TRANSNUMERATION AND DATA VISUALIZATION: A LOOK AT GRAPHIC ANALYSES OF DATA MEDIATED BY GAPMINDER

José Ronaldo Alves Araújo, Celina Aparecida Almeida Pereira Abar, and Zsolt Lavicza
Pontifical Catholic University of São Paulo
jronaldoaaraujo@gmail.com

This article reports on part of a study to explore the possibilities of graphical data analysis mediated by using technological resources. The research considers transnumeration and data visualization to present analyses about the use of Gapminder in a statistics class for teachers in initial training. The qualitative methodology and data analyses result from participatory observation and access to class activities performed by the future teachers. The results elucidate that the use of Gapminder allowed participants to select representations, including dynamic representations, of statistical data and to interact with graphic representations, resulting in the technology serving as a resource that facilitated their interpretations and understandings about real problems.

INTRODUCTION

This article reports on an investigation that problematizes the implementation of technological resources in activities related to statistical education. The larger research focuses on activities relating to graphical analysis of statistical data and focuses on research on the technological support of three resources: Gapminder (https://www.gapminder.org), GeoGebra (https://www.geogebra.org/), and SPSS (https://www.ibm.com/products/spss-statistics). We focus on analyses using Gapminder in this paper. Gapminder presents data for a diversity of problems that are relevant and discussed around the world—problems based in topics such as per capita income, life expectancy, and atmospheric pollution— and provides tools to generate graphic representations of the data.

Today's world brings a constant bombardment of data representations through various media that contribute to discussions in many different contexts (Prodromou & Dunne, 2017). Keim et al. (2006, p. 9) pointed out that progress in computational power and storage capacity resulted in an increase in the rate of production and data analysis. Laurentiz (2019) pointed out that increased amounts of information available in the form of data can be challenging for consumers and noted a need to generate strategies for analysing and understanding the data and representations of the data. According to Alexandre and Tavares (2007), computational resources that enable visualization and analysis can support the entire data analysis process.

Based on the issue, we proposed a research study that focuses on students' use of Gapminder in an introductory statistics course with undergraduate students from different areas of study such as mathematics, geography, and physical education. In this paper, we focus only on the work of students who are future teachers. The research reported in this paper aimed to identify and discuss possible benefits that result from future teachers' analyzing graphical representations produced by Gapminder. The work is part of the ongoing doctoral work of the first author and was carried out between the second semester of 2021 and the first semester of 2022 at an Austrian university. In particular, this manuscript describes work of implementing activities with future teachers using Gapminder to answer the following question: What possible contributions do Gapminder's graphical representations make to future teachers' graphical analysis of statistical data?

BACKGROUND

In reflecting on statistical education mediated through different technological resources and software, Biehler (1993) highlighted that "this reality increases the didactic relevance of examining these tools more closely" (p. 68). More recently, Andre and Lavicza (2019) stated that technological achievements and their integration into our society reveal a serious challenge for the education of future generations: "educational research on this subject must be intensified and studies on curricula must accompany this development" (p. 254). Laurentiz (2019), for example, stressed the question of how to present complex ideas with a large number of data and suggested the possibility of using data visualization activities in a clear, accurate, and efficient way to provide computational support.

Data visualizations, including tables, line graphs, histograms, and other representations, represent quantitative data in diagrammatic forms or by schemas (Khan & Khan, 2011). Creating data

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

visualizations requires designing graphic exhibitions to display data in descriptive ways, which may include transformed data displays that can facilitate interpretation and obtaining information from the data (Unwin, 2020). Prodromou and Dunne (2017) indicated that data visualization allows us to effectively explore and communicate relevant information about large volumes of data through graphical representations. They pointed out that data visualization encompasses not only the ability to understand the underlying messages that data potentially reveals but also the ability to critically examine statements that are based on data and data representations, including misleading or distorted visualizations that misrepresent data. They see such skills as a prerequisite for citizenship. Keim et al. (2006), pointed out that the basic idea of visual analysis is to "visually represent information, so that human beings interact directly with it, gain insights, draw conclusions, and ultimately improve decisions" (p. 9). The specific advantage of visual analysis is that "those who make decisions can concentrate their complete cognitive and perceptual abilities in the analytical process, while allowing the application of computational resources to increase the discovery process" (Keim et al., 2006, p. 9). They added that visual analysis can help transform the information overload generated by current applications into a useful asset.

We find a place for discussion about data visualizations in the field of statistics education based in the idea of having representations that allow for interaction, obtaining insights, and improving decision-making processes. This dynamic process of changing and transforming data representations to generate and facilitate understanding is how Wild and Pfannkuch (1999) define transnumeration. For them, transnumeration happens when we discover representations of data that help to convey new understandings about the real-world system the data are intended to reveal. Pfannkuch, Rubick, and Yoon (2002) identify transnumeration as a process with stages, as shown in Table 1,

Table 1. Transnumeration stages

1st stage	Quantitative or classification measures relevant to the problem are captured from real-world situations.
2nd stage	Several representations of the data are employed to understand what the data are saying about the situation of the real world.
3rd stage	Statistical summaries are communicated in an understandable and convincing way to the target audience and related to the original situation of the problem.

Next, we present a methodological conception and the procedures we adopted throughout the project.

METHODOLOGY

The research uses the qualitative approach described by Godoy (1995). We consider the future teachers' natural environment and the presence of the researcher in this environment for our research. We use an inductive approach to provide descriptive characterizations of the meaning that teachers give to using Gapminder data visualizations when they analyze data from activities that aim to promote statistical education. Gapminder provides databases for a variety of topics on their website and allows users to create graphic representations to represent the data.

In the light of our consideration of transnumeration and data visualization, we searched for contributions that Gapminder might offer for future teachers to understand the problems they choose to explore. The proposed activity (Figure 1) was designed for teachers to appropriate the tool initially. Their choice of a relevant topic provided an opportunity to conceive of statistics as a means to explore and analyze potentially significant problems. The last part of the activity offered the possibility of observing the future teachers' productions. Using the reports in which the future teachers presented their data analyses, we considered their appropriation of the technology tool, Gapminder, their choice of problems to discuss and analyze, and how representations mediated by the technological resource contributed to their statistical thinking. The descriptive character of the methodology adopted for this study is revealed in the analyses of the graphic representations created by the teachers and their work with the representations.

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.

Figure 1. Proposed activity

In our results, we highlight excerpts to describe the work done by a pair of participants, who we named with the pseudonyms Alice and Ben. Alice and Ben were working on the activity proposed in Figure 1. Their work focused on the analysis of an issue related to increases in CO_2 emission and increase in income.

RESULTS

Figure 2 displays a Gapminder screen used by the duo to discuss CO_2 emissions and to consider the beginning of a historical series that dates from the beginning of the 19th century. Gapminder can present the data dynamically by transitioning among graphs sequentially over time. The starting point of 1811 for addressing the problem reveals limitations with access to information on the subject. Gapminder only graphically represents the available data. When referring to the representation, the future teachers observe that: "The graph shows that very little CO_2 was emitted in 1811. However, you can also see that Europe was involved in the beginning. Over the years, CO_2 emissions have become increasing." Alice and Ben's observations are the result of observing the data representation transition over time and for the year displayed in Figure 2. Axes are clearly labeled to show the quantitative amount of CO_2 and the income for different countries in 1811. Gapminder allows the visualization of countries from a particular region using color. Yellow, for example, represents the European continent, which contributed to the statement that, at first, CO_2 emissions were more visible in European countries.

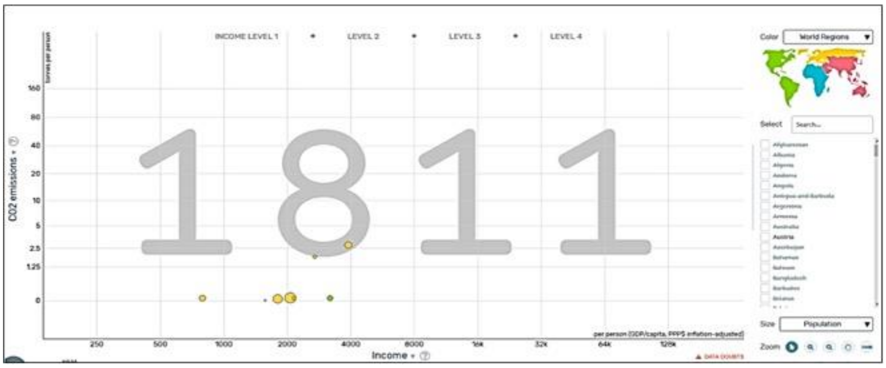

Figure 2. CO_2 emissions as a function of income at the beginning of the series

Alice and Ben continue with their historical analyses and stop at the graphical representation for 1924 that is displayed in Figure 3. In their analyses, they explain: "You can also see that this CO_2 emission increases depending on the income of the different countries, the higher the income." In this case, the pair indicates that there may be a functional relationship, the higher the income, the greater the amount of CO_2 emitted into the atmosphere. Other factors could have been highlighted, for example, there are countries from other regions that were of some relevance for the amount of CO_2 emitted.

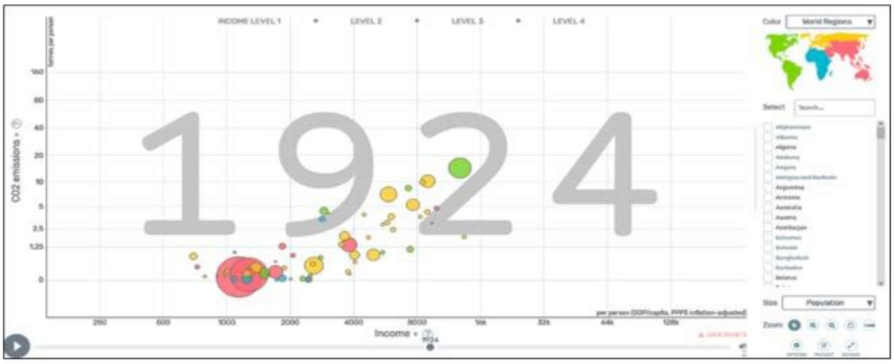

Figure 3. CO_2 emissions as a function of income in the middle of the series

At the end of the series, as shown in Figure 4, we can see that the representation that Alice and Ben generated supports their previous statement to relate the dependent variable [CO_2 emissions] and the independent variable [income]. The pair point out: "This probably explains why Europe is at the top of CO_2 emissions. As you can see from the chart, Africa emits the lowest CO_2, but its income is also lower." In fact, by the colors tied to the different regions of the planet, Alice and Ben made a relevant observation about the most recent context of the problem addressed by them. However, the graphic representation could still allow for further observations. For example, in all representations there are four established levels (see the top of the displays) that make it possible to visualize concentrations of countries at certain levels to suggest that a direct relationship between higher income and greater amount of CO_2 emitted might not necessarily be true.

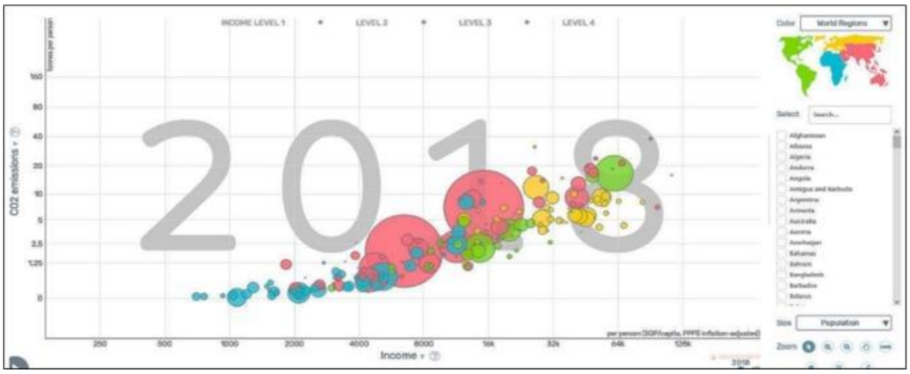

Figure 4. Graphical representation of CO_2 emissions in relation to income at the end of the series

We now present our analyses of Alice and Ben's activity within the context of data visualization and transnumeration. Gapminder allowed the possibility of pausing the representation at certain time points (moments of the series), which in the production of these future teachers, allowed them to point out relevant considerations of CO_2 emissions as a function of income. In this case, another observation is that these selected moments bring consideration of a third variable of time.

From the point of view of transnumeration (Pfannkuch et al., 2002), the duo established a problematic, real-world context. The pair, knowing the technological resource, concentrated on evaluating CO_2 emissions as a function of income but also displayed an interest in critically examining the data and visualized representations (Prodromou & Dunne, 2017). The aid of the graphic representation in the transnumeration process allowed them to understand what the data is saying, as evidenced by their statements: "you can also see that Europe was involved in the beginning [of CO_2 emissions]. Over the years, CO_2 emissions have become increasing." This observation is possible because Gapminder allows consideration for different regions of the planet and represents them in different colors. The bubbles and their colors represent countries from certain regions (and their populations based on the size of the bubble) to contribute to the data visualization process (e.g., Unwin, 2020). Furthermore, in the background, Gapminder allows for observing a historical series by indicating the time reference that can be observed in the graphic representation and the dynamic presentation of the graphs in sequence over time. We consider that the future teachers' observations were not based solely on previous knowledge. The dynamic nature of the graphic visualization to portray a historical series (that cannot be properly displayed using static figures) and the variety of visualizations, including the bubble chart, observed during the future teachers' use of the technological resource, are evidence of Gapminder's contribution to their graphical analysis of statistical data. Using Wild and Pfannkuch's (1999) definition of transnumeration, we consider that Gapminder's dynamic process of generating changes in data representations can facilitate developing understanding of the real problem.

Although Gapminder does not present statistical summaries, we have seen that Gapminder's representation, especially the scales on the axes, allowed the future teachers to make relevant communications from the point of view of CO_2 emissions. "[...] this CO_2 emission increases depending on the income of the different countries, the higher the emissions, the higher the income." And, with the support of the graphic representation, they elucidate that "Africa emits the lowest CO_2, but its income is also lower." These statements are anchored in the visualization of the data that the graphic representation allows. We conjecture that such understanding of the CO_2 emissions of each country in relation to their respective incomes is something that results from the fact that Gapminder allowed for the representation of data that favored the transnumeration process to some extent.

The descriptions of the report presented by the duo of Alice and Ben revealed that the representations of the data that Gapminder provided were relevant for supporting the future teachers' data analysis. In addition to the four levels that subdivide the chart vertically, the size and color of the bubbles and the country that each represents were available resources that could have contributed to deepen the teachers' understanding of the problem.

CONCLUSION

In this investigation, we observed Gapminder's possible contributions in graphical data analysis activities by considering how the technology supports data visualization and transnumeration. In the activity with the future teachers, we saw that they used Gapminder to establish a real-world problematic context of the real world [CO_2 emissions] and were interested in critically examining the data and its representations because they tangentially related the real problem with another factor [income]. From the future teachers' statements from the activity, we see evidence that the graphic representations were a resource that seem to have promoted their understandings about the data. Specifically, the teachers were able to interact with the graphic representation, a variety of visualizations, and the bubble chart depicting a series dynamically, which is the contribution Gapminder made to the visualization of data.

Although positive, in the results observed from the activity of Alice and Ben, we saw that other resources available by Gapminder were not used. This reveals a limitation of the activity in that it is necessary to expand access and mastery of the technological resource by users to foster additional conjectures about the problem addressed by the duo. In addition to this factor, from a methodological point of view, we acknowledge that in this report, we did not analyze the verbal discussions of the participants, which is a limiting factor that will be considered in future developments of the thesis project.

REFERENCES

Alexandre, D. S., & Tavares, J. M. R. S. (2007). Factores da percepção visual humana na visualização de dados [Factors of human visual perception in data visualization]. In J. C. S. R. Delgado, A. D. Santos, A. Rodriguez-Ferran, J. Oliver, P. R. M. Lyra, & J. L. Alves (Eds.), *Proceedings from Métodos Numéricos e Computacionais em Engenharia CMNE (Congresso de Métodos Numéricos em Engenharia)/CILAMCE (Congresso Ibero Latino-Americano sobre Métodos Computacionais em Engenharia) 2007*. APMTAC/FEUP.

Andre, M., & Lavicza, Z. (2019). Technology changing statistics education: Defining possibilities, opportunities and obligations. *Electronic Journal of Mathematics & Technology, 13*(3), 253–265.

Biehler, R. (1993). Software tools and mathematics education: The case of statistics. In C. Keitel & K. Ruthven (Eds.), *Learning from computers: Mathematics education and technology* (pp. 68–100). Springer. https://doi.org/10.1007/978-3-642-78542-9_4

Godoy, A. S. (1995). Pesquisa qualitativa: Tipos fundamentais [Qualitative research: Fundamental types]. *Revista de Administração de empresas*, *35*(3), 20–29. https://www.scielo.br/j/rae/a/ZX4cTGrqYfVhr7LvVyDBgdb/?format=pdf&lang=pt

Keim, D. A., Mansmann, F., Schneidewind, J., & Ziegler, H. (2006, July). Challenges in visual data analysis. In E. Banissi, R. A. Burkhard, A. Ursyn, J. J. Zhang, M. Bannatyne, C. Maple, A. J. Cowell, G. Y. Tian, & M. Hou (Eds.), *Tenth International Conference on Information Visualisation (IV'06)* (pp. 9–16). IEEE Computer Society. https://doi.org/10.1109/IV.2006.31

Khan, M., & Khan, S. S. (2011). Data and information visualization methods, and interactive mechanisms: A survey. *International Journal of Computer Applications*, *34*(1), 1–14. https://doi.org/10.5120/4061-5722

Laurentiz, S. (2019). Alguns desafios para a visualização de dados [Some challenges for data visualization]. In *18º Encontro internacional de Arte, Ciência Tecnologia—#18.ART*. Media Lab/Br. https://www.eca.usp.br/acervo/producao-academica/003055777.pdf

Pfannkuch, M., Rubick, A., & Yoon, C. (2002). Statistical thinking and transnumeration. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific. Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia, Auckland* (pp. 567–574). MERGA.

Prodromou, T., & Dunne, T. (2017). Data visualisation and statistics education in the future. In T. Prodromou (Ed.), *Data visualization and statistical literacy for open and big data* (pp. 1–28). IGI Global. https://doi.org/10.4018%2F978-1-5225-2512-7.ch001

Unwin, A. (2020). Why is data visualization important? What is important in data Visualization? *Harvard Data Science Review*, *2*(1), 1–7. https://doi.org/10.1162/99608f92.8ae4d525

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x