EU causes of Death

Sapienza University of Rome
Visual Analytics

Greta Aquilina
Daniele Calisi
Simone Giordano

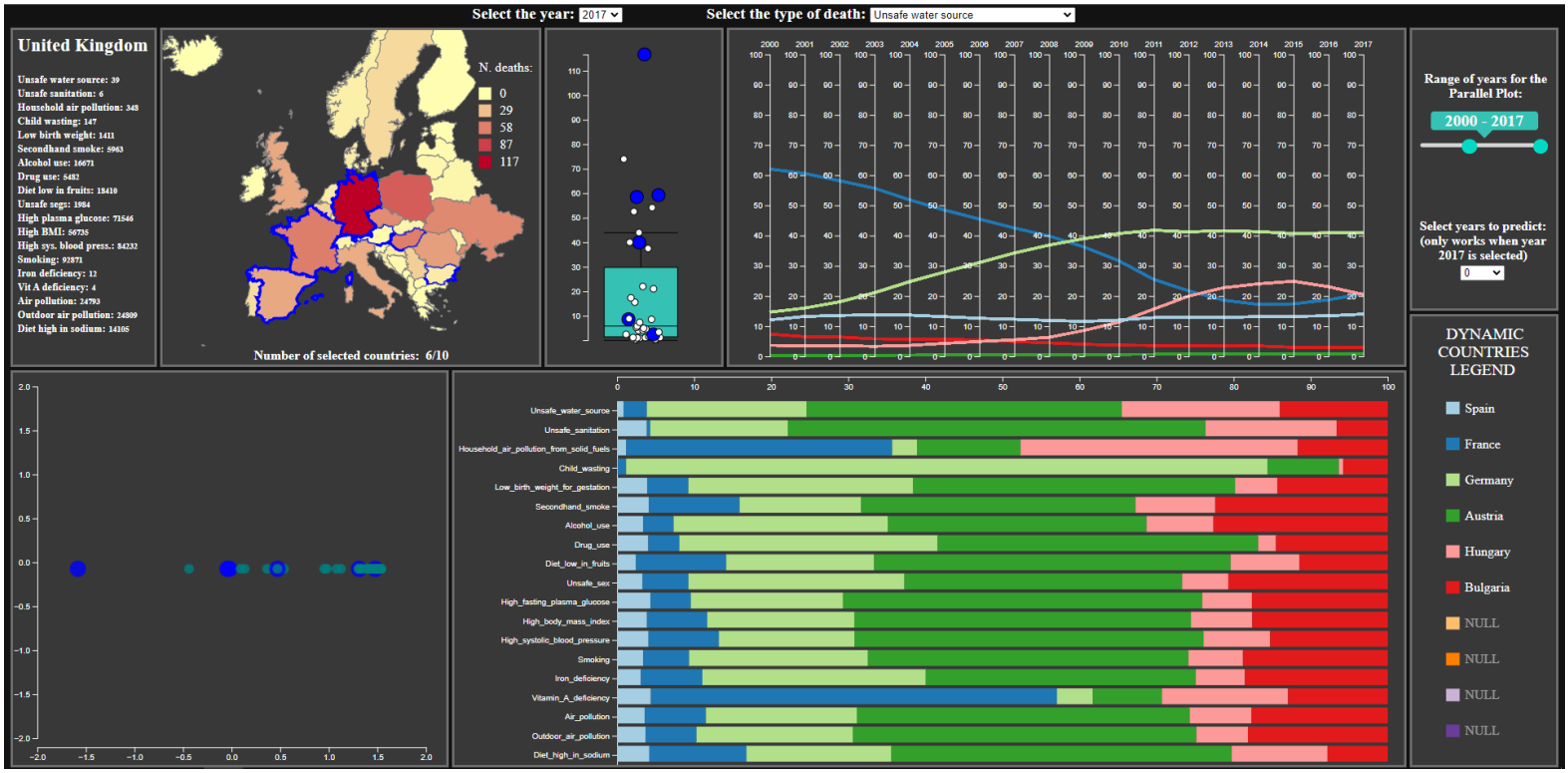

Figure 1: Whole tool

Contents

1. Introduction

The death trend for the last twenty years represents an interesting object of study. Through its analysis it is possible to predict future trends of a specific cause of death; it also allows a better understanding of past critical situations and correlations between different causes of death.
Our tool uses a refined **dataset** about multiple causes of death from 1990 to 2017 and provides different visualization techniques to improve the comprehension of the trend of a single or multiple causes of death in geographic Europe. The tool is intended for big **users** like the European Commission to visualize death trends in multiple countries of Europe. Indeed, the tool alone is great for a first analysis, but can be even better if integrated with other statistical datasets (es: population, economy, etc..), to understand a certain death trend with respect to other phenomena. Additionally, the tool is intended for individual users, for example someone who is moving out, that might be interested in an intuitive analysis of multiple causes of death in countries of Europe.
The tool is composed by the following **visualizations**, explained in detail in the respective chapter:

- **Map of geographical Europe**: to provide an intuitive accessibility to the data of individual countries for a given year, or to easily allow a multiple selection.
- **Boxplot:** to visualize, for a given year and cause of death, the distribution of data as well as outliers.
- **Parallel plot:** to show, for a given cause of death, its European trend during a selected years range.
- **Bar chart:** to provide a comparison between countries for each cause of death in a given year.
- **MDS**: for a selected cause of death and year, it shows the dissimilarity between each country.

In this section are briefly explained how we adopted some visualization techniques and the **design process** that led us to use them.

Moreover, the tool provides different types of **interactions** to the user, some of which can trigger specific **analytics**: recompute the MDS over different parameters or calculate a prediction for the trends in the parallel plot.

2. Dataset

The **original dataset** (6237 x 31) that we used as a starting point is available on Kaggle website [1]. The data were about 29 different causes of death, for 231 countries in the world, covering years from 1990 to 2017.

2.1 Data Pre-processing

Since our focus was geographical Europe, we started to shrink the original dataset by removing all the countries that were not in our target; the remaining ones were 39 European countries. After that, we cut out a subset of the initial causes of death, because we thought they were not relevant for our case or because too many values were missing. The 10 excluded columns were: 'No access to handwashing facility', 'Non-exclusive breastfeeding', 'Discontinued breastfeeding', 'Child stunting', 'Diet low in vegetables', 'Low physical activity', 'High total cholesterol', 'Diet low in whole grains', 'Diet low in nuts and seeds', 'Low bone mineral density'.
Therefore, the **refined dataset** is composed by 1092 rows and by 21 columns, leading to the following AngeliniSantucci index:

AS = 1092*21=22932

3. Visualization Techniques

3.1 Menus and Sliders

To start we needed an easy and intuitive way to make the user choose the year and the cause of death to consider. Thus, in the upper section of the page it is possible to select them from their respective menus. By changing the

reference year, all the main visualizations but the Parallel plot (because it shows an individual range of years) are influenced; while changing the cause of death to consider, all the main visualizations but the Bar chart (because it always shows all the causes of death) are influenced.
Moreover, we also used a menu to let the user select the number of years to predict in the parallel plot.
A slider instead, was our choice to let the user set the range of years to consider in the parallel plot. Indeed, by simply moving two circle pickers, the user can change the plot in real time.

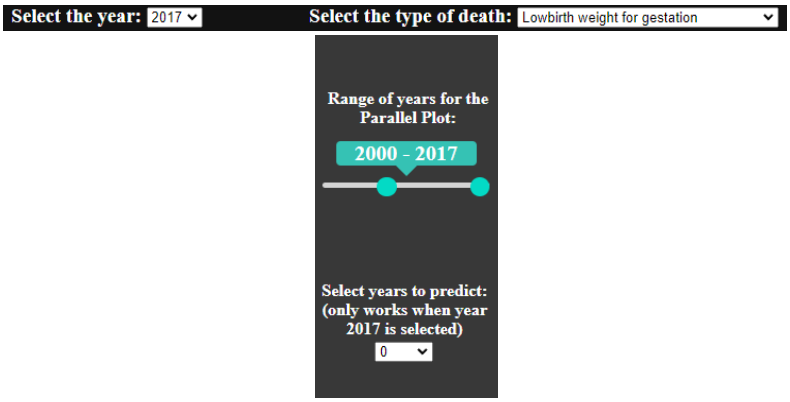

Figure 2: menus and sliders

3.2 Map of geographical Europe

This is the first main visualization that we thought of for our case, because we considered it a fundamental and simple way to show and interact with the countries.
For the selected year and cause of death, the map shows the number of deaths through a **sequential colour scale**. For the colour scale we choose a standard one, provided by *ColorBrewer* website [2]. Thus, each country in the map will be coloured according to the displayed scale giving the user a quick glance on the European situation.
If the user **hovers over a country**, the toolbox on the left side of the map, that shows the number of deaths for each cause, is updated to show the values about the hovered country.
Furthermore, it is possible to **select multiple countries** (up to ten) on the map by using the "CTRL" command and then clicking on them. The multiple selection triggers changes in the Bar chart and Parallel plot, i.e. making them show only the selected countries.
This visualization is **bidirectional coordinated with the MDS**, indeed the countries selected in the Map are selected and highlighted also in the MDS and vice versa.

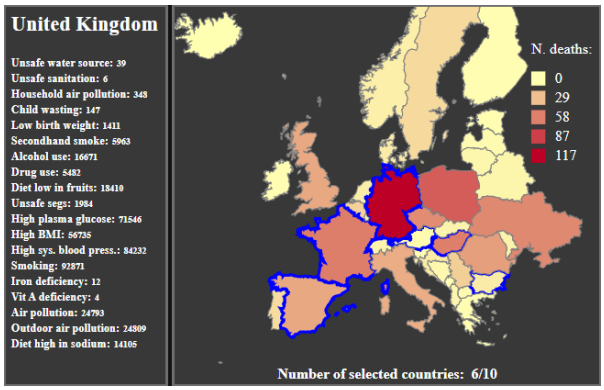

Figure 3: Map visualization

3.3 Box plot

To show the **distribution of our data** in a relatively small space we decided to use a Boxplot, which is based on a "five number summary": "minimum", first quartile (Q1), median, third quartile (Q3), and "maximum".

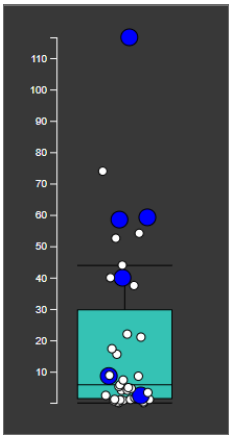

Figure 4: Boxplot visualization

Given the selected year and cause of death, these five values and the **outliers** are represented in a compact way; this visualization indeed excels in showing if the data is symmetrical and how tightly the data is grouped. It is also possible to visualize on the boxplot the individual data points (the quantitative values for each country), in this way it is easier to visualize the outliers and the range of values where the data are grouped. Moreover, when selecting one or more

countries from Map or MDS, the respective data points in the Boxplot are highlighted.

3.4 Parallel plot

To show **how the death trends evolved in time**, we decided on a Parallel plot. In the chart, for a selected cause of death and for the selected countries on the Map/MDS, it shows the behaviour along the years, thus there is a vertical axis for each year. Because of the different order of magnitude of the quantitative values for different causes of death, any attempt of comparison between these values would be misleading; therefore, we proceeded with the data **normalization** prior to plotting the chart. In this way, since each vertical axis has the same scale, it is possible to compare values of different death causes. The countries are distinguished with **categorical colours**, taken from ColorBrewer. These colours are assigned dynamically to the selected countries and shown in a legend on the right side of the screen.
The range of years to consider is selected by a slider, as said in the previous subchapter. Moreover, the tool gives the possibility to predict the values of death for the next few years, as explained better in the analytics section.

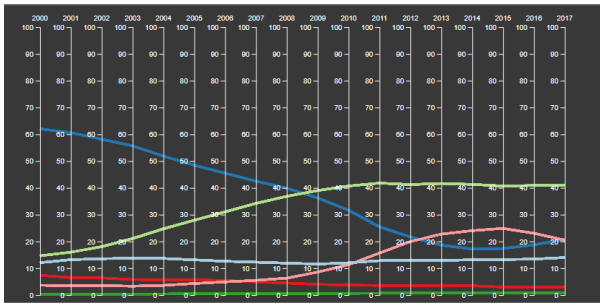

Figure 5: Parallel plot

3.5 Multidimensional Scaling

Multidimensional scaling is a visualization technique used to visualize in 2-dimensional space the **(dis)similarity** among objects. The starting point of our MDS was to build the dissimilarity matrix to determine how dissimilar from each other the couples of objects were. More specifically, in our MDS the objects are the countries, represented as green points. While for a given cause of death and year, the distances between each point represent the dissimilarity in number of deaths. In the MDS there is also the possibility to select up to 10 countries of interest by **brushing**. This selection is **coordinated**, as said before, with the one in the Map. Therefore, selecting multiple countries will trigger: the highlighting of the countries both in the map and in the Box Plot, the updating of the considered countries in the Bar chart and in the Parallel Plot.

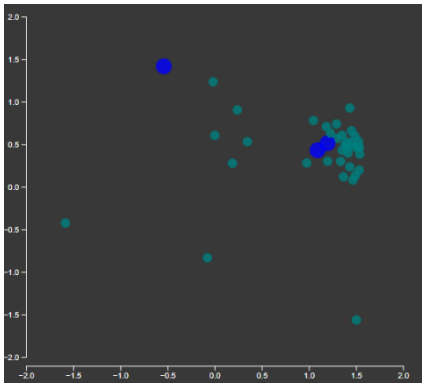

Figure 6: MDS

3.6 Horizontal Stacked Bar chart

We chose this visualization technique because we needed some way to let the user **compare the contribution**, in terms of the number of deaths, **of individual countries** for specific death causes; and at the same time a way to have an idea of this contribution for all the causes of death.
Due to the different order of magnitude of our data, we converted the initial quantitative values in a percentage scale to provide a better readability and to facilitate the **horizontal comparison** of the data.

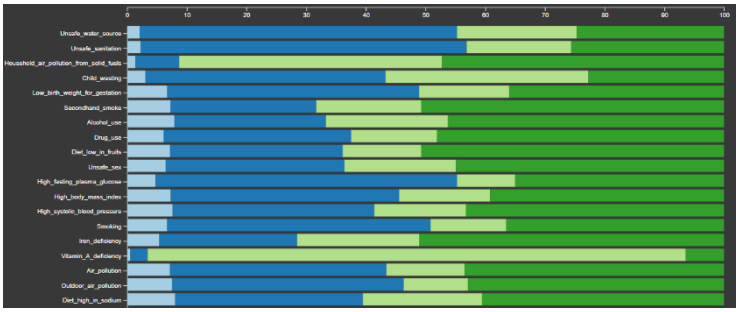

Figure 7: Bar chart

On the vertical axis we represented the different categories of data, which are the causes of death. For each cause of death, the

horizontal bars display the percentage of deaths, with respect to the total number of deaths of the considered countries, for each selected country.

4. Interactions and Analytics

In the previous section we described some interactions available to the user in different visualizations; in this section we will focus on the specific **interactions that trigger analytics**, with a short description of those analytics.

A change in the year or the cause of death, through menus, will lead to a change in the multidimensional scaling. This means that each time the user modifies one of these two parameters, the **MDS is recomputed in real-time**, and its new version is shown. To compute the MDS, we started by computing the proximity matrix, computing for each country and for the given death, the difference of the number of deaths; of course obtaining that the distance value between a country and itself, is zero. Once the computation of the proximity matrix is done, its values are passed to the function in charge of creating the MDS with a standard procedure.

Another interaction regards the parallel plot: if 2017 is the last year selected, on the right side of the chart there will be shown the possibility to **predict the percentages of death for up to three years**. Selecting the number of years to predict triggers the computation of the prediction, for the given cause of death, for each additional year. This computation updates the parallel plot so that it shows the range of years selected plus the new values. To predict the value of a year, the algorithm proceeds in the following way: given the selected range of years, for each pair of years, the algorithm will compute the difference between the values of deaths of the second and first year in the pair; the computed differences are summed together and are divided by the number of pairs of years. The final result is an index, and it may have a negative or positive value. This index will be added, for each country, to the number of deaths of the previous last year considered.

5. Related Works

There are multiple papers and informational websites on that examine the type of deaths in the European Union. Here are reported three **websites' articles related** with our work:

- Euromomo [3]: Graphs showing the pooled weekly total number of deaths in the data-providing Euromomo partner countries and subnational regions, all ages and by age groups.
- OECD-lib [4]: Two graphs in this case. The first one is a graph that shows the main causes of deaths just for one year and for one country (France), the second one shows the causes of deaths among all EU countries.
- EUROSTAT: Data explorer of the main causes of deaths among all European countries with a greater propensity to a better view of the data, as example [5].

Regarding **related papers**, it is particularly relevant "Socioeconomic Inequalities in Health in 22 European Countries" [6], that shows how in almost all countries, the rates of death and poorer self-assessments of health were substantially higher in groups of lower socioeconomic status. Inequalities in mortality were small in some southern European countries and very large in most countries in the eastern and Baltic regions. These variations among countries appeared to be attributable in part to causes of death related to smoking or alcohol use or amenable to medical intervention. This can be confirmed using the developed tool with the selection of different countries and visualized in the graphs, beyond the map visualization.

6. Discovered Insights

While using our tool, we discovered some interesting insights which are shortly reported in this section.

1- We noticed that before the 1997, among the majority of the European countries, also the ones of the 1st world, there was a constant increasing in deaths by "unsafe sex". Instead,

after 1997, all the European countries, except the ones in the western Europe, went through a constant decrease of the number of deaths for that specific type of death. By deepening this observation, we discovered a paper [7] that confirms that this decrease is due to the fact that on 14/10/1997, a pharmaceutical company found a cure, or better, a way to decrease the symptoms provoked by the AIDS illness, that, at least in the beginning, was really expensive and available just for the wealthier countries such as the ones in the western Europe.
2- For some causes of death, the countries with most deaths have remained constant through the years; for instance if we consider the deaths caused by smoking, the middle Europe countries remain constantly high through the years.
3- Remaining on smoking, since it affects, among the various factors, the blood pressure and heart rate, an intuitive insight that we observed was that the countries with the highest number of deaths by smoking are quite the same as the ones for high systolic blood pressure.
4- For some types of death we observed a progressive shift of the countries with the highest number of deaths. As an example, considering the number of deaths caused by unsafe water sources, in the 1990 the countries with the highest values of death were the ones in the eastern side of Europe, however over the years we noticed that the countries involved in the western part increased. This spread is not surprising since over the years, as suggested by researchers, climate changes and overpopulation are devastating the water supply in wider areas around the world.

7. Conclusions

Overall, we are satisfied with the outcomes of the entire project. The tool provides information through different techniques of visualizations, allowing different point of views on the analysed data, thus we believe that the final outcome gives a comprehensive overview of the dataset. We introduced this tool to give a better understanding of the several causes of deaths in Europe in order to support a better prevention in the future years. Moreover an individual user can investigate what are the main causes of deaths in each European countries and the dissimilarities among all of them. Finally, some interesting features that may be added in the future can be the following ones:

- Providing better scalability on multiple screens resolution
- Extending the tool focus not only on the geographical Europe but on the entire world: in this way we can provide a worldwide vision and then through the map the users can focus on specific continents. As a consequence it will be required to extend the dataset to the original one.

8. References

[1] Kaggle dataset: https://www.kaggle.com/varpit94/worldwide-deaths-by-risk-factors

[2] ColorBrewer website: https://colorbrewer2.org

[3] Euromomo: https://www.euromomo.eu/graphs-and-maps

[4] OECD-lib: https://www.oecd-ilibrary.org/sites/82129230-en/1/3/2/1/4/index.html?itemId=/content/publication/82129230-en&_csp_=e7f5d56a7f4dd03271a59acda6e2be1b&itemIGO=oecd&itemContentType=book

[5] EUROSTAT Causes and occurrence of deaths in the EU: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20190716-1

[6] "Socioeconomic Inequalities in Health in 22 European Countries", Johan P. Mackenbach, Ph.D., Mall Leinsalu, Ph.D., and Anton E. Kunst, Ph.D. for the European Union Working Group on Socioeconomic Inequalities in Health*, June 5, 2008

[7] Decline in the AIDS and death rates in the EuroSIDA study: an observational study: https://www.sciencedirect.com/science/article/abs/pii/S0140673603138020