

Design study project: Suicide Rates

Christian Rauch*

Daniel Hanzer†

Roman Schneglberger‡

University of Vienna
Faculty of Computer Science

1 MOTIVATION

1.1 Background information

Our Application deals with the suicide rate of a total of 41 countries. The goal of this project was to create a visualization that show a variety of graphs that are changeable by user interactions. The graphs are able to filter to get more precisely data out of it. It is possible to compare different countries with each other. Furthermore, it is possible to compare the suicide rates with unemployment rates, poverty rates as well as Gross domestic product (gdp) rates for each country. These comparisons should help to find possible parallel between the data in order to find any conclusions about the reasons for suicidal ones.

1.2 Tasks

Our tasks was to visualize the existing records as good as possible. One of our tasks was creating a heatmap. In addition, a timeline should be available to filter for certain periods of time. In general, our the main task was to install meaningful filters or groupings to give the graphs as much meaning as possible. Another task was to find an explanation for the reason for the increased suicide rates in certain countries or regions, therefore we compared them with gdp, unemployment and poverty rates and visualized these through a matrix scatterplot and a line chart over time. Another task was comparing different countries or regions with one another over certain years.

1.3 Users

On the one hand, our users are generally interested in our project, as well as users who conduct research in the field of suicide. Our users should be able to better understand the records by filtering them. For example, users might question why certain countries have a higher suicide rate than others and whether it involves economic or geographic causes. In particular, the course of suicide rates over time are relevant for this. Potential users may include journalists, researchers and psychologists.

1.4 Data

The Project focused mainly on the Dataset "Suicide rates" from OECD Data. The Data are from the year 1960 until to the year 2015 and include the amount of suicides per 100.000 in percentage. Altogether, the data of 41 countries are presented in this dataset. Furthermore, we used "Gross domestic product (GDP)", "Poverty rate", as well as "Unemployment rate" from OECD Data for comparing these datasets with the suicide rates. The Data from GDP are from the year 1960 until to the year 2015 and include 40 countries. The Data from Poverty rate are from the year 1976 until to the year 2015 and include 40 countries. The Data from Unemployment rate are from the year 1953 until to the year 2015 and include 40 countries.

*e-mail: a01202875@unet.univie.ac.at

†e-mail: a01349699@unet.univie.ac.at

‡e-mail: a01127050@unet.univie.ac.at

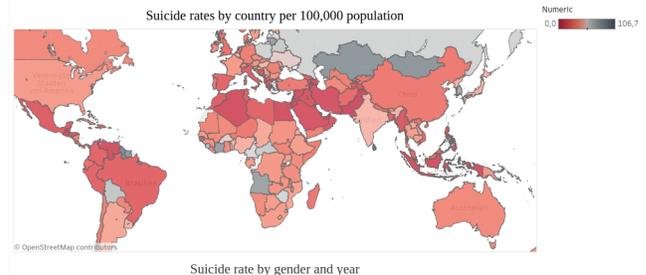


Figure 1: Vis from Tableau

Table 1. Sociodemographic Characteristics of Surveyed U.S. Oncologists

Characteristic	All Oncologists (n = 3299)	Medical Oncologists (n = 2501)	Surgical Oncologists (n = 2399)	Radiation Oncologists (n = 231)	Pediatric Oncologists (n = 228)
Mean age (range), y	47 (28-85)	47 (29-85)	49 (28-82)	48 (28-74)	45 (30-72)
Sex, %					
Male	80.6	81.8	85.4	77.9	66.5
Female	19.4	18.2	14.6	22.1	33.5
Religion, %					
Protestant	34.5	32.9	40.5	42.3	34.1
Catholic	26.6	26.6	31.5	25.3	23.5
Jewish	25.7	26.3	18.1	22.2	32.7
Other	13.2	14.2	9.9	10.2	9.7
Importance of religious belief, %					
Very important	34.0	34.2	38.3	35.0	26.3
Fairly important	33.0	32.2	36.2	36.5	33.3
Not important	33.0	33.7	25.5	28.5	40.4
Population of geographic setting, %					
<100 000 persons	12.1	13.4	5.1	12.8	4.4
100 000-500 000 persons	30.0	31.1	22.2	36.3	17.3
>500 000 persons	57.9	55.6	72.6	50.9	78.3
Type of practice, %					
Academic	34.1	27.7	63.6	28.5	80.3
Oncology specialty group	34.2	37.6	7.9	46.4	6.6
Other	31.8	34.7	28.5	25.2	13.2
Time spent in patient care, %					
<50%	13.4	13.1	11.2	5.8	29.2
50%-89%	35.2	31.6	56.0	34.5	54.0
≥90%	51.5	55.3	32.8	58.7	16.8
Patients who died in the past 12 months, %					
<25	37.5	31.1	74.4	20.0	93.2
25-49	26.9	30.0	19.3	24.3	5.0
≥50	35.6	39.0	6.3	55.7	1.8

Figure 2: Data usage in scientific papers

One problem with our Data was to merge them together. Another one was that not for each country are Data available for each year, so to speak there are a lot of null values in the existing records.

2 RELATED WORK

2.1 Other visualizations we discovered during this project:

We did a lot of research in advance on interesting large datasets. From data visualisation point of view(Figure 1) and also in scientific papers (Figure 2), which deal with the topic, as these also contain graphics to clarify the text/make a point. These papers had mostly poor graphics included, or tables with numerical values. There were also some visualizations on Tableau-Public - these showed small details of the total amount of data. For example only for a few countries, not linking the the data to other datasets or providing interactivity. OECD(Figure 3) website shows also graphs with their information. The solution can be used on many visualization topics.

2.2 Any previous visualization Ideas

One idea discarded was this detail visualization on male to female ratio(Figure 4) by limitations of Tableau.

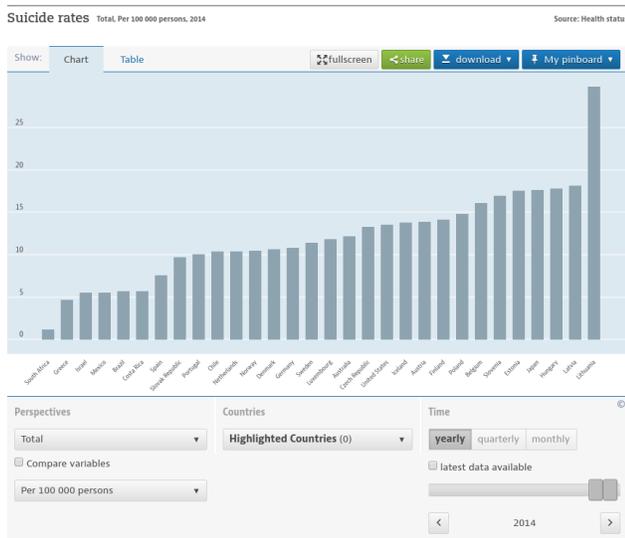


Figure 3: View from oecd webpage

2.3 References both academic and commercial tools used

Tools used in this project:

<https://public.tableau.com/en-us/s/>

Data taken from:

<https://data.oecd.org/healthstat/suicide-rates.htm>

<https://data.oecd.org/unemp/unemployment-rate.htm>

<https://data.oecd.org/inequality/poverty-rate.htm>

<https://data.oecd.org/gdp/gross-domestic-product-gdp.htm>

Other resources used:

<http://www.cs.ubc.ca/tmm/courses/533-11/resources.html>

3 APPROACH

3.1 Description (of our visualization design):

The approach was to build a tool for end users, so they can discover and get insight into suicide rates. And for us, to practice visualisation techniques from this course. For this reasons we choose a huge dataset containing lots of usable data. OECD website has a lot of public datasets available. We thought of combining suicide data with poverty rates, GDP and unemployment rates to see, if these matches together. We discussed a lot of ideas, talked about possible solutions and how to implement the course content. Then we build a prototype with many example views. Next step was to get feedback and improve the views. We picked the best visualization and refined them. So then we finally came up with a single dashboard.

3.2 Brief description of the dashboard views:

The dashboard(Figure 5) contains a huge map that is zoom and panable, shows data on the different countries with available information. The heatmap shows the number of suicides, a darker blue indicates a higher suicide rate. Top right are the different values shown, each of these 41 in a individual color. Bottom left is the timeline, a drop down menu can be used to show different data in comparison to the suicide rate(GDB, poverty and unemployment rate). A filters can be applied to the timeline. And then we have this scatter plot showing how this data correlates. On mouse over the users get a little tooltip window showing information. Multiple elements selected and be filtered.

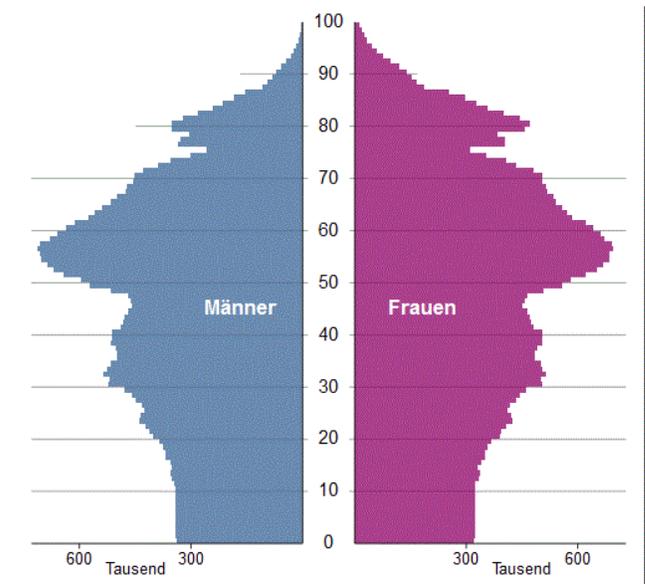


Figure 4: Gender gap

4 IMPLEMENTATION

4.1 Brief description of the Implementation

We decided to make use of Tableau after discussing other more technically challenging toolkits like d3 library. Our design study project on mind we have come to think about simplicity, focusing on the usability and interaction possibilities of a proven solution. Previously we worked with Tableau, that's why we decided to use it. On the other hand, there are some limitations that go along with it.

4.2 Implementation challenges

Implementation issues encountered, Public Tableau naturally has some restrictions to what can be accomplished. The used datasets were fairly huge. Joining the data together often resulted unexpected program behavior. The original datasets were missing some information, so there is no data provided for all countries. In the original data there were different scales used on the data. Some filters were hard to apply, also the selection in the drop down menu. Creating 41 different colors for the country list was also difficult.

4.3 Reasons for design choices

At the beginning we had focused on both, the data and ways to visualize them. The main questioned was, who is the audience. Which end users might want to use this visualization. In which facts they might be interested in? How they gonna use it? So we collected information and feedback and included it in our work.

For this reason one thing we focused on a clear design with high data-ink ratio. There shouldnt be a distraction from unnecessary chart chunk or from distracting colors. The design is simple, the user interface is easy to use and the charts gives easy to read information an overall good overview. The feedback we received was altogether pretty good. No one had problems using it, for most of them is seemed very familiar from the start. So this is an good entry point to explore the data further. On top of that the Functionality-Usability-Aesthetics Pyramide was implemented as well.

5 RESULTS

5.1 Scenarios of use examples

- Researching the impact of events on suicide rates

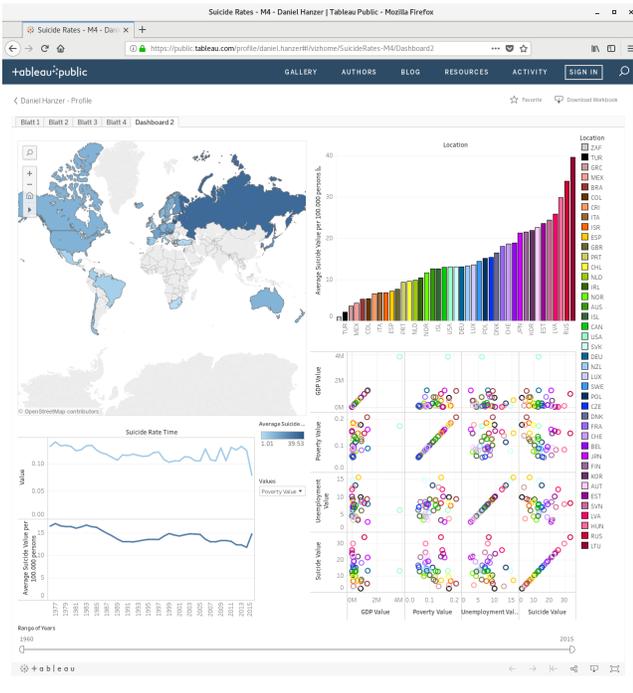


Figure 5: dashboard view

A researcher wants to view the impact on suicide rates in a country of a certain event that happened in 1985. For this, the researcher selects the country in the map view (Figure 8) and selects a appropriate time scale (Figure 7).

In the line chart (Figure 6) a slightly lower suicide rate in Canada in the year 1985 is visible.

- Researching how much / if economical factors influence the suicide rates in different countries

We select multiple European countries with CTRL + click on the map view as shown in Figure 9. In the color coded scatter plot matrix (shown in Figure 10) you can compare the correlation between various economical factors and suicide rates in the selected European countries.

- Simply comparing suicide rates in different countries

Without selecting anything, per default our visualization shows the suicide rates of all countries (with available data) in a bar chart. As you can see in Figure 11, the bar chart features color coding (identical to scatter plot matrix) and a mouse-over hint.

This default view should show the large differences in suicide rates in the various countries and encourage interest to the question "Why is that?". The user can now play around with the other charts and try to find possible explanations in economical factors.

5.2 Performance of the system

We designed, tested and improved our visualization to perform well with the intended use case scenarios, including but not limited to the ones detailed in 5.1.

The performance in speed is limited by our visualization tool of choice (Tableau) and the size of the dataset. We discussed limiting the dataset, but that would possibly lower the functionality for a rather small usability gain and could also violate the design principle to show context.

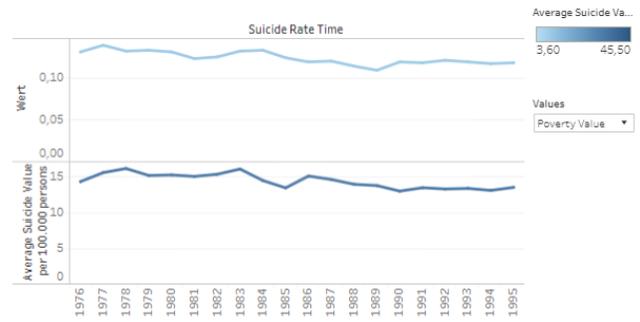


Figure 6: Line chart showing a dip in the value of suicide rates in Canada of 1985



Figure 7: Selecting a time scale for a more detailed view

5.3 Feedback from evaluations

User feedback was a vital element in the design and improvement of our visualization. We improved the color coding (Figures 10 and 11), implemented a scatter plot matrix instead of multiple scatter plots (or scatter plots selectable by a drop-down menu) and made a split line graph (Figure 6) all based on user feedback and evaluations.

With our final visualization the only user complaint remaining is the loading time when selecting multiple countries one after another and initial difficulties to interpret the scatter plot chart.

5.4 Findings

An interesting finding in our visualization is the increase in the suicide rate of Greece from 2010. A possible explanation for this increase could be the launch of the Greek crisis in late 2009. In 2010, Greece sought help from the European Union to prevent a national bankruptcy. Since then, unemployment rates have risen significantly, as well as a slight increase in poverty rates. How to assume the GDP value has fallen significantly. These relationships can be derived from our visualization, but it is not conclusive enough whether this is the reason for the increase in suicide in Greece, because many other unconsidered factors could play a role here. This basis could be enough to trigger further research.

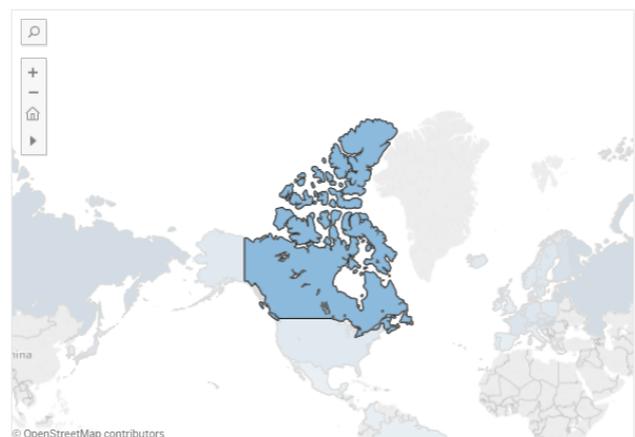


Figure 8: Map view

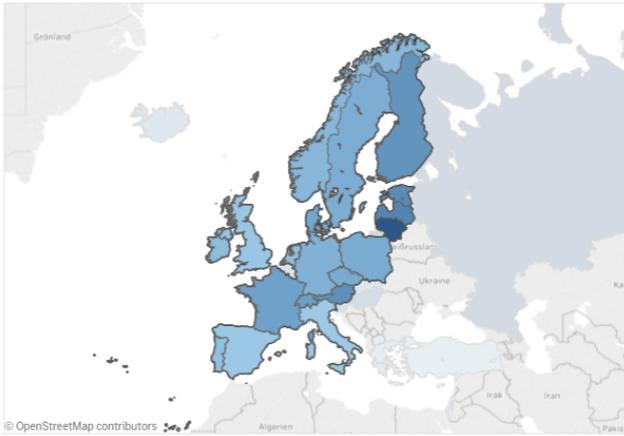


Figure 9: Multiple selection on map view

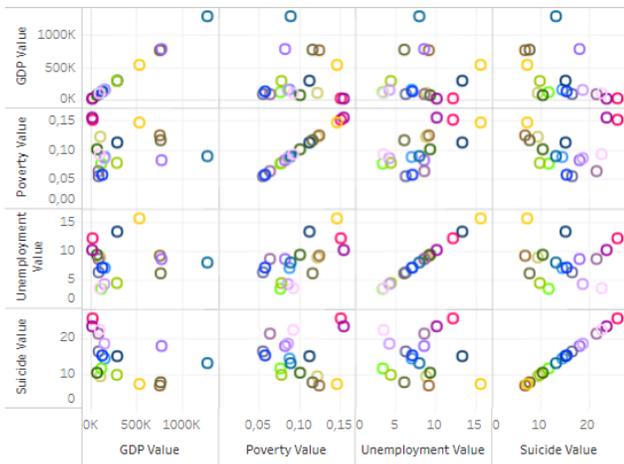


Figure 10: Scatter plot matrix

6 DISCUSSION

6.1 Strengths and weaknesses of our visualization

Strengths Every graph is necessary to visualize the intended concepts - no "chart clutter". Our visualization provides a quick overview of the different suicide rates in many OECD nations per default and makes it easy to compare them, which would probably be the most common use case. For a more detailed analysis it provides the necessary tools to find possible correlations to other factors (economical and developmental).

The scatter plot matrix makes it easier, especially for advanced users, to spot correlation between the visualized data on a grand scale - the line chart on a more time based / historical scale. It can be simplified by restricting the number of countries shown.

(Possible) Weaknesses There are only so many factors we could incorporate into our visualization and it is not guaranteed there are correlations between these factors and suicide rates. If we included more it could lead to a too complex visualization and possibly make it even less likely for users to spot correlations as a result.

Selecting countries in the map view leads to loading times. It is not a very significant problem, but if the user wants to select many countries (for example European countries) it becomes annoying

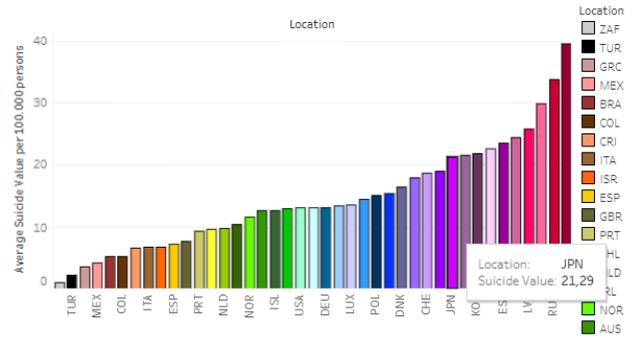


Figure 11: Default view of bar chart with mouse-over hint

pretty fast. We tried to improve the loading times, but it seems to be a problem with tableau and the size of the datasets.

The scatter plot matrix can be slightly difficult to read and interpret correctly because of its high data density. However we are confident this is the best way to visualize possible correlations between the various datasets and suicide rate among all countries and it enables the user to spot correlations on a grand scale in one graph.

It might be difficult for someone to distinguish the countries through the colors. We have tried to spread the color variation as well as to respect the brightness, but seeing up to 41 countries in the visualization can cause readability problems.

If the visualization was filtered by only one country, this leads to that the scatterplot is hardly readable, because the x and y axes adjust automatically depending on the values, if there is only one displayed value, this will show in the upper right corner. Unfortunately, we could not fix the axes, because the values of the different countries are very different and this would mean that the smaller values are all in one spot.

6.2 Lessons learned

- Visualization can be a time consuming process and good time planning is necessary.
- User feedback is a vital element for visualization. The developers are very likely to overlook possible flaws and shortcomings.
- Tableau is an easy to use tool if you want your visualization to work and look exactly as the developers intended it to. If you want to do something "non-standard" with this tool, the process can become quite complicated.
- A good visualization makes it possible to better understand large data sets.

7 SEPARATION OF TASKS: MILESTONE 4

Christian Rauch:

- "Results" section
- "Discussion" section
- "Separation of Tasks" section
- LaTeX document setup and coding

Daniel Hanzer:

- Improvements and implementation of feedback in visualization (correct color coding and scatter plot matrix)
- "Motivation" section

Roman Schneglberger:

- "Related Work" section
- "Approach" section
- "Implementation" section