

VISUALIZING DATA FOR GOOD

Andreas Kamilaris

Department of Computer Science, University of Twente, The Netherlands

Abstract – Good data facilitates acting upon pressing issues, while visualizations help bring the data to life, offering important insights to policy-makers and the public. A specific area where awareness-building is crucial, is the one of social inequalities around the world. This paper presents the work of 17 student team projects, developed during the “Data Visualization” course of the University of Twente at the spring semester 2018, focusing on the topic of enlightening the public over social inequalities, targeting different sustainable development goals, from the ones set by the United Nations. Based on truthful, functional and insightful data visualizations, social inequalities can be effectively communicated to stakeholders and the public, raising awareness on different unjust issues happening around the globe.

Keywords – Data Visualization, Sustainable Development Goals, Social Inequalities.

1. INTRODUCTION

Good data gives the world the information it needs to act upon pressing issues, and visualizations can help bring the data to life. Data visualization can better explain and communicate problems, inequalities, complex systems and designs, both to the stakeholders and to the general public [1]. Big data analytics and data visualization have been used together for the greater good [2], i.e. to model and understand congestion on roads and public transport, the potential impact of natural disasters or their consequences after they occur, city “story maps” for engaging citizens in municipal service delivery, small business decisions, land use planning etc.

Data analysis can be used for better decision- and policy-making, e.g. to better allocate funding to alleviate poverty, surveillance forces to reduce crime, to mitigate the impact of agriculture on the physical environment etc. [3]. Proper, self-explanatory graphs help to raise awareness over various problems happening around the world and promote action-taking [4]. A sensitive area where awareness-building is crucial, is the one of social inequalities around the world.

On September 25th 2015, countries adopted a set of 17 goals to end poverty, protect the planet and ensure prosperity for all, as part of a new sustainable development agenda set by the United Nations [5]. Each *sustainable development goal* (SDG) has specific targets to be achieved over the next 15 years, covering social and economic development issues including poverty, hunger, health, education, climate change, gender equality, water, sanitation,

energy, urbanization, environment and social justice. For the goals to be reached, everyone needs to do their part: governments, the private sector, civil society and the general public.

An effective way to promote these SDGs is by means of truthful, functional and insightful data visualizations, which communicate social inequalities to the public, increasing the knowledge and understanding on different unjust issues happening around the globe.

The contribution of this paper is to present the work of 17 student team projects, which focused on the topic of enlightening the public over social inequalities, targeting different SDGs.

The rest of the paper is organized as follows: Section 2 explains the methodology followed, while Section 3 analyzes the results of the team projects. Then, Section 4 demonstrates some exemplar visualizations and Section 5 concludes the paper.

2. METHODOLOGY

During the spring semester of 2018, the 74 students of the “Data Visualization” course of the “Creative Technologies” bachelor programme of the University of Twente were requested to form groups of five students, for a team project assignment. In total, 17 teams were formed.

Their task for each team was to choose one of the 17 SDGs of the United Nations [5], depicted in Figure 1, and locate relevant datasets with the goal of visualizing inequalities around the world. Students selected datasets from the online data sources listed in Table 1. Visualizations should be explanatory, with the general public being the

audience. They should follow the fundamental guidelines of good visualizations of data [6], and comply to the five qualities of good visualizations set by Alberto Cairo [7], i.e. truthful, functional, beautiful, insightful, enlightening.



Fig. 1 – Sustainable Development Goals of the United Nations.

Table 1 – Sources for relevant datasets.

No.	Source	Website
1.	ArcGIS Hub	http://hub.arcgis.com/pages/open-data
2.	ArcGIS Open Data	http://the-sdgs.opendata.arcgis.com
3.	World Bank	https://data.worldbank.org/
4.	UN Data Catalog	http://undatacatalog.org
5.	GapMinder	https://www.gapminder.org/data/
6.	Engage Data Project	http://www.engagedata.eu/
7.	Open Data Portals	http://dataportals.org

Each team was allowed to use any software tool for visualizing their data and/or creating interactive dashboards and stories. The following software tools were proposed (and also taught/described) during the work sessions of the course:

- Tableau¹ – Easy to use tool for developing visualizations, managing data and creating interactive dashboards and story-telling.
- D3² - JavaScript library for visualizing data with HTML, SVG, and CSS.
- Kendo UI³ - jQuery-based libraries and graphs for web and mobile development with HTML5 and JavaScript
- Vega-Lite⁴ - A high-level grammar of interactive graphics. It provides a concise JSON syntax for rapidly generating highly interactive visualizations.
- Data Illustrator⁵ - Aims to make

¹ Tableau. <https://www.tableau.com/>

² D3 Data-Driven Documents. <https://d3js.org>

³ Telerik Kendo UI. <https://www.telerik.com/kendo-ui>

visualization creation as easy as creating an illustration in a design tool.

The assignment had a duration of three weeks, and the students were provided with feedback about their intentions at the end of the first and second week. Finally, they had to present their team project in 5-minute presentations, during the final work session of the course.

3. ANALYSIS

The topics selected by the teams, together with the relevant SDGs, are presented in Table 2. A wide variety of topics was selected by the students, from various of the 17 SDGs. The most popular goals tackled were SDG1 (4 projects), SDG6 (3 projects) and SDG7 (2 projects), dealing with poverty, clean water/sanitation and clean energy respectively.

Table 2 – Topics and SDGs targeted.

Topic	SDG	No. of Projects
Populations living in poverty and people living in slums	1	2
Is McDonald's an indicator of inequality? Does McDonald's only settle in rich countries, or do they make a country more prosperous?	1	1
Poverty, wealth and health in the Middle East	1	1
No Hunger, food production and consumption	2	1
The influence of power inequality on health in Europe	3/5	1
Gender equality and salaries	5	1
Basic sanitation	6	1
Clean water and sanitation	6	2
Affordable and clean energy	7	2
Inequalities in the distribution of funds amongst provinces and municipalities of the Netherlands	8	1
Industry, innovation and infrastructure	9	1
Inequalities in the Netherlands on people with disabilities	10	1
Climate change	13	1
What is happening to our land and the life it sustains	15	1

⁴ Vega-Lite. <https://vega.github.io/vega-lite/>

⁵ Data Illustrator. <http://data-illustrator.com/>

In respect to software tools, most teams preferred to use Tableau (11 teams, 65%), as it is easy to use and it allows to develop interactive dashboards and stories quite fast. Three teams experimented with the D3 libraries, adjusting similar examples found in the D3 online gallery. The rest teams used a combination of Tableau, D3 and Kendo UI, while one particular team worked with Flourish⁶, which is another powerful platform for data visualisation and storytelling. None of the teams worked with Vega-Lite or Data Illustrator, although their use was promoted at the working sessions.

The students encountered various problems through their assignments, most common ones being the quality of the datasets discovered. In many cases, data for some countries or time periods were missing, or some values constituted outliers or non-sense. Data pre-processing and wrangling was needed, either manually or by means of specialized software such as Trifacta Wrangler⁷, which helps data analysts clean and prepare messy, diverse data quickly and accurately. Students were familiar with the tool as they used it in another parallel course of the same module of the programme.

4. DEMONSTRATIONS OF RESULTS

This section highlights some of the best visualizations created by the 17 teams. They have been organized according to the SDG they focus on.

SDG1: No Poverty

Figure 2 depicts the countries of the Middle East with their populations affecting the size of the flag. By clicking on each flag, a user can observe the Gross Domestic Product (GDP) per capita. Some countries such as Qatar and Kuwait, although small in population, they have higher GDP per capita in comparison to highly populated countries such as Saudi Arabia, Iran and Turkey.

In Figure 3, a tree map illustrates the percentage of people in sub-Saharan countries living below the international poverty line of \$1.90/day. The percentage ranges from 27% (Ethiopia) up to 77% (Madagascar), with percentages above 70% also for Burundi, Malawi and Congo. Although Nigeria has a percentage of 53%, due to its large population, it comes first in absolute numbers (99.5 million people), as shown in Figure 4. The second largest country in absolute numbers is Congo, with 60.7

million people below the poverty line.



Fig. 2 – Poverty, wealth and health in the Middle East.

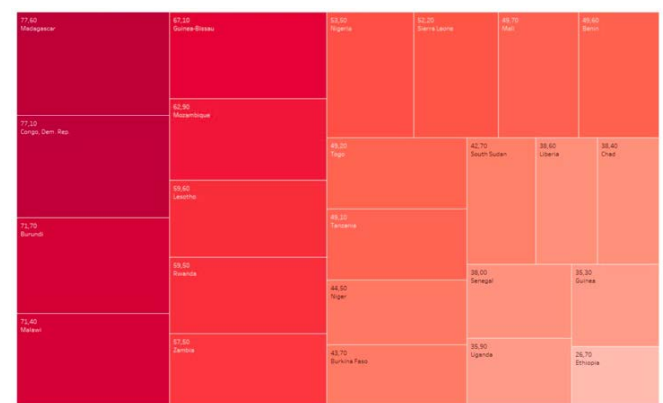


Fig. 3 – Sub-Saharan countries and the percentages of people living below the poverty line.

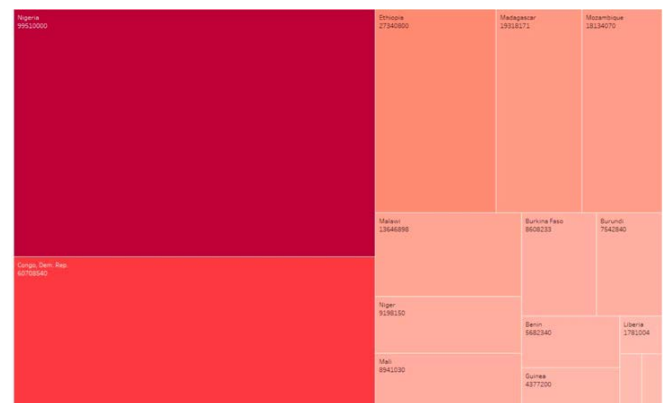


Fig. 4 – Absolute numbers of people at Sub-Saharan countries living below the poverty line.

SDG2: Zero Hunger

A bubble chart was selected by the students in Figure 5 to depict the food deficit per country in kilo-calories. The larger the bubble, the higher the deficit. Larger bubbles can be observed for developing countries such as Rwanda, Zambia, Haiti, Tanzania and Ethiopia. On the other hand, smaller

⁶ Flourish. <https://flourish.studio/>

⁷ Trifacta Wrangler.

<https://www.trifacta.com/products/wrangler/>

bubbles represent mainly countries of Europe and North America.

The relationship between undernourished population and yearly income per capita is presented in Figure 6, with different colors per continent, and size of each circle representing the country population. African and Asian countries have the lowest incomes, higher percentages of undernourished population, and at the same time larger populations. Exactly the opposite happens in Oceania, Europe and North America.



Fig. 5 – Average food deficit per country in kilo-calories.

SDG3: Good Health

The infant mortality rate around the world is presented in Figure 7. The deep purple color of the choropleth map in central Africa and south-east Asia emphasizes the shocking statistics of mortality rates between 30-50% in some of the developing countries located in these areas. Figure 8 targets the cases of AIDS identified in the Africal continent, comparing 1990 and 2016. Although there are better drugs to fight this disease today, the percentage of infected people has actually increased by 5-15% in these 26 years. For the case of Botswana, as Figure 8 shows, this increase has been 15%, from 7% in 1990 to 22% in 2016.

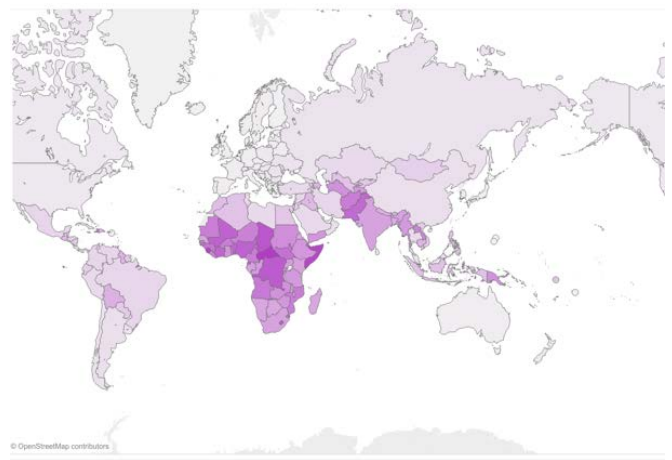


Fig. 7 – Infant mortality rate globally.

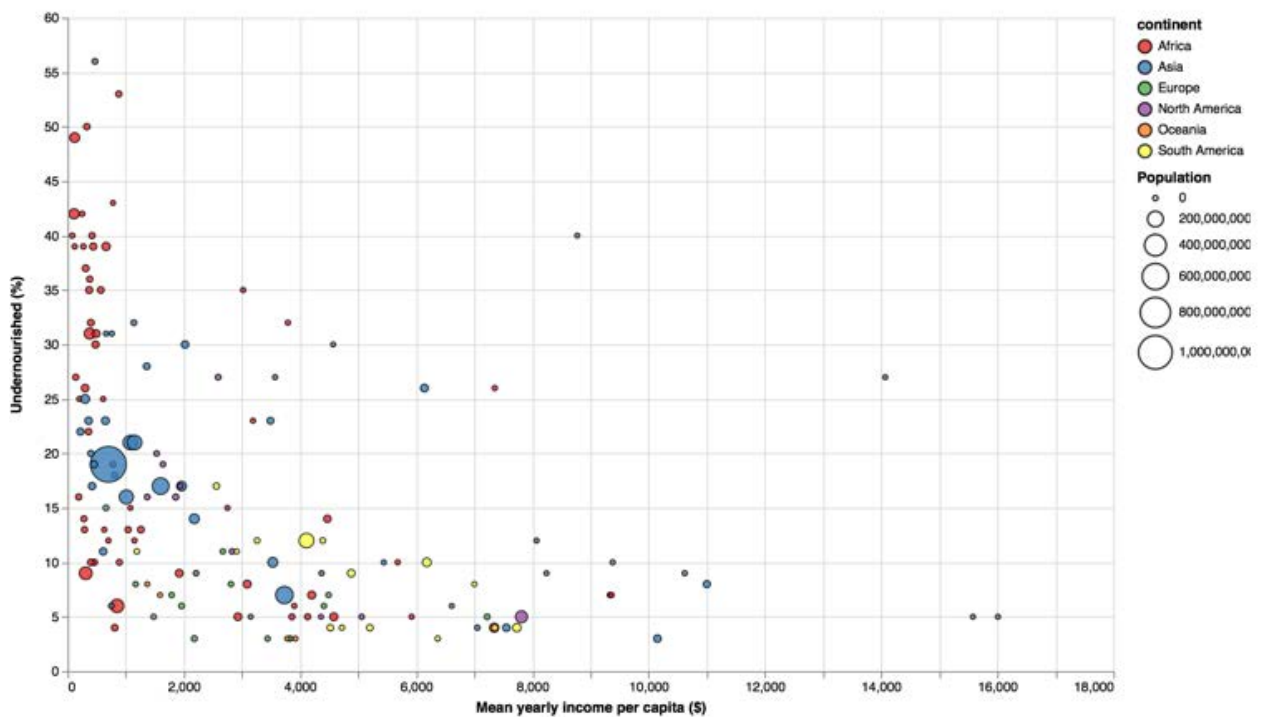


Fig. 6 – Undernourished population vs. income per capita.

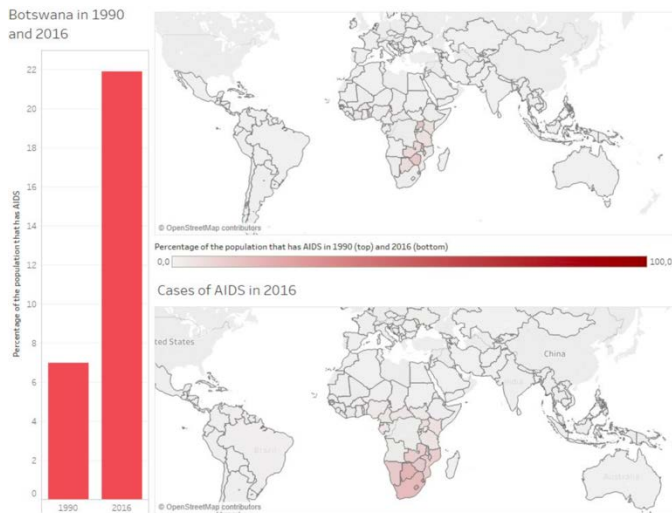


Fig. 8 – Cases of AIDS in Africa between 1990 and 2016.

SDG4: Quality Education

The scatterplot of Figure 9 aims to study the relationship between the percentage of students enrolled in secondary school and the expenditure of different countries in education per student. The inequalities are evident on this map, since the sub-Saharan African countries are placed in the bottom-left (yellow circles), while European countries are clustered in the middle-right side (green circles).

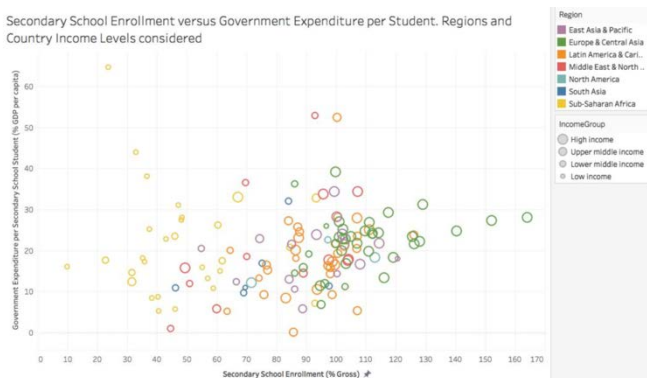


Fig. 9 – Secondary school enrollment vs. government expenditure in education per student, in different countries around the world.

SDG5: Gender Equality

The map of Figure 10 illustrates the power score of each country as calculated by the European Institute for Gender Equality⁸, and it is the average of political, social and economic equality scores. Scandinavian countries and France have the highest scores from the European countries involved. There is some inequality between the north Europe vs. the Balkans, Portugal and East European countries.

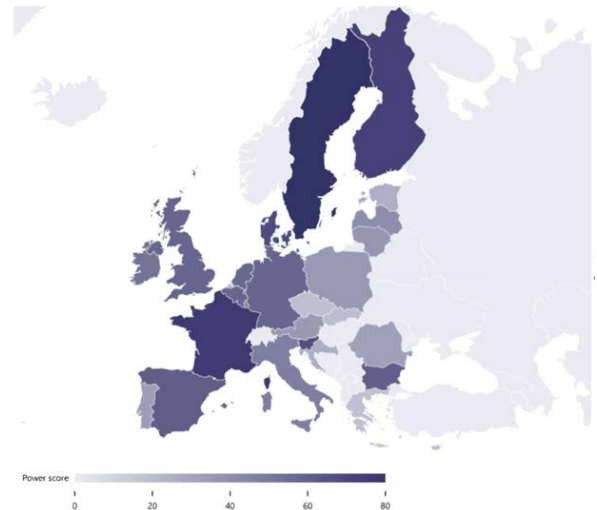


Fig. 10 – Power score in different European countries.

SDG6: Clean Water and Sanitation

The percentage of the population of each country without access to basic sanitation services (e.g. clean drinking water, waste disposal) is illustrated in Figure 11. Besides north America, Oceania and Europe, the rest of the world faces issues with limited access to sanitation services, with the problem being severe in sub-Saharan Africa (3-10% access) and south-east Asia (5-35%), and also in Bolivia (40%).

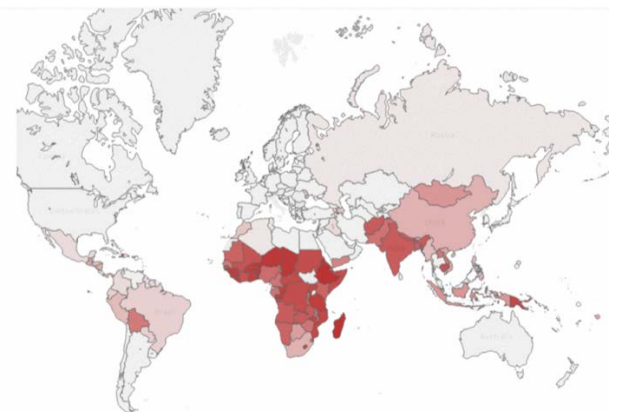


Fig. 11 – Access to basic sanitation services.

SDG7: Clean Energy

Figure 12 is very insightful, as it provides information about the use of renewable energy around the world, visualizing with green color the countries that rely on renewables and with red the ones that still use fossil fuels. It is apparent that the developed world is still investing on fossil fuels (i.e. 5-15% use of renewable sources for their total energy needs), while the developing countries (i.e.

⁸ European Institute for Gender Equality.

<http://eige.europa.eu>

central and South America, central Africa, south-east Asia) employ mainly renewable energy sources (i.e. photovoltaics, solar heaters, wind turbines) for their energy needs (40-90%). It is remarkable that from the developed countries, Scandinavian countries and Iceland constitute the exceptions, investing on renewable clean energy (35-70%).



Fig. 12 – Renewable energy use per country.

SDG8: Good Jobs and Economic Growth

In relation to work incomes and hours of work, a team aimed to highlight the existing inequalities using McDonald's meals as a reference value. The scatterplot of Figure 13 presents the relationship between salaries and hours of work needed to purchase a meal from McDonald's. Most developed countries are located at the bottom-middle section of the plot, however some developing countries in Africa and south-east Asia appear at the top-left area. This means that these countries need to work more than 1 hour to purchase a single meal, while their yearly salary is less than \$2,000.

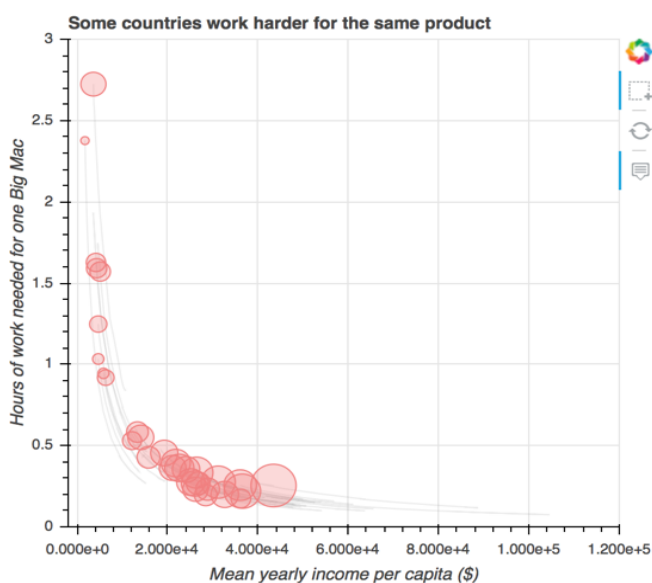


Fig. 13 – Hours of work for a McDonald's meal vs. yearly income per capita.

Similarly, Figure 14 shows more generally the mean hourly income per capita vs. the yearly working time per capita. Inequalities do exist between developed (big circles) and developing countries (small circles), as people in the developing world tend to work more and get paid less (i.e. bottom-right section of the graph). The size of the dots in Figure 14 denote the price of a Big Mac meal at each country. From the size, it becomes evident that pricing is fairer between prosperous countries and developing ones (smaller circles).

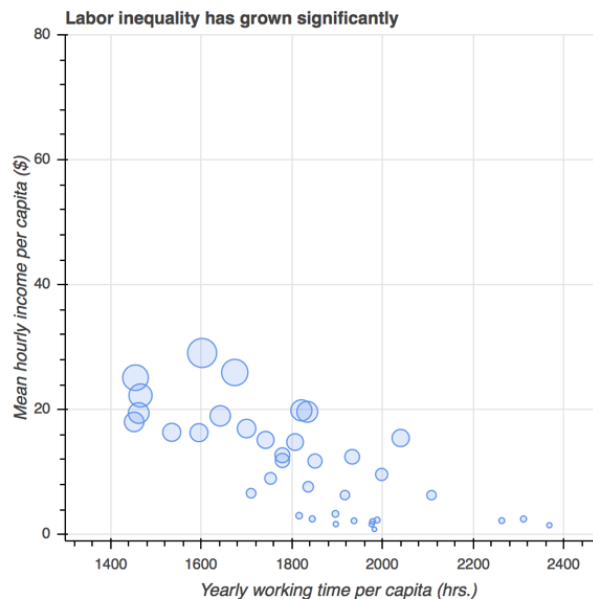


Fig. 14 – Hourly income per capita vs. working time per capita.

Finally, a team selected to visualize the workers in agriculture through time, and to show the difference between male (green color) and female (orange color). As Figure 15 indicates, in the last 20 years, there is an overall decrease of 190K people working at the agricultural sector, and this decrease has been mostly observed in the male population (165K workers less in 2016 than in 1996). This decrease has made the inequality between male and female workers in this domain lower.

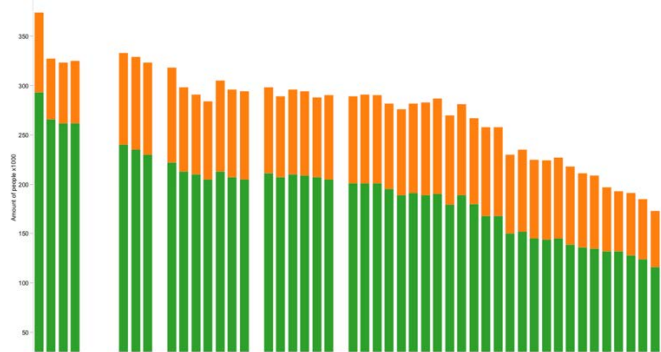


Fig. 15 – Male and female workers in agriculture in last 20 years.

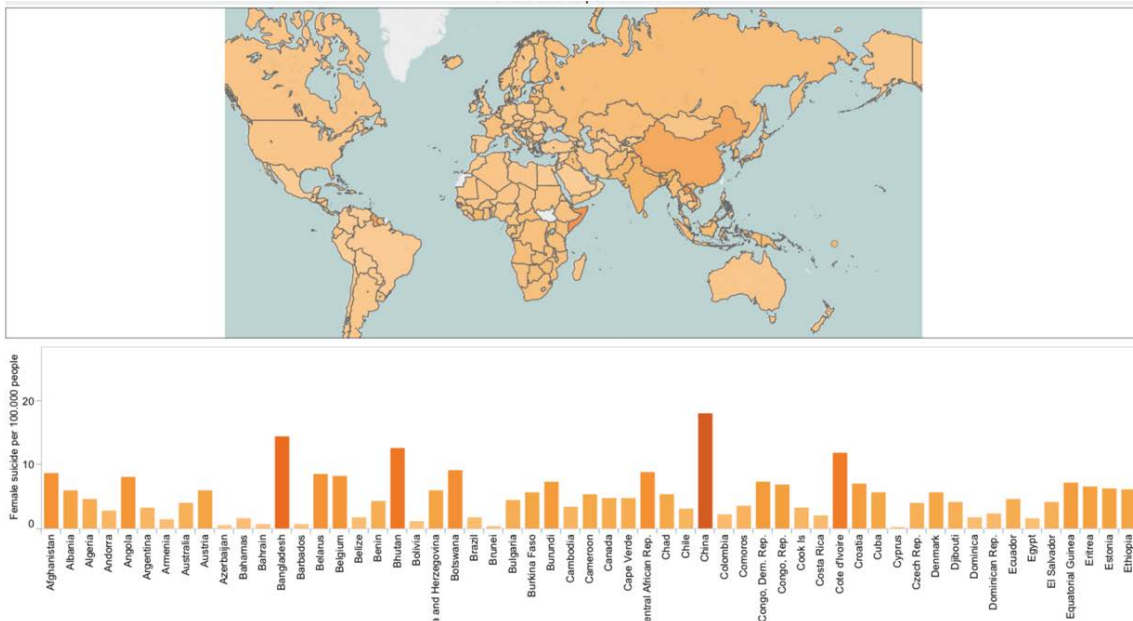


Fig. 17 – Female suicides around the world.

SDG9: Innovation and Infrastructure

The horizontal bars of Figure 16 compare the access to electricity between 1990 and 2016, only for countries whose population had less than 50% access to electricity in 1990 worldwide. The graph shows that in all the countries listed there is a substantial improvement in the access to electricity. Still however, there are numerous countries with 40-80% access for their citizens, while 20 developing countries such as Chad, Niger, Burundi and Liberia have percentages less than 40%. There is still a large inequality also on this aspect, in comparison to the developed world (95-100% access to electricity).

SDG10: Reduced Inequalities

This SDG is about reducing inequalities in general, for inequalities not listed in the other categories. One student team discovered an interesting visualization to demonstrate differences in female suicides around the world. This inequality is visualized in Figure 17, showing a choropleth map on the top, with the numbers of female suicides around the world (i.e. the darker the color, the more suicides at some country), while at the bottom the exact numbers per country are provided on a bar chart. Bangladesh, Bhutan, China and Cote d'Ivoire are the countries with the highest numbers of suicides (i.e. more than 1 per 10K female citizens), and generally higher numbers are recorded in the developing world. Similar observations have been made for the cases of male suicides, not presented in this paper.

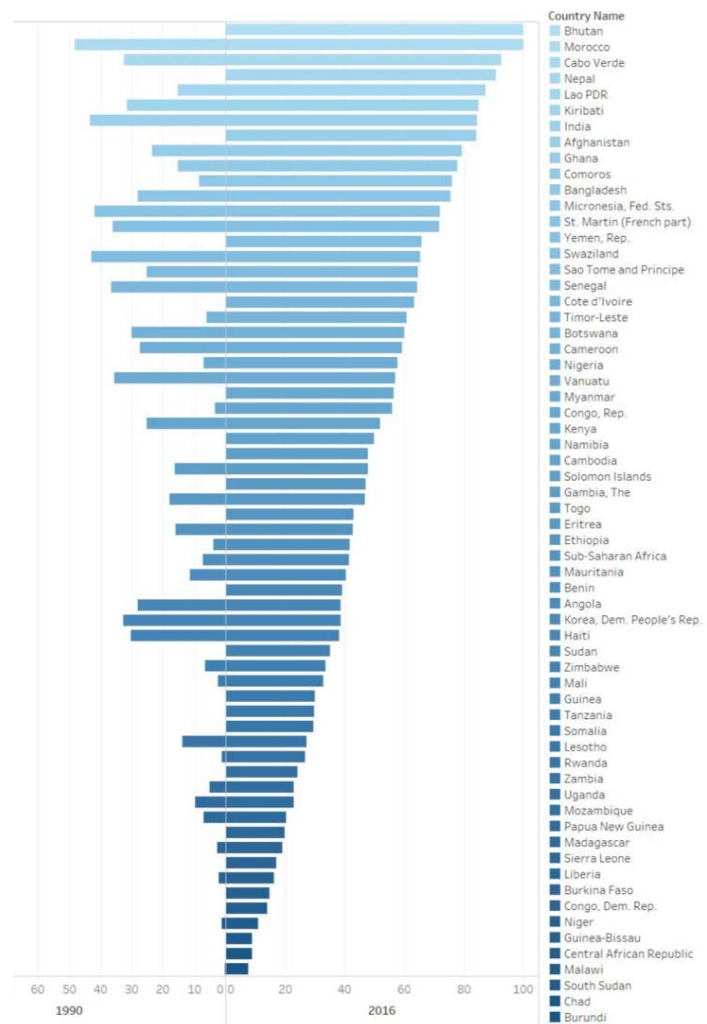


Fig. 16 – Comparison of access to electricity between 1990 and 2016, for countries with below 50% access in 1990.

5. CONCLUSION

This paper has presented the work of 17 student team projects during the “Data Visualization” course of the spring semester 2018 at the University of Twente, focusing on enlightening people over existing global social inequalities, targeting different sustainable development goals, from the ones set by the United Nations. Seventeen different visualizations from 10 different SDGs are depicted, selected from the students’ projects, demonstrating how social inequalities can be effectively communicated to stakeholders and the public, based on truthful, functional and insightful data visualizations, in order to raise awareness on different unfair issues happening around the globe.

ACKNOWLEDGEMENT

Special thanks to the students of the “Data Visualization” course of the spring semester 2018, part of the Module 8 of the “Creative Technologies” Bachelor programme of the University of Twente, for their excellent contributions and team projects. Selected visualizations prepared by the students have been presented in this paper. Many thanks also to the student assistants of the course Erik Kemp and Adam Baco, for the excellent feedback and support provided to the students during the assignment.

REFERENCES

- [1] S. R. Sheppard, "Landscape visualisation and climate change: the potential for influencing perceptions and behaviour," *Environmental Science & Policy*, vol. 8, no. 6, pp. 637-654, 2005.
- [2] C. V. and E. A., "Big Data for the Greater Good: An Introduction," *Studies in Big Data*, vol. 42, pp. 1-18, 2019.
- [3] A. Kamilaris and F. Ostermann, "Analyzing the Relationship between Geospatial Analysis and the Internet of Things," *ISPRS International Journal of Geo-Information*, Special issue on "Geospatial Applications of the Internet of Things (IoT)", vol. 7, 2018.
- [4] E. Lange, "Issues and Questions for Research in Communicating with the Public through Visualization - Trends in Real-Time Landscape Visualization and Participation," Proceedings at Anhalt University of Applied Sciences, Dessau, Germany, 2005.
- [5] United Nations, "Sustainable Development Goals (SDGs)," 2018. [Online]. Available: <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>.
- [6] M. O. Ward, G. Grinstein and D. Keim, Interactive data visualization: foundations, techniques, and applications, AK Peters/CRC Press, 2015.
- [7] A. Cairo, The truthful art: Data, charts, and maps for communication, New Riders, 2016.