Suicide: A Concern For The Future Generations

Bazil Muzaffar Kotriwala, Luc Antonissen

Fig. 1. don't give up. You are not alone, you matter signage on metal fence. Photo by Dan Meyers on Unsplash

Abstract— Suicide is a significant global issue that is the second highest cause of death among people aged between 15 and 24 years old. With the help of visualization techniques, this research aims to show the significance of this issue and to gain insights in the causes that lead people to commit suicide. Statistics on population, unemployment, gender, income and life expectancy are considered.

Index Terms—Suicide, Analysis, Forecast

1 Introduction

Suicide is among the top 15 causes of death nearly each year. It has reached epidemic proportions globally, taking lives of approximately 800,000 people annually or about 1 person every 40 seconds. It is the second highest leading cause of death amongst people aged between 15 - 24 years [4, 5]. This is a serious concern for the survival and mental well-being of our current as well as future generations around the world.

A considerable amount of research is being done to prevent and find adequate cures for cancer and cardiovascular disease, which are major causes of death globally [2]. Due to the stigma attached to the topic of suicide in different cultures, it is not openly talked about as often as it should be [5]. However, in recent times, there has been a shift and it is now being more widely spoken about with efforts being made to reduce suicide attempts. Research is being done in understanding what causes a person to commit this act and how to prevent it. Specifically, the World Health Organization has done extensive research on preventing suicide globally [4].

2 Problem Description and Task Analysis

There is not one but various reasons which typically lead to people committing suicide. A few of the reasons include drug addiction, alcohol addiction, life stress, financial stress, lack of social support, etc [4, 5]. Exploring specific factors would allow us to understand the situation better. Therefore, we would like to investigate if there are any significant relationships between specific factors contributing to people committing suicide around the world.

• *Bazil Muzaffar Kotriwala, Luc Antonissen are with Eindhoven University of Technology. E-mail: {b.m.kotriwala, l.antonissen}@student.tue.nl.*

2.1 Data

We primarily use the dataset, 'Suicide Rates Overview 1985 - 2016' [6], which has been consolidated using data released by The World Bank, World Health Organization, and United Nations Development Program [9, 10, 12]. The primary dataset contains suicide data on 101 countries from 1985 - 2016. It has 27,820 observations with features shown in Table 1.

Feature	Value Range
Country	101 countries e.g Albania
Year	1985-2016
Gender	Male, Female
Age (in years)	5-14, 15-24, 25-34, 35-54, 55-74, 75+
Number of suicides	0 - 22338
Human Development Index	0.483 - 0.944
Gross Domestic Profit ($)	251 - 126352
Generation	Generation X, Silent, G.I., Generation, Boomers, Millenials, Generation Z

Table 1. Overview of Primary Dataset

We identified additional data from the repository, 'World Development Indicators' [10], which we augment our primary dataset with. We use some selected features from this source which seem interesting to investigate with respect to suicide shown in the table below. [11].

Feature	Value Range
Unemployment Rate	0.15 - 33.47
Life Expectancy	51.06 - 86.99
Labor Rate (Female to Male)	32.76 - 90.99
Region	7 continents

Table 2. Overview of Additional Data

The resulting dataset contains some missing data for the continuous variables "Unemployment Rate", "Life Expectancy" and "Labor Rate (Female to Male)". This data can still be used with aggregation functions. The column "Human Development Index" has half of the observations missing and should be carefully used in gathering insights from the visualizations, as the data may look quite strange. Not all years are present for each country, but there are enough to be able to explain possible differences between the years. 95 countries are missing in the dataset. The most notable countries are in the data-set apart from China and India, which contribute to more than half the world population. Thus, this population is not included in this analysis. However, enough data exists such that the data can still be used to look for countries, their features, patterns and relationship with suicide rate, etc.

The final augmented dataset consists of the suicide rates across the world aggregated per year, country, age and gender. These four features and the other features that are added are all selected in their usefulness for the visualisation task. The additional features make this dataset particularly interesting as it helps us to identify whether any of these features have any unexpected relationships or patterns with respect to the suicide rate. Thus, if we could identify such unexpected patterns, it could have a high impact on preventing suicide by improving the state of those factors in the respective countries. For each of them it is not unlikely that they have an impact on the suicide rate and therefore they are included in the dataset to analyse if they actually have an impact and to quantify that impact.

2.2 Analysis

The data used is from 1991 - 2016 as the augmented data in Table 2 starts at 1991. We investigate whether the features in Table 1 have any relationships, patterns and correlations with respect to the suicide rate. The following questions will be addressed:

- Which region has had the highest suicide rate over the years?
  - Which countries have the highest suicide rate within the region?
  - Which age groups and genders contribute to the highest suicide rate within these countries?
  - Do countries in which euthanasia is legal have higher suicide rates?
- What indicators can be found that contribute to the highest and lowest suicide rate?

The tasks performed will consist of interaction between the plots, selecting data on the plots, applying multi-select filters for age, gender, income and region. The plots can also measure the relation between other features such as unemployment rate, life expectancy, GDP with respect to the suicide rate across continents and countries.

3 Visualization Design

We make use of "Dash by Plotly" to create the visualization dashboard (insert reference). The choice of visual encoding and interaction design techniques were motivated based on the questions we are trying to answer mentioned in Sect. 2.2. For the visual encodings, we make use of point and line based techniques to discover overall trends, patterns and relationship of attributes with respect to the total number of suicides. For the interaction design, we use the "Overview + Detail" approach [1]. The plots can be manipulated using navigation techniques such as item reduction (zoom, constrained) and attribute reduction (slice, cut).

3.1 Visual Encodings and Interaction

The suicide data consists of statistics on suicide over the past 25 years for a 101 countries. Therefore, it is a time-series data which influenced our decision to use certain types of visualizations which can represent the set of values changing over time. Thus, we make use of maps and statistical distributions such as the choropleth map, box-and-whisker, parallel coordinate and scatter plot [3]. All plots shown in this section represent the suicides per 100k people (1991 - 2016).

For interaction, we primarily use the overview first, zoom and filter, details on demand technique [8]. On loading the dashboard, the overview is shown with all filters selected such as all years, age groups, income groups and regions. To dive into specific details of the plots based on your requirement, filters can be removed, plots can be zoomed into, plots have click, box select and lasso features available to lookup, explore, compare and summarize the data for all plots on the dashboard. To answer the questions, mentioned in Sect. 2.2, we will need to make use of all these techniques.

3.1.1 Scatter Plot

The scatter plot is used to identify any sort of trends, clusters, correlations, extreme values for the total number of suicides with respect to the different features such as countries, regions, age, etc. [7]. It has a broad scope and is updated based on the interaction with the world map and/or the filters. It is also used to compare features across different countries for e.g. as the total number of suicides for the selected countries with respect to age groups over the years. It is also possible to change the x-axis to see the relationship, patterns and direction of continuous variables such as "Gross Domestic Profit" "Unemployment rate" etc with respect to the suicide rate. This is shown in Fig. 3.

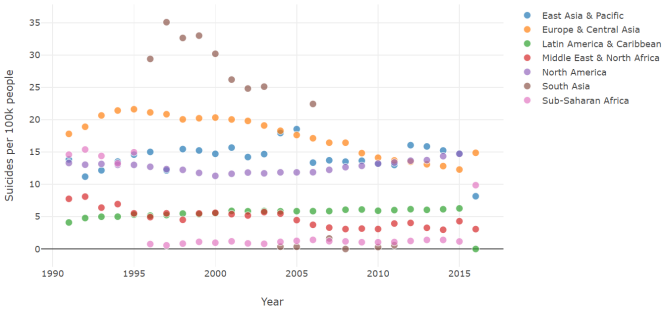

Fig. 2. Scatter Plot (Continent Based)

In Fig. 2, we can see the trend of suicide rates of each region over the last 25 years. The scatter plot helps us to identify which region to look at for further analysis. This can now be used to select the countries on the world map using a lasso or box selection to further identify which countries contribute to the suicide rate of the respective region.

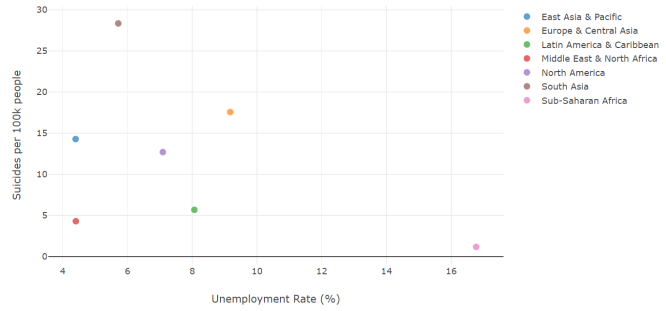

Fig. 3. Scatter Plot (Average Unemployment Rate)

3.1.2 Choropleth Map

The choropleth map enables us to visualize our geographic data based on different countries and regions respectively. A color encoding of the countries is used to indicate the intensity of the suicide rate over the selected years as shown in Fig. 4. Luminance contrasts of red are used to represent the intensity, with darker shades of red indicating higher suicide rates.

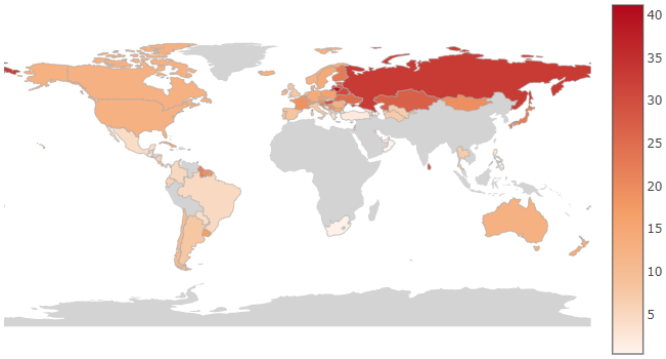

Fig. 4. Choropleth Map (Country Based)

Now we make use of the click, lasso or box select feature to select a single or group of countries on the world map to see which countries contribute significantly to the suicide rate of the respective region. Once we select the countries on the world map, all the other plots on the dashboard are updated as well with their respective suicide rates. Furthermore, to this comparison, we can further apply filters such as age groups, income groups and gender and compare those across these countries as shown in Sect. 4.

3.1.3 Box-and-Whisker Plot

A box plot (Fig. 5) is used to display the median, quantiles and 5% upper and lower range of the suicide rates [3]. It helps to understand the overall distribution of the respective countries and regions and helps to see if any extreme values exist which may be pulling the averages in a certain direction. Initially the data shows a box plot for the regions. It is possible interact with the plot using the world map. When a selection of countries is made on the world map, these will show up as a separate box, so they can be compared to the regions and how different the countries are as a collective to the entire region they belong to.

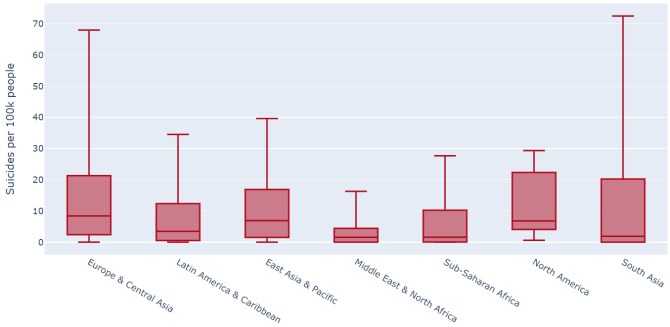

Fig. 5. Suicide Rate per 100k People per Country (1991 - 2016)

Another option is to select categories other region. The other categories are “Age”, “Gender” and “Income Group”. In this case, all countries are combined to show the plots for the selected category. In this case, when a region is selected on the world map, for each “Age”, “Gender” or “Income Group” a plot is added next to the plot for all countries for the selected data. This is shown in Fig. 14.

3.1.4 Parallel Coordinate Plot

As we have many features in the data-set with respect to the suicide rate, we decided to use at least one plot which would be able to show many variables together whilst being relatively compact. The parallel coordinate plot allows us to find patterns within the features and how they relate to each other with respect to the suicide rate [3].

Fig. 6. Parallel Coordinate Plot

In Fig. 6, we show 6 features with respect to the suicides per 100k people. These features can be ordered differently by interacting with the plot. The world map and scatter plot can interact with the parallel coordinates plot at the same time. Countries that are not in the selection in the world map will be greyed out and additionally hiding a category in the scatter plot will also make those lines grey (Fig. 15). This allows to make a very specific selection of highlighted traces, while keeping the other lines as grey to see relatively how the highlighted lines compare.

The filters on top also interact with the plot. Like all the other plots, only the categories selected in the filters will show up. However this plot also changes when only one value for a certain option is left in the plot. So for example, when only male is selected, the whole gender axis will be removed from the parallel coordinates plot (Fig. 10), since it is already known that all the data in the plot is about males.

Due to the standard functionality of the plot, it is possible to move the axes around and to highlight the traces of part of the plot. The traces which are not highlighted are a darker grey than the deselected traces by the world map and the scatter plot to distinguish between the parts that are deselected.

3.1.5 Alternate Plots Considered

Region based techniques such as heatmap and stacked bar charts were considered.

The heatmap was initially added to the dashboard and seemed useful. However, the scatter plot and the heatmap were giving similar insights in terms of clusters and which features had higher intensity or values respectively against the suicide rate. Scatter plot was preferred over the heatmap in the dashboard as we were able to add more filters to it such as age, income and regions along with continuous variables simultaneously such as unemployment rate, life expectancy etc.

The stacked bar chart was not used as we are using suicide ‘rate’ which is not optimal to visualize on it. Additionally, the box-and-whisker plot gave us more flexibility to visualize the whole distribution. Similarly, parallel coordinate plot also had preference over stacked bar chart as it is able to many more features in a compact format. [3]

4 Use Cases

The dashboard with the visualization techniques described in Sect. 3 are used to answer the research questions (Sect. 2.2).

4.1 Which region has had the highest suicide rate over the years?

As shown in Fig. 2, Europe consistently seems to have the highest suicide rate apart from a few years. This is further affirmed by looking at the world map in Fig. 4, Lithuania, Russia and the surrounding countries seem to have the highest suicide rate over the past 25 years. These countries belong to the biggest region, namely: “Europe & Central Asia”.

4.1.1 Which countries have the highest suicide rate within this region?

We find this out by selecting the 'Europe & Central Asia' region on the world map using box select or lasso as shown in Fig. 7.

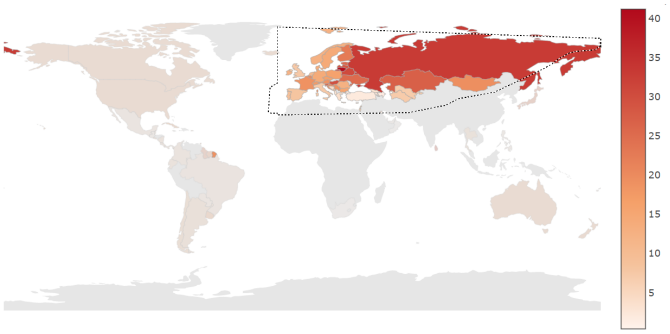

Fig. 7. Choropleth Map Country Selection

This selection updates the scatter plot, box-and-whisker plot and the parallel coordinate plot on the dashboard.

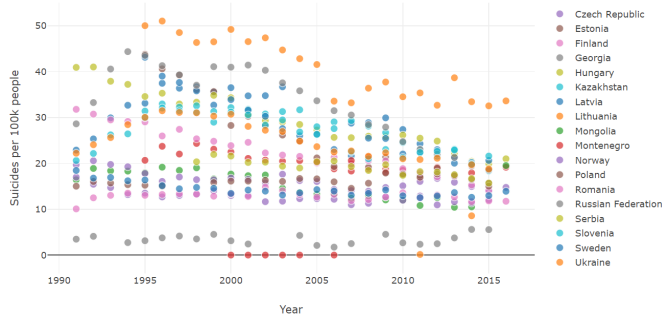

Fig. 8. Comparison Scatter Plot (Selected Countries)

We obtain the scatter plot in Fig. 8 by selecting the group of countries on the world map. This plot can be used to compare the countries within the region to identify their suicide trend over the past 25 years. This shows over the years that the suicide rate is steady or mostly decreasing with Lithuania and Russia contributing to the highest values. While Georgia and Bosnia contribute to the lowest rates, potentially because we have less data available on them.

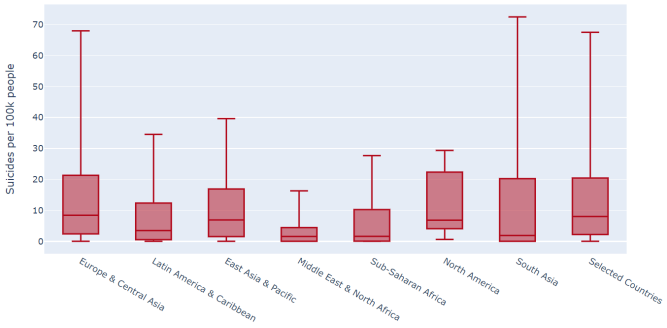

Fig. 9. Box-and-Whisker Plot (Selected Countries)

The box plot (Fig. 9) indicates this by showing that the regions "Europe & Central Asia", "East Asia & Pacific" and "North America" have a higher median suicide rate compared to the other regions.

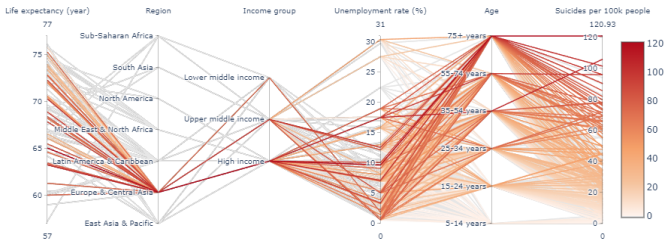

Fig. 10. Parallel Coordinate Plot (Selected Countries

The grey traces in the parallel coordinates plot (Fig. 10) show that the deselected countries mainly belong to a lower suicide rate on the range.

4.1.2 Which age groups and gender contribute to the highest suicide rate within these countries?

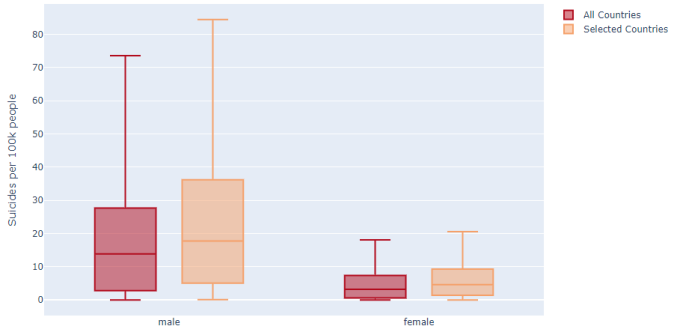

Fig. 11. Country Scatter Plot (Gender Based)

We see that the men in the European region dominate the suicide rates as opposed to women (Fig. 11). Women seem to nearly have one-third the suicide rate of men approximately over the years (Fig. 12).

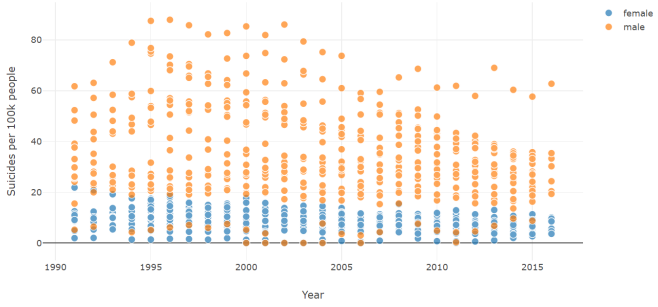

Fig. 12. Country Scatter Plot (Gender Based)

4.1.3 Do countries in which euthanasia is legal have higher suicide rates?

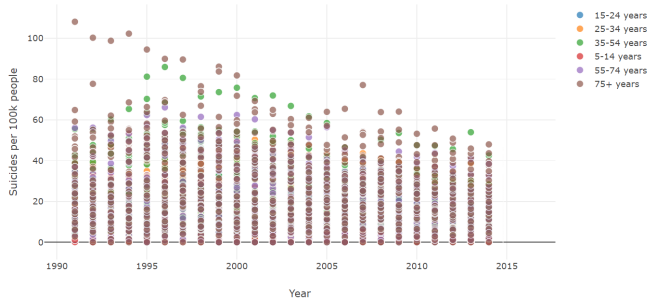

Fig. 13. Country Scatter Plot (Age Based)

Looking at the scatter plot (Fig. 13), there is mainly one country, age combination sticking out. These are people from Lithuania 75 years or older. The likely reason for this is that euthanasia is illegal there, older people commit suicide instead.

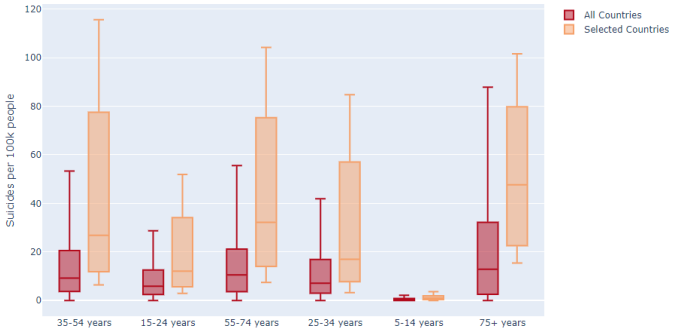

Fig. 14. Box-and-Whisker Plot (Age Based & Euthanasia Illegal)

Looking at countries where euthanasia is illegal (like Lithania and Russia) (Fig. 14), It is clear that the medians are different for older people, which supports the reasoning that this is potentially because of euthanasia laws.

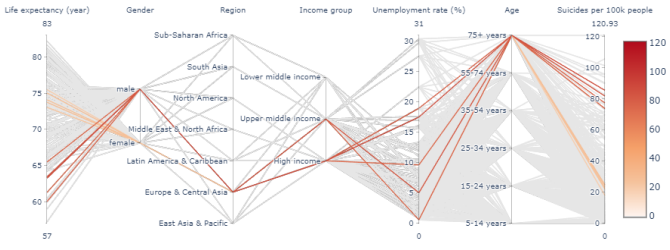

Fig. 15. Parallel Components Plot (75+ Years Old in Countries where euthanasia is illegals)

Lastly, the parallel coordinated plot shows that 75+ year old people in countries where euthanasia is illegal, have a higher suicide rate.

4.2 What indicators can be found that contribute to the highest suicide rate?

The most prominent indicators found so far are age, namely older people in countries where euthanasia is illegal commit suicide relatively often, and gender (male suicide rate is higher than the female suicide rate). Additionally over the years suicide has slightly declined overall. Countries with a high income have a relatively high suicide rate. Similarly, a low unemployment rate results in a lower suicide rate, however most likely this is just a correlation and not causation, since is is reasonable to assume unemployment rate is directly related to the income group of a country.

5 Discussion

Our approach primarily made use of drop-downs, multi-select boxes, range sliders and check-boxes. We used a horizontal design approach with no vertical panel but rather all options placed in a row format. This helped us to create a neat horizontal design on a single page without any scrolling required to view all four plots. This enabled us to interact with any of the plots and see them all update together without any scrolling. The geographical plots fitted our data perfectly as the observations in our data-set were based on countries. All the multi-selects, check-boxes and range slider was pre-selected to show the user the overview of the entire situation surrounding suicides through four plots. Based on his requirements, he could remove the filters he does not need to see the relationships he may be interested to find out such as suicide rates for people aged 75+. We believe that this design and approach was quite convenient to use, showing various depth if required and neat interaction between all the plots.

However, one of the limitations we faced greatly was to fit more than 4 plots on a single page dashboard without compromising their size which would have affected their view-ability. Similarly, if kept the same size and added more graphs then the page would require scrolling. We would have liked to include more region based techniques and even some more basic visualizations such as bar charts to get a better overview. These could be some of the potential improvements which could be made to the dashboard in future development.

6 Conclusion

The main contributions was to identify the trend of the suicide rate over the last 25 years. We saw that the European region contributed consistently to higher suicide rates over this period of time. The lowest rate over the years was in Sub-Saharan Africa and Middle East Region. Now this may be due to having less data available, poor data collection in those countries or even some other factor such as dominant religious beliefs which may be a factor worth investigating in future analysis.

Additionally, we saw that Russia and Lithuania have had the highest suicide rates on average over this period of time, however, with the trend indicating that the suicide rates have been decreasing year on year. These and other European countries also showed that people aged 75+ and 35-54 contributed to the majority of the suicides. This possibly is due to the fact that euthanasia is illegal in some European countries. The people aged 35-54 would be classified as working class individuals, and these compromise of males mostly. This is possibly because more males are employed than females throughout history. For unemployment rate, we were unable to find significant relation with respect to the suicide rate as some countries with higher unemployment rate showed lower suicide rate whilst others did not. This was inconclusive. Similarly, for life expectancy no significant patterns were observed either.

This analysis and visualization will help us to understand the historical situation of suicides across the world and enable organizations and societies to understand how much work needs to be done to help them prepare to prevent suicides. The insights gained from this can be useful for suicide prevention programs and organizations. The World Health Organisation also provides global data on other related topics, such as mortality, cause of death and statistics on children that might be worth investigating in a future study. [11].

Work Division

The work was divided equally between both team members. Bazil worked on the world map, click, lasso interaction, scatter plot and integrating all their respective filters, the year slider and a few drop-downs. Luc worked on box plot, parallel coordinate plot and connecting them to the scatter plot and world map and making the layout of the dashboard.

REFERENCES

[1] A. Cockburn, A. Karlson, and B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.*, 41, 12 2008. doi: 10.1145/1456650.1456652

[2] S. Harikrishnan, P. Jeemon, G. Mini, K. Thankappan, and P. Sylaja. Gbd 2017 causes of death collaborators. global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the global burden of disease study 2017. 2018. doi: 10.1016/S0140-6736(18)32203-7

[3] J. Heer, M. Bostock, and V. Ogievetsky. A tour through the visualization zoo. *ACM Queue*, 8:20, 01 2010. doi: 10.1145/1794514.1805128

[4] W. H. Organization et al. *Preventing suicide: A global imperative*. World Health Organization, 2014.

[5] A. J. Pumariega and N. Sharma. *Suicide Among Diverse Youth*. Springer, Cham, 2018. doi: 10.1007/978-3-319-66203-9

[6] Rusty. *Suicide Rates Overview 1985 to 2016*. Kaggle, 2018. Accessed on: December 22, 2019 [Online]. Available: `http://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016`.

[7] A. Sarikaya and M. Gleicher. Scatterplots: Tasks, data, and designs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):402–412, Jan 2018. doi: 10.1109/TVCG.2017.2744184

[8] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pp. 336–343, Sep. 1996. doi: 10.1109/VL.1996.545307

[9] United Nations. *Human Development Data*, 2018. Accessed on: December 22, 2019 [Online]. Available: `http://hdr.undp.org/en/data`.

[10] The World Bank. *World Development Indicators*, 2019. Accessed on: December 22, 2019 [Online]. Available: `http://datacatalog.worldbank.org/dataset/world-development-indicators`.

[11] World Health Organization. *Disease burden and mortality estimates*, 2016. Accessed on: December 22, 2019 [Online]. Available: `https://www.who.int/healthinfo/global_burden_disease/estimates/en/`.

[12] World Health Organization. *Suicide Data*, 2019. Accessed on: December 22, 2019 [Online]. Available: `http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/`.