# WHAT KINDS OF QUESTIONS ARE FORMULATED BY A GROUP OF STUDENT TEACHERS WHEN DEALING WITH DATA REPRESENTATIONS THAT ADDRESS SOCIAL ISSUES?

Francisca M. Ubilla and Núria Gorgorió
Autonomous University of Barcelona
francisca.manriquez@uab.cat

*In this study we analyze the questions generated by a group of student teachers when dealing with summary tables or summary graphs that address social issues linked to civic statistics. Drawing on different authors who define various levels of understanding of statistical representations, we established a characterization of the questions formulated by these students. We help to fill a theoretical gap, advancing in the characterization of the questions that students ask when confronted with data presented in a complex way in summary tables and summary graphs. This study can provide guidance on how to teach students to decode tables and representations typical of civic statistics.*

## INTRODUCTION

Reading and questioning the world through the *lenses of critical statistics* (Weiland, 2017) demands a knowledge of the social context in which the data was generated as well as an understanding of the statistical procedure that allowed the collection of the data. Taking this as a starting point, we consider that teacher education should offer learning opportunities that enable teachers to put themselves in the role of consumers and producers of data so that they are able in this way to read and write about the world through statistics (Weiland, 2017). In line with Sousa et al. (2020), we consider it necessary to promote the role of future teachers as data producers in order to develop and enhance their ability to make decisions when handling social phenomena presented to them through statistics. To this end, it is necessary in statistical education to address the ability to generate questions, either to carry out research or to examine a set of data.

Accordingly, civic statistics emerges as a subdiscipline of statistics focused on contexts and topics that are relevant to society (ProCivicStat Partners, 2018). It aims to generate educational instances that enable citizens to look beyond the data and identify the political and social implications of statistical information (Engel et al., 2021). On the other hand, civic statistics involves the use of complex visualizations and data that do not usually appear in the contexts of school statistics (Kwon et al., 2021). Thus, the need arises during teacher education to present diverse representations with the aim of fostering a process that decodes the components of these visualizations, as well as interpreting these components in a particular social context.

Within the framework of civic statistics, we were interested in finding out what questions a group of students asked themselves at the beginning of their training as elementary school teachers when they had to deal with tables and graphs that address social issues.

## DATA REPRESENTATION

In accordance with Arteaga et al. (2011), we understand tables and graphs as cultural objects in view of their notable presence in both the school environment and the media. Therefore, the need arises for future teachers to become conversant with the different types of data representations so that, on the one hand, they can use them as data organizers, and on the other, so that they can learn to distinguish the nature of each representation and its corresponding reading and interpretation.

With regard to data producers, Schield (2001) states that "a goal of statistical literacy is to construct readily understandable ratio-based *comparisons* that follow directly from data, take into account multiple factors, and can support arguments about causation" (p.1). On this basis he argues that tables are representations that, given their organization into rows and columns, facilitate the comparison of different elements. However, not all statistical tables have the same format or the same purpose. Schield (2001) identifies *summary*, *demonstration*, and *reference tables*, which refer to tabulated data about groups of subjects, whereas *detail list tables* are tables with lists of data about each individual subject. Graphical representations can similarly display a frequency count or the representation of more complex indexes, rates, or indicators. In the context of compulsory education, statistical tables and graphs are the most commonly used representations, with a major presence in textbooks (Pallauta et

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 

al., 2021; Shreiner, 2018). However, the tables and graphs typical of the school context differ greatly from those that appear in the media and in statistical reports (the reader can find this complexity reflected in https://guides.library.duke.edu/datavis/vis_types). Continuing with the terminology proposed by Schield (2001), summary tables are the most studied in school, particularly frequency tables with one variable, whereas double-entry tables and *detail list tables* are rarely encountered. However, in the media and/or in statistical reports there is a tendency to use summary tables that show percentages or ratios, as well as two or more variables in a single table, and also graphs and/or infographics that represent rates or ratios obtained through a mathematical procedure. The wage gap calculation infographic in Figure 1 is an example of this type of representation.

Due to the complicated nature of the information appearing in representations, their reading and interpretation requires a decoding process that can vary in complexity depending on the information represented and the question formulated. Curcio (1987) establishes three levels of graph reading: *reading the data*, *reading within the data*, and *reading beyond the data*. On the other hand, Friel et al. (2001) propose three levels of questions that reveal different levels of understanding of the representations: the *elementary level*, focused on extracting information from the data; the *intermediate level*, focused on establishing relationships and interpolating data; and the *overall level*, which requires extrapolation from the data and the establishment of relationships not explicit in the representation. In the proposal of Friel et al. (2001), the level of the question posed allows teachers to obtain information on how students understand the graphic representation based on how they answer it. These three levels of questions are related one by one to the levels of graph reading proposed by Curcio (1987). On the basis of these levels, Shaughnessy (2007) proposes a fourth level of graph reading, called *reading behind the data*, and adds two behaviors associated with graph sense to the six identified by Friel et al. (2001). Table 1 shows Shaughnessy's (2007) final proposal.

Table 1. Levels of understanding of representations (Shaughnessy, 2007, p. 991)

| Levels | Characteristics |
|---|---|
| *Reading* the data | Recognizing components of the graph.<br>Speaking the language of graphs. |
| Reading *within* the data | Understanding relationships among tables, graphs, and data.<br>Making sense of a graph but avoiding personalization and maintaining an objective stance while talking about the graph. |
| Reading *beyond* the data | Interpreting information in a graph and answering questions about it.<br>Recognizing appropriate graphs for a given data set and its context. |
| Reading *behind* the data | Looking for possible causes of variation.<br>Looking for relationships among variables in the data. |

Friel et al. (2001) argue that asking questions is an essential part of understanding representations. In line with these authors and under the idea of the social construction of statistics (Schield, 2007), we consider that in order to understand data representations in depth, it is necessary to develop the ability to generate questions that allow readers to probe and interpret both the representations and their own knowledge of the represented information.

## STATISTICAL QUESTIONS

Generating questions is crucial to the teaching and learning of statistics. Hence, Arnold and Franklin (2021) ask what makes a good statistical question. Taking as references the work done by Arnold (2013), the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) proposal (Franklin et al., 2007) for teaching statistics in schools, and the SET document (Franklin et al., 2015) for teacher education, we decided it was interesting to characterize the questions that student teachers formulate when solving statistical problems and, in particular, those that follow the structure of a cycle of statistical inquiry (Wild & Pfannkuch, 1999).

Arnold (2013) affirms that there are two processes where questions are generated: *question posing*, when questions are generated in a structured way, and *question asking*, when the questions result from a continuous questioning process during problem solving. The *question-posing* process includes, on the one hand, the *investigative questions*, which are the statistical questions to be answered

or the problem to be solved, i.e., those questions that must be answered with the data. On the other hand, there emerge *survey/data collection questions*, which are the questions that make up the data collection instrument; these questions are the ones that serve to obtain the data that is used to develop statistical research. In a *question-asking* process, *interrogative questions* and *analysis questions* appear. *Interrogative questions* are questions in attendance throughout the development of a statistical problem, and their purpose is to check each decision made during problem solving. On the other hand, *analysis questions* are those questions that are asked about the statistical procedures carried out during the resolution of a problem. Questions about tabular and graphical representations are examples of *analysis questions*.

Arnold (2013) affirms that a good *investigative question* clearly indicates the variable(s) of interest, the population or sample, and the purpose of the question—which may be a *summary/description, comparison,* or *association*. Furthermore, a good *investigative question* should be answerable with the data, be interesting to the questioner and to others, and permit the analysis of a whole group rather than just isolated individuals. Ubilla et al. (2021), in their study within the framework of the statistical research cycle, found that the majority of a group of elementary school student teachers with whom they were working formulated *investigative questions* of an essentially descriptive nature. They also observed that most of the participants included their *investigative questions* in the data collection instrument, which showed that they mistook them for *survey/data collection questions*.

On the other hand, Puloka et al. (2021) studied the questions formulated by a group of students aged 13–14 when confronted with categorical data representations: a *detail list table*, a *summary table* (double-entry table), and different bar graphs from CensusAtSchool©. Among the questions that Puloka et al. (2021) characterized are the following: questions aimed at understanding task terminology and/or representations, survey background questions, questions aligned with the *reasoning behind the data*, and quantifying questions.

All things considered, we have not yet found any research that addresses what kind of questions students ask teachers when they have to deal with statistical tables that address social issues.

## METHODOLOGY

We designed an activity based on the cycle of learning from data (International Data Science in Schools Project, 2019), which consists of following a research cycle that starts with second-order data, i.e., data and representations produced by others. (For the details of the activity, see Ubilla & Gorgorió, 2021.) In preparing this task, we generated four data packages and representations from EUROSTAT and from a study entitled "The Lives of Women and Men in Europe: A Statistical Portrait," carried out by the Spanish National Institute of Statistics. The social issues organized by the data packages were as follows: education and work; work and family; habits and health; and life expectancy, health, and retirement.

The activity was carried out by 134 first year students of the Primary Education Degree of the Universidad Autónoma de Barcelona. In this paper we focus on Task 4 of the activity: *From the chosen topic, make a list of questions that can be answered with the data. Choose one (or more than one) that goes beyond a direct reading of the data. Justify your choice.* Our data comes from the written answers of 38 working groups, made up of three to four students each.

## ANALYSIS AND RESULTS

Each group wrote an average of four questions. In this article, we present examples of the questions formulated by the groups that addressed the social issue of education and work. The representations related to education and work consisted of three summary tables presenting information about the countries of the European Union. This included the following: population by gender, percentage of population between 25 and 64 years old with the highest level of education by gender, and percentage of unemployed working population by age (15 to 24 years old and 25 to 74 years old) and by gender. They also worked with three graphical representations: a double bar chart showing by gender the age at which people start their first job (vertical axis) in European Union countries (horizontal axis), and the two representations shown in Figure 1.

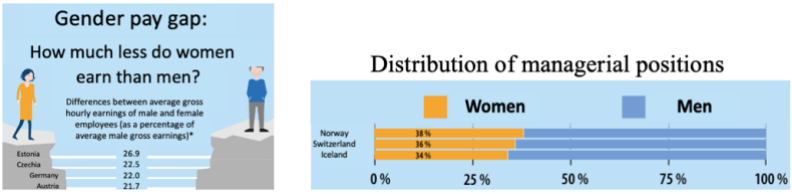


Figure 1. Extract from the graphical representations in the "education and work" data package

Friel et al. (2001) allows characterizing the way in which students understand graphs when answering questions at different levels. On the other hand, Shaughnessy (2007, p. 991) proposes four levels of graph reading. In our study we characterize the questions posed by students when confronted with different types of representations with features of civic statistics. We draw on Friel et al. (2001) and Shaughnessy (2007) to guide our analysis.

When characterizing the students' questions during the process of deductive analysis, we found characteristics that match up with the proposals of both Friel et al. (2001) and Shaughnessy (2007), but also some differentiating aspects because we are interested in the intention of the questions posed by the students in relation to different types of representations. Thus, in the analysis process we used the following categories:

- *Asking the data.* This type of question requires a direct reading of the information present in the representations. For example, in the case of Figure 1, the G2 group asked: "Which country has the highest percentage of female managers?" And for its part, the G7 group asked: "Which country has the widest wage gap?" Another type of question in this same category seeks to compare two values of the same variable or the same value of two related variables. For example, the G3 group asked: "How does the wage gap in Spain compare with the European Union?" The questions in this category are reminiscent of those defined by Friel et al. (2001) as elementary questions.
- *Asking between the data.* The questions in this category consisted of those that asked for a calculation using the data appearing in the representations. For example, the G8 group asked the question: "What is the average wage gap between men and women in the European Union?" The questions in this category could be associated with those defined by Friel et al. (2001) as intermediate questions.
- *Asking beyond the data.* This type of question aims to identify relationships between different variables present in the representations. For example, the G5 group asked: "Does the level of studies achieved influence the age at which people start their first job?" Another type of question in this same category consisted of those that sought to identify trends between different variables. For example, the G2 group asked: "What trends are observed linking the level of higher education and management position in the countries where the gender gap is the largest, smallest, and closest to the average in the EU?" The questions in this category would correspond to those defined by Friel et al. (2001) as overall questions.
- *Asking behind the data*. This type of question seeks to find explanations for the relationships or trends identified in the data representations. For example, the G1 group asked: "What are the reasons for the wage gap?" Note that Friel et al. (2001) does not consider this category.

## DISCUSSION AND CONCLUSIONS

Given the format of the activity carried out with the students, the questions they formulated can be classified as analysis questions within the categories of Arnold (2013), because they are questions asked based on data representations. However, while we consider that questions of the type *ask the data* and *ask between the data* are analysis questions, questions of the type a*sk beyond and behind the data* could be considered *investigative questions* because they can generate new cycles of investigation. But the last two can still be considered analysis questions for consumers. Consumers can think about these questions without having to produce new data. The distinction between categories of questions highlights the need to move between being data consumers and being data producers or vice-versa. An active attitude can be fostered in this way, which begins with understanding the information after reading it, then generating questions about the data and beyond the data, and linking these questions to the specific context.

Schield (2007) reflects on the social construction of statistics and proposes the need to develop hypothetical thinking to question statistical messages. Following this line, from the position of data consumers, the categories proposed by Puloka et al. (2021) *about the terminology of the representations* and *about the background of the survey* could be part of a new category, namely *asking about social construction of data*, which would be part of the question-asking process proposed by Arnold (2013) (see Figure 2).

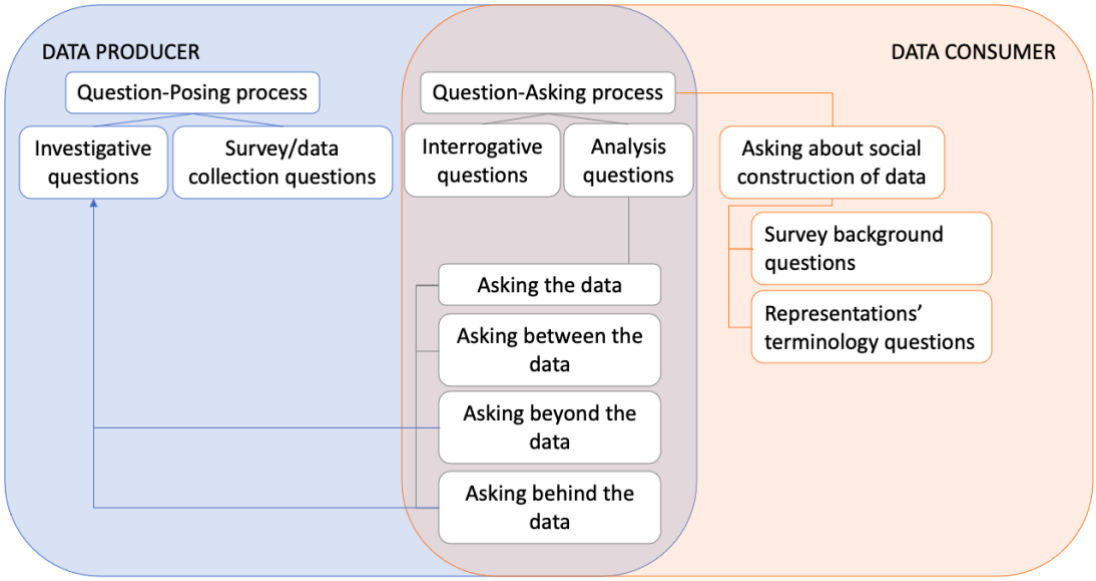


Figure 2. Types of questions in the development of statistical problems from the perspective of data producers and/or data consumers (Created by the authors)

To sum up, we would like to highlight that questioning activities such as the one proposed help students to reflect on and discuss social issues by facilitating questions that go beyond a direct reading of the representations. However, we consider it appropriate to reflect on the relevance and feasibility of providing answers to the different types of questions that arise when working with data in social contexts. Depending on the educational level of the students, the questions they formulate may vary in depth and some of them cannot be answered with the tools available to them. However, there is no reason why this should hinder reflection on the social issues reflected in the data.

ACKNOWLEDGEMENTS

This study was carried out under the umbrella of the project "Study of the Requirements for Admission to Primary Education Teacher Degrees from the Perspective of Mathematical Knowledge," funded by the Directorate General of Research, Development, and Innovation, of the Ministry of Science, Innovation, and Universities of Spain, with reference EDU2017-82427-R, and with the support of a Postgraduate Scholarship Abroad funded by the Chilean National Agency for Research and Development (ANID), whose reference is ANID PFCHA/DOCTORADO BECAS CHILE/2018 - 72190313.

REFERENCES

Arnold, P. (2013). *Statistical investigative questions—An enquiry into posing and answering investigative questions from existing data* [Doctoral thesis, The University of Auckland]. ResearchSpace@Auckland. https://researchspace.auckland.ac.nz/handle/2292/21305

Arnold, P., & Franklin, C. (2021). What makes a good statistical question? *Journal of Statistics and Data Science Education, 29*(1), 122–130. https://doi.org/10.1080/26939169.2021.1877582

Arteaga, P., Batanero, C., Cañadas, G., & Contreras, M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números. Revista de Didáctica de las Matemáticas, 76*, 55–67.

Curcio, F. R. (1987). Comprehension of mathematical relationships experienced in graphs. *Journal for Research in Mathematics Education, 18*(5), 382–393. https://doi.org/10.2307/749086

Engel, J., Ridgway, J., & Weber, F. (2021). Educación estadística, democracia y empoderamiento de los ciudadanos. *PARADIGMA, 42*(e1), 1–31. https://doi.org/10.37618/PARADIGMA.1011-2251.2021.p01-31.id1016

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R., & Spangler, D. (2015). *The statistical education of teachers*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/EDU-SET.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Pre-K–12 Guidelines for assessment and instruction in statistics education (GAISE)*. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseprek-12_full.pdf

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124–158. https://doi.org/10.2307/749671

International Data Science in Schools Project (IDSSP) Curriculum Team. (2019). *Curriculum frameworks for introductory data science*. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Kwon, O. N., Han, C., Lee, C., Lee, K., Kim, K., Jo, G., & Yoon, G. (2021) Graphs in the COVID-19 news: A mathematics audit of newspapers in Korea. *Educational Studies in Mathematics 108*(1–2)**,** 183–200. https://doi.org/10.1007/s10649-021-10029-0

Pallauta, J. D., Gea, M. M., & Arteaga, P. (2021). Caracterización de las tareas propuestas sobre tablas estadísticas en libros de texto chilenos de educación básica. *PARADIGMA, 41*(e1), 32–60. https://doi.org/10.37618/PARADIGMA.1011-2251.2021.p32-60.id1017

ProCivicStat Partners. (2018). *Engaging Civic Statistics: A call for action and recommendations.* https://iase-web.org/islp/pcs/documents/ProCivicStat_Report.pdf

Puloka, M., Budgett, S., & Pfannkuch, M. (2021). What questions do novices pose about categorical data? In R. Helenius & E. Falck (Eds.), *Statistics education in the era of data science: Proceedings of the satellite conference of the International Association for Statistical Education.* IASE. https://doi.org/ 10.52041/iase.lciru

Schield, M. (2001, August). Statistical literacy: Reading tables of rates and percentages. In *Proceedings of the Annual Meeting of the American Statistical Association, Statistical Education Section*.

Schield, M. (2007, April 6). *Teaching the social construction of statistics* [Paper presentation]. Midwest Sociological Association, Chicago, IL. http://www.statlit.org/pdf/2007SchieldMSS.pdf

Shaughnessy, J. M. (2007). Research on statistics' reasoning and learning. In. F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Information Age.

Shreiner, T. L. (2018). Data literacy for social studies: Examining the role of data visualizations in K–12 textbooks. *Theory & Research in Social Education, 46*(2), 194–231. https://doi.org/10.1080/00933104.2017.1400483.

Sousa, L. D., Lopes, C. E., & Fitzallen, N. (2020). Creative insubordination in statistics teaching: Possibilities to go beyond statistical literacy. *Statistics Education Research Journal, 19*(1), 73–91. https://doi.org/10.52041/serj.v19i1.120

Ubilla, F. M. & Gorgorió, N. (2021). From a source of real data to a brief news report: Introducing first-year preservice teachers to the basic cycle of learning from data. *Teaching Statistics, 43*, S110–S123. https://doi.org/10.1111/test.12246

Ubilla, F. M., Vásquez, C., Rojas, F., & Gorgorió, N. (2021). Santiago–Villarrica–Barcelona: The statistical investigative cycle in primary education teacher training. *Statistics Education Research Journal, 20*(2). https://doi.org/10.52041/serj.v20i2.392

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics, 96*(1), 33–47. https://doi.org/10.1007/s10649-017-9764-5.

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