# TEACHING SOCIAL SCIENCE STUDENTS TO THINK WITH DATA

Joachim Engel and Florian Weber-Stein
Ludwigsburg University of Education
engel@ph-ludwigsburg.de

*A vibrant democracy that wants to prove resilient to expertocratic strategies of rule needs citizens who participate in public deliberations and intervene in political affairs. Without a basic understanding of statistical concepts and handling of data, it is difficult to follow media coverage of, e.g., the pandemic or climate change, let alone intervene in political processes. However, the high relevance of statistical literacy as an integral part of civic education contrasts with its actual value in classroom practice at schools and universities. Social science teachers are usually unprepared to include quantitative evidence as a source in their teaching. Our concept of teaching civic statistics to social science students is based on ProCivicStat and integrates statistical methods and socio-political content. The focus is on critical thinking and conceptual understanding to teach students how to be better consumers and analysts of the kinds of quantitative information and arguments that they will encounter.*

## DATA LITERACY IN THE SOCIAL SCIENCE CURRICULUM

The increased importance of data for political decisions implies a responsibility of the education system to equip citizens with the skills to competently engage with data and statistics. Various education policy strategy papers such as the report "Data Literacy for the Public Sector: Lessons from Early Pioneers in the U.S." (https://www.datafoundation.org/data-literacy-report-2022) or, in a European version, the "Data Literacy Charta" (https://www.stifterverband.org/charta-data-literacy) call for data literacy to be taught in all areas of education.

In the social sciences, special conditions prevail for teaching data literacy. On the one hand, research has shown that students here have significantly poorer statistical skills than, for example, in economics or medicine (Berndt et al. 2021, p. 5). It can be assumed that some of the students have explicitly chosen to study social sciences to largely avoid mathematical and technical issues. Approaches to data analysis must be motivated by the relevance of the social or political issues on which students often have "strong opinions" (Ruediger et al. 2022, p. 8) and should minimize computational aspects as much as possible in favor of visual statistics.

With regard to preservice teacher students in the field of social studies, the above-mentioned conditions apply to a particular extent. It can be assumed that the aversion to technological-mathematical study components is rather more pronounced. Moreover, in contrast to regular social science students, their curricula include few or no courses on statistics and empirical social research. At the same time, however, they have an outstanding importance as multipliers for teaching data literacy as "future skills" (Schüller et al., 2019). Prospective educators, especially in civics, are confronted with the challenge of teaching students the skills to critically reflect on political decisions and to show them ways and means of participation. In a datafied society, this includes abilities to read statistics in social and political contexts, but also to critically question them and, if necessary, to initiate their own data research and surveys. This requires a set of skills that are referred to in the discourse on statistics education as "civic statistical literacy" (Gal et al., 2022a).

## CHALLENGES FOR TEACHER EDUCATION

With the growing importance of quantitative methods for knowledge generation in the social sciences, the need to implement them in university teaching has also increased. Traditionally, empirical social research is taught in special methods courses that often have little relation to the subject-specific content of the subject. In its report "Political Science in the 21st Century," the American Political Science Association diagnoses the structural problem that methods courses with a focus on inferential statistics and research methodology are tailored primarily to the needs and interests of young academics in the cadre of elite universities, but less so to 'regular' BA students with other professional goals (APSA, 2011, p. 20). The study recommends strengthening the "inclusiveness of the political science curriculum" (APSA, 2011, p. 35ff.) by integrating methodological training into subject-based (i.e., non-methods specific) courses, emphasizing action-oriented learning exercises, using data relevant to real world problems, and making use of low-threshold digital technologies. At Ludwigsburg University of

Education in Germany, we (a math/ statistics and a political science educator), came together to design a teaching program to prepare future secondary teachers of social science to think with data. The course has been tested repeatedly in mixed groups of students majoring in social science or mathematics (Weber-Stein & Engel, 2021; Engel et al. 2021). The mathematics students had not taken any prior statistics class. This course does not assume any knowledge in statistics beyond high school, hence may be suitable for courses composed of social science students only. As an introduction to data science and statistics the course focuses, first and foremost, on critical thinking and conceptual understanding. While the social science content may slightly vary (see below), there is an emphasis on visual representations of data facilitated by low-threshold open access webtools, an evaluation of the provenance and trustworthiness of the data as sources of information, and an assessment if a perceived relationship in the data reflects a real relationship in the world (de Mesquita & Fowler, 2021). The course faces the three-fold challenge of combining statistical and socio-political learning with providing a background in the pedagogy of teaching these fields. To address these challenges, the sessions were split in an instructional part and in-class project work in small mixed groups of two to four students to emphasize active learning.

FRAMEWORK AND DESIGN PRINCIPLES

The overriding goal of the course is for students to learn how to process and reflect on statistical information about topics relevant for society and to develop their capacity for evidence-informed decision making as responsible citizens—in short, to foster their skills in Civic Statistics. The course design is based on principles associated with the Statistical Reasoning Learning Environment (Garfield and Ben-Zvi, 2009) as well as the ProCivicStat framework. Contexts are no longer just a pretext to learn statistical methods, but socio-political reflections based on quantitative knowledge are the key to building a critical citizenship in which students are no longer observers of the world, but participants and transformers of social crises. Teaching statistics using authentic and complex social contexts is proven to be more sustainable (Garfield & Ben-Zvi, 2009) and contributes to the education of empowered citizens in the information society.

Statistical information about society is often quite complex. Data are usually multivariate; aggregated data and indicator systems are common; variables interact in a complex network of correlations; data may be time critical, i.e., may change rapidly over time. Data of this type do not form the core of regular statistics instruction in schools or universities. Civic Statistics reflects on the meaning and consequences of analyses and decisions for society. This requires knowledge about the processes of knowledge generation, i.e. how data are being used (and abused) to create new knowledge which implies: developing a sophisticated approach to questions of data provenance and data quality; understanding the uses and abuses of a wide range of methods for presenting and analyzing social data from a variety of sources; ways to represent and model situations; understanding risk; adopting an investigative stance and knowing what questions to ask about the nature, limitations, or credibility of different data sources, statistical statements, and conclusions. When engaging with data about complex social phenomena, there is rarely a simple and clear right or wrong answer. Students must find ways to triangulate the problem and gain insight through different approaches, alternative representations, and examination of related datasets.

Gal et al. (2022a) developed a theoretical framework for Civic Statistics and identified specific skills for understanding statistics about society. The construct civic statistical literacy has been operationalized by distinguishing eleven facets that underpin the ability to engage with social issues. The eleven facets of Civic Statistics as identified by Gal, Nicholson and Ridgway can be organized in three dimensions or groups (see Figure 1; for more details we refer to Gal et al., 2022a, p. 39):

- *Engagement & Action,* comprising: Meaning for society and policy; Critical evaluation and reflection; and Dispositions
- *Knowledge,* comprising: Statistics and risk; Representations, patterns and models; Methodology and enquiry processes; Extensions in official statistics; and Contextual civic knowledge
- *Enabling Processes,* comprising: ICT and search; Quantitative core; and Literacy and communication

Figure 1: A Conceptual model for Civic Statistics: 11 facets
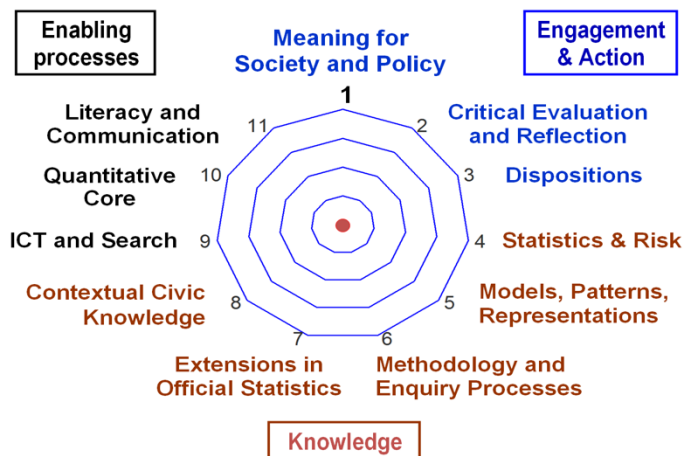
The ProCivicStat operationalization of statistical competence via civic statistical literacy is particularly suitable for the implementation of statistical education in the field of social studies education. The conceptual proximity to social science thinking is shown by the fact that six of the eleven competence facets (1, 2, 3, 6, 8, 11) have strong points of contact with social science skills across all three dimensions (cognitive, dispositional, and processual).

SKETCH OF THE COURSE DESIGN AND IMPLEMENTATION

A central element is the integration of methodological learning about data and statistics with subject-based investigations of "hot" real world problems, in particular the development of statistical concepts and data visualizations along the way as tools for deeper understanding of the context. Main statistical contents being addressed include standard elements of descriptive statistics (location and dispersion parameters, regression and correlation), operationalization of variables and indices, and multivariate data (confounders, Simpson's paradox). Subject-related "hot" contents in our course covered a variety of topics, ranging from investigating income distributions among countries, the difference between the remuneration for employed men and women (aka gender pay gap), global warming and its environmental impact, to exploring the quality of democracy. In our ICOTS-11 paper we focused on how the Covid-19 pandemic challenged democratic and if democratically governed countries dealt as efficiently with the crisis as countries with more autocratic rulers (Weber-Stein & Engel, 2022). In this paper we focus on the course design following the recommendations given by ProCivicStat (Gal et al., 2022b, p 84 ff)

*Develop activities which promote engagement with social issues and develop learners' critical understanding of statistics about key civic phenomena (PCS Recommendation 1)*

Societal relevance and context are of paramount importance for statistics education. Statistics as a discipline is valuable because it can empower people to address real- world problems, and enable engagement with complex social phenomena. Hence, the curriculum should be revised so it introduces specific statistical ideas and techniques in a way that can empower and engage learners in addressing issues of relevance to society. In spring of 2022, the time of teaching the course, the corona crisis was far from being over. Governments across the world responded quite differently to contain the virus, from Nicaragua's denial of the existence of any pandemic to China's strict no-covid policy. In almost all countries, state measures were accompanied by (at least temporary) restrictions on civil rights. Those days of the corona lockdowns incited a strong interest among our students for a discussion of the impact of the pandemic on democratic rights across the globe. The pandemic had obviously exacerbated the simmering crisis of democracy (Parry et al., 2020): on the one hand, with a view to China's 'zero COVID' policy, which had been implemented with authoritarian measures and proven successful for a long time, there was a discussion as to whether democratic governments are able to fight the pandemic as effectively as autocratic systems (Karabulut et. al., 2021). On the other hand, observers warned that the quality of democracy in many countries had declined in the wake of the COVID crisis because fighting the virus had been given priority over safeguarding civil liberties.

*Use socially-relevant data and texts, and highlight the many features of civic statistics, e.g., multivariate, dynamic and aggregated nature of social phenomena (PCS Recommendation 2)*

Students are motivated by interesting subject areas and bring some knowledge (and assumptions) before the actual analysis begins. Understanding social phenomena and making decisions

about them requires dealing with multivariate data. Activities should therefore consider data from important providers, and multiple and novel data sources should be used to triangulate problems. To develop metacognition, students can be asked to reflect on the characteristics of data used in different subject areas, and in a variety of political arguments. Table 1 gives an overview of data sets and tools used in our course as well as contents and statistical ideas taught.

|  | *Module 1* | *Module 2* | *Module 3* | *Module 4* |
|---|---|---|---|---|
| *Content* | Social inequality | Epidemiological basics | Vaccination policies | Democracies in the Pandemic |
| *Database(s)* | American Community Survey[1] | Codap Webinar[2] | Coronavirus (COVID-19) Vaccinations (Our World in Data)[3] | Oxford Covid-19 Government Response Tracker[4] <br><br> Pandemic Backsliding: Democracy During COVID-19 (V-Dem)[5] |
| *Technology* | CODAP (Introduction) | CODAP | Gapminder / CODAP | CODAP |
| *Statistical Concepts* | • measures of location <br> • measures of dispersion | • incidence <br> • reproduction number <br> • herd immunity | • correlation / regression <br> • confounders | • time series data <br> • aggregation and disaggregation of data (indices) |

Table 1: Tabular overview of course content, technologies used and statistical learning content.
[1]https://codap.concord.org/releases/latest/static/dg/en/cert/index.html?di=https://codap.concord.org/sdlc/plugin/index.html (with thanks to Tim Erickson)
[2] https://iase-web.org/Webinars.php?p=211108_0800
[3] https://ourworldindata.org/covid-vaccinations
[4] https://covidtracker.bsg.ox.ac.uk/
[5] https://www.v-dem.net/pandem.html

*Embrace technologies that enable rich visualizations and interactions with data about relevant social phenomena (PCS Recommendation 3)*

There is a variety of technological tools to help learners develop understanding and reasoning. They relieve students from cognitive overload associated with detailed calculations so that they can concentrate on more important tasks such as selecting methods or presenting, evaluating and discussing the results. Relevant and innovative software should include dynamic and interactive data visualizations that encourage learners to use a variety of tools.

*Adopt teaching methods to develop skills of critical interpretation, applicable to a wide variety of data and textual sources (PCS Recommendation 4)*

An important component is the use of carefully designed activities that promote learning through collaboration, interaction, discussion, and interesting problems. Discussions are a central part of the learning environment. Students should ask each other questions and, if possible, answer the questions among themselves. Questions should encourage new conjectures and independent thinking. The activities focus on the development of competences in the field of statistical reasoning. Different teaching methods should be used, with a focus on asking appropriate questions, searching for answers and interpretations in a reasoned way, using appropriate analysis methods. The aim is to develop skills in critical interpretation and to stimulate communication on social issues, including through narrative reports. Through early argumentation with non-linear, multivariate data, modelling competences can also be developed so that students can recognize the respective strengths and weaknesses of modelling in social contexts. Indices have gained enormous importance in social sciences in recent decades. They offer advantages when it comes to compression and thus a reduction in the complexity of individual information. However, a focus on the critical interpretation of data requires that students are also made aware of possible dangers of indices, which can be found at the level of conceptualization and

aggregation. An aha experience may be triggered by de-composing and re-aggregating an index, with giving different weights to individual variables. In our course, this applied to discussing and re-aggregating various sub-indices for summarizing components of democratic societies such as civil liberties on the one hand and accumulation of power in the executive branch to explore possible correlations.

*Use alternative assessments that examine the ability to investigate and critically understand data (PCS Recommendation 5)*

The assessments that learners face during and at the end of any course are the strongest guide for learners about what is to be learned, and the extent to which they are making progress. Using formative and summative assessments to improve and evaluate student learning should ensure that students build the ability to draw from their data analyses implications for society and policy. In our course, we chose to assess student learning via a video production instead of the more traditional paper or oral presentation. The students are asked to investigate a socio-political question relevant to society based on data and to present their results in a video (approx. 10 min). Teams of two or three students, after consultation with the instructors, work on a topic of their own choice. During the process of the video production, the instructors are available as a for consultation and feedback to provide formative assessment, while the end product of the video together with additional homework assignments during the course serves as a summative evaluation determining the students' grades.

Producing the video requires the students to search for a suitable data set and to identify questions to address the chosen topic and its relevance, followed by data visualizations, analyses, and final conclusions. The videos include an introduction to the social context and a characterization of the data, e.g., by describing the data source, data collection methods employed, and explaining the variables involved, so that the viewer receives an overview of the topic and an impression of the quality of the data set. The data analysis around the chosen questions is carried out with the help of software, notably CODAP, Excel or Gapminder. The students document interesting discoveries within the data (e.g., conspicuous patterns) and explain them by placing them in context. Based on relevant calculations and graphical representations, students draw their conclusions. A critical review, limitations of the conclusions and possible subsequent questions are addressed at the end of the video.

CONCLUSION: LESSONS LEARNT

In this type of course, students learn statistical concepts from applying them in the search for solutions in concrete situations, not from following a systematic statistics syllabus. This innovative course connecting introductory statistics with socio-political topics puts specific demands on students as well as on the instructor. Carefully planned and implemented, it addresses an important aim of general education, namely to help students to think with data. Our take-away messages are:

- It is essential to find a balance between teaching important statistical concepts and teaching a critical attitude towards statistics. Therefore, the instructor may limit the statistical inputs to some more basic information and provide time for the students to explore the data sets themselves and to discuss their explorations together.
- Many of the students' analyses are based on common sense, sometimes colored with some misconceptions. Therefore, the instructor needs to listen carefully to the students' reasoning and intervene to clear up misunderstandings.
- A challenging and interesting topic of relevance for the students' life and for society is crucial for motivating the students. Topics such as personal happiness, human development, health, income inequalities, racism, natural disasters, and crime proved to enhance students' interest in exploring data, learning basic statistical concepts and relating the conclusions to the context.
- The use of real and relevant data sets in the teaching of statistics not only ensures a more realistic learning of statistical skills, but also makes them more applicable in situations in everyday life.
- Videos offer opportunities for team work and are a source for motivation, and enhance media literacy. Yet the processes of creating videos are time consuming both for students and instructor. If used as a student project, it is important to focus on a specific topic investigating a single precise question. Videos can be used for formative and summative assessment.

REFERENCES

APSA – American Political Science Education (2011). *Political Science in the 21ˢᵗ Century.* https://www.apsanet.org/portals/54/Files/Task%20Force%20Reports/TF_21st%20Century_AllPgs _webres90.pdf .

Ben-Zvi, D. , & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: goals, definitions, and challenges. In: D. Ben-Zvi & J.Garfield (Eds.),*The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (3-15). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Berndt, M. Schmidt, F., Sailer, M., Fischer, F., Fischer, M. & Zottmann, J. M. (2021). Investigating statistical literacy and scientific reasoning & argumentation in medical-, social sciences-, and economics students. *Learning and Individual Differences* 86 (2021), DOI: 10.1016/j.lindif.2020.101963

De Mesquita, E., & Fowler, A. (2021). Thinking clearly with data. A guide to quantitative reasoning an analysis. Princeton University Press.

Engel, J., Ridgway, J., & Weber-Stein, F. (2021). Educación Estadística, Democracia y Empoderamiento de los Ciudadanos. *Revista Paradigma* (Edición Especial – Educación Estadística), 42:1 (2021), S. 1-31. DOI: 10.37618/PARADIGMA.1011-2251.2021.p01-31.id1016

Gal, I., Nicholson, J., & Ridgway, J. (2022a). A Conceptual Framework for Civic Statistics and Its Educational Applications. In: Ridgway, J. (ed) *Statistics for Empowerment and Social Engagement.* Springer, DOI: 10.1007/978-3-031-20748-8_3

Gal, I., Ridgway, J., Nicholson, J., & Engel, J. (2022b). Implementing Civic Statistics – An agenda for action. In J. Ridgway (Ed.), *Statistics for empowerment and social engagement.* Springer, DOI: 10.1007/978-3-031-20748-8_4

Garfield, J., & Ben-Zvi, D. (2009). Helping Students Develop Statistical Reasoning: Implementing a Statistical Reasoning Learning Environment. *Teaching Statistics, 31* (3), 72–77. DOI: 10.1111/j.1467-9639.2009.00363.x

Karabulut, G., Zimmermann, K., Bilgin, M., & Doker, A. (2021): Democracy and COVID-19 outcomes, *Economic Letters*, DOI: 10.1016/j.econlet.2021.109840

Parry, L.J., Asenbaum, H.,& Ercan, S.A. (2020). Democracy in flux: a systemic view on the impact of COVID-19. *Transforming Government: People, Process and Policy*, DOI: 10.1108/TG-09-2020-0269

Ridgway, J. (ed) (2022) *Statistics for Empowerment and Social Engagement*. Springer, DOI: 10.1007/978-3-031-20748-8

Ruediger, D. et al. (2022). Fostering Data Literacy: Teaching with Quantitative Data in the Social Sciences. Ithaka S+R. DOI: 10.18665/sr.317506

Schüller, K., Busch, P., & Hindinger, C. (2019). Future Skills: Ein Framework für Data Literacy – Kompetenzrahmen und Forschungsbericht (Arbeitspapier Nr. 47). Hochschulforum Digitalisierung. DOI: 10.5281/zenodo.3349865.

Weber-Stein, F., & Engel, J. (2021). Civic Statistical Literacy und Politische Bildung im Informationszeitalter. Kooperative statistik- und politikdidaktische Erkundungen im Feld der Demokratiemessung. In: C. Deichmann & M. Partetzke (Eds.): *Demokratie im Stresstest? Reaktionen von Politikdidaktik und politischer Bildung,* Springer VS 2021, S. 165-192. DOI: 10.1007/978-3-658-33077-4_10

Weber-Stein, F., & Engel, J. (2022). The Covid-19 Crisis as a challenge for the integration of statistical and citizenship education. 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. DOI: 10.52041/iase.icots11.