Statistical literacy and society: What is civic statistics?[1]

Cultura estadística y sociedad: ¿Qué es la estadística cívica?

Joachim Engel

Ludwigsburg University of Education, Germany

Abstract

How can we prepare students to understand statistical data and get insights regarding trends and changes on key societal issues such as demographic change, crime, unemployment, pay equity, migration, health, racism, and other areas of concern to society? This paper, summarizing issues of my plenary talk, introduces by extending the notion of statistical literacy a sub-discipline we call civic statistics. Civic statistics focuses on understanding statistical information about society, as provided by the media, statistical offices and other statistics providers. Civic statistics skills are required for participation in democratic societies, but include data that are open, official, multivariate and dynamic, and that are usually neglected in regular statistics education. I present some specific features of civic statistics, provide examples and describe implications for curricula, teacher activities and the future of statistical education. In my presentation, I will give a short online demonstration exploring data using the browser-based software CODAP. The audience is encouraged to do some of these analyses on their own computer. A short video is provided to help follow or redo the single steps in using the software.

Keywords: Statistical literacy, open data, multivariate thinking, citizenship, curriculum

Resumen

¿Cómo podemos preparar a los estudiantes para que comprendan datos estadísticos e información sobre tendencias y cambios en temas sociales clave como el cambio demográfico, el crimen, el desempleo, la igualdad salarial, la migración, la salud, el racismo y otras áreas de interés para la sociedad? Este documento, que resume los temas de mi conferencia plenaria, introduce una subdisciplina que llamamos estadística cívica, al extender la noción de cultura estadística (o statistical literacy). La estadística cívica se centra en comprender la información estadística sobre la sociedad, tal como lo proporcionan los medios de comunicación, las oficinas de estadística y otros proveedores de estadísticas. Las habilidades de estadísticas cívicas son necesarias para la participación en sociedades democráticas, que incluyen datos que son abiertos, oficiales, multivariados y dinámicos, y que no son fundamentales para la educación habitual en estadística. Voy a presentar algunas características específicas de las estadísticas cívicas, proporciono ejemplos y describo las implicaciones para los planes de estudio, las actividades de los docentes y el futuro de la educación estadística. En mi presentación, daré una breve demostración en línea que explora los datos utilizando el software CODAP basado en el uso de un navegador. Se alienta a la audiencia a hacer algunos de estos análisis en su propia computadora. Se proporciona un breve video para ayudar a seguir los pasos individuales en el uso del software.

Palabras clave: Statistical Literacy, Datos abiertos, pensamiento multivariado, ciudadanía, currículum.

1. Introduction

Democracy thrives on arguments based on evidence. Misinformation, fake news, lack of information and ignorance are threats to our way of life. In an increasingly complex world, the involvement of citizens is an essential resource for public decision-making at

---

[1] This is an adapted and translated version of a paper "Statistical Literacy und Gesellschaft: Was ist Zivilstatistik?", published in *Stochastik in der Schule*,2019, 39(1), 2-12

international, national and local levels. Following the term "statistical literacy", the international strategic partnership ProCivicStat (with the participation of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged from UK, Israel, Germany, Hungary and Portugal) has conceptualized a sub-discipline called Civic Statistics, which focuses on getting of meaning from data that inform about social processes, the social and economic well-being and the realization of civil rights. Understanding such issues is of crucial importance to civic engagement in modern societies, but often relies on complex multivariate data whose interpretation and development requires knowledge that is usually not taught in regular math and statistics education, let alone in politics or social studies. My plan is to present core ideas and concepts of civic statistics illustrated with examples. Iddo Gal, who is also part of ProCivicStat, in his keynote CIVEEST presentation will explore in depth the role of contexts and modelling for statistical literacy and citizens' engagement in civic statistics.

Our analyses from the project ProCivicStat suggest that civic statistics competence has special features, such as the understanding of multivariate, dynamic (that is, rapidly changing) phenomena in meaningful social contexts and the understanding of rich texts and visualizations. Based on these analyses, we have developed a model in ProCivicStat that describes eleven knowledge bases and processes needed to understand, evaluate, and reason with statistics about social phenomena. This paper aims to give a brief introduction to the characteristics and competence requirements of civic statistics, to present typical examples and to point out curricular implications. For the interested reader, extensive resources consisting of teaching materials, datasets, conceptual representations and much more are freely available via the website http://iase-web.org/islp/pcs.

2.   What is civic statistics?

As early as 1792, Condorcet (1994), a French mathematician and philosopher during the time of the French Revolution, argued for the importance of informing citizens about governance and presenting evidence about the state of society in order to raise awareness of injustices and structural social inequalities. He believed in savoir liberateur - knowledge that would enable people to free themselves from social oppression.

In our modern mass societies, knowledge and skills to adequately argue with data is an important prerequisite for the functioning of democracy. In a society that that aims to keep up with promises of equity and fairness to all its citizens, questions, e.g. if women, minorities or people with disabilities are disadvantaged in their career options or access to education, can only be largely assessed on a quantitative level, because each individual case may be unique. In an increasingly complex world, citizens' contribution is an important resource for national and local policymakers. Recent government initiatives (see for example in the US https://data.gov, in the UK https://data.gov.uk, or in Germany https://govdata.de) explicitly mention political goals to encourage citizen participation, in particular for the promotion of the democratic process by giving citizens access to data that can stimulate debate and influence policymaking. In order to base the public debate on facts rather than emotions and promote evidence-informed policy making, statistics education must encompass two areas neglected in current secondary and tertiary education: understanding multivariate phenomena, and thinking and learning from complex data (Engel 2016). In addition to basic knowledge, civic

statistics requires special statistical knowledge and is located at the intersection of statistics, social sciences (social studies, politics, etc.) and education (on the basis of its educational mission for citizen participation in democratic processes) (see Fig. 1).
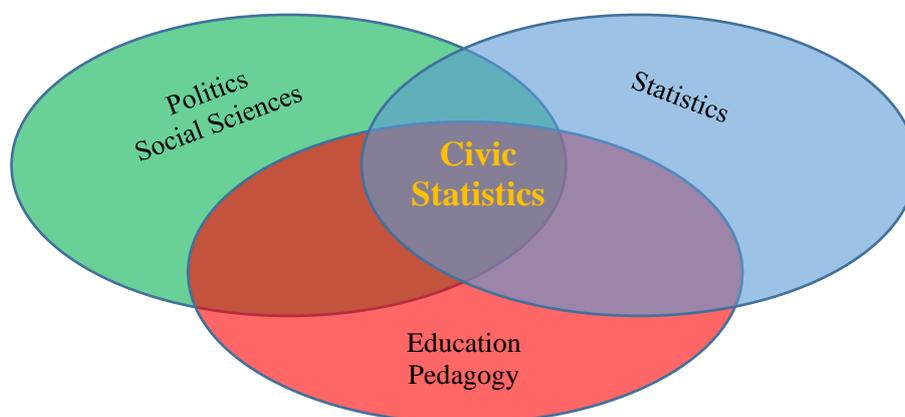

Figure 1. Civic statistics at the intersection of statistics, social and educational science

While statistical literacy as a basic competence aims at the general ability (see, for example, Gal 2002; Wallman 1993) to assess statistical data from all kinds of contexts, civic statistics focuses on issues of social relevance for society. Already earlier approaches (Fischer 1984, Borovcnik & Ossimitz 1987) have linked exploratory data analysis with goals of political education and consciousness raising. López-Martín, Batanero & Arteaga (2016) propose investigating life expectancy and gender inequality data to increase interest in social data and data analysis among prospective secondary teachers in Spain. Lesser (2007) describes statistics as grammar for social justice and discusses how to treat a number of statistical concepts in the context of social justice issues. The ProCivicStat project developed a theoretical framework for civic statistics, identified specific skills for understanding statistics about society, and developed numerous concrete, open-source teaching materials that offer many opportunities by accessing existing records and powerful, educationally designed tools to provide more in-depth data analysis than previous approaches.

Statistics about social phenomena often have special characteristics that are different from the contents of other statistics. These include:

(i) Multivariate Phenomena

Data about social phenomena usually do not stand in isolation; their description and understanding involves several variables that are correlated, interacting with each other, or having nonlinear relationships among themselves. This point was emphasized by different authors in different ways. The American Statistical Association Guidelines for Assessment and Instruction in Statistics Education (Carver et al., 2016), for example, corroborate the fact that phenomena in the modern world co-vary, i.e. a phenomenon is influenced or caused by other variables or factors and interacts with or otherwise influences other phenomena making it challenging to describe and understand what depends on what. Civic statistics data (as opposed to data from experimental studies) come mostly from observational studies or surveys, which is why a reliable identification of cause-and-effect relationships is usually very difficult to establish.

(ii) Aggregated data

Statistics about society often include data grouped in several ways - e.g. due to nominal variables (men vs. women) or ranks (e.g. level of educational qualifications). Some data are reported for multiple subgroups at different levels of aggregation. Possible conclusions may be influenced by the level of aggregation, i.e. patterns in the data may change when data within subgroups are aggregated or disaggregated.

(iii) Dynamic Data

Official statistics are often based on data collected regularly (e.g., every month, quarter, year) or on a comparative basis (e.g., in several countries), sometimes from different sources and with different methodologies and baselines. Data are often reported as a trend over time and can be updated as new data become available or old data are re-evaluated.

(iv) Rich texts

Statistics about society are made available to the public through texts published by official statistics offices (such as press releases, short reports) or articles in print media or digital media, often consisting of a combination of graphic, tabular and textual elements. Text is a primary medium for communicating statistics, and the public needs to understand various ways of writing, e.g. the formalized language that is often used in official reports, journalistic texts and more.

(v) Innovative visualizations:

Civic statistics are often associated with rich, novel data visualizations. Users need abilities to understand data presented in a novel way, and they need skills to access and manipulate computer-mediated illustrations.

Not all of these features are present in all civic statistics. However, they are unique in their combination. Below are two examples that illustrate how basic features of civic statistics manifest themselves in different types of messages.

Example: risk of poverty

Figure 2 is an excerpt from a one-page press release of EUROPA PRESS, displaying data collected by INE, the National Statistics Office of Spain. It shows some of the characteristics of civic statistics listed above. Poverty or social exclusion is a multivariate phenomenon as several factors influence it. It is summarized in the AROPE index (At Risk Of Poverty or social Exclusion). The extent of poverty cannot be described or understood without including additional variables such as the age or number of children in a household. The monthly surveys of the poverty rate, always with reference to the previous month or the year, illustrate the dynamic nature of the data, displayed as time series in the graph. While the text briefly explains what at risk of poverty and social exclusion mean, the term material deprivation and social exclusion are not clearly defined in the release. Further information can be obtained from the website of the National Statistical Office: Social Exclusion exists if, according to self-assessment, 4 out of 9 criteria are met. But who selected the criteria, how valid are they and how meaningful are self-assessments? In addition, the place of residence and the associated cost of living also have a major impact on a possible risk of poverty or social exclusion. A detailed description by means of disaggregated data - who is particularly at risk from poverty - would be more revealing in a next step than a global statement of

26.6% of the total population. Further, the text does not show how the different percentages for at risk of poverty (21,6%), severe material deprivation (5,1%), low employment (12,8%) and social exclusion relate to each other. Also, the precision provided in the graph for the AROPE index for 2017 given with up to 10 decimals (!) suggests a level of precision that is far beyond any reason.

Evolución de la tasa AROPE de probreza - estadística

El 26,6 por ciento de la población residente en España estaba en riesgo de pobreza o exclusión social en 2017, una variación de -1,3 puntos porcentuales comparado con el año anterior, según la **Encuesta de Condiciones de Vida del Instituto Nacional de Estadística.**

Se trata del **indicador europeo AROPE**, que mide tanto el riesgo de pobreza (un 21,6 de españoles se encuentra en esta situación) como la carencia material severa (5,1% de españoles) y la baja intensidad en el empleo (12,8% de españoles) y lo cruza con nueve ítems sobre la capacidad de maniobra de los hogares, como su posibilidad de afrontar imprevistos o de caldear la casa en invierno, para estimar el volumen de ciudadanos que están en riesgo.

El siguiente gráfico muestra cómo ha evolucionado la pobreza desde 2008.

Población en situación de pobreza y exclusión social o en riesgo de sufrirlo

Fecha	2017
Tasa AROPE	26,6000003815
Dato	Definitivo

■ Tasa AROPE   **Fuente:** INE, www.epdata.es

Figure 2. Press release from Europa Press with data from the National Statistics Office of Spain

Example: literacy and reproduction rate

Is there a correlation between the percentage of women per country who can read and write and the average number of children they give birth to? Figure 3 shows a Gapminder diagram of the two variables literacy rate and mean number of babies per woman, aggregated at country level, for Iran, Peru and Thailand. Gapminder is a foundation founded by the Swedish physician and public health expert Hans Rosling, who pioneered in the visualization of complex data. Note that (the static) Figure 3 represents four different variables: Each bubble represents a country in one of the years between 1976 and 2015. The central variables are the percentage of adult women (15 years and over) who can read and write (horizontal axis) and the average rate of children giving birth to women in the respective country (vertical axis); the size of the bubbles is proportional to the variable "number of physicians per 1000 inhabitants".

Finally, the colour of the dots represents the prevailing religion of the country (Muslim, Christian or Eastern religion). The trend is always the same, regardless of where the women are geographically at home and what religion the country is shaped by: the more educated women are, the lower the average number of children they give birth. A fifth variable can be dynamically visualized: at the press of a button, the development of the variables represented are animated over the period from 1976 to 2015. To enable, this requires a visit to the website www.gapminder.org.
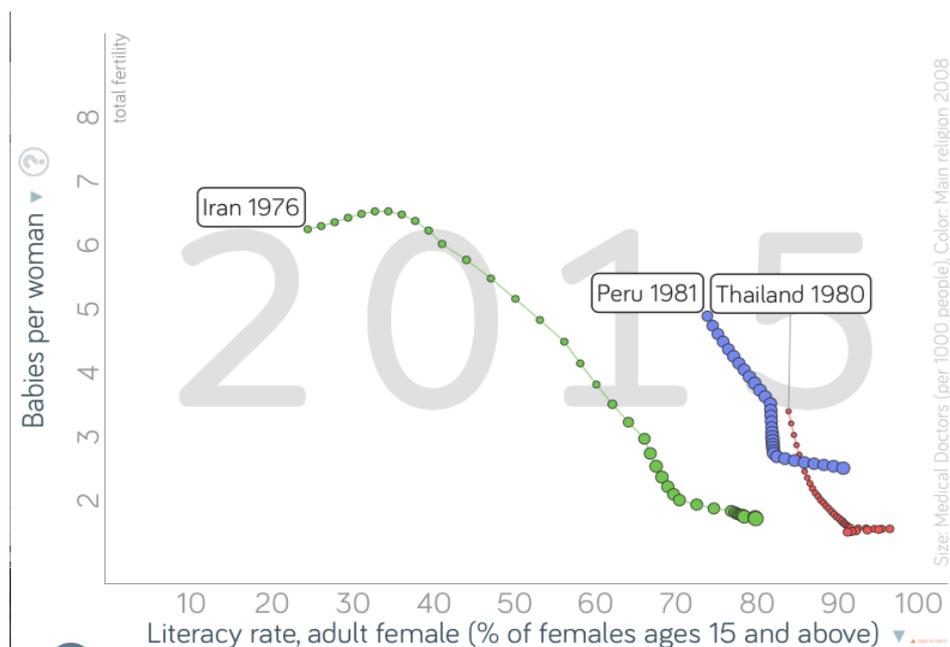
Figure. 3. Graphical representation of the variables literacy rate of women and average number of babies per woman[2] (source: free material from www.gapminder.org).

3.   A conceptual framework for civic statistical competences

The above examples illustrate that a competent handling of socially relevant data requires an interdisciplinary perspective, contains special knowledge elements and basic skills and also includes non-cognitive dispositional elements.

Civic statistics is not simply about mastery of techniques. It involves habits of mind – a willingness to engage with data – and enabling processes – the ability to reason, search and communicate with numerical evidence. Nicholson, Gal and Ridgway (2018) describe eleven facets of statistical literacy that underpin the ability to engage with social issues – i.e. civic statistics. The core facet is *meaning for social policy*; every other aspect of statistical literacy contributes to this facet. At the centre, as the starting and end point of any involvement with civic statistics, is the critical evaluation and reflection of a socially significant topic. Obviously, general and specific statistical knowledge, e.g. to representations and models and the methodology of the data collection is just as necessary as contextual social or socio-economic knowledge. Iddo

---

[2] Each point represents one data point per year for the three countries shown. The point size is proportional to the number of physicians per thousand inhabitants, while the colour is the predominant religion of the country

Gal in his CIVEEST keynote targets and deepens the role of contextual knowledge and modelling for understanding statistics about socio-economic issues and civic statistics. In order to be able to activate or acquire existing special knowledge and interest in civic-statistical topics, certain skills in data recherché and in the preparation and cleaning of data with suitable software and hardware may still be required. In addition, basic numerical abilities (e.g., understanding of percentages, a reasonable notions of numerical magnitudes) are as basic as appropriate reading and text comprehension, as well as communication skills to adequately communicate elaborated positions and conclusions. Nicholson et al. identify three groups or dimensions of facets.

- *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
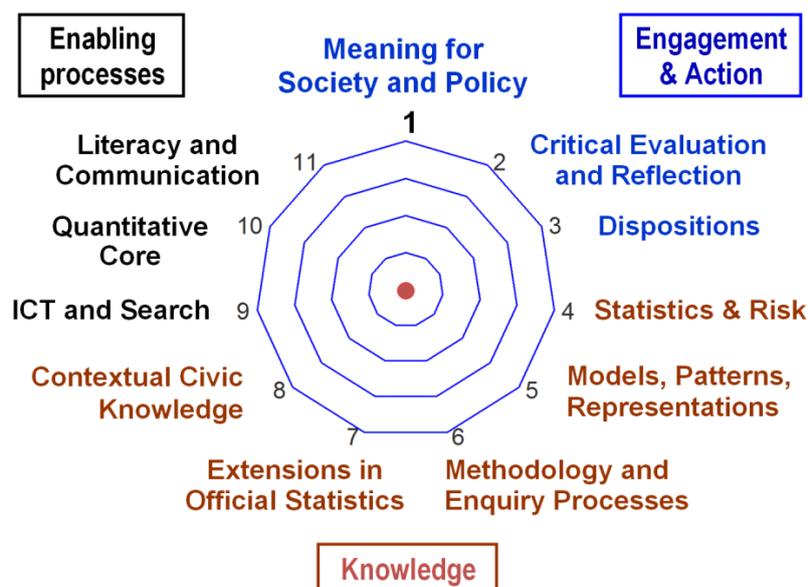
Figure 4. A conceptual framework for civic statistics

Facet 1: Readiness for social engagement,

This facet is at the heart of civic statistics – students and citizens need to be able to gain insight into what can and should be done to discuss and address some policy issue, grounded in evidence. In concrete situations, decision-making involves weighing existing evidence with perceptions of risk and probabilities, costs and benefits, expected value, and subjective utilities.

Facet 2: Critical evaluation & reflection

'Post-truth' and 'alternative facts' are attacks on the core of informed decision-making. Even when using data from well-authenticated sources, there is a need for critical

evaluation and reflection. Critical evaluation and reflection should be habits of mind – questions one triggers routinely without much conscious effort. For example:

- What is the story being told – whose story, and why are they telling it?
- Has the problem been identified appropriately?
- What evidence is being presented? From a credible source?
- Have appropriate statistical models been chosen?
- What else could be going on – a confounding variable?
- Are the conclusions consistent with the evidence?

Facet 3: Dispositions

Attitudes reflect a complex web of values, motivations, beliefs and attitudes. They have a social dimension, such as the willingness to share interpretations with others. They have an ethical dimension ("Why should I care?"). They have components that influence one's commitment both positively and negatively (e.g. self-efficacy, trust). Ignoring evidence based on beliefs, uncritical acceptance of new information or the belief that social phenomena can only be understood by experts are all symptoms of problematic attitudes.

Facet 4: Statistics and risk

This facet contains much of what is commonly taught in introductory university statistics courses: sampling, populations, and representativeness; variability; description and comparison of distributions; association and correlation; regression; nonlinearity; signal and noise; interaction; Bayesian reasoning. Civic statistics also requires an understanding of some of the ideas around Big Data - such as familiarity with a variety of data sources and associated techniques of analysis, especially those used for detecting patterns.

Facet 5: Models and representations

In modelling complex social phenomena, qualitatively different mathematical models can often be used to model one and the same phenomenon. For example, an economist and a sociologist may use very different methods to study "poverty" and have very different theories for defining and studying poverty. Civic statistics requires the ability to identify and understand the use of models as well as the basic assumptions underlying a model.

Representation is a core competence for understanding phenomena. Civic statistics requires familiarity with sophisticated representations and visualizations, including those that are dynamic and interactive. Recently, there has been an explosion in the use of technology-based representations and visualizations. Civic statistics requires the ability to understand and critically evaluate innovative representations.

Facet 6: Methodology and research process

Civic statistics requires an understanding of the strengths and weaknesses of different research methods and some procedural skills. Quantitative methods include: survey research (survey types, sampling methods), observational studies, quasi-experiments. Topics include: bias, variability, randomization. An understanding of ethical issues related to data production and the use of different research methods is as much an

integral part of civic statistics as the need to be aware of issues of confidentiality and the protection of citizens' identity.

Facet 7: Extensions in the area of official statistics

Official statistics producers (Statistical Offices, EuroStat, UN etc.) are an important source of data-based evidence regarding socially relevant issues. Many of the key ideas they use receive little attention in traditional statistics courses such as: survey design (and related issues such as non-response or response bias); Measurement problems (reliability and validity, metadata definitions); Operationalization of variables as well as definition and meaning of indices.

Facet 8: Contextual social knowledge

Statistics is about modelling; but to model, one must have at least a basic understanding of the phenomena that are to be modelled. Contextual societal knowledge includes, for example: general knowledge such as size of population, size of gross national product, national debt and resources; demographics; human rights and vulnerability of minorities, history and geography; regional and geopolitics. An advantage of contextual societal knowledge is that one can search for alternative explanations and analyses based on the knowledge of plausible covariates.

Facet 9: Information and communication technology and information research

Many official statistical providers (e.g. Federal Statistical Office, EuroStat, UN) make data publicly available - but the access and direct handling of these data sets often requires a considerable amount of specialized knowledge and information technology skills. Data have to be downloaded and imported into suitable software. Data may need to be cleaned, tidied, transformed, aggregated or disaggregated. Fortunately, nowadays more and more use-friendly software, some with a distinct educational design, are freely available such as CODAP (Finzer 2018).

Facet 10: Quantitative core

Quantitative skills are the foundation of all aspects of Statistical Literacy. The components include an understanding of numbers, ratios, percentages, rates, and fractions. In civic statistics, very large numbers are common and seemingly large resources can actually be small in context. It is easy to find examples where an author has deliberately chosen to report (accurate) data that are misleading - for example, when reporting a percentage increase, where the rise in absolute numbers is very small.

Facet 11: Reading and text comprehension and communication

In many cases, information is also presented as a text, diagram or image in printed form. The text is often very dense. It is then an essential ability to understand the text, to associate it with the displayed representations and to absorb the overall meaning of an article. Likewise, important is the skill to communicate the conclusions of an analysis in a comprehensible and transparent way.

4.   Examples, teaching and learning materials

The competences presented in Section 3 give a comprehensive view of the manifold requirements in the context of civic statistics. They set ambitious goals, even for university education. Part of these facets can already be addressed in high school after some adaption and didactic reduction. In ProCivicStat, we created 41 units of

instruction in various languages (English, German, Portuguese and Hungarian) for teaching in schools and colleges, and for the most part they have also been tested in seminars. All the materials have a distinct socially relevant topic or question as a starting point and are designed for a range of 3 to 8 hours of class time. The wide range of topics covered includes issues such as: *How is the earnings of employees and workers distributed in European countries and worldwide? Do female employees receive less salary than their male counterparts? Are people happier in some countries and regions of the world than elsewhere? Has the crime rate in Germany increased with the influx of refugees? Is humanity getting fatter? Do dark-skinned players in European football get more red cards than their fair-skinned teammates? Under what conditions is fine dust pollution particularly high?*

Preceded by a context-related introduction and additional references to provide more detailed context information, these materials present comprehensive datasets with tasks for data analysis, using didactically designed and mostly freely available digital tools such as Fathom, CODAP (Finzer 2018), iNZight, Tableau, Gapminder or apps available on the web. CODAP (Common Online Data Analysis Platform, available at http://codap.concord.org) is a data management and visualization environment similar to Fathom. CODAP is web-based, therefore it can be used without the installation of local software and enables the provision of electronic worksheets through the web.

In order to facilitate the selection of appropriate teaching materials, the special platform *CivicStatMap* (available at http://iase-web.org/islp/pcs, see Figure 5) was developed by ProCivicStat. After selecting a topic of interest (e.g. Migration, Climate and Environment, Income Inequalities, Education etc.), a statistical content (e.g. Mean, Boxplots, p-Value, Correlation etc.) and a preferred digital tool for analysis and visualization (e.g., CODAP, Fathom, Gapminder, iNZight, Tableau, R) you are directly referred to concrete teaching materials including datasets for use in the classroom. Working on these materials connects contextual knowledge with basic mathematical and statistical knowledge and promotes statistical skills that are rarely addressed in traditional classrooms, such as multivariate thinking, the search for confounding variables and interactions, Simpsons paradox, comparing distributions, the exploration of functional relationships between variables and the use of different representations and visualizations. Engaging with these tasks challenges and stimulates critical thinking: How are variables defined? How are constructs operationalized (e.g. risk of poverty or unemployment)? In what way, for which purpose and by whom were the data collected?

Figure 5. CivicStatMap, freely available through the website https://iase-web.org/islp/pcs

In the following, we present a specific teaching and learning material used in classes for students preparing to be secondary mathematics teachers. This to illustrate how aspects of civic statistics can be addressed in secondary school. A detailed discussion of this example with teacher tips and a student worksheet can be downloaded using *CivicStatMap* with the following filter settings (see Figure 5): Language: English, Theme: Income Inequalities, Tools: CODAP,iNZight or iNZight Lite; Material Type: Students.

Example: Deepening income inequality worldwide

A balance in earning potentials and a modest prosperity gap are important safeguards for social peace in our societies. On the other hand, overly egalitarian societies often lack incentives to make an effort or be willing to take risks for economic success. Differences between rich and poor provide fuel for social tensions everywhere in the world. The World Economic Forum in Davos, a meeting not known for social romanticism or revolutionary rhetoric, declared in 2017: "*Inequality is one of the key challenges of our time. In particular, income inequality is one of the most visible aspects of a broader and more complex issue, one that creates inequality of opportunity and extends to gender, ethnicity, disability, and age*."

The worksheet "Deepening Income Inequality Worldwide" (available through CivicStatMap) assumes some familiarity with location parameters (arithmetic mean, median, percentiles), boxplots, scatter plots, time series, and the notion of distribution and can be dealt with in about 4 -5 hours of class time. In terms of context, it deals with studies of economic inequality around the world and particularly motivates students interested in issues of economic justice and human rights.

The student worksheet opens with the above citation from the World Economic Forum and additional information highlighting the urgency of the issue of income distribution. It provides additional references (including resources on the Internet) to introduce and sensitize students for the socio-economic explosiveness of the subject. As an introduction to the topic, students are asked, "*Many think that a reasonably egalitarian distribution of income is a very worthwhile societal goal, while others point out that*

achieving an equal distribution undermines incentives for acquiring higher average standards of living. How do YOU position yourself on this issue?"

The section following explains how to measure inequality and how statistics agencies collect the necessary data from household surveys. Inequality is usually summarized in an index, of which there are quite a few. A relatively simple but informative way of measuring inequality is to look at the share of the total income that different groups of society receive. For example, households from the poorest to the richest can be grouped into ten parts (deciles) (the 10% lowest earners, the next group, which is between 10% and 20%, and so on). Similarly, consider the total of all incomes. The income share of the 10% lowest income earners defines the proportion of the first decile, and so on. Income shares that belong to a certain decile can then be examined within and compared between countries. Analogously to deciles, the data can be divided into five groups (quintiles). Another, but more advanced concept for measure inequality is the Gini index.

Subsequently, the data set underlying these materials is described on the worksheet (Source: United Nation University UNU-WIDER, World Income Inequality Database, WIID4). In the version provided here, it includes income data for 42 countries in Europe, North and South America over the period from 1980 to 2015 and comprises 26 variables: year, country, region, sub-region, population size, arithmetic mean and median of income, 10 deciles, 5 quintiles, Gini-index, EU (yes/ no), OECD (yes/ no), income group (upper, middle, low). The worksheet then provides some concrete questions for a guided exploration of the data before students are asked for more "free", unguided data explorations.

The link https://tinyurl.com/desigual-salario leads to the CODAP based electronic worksheet containing the data. As a first approximation of the data, it may be useful to look at the distribution of individual deciles over all the years (see Figure 6 for d10). We notice several cases (=country, year) where the best earning 10% make more than half of all the total income (Colombia in 1997 and Belize between 1989 to 1997).
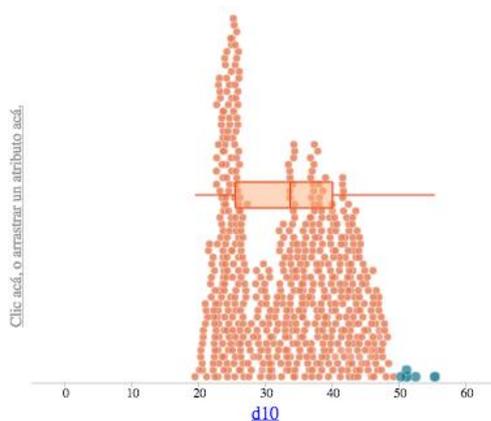

Figure 6. Income shares of the 10% best paid in 42 countries of Europe and the Americas between 1980 and 2015 with superimposed boxplot.

A scatter plot of d10 versus d1 (Figure 7) shows a clear correlation: the larger d10 (the share of income of the 10% best earner) the lower d1 (the share of income of the 10% lowest earner). Why would it make sense to model the relationship between d10 and d1 with a falling line? If a straight line proves to be a suitable model, how can we explain

deviations from the exact linear relationship? What significance does the slope of this line have?
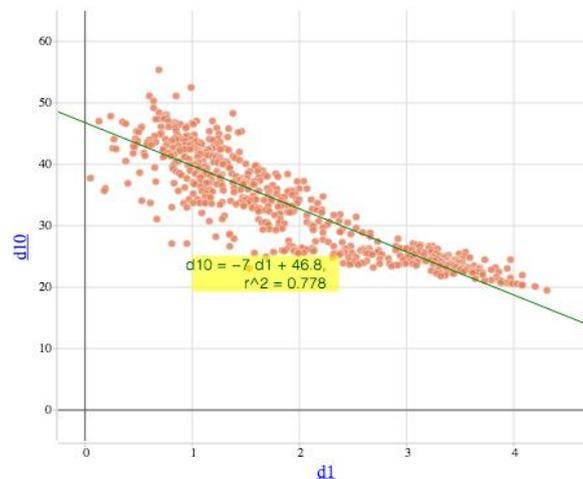

Figure 7. Scatter plot D10 versus D1 with the least squares line

As a possible measure of inequality, the quotient d10/d1 can be used, which indicates by which factor the proportion of the 10% Best earners is higher than the proportion of the 10% with the lowest income. CODAP makes it easy to define a new variable d10/d1. Why is d10/d1 a plausible measure of inequality? What are the weaknesses of this measure, and would q5/q1 be a better measure? For advanced students, the Gini index can be considered a measure of inequality and compared to the variable d10/d1. In civic statistics, it is recommendable to look at the data from different angles and apply different measures to investigate a phenomenon.

How reliable is d10/d1 as a measure of inequality? One way to assess is to compare d10/d1 with other approaches to measure inequality like the Gini index. A Gini index of zero expresses perfect equality, where all values are the same (for example, where everyone has the same income). A Gini coefficient of 1 (or 100%) expresses maximal inequality among values (e.g., one person has all the income and others have nothing, the Gini coefficient will be very nearly one).

Figure 8 shows a strong curvilinear relationship between the Gini index and d10/d1.
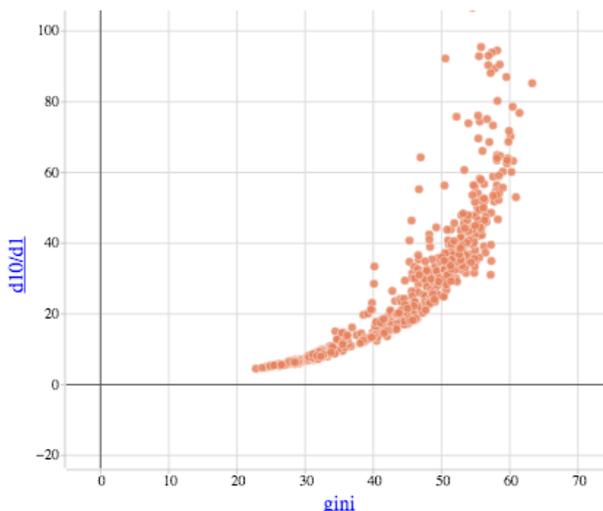
Figure 8. Two measures of income inequality: d10/d1 and the Gini-Index

To explore similarities of countries with similar d10/d1 scores, Figure 9 shows a scatter plot with the median earnings versus d10/d1. The most important message of Figure 9 is that a higher median income is not associated with increased income inequality.

It can be seen that, above all, the countries of Eastern Europe are characterized by low median earnings and high inequality, while countries in Northern Europe are characterized by high incomes and less inequality. Countries in Western Europe have a high median income (with Switzerland and Luxemburg on the top) and a relatively low d10/d1 score. Canada and the US have similarly high median earnings, but with higher inequality. Countries in Southern Europe and South America are quite scattered, while central American countries have low median income and extreme income inequality. Certainly, the phenomena of mean income and inequality and is highly multivariate and many other factors are of strong influence. For example, the wide dispersion of data for countries in Southern Europe in Figure 9 can be explained through recent history. Countries formerly belonging to the Eastern bloc (Serbia, Kosovo, Macedonia etc.) are more like countries in South America with regard to median income and inequality, while Italy and Spain are close to Western European countries.
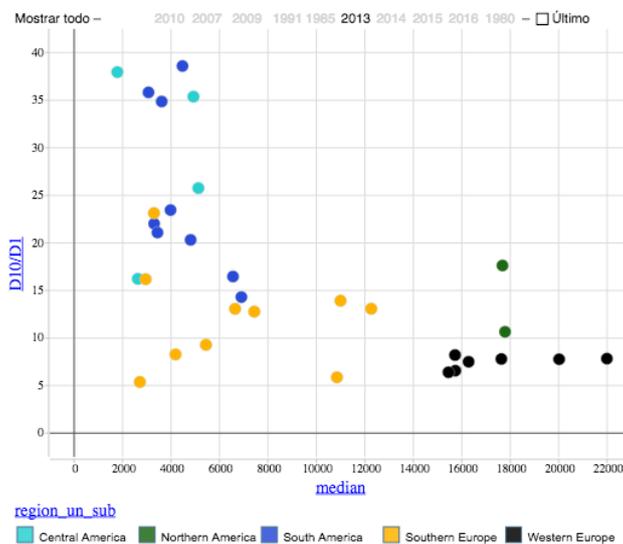

Figure 9. Scatter plot of D10/D1 versus median earning for 2013. The colouring corresponds to different world regions.

An explosive socio-political question is whether inequality in earnings will increase over time, i.e., if the rich get richer and the poor get poorer. Figure 10 shows a diagram of the Gini index time series for Brazil, Mexico, Spain and the United States. Despite clearly discernible fluctuations, one can discern that Brazil had extreme income inequalities in particular in the 80's and 90's with notable decline since 2000 while the Gini index in the United States and Spain is slowly increasing. While there are notable differences in income inequality between the four countries, a trend of getting more alike is noticeable between these countries.
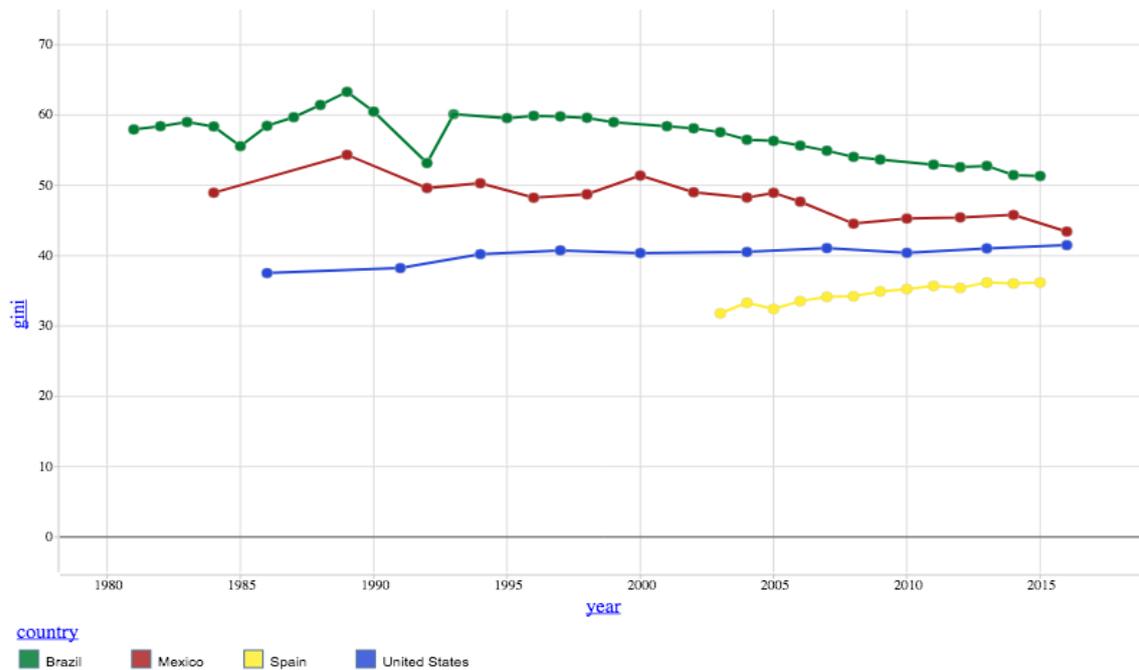

Figure 10. Time series of the variable Gini Index for Brazil, Mexico, Spain and the United States

These exemplary analyses show the potential of the worksheet "Deepening income inequality worldwide". Many of the competences presented in Section 3 are addressed by this example. The radar plot representing the facets discussed in Section 3 allows an evaluation of the extent to which specific facets of civic statistics are addressed (from 0 = not at all to 8 = very strong). The explosiveness of the topic demands engagement with the issue, on the basis of socio-economic context knowledge different graphics and analyses are created, which require certain skills in the use of handling the software, different measures of inequality need to be modelled and critically reflected, leading to an evaluation of this task as in Figure 11.
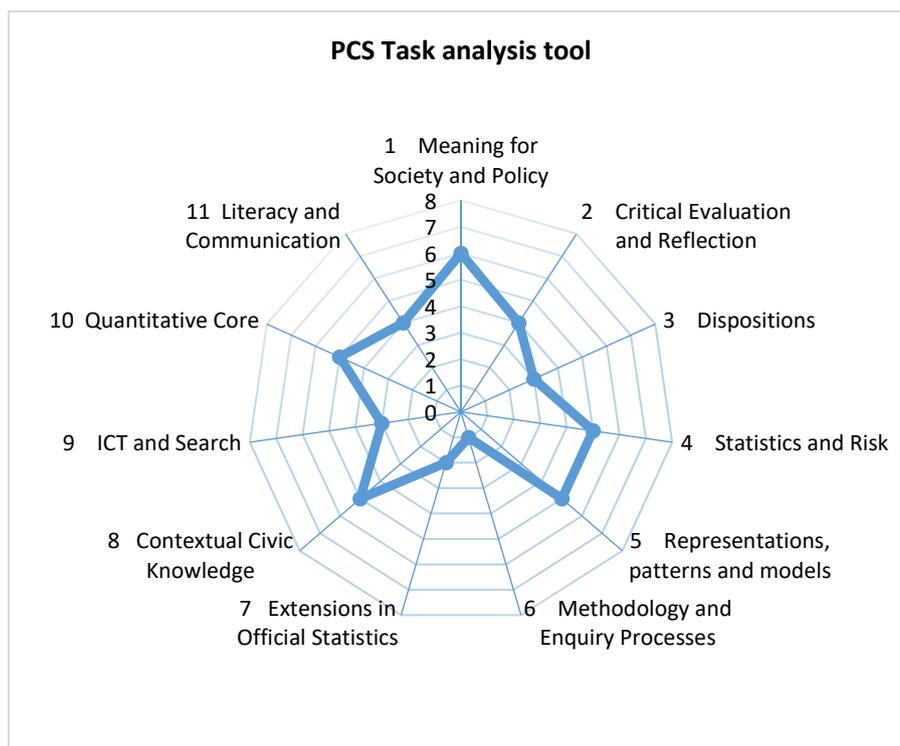
PCS Task analysis tool

1 Meaning for Society and Policy
2 Critical Evaluation and Reflection
3 Dispositions
4 Statistics and Risk
5 Representations, patterns and models
6 Methodology and Enquiry Processes
7 Extensions in Official Statistics
8 Contextual Civic Knowledge
9 ICT and Search
10 Quantitative Core
11 Literacy and Communication

Figure 11. Radar plot for the task "income inequality worldwide"

5.   Curricular implications

The competences and content described here cannot be delegated to a single discipline or subject in school. Making sense of data is also part of teaching in geography, history, civics and economics as well as in biology and physics. Mathematics provides important analytical know-how and conveys certain methods and representations of exploratory data analysis such as scatterplots or boxplots, introduces concepts such as distribution, correlation, regression, explanatory confounding variables etc. and discusses how new insights can be gained by changing various numerical and graphical forms of representation. A possible integrative function for the school curriculum could be provided by computer science education in the context of a data science curriculum.

Central to civic statistics are issues of public interest concerning the social and economic well-being of all citizens and the functioning of democracy. This puts high demands on the instructors: Civics and geography teachers need basic knowledge in descriptive statistics and their pedagogy. Mathematics teachers should be able to interpret the results of statistical representations and analyses against the background of the social context and to facilitate discussion processes on controversial issues in the classroom. Interdisciplinary teaching should emphasize interest in burning issues related to the social and economic well-being and the realization of civil rights. Students need to experience the role of evidence and data in understanding these issues. To do this, it will be important to use authentic data and textual documents that match the specific nature of data (see Section 2) on social phenomena. Fortunately, there are pedagogically designed digital tools such as Fathom, CODAP or iNZight as well as visualizations on the Internet such as Gapminders or animations on the official statistics web pages, which enable students to explore multivariate social data in an innovative way.

Our analyses in the project ProCivicStat point to shortcomings and problems of the current education in statistics at schools and colleges. To help young people to competently and informedly engage in public debates on societal issues and engage them in resolving burning issues, we consider curricular changes in the way statistics is taught in schools to be necessary. Data analysis, interpretation of graphics and (dynamic) visualizations are not just related to mathematical knowledge, but are part of the knowledge gained in many subjects and should therefore be part of a broader scope of disciplines. Mathematics instruction should not be confined to the teaching of techniques and formal terms with little relevance to content, but may in civic-statistical topics illustrate the relevance of mathematical analysis and thus empower young people to form their own evidence-based position and engage in public decision-making processes.

References

Borovcnik, M., & Ossimitz, G. (1987). *Materialien zur Beschreibenden Statistik und Explorativen Datenanalyse*. Wien, Austria: Hölder-Pichler-Tempsky.

Carver, R., Everson, M., Gabrosek, J., Rowel, G.H., Horton, N., Lock, R., Mocko, M., Rossman, A., Homes, G., Velleman, P., Witmer, J. y Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report*. Reston, VA: American Statistical Association. Available from: http://www.amstat.org/education/gaise

De Condorcet, A. (1994). *Foundations of social choice and political theory*. Brookfield, VT, USA: Edward Elgar (Original published in 1792).

Engel, J. (Ed.) (2016). *Promoting understanding of statistics about society. Proceedings of the IASE Roundtable Conference*. Berlin 2016 http://www.iase-web.org/Conference_Proceedings.php?p=Promoting_Understanding_of_Statistics_about_Society_2016. (Accessed 18.11.2018).

Finzer, W. (2018). Common online data analysis platform. Emeryville, CA. Available from: https://codap.concord.org.

Fischer, R. (1984). Offene Mathematik und Visualisierung, *Mathematica Didactica 7*, 139-160.

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review*, *70*(1), S. 1-51.

Lesser, L. (2007). Critical values and transforming data: Teaching statistics with social justice. *Journal of Statistics Education*, *Volume 15*(1), 1-21.

López-Martín, M., Batanero, C., & Arteaga, P. (2016). Using united nation data in the training of teachers to teach statistics. In J. Engel (Ed.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE),* July 2016, Berlin, Germany.

Nicholson, J., Gal, I., & Ridgway, J. (2018). Understanding civic statistics: A conceptual framework and its educational applications. A product of the ProCivicStat Project. Available from: http://IASE-web.org/islp/pcs.

Ridgway, J. (2015). Implications of the data revolution for statistics education. *International Statistical Review*. 84(3), https://doi.org/10.1111/insr.12110.

Wallman K. (1993). Enhancing statistical literacy: enriching our society. *Journal of the American Statistical Association, 88*(421), 1-8.