STATISTICS EDUCATION AND MONITORING PROGRESS TOWARDS CIVIL RIGHTS

Joachim Engel[1], Achim Schiller[1], Daniel Frischemeier[2] and Rolf Biehler[2]
[1]Ludwigsburg University of Education, Germany
[2]University of Paderborn, Germany
engel@ph-ludwigsburg.de

Monitoring societal progress towards respect of civil, economic, social and cultural human rights is mainly about analyzing trends in mass phenomena that may contradict the purposes and visions of an open society which promises equity and fairness to all its members. To assess these trends requires statistical knowledge and understanding of multivariate phenomena. For educators, these topics address issues beyond teaching technical skills for analyzing data and concern matters of value clarification, understanding and embracing the principles of human equality and dignity - topics that address the mind and the heart. Moreover, students experience statistical analyses playing a role in understanding pressing social and political issues of our time. Exemplified by data from the EU Agency for Fundamental Rights and by a large multivariate data set on racial bias in European football we discuss potentials and implications of taking this topic to the classroom.

INTRODUCTION

Knowledge and skills to reason adequately with data are an important prerequisite for the functioning of democracy in our mass societies. Active democrats need skills in reading statistics and charts as well as in interpreting and critically evaluating data. The questions as to whether women are disadvantaged in their careers, ethnic minorities such as Roma or people of color are discriminated against in the housing market or work place, or whether access to higher education is too strongly determined by socio-economic background – all have to be judged largely on a quantitative level whether a society keeps up with the promises of equity and fairness to everyone. In its charter the European Union commits its member states to *respect of human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail*. (Article 2, the Treaty on European Union).

Topics of democratic values, fairness and equity to all members of society provide ample opportunities to apply statistical knowledge and critical thinking to pressing real-world problems, on the basis of authentic data. Written by (mathematics) educators, this paper pursues a twofold pedagogical purpose: 1) to draw attention to the fact that monitoring progress with respect to fundamental rights requires - besides civil courage, critical thinking and commitment – also quantitative skills in understanding and analyzing data, and 2) to give students a strong experience that statistics matters, that statistical analyses play a role in understanding pressing social and political issues of our time.

Any serious discussion of the application of statistics in the area of civil rights has to address the issue of reliable and valid measurement and operationalization. How can one develop concrete measurable definitions for the concepts involved? How does one define deprivation of cultural rights, economic discrimination or restricted access to public services? Even commonly used concepts like the unemployment rate are far from trivial to define. The operationalization of a complex concept such as discrimination is not easy and involves philosophical, ethical and cultural aspects.

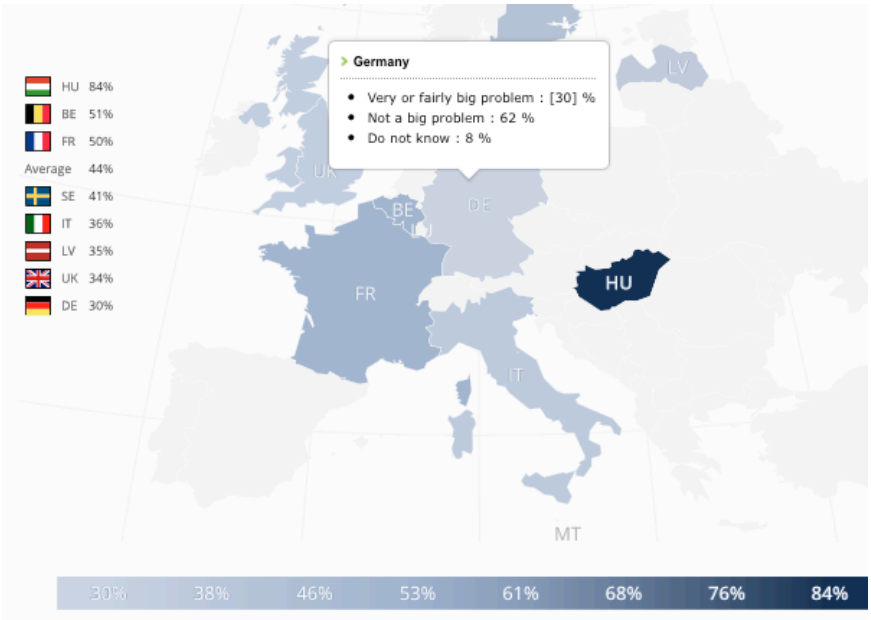
While addressing knowledge elements associated with statistical literacy (mathematical and statistical skills, socio-historical awareness and knowledge) the topic of civil rights and statistics foregrounds dispositional elements (Gal, 2002) requiring a critical stance with a focuses on beliefs and attitudes related to one's own values. By drawing attention to civil right topics implies that we address the responsibility of statistically literate citizens to get involved in ensuring and promoting

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. iase-web.org/Conference_Proceedings.php

human dignity, respect and progress of the human family. Human rights education is not just an academic affair but engages the heart as well as the mind (de Maio, 2008; Engel, 2013). It challenges students to ask what these rights actually mean to them personally and encourages them to translate caring into informed advocacy and action. Why are civil rights important and why should students even care about these issues? The questions involved touch on broad issues, such as how prejudice affects society and how the effects of prejudice, as detected in laboratory settings, show up in the real world.[1] This discussion can enrich the discourse in the classroom but presents particular challenges teachers in high school are usually not prepared for: Mathematics and statistics teachers are usually not trained in teaching social studies, and even less in moderating learning processes involving value clarification. Social science and ethics teachers, on the other hand, commonly are not prepared to include quantitative analyses in their teaching.

FUNDAMENTAL CITIZEN'S RIGHTS

The European Union Agency for Fundamental Rights (FRA, founded in 2007) has the task to provide independent, evidence-based advice on fundamental civil rights such as the rights to be free from discrimination or degrading treatment, the right to access healthcare services, the right to the protection of privacy etc. To fulfill its mission the agency collects "objective, reliable and comparable" information and data through surveys on particular themes across the EU. For example, the European Union Minorities and Discrimination Survey (EU-MIDIS) interviewed 23,500 people from selected immigrant and ethnic minority groups in all 27 EU Member States. The results allow the agency to provide useful and relevant indicators of discrimination, such as the proportion of victims who had experienced discrimination or racially-motivated crimes. This, in turn, revealed the levels of discrimination on different states felt by different ethnic minorities.



In your opinion, how big a problem, if at all, is antisemitism in the political life in your country today?

Figure1: Sample question and interactive survey map on antisemitism in Europe, from EU agency for fundamental rights

[1] To test one's own attitudes and beliefs could be a sensitive start for this teaching unit. The "project implicit", an international collaborative network of researchers that produced new ways of understanding attitudes, stereotypes and other hidden biases that influence perception, judgment and action, offers a short online test to explore one's own biases towards various social groups in society. This test is freely available under http://projectimplicit.net/index.html

From their websites[2] updated aggregated macro data are available on topics like violence against women, antisemitism, attitudes towards Roma, or discrimination of sexual minorities. Interactive displays of aggregated data allow the user to investigate issues and compare various European countries in further detail. The country detail visualization shows the values for all answer categories for the selected Member State, survey question and subset of respondents. The map visualization in Figure 1 presents an overview of the results of one question of a survey in 8 European countries on Jewish people's experiences and perceptions of discrimination, hate-motivated crime and antisemitism. The data explorer also allows the filtering of responses by age, gender and strength of Jewish identity of respondents.

Another survey investigated violence against women, asking 42,000 women in the EU about their experiences of physical, sexual and psychological violence, including domestic violence, since the age of 15. Questions were also asked about incidents of stalking, sexual harassment, and the role played by new technologies in women's experiences of abuse, see Figure 2 for an example result.

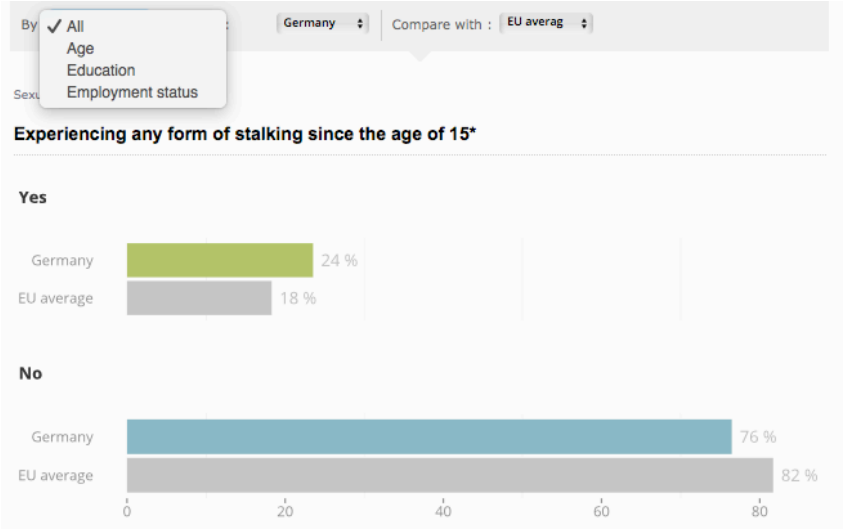


Figure 2: Comparing percentage of women in Germany and EU surveyed who reported being victim of stalking

Other data visualizations on civil rights within the EU provided by the Fundamental Rights Agency relate to the situation of Roma in various European countries and to attitudes towards sexual minorities. The Roma survey gives data on the socio-economic conditions, experiences of discrimination of Roma with a focus on respondents' situation with respect to employment, education, housing and health. Another group of displays visualizes data about discrimination of sexual minorities in the EU, their experiences of violence, verbal abuse or hate speech on the grounds of their sexual orientation or gender identity. The data explorer allows the filtering of responses according to a variety of covariates. For example, responses to the question *In your opinion, how widespread is offensive language about lesbian, gay, bisexual and/or transgender people by politicians in the country where you live?,* collected information from more than 93,000 LGBT people in the EU, can be explored country by country and conditioned on several covariates (see Figure 3).

[2] http://fra.europa.eu/en/publications-and-resources/data-and-maps/survey-data-explorers

Figure 3: Panel allowing investigations of responses for various subgroups of the LGBT survey

IS THERE RACIAL BIAS AMONG REFEREES IN EUROPEAN FOOTBALL?

The FRA platform is easy to use and very informative about the compliance of EU countries to civil right standards. The variables presented are operationalized through a questionnaire comprised of several items. By filtering and conditioning on subpopulations the impact of different variables on the target variable can be investigated in further detail. However, a detailed exploration of the data is limited by the fact that the data are aggregated macro data and (user-friendly) visualizations are restricted by the options the data explorer provides. In contrast, large micro data sets afford much more flexibility to explore complex issues such as discrimination or racism. For that purpose we introduce a large multivariate data set referring to potential racial bias in European football and discuss its didactical potential. The leading question is whether players with dark skin tone are more likely than light skinned players to receive red cards from referees. The decision to give a player a red card results in the ejection of the player from the game. Red cards are given for aggressive behavior such as violent tackle, a foul intended to deny an opponent a clear goal scoring opportunity, hitting or spitting on an opposing player, or threatening and abusive language. However, despite a standard set of rules and guidelines, referees are often faced with ambiguity (e.g., was that an intentional foul or was the player only going for the ball?) It is inherently a judgment call on the part of the referee as to whether a player's behavior merits a red card. The rich dataset contains demographic information with 19 variables about soccer players (n=1419) playing in the first male division of England, France, Germany, and Spain[3]. The original even further disaggregated data set was analyzed in a crowdsourced research project involving 29 research teams analyzing these data (Silverzahn et al., 2014). The data include referee calls, player demographics such as team position, height and weight, number of games played, wins, losses and ties and number of goals scored. It also includes a rating of players' skin color.

Any serious exploration of these data has to address the question of how the variables were operationalized and measured, what the relevant covariables are. To investigate potential racial bias we defined the target variable REDCARDSRATE as the ratio of the number of red cards a player received in his career divided by the number of games played. The variable SKINTONE was coded by two independent raters blind to the research question who, based on the player's photo, categorized players on a 5-point scale ranging from 1= very light skin to 5=very dark skin (r=0.92; rho=0.86). Obviously, there are relevant covariates that may influence a player's likelihood to receive a red card. A serious analysis has to take into account that, for example the influence of the variable POSITION. Defenders may receive many more cards than the average forward since their role of preventing the other teams from scoring may cause them to make more fouls. Other possibly relevant variables to consider are a player's physical condition (i.e., weight and height) which may be confounded with SKINTONE as darker players may be stronger or weaker and

[3] The dataset considered here is a condensed version of an even much larger dataset which contained data about all interactions of those players with all referees (N=3,147) they encountered in their professional career amounting to 146,028 dyads of players and referees. The data are publicly available from https://osf.io/47tnc/). For our purposes data were aggregated to a players level.

heavier players may be a bit slower in their motions and hence seem to have a higher chance of getting a red card. Another relevant variable to be controlled for may be the league COUNTRY. SKINTONE is not evenly distributed across league countries, and in some leagues the rules may be applied stricter than in others, a cause for potential bias due to a different distribution of skin tone and the different distribution of red cards within each league. As illustrated in Figure 4, player's skin color is not evenly distributed across positions (left display), nor across the European Leagues (right display);

However, as reasonable as it sounds at first, the inclusion of league country as a covariate may also be questionable given that the variable COUNTRY was only available at the time of data collection, but the variable REDCARDS refers to a players' total number of red cards throughout his whole career. Many players have changed the country league in which they played over the course of their careers. For similar reasons the position of a player may be confounded with its skin tone, thus a possible cause of bias if this variable is not taken into consideration.

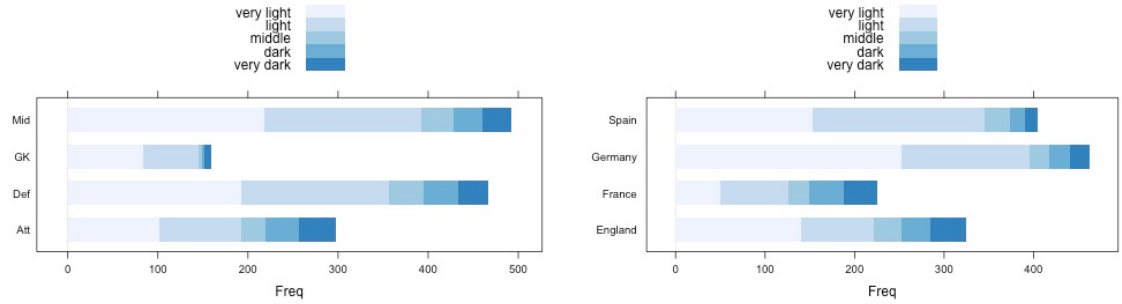


Figure 4: Skin color as a function of position and league

The relevance of these covariates is confirmed by a CART-style regression tree (Breiman et al., 1984), created with the R package `rpart` (Therneau, 1997). Classification and Regression trees are an exploratory methods for prediction and for discovering structure in multivariate data as usually encountered in social sciences: many covariates, mixed types variables, non-linear relationships, confounders, co-linearity etc. The tree models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the partitioning can be represented graphically as a decision tree. In a variety of medical, sociological and other investigations they give an illuminating way of looking at data. For data analysis tree-based methods are not recommended at the exclusion of other methods, but they often give an interesting insight into the predictive structure of the data. The binary tree in Figure 5 splits the sample into increasingly more homogeneous subsamples. Each node lists the corresponding sample size and average value for the corresponding (sub-)population, edges denote the optimal split to partition the sample recursively into increasingly more homogeneous subsamples. Optimal splits are obtained along the variables POSITION and COUNTRY, with HEIGHT also of some subordinate relevance. The fact that the variable SKINTONE does not appear in the tree does not imply that this variable is insignificant. It just highlights the importance of these other variables.

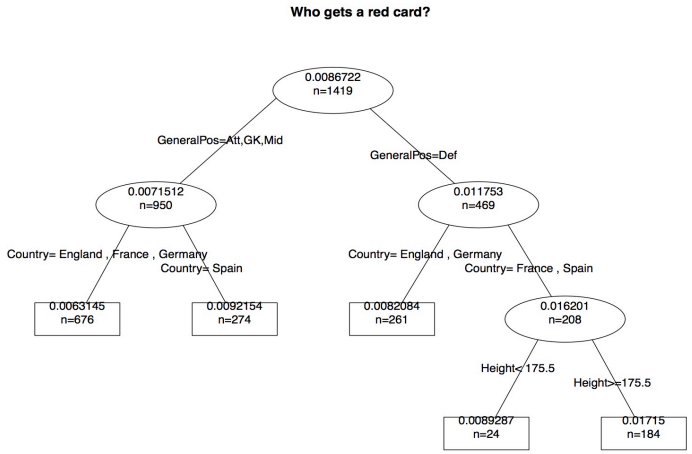


Figure 5: Regression tree illustrating the relevance of the variables POSITION, COUNTRY and HEIGHT for the target variable REDCARDSRATE.

To explore the impact of covariates on the target variable REDCARDSRATE, the advantages and challenges of various digital tools, from easy-to-use educational tools like Tinkerplots (Konold & Miller, 2011) or Fathom (Finzer, 2001) to powerful professional packages like R, are discussed by Frischemeier et al. (2016). For the purpose of visualizing multivariate social data the R package *Lattice* (Sarkar, 2008) is very helpful. As most covariates are nominal variables, boxplots and density plots for various subpopulations provide some insight and may guide the exploration of relationships between relevant variables. Figure 6 shows boxplots of distributions of the variable "REDCARDSRATE" which can easily be produced in TinkerPlots. The left plot in Figure 6 shows distributions of the variable "REDCARDSRATE" separated by dark and non-dark players for the German "Bundesliga", the right plot in Figure 6 shows the situation for the English "Premier League".

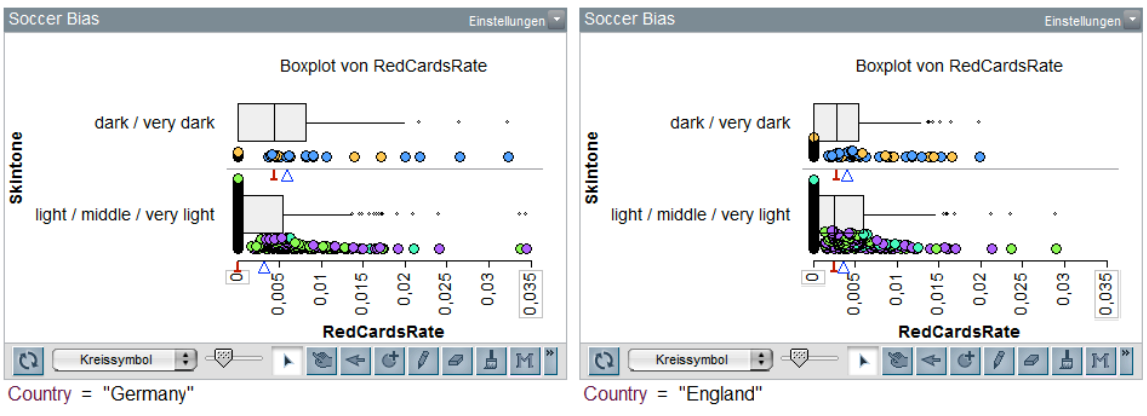


Fig 6: Boxplots of distributions of the variable "REDCARDSRATE", separately for dark and non-dark players and distinguished by league country "Germany" (left) and "England" (right)

IMPLICATIONS FOR PRACTICE

Exploring authentic data on civil rights topics incites discussion and reflection about central statistical issues like operationalization of relevant variables and their measurement, the choice of relevant covariates and the role of potential confounders. On a more personal level, the topic of racism sensitizes students to reflect about own potential biases.

The topic of civil rights and statistics involves issues of value clarification, understanding and embracing principles of human equality and dignity as well as skills for analyzing situations

and data in a socially sensitive context. This implies particular challenges for teachers whose preparation may have focused on technical skills and who are not trained in moderating It lets students experience statistical analyses play a role in understanding pressing social and political issues of our time.

Learning about civil rights is largely cognitive, including civil rights movements and civil rights history, documents, etc. Analyzing data and using statistics to explore these issues certainly belongs to the cognitive side. Statistical methods are an important addition to the more anecdotal reports and case studies, because they provide evidence-based insights into respect or neglect of universal rights. Statistical analysis can make a significant contribution to the professionalization of monitoring civil rights, thus be an important instrument to help our societies move forward to respect these rights. Carefully and sensibly introduced into the classroom, it is a strong enrichment for the statistics class demonstrating the importance of quantitative analysis for progress and human dignity.

EPILOG

The study by (Silverzahn & Uhlmann, 2015) showed a complex pattern of results. Of the 29 teams, a statistically significant correlation between skin color and red cards was found for 20 of the teams. The median result was that dark-skinned players were 1.3 times more likely than light-skinned players to receive red cards. But findings varied enormously, from a slight (and non-significant) tendency for referees to give more red cards to light-skinned players to a strong trend of giving more red cards to dark-skinned players.

ACKNOWLEDGMENT:

The work reported in this paper was supported in part by ProCivicStat project, a strategic partnership of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged, funded by the ERASMUS+ program of the European Commission. However the views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.I. (1984). *Classification and regression trees*. Belmont, California: Wadsworth.

De Maio, F. (2008). Pedagogical notes on statistics and human rights. *DataCritica* 2 (1), 19-26.

Engel, J. (2013) Statistics Education and Human Rights Monitoring. IASE Satellite Conference, http://iase-web.org/documents/papers/sat2013/IASE_IAOS_2013_Paper_2.5.1_Engel.pdf (retrieved Feb 4, 2016)

Finzer, W. (2001). Fathom Dynamic Statistics (v1.0) [Current version is 2.1]: Key Curriculum Press.

Frischemeier, D., Biehler, R. & Engel, J. (2016). Competencies and dispositions for exploring micro data with digital tools. (Paper in preparation for IASE Roundtable conference 2016)

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

Konold, C. & Miller, C. (2011). TinkerPlots 2.0. Emeryville, CA: Key Curriculum Press.

Sarkar, D. (2008). *Lattice. Multivariate Data Visualization with R*. Springer: New York.

Silverzahn, R. and Uhlmann, E. (2015). Many hands make tight work. Nature, (526), 189-191. http://www.nature.com/news/crowdsourced-research-many-hands-make-tight-work-1.18508 (retrieved Jan 8, 2016)

Silberzahn, R. Uhlmann, E.L, Martin, D. & Nosek, B. (2014). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. Online: https://osf.io/47tnc/ (retrieved Feb 4, 2016)

Therneau, T., Atkinson, E. & Ripley, B. (2015). Package 'rpart'. https://cran.r-project.org/web/packages/rpart/rpart.pdf (retrieved Feb 4, 2016)