A COLLABORATIVE PROJECT OF THE SUBJECTS GEOGRAPHY AND COMPUTER SCIENCE: IMPLEMENTING STATISTICAL INVESTIGATIONS WITH GAPMINDER IN SECONDARY SCHOOLS

Andreas Prömmel and Daniel Frischemeier
Gymnasium Ernestinum Gotha
University of Muenster
aproemmel@me.com

Asking good statistical questions is an important prerequisite for a good statistical investigation. In a joint collaboration of the subjects of geography and computer science, students of a 12th grade geography course of a high school worked on statistical projects and had to deal with resources and their global inequality. The Gapminder software was used as a data base and visualization tool to explore multivariate data. In small teams, the students had to explore the data using Gapminder and to present their topics in short statistical enlightenment videos. We collected working sheets of the student's project work and analysed them with qualitative methods. To get a better instruction for asking statistical questions, some results regarding a category system are presented in this paper.

BACKGROUND

In school subjects, the handling of data plays an important role. In the natural sciences, measurement data are collected and evaluated. In mathematics, statistics runs as a curricular strand through all grade levels. Geography lessons in particular need up-to-date data to motivate students to deal with it. But textbooks are often outdated and not suitable for that. Therefore, the idea of linking computer science content with geography content in a multidisciplinary project emerged. The handling of application software is typical for computer science lessons. Using the Gapminder software, students should gain initial experience with more complex statistical contexts. We have chosen this program because current statistical data can be used in different ways and contexts. In particular, data are relevant for geography teaching, such as dimensions of globalization, development of world trade, economic sectors, structural change, tourism, and terms of trade. The idea for the project was first to learn how to use the Gapminder software in computer science lessons on some statistical topics. In a second part, the participants were asked to create a statistical educational video on a topic of their own choice. These videos should then be used and discussed in geography lessons. Many projects of this kind have shown that the process of developing statistical questions is very important for a successful and qualitatively valuable statistical investigation. Therefore, we want to focus on this aspect in this article to be able to support students more appropriately in the design of statistical questions. Our research question is: How far are secondary school students able to generate adequate statistical questions with minimal help in the instruction?

LESSON PLAN AND LEARNING TRAJECTORIES FOR GAPMINDER

In the first part of the project students should get to know how to use the software Gapminder and make some statistical investigations in the frame of the PPDAC cycle—problem-plan-data-analysis-conclusion (see Wild & Pfannkuch 1999). Only a few parts of the ideal-typical cycle were addressed in the units. After an introductory unit on the Gapminder software, the focus was on addressing two problems, population as well as poverty and inequality. The focus was on analysis and conclusion to explore the possibilities of the Gapminder tool with the guidance of teacher instructions and adapted working material. The basis was the material (working sheets with tasks, information, and classroom activities) developed by the ProCivicStat project for students at high schools and universities (see https://rstudio.up.pt/shiny/users/pcs/civicstatmap/). ProCivicStat is part of the International Statistical Literacy Project and supports the learning and teaching of statistics by developing materials for teacher and students to deal with current social issues (see https://iase-web.org/islp/pcs/). An overview of the lesson plan of the first part of the statistical project can be found in Table 1.

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

Table 1. Lesson plan with teacher input and classroom activities

Lesson	Content/Process	Classroom activities
Unit 1	• Getting to know Gapminder • Teacher input: What does Civic Statistics mean? • Gapminder-Quiz	How to use Gapminder for exploring world's population: • Which are the 5 most populous countries in the world? • Life expectancy and infant mortality rate have an influence on population development, find graphical representations.
Unit 2	• Population—How will the world's population develop by the year 2100? • Getting started with short videos from Gapminder, e.g. "How Did the World Population Change?" • Teacher input: Which variables are available under the topic Population? Which variables are related to this?	Create meaningful graphics on the specified topics for a presentation. Find statistical statements that match your graphics. 1. Development of the world population by age group (1950, 2018, 2050, 2100) 2. Development of the number of newborns per woman 3. Relationship between life expectancy and the number of newborns per woman

Unit 3	• Presentation and discussion of the results from unit 2: Population—How will the world's population develop by the year 2100
	• Hints for presentation and discussion, e.g.,
	  • What is displayed? (title, chart type, scale)
	  • What categories or values are there? (categorical or numerical variables)
	  • Which data developments and which correlations can you describe? (largest/smallest value, relative increase or decrease, trends, ...)
	  • Which questions can be answered with the help of graphics? Which ones don't?	

Unit 4	• Poverty and inequality—How are poverty and wealth distributed in the world?
	• Teacher input: Indicators on poverty and income inequalities in Gapminder, e.g.,
	  • Which countries are rich?
	  • Are there regional differences?
	  • Are populous countries rich?
		Explore the link between extreme poverty and food supply.
		• Which indicators are most strongly associated with extreme poverty? Describe their context.

Unit 5	• Presentation and discussion of the results from unit 4: Poverty and inequality—How are poverty and wealth distributed in the world? • Hints for presentation and discussion as in unit 3 • Summary video by Hans Rossling "How to end poverty in the world"	
Unit 6	Project meeting—Video on a topic that can be explored with the statistical data of Gapminder: 1. Team finding 2. Collection of ideas for topics (HDI, climate, economic growth, ...) 3. Design of the statistical questions of investigation 4. Agreements for submission, evaluation, and grading	

DATA ANALYSIS & RESULTS OF THE STUDENTS PROJECTS

In the first part of the project, the students practiced how to use Gapminder for data analysis and conclusions on two problems on civic statistic issues. In the second part of the project, the students now have to work on a self-chosen topic in the frame of the PPDAC cycle. In the computer science course, six teams have been formed, which have dealt with partly very different geographical topics: Nuclear energies (Group 1), Arms exports and Arms imports (Group 2), Ecological footprint (Group 3), Drinking

water supply (Group 4), Tourism in times of crisis (Group 5), International tourism and economic sectors (Group 6). The teams consisted of two to three students.

In this part of the paper, we will focus on the first two parts of the PPDAC cycle, and we will investigate in which way our students are able to generate adequate statistical questions within their project work. Each team had to fill out a worksheet with statistical questions and associated variables before they could start the investigation on the chosen topic. As minimal support the teams got some hints on a worksheet to find suitable research questions, e.g., "Search the Gapminder variable list to find the topic you have chosen with your team. Create as many questions as possible in your group about the specified variables. Also link two or more variables in a question. Choose the three most exciting ones from all the questions you created. In addition to these questions, also specify the variables which are necessary for your analysis."

As Arnold (2013) and Leavy and Frischemeier (2022) have identified, the quality of the data exploration, the analysis, and the results depend heavily on the quality of the statistical questions. To evaluate the statistical questions, we have taken the category system from Leavy and Frischemeier (2022) and adapted this for our purposes. The category system from Leavy and Frischemeier (2022) was oriented to the four levels of Arnold (2013). In this specific case, we will only concentrate on specific categories. We use qualitative content analysis (Mayring 2015) to analyze the quality of the statistical questions. For our coding we use a mixed approach, in the sense that we generated categories in a deductive way taking into account already existing categories from Arnold (2013) and Leavy and Frischemeier (2022) and refined and added categories in an inductive way. Gapminder enables students to discover relationships between variables, therefore we investigate in which way our students are able to create statistical questions covering more than one variable.

All categories with their definitions and key examples can be found in Table 2. Furthermore, all questions generated by the six groups (G1-G6) can be found in Table 2. Additionally, we made a frequency analysis of the occurrence of several categories (see first column in Table 2). One first interesting observation is that the most questions include two variables. This is a positive fact that our participants take into account two variables for their exploration. In addition to that we can also say that seven of all 17 questions are relationship questions that investigate the relationship between two variables. These relationship questions can be identified by signal/trigger words such as "in general," "in how far," etc. (see trigger words italicized in Table 2). More than 40 % of the questions are coded as "relationship/trend" questions. Only a few groups (G1, G6, 2 of 17 questions) take into account the relationship over time (although Gapminder facilitates this kind of exploration). Group 2 exclusively generated questions that can be answered with a single, specific value.

CONCLUSION

This statistics course has two special features. On the one hand, the two subjects of geography and computer science have worked together in a collaborative project. On the other hand, the project consisted of two parts, a more teacher-oriented first part and a more student-oriented second part. The students had to apply their experience in using Gapminder for civic statistic issues in their own statistical enlightenment video.

For the design of adequate statistical questions and required variables, the students received minimal support by a worksheet. It seems that Gapminder in comparison to previous studies with other software promotes the use of more than two variables and their relationship.

All in all, we can say that our participants generate adequate statistical questions, which allowed them to conduct elaborated statistical investigations and explorations. Therefore, this a positive and remarkable observation that more than half (9 of 17) of the statistical questions can be considered as questions that tackles development over time or a relationship/trend between variables. For the future one might consider providing students with further support such as a checklist or specific feedback. In further research we will also investigate the quality of the statistical investigations.

With a further preparation and modification of the materials for high schools in the frame of design-based research, this project could be repeated in a similar manner. This would be an important contribution to the development of statistical literacy for students at the content of social issues.

Table 2. Overview on the statistical questions generated by our students

Question type	Questions from the six groups
No question (2 of 17) A statistical question cannot be identified. The students have rather formulated an expression without a question.	• Decline of the economy in various countries due to crisis (G5) • Declining tourism in crisis countries (G5)
Only Yes/no without relationship (3 of 17) The statistical question can be answered with yes or no and it is no relationship between variables mentioned	• Do countries which create much nuclear power use these powers? (G3) • Does arms export lead to higher GDP? (G4) • Do ended crises have an impact on countries? (G5)
Specific Value (3 of 17) The statistical question can be answered with a single, specific value	• How much do several countries spend in the tourism sector on average? (G2) • How many employees work in tourism-related service areas on average? (G2) • What are the proportions of men and women in the tourism sector in Europe? (G2)
Development over time (2 of 17) The statistical question includes the consideration of the development of a variable over time	• Has the world's drinking water supply improved over time? (G1) • How does the share of alternative energies develop over time in different countries? (G6)
Relationship questions/Trend (7 of 17) The statistical question incorporates a relationship between variables which can be mentioned by expressions like "generally," "in how far," "related," "tendency"	• Do poorer countries around the world *generally* have poorer drinking water supplies? (G1) • *To what extent* do nuclear disasters influence production with nuclear energy sources? (G3) • Is CO2 emission *related* to average income? (G6) • Is the drinking water supply worldwide *higher the more* drinking water resources one has available? (G1) • How does the level of development of a country influence energy production with nuclear energy sources? (G3) • How does the export of weapons influence the number of people killed in combat? (G4) • How does the political system of a country influence the arms exports of a state? (G4)

REFERENCES

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

Leavy, A. & Frischemeier, D. (2022). Developing the statistical problem posing and problem refining skills of prospective teachers. *Statistics Education Research Journal, 21*(1). Article 10. https://doi.org/10.52041/serj.v21i1.226

Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 365–380). Springer. https://doi.org/10.1080/10986065.2016.1151294

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