Visual analysis of socio-economics effects on suicides, over the years, in different countries and generations

1st Manuel Ivagnes
Engineering in Computer Science
Sapienza University of Rome
Rome, Italy
ivagnes.1698903@studenti.uniroma1.it

2nd Riccardo Bianchini
Engineering in Computer Science
Sapienza University of Rome
Rome, Italy
bianchini.1700170@studenti.uniroma1.it

Abstract*—Mental heath: "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community". Nowadays, not a week passes without a story in the press about the impact of covid-19 on mental health and more specifically on suicides. Claims on social media seem to appear daily and discussions on these topics are getting a huge attention. Therefore, we want to analyze the socio-economics effects on suicides, over the years[1], in different countries and generations, so that to better understand some of the reasons behind this illness. The repo containing all the material is accessible at the following link: *https://github.com/BCI-ECS-sapienza/Suicide-rate-visualization

***Index Terms*—Visual Analytics, Mental Health, Suicides**

I. Introduction

Mental health is an integral and essential component of health. The WHO constitution states: "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity". An important implication of this definition is that mental health is more than just the absence of mental disorders or disabilities [1].

In recent years, there has been increasing acknowledgement of the important role mental health plays in achieving global development goals, as illustrated by the inclusion of mental health in the Sustainable Development Goals. Depression is one of the leading causes of disability. Suicide is the second leading cause of death among 15-29-year-olds. People with severe mental health conditions die prematurely – as much as two decades early – due to preventable physical conditions.

Despite progress in some countries, people with mental health conditions often experience severe human rights violations, discrimination, and stigma. Many mental health conditions can be effectively treated at relatively low cost, yet the gap between people needing care and those with access to care remains substantial. Effective treatment coverage remains extremely low [2].

While the link between suicide and mental disorders (in particular, depression and alcohol use disorders) is well established in high-income countries, many suicides happen impulsively in moments of crisis with a breakdown in the ability to deal with life stresses, such as financial problems, relationship break-up or chronic pain and illness. However, suicide does not just occur in high-income countries, but is a global phenomenon in all regions of the world. In fact, over 77% of global suicides occurred in low- and middle-income countries in 2019 [3].

Unfortunately, in this article, we cannot analyze all the reasons behind mental disorders or mental health conditions in general. However, our visual analytics system wants to give a first insight into how the different countries' socio-economic conditions can influence the number of suicides. We accomplish this task by providing the user with a complete interactive visual analytics environment, in which they can filter the data as needed, by simply selecting the corresponding components.

II. Related Work

Analyzing this phenomenon, we observed that extensive literature is devoted to studying and analyzing mental health conditions and suicides. Therefore, as the first step, we want to understand how can this work locate concerning the existing literature.

The first paper that we want to take into consideration is the *"Archives of Suicide Research"* [4], that uses the WHO databank to analyze suicides in almost 100 different countries, taking into particular consideration the role that biological sex take in this phenomenon.

The study shows that suicide rates among the countries differ widely, and since the beginning of the official registration, Hungary has been the country with the highest suicide rates in Europe (if not in the World). However, more recent data shows that Hungary is being surpassed by some new Russian and Baltic states.

Nordic and Eastern European countries also have somewhat higher suicide rates, while the southern parts of Europe have

[1] Unfortunately, we did not find any suitable/updated dataset to analyse the panemic effect. Our project will instead cover past years, that still consider less relevant crisis

comparatively low suicide rates. America and Asia generally have lower rates than most of the European countries.

The male rates are, with one exception (China), in all countries - even the countries with very low rates - higher than the female rates.

Finally, the article points out that the quality and reliability of official mortality reporting varies tremendously among the various countries. Therefore, the available figures give only a rough picture of the situation of the problem in the world.

Thus, this study gives some interesting insights and approaches that we can use, compare and extend in our work. First of all, we can notice that the data used inside this paper are not updated. Moreover, there is no clear distinction between values during different years and peer groups. Therefore, following the same approach, we can extend this survey, giving also the opportunity to analyze the data in more details, distinguishing by years and peer groups.

The second paper that we found particularly related to our research is "Worldwide impact of economic cycles on suicide trends over 3 decades: differences according to level of development. A mixed effect model study" [5], that directly compares the suicides rates with the GDP of the different countries. It also manages to refine the direction of the correlation between suicide rates and GDP depending on the overall trends of economic expansion versus recession and distinguishing between developed and developing countries.

The study examined worldwide trends and correlations of PPP-adjusted GDP per capita and suicide rates in 10 WHO subregions during a period of about 30 years (from 1980 to 2007). PPP-adjusted GDP increased over the study period in all regions examined with the exception of the EUR C[2] countries. Suicide rates increased in developing Latin American and Caribbean countries (AMR B[3]) and high-income Asian countries (WPR A[4]), while they decreased in most European countries and Canada.

The article shows that GDP is strongly correlated to suicide rates worldwide and that the direction and magnitude of the correlation differs between developing and developed countries. Moreover, the paper suggests that prevention strategies should be tailored to each WHO region and PPP-adjusted GDP per capita offers a simple measure to help decide the type of preventive measure that is more suitable across countries. In order to reduce population suicide rates, macroeconomic public health interventions - for example, the provision of basic needs, the reduction of socioeconomic inequalities - might be more suitable for developing countries where socioeconomic and cultural factors appear to play a major contributing role in suicide. In high-income countries, where the medical model prevails and suicide is understood as an unfortunate consequence of psychiatric illness - particularly depressive disorders - preventive measures based on the medical model - for example, incrementing the number of psychiatric or counselling services - might prove more fruitful in helping to decrease the daunting suicide rates worldwide.

Thus, following this approach, we want to provide visualizations that reveal the relationship between GDP and suicide rates. Moreover, we also wish to show how increase and decrease in general income influence this illness.

In conclusion, we found plenty of factors influencing such behaviour and a large amount of paper and studies dealing with mental health and suicides. However, in this project, we want to keep our analysis limited only to those elements that we found to be the most relevant and easily observable. In the final chapter - Future work - we will give some insights on how to extend this work to provide a more comprehensive analysis.

III. Dataset

Before we started implementing our system, we needed a lot of information about suicides over the world and therefore we took the *Suicide Rates Overview 1985 to 2016* dataset from Kaggle [6], that was in turn taken from the World Health Organization databank [7].

In fact, this dataset was built to find signals correlated to increased suicide rates among different cohorts globally, across the socio-economic spectrum in an interval of time starting from 1985 to 2016. It contains 27.8k valid entries with the following attributes:

- *Country*: the country to which the tuple is referred;
- *Year* the year to which the tuple is referred;
- *Sex:* the biological sex of the people taken into consideration in the tuple;
- *Age:* the age group of the people taken into consideration in the tuple. There are six different classes: 05-14 years, 15-24 years, 25-34 years, 35-54 years, 55-74 years, 75+ years;
- *# Suicide_no:* number of suicides inside the "Country" during the relative "Year";
- *# Population:* population of the "Country" during the relative "Year";
- *# Suicides/100k pop:* suicides ratio that specifies the number of suicides every 100k inhabitants;
- *Country-year:* composite key with the name of the "Country" and the "Year";
- *HDI for year:* the Human Development Index is a static composite index of life expectancy, education and per capita indicators, which are used to rank countries;
- *Gdp_for_year:* is the GDP of the "Country" in the relative "Year";
- *Gdp_per_ capita:* is the average GDP per capita of the population of a "Country" in the relative "Year";
- *Generation:* is the generation relative to the age of the portion of the population taken in consideration in the tuple.

[2]Estonia, Hungary, Latvia and Lithuania

[3]Argentina, Brazil, Chile, Colombia, Costa Rica, El Salvador, Mexico, Panama, St. Lucia, Trinidad and Tobago and Venezuela

[4]Australia, New Zealand, Japan and Singapore

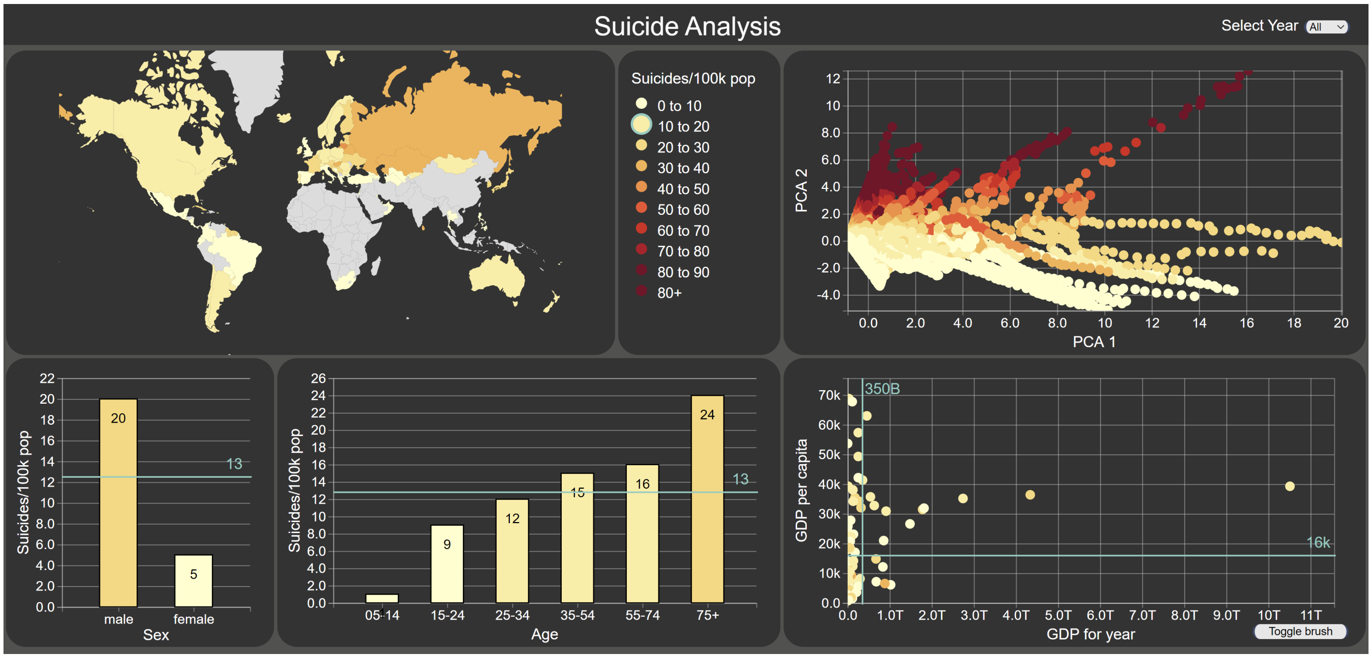

Fig. 1. Overview all data

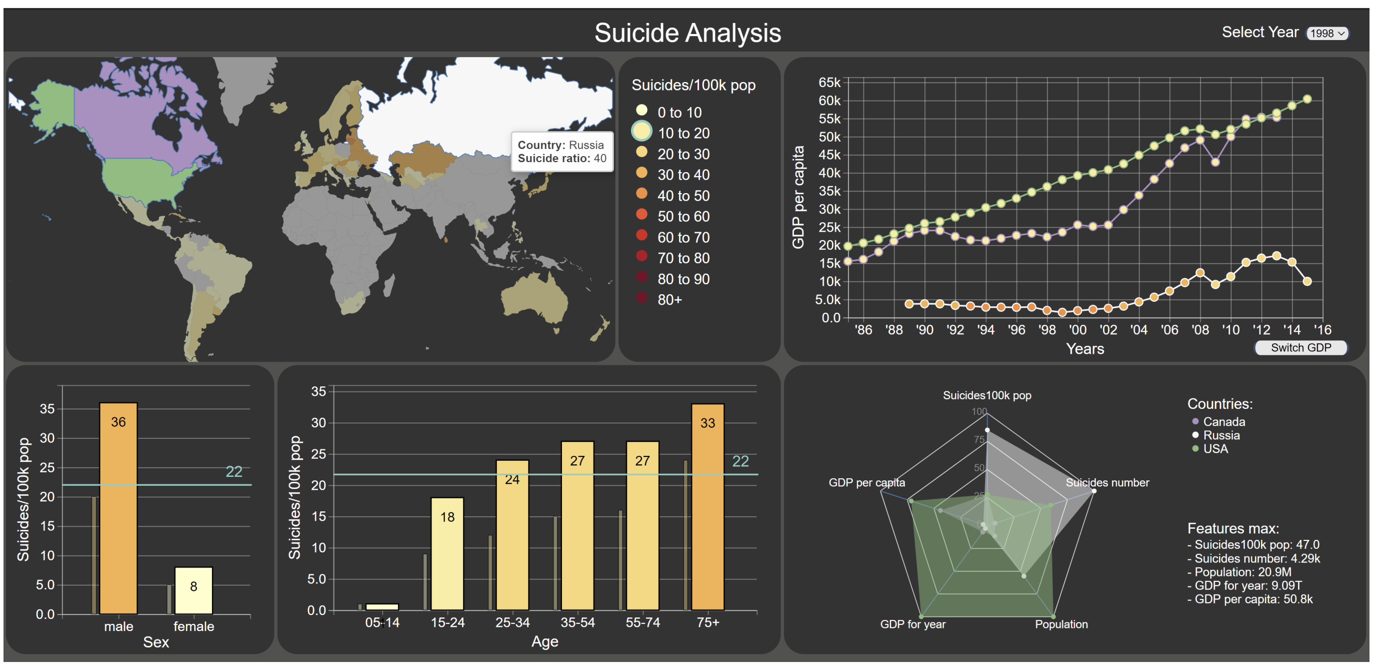

Fig. 2. Overview single countries

A. *Preprocessing*

The dataset, as provided by Kaggle, required some preprocessing to become entirely suitable for our purpose. We applied the following steps:

1) We checked for the presence of tuples with *null* values in one of the attributes and found that "HDI for year" has mostly null values. Thus, we decided to remove it;
2) We removed "Country-year" and "Generation" that are redundant for our purpose;
3) We modified the encoding of the name relative to the age groups so that to remove the suffix "year". This step is needed to avoid overlap on the visualization;
4) We searched for differences in the names of countries inside the csv file and the geojson used by the map. Then, we modified the names to have a perfect match. The name that we modified are:

- United States → USA
- United Kingdom → England
- Republic of Korea → South Korea
- Russian Federation → Russia
- Serbia → Republic of Serbia
- Bahamas → The Bahamas

B. *Dimensionality reduction*

Dimensionality reduction, or dimension reduction, is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension [8].

Specifically, for this task, we decided to use PCA (Principal component analysis), a linear technique for dimensionality reduction that performs a linear mapping of the data to a lower-dimensional space so that the variance of the data in the low-dimensional representation is maximized. Therefore, this method allows us to plot each multidimensional tuple on a bidimensional space, still maintaining all the underlying properties. The algorithm is applied on the following attributes: suicide_no, population, suicides_pop, gdp_for_year, gdp_per_capite. The results are saved inside the datasheet, under the attributes PCA_1 and PCA_2.

IV. Technologies

For the development of our project we used the following technologies:

A. *Python*

Python [9], along with the Jupyter notebooks [10] on Colab [11], has been used for the entire step of data preprocessing. Then, it has been used to perform the dimensionality-reduction algorithms, and finally, run the simple HTTP server that loads the service.

B. *Javascript and D3.js*

Javascript [12] and the D3.js framework [13] have been used for the development of the visualizations that compose the service.

C. *HTML and CSS*

These basic web technologies are used to render and beautify the visualizations.

V. Visualizations

The final service that we implemented is composed by a set of different visualizations. Therefore, in this chapter, we want to analyze each component individually to understand what it is showing and how it interacts with the others.

There are two main views that we are using:

- The first view shows data relative to all the countries. It is shown in fig. 1;
- The second view shows data relative to the selected countries only. It is shown in fig. 2;

A. *Interactive Legend*

In general, legends are used inside single visualizations to summarize distinctive visual properties such as colours or texture. However, in our case, this component is not a simple static item inside a single visualization. It is an interactive visualization itself. Indeed, it includes the following functionalities:

- On mouse over a legend circle, the Map will decrease the opacity of points that are not inside the range;
- On mouse over a legend circle, the ScatterPlot will decrease the opacity of points that are not inside the range;
- On mouse over a legend circle, the service shows a tooltip containing all the countries that are inside that range;
- When other visualizations update the values of suicide ratio, the blue outline used to denote the mean is updated;

The color encoding is used by: Choroplet Map, Sex Bar chart, Age bar chart, Scatterplot, PCA and Line chart. The legend is shown in fig. 3.

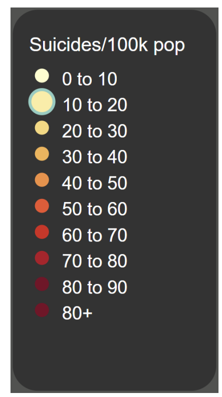

Fig. 3. Legend

B. *Choropleth Map*

A choropleth map is a type of thematic map in which a set of pre-defined areas, in our case the countries, is colored or patterned in proportion to a statistical variable that represents an aggregate summary of a geographic characteristic within each area [14], in our visualization we used the number of *suicides every 100k people*.

It is a simple view that allows to visualize in an effective way how suicides rates varies across the world and has the following interactions:

- It allows zoom and pan;
- On country selected, Age chart will show only the related deaths;
- On country selected, Sex chart will show only the related deaths;
- On country selected, PCA will be replaced by the Line chart that shows the Suicide radio over the years;
- On country selected, Scatterplot will be replaced by the Radar chart that shows all the main information about the selected country;

- On country selected, the other countries inside Map will reduce their opacity. We can select at most three countries;
- On mouse over one country, the service shows a tooltip containing all the related information;
- On mouse over one country, the relative circle on PCA chart increases its size and gets a blue stroke;
- On mouse over one country, the relative circle on Scatterplot increases its size and gets a blue stroke;

The Choropleth Map is shown in fig. 4.

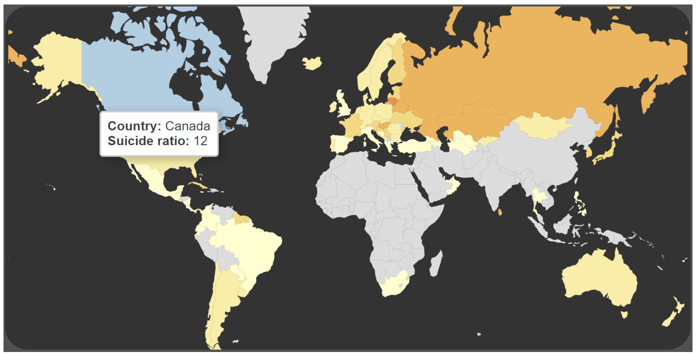

Fig. 4. Map

C. Scatterplot

A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. If the points are coded (color/shape/size), one additional variable can be displayed. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis [16].

In this case, we have *"GDP for year"* on the horizontal axis and *"GDP per capita"* on the vertical axis. Each point corresponds to a single country, and the colour depends on the values of the *suicide ratio*. It has the following interactions:

- On brush, Scatterplot will zoom over the selected points and the mean bars for the axis values are updated;
- On brush, Age chart will show only the related deaths;
- On brush, Sex chart will show only the related deaths;
- On brush, Map will show colours referring to related deaths;
- On mouse over one point, the service shows a tooltip contains all the related information;
- On mouse over one point, the relative country on Map becomes blue;
- On mouse over one point, the relative circle on PCA chart increases its size and gets a blue stroke;

Two bars denote the mean values for *"GDP for year"* and *"GDP per capita"*. These are updated any time there is a brushing interaction. The Scatterplot is shown in fig. 5.

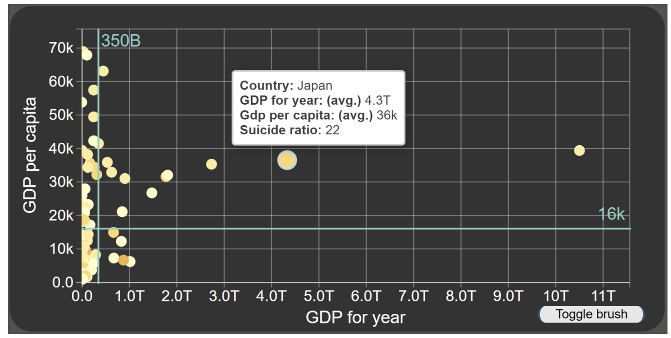

Fig. 5. Scatterplot

D. PCA chart

Same as the Scatterplot, but here the chart shows all the tuples inside the dataset. The plot is based on the values of *PCA_1*, on the X axis, and *PCA_2*, on the Y axis. The color of the points is given by their suicide rates.

This visualization can become very useful to find clusters and outliers inside the dataset. It has the following interactions:

- On mouse over one point, the service shows a tooltip containing all the related information;
- On mouse over one point, the relative country on Map becomes blue;
- On mouse over one point, the relative circles on Scatterplot increases its size and gets a blue stroke;

The PCA chart is shown in fig. 6.

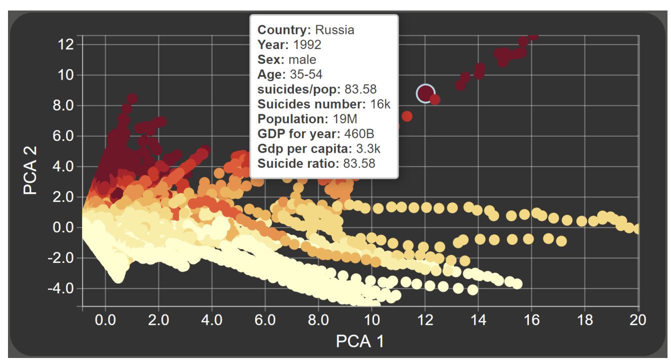

Fig. 6. PCA

E. Sex bar chart

A bar chart is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. Therefore, it is used to show comparisons among discrete categories. One axis of the chart shows the specific categories being compared, and the other axis represents a measured value [15].

In this case, we have the *biological sex* on the categorical axis and the *suicide ratio* on the other axis. The bars have the following interactions:

- On sex bar selection, Age bar chart will show only the related deaths;

- On sex bar selection, Scatterplot will show colours referring to related deaths;
- On sex bar selection, Map will show colours referring to related deaths;
- On mouse over one bar, the other bar shows the divergence to the current suicide ratio value;
- When the year selector has a value that is different than “All”, a slim bar is shown to compare the current data with the global data;
- Every time a filter is applied by other visualizations, the mean bar is updated;

The color of the bars depends on the value of suicide ratio and a blue bar denotes the current mean value. The Sex bar chart is shown in fig. 7.

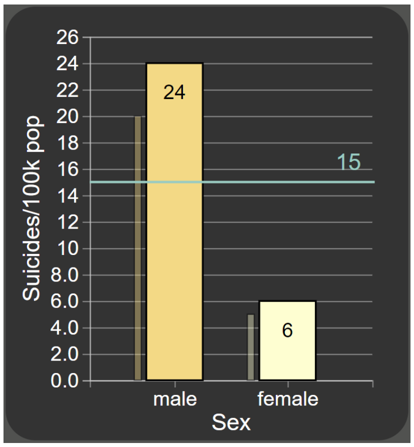

Fig. 7. Sex bar chart

F. Age bar chart

Same as the Sex bar chart, but here we have the *age group* on the categorical axis and the *suicide ratio* on the other axis. The bars have the following interactions:

- On age group selection, Sex bar chart will show only the related deaths;
- On age group selection, Scatterplot will show colours referring to related deaths;
- On age group selection, Map will show colours referring to related deaths;
- On age group selection, the average bar for only-selected-groups is updated;
- On mouse over one bar, the other bars show the divergence to the current suicide ratio value;
- When the year selector has a value that is different than “All”, a slim bar is shown to compare the current data with the global data;
- Every time a filter is applied by other visualizations, the mean bar is updated;

The color of the bars depends on the value of suicide ratio and a blue bar denotes the current mean value. The Age bar chart is shown in fig. 8.

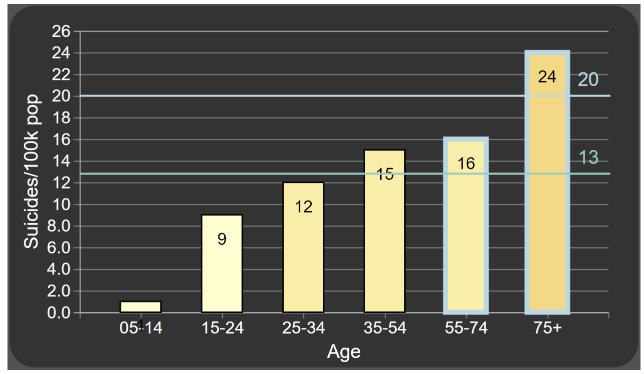

Fig. 8. Age bar chart

G. Line chart

A line chart is a type of chart which displays information as a series of data points called ’markers’ connected by straight line segments [18].

We can use this visualization to understand whether or not there is a link between *GDP* and *suicide rate* over the years. In fact, the years are represented on the horizontal axis , while GDP values (we can select between *GDP per capita* and *GDP for year*) are represented on the vertical axis. The colour of the markers follows the suicide rate encoding, while the colour of path and stroke is relative to the radar encoding for the countries.

It has a single interaction:

- On mouse over one point, the service shows a tooltip containing all the related information;

The Line chart is shown in fig. 9.

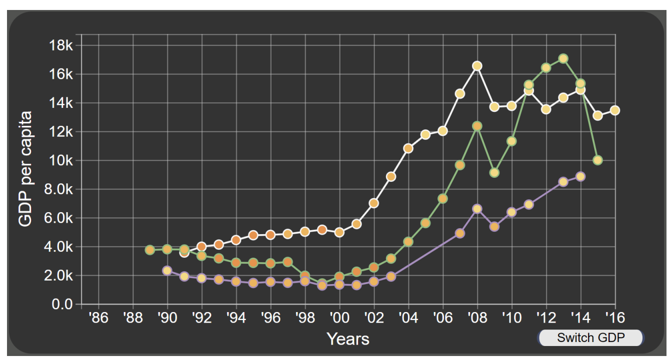

Fig. 9. line chart

H. Radar chart

A radar chart is a graphical method of displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point [17]. In our case we selected 5 variables: *population, gdp_per_capita, gdp_for_year, #suicides and suicides rates*. The relative position and angle of the axes is uninformative.

We used this kind of visualization to represent all the data of the selected countries in a unique graph and compare the defined variables between different countries.

On the top right of the container, we can visualize a legend containing the colours associated with the selected countries. While on the bottom right of the container, we can visualize the max value for each represented variable, based on the applied filters applied through the bar charts or by selecting a year. On mouse over, the opacity of the shape is increased to highlight it.

The Radar chart is shown in fig. 10.

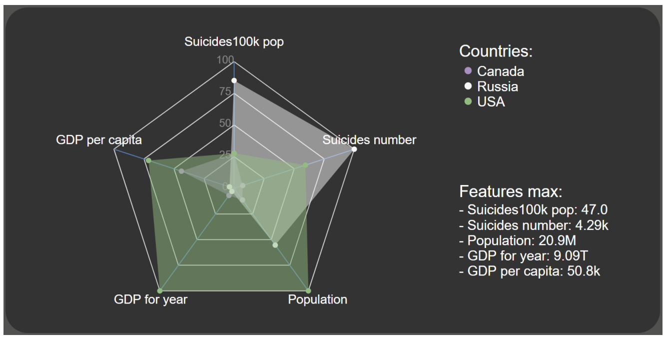

Fig. 10. radar

VI. CONCLUSION

In conclusion, our visual analytics system wants to support all those people who are analyzing this issue and want to understand how to mitigate this illness. Indeed, the simple interface provides the user with a complete interactive visual analytics environment, in which they can filter the data as needed and perform different complex analyses.

About the results, our outcomes are very similar to those we found in the literature. Suicide rates among the countries differ widely. Eastern European, Russian and Baltic states tend to have noticeably higher suicide rates than the rest of the world. The male rates are significantly higher than the female rates. Moreover, increasing the age, the suicide rates considerably rises.

The general country income may have a significant impact on this illness. However, we must consider that the reasons behind this correlation are much more complex than what we can summarize in this superficial analysis. Indeed, some low- and middle-income countries still have significantly lower suicide rates than high-income countries, even if, in general, high-income countries tend to have low suicide rates. The main other factors that we can analyze to better understand this issue are related to the mental health conditions, humanitarian crises and other forms of adversity that affect people in these countries.

VII. FUTURE WORK

In this final chapter, we want to discuss some possible elements that we might include to extend our system and provide a more exhaustive analysis.

- A comparison between suicides and other causes of death. Indeed, even if we found that younger people commit fewer suicides than older people, the percentage level, compared to other causes of death, is significantly higher in younger generations.
- Clinical studies definitively establish that psychiatric disorders are a major contributing factor to suicide. Therefore, it might be helpful to find the incidence of mental health conditions inside the population and the policies adopted by the countries to heal these conditions.
- A deeper analysis of living conditions inside the different countries. For example, we can consider income inequality, unemployment levels, social regulations, integration and so on.

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response
[2] https://www.who.int/health-topics/mental-health#tab=tab_1
[3] https://www.who.int/news-room/fact-sheets/detail/suicide
[4] Schmidtke, A., Weinacker, B., Apter, A., Batt, A., Berman, A., Bille-brahe, U., … Wasserman, D. (1999). Suicide rates in the world: Update. Archives of Suicide Research, 5(1), 81–89. doi:10.1080/13811119908258318
[5] Blasco-Fontecilla, H., Perez-Rodriguez, M. M., Garcia-Nieto, R., Fernandez-Navarro, P., Galfalvy, H., de Leon, J., Baca-Garcia, E. (2012). Worldwide impact of economic cycles on suicide trends over 3 decades: differences according to level of development. A mixed effect model study. BMJ Open, 2(3), e000785. doi:10.1136/bmjopen-2011-000785
[6] https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016?select=master.csv
[7] https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/mental-health
[8] https://en.wikipedia.org/wiki/Dimensionality_reduction
[9] https://www.python.org/
[10] https://jupyter.org/
[11] https://colab.research.google.com/
[12] https://www.javascript.com/
[13] https://d3js.org/
[14] https://en.wikipedia.org/wiki/Choropleth_map
[15] https://en.wikipedia.org/wiki/Bar_chart
[16] https://en.wikipedia.org/wiki/Scatter_plot
[17] https://en.wikipedia.org/wiki/Radar_chart
[18] https://en.wikipedia.org/wiki/Line_chart