variáveis nos diferentes países, a representação gráfica de regressão linear possibilitada pelo GeoGebra permite que a dupla de futuros professores chegue a uma conclusão sobre a possível relação existente entre alturas médias e pesos médios da população dos 134 países.

No relatório da **Dupla 1**, são identificadas as representações gráficas de gráfico de colunas, *box-plot* e regressão linear como contribuições do GeoGebra que possibilitaram aos futuros professores visualizar e transnumerar os dados.

As possibilidades de análises univariadas e bivariadas utilizadas pela dupla de futuros professores podem ser identificadas como ferramentas presentes no GeoGebra que, quando integradas às atividades de análise gráfica de dados, oferecem oportunidades de compreender os problemas subjacentes a esses dados, como foi o caso da **Dupla 1** ao estudar os diferentes pesos médios e alturas médias da população humana de 134 países.

Dupla 2 – GeoGebra

O relatório da **Dupla 2**, como no caso da **Dupla 1**, não oferece detalhes sobre as primeiras explorações realizadas no GeoGebra e sobre a experiência dos futuros professores com esse recurso tecnológico digital. A **Dupla 2** inicia seu relatório indicando o primeiro tópico abordado, a pandemia de Covid-19, especificamente, a taxa média de vacinação em função de número relativo de mortes acumuladas.

For our first visualisation and exploration, we decided to use the datasets from the Humanitarian Data Exchange program (<https://data.humdata.org/event/covid-19>), which collected data from all the countries around the world regarding COVID-19 cases and death tolls. We wanted to compare this data with the data depicting the Vaccination progress for the COVID-19-Virus (<https://data.humdata.org/dataset/covid-19-vaccinations>), to show, that in general, the vaccination is reducing the relative percentage of the population dying due to an infection with the virus.

Tradução: Para nossa primeira visualização e exploração, decidimos usar os conjuntos de dados do programa de Intercâmbio de Dados Humanitários (<https://data.humdata.org/event/covid-19>), que coletou dados de todos os países ao redor do mundo sobre casos de COVID-19 e números de mortes. Queríamos comparar esses dados com os dados que descrevem o progresso da vacinação para o vírus COVID-19 (<https://data.humdata.org/dataset/covid-19-vaccinations>), para mostrar que, em geral, a vacinação está reduzindo a porcentagem relativa da população que morre devido a uma infecção pelo vírus.

A dupla de futuros professores delimita os possíveis efeitos da vacinação em relação ao número de mortes em decorrência da Covid-19 como um problema a ser investigado

(primeira fase do ciclo investigativo). Ao selecionar locais de origem dos dados referentes às variáveis de interesse, identifica-se uma escolha que tem a ver com o planejamento. Foram esses bancos de dados que possibilitaram à dupla de futuros professores realizar o estudo. A **Dupla 2** explica que:

Though one might think, that with the right data, this would be an easy task, it was a long and difficult journey and the results that came out of the research were still not really satisfying. To make a fair comparison, we selected a specific timeframe which we wanted to analyse. This timeframe that we also selected for economic reasons, is the timeframe from July 2021 (the 1st of July that is) until now.

Tradução: Embora se possa pensar que, com os dados certos, essa seria uma tarefa fácil, foi uma jornada longa e difícil e os resultados que saíram da pesquisa ainda não foram realmente satisfatórios. Esse cronograma, também selecionamos por questões econômicas, é de julho de 2021 (ou seja, 1º de julho) até agora.

O termo “até agora” se refere a meados de novembro de 2021. O relato apresentado pela **Dupla 2** exemplifica os procedimentos para acessar e limitar uma amostra de dados, que permitisse a realização do estudo, o planejamento e levantamento dos dados, características da segunda fase do ciclo investigativo.

Considerando o referencial de Pfannkuch, Rubick e Yoon (2002), a dupla de futuros professores realizou o primeiro estágio da transnumeração ao delimitar amostras, espaços temporais. Mas a complexidade dos dados exigiu que a **Dupla 2** explorasse apenas parte dos dados, em um recorte temporal, cenário que revela o caráter dos dados produzidos na era *OPEN Data*.

Realizados os recortes e o acesso aos dados, a organização e gestão deles foi descrita da seguinte forma:

Now we imported the .csv file from the website into Excel and decided to hand-pick (indeed very unscientific) a variety of countries from all across the world (Austria, Germany, Italy, India, China, Japan, USA, Brazil, Colombia, Somalia, Egypt and South-Africa). We deliberately picked countries from Africa, Europe,

America, and Asia. We then selected all the entries of all the total fully vaccinated people in each country per day and copied those in a new excel file to calculate the mean of the fully-vaccinated inhabitants in the selected timeframe.

Tradução: Agora importamos o arquivo.csv do site para o Excel e decidimos escolher a dedo (na verdade não muito científico) uma variedade de países de todo o mundo (Áustria, Alemanha, Itália, Índia, China, Japão, EUA, Brasil, Colômbia, Somália, Egito e África do Sul). Escolhemos deliberadamente países da África, Europa, América e Ásia. Em seguida, selecionamos todas as entradas de todas as pessoas totalmente vacinadas em cada país por dia e as copiamos em um novo arquivo Excel para calcular a média dos habitantes totalmente vacinados no período selecionado.

A descrição revela o trabalho realizado a partir da seleção de amostra dos dados e apresenta os procedimentos adotados para possibilitar uma análise, característica da terceira fase do ciclo investigativo. As escolhas por dados de determinados países, vista a diversidade de regiões do planeta, explicita uma intenção da dupla de futuros professores em compreender o processo de vacinação e o avanço do número de mortos por Covid-19 ao redor do mundo. Uma explicação da **Dupla 2** nesse sentido foi a seguinte:

We then researched the total population for each country and calculated the relative average full-vaccination rate. The second variable we looked at was the death-toll each day since July 1st to sum them up for each country and then calculate the percentage in relativity to the population of each country.

Tradução: Em seguida, pesquisamos a população total de cada país e calculamos a taxa média relativa de vacinação completa. A segunda variável que analisamos foi o número de mortos a cada dia desde 1º de julho para somá-los para cada país e, em seguida, calcular a porcentagem em relação à população.

O processo descrito apresenta os procedimentos que possibilitaram seguir no ciclo investigativo, para abordar a taxa média de vacinação em função de número relativo de mortes acumuladas. A planilha (Figura 22) mostra parte dos dados organizados pela dupla de futuros professores, que salienta para a dificuldade de interpretá-los da forma como estão dispostos.

Figura 22 – Planilha com dados estatístico sobre a taxa média de vacinação e o número relativo de mortes acumuladas

	A	B	C	D	E	F	G	H
1		Austria	Germany	Italy		China	India	Japan
2								
3	Total fully Vacc. (since 1st of July 2021)	3235126	31662203	19286742			59048568	18439014
109		5505649	54658274	42128182			277747310	83656184
110	Mean of total vaccinations	4862591,613	47017403	34229039,9		933999200	152177616	49299926,01
111	Population Total	8971000	83240000	59000000		1402000000	1380000000	125000000
112	Percentage of Population	0,545316992	0,564841458	0,580153219		0,666190585	0,110273635	0,394399408
113								
114								
115	Daily death toll (since July 2021)	1	63	24		5	1005	41
116		0	69	21		13	853	21
221		17	65	40		0	379	31
222		405	3651	3919		205	53360	3311
223		4,51455E-05	4,38611E-05	6,64237E-05		1,4622E-07	3,86667E-05	16.

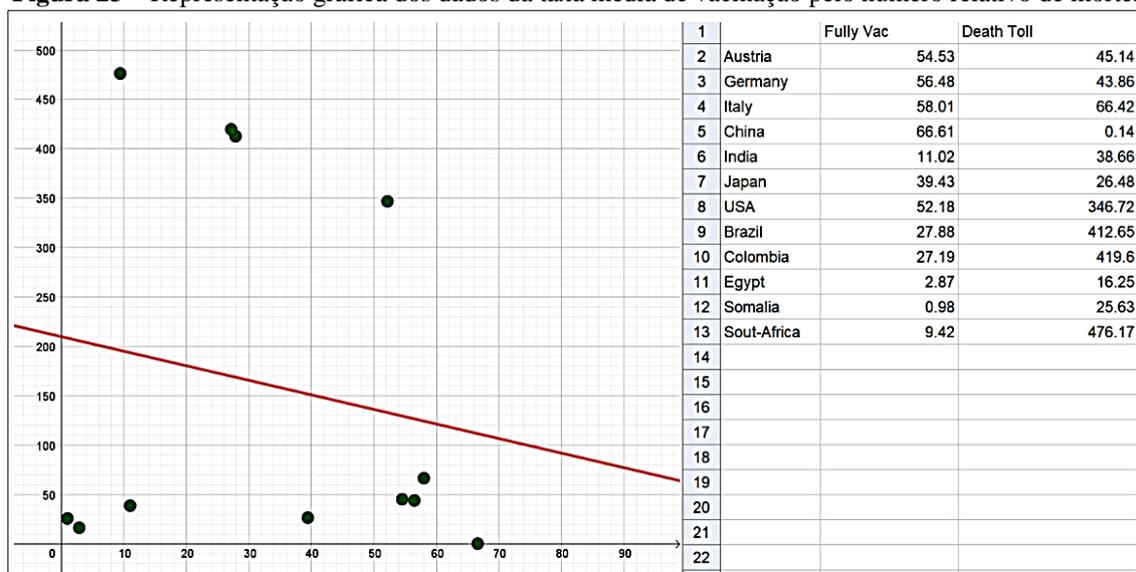
Fonte: Dados da pesquisa.

The data already showed some weird numbers that we could not explain, but by plotting the data on a scatter-plot, we can see that there is something weird going on.

Tradução: Os dados já mostravam alguns números estranhos que não conseguíamos explicar, mas ao plotar os dados em um gráfico de dispersão, podemos ver que há algo estranho acontecendo.

No sentido de obter melhores condições para analisar os dados, a escolha por importar e realizar manipulações com eles no GeoGebra indica que o recurso gráfico pretendido pela dupla de futuros professores foi relevante para o processo de análise gráfica. Além disso, a escolha pela representação gráfica utilizada para entender os dados condicionou aspectos relacionados ao segundo estágio de transnumeração, conforme Pfannkuch, Rubick e Yoon (2002).

Figura 23 – Representação gráfica dos dados da taxa média de vacinação pelo número relativo de mortes



Fonte: Dados da pesquisa

Um resultado da terceira fase do ciclo investigativo é a construção de uma representação gráfica (Figura 23), e a explicação sobre ela explicita a visualização da dupla de futuros professores sobre a disposição dos dados, ou seja, no sentido do segundo estágio de transnumeração, a representação gráfica foi empregada para entender o que os dados revelam sobre o problema. Objetivamente, a **Dupla 2** observa uma possível correlação entre a taxa média de vacinação e o número relativo de mortes.

In this chart we see the relative full-vaccination-rate in percent on the x-Axis and the relative death-toll in parts per million on the y-Axis (Deaths per 1 million inhabitants in the selected timeframe). By using the command Trendline, we can see, that through the data that we selected and calculated, there seems to be the relation, that we wanted to show: the higher the full-vaccination-rate in a country is, the smaller the death-toll is.

Tradução: Neste gráfico, vemos a taxa relativa de vacinação total em porcentagem no Eixo x e o número relativo de mortes em partes por milhão no Eixo Y (Mortes por 1 milhão de habitantes no período selecionado). Usando o comando linha de tendência, podemos ver que, através dos dados que selecionamos e calculamos, parece haver a relação que queríamos mostrar: quanto maior a taxa de vacinação total em um país, menor é o número de mortes).

A explicação apresentada descreve a representação gráfica dos dados e a linha de tendência (reta de regressão), que permite à **Dupla 2** indicar uma possível relação existente entre as duas variáveis, explicação condicionada ao resultado da manipulação de dados que, no sentido da visualização, permitiu à dupla de futuros professores uma *compreensão* sobre os dados e seguir para as análises, a quarta fase do ciclo investigativo.

É relevante salientar que a interpretação apresentada é complementada no trecho seguinte, pois os aspectos relativos à variabilidade dos dados amostrados são evidentes na representação gráfica (Figura 23) e destacados no discurso da dupla de participantes.

But if we look more closely at the data, we see some huge discrepancies: Somalia, a country with an average vaccination-rate of nearly 1% in the selected timeframe, only counted a death-toll of approx. 26 inhabitants per 1 million people, while Austria with a vaccination-rate of nearly 55% has a relative death toll of ~45 people per 1 million inhabitants. How is this possible? Are all the conspiracy-theorists right? Does vaccination cause more death, than it prevents?

Tradução: *Mas se olharmos mais de perto para os dados, vemos algumas enormes discrepâncias: a Somália, um país com uma taxa média de vacinação de quase 1% no período selecionado, contou um número de mortes somente de aproximadamente 26 habitantes por 1 milhão, enquanto a Áustria, com uma taxa de vacinação de quase 55%, tem um número relativo de mortes de aproximadamente 45 habitantes por 1 milhão. Como isso é possível? Todos os teóricos da conspiração estão certos? A vacinação causa mais mortes do que evita?*

A análise gráfica dos dados realizada pela **Dupla 2** explora os desvios dos dados em relação à reta de regressão, uma *finalidade* observável da representação gráfica visualizada. Esses desvios explicitam a baixa correlação existente entre os dados amostrados. A possibilidade de comparação entre os dados, por exemplo, da Somália e da Áustria revela uma contribuição da representação gráfica construída para a análise da quarta fase do ciclo investigativo. Os questionamentos em relação à maior taxa de vacinação parecem causar maior número relativo de vítimas, o que seria uma conclusão superficial sobre o problema. De fato, a **Dupla 2** amplia a discussão e apresenta *verbalizações* sobre as possíveis razões para as discrepâncias observadas nos dados.

We can only guess what caused this weird data inaccuracy (because we do not believe that vaccination is contributing to a higher death-toll), but in our opinion, the big differences in death-toll could be related to how countries count COVID-19-deaths and how good the healthcare in each country is. Also, we simply cannot explain how China, although they have a relatively high rate of vaccination, could only report 0.14 COVID-19-deaths per 1 million inhabitants. So, we think that the source data is just skewed, due to the fact that the health-system in Africa just could not count the COVID-19-deaths properly and therefore the unrecorded cases could be quite high.

Tradução: Podemos apenas supor o que causou essa estranha imprecisão de dados (porque não acreditamos que a vacinação esteja contribuindo para um maior número de mortes), mas, em nossa opinião, as grandes diferenças no número de mortes podem estar relacionadas à forma como os países contam as mortes por COVID-19 e quão bons são os cuidados de saúde em cada país. Além disso, simplesmente não podemos explicar como a China, embora tenha uma taxa relativamente alta de vacinação, só poderia relatar 0,14 mortes por COVID-19 por 1 milhão de habitantes. Então, pensamos que os dados de origem estão apenas comprometidos, devido ao fato de que o sistema de saúde na África simplesmente não conseguiu contar as mortes por COVID-19 adequadamente e, portanto, os casos não registrados podem ser bastante altos.

A conclusão apresentada pela **Dupla 2** traz elementos que explicam o porquê de os dados analisados revelarem certas discrepâncias da taxa de vacinação em relação ao número relativo de mortes por Covid-19. Como propõe Pfannkuch, Rubick e Yoon (2002) para o terceiro estágio de transnumeração, a análise conclusiva realizada pela dupla de futuros professores se configura como uma possível explicação convincente para a situação original do problema estudado.

No sentido do letramento estatístico na era *OPEN Data*, as considerações sobre as condições de saúde de cada país, as formas de contagem e de divulgação das taxas de vacinação e do número relativo de mortes são conhecimentos mobilizados que indicam *inferências plausíveis em relação aos dados* e sobre os resultados observados. É, também, um indício do conhecimento da dupla de futuros professores sobre o *contexto* e sobre a necessidade de refletir e *avaliar a qualidade da evidência* dos dados e sobre a origem deles.

Em relação ao problema inicial, investigar a relação existente entre a taxa de vacinação e o número relativo de mortes por Covid-19, foi possível identificar que a **Dupla 2** obteve um resultado relevante. Ao salientarem a discrepância dos dados, os futuros professores revelaram não apenas a relação direta entre uma variável e a outra, mas a possível fragilidade existente nos dados amostrados. Os conhecimentos relativos ao letramento estatístico na era *OPEN Data* se evidenciaram e permitiram supor as razões de a representação gráfica visualizada indicar tamanha discrepância nos dados.

O segundo tópico abordado pela **Dupla 2** ainda seguiu na problemática da pandemia de Covid-19 e tratou do total de vacinados contra a enfermidade ao longo do tempo na Áustria. Essa possibilidade de os participantes trabalharem com dados próximos de sua realidade pode contribuir para a compreensão sobre problemas relevantes para eles. Sobre isso, a **Dupla 2** explica:

For the second exploration we wanted to look at the development of the vaccination-rate in Austria. Specifically, we wanted to see whether we could see if we could spot a huge rise in vaccination-rate around May and June, where vaccinations were firstly available to a broader part of the population.

Tradução: Para a segunda exploração, quisemos analisar a evolução da taxa de vacinação na Áustria. Especificamente, queríamos ver se poderíamos detectar um enorme aumento na taxa de vacinação por volta de maio e junho, onde as vacinas estavam inicialmente disponíveis para uma parte mais ampla da população.

O conhecimento do contexto no qual estão inseridos permitiu à dupla de futuros professores investigar um possível comportamento de uma população e determinar uma questão a ser estudada, a primeira fase do ciclo investigativo. De fato, estudar a evolução do total de vacinados, dada a disponibilidade das vacinas para a população, pode indicar a sua aderência em relação ao processo de vacinação.

We took the same data as in exploration 1 but used the data from the whole timeframe the website could offer to visualize a maximal timespan. We first plugged in all the data into GeoGebra to visualize every day as a dot in the graphical interface. We then used the list of points to generate a Spline-Curve

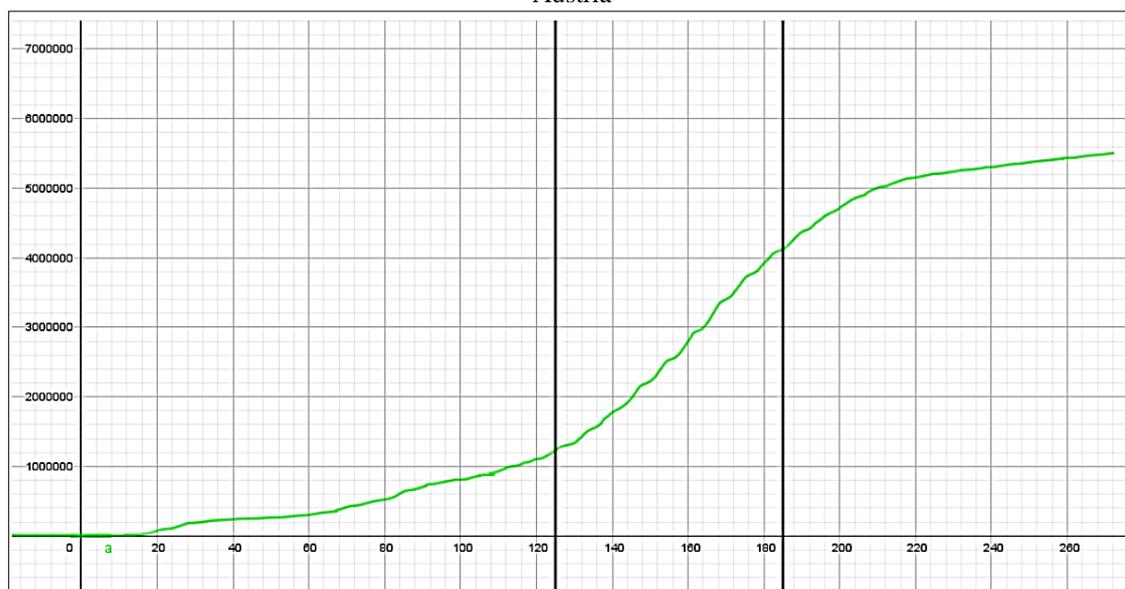
depicting the development of the total amount of fully vaccinated inhabitants in Austria.

Tradução: Pegamos os mesmos dados da exploração 1, mas usamos os dados de todo o período que o site poderia oferecer para visualizar um período máximo. Primeiro, importamos todos os dados ao GeoGebra para visualizar todos os dias como pontos na interface gráfica. Em seguida, usamos a lista de pontos para gerar uma Curva Spline descrevendo o desenvolvimento da quantidade total de habitantes totalmente vacinados na Áustria.

A explicação que descreve a utilização dos mesmos dados da exploração anterior traz consigo escolhas e procedimentos de planejamento, característicos da segunda fase do ciclo investigativo, realizada para delimitar a amostra de dados a ser estudada. Essa explicação da **Dupla 2** também evidencia, no sentido de Pfannkuch, Rubick e Yoon (2002), as medidas quantitativas relevantes para o problema, características do primeiro estágio de transnumeração.

O trabalho com os dados referente à terceira fase do ciclo investigativo fica evidente na importação ao GeoGebra e na escolha de uma forma de representar tais dados graficamente. A representação gráfica visualizada (Figura 24) apresenta o comportamento acumulativo do total de pessoas vacinadas na Áustria ao longo do tempo.

Figura 24 – Representação gráfica da curva do total acumulado de pessoas vacinadas ao longo do tempo na Áustria



Fonte: Dados da pesquisa

A curva representada pela linha verde (Figura 24) mostra o avanço no número total de vacinados ao longo do tempo. A data de início da vacinação não é evidente, todavia, a análise (quarta fase do ciclo investigativo) realizada pela **Dupla 2** indica que o eixo x se refere aos dias, desde que a vacinação foi iniciada e as retas perpendiculares ao eixo x delimitam o período de interesse da dupla de futuros professores.

Essa representação gráfica (Figura 24) pode ser identificada como uma ferramenta visual que a **Dupla 2**, ao manipular os dados, obteve e que poderia permitir inferências sobre o problema estudado. Na ótica dos estágios de transnumeração de Pfannkuch, Rubick e Yoon (2002), a representação foi empregada para entender o que os dados mostram sobre a situação em análise.

The two vertical lines are showing the first of May of 2021 and the 30th of June (the same year). As we can see, there is a huge growth in vaccinated population in this timeframe, as suspected.

Tradução: As duas retas verticais estão mostrando o 1º de maio e 30 de junho de 2021. Como podemos ver, há um enorme crescimento da população vacinada nesse período, como se suspeitava.

A representação gráfica visualizada permitiu uma análise sobre o total de pessoas vacinadas contra a Covid-19 ao longo do tempo na Áustria. Especificamente, para o período de interesse estipulado pela dupla de futuros professores, delimitado pelas retas perpendiculares, tornou evidente o comportamento crescente acentuado do total de vacinados entre os meses de maio e junho de 2021, possibilitando uma conclusão (quinta fase do ciclo investigativo) sobre como foi a evolução do total de vacinados, dada a disponibilidade das vacinas para a população.

No que se refere à transnumeração, ela pode ser vista como resultante da representação gráfica, com o auxílio das retas perpendiculares, que levou a dupla de futuros professores a uma compreensão sobre os dados estudados.

A afirmação conclusiva “*como podemos ver, há um enorme crescimento da população vacinada nesse período, como se suspeitava*” é, à luz da representação visualizada (Figura 24), um discurso conclusivo e convincente sobre o comportamento dos

dados relativos ao problema abordado, caracterizando o terceiro estágio de transnumeração, em conformidade com Pfannkuch, Rubick e Yoon (2002).

A partir do trecho citado acima, pautado na visualização de dados de Prodromou e Dunne (2017), é possível identificar que a representação gráfica visualizada permitiu *compreender* como o avanço no número de vacinados ocorreu ao longo do tempo. Ao considerar a análise gráfica de dados, o processo realizado, elucida a *finalidade* da representação gráfica, invocada pela dupla de futuros professores ao *verbalizar* sua conclusão sobre o problema.

Uma ampliação na exploração das ferramentas disponíveis para a análise gráfica dos dados foi apresentada no relatório da **Dupla 2** e trata da possibilidade de utilização das ferramentas de controle deslizante atrelada à representação de um gráfico de setores.

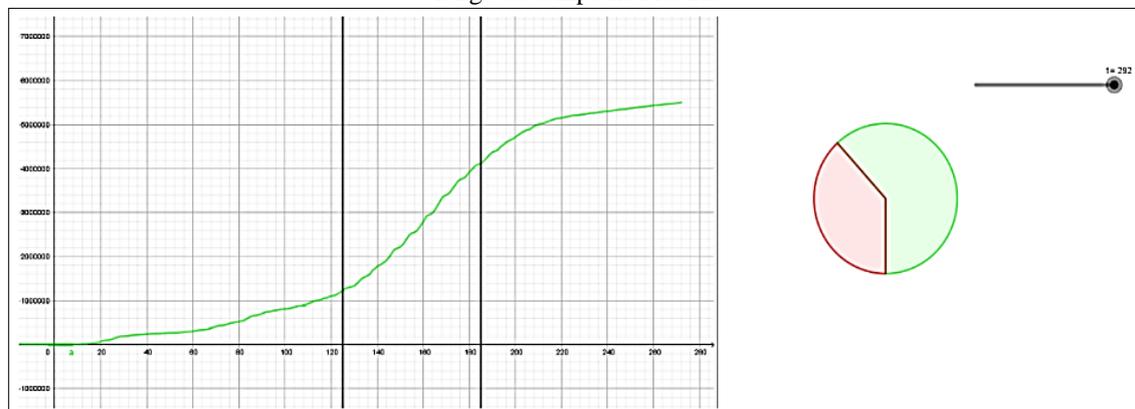
*To take the visualization one step further, we implemented a variable that could vary from 1 to 292 (=:t for "Tage") by a slide in GeoGebra. With the command Zelle (2, t) (=: b), we then generated a constant, that was counting the total amount of fully vaccinated people at that day. Then with the command Angle (Point, Point, Degrees) with Degrees: = (360*b/8917000) we generated a third point for our circle-sector that should depict the relative amount of people that are fully vaccinated. So, in the end we generated a dynamic pie-chart that depicts the relative amount of people that are fully vaccinated every day since the start of the data-collection.*

*Tradução: Para levar a visualização um passo adiante, implementamos uma variável que poderia variar de 1 a 292 (=:t para "Dias") por uma janela no GeoGebra. Com o comando Célula (2, t) (=:b), geramos então uma constante, que estava contando a quantidade total de pessoas totalmente vacinadas naquele dia. Em seguida, com o comando Ângulo (Ponto, Ponto, Graus) com Graus:=(360*b/8917000) geramos um terceiro ponto para o nosso setor circular que deve representar a quantidade relativa de pessoas que estão totalmente vacinadas. Então, no final, geramos um gráfico de pizza dinâmico que descreve a quantidade relativa de pessoas que são totalmente vacinadas todos os dias desde o início da coleta de dados.*

Tais comandos resultaram na representação gráfica que promoveu a visualização de como a vacinação avançou ao longo do tempo. Como descrito, o *t* máximo era de 292 e isso

indica que, no dia da realização da análise da dupla de futuros professores, esse era o total de dias com vacinação registrada na Áustria.

Figura 25 – Representação da curva e dos setores dinâmicos para o total de vacinados contra Covid-19 ao longo do tempo na Áustria



Fonte: Dados da pesquisa.

Embora a Dupla 2 não tenha ampliado a discussão sobre o avanço no número de austríacos vacinados ao longo do tempo, as representações gráficas (Figura 25) são exemplos das ferramentas disponíveis no GeoGebra que permitiram observar o comportamento do fenômeno ao longo do tempo.

No sentido do letramento estatístico na era *OPEN Data*, essas ferramentas poderiam contribuir para a concepção de possíveis inferências, para a organização de uma linguagem e, possivelmente, para a admissão de um modelo que representaria o fenômeno do processo de vacinação da população austríaca.

Os resultados analisados a partir da *Homework 2*, desenvolvida pelas **Dupla 1** e **Dupla 2**, possibilitam retomar a questão de pesquisa desta tese. Nesse contexto, as vantagens provenientes da utilização do GeoGebra para análise gráfica de dados podem ser observadas. Ao estudar os dados relacionados às variáveis de altura média e peso médio da população humana em 134 países, a **Dupla 1** utilizou representações gráficas como o gráfico de colunas e o *box-plot*, obtendo *insights* sobre a variação das amostras para essas duas variáveis.

Ainda, no relatório da **Dupla 1**, a representação gráfica de regressão linear possibilitada pelo GeoGebra revela como a utilização desse recurso tecnológico digital contribuiu para investigar a possível relação entre essas variáveis nos diferentes países e permitiu à dupla de futuros professores chegar a uma possível conclusão sobre a relação existente entre alturas médias e pesos médios da população dos 134 países.

Ao analisar os resultados emergentes no relatório da **Dupla 2**, as possibilidades oferecidas pela utilização do GeoGebra se configuraram como favoráveis para análise gráfica de dados. A representação gráfica (Figura 23) contribuiu para a percepção de como os dados amostrais, ao serem correlacionados, estavam dispersos, colaborando para uma possível compreensão sobre o problema estudado.

As representações do gráfico de colunas, do *box-plot* e da regressão linear podem ser consideradas contribuições do GeoGebra, permitindo aos futuros professores da **Dupla 1** visualizarem e transnumerarem os dados. As possibilidades de análises univariadas e bivariadas utilizadas pela **Dupla 1**, identificadas como ferramentas presentes no GeoGebra quando integradas às atividades de análise gráfica de dados, oferecem oportunidades para compreender os problemas subjacentes a esses dados, como foi o caso ao estudar os diferentes pesos médios e alturas médias da população humana de 134 países.

A análise realizada a partir do relatório da **Dupla 2** permite ampliar o total de ferramentas que possibilitaram as análises gráficas de dados mediadas pelo GeoGebra, como se observa nas Figura 24 e Figura 25. Os gráficos da curva *Spline* e de setores dinâmicos são exemplos de como as representações gráficas podem contribuir para esse processo.

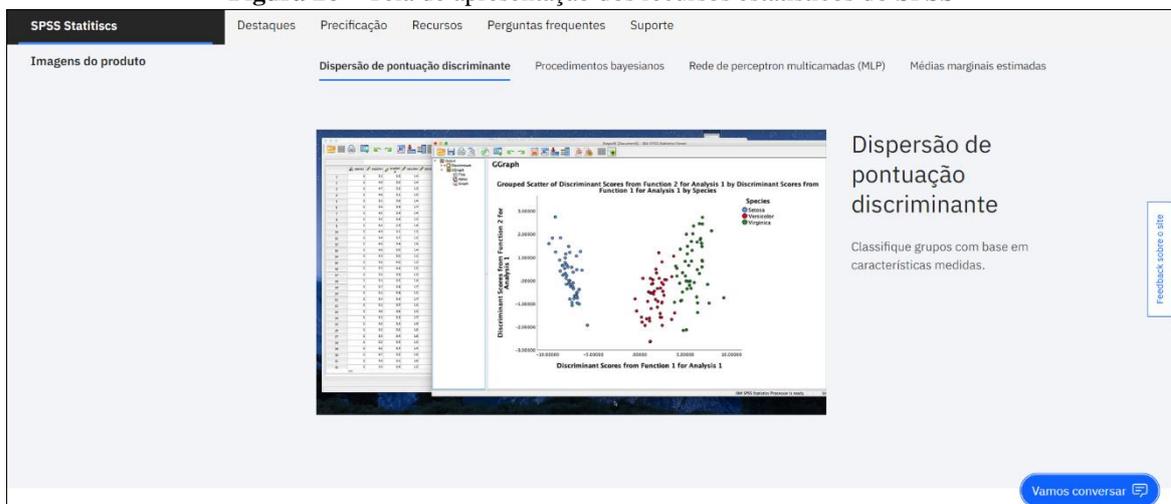
À luz dos referenciais teóricos, como analisado ao longo das atividades das duas duplas de futuros professores, foram as ferramentas de representação gráfica nos processos de transnumeração dos dados que permitiram aprimoramentos nas análises apresentadas por ambas as duplas. No contexto da visualização de dados, essas representações gráficas contribuíram para promover a compreensão dos problemas estudados.

6.3 Atividades desenvolvidas com o SPSS

Neste tópico são apresentados os resultados da atividade proposta com a utilização do SPSS. Com base nos referenciais teóricos desta tese, analisamos as atividades para explicitar as possíveis contribuições do SPSS para a análise gráfica dos dados abordados pelas **Dupla 1 e Dupla 2**.

Ao considerar o referencial metodológico do ciclo investigativo – PPPDAC, a atividade teve como objetivo permitir aos futuros professores realizarem análise de dados por meio de explorações com o SPSS. Ao longo de algumas aulas, o professor Zsolt Lavicza apresentou algumas aplicações do recurso, tecnologia que possui funcionalidades importantes para o trabalho com dados (Figura 26).

Figura 26 – Tela de apresentação dos recursos estatísticos do SPSS



Fonte: SPSS, 2023.

Com o intuito de proporcionar acesso a bancos de dados que permitissem aos futuros professores abordarem algum problema, foram fornecidos alguns links de sites na atividade, como <https://www.gapminder.org/tools/>, <https://data.worldbank.org/> e <https://www.data.gv.at>.

Os participantes poderiam acessar esses *websites* ou outros, e era esperado que eles encontrassem alguns bancos de dados relativos a problemáticas relevantes, por exemplo, envolvendo questões sociais, econômicas, saúde etc., de modo que, para explorar esses dados, fizesse sentido estabelecer problemas a serem abordados, aspecto importante para a fase inicial do ciclo investigativo.

Ao solicitar que os participantes importassem dados de pelo menos três variáveis para o SPSS, foi considerado que, com base nos estudos de correlação abordados nas aulas, os futuros professores poderiam fornecer descrições adequadas dos dados importados, utilizando as visualizações disponíveis no SPSS. Esses procedimentos contribuiriam para a segunda fase de planejamento do ciclo investigativo.

A atividade também envolvia a execução de análises de correlação no SPSS, seguindo etapas como delimitação das questões a serem estudadas, suposições sobre elas, formulação de hipóteses, cálculo de valores estatísticos, significância estatística e tamanho do efeito. O objetivo era explorar as ferramentas disponíveis para trabalhar com os dados, o que corresponde à terceira fase do ciclo investigativo.

Ao final da proposição da atividade, ao solicitar que os grupos interpretassem os resultados das correlações de pelo menos três variáveis, era esperado que as análises

características do ciclo investigativo emergissem. Essas análises seriam relevantes para a compreensão dos dados sobre os problemas estudados e úteis para formalizar uma possível conclusão sobre o contexto de origem dos problemas.

Após apresentar os delineamentos metodológicos que promoveriam o desenvolvimento da atividade *Homework 3* com a utilização do SPSS, seguem os resultados observados nos relatórios das **Dupla 1** e **Dupla 2** e as análises correspondentes, relacionadas aos estágios de transnumeração propostos por Wild e Pfannkuch (1999) e Rubick, Pfannkuch e Yoon (2002), e à visualização de dados, conforme Prodromou e Dunne (2017). Assim como nas demais atividades realizadas pela dupla, o desenvolvimento da *Homework 3* foi originalmente em inglês e traduzido para o português.

Dupla 1 - SPSS

Durante a atividade, a **Dupla 1** selecionou dados sobre as variáveis de emissões de gás carbônico (CO₂), renda per capita e população humana de 192 países no ano de 2018, obtidos a partir do *website Gapminder* (<https://www.gapminder.org/tools/>), no qual cada respectivo dado dos 192 países se refere a um valor médio para cada variável. Como mencionado na proposta da atividade, todos os bancos de dados disponíveis no *website Gapminder* foram utilizados. A dupla já estava familiarizada com o acesso a esses dados, o que possibilitou o *download* das planilhas do *Excel* contendo os dados, para posterior importação para o SPSS.

Ao delimitar as variáveis a serem estudadas, a **Dupla 1** evidencia características da fase inicial do ciclo investigativo. Além disso, ao considerar a amostra de dados dos 192 países, são observados os delineamentos necessários para o primeiro estágio de transnumeração. Embora os dados sejam médias sobre as variáveis em cada país, eles podem ser relevantes para uma compreensão mais ampla das variáveis em um contexto mundial, o que é importante no contexto do letramento estatístico na era *OPEN Data*.

Não foi possível observar no relatório da **Dupla 1** o trabalho com os dados importados para o SPSS, o que diz respeito à terceira fase do ciclo investigativo. No entanto, a dupla de futuros professores considerou as representações gráficas e as estatísticas resultantes dessa fase, conforme visualizadas na janela do SPSS.

A primeira análise realizada pela **Dupla 1** foi a suposição de normalidade dos dados, como apresentado na Figura 27. É importante observar que a configuração da versão do SPSS utilizada pela dupla estava em alemão.

Na Figura 27, a **Dupla 1** apresenta resultados sobre a distribuição dos dados das três variáveis. Estatísticas como desvio padrão, curtose e desvio padrão da curtose, podem auxiliar na compreensão da distribuição dos dados, embora o termo mais adequado para curtose seria assimetria.

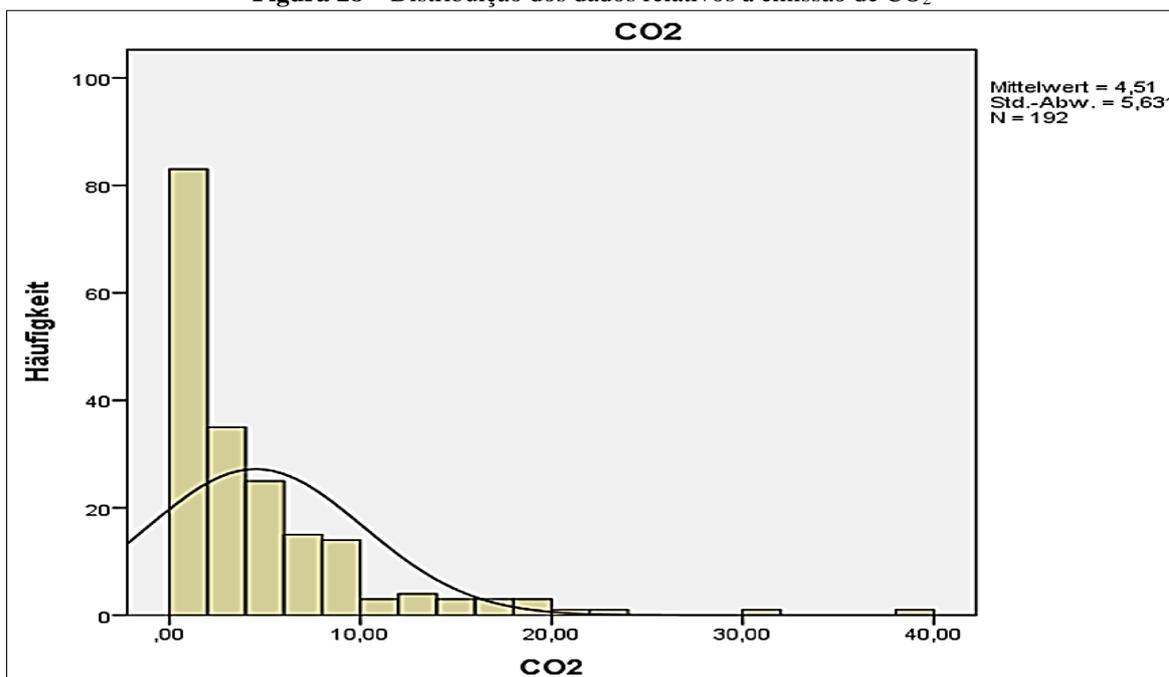
Figura 27 – Estatísticas das distribuições dos dados estudados

		Statistiken		
		CO2	Income	Population
N	Gültig	192	192	192
	Fehlend	0	0	0
Schiefe		2,595	1,604	9,733
Standardfehler der Schiefe		,175	,175	,175
Kurtosis		9,268	2,780	93,785
Standardfehler der Kurtosis		,349	,349	,349

Fonte: Dados da pesquisa.

Em complemento às estatísticas, a **Dupla 1** recorre às representações gráficas que permitem visualizar como se distribuem os dados sobre a emissão de CO₂, sobre renda *per capita* e sobre a população humana dos 192 países.

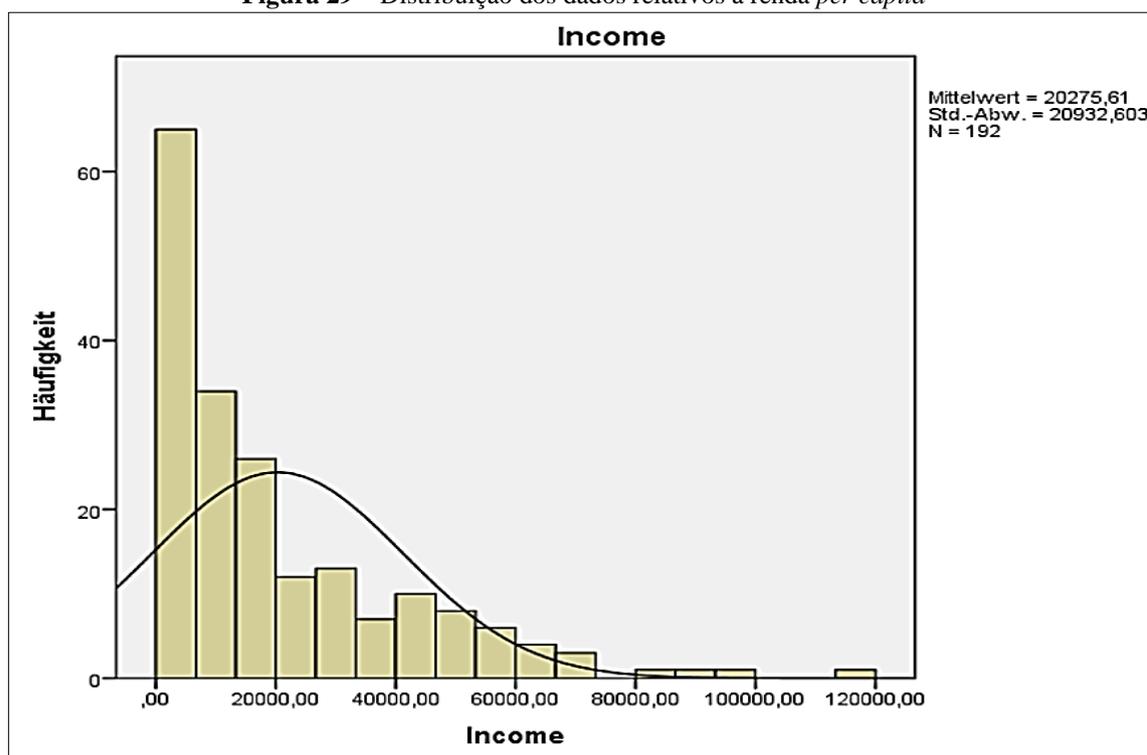
Figura 28 – Distribuição dos dados relativos à emissão de CO₂



Fonte: Dados da pesquisa.

Na representação gráfica (Figura 28), é possível visualizar a distribuição da frequência dos países em relação à quantidade de CO₂ emitida na atmosfera. Embora não seja evidente na figura, é importante destacar que a quantidade de CO₂ é medida em toneladas por pessoa, como informado no *link* de origem dos dados. Nesse sentido, a distribuição de frequência permite observar a distribuição dos países de acordo com a quantidade de CO₂ emitida por habitante.

Na Figura 29, o relatório da **Dupla 1** apresenta uma representação gráfica para a distribuição dos dados relacionados à renda *per capita* em dólares americanos (US\$) nos 192 países.

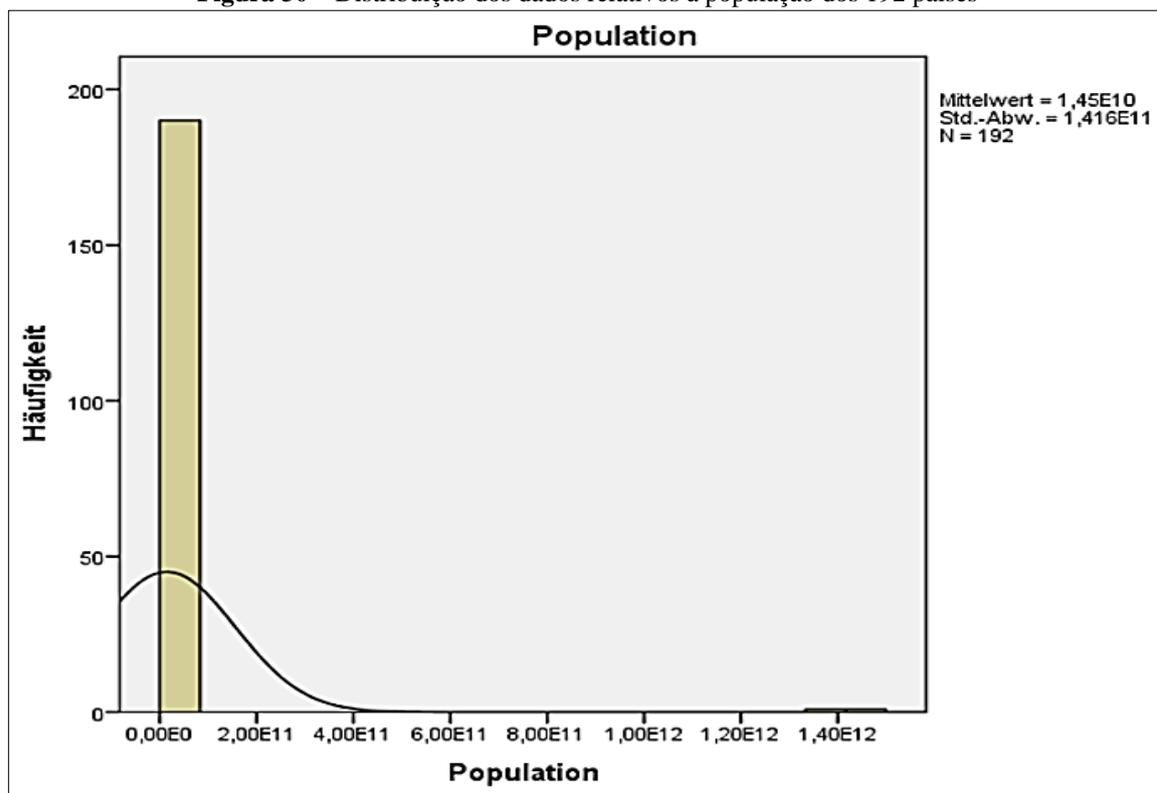
Figura 29 – Distribuição dos dados relativos à renda *per capita*

Fonte: Dados da pesquisa.

A distribuição das frequências de países, conforme a renda *per capita* (Figura 29) permite visualizar como nos diferentes países a renda pode variar e como a curva representada indica uma curtose (assimetria) à direita, mostrando que há uma concentração de países com renda *per capita* menor que 2000,00 US\$.

Na Figura 30, está representada a distribuição de frequência dos países, conforme suas populações.

Figura 30 – Distribuição dos dados relativos à população dos 192 países



Fonte: Dados da pesquisa.

Ao visualizar a representação gráfica da distribuição de frequência dos dados sobre as populações dos diferentes países, assim como as distribuições sobre a emissão de CO₂ e de renda *per capita*, a **Dupla 1** considera:

One can see in the three pictures, that the data are approximately but not exactly normal distributed, as assumed.

Tradução: Pode-se ver nas três figuras, que os dados são aproximadamente, mas não exatamente normais distribuídos, como assumido.

Algumas outras estatísticas poderiam ter sido exploradas, como média e desvio padrão, que acompanham cada representação gráfica visualizada. Mas, ao observar as distribuições presentes nessas representações, a **Dupla 1** inferiu que os dados não representavam, necessariamente, distribuições normais, pois pode-se verificar que a representação gráfica não oferece condições para inferir sobre a normalidade da distribuição dos dados. Tecidas essas observações, que resultam do trabalho com os dados, terceira fase

do ciclo investigativo, algumas hipóteses foram apresentadas pela dupla de futuros professores.

The higher the human population the higher the CO₂ emissions, namely there is a correlation between the CO₂ emissions and the human population.

The higher the income the higher the CO₂ emissions, namely there is a correlation between the income and the human population.

The higher the population the lower the income, namely there is a correlation between the population and the income.

Tradução: *Quanto maior a população humana, maiores as emissões de CO₂, ou seja, existe uma correlação entre as emissões de CO₂ e a população humana.*

Quanto maior o rendimento, maiores as emissões de CO₂, ou seja, existe uma correlação entre o rendimento e a população humana.

Quanto maior a população, menor a renda, ou seja, há uma correlação entre a população e a renda.

A partir dessas hipóteses, os procedimentos para calcular valores estatísticos adequados para verificar a força das correlações, a significância e o tamanho do efeito foram realizados por intermédio do SPSS. Todos esses procedimentos não estão evidentes no relatório da **Dupla 1**, mas deles é que resultam os valores apresentados na Figura 31.

Figura 31 – Valores estatísticos relativos ao estudo de correlação das variáveis estudadas

		Korrelationen		
		CO2	Income	Population
CO2	Korrelation nach Pearson	1	,683**	,001
	Signifikanz (2-seitig)		,000	,986
	N	192	192	192
Income	Korrelation nach Pearson	,683**	1	-,046
	Signifikanz (2-seitig)	,000		,529
	N	192	192	192
Population	Korrelation nach Pearson	,001	-,046	1
	Signifikanz (2-seitig)	,986	,529	
	N	192	192	192

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Fonte: Dados da pesquisa.

Os valores presentes na tabela visualizada pela dupla de futuros professores (Figura 31) permitiram as seguintes análises:

Based on the results of our analysis with SPSS, there is a high positive correlation between the income and the CO₂ emissions of the countries. There is no remarkable correlation between the CO₂ and human population. There is a slightly negative correlation between the income and human population.

Tradução: Com base nos resultados de nossa análise com o SPSS, há uma alta correlação positiva entre a renda e as emissões de CO₂ dos países. Não há correlação notável entre a emissão de CO₂ e a população humana. Há uma correlação ligeiramente negativa entre a renda e a população humana.

Os apontamentos apresentados pela **Dupla 1** elucidam análises, características da quarta fase do ciclo investigativo. O conhecimento das propriedades relativas aos valores das estatísticas de correlação, por exemplo, relacionadas ao nível de significância, possibilitou à dupla de futuros professores estipular a força de correlação entre as variáveis.

Sob o ponto de vista da transnumeração, as representações gráficas que buscavam supor a normalidade dos dados estudados possibilitaram que a dupla compreendesse com se distribuía as frequências desses dados, a depender da variável. A escolha dessas representações gráficas, assim como as medidas apontadas na tabela (Figura 31), foi relevante para a dupla de futuros professores entender o que os dados mostravam sobre os 192 países, considerando as emissões de CO₂, a renda *per capita* e a população e as possíveis correlações existentes entre eles. Esses aspectos podem ser caracterizados como elementos do segundo estágio de transnumeração, em conformidade com Rubick, Pfannkuch e Yoon (2002).

Cabe salientar que o SPSS disponibiliza a representação gráfica que permite visualizar as dispersões dos dados no estudo das correlações. Mas, para a **Dupla 1**, apenas os valores apresentados na tabela foram suficientes para transnumerar e organizar, no sentido do terceiro estágio de transnumeração, definido por Rubick, Pfannkuch e Yoon (2002), um

discurso sucinto de análise, de forma convincente, sobre as correlações existentes entre as variáveis emissões de CO₂, sobre a renda *per capita* e sobre a população nos 192 países.

Sob o ponto de vista da visualização de dados, de acordo com Prodromou e Dunne (2017), a representação gráfica das frequências das variáveis auxiliou nas suposições de normalidade dos dados, contribuindo para a compreensão dessa propriedade estatística a partir dos dados. No discurso escrito “*Pode-se ver nas três figuras, que os dados são aproximadamente, mas não exatamente normais distribuídos, como assumido*”, percebe-se a *finalidade* das representações gráficas resultantes da manipulação dos dados no SPSS, que auxiliaram na *verbalização* das distribuições das variáveis relacionadas aos diferentes países.

O recurso da representação gráfica poderia ter sido utilizado no estudo das correlações de forma pontual e assertiva, mas, pelos valores estatísticos de significância entre as variáveis correlacionadas, a **Dupla 1** tece uma análise conclusiva sobre as correlações estudadas.

Na perspectiva do letramento estatístico na era *OPEN Data*, o próprio contexto a que se referem os dados dos 192 países, acessados a partir do *Gapminder*, aponta para a dimensão representativa que esses dados têm.

Ao identificar as diferentes distribuições dos dados e apresentar algumas hipóteses sobre como essas variáveis poderiam se correlacionar, a dupla identifica a possibilidade de realizar *inferências plausíveis* a partir desses dados.

Os resultados emergentes da manipulação com os dados, os valores que apontam para os níveis de significância no estudo de correlação, por exemplo, foram relevantes para a compreensão do *contexto* de origem. De fato, esses valores estatísticos permitem avaliar a qualidade da evidência em relação às correlações existentes entre as variáveis.

Em relação à questão de pesquisa que guia esta tese, o relatório da *Homework 3*, elaborado pela **Dupla 1**, indica que a utilização do SPSS contribuiu para realizar suposições sobre as distribuições dos dados no contexto da análise gráfica. A tabela (Figura 18) permite observar que, por meio de uma única manipulação, a dupla de futuros professores obteve resultados estatísticos para o estudo das correlações entre as três variáveis: emissão de CO₂, renda *per capita* e população humana dos 192 países. Seria possível ampliar o estudo para incluir mais variáveis, revelando o potencial da utilização do recurso tecnológico digital SPSS na atividade.

Assim, as representações gráficas para as distribuições dos dados estudados e a tabela resultante do estudo das correlações, trazidos nos resultados obtidos a partir do relatório, são vistos como contributos do SPSS para análise dos dados pela **Dupla 1**.

Dupla 2 – SPSS

Em relação à primeira solicitação da *Homework 3*, para estabelecer um banco de dados, a **Dupla 2** acessou o site <https://data.worldbank.org/>, que foi uma das sugestões apresentadas na atividade. O interesse da dupla de futuros professores era explorar dados sobre variáveis educacionais. O tópico que a **Dupla 2** apresentou em seu relatório trata da razão aluno-professor e seus possíveis efeitos em outras variáveis.

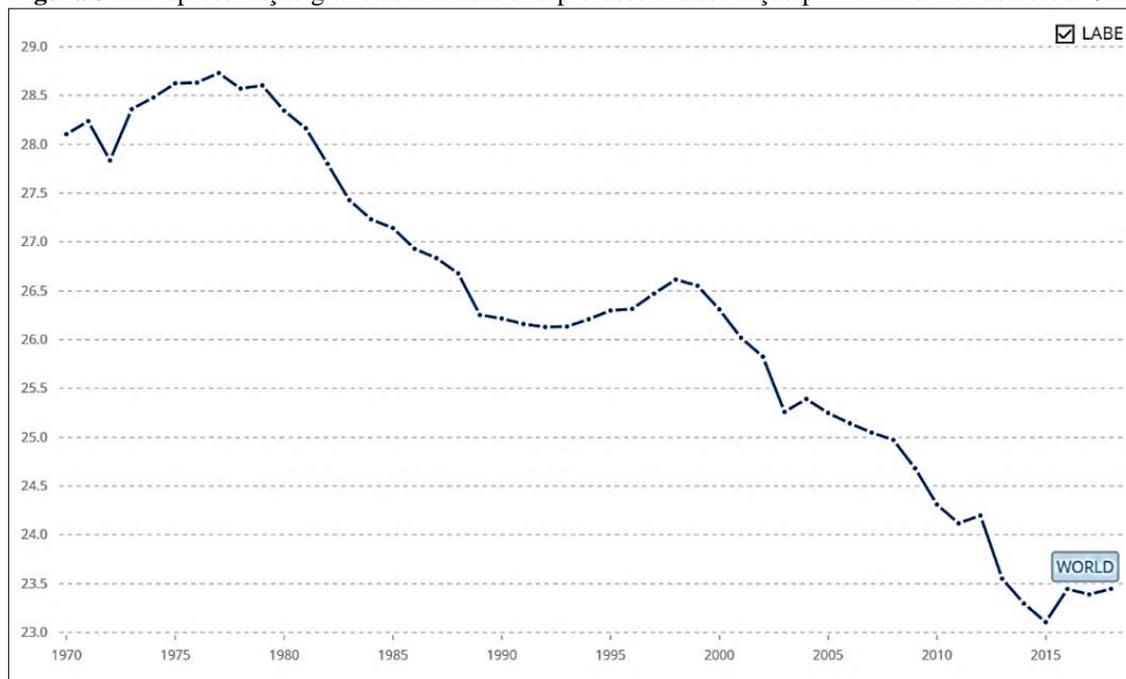
One of the most interesting variables and datasets for us was the variable of “pupil-teacher ratio” in primary and secondary education. It states, how many pupils a teacher has to teach in class, on average. One could assume that the lower the ratio, the easier it is for the teacher to focus on individual pupils and the better the scholastic performance of each pupil could be. Because the specific dataset for Austria is lacking completeness, we chose to look at the pupil-teacher ratio and its development regarding the whole world.

Tradução: Uma das variáveis e conjuntos de dados mais interessantes para nós foi a variável "razão aluno-professor" no ensino primário e secundário. Ela diz, quantos alunos um professor tem que ensinar em sala de aula, em média. Pode-se supor que quanto menor a proporção, mais fácil é para o professor se concentrar em alunos individuais e o desempenho escolar de cada aluno poderia ser melhor. Uma vez que o conjunto de dados específico para a Áustria carece de exaustividade, optamos por analisar a razão aluno-professor e o seu desenvolvimento em relação ao mundo inteiro.

É relevante salientar que a escolha por estudar o problema com base em dados em um nível mundial é compreendida como uma forma de representar esses dados e atribuir significados aos resultados obtidos, o que pode permitir identificar elementos relacionados às proposições de Prodromou e Dunne (2017) sobre questões do letramento estatístico na era *OPEN Data*.

Como resultado da exploração inicial realizada pela **Dupla 2**, o relatório traz uma representação gráfica disponibilizada pelo *site* acessado, na qual é possível visualizar a variação da média de aluno por professor na educação primária desde 1970.

Figura 32 – Representação gráfica da razão aluno-professor na educação primária no mundo desde 1970



Fonte: Dados da pesquisa.

A representação gráfica visualizada (Figura 32) permite identificar que, ao longo do tempo, a média de alunos por professor em sala de aula tem reduzido. E, diante desse resultado, para o estabelecimento de uma problemática a ser estudada com mediação do SPSS, seguindo as solicitações da atividade, a dupla de futuros professores indicou:

We wanted to know whether the pupil-teacher-ratio in some way correlates with different variables such as overall literacy and secondary school-enrolment (perhaps having more individual teaching results in one being more likely tending to secondary education).

Tradução: *Queríamos saber se a razão aluno-professor de alguma forma se correlaciona com diferentes variáveis, como a alfabetização geral e a matrícula no ensino secundário (talvez ter mais ensino individual resulte em uma tendência mais provável para o ensino secundário).*

Na descrição apresentada, é possível identificar a problemática a ser abordada pela dupla de futuros professores, característica da primeira fase do ciclo investigativo. Como sugere a atividade, o estudo de correlações seria pertinente, e a seleção de dados relativos a pelo menos três variáveis deveria possibilitar esse estudo, ou seja, no sentido da segunda fase do ciclo investigativo – PPDAC, algo esperado ao longo do planejamento da dupla de futuros professores para analisar os dados.

Em relação aos estágios de transnumeração de Pfannkuch, Rubick e Yoon (2002), ao delimitar as amostras relativas a pelo menos três variáveis relacionadas ao problema e admitir a possibilidade de estudá-las por meio de correlações, esses procedimentos trazem consigo o interesse de parametrizar como as variáveis se correlacionam, como prevê o primeiro estágio de transnumeração. O planejamento realizado expõe o processo seguinte, que sugere o levantamento, a organização e a gestão dos dados.

We first downloaded the datasets from the website and copied the world-specific rows into a new excel-sheet in order to transpose them from horizontal to vertical alignment. After this transposition we inserted the selected data into SPSS for further analyzation.

Tradução: Primeiro, baixamos os conjuntos de dados do site e copiamos as linhas específicas do mundo em uma nova planilha do Excel para transpô-las do alinhamento horizontal para o vertical. Após essa transposição, inserimos os dados selecionados no SPSS para posterior análise.

Diante dos dados importados e organizados no SPSS, a **Dupla 2** estabeleceu um conjunto de hipóteses que norteou o trabalho com os dados, característica da terceira fase do ciclo investigativo, de modo a analisá-los por meio do estudo de correlações.

There is a significant negative connection (correlation) between pupil-teacher-ratio in primary education and general literacy (because we assume, that having less pupils makes teaching individually easier for the teacher resulting in better effects on learning for the pupils)

There is a significant negative connection (correlation) between pupil-teacher-ratio in primary education and enrolment in secondary education (because we

assume that being taught more individually results in a positive connotation for education in general, making it more attractive to enrol in institutions for secondary education).

Pupil-teacher-ratio in primary education correlates positively with pupil-teacher-ratio in secondary education (because we assume, that the changes in pupil-teacher-ratio are due to general changes in education-politics and should therefore change in the same way).

Tradução: Existe uma significativa conexão negativa (correlação) entre a razão aluno-professor no ensino primário e o letramento geral (porque assumimos que ter menos alunos torna o ensino individual mais fácil para o professor, resultando em melhores efeitos na aprendizagem para os alunos).

Existe uma significativa conexão negativa (correlação) entre a razão aluno-professor no ensino primário e a matrícula no ensino secundário (porque assumimos que ser ensinado mais individualmente resulta em uma conotação positiva para a educação em geral, tornando mais atraente matricular-se em instituições para o ensino secundário).

A razão aluno-professor no ensino primário correlaciona-se positivamente com a razão aluno-professor no ensino secundário (porque assumimos que as alterações na razão aluno-professor se devem às mudanças gerais na educação-política e, por conseguinte, devem mudar da mesma forma).

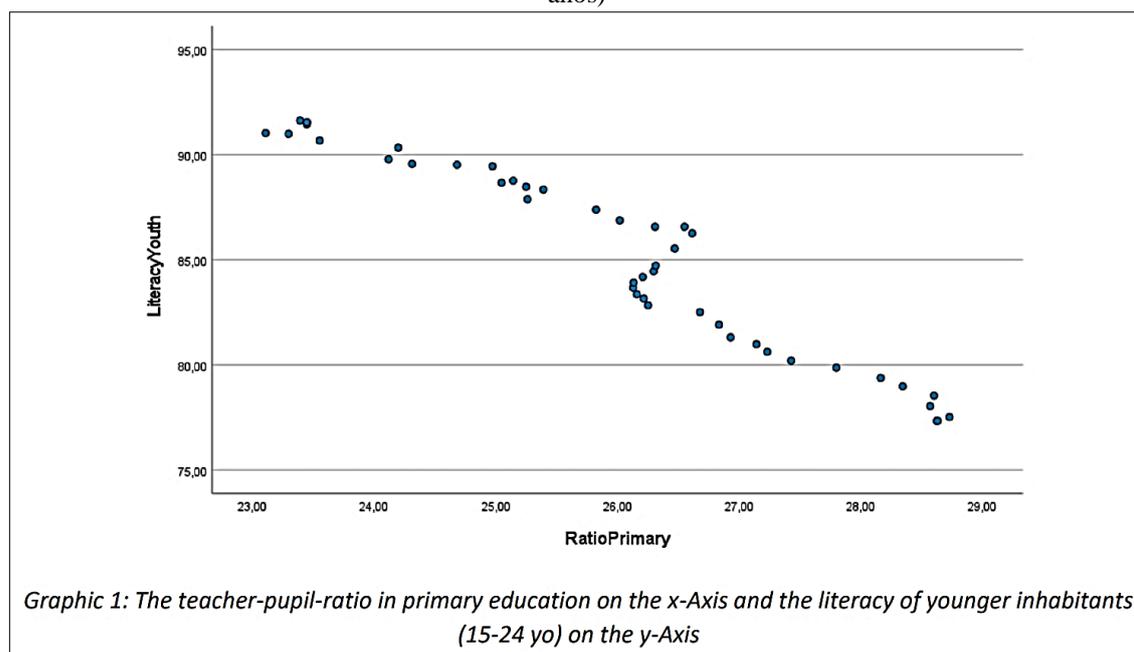
O conjunto de hipóteses estabelecido viabilizou escolhas quanto à terceira fase do ciclo investigativo. A descrição no relatório da **Dupla 2** mostra que o trabalho com a utilização dos dados com o SPSS possibilitou a construção de algumas representações gráficas para visualizar as suas dispersões. Como sugere Pfannkuch, Rubick e Yoon (2002), escolhas dessa natureza dizem respeito aos procedimentos realizados no segundo estágio de transnumeração.

First of all, we wanted to visualize some scatterplots in order to vaguely check some of our assumptions and to see if we would have to eliminate statistical anomalies. In order to do that we used the scatter-plot tool in SPSS.

Tradução: *Primeiro de tudo, queríamos visualizar alguns gráficos de dispersão para verificar vagamente algumas de nossas suposições e ver se teríamos que eliminar anomalias estatísticas. Para isso, utilizamos a ferramenta gráfico de dispersão no SPSS.*

O relatório da **Dupla 2** não explica os procedimentos realizados para a manipulação das amostras, mas atendendo a seu anseio por visualizar graficamente as possíveis dispersões dos dados, as Figuras 33, 34 e 35 mostram o resultado desses procedimentos, ao correlacionar as variáveis razão aluno-professor no ensino primário, letramento de jovens (15-24 anos), matrícula no ensino secundário (em porcentual) e razão aluno-professor no ensino secundário.

Figura 33 – Representação gráfica da dispersão entre razão aluno-professor e o letramento de jovens (15-24 anos)

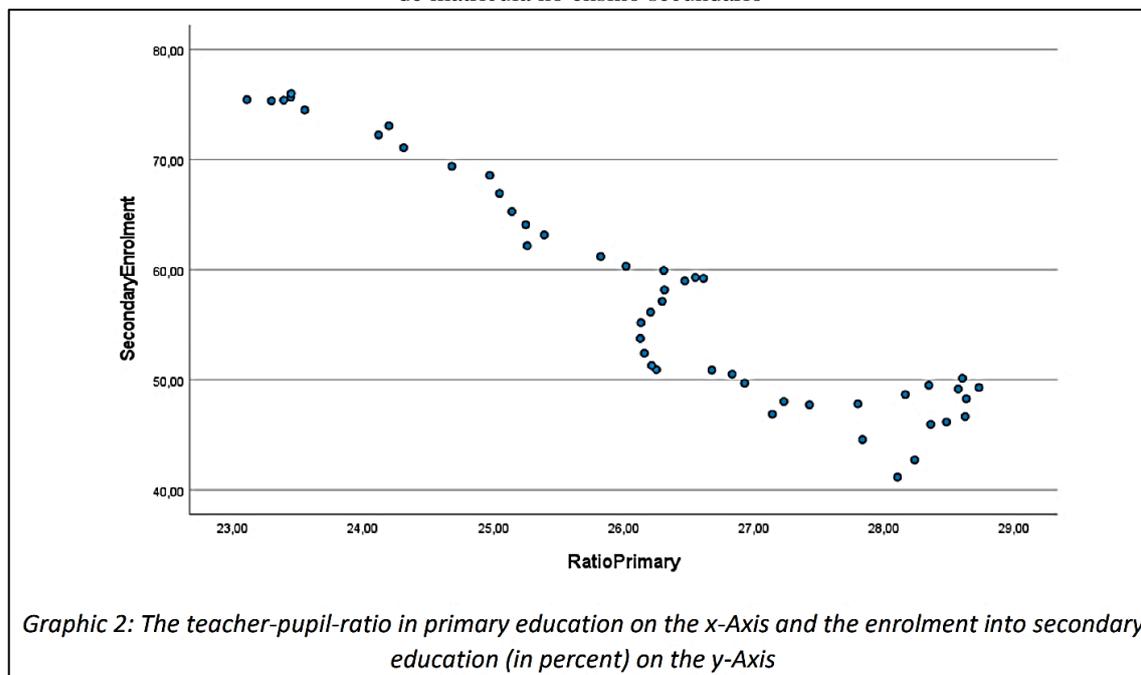


Fonte: Dados da pesquisa.

A representação gráfica (Figura 33), resultado da manipulação dos dados pela **Dupla 2** no SPSS, mostra a dispersão dos dados relativos à razão aluno-professor em relação ao letramento observado entre jovens de 15 a 24 anos. Como é possível visualizar, a representação indica a existência de uma correlação negativa entre as duas variáveis.

Na Figura 34, está uma representação gráfica sobre a dispersão resultante da manipulação dos dados das variáveis razão aluno-professor no ensino primário e porcentual de matrícula no ensino secundário.

Figura 34 – Representação gráfica da dispersão entre razão aluno-professor no ensino primário e porcentual de matrícula no ensino secundário



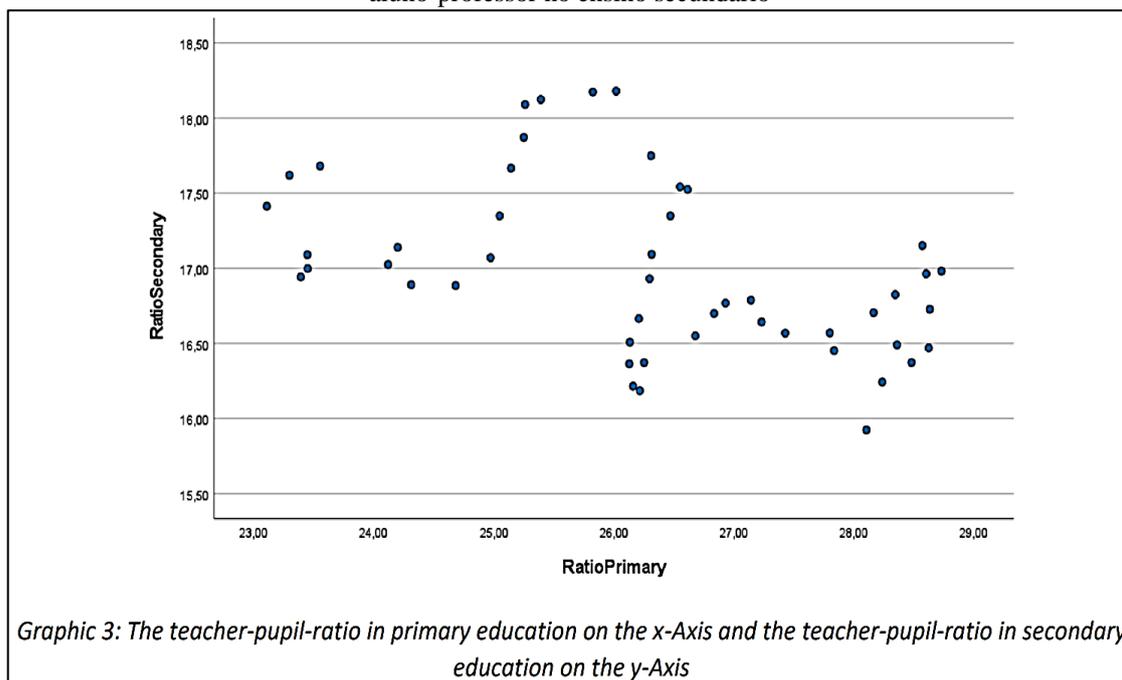
Fonte: Dados da pesquisa.

A representação gráfica (Figura 34), disponível a partir da manipulação no SPSS com os dados das variáveis entre razão aluno-professor no ensino primário e porcentual de matrícula no ensino secundário, permite visualizar, assim como no caso da Figura 33, a existência de uma correlação negativa entre as variáveis.

Seguindo o trabalho com os dados amostrados, a representação gráfica seguinte (Figura 35) é resultado do estudo envolvendo os dados das variáveis razão aluno-professor no ensino primário e razão aluno-professor no ensino secundário.

Como é possível observar, a correlação entre as duas variáveis, embora possa existir, na representação disponibilizada pelo SPSS e apresentada pela **Dupla 2** em seu relatório, não se revela claramente qual o comportamento dessa possível correlação.

Figura 35 – Representação gráfica da dispersão entre a razão aluno-professor no ensino primário e a razão aluno-professor no ensino secundário



Fonte: Dados da pesquisa.

As representações gráficas visualizadas, resultado do trabalho com os dados (terceira fase do ciclo investigativo), podem contribuir para a compreensão sobre as possíveis correlações entre as variáveis estudadas. No sentido da transnumeração definida por Wild e Pfannkuch (1999), essas mudanças de representação dos dados oferecem meios para melhor compreendê-los. No relatório da dupla de futuros professores, há algumas considerações sobre isso:

The first two charts could be a hint, that our first two hypotheses could be right (note that this graphical analysis is, like we stated, only a hint in the right direction). The last one is not that easy to interpret though. In order to really say something significant about these hypotheses, we have to calculate the respective correlations. So again, we worked with SPSS to see if we could show significant correlations:

Tradução: Os dois primeiros gráficos poderiam ser uma dica, que nossas duas primeiras hipóteses poderiam estar certas (note que esta análise gráfica é, como afirmamos, apenas uma dica na direção certa). A última, porém, não é tão fácil de interpretar. Para realmente dizer algo significativo sobre essas hipóteses,

temos que calcular as respectivas correlações. Então, novamente, trabalhamos com o SPSS para ver se poderíamos mostrar correlações significativas

A análise apresentada pela **Dupla 2** salienta a contribuição das representações gráficas visualizadas. Ao direcionar a análise para o conjunto de hipóteses previamente delimitado, observa-se que essas representações são indicativos prévios sobre a possível aceitação ou rejeição das respectivas hipóteses. No contexto das fases do ciclo investigativo, a utilização das ferramentas de representação gráfica do SPSS foi relevante para a dupla de futuros professores seguir em suas análises.

No que se refere à visualização de dados, as representações gráficas permitiram à dupla de futuros professores uma *compreensão* sobre como os dados amostrados se dispersam. Uma *finalidade* da representação gráfica foi oferecer indicativos sobre as dispersões, e isso resulta na *verbalização* presente no discurso (trecho anterior) da **Dupla 2**.

Assim como a **Dupla 1**, para a executar o estudo de correlação entre as variáveis, a **Dupla 2** recorreu não só às ferramentas de representação gráfica do SPSS, mas também aos resultados obtidos para valores críticos, representados de forma tabular, que favoreceram as análises tecidas pela dupla de futuros professores.

A Figura 36 explicita os resultados obtidos para o estudo de correlação dos dados entre as variáveis razão aluno-professor em relação ao letramento observado entre jovens de 15 a 24 anos. Como no caso da **Dupla 1**, a configuração do recurso tecnológico digital SPSS, utilizado pela **Dupla 2**, também estava em idioma alemão.

Figura 36 – Valores estatísticos relativos ao estudo de correlação entre razão aluno-professor e letramento entre jovens de 15 a 24 anos

Korrelationen			
		RatioPrimary	LiteracyYouth
RatioPrimary	Pearson-Korrelation	1	-,962**
	Sig. (2-seitig)		<,001
	Quadratsummen und Kreuzprodukte	134,733	-303,994
	Kovarianz	2,807	-7,070
	N	49	44
LiteracyYouth	Pearson-Korrelation	-,962**	1
	Sig. (2-seitig)	<,001	
	Quadratsummen und Kreuzprodukte	-303,994	967,935
	Kovarianz	-7,070	21,510
	N	44	46

Fonte: Dados da pesquisa.

Uma diferença observada nas manipulações realizadas pelas **Dupla 1** e **Dupla 2** foi que a primeira fez o estudo das correlações tomando os dados de todas as variáveis de uma única vez, enquanto a segunda dupla correlacionou os dados das variáveis, tomando duas a duas.

Em complemento às análises realizadas, a partir das representações gráficas, no sentido da quarta fase do ciclo investigativo, com base nos valores críticos visualizados na tabela de resultados do estudo de correlações, apresentada pelo SPSS, identifica-se um discurso analítico que leva em conta esses valores críticos.

Com base nesses resultados, as análises emitidas pela **Dupla 2** sobre a possível correlação existente entre as variáveis razão aluno-professor em relação ao letramento observado entre jovens de 15 a 24 anos foram as seguintes:

The first correlation that we calculated is the correlation between the variable of the pupil-teacher-ratio in primary education and the youth-literacy (15-24 yo). In the chart we see that there is a strong negative correlation between pupil-teacher-ratio and literacy of the young population (15-24 yo), $r = -.962$, $p < 0.001$.

Tradução: A primeira correlação que calculamos é a correlação entre a variável da razão aluno-professor no ensino primário e o letramento de jovens (15-24 anos). No gráfico vemos que há uma forte correlação negativa entre a relação aluno-professor e o letramento da população jovem (15-24 anos), $r = -.962$, $p < 0,001$.

A análise de correlação entre a razão aluno-professor em relação ao letramento observado entre jovens de 15 a 24 anos apresentada pela **Dupla 2** salienta a visualização gráfica (Figura 33) e os valores críticos, coeficiente de correlação r de Pearson e valor- p , resultantes na representação tabular, úteis para parametrizar a possível correlação existente, embora a dupla de futuros professores, nesse trecho da análise, não tenha discutido esses valores críticos.

Na sequência, estão os resultados apresentados pela **Dupla 2** ao integrar o SPSS para o estudo de correlação entre os dados das variáveis razão aluno-professor no ensino primário e porcentual de matrícula no ensino secundário.

Figura 37 - Valores estatísticos relativos ao estudo de correlação entre razão aluno-professor no ensino primário e porcentual de matrícula no ensino secundário

		RatioPrimary	SecondaryEnrolment
RatioPrimary	Pearson-Korrelation	1	-.944**
	Sig. (2-seitig)		<.001
	Quadratsummen und Kreuzprodukte	134,733	-792,757
	Kovarianz	2,807	-16,516
	N	49	49
SecondaryEnrolment	Pearson-Korrelation	-.944**	1
	Sig. (2-seitig)	<.001	
	Quadratsummen und Kreuzprodukte	-792,757	5571,870
	Kovarianz	-16,516	113,712
	N	49	50

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Fonte: Dados da pesquisa.

Como é possível visualizar na tabela resultante (Figura 37), os valores críticos indicam a existência de correlação entre duas variáveis estudadas. Sobre isso, a dupla de futuros professores apresenta a seguinte análise:

We calculated the correlation between the variable of the pupil-teacher-ratio in primary education and enrolment in secondary education (in percent). We calculated that there is a strong negative correlation between pupil-teacher-ratio and enrolment in secondary education, $r = -.944$, $p < 0.001$.

Tradução: Calculamos a correlação entre a variável da relação aluno-professor no ensino primário e a matrícula no ensino secundário (em porcentagem). Calculou-se que há uma forte correlação negativa entre a relação aluno-professor e a matrícula no ensino secundário, $r = -.944$, $p < 0,001$.

A análise da **Dupla 2** salienta a existência de correlação entre as variáveis razão aluno-professor no ensino primário e porcentual de matrícula no ensino secundário. Ao indicar essa correlação, seu comportamento e admitir a intensidade, é possível supor que a **Dupla 2**, embora não tenha mencionado, atribuiu esse comportamento e intensidade considerando o valor de r observado. Além disso, poderia levar em consideração o valor- p para atribuir o nível de significância da correlação existente.

Sobre o estudo de correlação entre as variáveis razão aluno-professor no ensino primário e razão aluno-professor no ensino secundário, os resultados da manipulação dos dados foram apresentados na Figura 38.

Figura 38 - Valores estatísticos relativos ao estudo de correlação entre razão aluno-professor no ensino primário e razão aluno-professor no ensino secundário

		RatioPrimary	RatioSecondary
RatioPrimary	Pearson-Korrelation	1	-,459**
	Sig. (2-seitig)		<,001
	Quadratsummen und Kreuzprodukte	134,733	-20,922
	Kovarianz	2,807	-,436
	N	49	49
RatioSecondary	Pearson-Korrelation	-,459**	1
	Sig. (2-seitig)	<,001	
	Quadratsummen und Kreuzprodukte	-20,922	15,399
	Kovarianz	-,436	,321
	N	49	49

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Fonte: Dados da pesquisa.

Diante dos resultados obtidos, no trecho seguinte a **Dupla 2** indica a existência da correlação entre as variáveis razão aluno-professor no ensino primário e razão aluno-professor no ensino secundário.

We calculated the correlation between the variable of the pupil-teacher-ratio in primary education and pupil-teacher-ratio in secondary education. We calculated that there is a negative correlation between pupil-teacher-ratio in primary education and pupil-teacher-ratio in secondary education, $r = -.456$, $p < 0.001$.

Tradução: Calculamos a correlação entre a variável razão aluno-professor no ensino primário e a razão aluno-professor no ensino secundário. Calculou-se que há uma correlação negativa entre a razão aluno-professor no ensino primário e a razão aluno-professor no ensino secundário, $r = -.456$, $p < 0,001$.

A admissão da dupla de futuros professores sobre a correlação negativa existente entre as duas variáveis é uma evidência de que o coeficiente de correção r exibido na tabela contribuiu. O conhecimento sobre o significado e as propriedades relativas ao valor de r se mostraram relevantes, visto que, na representação gráfica (Figura 35), essa correlação não estava explícita.

No segundo estágio de transnumeração definido por Pfannkuch, Rubick e Yoon (2002), o r e valor- p obtidos, assim como a representação gráfica mencionada pela dupla de futuros professores, em cada um dos estudos de correlação, podem ser concebidos como resultados que emergem da utilização do SPSS, contribuindo para entender a correlação existente entre os dados amostrais das variáveis razão aluno-professor e letramento entre jovens de 15 a 24 anos.

Sob o ponto de vista da visualização de dados, como definida por Prodromou e Dunne (2017), os valores críticos, como as representações gráficas apresentadas inicialmente, foram relevantes para a *compreensão* de como os dados amostrados se correlacionavam. No sentido das *finalidades* dos resultados numéricos e gráficos visualizados, observa-se que, para *verbalizar* as correlações entre as variáveis, tais resultados se mostraram efetivos no discurso da dupla de futuros professores.

Ao apresentar conclusões para o problema abordado (quinta fase do ciclo investigativo) e investigar as possíveis correlações entre a razão aluno-professor com diferentes variáveis, como o letramento geral e a matrícula no ensino secundário, a **Dupla 2** retoma cada hipótese delimitada no início da atividade e tece algumas considerações. Sobre a primeira hipótese, a dupla de futuros professores relata:

The first hypothesis, that there is a negative connection between the pupil-teacher-ratio in primary education and the literacy in the younger part of the population (15–24-year-olds), could be assumed to be true, due to the significant correlation-coefficient of -0.962 . Why is that the case? Because of the nature of correlation, we cannot say anything about the causation of this connection, but we could still hypothesize why this connection came to be. As we stated, we think that there even could be a causal-relationship in the sense, that the lower the pupil-teacher-ratio is, the higher the learning outcome (in this case literacy) is. Here we want to emphasize that this is only a hypothesis of ours. In general, we can say, that we showed, that lower pupil-teacher-ratio is more likely to coexist with high literacy at a young age.

Tradução: A primeira hipótese, de que existe uma conexão negativa entre a razão aluno-professor no ensino primário e o letramento na parte mais jovem da população (15-24 a anos), pode ser assumida como verdadeira, devido ao significativo coeficiente de correlação de -0.962 . Por que isso acontece? Devido à natureza da correlação, não podemos dizer nada sobre a causalidade dessa conexão, mas ainda poderíamos hipotetizar porque essa conexão surgiu. Como afirmamos, pensamos que poderia mesmo haver uma relação causal no sentido de que, quanto menor for a razão aluno-professor, maior será o resultado da aprendizagem (neste caso, o letramento). Aqui queremos enfatizar que esta é apenas uma hipótese nossa. No geral, podemos dizer que mostramos, quanto menor a razão aluno-professor, é mais provável coexistir com o alto letramento em uma idade jovem.

Ao analisar o discurso conclusivo apresentado pela **Dupla 2**, no qual se observa que a hipótese de haver uma correlação negativa entre a razão aluno-professor no ensino primário e o letramento na parte mais jovem da população (15-24 anos) é verdadeira, destaca-se a relevância do valor encontrado para o coeficiente de correlação r de Pearson. Todavia, as razões para tal resultado não são evidentes, mas a hipótese de que “quanto menor for a razão aluno-professor, maior será o resultado da aprendizagem” é um direcionamento da dupla de futuros professores que explicaria o resultado observado e que fundamentaria a conclusão de que “quanto menor a razão aluno-professor, é mais provável coexistir com o alto letramento em uma idade jovem”. Ou seja, nessas condições, a razão aluno-professor teria efeito sobre o nível de letramento dos jovens.

Sobre a segunda hipótese, as conclusões apresentadas pela **Dupla 2** foram as seguintes:

The second hypothesis, that there is a negative connection between the pupil-teacher-ratio in primary education and the enrolment into secondary education could also be assumed to be true, due to the calculated data. So, in general, one could say that it is highly likely, that low pupil-teacher-ratios coexist with higher enrolment into secondary education. We hypothesize, that this is due to the individual attention a pupil gets, when the pupil-teacher-ratio is low, resulting in the pupils being more prone to enrolling into secondary educational institutions.

Tradução: A segunda hipótese, de que existe uma relação negativa entre a razão aluno-professor no ensino primário e a matrícula no ensino secundário, também pode ser assumida como verdadeira, devido aos dados calculados. Assim, em geral, pode-se dizer que é altamente provável que baixa razão aluno-professor coexista com maior matrícula no ensino secundário. Hipotetizamos que isso se deve à atenção individual que um aluno recebe, quando a razão aluno-professor é baixa, resultando em que os alunos sejam mais propensos a se matricular em instituições de ensino secundário.

Ao analisar a segunda hipótese, as conclusões da dupla de futuros professores indicam que o ciclo investigativo realizado permitiu identificar uma tendência de que, quanto menor for a razão aluno-professor no ensino primário, maior será o percentual de matriculados no ensino secundário. No sentido de concluir e justificar os resultados obtidos, ao salientar “*isso se deve à atenção individual que um aluno recebe, quando a razão aluno-professor é baixa, resultando em que os alunos sejam mais propensos a se matricular em instituições de ensino secundário*”, a dupla complementa e oferece um discurso plausível sobre um possível efeito da razão aluno-professor em relação ao percentual de matriculados no ensino secundário.

Ao se referir à possível correlação existente entre a razão aluno-professor no ensino primário e a razão aluno-professor no ensino secundário, a terceira hipótese estabelecida, a dupla de futuros professores tece a seguinte conclusão:

The third hypothesis, that there is a positive connection between the pupil-teacher-ratio in primary education and the pupil-teacher-ratio in secondary education (due to educational-politics, as we argued) can surprisingly not be accepted, due to the calculations with the data that we collected. In fact, the result was a negative connection (alas not too strong) between the two variables. For this result, we are not really able to come up with an explanation.

Tradução: A terceira hipótese, de que existe uma conexão positiva entre a razão aluno-professor no ensino primário e a razão aluno-professor no ensino secundário (devido à política educacional, como argumentamos) pode surpreendentemente não ser aceita, devido aos cálculos com os dados que coletamos. De fato, o resultado foi uma conexão negativa (infelizmente não muito forte) entre as duas variáveis. Para este resultado, não somos realmente capazes de chegar a uma explicação.

Quanto à terceira hipótese, o discurso conclusivo da **Dupla 2** tende a rejeitar o fato de que quanto menor a razão aluno-professor no ensino primário, menor seria a razão aluno-professor no ensino secundário. São os resultados obtidos ao integrar o SPSS, seja a representação gráfica ou os valores disponíveis no estudo de correlação, que permitem à dupla de futuros professores rejeitar a hipótese estabelecida. Cabe destacar que, embora rejeite a hipótese, os futuros professores não dispunham de elementos para explicar a correlação entre as variáveis.

Com base nos resultados obtidos e analisados, as considerações conclusivas para as hipóteses estabelecidas revelam que as representações gráficas visualizadas e os valores resultantes do estudo de correlações, emergentes pela utilização do SPSS, foram relevantes para as análises e contribuíram para as conclusões, a quinta fase do ciclo investigativo – PPDAC.

As conclusões apresentadas pela **Dupla 2**, sob o ponto de vista do terceiro estágio de transnumeração proposto por Pfannkuch, Rubick e Yoon (2002), fornecem uma compreensão aprofundada sobre os dados estudados. Com base nos resultados obtidos, as considerações tecidas resumem o que os dados mostram sobre o problema abordado, ou seja, os possíveis efeitos da razão aluno-professor em outras variáveis educacionais.

Ao longo das análises realizadas pela **Dupla 2**, e segundo Prodromou e Dunne (2017), foram identificadas as contribuições das representações gráficas e dos valores emergentes do estudo de correlações, mediados pelo SPSS, para a visualização de dados.

Além disso, no sentido dos elementos do letramento estatístico na era *OPEN Data*, proposto por esses mesmos autores, as conclusões apresentadas pela **Dupla 2** oferecem *inferências plausíveis em relação aos dados* estudados, por exemplo, nas conclusões relativas à correlação forte observada entre as variáveis envolvidas na segunda hipótese: *“Hipotetizamos que isso se deve à atenção individual que um aluno recebe, quando a razão aluno-professor é baixa, resultando em que os alunos sejam mais propensos a se matricular em instituições de ensino secundário”*.

Discursos dessa natureza revelam a presença da *linguagem* como resultado da mobilização de conhecimentos estatísticos para emitir uma compreensão sobre os dados. O conhecimento do *contexto* também contribuiu para as conjecturas *causais* dos resultados obtidos e apresentados pela **Dupla 2**, visto que, graficamente ou por valores críticos,

indicavam a existência, o comportamento, a intensidade e o nível de significância das correlações entre os dados, mas não do contexto deles.

No que se refere à *avaliação dos dados* amostrados, a **Dupla 2** não os questiona, mas, em trechos como “*o resultado foi uma conexão negativa (infelizmente não muito forte) entre as duas variáveis. Para este resultado, não somos realmente capazes de chegar a uma explicação*”, questiona a correlação observada entre as variáveis razão aluno-professor no ensino primário e razão aluno-professor no ensino secundário. É possível supor que um aprofundamento na direção dessas variáveis permitiria melhor compreensão sobre a correlação observada e isso geraria outros ciclos investigativos.

Diante do exposto, conclui-se que em relação à questão de pesquisa que guia esta tese os relatórios da *Homework 3*, elaborados pelas **Dupla 1** e **Dupla 2**, indicam que a utilização do SPSS favoreceu os futuros professores a realizarem, respectivamente, suposições sobre as distribuições e as dispersões dos dados no contexto de análise gráfica.

Além da representação gráfica, como apresentado pela **Dupla 2** (Figuras 33, 34 e 35), os estudos de correlações executados pelas duplas, mediados pela utilização do SPSS, resultaram em valores numéricos em representações tabulares (Figuras 36, 37, e 38, por exemplo), que permitiram observar, por meio de uma única manipulação, resultados estatísticos para o estudo das correlações entre variáveis envolvidas.

As possibilidades de representar como dados se distribuem (no caso da **Dupla 1**), ou, ao serem correlacionados, como eles se dispersam (no caso da **Dupla 2**) exemplificam formas de como as ferramentas disponíveis no SPSS podem contribuir para a análise gráfica de dados.

Além dos recursos de representação gráfica, emergentes do estudo de correlações nos dois relatórios analisados, existindo o conhecimento sobre as propriedades dos valores numéricos resultantes das correlações, a representação tabular revela outra possibilidade de ampliar o estudo, uma potencialidade da utilização do recurso tecnológico digital SPSS na atividade.

Como analisado ao longo das discussões apresentadas pelas duas duplas de futuros professores, à luz dos referenciais teóricos, os resultados obtidos nos relatórios permitiram identificar as representações gráficas para as distribuições e dispersão dos dados estudados. Além disso, as tabelas resultantes do estudo das correlações, possibilitadas pela utilização do SPSS, foram relevantes para aprimoramentos nas análises gráficas realizadas.

Os resultados presentes no relatório da dupla de futuros professores, sob o ponto de vista dos referenciais teóricos, transnumeração e visualização de dados, permitem identificar que a exploração do recurso SPSS, integrado na atividade, ofereceu possibilidades de aprimoramento na análise gráfica dos dados manipulados.

7 CONSIDERAÇÕES

Esta pesquisa concentra-se na discussão sobre a implementação de recursos digitais na educação estatística e está relacionada aos resultados obtidos na pesquisa de mestrado do autor (ARAÚJO, 2018), a qual mostrou as potencialidades do GeoGebra para o estudo de objetos estatísticos, especificamente as medidas de tendência central, por meio de uma sequência de atividades com alunos ingressantes do ensino médio.

A dissertação de mestrado realizada inspirou o projeto de pesquisa para o doutorado, aprovado no primeiro semestre de 2019, para o curso do programa de Estudos Pós-Graduados em Educação Matemática da Pontifícia Universidade Católica de São Paulo, o qual tinha como objetivo realizar uma investigação acerca das possíveis contribuições do GeoGebra, de suas representações gráficas em atividades relativas à análise gráfica de dados, desenvolvendo uma pesquisa com alunos do ensino médio para verificar, por meio de atividades de análise de dados, especialmente análise gráfica, quais seriam os possíveis contributos do GeoGebra nesse processo.

No início do ano letivo de 2020, foi decidido organizar os meios para implementar o projeto de pesquisa em uma escola pública do Estado de São Paulo. E assim, com autorizações da instituição de ensino e das solicitações éticas legais para o desenvolvimento da pesquisa, iniciou-se a mobilização dos potenciais participantes para o estudo.

De forma inesperada, em meados de março de 2020, a infecção respiratória causada pelo vírus SarsCov-19 provocou uma pandemia, e as escolas, assim como toda a sociedade, foram afetadas. Como medida de prevenção ao aumento de infecções, as escolas foram fechadas e todo o processo de ensino passou a ser realizado, em caráter emergencial, de forma remota. Dado o contexto das aulas remotas, a pesquisa sofreu alterações e a investigação das possíveis contribuições do GeoGebra em atividades de análise de dados com alunos do ensino médio não foi possível devido à perda de contato com os estudantes e ao desligamento institucional do pesquisador da Escola Pública do Estado de São Paulo.

No entanto, ao longo de 2020 e 2021, houve aproximação e participação nas discussões realizadas na *Linz School of Education* da *Johannes Kepler Universität*, na Áustria, por meio do professor Zsolt Lavicza, o qual formalizou um convite para a realização de um estágio doutoral junto ao seu grupo de pesquisa. Em relação ao contexto pandêmico na Áustria, naquela época, uma vez que estava controlado, foi possível realizar o estágio doutoral.

Durante o estágio de doutorado, realizado entre outubro de 2021 e janeiro de 2022, o pesquisador teve permissão para participar das aulas da disciplina de Métodos de Pesquisa e dos seminários conduzidos pelo professor Zsolt Lavicza, destinados aos alunos de mestrado e doutorado da *Linz School of Education*. Além disso, foi possível acompanhar e participar das aulas de Estatística ministradas por Lavicza para os alunos de graduação em matemática, geografia, educação física e outros cursos de formação inicial de professores.

Com o apoio do professor Zsolt Lavicza, responsável pela disciplina de Estatística, a pesquisa foi redirecionada para esse público, e, ao longo dela, foram propostas atividades em que os futuros professores utilizaram recursos tecnológicos para analisar dados de problemas reais diversos, processo relevante, pois, de acordo com Lopes (2008), problemas próximos à realidade dos participantes complementam de forma significativa a aprendizagem de conceitos estatísticos.

Ao considerar o planejamento do professor para a disciplina de Estatística, verificou-se que a proposta da pesquisa poderia ser ampliada, pois havia previsão de utilização dos recursos *Gapminder*, GeoGebra e SPSS. Diante dessa oportunidade, a pesquisa manteve como objeto de estudo a análise gráfica de dados, mas foi ampliada para explorar as possibilidades de investigação proporcionadas pela implementação de outros dois recursos digitais, além do GeoGebra.

Com base nas oportunidades de investigação oferecidas pelo *Gapminder*, GeoGebra e SPSS como recursos em atividades, o objetivo focou-se em identificar como essas ferramentas podem contribuir para a análise gráfica de dados, como parte das discussões sobre a viabilidade da incorporação de recursos tecnológicos na promoção da educação estatística. Nesse sentido, ao considerar a problemática desta tese, foi delimitado como problema:

Como a utilização de recursos digitais pode contribuir para análise gráfica de dados na disciplina estatística em uma formação inicial de professores?

Assim, nesta tese, o estudo se restringe aos recursos *Gapminder*, GeoGebra e SPSS, devido à disponibilidade de implementação e sua relevância na construção de representações e análises gráficas de dados, objeto de estudo da educação estatística, estabelecendo-se os seguintes objetivos para o estudo, oriundos nos relatos dos futuros professores participantes:

- 1) Identificar e analisar contribuições do *Gapminder*, GeoGebra e SPSS que se configuram como favoráveis para análise gráfica de dados;

- 2) Identificar recursos presentes nessas ferramentas que caracterizam tais contribuições;
- 3) Identificar nesses recursos possibilidades que permitam aprimoramentos na análise gráfica de dados.

Como Fiorentini e Lorenzato (2012) salientam, ao delimitar objetivos do estudo, esses nortearão as escolhas metodológicas. Nesse sentido, embora a aplicação do estudo tenha passado por redirecionamentos, seja pelo contexto da pandemia de COVID-19, seja pela oportunidade de implementação em contexto diferente do Brasil, considera-se que, ao investigar as possíveis contribuições desses recursos tecnológicos implementados em atividades, pode-se evidenciar contribuições que atendam à problemática da educação estatística, inclusive em outros contextos, como o brasileiro.

No Brasil, a Educação Matemática tem sido a área maior em que são observados estudos relativos à educação estatística. Cai et al. (2019a) sugerem que pesquisas na área da Educação Matemática devem ter uma conexão próxima com a sala de aula e impactar diretamente a prática. Ao investigar como o *Gapminder*, o GeoGebra e o SPSS podem contribuir para a análise gráfica de dados, os resultados obtidos podem fornecer subsídios para práticas a serem implementadas em sala de aula.

Os resultados de pesquisas preliminares realizadas indicam a escassez de estudos para o uso de tecnologias no ensino de Estatística, e, como apontou Pontes (2019), ao inserir recursos dessa natureza na prática educativa, é necessário realizar reflexões e possíveis adaptações da prática, levando em consideração os objetivos desejados, porque algumas dessas tecnologias não foram originalmente desenvolvidas com finalidade educativa.

Os trabalhos levantados a partir do Catálogo de Teses e Dissertações da Capes e nos anais de 2018 e 2022 da ICOTS revelaram, em diferentes níveis, fragilidades em relação aos conhecimentos necessários para a construção e para a análise gráfica de dados, como as pesquisas de Oliveira (2016) e Albuquerque (2018), nas quais problemas relacionados à análise gráfica de dados foram vistos como percalços da aprendizagem estatística. O estudo de Woodard (2018) salienta que há poucas pesquisas avaliando como os recursos computacionais para a Estatística afetam as habilidades dos alunos para pensar e raciocinar estatisticamente. Para esta tese, os pontos destacados por Bolch e Jacobbe (2018) podem ser vistos como relevantes, ao proporem a inserção de tecnologias para atividades no campo da educação estatística, em atividades relativas à análise gráfica de dados.

As bases teóricas e metodológicas utilizadas na aplicação e análise dos resultados das atividades desenvolvidas com a implementação dos recursos *Gapminder*, GeoGebra e SPSS forneceram suporte para a análise gráfica de dados. Os resultados obtidos indicam que esses recursos tecnológicos têm o potencial de promover a educação estatística de forma promissora.

Fazem parte dos referenciais teóricos a proposição de Prodrômou e Dunne (2017) sobre a visualização de dados, visto que esses autores consideram que ela permite explorar e comunicar efetivamente informações relevantes sobre grandes volumes de dados por meio de representações gráficas.

Além disso, o conceito de transnumeração, discutido por Wild e Pfannkuch (1999), também foi adotado, e é definido por eles como “um processo dinâmico de mudança de representações para gerar compreensão” (WILD; PFANNKUCH, 1999, p. 227). A transnumeração refere-se à ideia de formar e alterar representações de dados de um sistema para chegar a um melhor entendimento deles. O modelo dimensional proposto por Wild e Pfannkuch (1999) descreve quatro dimensões para o pensamento estatístico:

- Na dimensão um, *o ciclo investigativo*, referencial metodológico desta tese, aborda a resolução de um problema real e propõe o modelo PPDAC (problema, plano, dados, análises e conclusões), objetivando, geralmente, mudar o sistema para melhorar algo. Wild e Pfannkuch (1999) ressaltam que esse ciclo é iniciado para promover a aprendizagem, de modo que certos objetivos devem ser alcançados para que seja possível chegar a um nível desejado de compreensão do problema.

- Na dimensão dois, são apresentados os *tipos de pensamento* estatístico, incluindo os tipos geral e fundamental. Nesta tese, o referencial se concentra nos tipos fundamentais de pensamento estatístico presentes no *framework* dos autores, quais sejam: *reconhecimento da necessidade de dado; transnumeração; consideração da variação; raciocínio com modelos estatísticos e integração estatística e contextual*.

Após apresentar os tipos fundamentais de pensamento estatístico, na dimensão três do *framework* de Wild e Pfannkuch (1999), têm-se o ciclo interrogativo, “um processo de pensamento genérico em uso constante na resolução de problemas estatísticos” (p. 231). Para os autores, o aprendiz está sempre em um dos estados interrogativos durante a resolução de problemas, seja para gerar ideias, buscar informações, interpretar resultados de buscas, criticar a consistência de ideias e informações recebidas e julgar com base nas críticas realizadas.

Na dimensão quatro, as *disposições* se referem ao compromisso com um problema que se submete a resolvê-lo. Conforme Wild e Pfannkuch (1999), nessa dimensão discutem-se as qualidades pessoais que afetam ou mesmo iniciam a entrada em um modo de pensamento no contexto da resolução de problemas estatísticos.

As dimensões apresentadas e propostas por Wild e Pfannkuch (1999) permitem situar que a transnumeração é parte de um marco teórico mais amplo, o qual, como um tipo fundamental de pensamento estatístico, é adotado como referencial para esta tese.

De forma a delinear como a transnumeração se insere como aporte teórico nesta tese, são consideradas as propostas de Pfannkuch, Rubick e Yoon (2002) e Pfannkuch, Rubick (2002), que admitem a transnumeração como um processo em três estágios: no primeiro estágio, o foco está na captura de qualidades ou características da situação real; no segundo estágio, para que sejam apresentadas representações de dados, o foco está na necessidade de os dados serem coletados, transformados de brutos em múltiplas representações gráficas, resumos estatísticos, de modo a permitir a obtenção de significado da situação real a partir deles; no terceiro estágio, dada a obtenção de significados, o foco reside na comunicação a partir dos dados e julgamentos de forma a serem compreendidos, em termos das situações reais, por outros. Nesta tese, os estágios de transnumeração possibilitaram analisar o percurso dos participantes em cada atividade proposta, de modo a elucidar as contribuições dos recursos tecnológicos implementados.

Como esta tese investiga as possíveis contribuições de recursos digitais integrados para a análise gráfica de dados, a visualização de dados é outro referencial para as análises das atividades propostas, pressupondo esse processo assistido pelos recursos *Gapminder*, GeoGebra e SPSS. A proposição de Prodromou e Dunne (2017) para a visualização de dados enfatiza as possibilidades de aprendizagem que podem ocorrer a partir do desenvolvimento tecnológico que, no campo da educação estatística, engloba a análise exploratória de dados assistida por computador, o que possibilita explorar e comunicar efetivamente informações relevantes por meio de representações gráficas.

Embora Prodromou e Dunne ressaltem que a visualização de dados não reside nos recursos que permitem a manipulação diversificada dos dados, eles consideram que a inovação está nas formas de utilização dos dados e afirmam que “a grande importância da disponibilidade de *software* e da popularidade associada, na determinação de quais análises são realizadas e como elas são apresentadas, é um tópico de grande relevância” (PRODROMOU; DUNNE, 2017, p. 2, tradução do autor). Os pesquisadores apresentam

elementos de discussão sobre questões relacionadas à visualização, seguidas pela organização teórica proposta por eles, como a compreensão, finalidades, verbalização, linguagem, contexto, mudança e causa, e avaliação dos dados. Eles abordam a importância da visualização no ensino da Estatística em um contexto de *OPEN Data*.

O *framework* de Prodrômou e Dunne (2017) leva em conta toda a complexidade dos dados e argumenta pela necessidade de uma mudança na forma como se abordam o controle e a manipulação de dados.

Wild e Pfannkuch (1999) apontam que um ciclo investigativo pode ser definido para atingir objetivos de aprendizagem. O conhecimento adquirido e as necessidades identificadas nesses ciclos podem iniciar novos ciclos investigativos, de modo que as conclusões das investigações alimentam uma base de conhecimento sobre o contexto e ao ser expandida pode informar quais ações tomar, indicando uma tomada de decisão.

Como os autores o modelam, o PPDAC é um ciclo composto por fases: problema, plano, dados, análises e conclusões cujo “objetivo final da investigação estatística é aprender no domínio do contexto de um problema real” (WILD; PFANNKUCH, 1999, p. 244, tradução o autor)

A parte experimental deste estudo, uma proposta de atividades que envolve o estudo de situações reais, sob ciclos investigativos, busca identificar contribuições dos recursos *Gapminder*, GeoGebra e SPSS em análise gráfica de dados, utilizados na disciplina de Estatística oferecida pelo professor Zsolt Lavicza.

O *Gapminder* (<https://www.gapminder.org>) é uma fundação sueca independente sem afiliações políticas, religiosas ou econômicas e tem como missão promover uma visão de mundo baseada em fatos, de modo que todos possam acessá-los. No website do *Gapminder* os materiais disponíveis apresentam dados sobre uma diversidade de temas, como educação, renda, pobreza, entre outros, e oferece representações de dados sobre esses temas. Suas ferramentas e materiais didáticos são de acesso gratuito.

A descrição do *software* GeoGebra, apresentada no *site* oficial indica que se trata de *software* dinâmico de matemática para o ensino em todos os níveis. Dentre suas ferramentas, ele permite abordar tópicos de geometria, álgebra, estatística, probabilidade em uma única plataforma. O trabalho com planilhas e construção de representações gráficas de dados são funcionalidades disponíveis nesse recurso tecnológico digital. Além de ser de acesso gratuito e em vários idiomas, o GeoGebra dispõe de versões que podem ser utilizadas de forma *online* ou *off-line*, acessível também em dispositivos móveis.

O SPSS (<https://www.ibm.com/products/spss-statistics>) é um *software* estatístico desenvolvido pela IBM®, que oferece ferramentas para análise estatística e acessível aos usuários de todos os níveis de habilidade. O uso do SPSS é possível a partir de uma licença, ou seja, não é um recurso tecnológico digital de acesso livre, todavia, a IBM® o disponibiliza para acesso experimental por um mês. Como destaca no seu *site* oficial, o SPSS é um recurso de fácil utilização, possui uma ampla biblioteca de algoritmos de *machine learning*, análise de texto, extensibilidade de *software* livre, integração com *big data*, que permitem abordar problemas em diferentes níveis.

O desenvolvimento do experimento desta tese foi conduzido em inglês, e neste trabalho, foi realizada a tradução para o português, e consistiu na proposição de três atividades, cada uma utilizando um dos recursos tecnológicos: o *Gapminder*, o GeoGebra e o SPSS, de modo que todos os participantes tiveram a oportunidade de realizar essas atividades.

O total de alunos matriculados na disciplina de Estatística foi de 72. Por se tratar de alunos de cursos de formação inicial de professores, eles foram denominados como futuros professores nesta tese, e, organizados em duplas ou trios, tiveram a oportunidade de explorar e se familiarizar com os recursos digitais, utilizando dados disponíveis para definir problemas a serem estudados. A autorização para a utilização dos dados dos relatórios dos alunos foi obtida pelo professor Zsolt Lavicza. A disciplina tinha uma carga horária de três horas semanais, com aulas teóricas e práticas, o que possibilitou a proposição das atividades do experimento desta tese, seguindo a dimensão do ciclo investigativo - PPDAC, de acordo com Wild e Pfannkuch (1999).

As atividades propostas, denominadas *Homework*, foram realizadas com cada grupo utilizando um dos recursos digitais específicos. Os futuros professores delimitaram problemas relevantes para eles, realizaram buscas, limpeza e organização dos dados relacionados aos problemas escolhidos, processo que seguiu o ciclo investigativo - PPDAC, com sugestões de acesso e utilização dos recursos para garantir a relevância dos problemas abordados pelos participantes.

Após acessar os relatórios dos participantes, foram criadas pastas para o material de cada grupo e realizada a leitura do desenvolvimento das atividades de cada um deles. O objetivo dessa leitura foi identificar e selecionar quais produções seriam utilizadas como material de análise para esta pesquisa. Com base na escolha metodológica, foram considerados os grupos que mantiveram a mesma composição e cujas produções

demonstraram o uso adequado das respectivas ferramentas dos recursos, além de realizarem as análises solicitadas nas três atividades propostas.

Dessa forma, o recorte metodológico resultou na seleção dos relatórios produzidos por cinco duplas de futuros professores. Para cada atividade, foi realizada a leitura do desenvolvimento das atividades de cada grupo, visando identificar e selecionar as produções que seriam utilizadas como material de análise para esta pesquisa. No entanto, o *corpus* do material de análise desta pesquisa consiste nos resultados dos relatórios de dois dos grupos, visto que eles se adequaram melhor aos objetivos desta tese, levando em consideração os referenciais teóricos da transnumeração e da visualização de dados.

Na primeira atividade, denominada *Homework 1*, os participantes foram solicitados a acessar o *website* do *Gapminder* e realizar uma atividade de exploração de dados, a qual poderia ser mediada pelas diversas representações dos dados relacionados aos problemas escolhidos pelas duplas ou trios. Foram sugeridos vídeos e ferramentas para familiarizar os participantes com o *Gapminder*.

O objetivo dessa atividade era permitir que os participantes identificassem variáveis e estudos realizados com base nos dados relacionados a essas variáveis, de modo que eles foram incentivados a explorar, selecionar tópicos (temas envolvendo as variáveis estudadas) e dissertar sobre as representações gráficas, utilizando as ferramentas interativas do *Gapminder*, a fim de promover uma análise exploratória dos dados, em que a análise gráfica seria um processo importante para a compreensão das variáveis em estudo.

Os resultados da atividade, conforme analisados nos relatórios das **Dupla 1**, e **Dupla 2** demonstraram o potencial interativo do *website* do *Gapminder*. Nesse sentido, a possibilidade de escolher diferentes variáveis e utilizar diversas formas de representação foi relevante para que as duplas de futuros professores realizassem análises e reflexões sobre os dados.

A *Homework 2* proposta, a partir da implementação do GeoGebra para o estudo de dados, tinha como objetivo que os participantes pudessem importar, organizar, e construir representações de dados mediados pelo GeoGebra e suas respectivas análises, novamente, seguindo as fases metodológicas do ciclo investigativo – PPDAC, conforme Wild e Pfannkuch (1999).

Os resultados da utilização do recurso tecnológico digital GeoGebra, como analisado nos relatórios das **Dupla 1** e **Dupla 2**, apontam para possibilidades gráficas de análise univariada e bivariada, como as representações do gráfico de colunas, do *box-plot*, do gráfico

da reta de regressão e da curva *Spline* que, ao serem utilizadas, contribuíram para a compreensão dos participantes sobre os dados estudados.

Na terceira atividade, *Homework 3*, foi proposta a utilização do SPSS. Levando em consideração o estágio dos futuros professores na disciplina, uma das possibilidades de uso do *software* era o estudo de correlações cujo objetivo era realizar análises de correlação entre pelo menos três pares de variáveis e interpretar os resultados, seguindo as fases do ciclo investigativo - PPDAC, de acordo com Wild e Pfannkuch (1999).

A partir dos relatórios das duplas de futuros professores, o recurso da representação gráfica, resultante do estudo sobre a distribuição dos dados, contribuiu para inferências sobre a normalidade desses dados, realizadas pela **Dupla 1**. Da mesma forma as representações gráficas favoreceram à **Dupla 2** inferir indícios de possíveis correlações entre as variáveis estudadas. Além disso, os valores resultantes do estudo das correlações, suficientes para as duplas de futuros professores tecerem análises sobre as variáveis, apontam para a viabilidade de utilização do SPSS em atividade de análise de dados.

Considerando a questão de pesquisa desta tese: **Como a utilização de recursos digitais pode contribuir para análise gráfica de dados na disciplina estatística em uma formação inicial de professores?** concluímos que:

As oportunidades de investigação condicionadas pelas ferramentas do *Gapminder*, do GeoGebra e do SPSS contribuíram para análise gráfica de dados e para a viabilidade da utilização de recursos tecnológicos na promoção da educação estatística.

As bases teóricas e metodológicas utilizadas na aplicação e análise dos resultados dos *Homework* desenvolvidos, forneceram suporte para a análise gráfica de dados e indicaram que esses recursos tecnológicos têm o potencial de promover a educação estatística de forma promissora.

Os relatórios das demais duplas participantes, analisados à luz dos referenciais teóricos da transnumeração e da visualização de dados, serão utilizados em futuros artigos, nos quais serão apresentados os resultados obtidos. Esses relatórios também evidenciam os benefícios de cada um dos recursos digitais, *Gapminder*, GeoGebra e SPSS, e indicam que sua utilização pode complementar de forma significativa a aprendizagem de conceitos estatísticos.

REFERÊNCIAS

- ABAR, C. A. A. P.; LAVICZA, Z. Underlying theories for use of digital technologies in mathematics education. **Acta Scientiae**, Canoas, v. 21, n. 1, p. 39-54, jan./fev. 2019. Disponível em: <http://www.periodicos.ulbra.br/index.php/acta/article/view/4913/pdf>. Acesso em: dez. 2021.
- ANDRE, M. **Implementing the statistical investigative process in secondary school education**. Doctoral Thesis. 131 f. 2020. Doctorate degree in Science (Didactics of Mathematics). Johannes Kepler University, Linz, Austria. 2020.
- ANDRE, M.; LAVICZA, Z. Technology changing statistics education: defining possibilities, opportunities and obligations. **Electronic Journal of Mathematics & Technology**, v. 13, n. 3, 2019. Disponível em: <https://go.gale.com/ps/i.do?id=GALE%7CA673853528&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=19332823&p=AONE&sw=w&userGroupName=anon%7Ef2944590&aty=open+web+entry>. Acesso em: jun. 2023.
- ANDRE, M.; LAVICZA, Z.; PRODROMOU, T. Formalizing students' informal statistical reasoning on real data: using Gapminder to follow the cycle of inquiry and visual analyses. **Eleventh Congress of the European Society for Research in Mathematics Education (CERME11)**. Freudenthal Group; Freudenthal Institute; ERME, 2019. Disponível em: <https://hal.science/hal-02410839/document>. Acesso em: jun. 2023.
- ANDRE, M.; SCHNEIDER, C.; FRISCHEMEIER, D. Covariational reasoning in primary school: a qualitative approach. **Twelfth Congress of the European Society for Research in Mathematics Education**, p. 2022. Disponível em: <https://hal.science/hal-03754724/document>. Acesso em: jun. 2023.
- ALBUQUERQUE, M. R. G. C. **Escala apresentada em gráficos: conhecimentos matemáticos para o ensino dos anos iniciais do ensino fundamental (crianças e Eja)**. 294 f. 2018. Tese (Doutorado em Educação Matemática e Tecnológica) - Universidade Federal de Pernambuco, Recife, 2018. Disponível em: <https://repositorio.ufpe.br/handle/123456789/33605>. Acesso em: mar. 2021.
- ARAÚJO, J. R. A. **Atividades para o estudo das medidas de tendência central: uma proposta com o apoio do GeoGebra**. 2018. 145 f. Dissertação (Mestrado em Educação Matemática) – Programa de Estudos Pós-Graduados em Educação Matemática, Pontifícia Universidade Católica de São Paulo, São Paulo, 2018. Disponível em: <https://repositorio.pucsp.br/jspui/handle/handle/21308>. Acesso em: mar. 2021.
- ARAÚJO, J. R. A. O GeoGebra como recurso para o estudo de objetos estatísticos: reflexões a partir de um percurso investigativo em desenvolvimento. **Boletim Cearense de Educação e História da Matemática, [S. l.]**, v. 7, n. 21, p. 59–70, 2021. Disponível em: <https://doi.org/10.30938/bocehm.v7i21.3393>. Acesso em: jul. 2021.
- ARAÚJO, J. R. A.; ABAR, C. A. A. P. Contribuições do GeoGebra nas dialéticas de uma situação didática para o estudo das medidas de tendência central. **Educação Matemática Debate**, v. 3, p. 282-302, 2019. Disponível em:

<https://www.periodicos.unimontes.br/index.php/emd/article/view/92/97>. Acesso em: abr. 2020.

ARAÚJO, J. R. A.; ABAR, C. A. A. P. Estudo dos processos de transnumeração mediados pela utilização de uma tecnologia. **Revista de Produção Discente em Educação Matemática**. v. 9, n. 2, 2020. Disponível em: <https://revistas.pucsp.br/index.php/pdemat/article/view/51081>. Acesso em: mar. 2022.

ARAÚJO, J. R. A.; ABAR, C. A. A. P. LAVICZA, Z. Transnumeration and data visualization: a look at graphical analyses of data mediated by Gapminder. In: **Bridging the Gap: Empowering and Educating Today's Learners in Statistics**. Proceedings of the Eleventh International Conference on Teaching Statistics. International Association for Statistical Education, 2022, p. 1-6. Disponível em: https://iase-web.org/icots/11/proceedings/pdfs/ICOTS11_315_ARAJO.pdf?1669865554. Acesso em: mar. 2022.

BEN-ZVI, D.; FRIEDLANDER, A. Statistical thinking in a technological environment. In: GARFIELD, J. B.; BURRILL, G. (Eds.). **Research on the role of technology in teaching and learning of statistics**. Voorburg, The Netherlands: International Statistical Institute, p. 45-55, 1997. Disponível em: <https://iase-web.org/documents/papers/rt1996/4.Ben-Zvi.pdf>. Acesso em: maio. 2019.

BIEHLER, R. Software for learning and for doing statistics. **International Statistical Review**, n. 65, v. 2, p. 167–189, 1997. Disponível em: <https://www.stat.auckland.ac.nz/~iase/publications/isr/97.Biehler.pdf>. Acesso em: jan. 2022.

BIEHLER, R. Design principles, realizations and uses of software supporting the learning and the doing of statistics-a reflection on developments since the late 1990s. In: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)**. Kyoto, Japão, 2018. p. 1-3. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_1B1.pdf. Acesso em: jun. 2021.

BOLCH, C.; JACOBBE, T. Students' Understanding of Data Visualizations. In: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)**. Kyoto, Japão, 2018. p. 1-6. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3C3.pdf?1531364257. Acesso em: jun. 2021.

BORTOLOSSI, H. J. O uso do software gratuito GeoGebra no ensino e na aprendizagem de estatística e probabilidade. **VIDYA**, Santa Maria, v. 36, n. 2, p. 429-440, jul./dez. 2016. Disponível em: <https://periodicos.ufn.edu.br/index.php/VIDYA/article/view/1804>. Acesso em: fev. 2021.

BRASIL. Secretaria da Educação Básica. **Base Nacional Comum Curricular**. Brasília-DF, 2018.

CABRERA, G. B. et al. La transnumeración como insumo del pensamiento crítico. **Bridging the Gap: Empowering and Educating Today's Learners in Statistics**. Proceedings of the Eleventh International Conference on Teaching Statistics. International Association for

Statistical Education, Rosario, Argentina, 2022. p. 1-6. Disponível em: https://iase-web.org/icots/11/proceedings/pdfs/ICOTS11_179_CABRERA.pdf?1669865532. Acesso em: mar. 2023.

CAI, J. et al. Posing significant research questions. **Journal for Research in Mathematics Education**, v. 50, n. 2, p. 114-120, 2019a. Disponível em: <https://doi.org/10.5951/jresematheduc.50.2.0114>. Acesso em: ago. 2021.

CAI, J. et al. Choosing and justifying robust methods for educational research. **Journal for Research in Mathematics Education**, v. 50, n. 4, p. 342-348, 2019b. Disponível em: <https://doi.org/10.5951/jresematheduc.50.4.0342>. Acesso em: ago. 2021.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Serviços**: Catálogo de teses e dissertações. Disponível em: <https://catalogodeteses.capes.gov.br/catalogo-teses/#/>. Acesso em: mar. 2022.

CARVER, R.; KRAFT, V. Keeping it real with data visualization. *In*: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward**. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10). Kyoto, Japão, 2018. p. 1-4. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3C2.pdf?1531364257. Acesso em: jul. 2021.

CHERNOFF, H. **Graphical Representations as a discipline**. Massachusetts Inst of Tech Cambridge Dept of Mathematics. 1978. Disponível em: <https://apps.dtic.mil/sti/citations/ADA056633>. Acesso em: mar. 2021.

CHICK, H. Tools for transnumeration: early stages in the art of data representation. **Mathematics education for the third millennium**: Towards 2010. Proceedings of the twenty-seventh annual conference of the Mathematics Education Research Group of Australasia. 2004. p. 167-174. Disponível em: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8dbef785de9b180bcf3eebc7dd5d306f225b7de8>. Acesso em: dez. 2020.

DAIGA, M. C. **Preservice Teachers' Knowledge and use of Transnumeration**. Thesis Doctoral. Doctoral Thesis. Doctorate degree in Philosophy of Education. Indiana University. Bloomington, USA. 2018. Disponível em: <https://iase-web.org/documents/dissertations/18.MichaelDaiga.Dissertation.pdf>. Acesso em: out. 2020.

ECKERT, A. Implementing research results in technology induced teaching of statistics – a literature review. *In*: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward**. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10). Kyoto, Japão, 2018. p. 1-4. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_9J1.pdf?1531364302. Acesso em: ago. 2021.

FEY, J. T. Technology and mathematics education: a survey of recent developments and important problems. **Educational Studies in Mathematics**. Information Technology and Mathematics Education, v. 20, n. 3, p. 237-272, 1989. Disponível em: <https://link.springer.com/article/10.1007/BF00310873>. Acesso em: maio. 2019.

FIorentini, D.; Lorenzato, S. **Investigação em educação matemática**: percursos teóricos e metodológicos. Campinas: Autores Associados, v. 3, 2012.

FONSECA, J. J. S. **Apostila de metodologia da pesquisa científica**. Rio de Janeiro: UFRJ, Apostila. 2002. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=oB5x2SChpSEC&oi=fnd&pg=PA6&dq=+Apostila+de+metodologia+da+pesquisa+cient%C3%ADfca.&ots=ORW-Xx8ki4&sig=Y7dXNkiuWHgh9-vxpXcN6ajluTg#v=onepage&q&f=false>. Acesso em: jun. 2019.

FRANCISCO, V. R. **Interpretação de dados estatísticos**: um estudo com alunos do ensino médio da Educação de Jovens e Adultos. 131 f. 2016. Dissertação (Mestrado em Educação Matemática e Tecnológica) - Universidade Federal de Pernambuco, Recife-PE, 2016. Disponível em: <https://repositorio.ufpe.br/handle/123456789/22287>. Acesso em: maio. 2020.

FUJII, Y. Current situation and issues for effective use of statistical software in Japan. *In*: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward**. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10). Kyoto, Japão, 2018. p. 1-4. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_9E1.pdf?1531364300. Acesso em: ago. 2021.

GAL, I. Adults' statistical literacy: meanings, components, responsibilities. **International Statistical Review**, Netherlands, v. 70, n. 1, p. 1-50, 2002.

GAL, I.; GARFIELD, J. **The assessment challenge in statistics education**. Amsterdam: IOS Press and International Statistical Institute, 1997.

GIORDANO, L. V. O. **Panorama da competência estatística no ensino médio brasileiro**: das ideias e práticas dos professores ao desempenho dos alunos no ENEM. 2017. 274f. (Tese de doutorado) - Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin, Campinas, SP, 2017. Disponível em: <https://www.repositorio.unicamp.br/acervo/detalhe/991250>. Acesso em: abr. 2020.

GODOY, A. S. Introdução à pesquisa qualitativa e suas possibilidades. **Revista de administração de empresas**, v. 35, p. 57-63, 1995. Disponível em: <https://doi.org/10.1590/S0034-75901995000200008>. Acesso em abr. 2022.

GOMES, J. **Programas de SPSS e R como ferramenta no ensino de probabilidades e estatística no 12º ano de escolaridade em Timor-Leste**. 2016. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/82456/2/38022.pdf>. Acesso em: maio. 2023.

GONZALES, F.; LOPES, C. E. Estudo de viabilidade do uso do software GeoGebra, no ensino estatístico. FESPM, Federação Espanhola de Sociedades de Professores de Matemática (Ed.), **VIII Congresso Ibero-Americano de Educação Matemática**, Madri, Espanha: FESPM, p. 14-22, 2017. Disponível em: <http://funes.uniandes.edu.co/19834/1/Gonzales2017Estudo.pdf>. Acesso em: maio. 2023.

HABIBULLAH, S. N. A workshop in Pakistan on teaching statistical inference through a freely available online software. *In*: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward**. Proceedings of the Tenth International Conference on Teaching

Statistics (ICOTS10), Kyoto, Japão, 2018. p. 1-4. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_2D2.pdf. Acesso em: jul 2021.

KEIM, D. A.; MANSMANN, F.; SCHNEIDEWIND, J.; ZIEGLER, H. Challenges in visual data analysis. **Tenth International Conference on Information Visualization**, v. IV, n. 6, p. 9-16. IEEE. 2006. Disponível em: <https://kops.uni-konstanz.de/server/api/core/bitstreams/3a0170ca-a333-4498-8c3f-eb145168ad9a/content>. Acesso em: out. 2021.

LEE, H.S. et al. Teachers' use of transnumeration in solving statistical tasks with dynamic statistical software. **Statistics Education Research Journal**, v. 13, n. 1, p. 25-52, 2014. Disponível em: <https://doi.org/10.52041/serj.v13i1.297>. Acesso em: abr. 2019.

LÉVY, P. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. Rio de Janeiro, Ed. 34, 1993.

LOPES, C. E. O ensino da estatística e da probabilidade na educação básica e a formação dos professores. **Cadernos Cedes**, v. 28, n. 74, p. 57-73, 2008. Disponível em: <https://doi.org/10.1590/S0101-32622008000100005>. Acesso em: maio. 2022.

LUCAS, G. B. **Estudo dos níveis de letramento estatístico e dos estádios de desenvolvimento cognitivo no Programa LeME**. 2019. Dissertação (Mestrado em Educação em Ciências Química da Vida e Saúde) (UFSM - FURG) - Universidade Federal do Rio Grande, Porto Alegre-RS, 2019. Disponível em: <http://www.repositorio.furg.br/handle/1/9096>. Acesso em: abr. 2020.

MACEDO, R. C. **Conhecimentos de professores de matemática sobre o processo de ensino e de aprendizagem de noções estatísticas—curva normal**. 2016. Dissertação (Mestrado em Educação Matemática) - Universidade Anhanguera de São Paulo, São Paulo, SP, 2016. Disponível em: <https://repositorio.pgsscogna.com.br/handle/123456789/21816>. Acesso em: maio. 2020.

MARTINS, G. A.; THEÓPHILO, C. R. **Metodologia da investigação científica para ciências sociais aplicadas**. 3. ed. São Paulo: Atlas, 2016.

MCNAMARA, A. Imagining the Future of Statistical Education Software. In: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)**. Kyoto, Japão, 2018. p. 1-3. Disponível em: http://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_1B2.pdf. Acesso em: maio. 2021.

MELO, K. M. F. **O pensamento estatístico no ensino fundamental: uma experiência articulando o desenvolvimento de projetos de pesquisa com os conceitos básicos da estatística implementados em uma sequência didática eletrônica**. 2017. 432f. Tese (Doutorado em Ensino de Ciências e Matemática) - Universidade Luterana do Brasil, Canoas-RS, 2017. Disponível em: <http://www.ppgecim.ulbra.br/teses/index.php/ppgecim/article/view/293>. Acesso em: abr. 2020.

MANTILLA, C. Using ICT's for to Introduce to Statistic. In: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward**. Proceedings of the Tenth International

Conference on Teaching Statistics (ICOTS10). Kyoto, Japão, 2018. p. 1-4. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_2D3.pdf?1531364242. Acesso em: jul. 2021.

NETO, P. L. O. C. **Estatística**. São Paulo: Edgard Blücher, 2002.

OGBONNAY, C. Teaching Statistics with MATLAB: Why and how? **Bridging the Gap: Empowering and Educating Today's Learners in Statistics**. Proceedings of the Eleventh International Conference on Teaching Statistics. International Association for Statistical Education, Rosario, Argentina, 2022. p. 1-6. Disponível em: https://iase-web.org/icots/11/proceedings/pdfs/ICOTS11_314_OGBONNAY.pdf?1669865554. Acesso em: mar. 2023.

OLIVEIRA, S. A. P. **Educação Estatística em escolas do povo Xukuru do Ororubá**. 2016. Dissertação (Mestrado em Educação Matemática e Tecnológica) - Universidade Federal de Pernambuco, Recife, PE, 2016. Disponível em: <https://repositorio.ufpe.br/handle/123456789/18717>. Acesso em: nov. 2020.

PATSIOMITOU, S. Objetos híbrido-dinâmicos: ambientes DGS e transformações conceituais. **Revista Internacional de Estudos Educacionais e Vocacionais**, v. 1, n. 1, p. 31-46, 2019. Disponível em: <https://doi.org/10.29103/ijevs.v1i1.1416>. Acesso em: maio. 2022.

PFANNKUCH, M.; RUBICK, A. An exploration of students' statistical thinking with given data. **Statistics Education Research Journal**, v. 1, n. 2, p. 4-21, 2002. Disponível em: <https://iase-web.org/ojs/SERJ/article/view/562/424>. Acesso em: dez. 2021.

PFANNKUCH, M.; RUBICK, A.; YOON, C. Statistical thinking and transnumeration. In: B. BARTON, B.; IRWIN, K. C.; PFANNKUCH, M.; THOMAS, M. O. J. (Eds.), **Mathematics Education in the South Pacific**. Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia, Auckland. Sydney: MERGA, 2002. Disponível em: https://merga.net.au/Public/Publications/Annual_Conference_Proceedings/2002_MERGA_C P.aspx. Acesso em: ago. 2021.

PONTES, M. E. N. **Aprendizagem de gráficos com e sem uso do Excel por alunos do 5º ano Ensino Fundamental**. 2020. Dissertação (Mestrado em Educação Matemática e Tecnológica) – Universidade Federal de Pernambuco, Recife, 2020. Disponível em: <https://repositorio.ufpe.br/handle/123456789/37952>. Acesso em: abr. 2020.

PRODROMOU, T.; DUNNE, T. Data visualisation and statistics education in the future. Data visualization and statistical literacy for open and big data. **IGI Global**, 2017. p. 1-28. Disponível em: <https://www.igi-global.com/chapter/data-visualisation-and-statistics-education-in-the-future/179958>. Acesso em: out. 2021.

RENAUX, C. D. Z. **O uso de objetos de aprendizagem de estatística em um curso de pedagogia: algumas possibilidades e potencialidades**. 2017. 110f. Dissertação (Mestrado em Educação em Ciências e em Matemática) - Universidade Federal do Paraná, Curitiba-PR, 2017. Disponível em: <https://acervodigital.ufpr.br/handle/1884/53049>. Acesso em: abr. 2020.

SANTOS, A. A. **A construção do letramento estatístico em estratégias com o uso de tecnologias digitais em aulas de estatística de cursos de graduação**. 2019. 134 f. Dissertação (Mestrado em Educação Matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo, SP, 2019. Disponível em: <https://repositorio.pucsp.br/jspui/handle/handle/22764>. Acesso em: abr. 2020.

SANTOS, P. G. et al. The use of technological resources in teacher education as a contribution to the teaching of statistics and probability in Brazil. **Bridging the Gap: Empowering and Educating Today's Learners in Statistics**. Proceedings of the Eleventh International Conference on Teaching Statistics. International Association for Statistical Education, Rosario, Argentina, 2022. p. 1-6. Disponível em: https://iase-web.org/icots/11/proceedings/pdfs/ICOTS11_266_GERMANOD.pdf?1669865546. Acesso em: mar. 2023.

SERA, E. K. **Conhecimento de professores para o ensino da leitura e construção de gráficos estatísticos na educação básica**. 2016. Dissertação (Mestrado em Educação Matemática) - Universidade Anhanguera de São Paulo, São Paulo-SP, 2016. Disponível em: https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=3815938. Acesso em: jun. 2020.

SILVA, C. B. **Pensamento estatístico e raciocínio sobre variação: um estudo com professores de matemática**. 354 f. 2007. Tese (Doutorado em Educação Matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo, SP, 2007. Disponível em: <https://tede2.pucsp.br/handle/handle/11206>. Acesso em: mar. 2019.

SOARES, S. V.; PICOLLI, I. R. A.; CASAGRANDE, J. L. Pesquisa bibliográfica, pesquisa bibliométrica, artigo de revisão e ensaio teórico em administração e contabilidade. **Administração: Ensino e Pesquisa**, v. 19, n. 2, p. 308-339, 2018. Disponível em: <https://raep.emnuvens.com.br/raep/article/view/970>. Acesso em: maio. 2020.

SOUZA, J. M. G. **Interpretação de gráficos: explorando o Letramento estatístico dos professores de escolas públicas no campo nos espaços de oficinas de formação continuada**. 144 f. 2019. Dissertação (Mestrado em Educação Matemática e Tecnológica) - Universidade Federal de Pernambuco, Recife-PE, 2019. Disponível em: <https://repositorio.ufpe.br/handle/123456789/34197>. Acesso em: abr. 2020.

SOUZA, R.; CALEJON, L. M. Uso da tecnologia da informação e comunicação em uma sequência didática incluindo *software* Geogebra no ensino da estatística descritiva. **REnCiMa. Revista de Ensino de Ciências e Matemática**, v. 10, n. 4, p. 227-244, 2019. Disponível em: <https://revistapos.cruzeirodosul.edu.br/index.php/rencima/>. Acesso em: maio. 2023.

STRELHOW, M. R. W.; CÂMARA, S. G. Descobrimo a estatística usando o SPSS. **Aletheia**, n. 35-36, 2011. Disponível em: <http://www.periodicos.ulbra.br/index.php/aletheia/article/viewFile/3435/2571>. Acesso em: maio. 2023.

TUKEY, J. W. **Exploratory Data Analysis**. Addison-Wesley Publishing Company. 1977. Disponível em: http://theta.edu.pl/wp-content/uploads/2012/10/exploratorydataanalysis_tukey.pdf. Acesso em: jun. 2023.

UNWIN, A. “Why is Data Visualization important? What is Important in Data Visualization?” **Harvard Data Science Review** v. 2, n. 1, p. 1–7, 2020. Disponível em: <https://hdsr.duquduq.org/pub/zok97i7p/release/4?readingCollection=03f9b00c>. Acesso em: out. 2021.

VIALI, L.; ODY, M. C. A produção brasileira em educação estatística avaliada pela análise das teses. **Educação Matemática Pesquisa**, v. 22, n. 1, p. 68-94, 2020. Disponível em: <https://doi.org/10.23925/1983-3156.2020v22i1p068-094>. Acesso em: out. 2021.

VOTTO, T. R. **As potencialidades lúdicas nas estratégias para o ensino e a aprendizagem estatística nos anos iniciais do ensino fundamental**. 175 f. 2018. Dissertação (Mestrado em Educação em Ciências Química da Vida e Saúde). - (UFSM - FURG), Universidade Federal do Rio Grande, Porto Alegre-RS, 2018. Disponível em: <http://repositorio.furg.br/handle/1/8540>. Acesso em: maio. 2020.

WATSON, J.; CALLINGHAM, R. Statistical literacy: a complex hierarchical construct. **Statistics Education Research Journal**. v. 2, p. 3-46, nov. 2003. Disponível em: <https://iase-web.org/ojs/SERJ/article/view/553>. Acesso em: abr. 2020.

WILD, C. J.; PFANNKUCH, M. Statistical Thinking in Empirical Enquiry. **Internacional Statistical Review**. v. 67, n. 3. p. 223-265, 1999. Disponível em: <https://iase-web.org/documents/intstatreview/99.Wild.Pfannkuch.pdf>. Acesso em: jan. 2021.

WOODARD, V. W. Defining the relationship between statistical thinking and statistical computing. *In*: SORTO, M. A.; WHITE, A.; GUYOT, L. (Eds.) **Looking back, looking forward**. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10). Kyoto, Japão, 2018. p. 1-6. Disponível em: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3F2.pdf?1531364258. Acesso em: jul. 2021.

ANEXO



Linz, 11 July 2023

Authorisation for Mr José Ronaldo Alves Araujo

To Whom It May Concern,

I am writing this letter to provide further information and clarification regarding Mr José Ronaldo Alves Araujo, a doctoral student enrolled in the Post-Graduate Studies Program in Mathematics Education at the Pontifical Catholic University of São Paulo. I, Professor Dr Celina Abar, have been serving as his supervisor during his doctoral internship at the School of Education, Johannes Kepler Universität in Austria, which took place from October 2021 to January 2022.

Throughout his internship, Mr Alves Araujo actively participated as both an assistant and a professor in the Research Methods and Statistics classes that I taught at the graduate and undergraduate levels, respectively. The Research Methods course was specifically designed for students pursuing their master's and PhD degrees at the Linz School of Education, while the Statistics course catered to undergraduate students in various teacher training programs.

This internship facilitated valuable collaboration and research output, as Mr. Alves Araujo had the opportunity to engage with students and gather data from their activities. These students granted him the necessary authorization to utilize their reports and findings in his doctoral thesis.

Therefore, I hereby grant José Ronaldo Alves Araujo full authorization to proceed with his doctoral research, utilizing the data acquired during his internship period and under my ongoing supervision.

Should you require any further information or documentation regarding this matter, please do not hesitate to contact me.

Sincerely yours,

Univ.-Prof. Dr. Zsolt Lavicza

Univ.-Prof. Dr. Zsolt Lavicza
Department of STEM Education
Linz School of Education

P +43 732 2468 6872
Zsolt.lavicza@jku.at

Office:
Daniela Roidinger
+43 732 2468 6888
daniela.roidinger@jku.at

**JOHANNES KEPLER
UNIVERSITY LINZ**
Altenberger Str. 69
4040 Linz, Austria
www.jku.at
DVR 0093696