UNIVERSITÀ DEGLI STUDI DI MILANO
BICOCCA

DEPARTMENT OF INFORMATICS, SYSTEMS AND COMMUNICATION

Master Degree in Data Science

Data Visualization Project

A temporal and geographical analysis of the primary causes of death in the world

PAOLO CAGGIANO - DAVIDE GIARDINI

Contents

1 Introduction

In recent years the life expectancy in the world is growing due to many reasons. It is clear that the situation is not equally the same in the different areas of the world. As the Life Expectancy research by Max Roser, Esteban Ortiz-Ospina and Hannah Ritchie [1] points out:
Life expectancy has increased rapidly since the Age of Enlightenment. In the early 19th century, life expectancy started to increase in an unequal way, favouring the industrialized countries while remaining low in the rest of the world. Over the last decades this global inequality decreased: now no country in the world has a lower life expectancy than the countries with the highest life expectancy in 1800. Since 1900 the global average life expectancy has more than doubled and is now above 70 years.

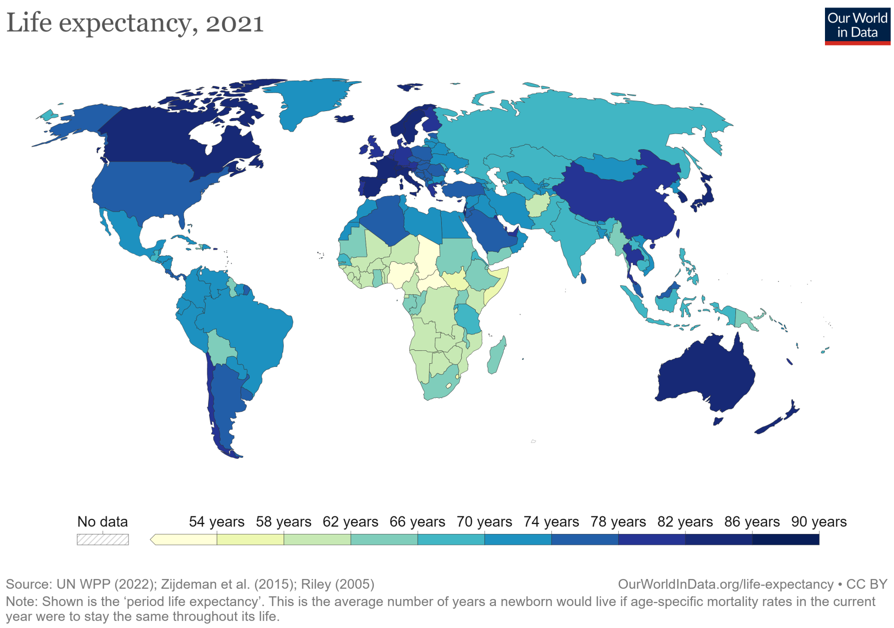


Figure 1: Life Expectancy in the World (2021)

As the previous graphic shows, there are regions, like Australia, Western Europe and Canada, where life expectancy is very high, spanning between 82 and 85 years. Instead, in the Center and in South regions of Africa, life expectancy is commonly less than 60 years.
These changes in life expectancy through time and space are rooted in numerous causes of death that with different intensities have affected distinct countries. The goal of our work is to evaluate if and how the causes of death changed during the past decades throughout the world. In particular, the time span that has been taken into account is the range 1990-2019, before the rise of the Covid-19. Our curiosity is to evaluate how are distributed the various diseases in the world.

2 Objectives

The objective of our research is to analyse the changes of the principal causes of death temporally and geographically. This investigation will be developed by addressing further sub-questions.
First of all, as specified in the introduction, our work places itself before the rising of Covid-19. Therefore, we want to have a first look into the geographical distribution of these causes of death in this 30 years span, i.e. 1990-2019.
Then, we will move on analysing, between them, which have been the most deadly through time. It is probable that this distribution is different in every continent, so we will have a deeper look into every territory.
Then, for addressing the temporal part of this research, we will move on to analysing how every cause have changed through the years.
Lastly, we want to have a close look into natural disasters and see the impact of big natural events on the affected countries. We chose this cause of death among the others because it stands out: it is easy to relate them to a specific event and they do not have a similar trend to the others, since they are characterized by peaks.

3 Data Description

The data that we have taken into account are coming from two different sources.

3.1 Causes of Death

The first of these, is taken from the Kaggle platform [1]. The dataset has been cited in a research written by the non-profit organisation "our world in data"[2]. In this research the data were originally taken by the Global Burden of Disease Study carried out by the "Institute for Health Metrics and Evaluation (IHME)". For this study, a worldwide consortium of 500 researchers, coordinated by IHME, measured the impact of more than 290 health conditions and 67 health risk factors worldwide. Part of this research has involved conducting in-person surveys in several countries and gathering health information through website-surveys. The dataset we use, contains the information about the number of deaths (from 1990 to 2019) due to the different diseases in the various countries of the world. In particular the dataset is composed in this manner:

- Country
- Code
- Year
- Meningitis - No. of People died from Meningit
- Alzheimer's Disease and Other Dementias- No. of People died from Alzheimer's Disease
- Parkinson's Disease - No. of People died from Parkinson's Disease
- Nutritional Deficiencies - No. of People died from Nutritional Deficiencies
- Malaria - No. of People died from Malaria
- Drowning - No. of People died from Drowning
- Interpersonal Violence - No. of People died from Interpersonal Violence
- Maternal Disorders - No. of People died from Maternal Disorders
- Drug Use Disorders - No. of People died from Drug Use Disorders
- Tuberculosis - No. of People died from Tuberculosis
- Cardiovascular Diseases - No. of People died from Cardiovascular Diseases
- Lower Respiratory Infections - No. of People died from Lower Respiratory Infections
- Neonatal Disorders - No. of People died from Neonatal Disorders
- Alcohol Use Disorders - No. of People died from Alcohol Use Disorders
- Self-harm - No. of People died from Self-harm

[1] https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world

- Exposure to Forces of Nature - No. of People died from Exposure to Forces of Nature
- Diarrheal Diseases - No. of People died from Diarrheal Diseases
- Environmental Heat and Cold Exposure - No. of People died from Environmental Heat and Cold Exposure
- Neoplasms - No. of People died from Neoplasms
- Conflict and Terrorism - No. of People died from Conflict and Terrorism
- Diabetes Mellitus - No. of People died from Diabetes Mellitus
- Chronic Kidney Disease - No. of People died from Chronic Kidney Disease
- Poisonings - No. of People died from Poisoning
- Protein-Energy Malnutrition - No. of People died from Protein-Energy Malnutrition
- Chronic Respiratory Diseases - No. of People died from Chronic Respiratory Diseases
- Cirrhosis and Other Chronic Liver Diseases - No. of People died from Cirrhosis and Other Chronic Liver Diseases
- Digestive Diseases - No. of People died from Digestive Diseases
- Fire, Heat, and Hot Substances - No. of People died from Fire or Heat or any Hot Substances
- Acute Hepatitis- No. of People died from Fire or Heat

Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Protein-Energy Malnutrition	Road Injuries	Chronic Respiratory Diseases	Cirrhosis and Other Chronic Liver Diseases	Digestive Diseases	Fire, Heat, and Hot Substances	Acute Hepatitis
Somalia	SOM	2008	4154	296	89	8756	1634	756	1817	...	1863	1346	494	8598	2729	2973	2851	4535	507	1110
Cote d'Ivoire	CIV	2018	2367	1137	371	809	27758	609	2323	...	3450	3398	377	758	3404	3638	3958	6661	520	234
Turkmenistan	TKM	1999	146	295	54	23	1	481	311	...	372	477	58	11	446	613	1242	1507	352	100
Solomon Islands	SLB	2005	24	14	14	33	220	101	44	...	277	96	14	33	127	282	101	152	17	13
Lebanon	LBN	1994	50	418	90	13	0	73	279	...	462	650	33	12	370	737	377	586	89	163
Papua New Guinea	PNG	2018	609	387	191	206	945	378	972	...	4505	716	150	202	2185	6954	643	1336	708	91
Kuwait	KWT	2019	12	425	58	1	0	23	42	...	365	322	19	0	583	230	204	338	34	1
Saint Kitts and Nevis	KNA	2015	1	8	3	4	0	3	14	...	26	26	0	3	7	9	10	19	2	0
Italy	ITA	2006	216	34184	5876	469	0	419	666	...	18535	11204	100	325	7154	26286	13429	26019	436	97
Uganda	UGA	2004	6737	1048	261	6462	57092	793	1634	...	3518	2372	548	6326	4804	4639	4181	7092	595	392

3.2 Continent Dataset

The second source of data is composed by the information about the continent of the different countries. This dataset, downloaded from Kaggle[2], has been taken into account because one of our goals is to verify how the phenomenon is distributed among the continents. In particular the dataset is composed as follows:

- Code
- continent
- sub_region(like Southern Asia, Northern Europe)

This data is going to be used to enrich the first dataset with the information of the continent of each country.

3.3 Data Preprocessing

After the merging of the two datasets, done using the countries' codes, we have to make some changes at the data in order to get the most out of them once in Tableau.
Firstly, we change the disposition of the data. The dataset, in fact, presented a row for every combination {country, year} and arranged all the causes of deaths in columns. Since we want to make a visualization that gave to the users the power to change between illnesses, we have to use Tableau's filters, and these are applicable only on rows. To resolve this problem, we change the structure of the dataset in order to have a new row for every {country, year, cause of death} combination, passing from 6 thousands records to 186 thousands. Doing so, we obviously increase repetition at the disadvantage of storage memory, but we make the dataset more easy to work with. The number of columns, instead, lowered to seven: country, code, continent, sub-region, year, cause of death, number of deaths.
Lastly, we add a column called "Percentage of Deaths" in which we store the number, in percentage, of the ratio between the number of deaths for the row's cause and the total number of deaths occurred that year in that country. This is done because in the visualizations we want to focus on the impact of a cause on the total deaths, and not use absolute numbers that would be impacted by the population's size. This measure is also called the "Burden of disease": the weight that a cause of death has on the mortality of a population.

[2]https://www.kaggle.com/datasets/statchaitya/country-to-continent

4 Infographic

For this visualization we want to give to the users the maximum level of flexibility. Therefore, all the infographics are thought in a manner that facilitates the users' interactions with the visualization. This is also reflected in the decision of employing a dashboard, since it is the best way to autonomously explore the data.
The dashboard is built taking into account the seriousness of the topic. For this reason, we use a simple snow white color for the background and we change the font to "Times New Roman". We also use some separators to divide the dashboard into four different subsection, each of which will be used to address a different subquestion.

4.1 Map Chart

For what regards the first subsection, we want the user to have a first look by approaching the problem geographically. To do this, we use plot on a map the average percentage of deaths caused by the different illnesses. In Figure 2, is reported an example using one of the principal causes of death: the distribution of "Cardiovascular" deaths across the world.

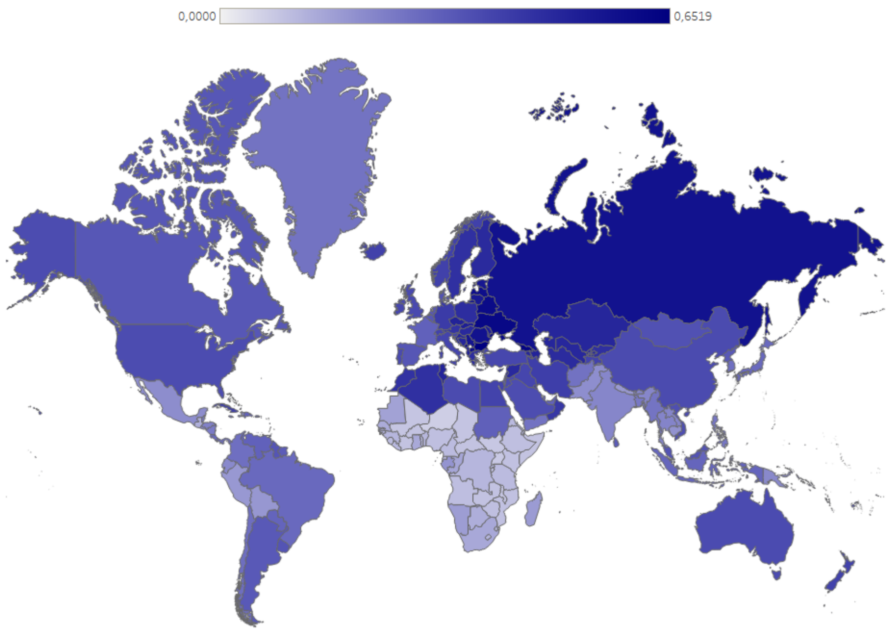


Figure 2: Average Cardiovascular Deaths Percentage

In this subsection we decide to focus on the geographical side of the problem. To do this we summarize the temporal information by computing an average from all the years.
This first look at the data points out a deep inequality in the way in which causes of deaths are distributed through the world. While industrialized countries fight

with senile diseases like Cardiovascular diseases, Neoplasms and Alzheimer, the majority of developing countries still struggles with many curable illnesses like Diarrheal and Neonatal diseases, Maternal disorders and Tubercolosis. Non industrialized countries also have to face phenomenon like terrorism, interpersonal violence and malnutrition. Indeed, drug and alcohol disorders are mainly related to wealthier countries.
This brief description is added to the dashboard.

For what regards the legend, we utilized a sequential fade from white to Navy Blue. We use this color in order to grant the same view also for people affected by Protanopia and Deuteranopia, which are the most common color vision deficiencies among colorblind people. In the next three figures, we show how people affected by these problems may see our color palette:

Figure 3: Final legend color palette

Figure 4: Deuteranopia legend palette

Figure 5: Protanopia legend palette

As we can see the palette does not have important changes between the three different situations.

4.2 Bubble & Donut chart

As previously exposed, one of our goals is to investigate on the principal causes of deaths in time, and how they are distributed in the five continents. In the Figure 6 is represented the situation in 2019, across the world.

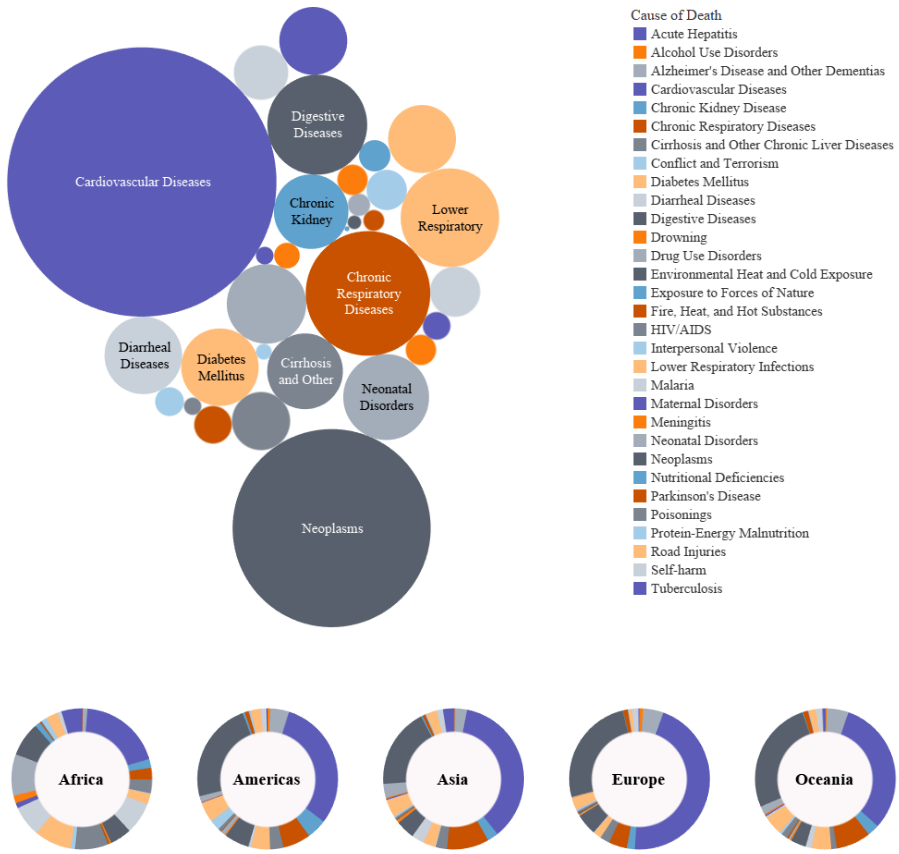


Figure 6: Cause of Deaths in 2019

From this infographic emerges that, in 2019, the most frequent cause of death in the world is related to Cardiovascular diseases. Neoplasms and chronic respiratory diseases have also a great impact in the people mortality.
Using the year's filter at his disposition, the user will be able to see that:
During the years, Cardiovascular Diseases and Neoplasms remain the principal causes of deaths in the world. This phenomenon is very similar between all the continents. Other diseases follow, between different continents, the same pattern discovered in the previous section. Diseases like HIV have risen into the most deadly causes in the mid 90s and then slowly faded away 20 years later thanks to advancements in the medical field. These same advancements that lowered the mortality of curable diseases, come at the price of an higher mortality for senile diseases like Cardiovascular diseases and Neoplasms. These two in 2019 accounted for respectively four times and double the number of deaths of the third cause of death: "Chronic Respiratory diseases."
This type of description, written by us, is added to the dashboard in order to

help the user to understand what kinds of changes to look for.

A big problem that we have to address with this type of infographic is the color palette. Since we have 31 causes of death, it is not possible to build a color palette that assigns for every illness a color that is sufficiently distinguishable from the others.
We decide to use Tableau's color blindness palette. This palette is composed of 10 colors and ensures readability by colorblind people. To cover all the other 21 causes of death we simply use repetition. This is done having in mind that users' will be able to interact with the dashboard and see through labels the cause of death referring to each bubble. In our opinion, a 31 color palette would not ensure distinction between the colors, and therefore understanding the illness only searching the matching color in the legend would be very hard anyway, especially for colorblind people.

Figure 7: Final color palette

Figure 8: Deuteranopia palette

Figure 9: Protanopia palette

People affected by Deuteranopia and Protanopia will see differently the red shades, but the differentiation from the other colors will be ensured anyway.

4.3 Stacked Area Chart

Utilizing five of the ten colors of the palette created, we build a stacked area chart to have a temporal perspective about the developments in the impact of the causes of death.

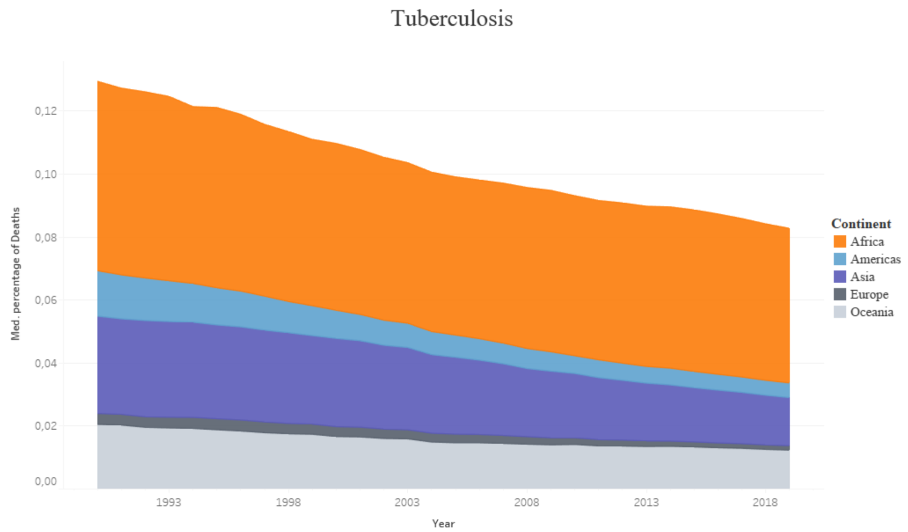


Figure 10: Tubercolosis Stacked Area Chart

In the chart in Figure 10 we can see the downward development in Tubercolosis' impact in the five continents. Africa is the most afflicted continent, and the only one that does see nearly no improvement in terms of percentage of deaths.
In the dashboard, the user is able to change the disease to which the area chart correspond. Moving through the different causes of deaths, the user can analyse how every disease developed through time in each continent. This provides him with an overview on each continent's advancements and struggles with every cause of death. In general, the user will be able to see four kinds of developments:

1. Diseases like Chronic Respiratory disease that remain constant through time.

2. Disorders like those related to Drug Use have constantly grown during these 30 years.

3. Other causes like Neonatal Disorders that have decreased thanks to advancements in the quality of life.

4. Illnesses like HIV that have shown a peak of cases and then got under control.

We add this description to the dashboard, in order to further guide the user in the visualization.

4.4 Forces of Nature

Lastly, we get a closer look into causes of death provoked by forces of nature. The resulting map available in Figure 11 shows the most hit countries in the 30 years span by computing the average of the percentage of deaths in each country.

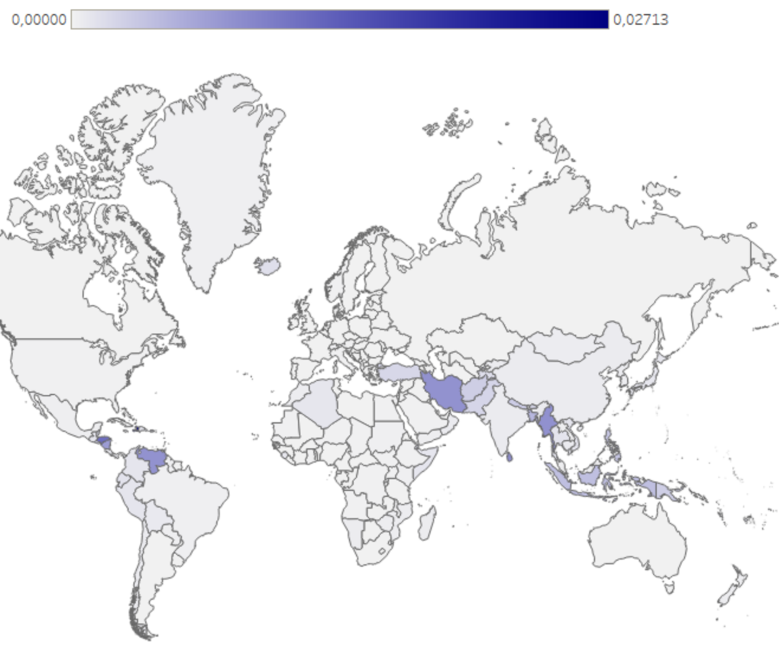


Figure 11: Average Percentage of deaths by Forces of Nature

In particular, as we can see, Iran, Venezuela and Myanmar are those that present the highest relative frequency of deaths caused by forces of nature.
In the dashboard the user is able to evaluate this phenomenon during the years. As we already said, this phenomenon is interesting because it is easy to relate it to specific events that already had a big impact on the media. For this reason, we enrich the user's experience with suggestions that show, for some of the years in the filter selector, the corresponding natural event. We therefore add the following comments on the dashboard:

- 1990: Iran. Earthquake (magnitude: 7.4).
- 1998: Honduras. Hurricane.
- 1999: Venezuela. Floods and landslides.
- 2004: Indonesia and Sri-Lanka. Earthquake and Seaquake (magnitude: 9.3).
- 2008: Myanmar. Cyclone.
- 2015: Nepal. Earthquake (magnitude: 7.8)

Lastly, we fix the legend for this last section. This is done because, by default, Tableau changes the legend every time the filter changes, and adjust the map's color accordingly. Therefore, by changing the year of the visualization, the legend span changes based on the lower and higher percentage present that year. This makes it so that the colors are not comparable between different years. We fix the lower margin of the legend at zero and the higher margin at 0.32. This second value corresponds to the higher value ever, registered in Honduras in 1998.

5 Quality Check

For the quality check of our data visualization we used the following techniques:

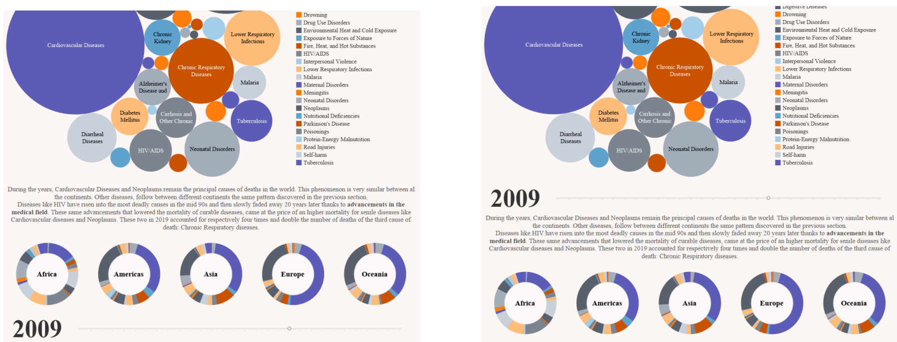
- Heuristic Evaluation
- Psychometric Questionnaire
- User Test

5.1 Heuristic Evaluation

In a Heuristic Evaluation, we observe 3 experts interacting with our dashboard. We ask them to think “aloud” and detect usability problems (i.e. opportunities for improvement).
From users’ behaviours and complaints we identify four possible improvements to our visualization:

1. Following some difficulties experienced while reading the plots’ description, we decide to increment the font’s size.

2. In the second subsection, the position of the filter at the bottom of the page caused a worse visualization. The users in fact were obliged to scroll up and down in order to reach the filter and then visualizing the changes in the bubble chart. To address this problem, we change the location of the year filter, positioning it between the bubble chart and the description. In this way, it is possible to have simultaneously on screen both the filter and one of the charts between the bubble chart and the donut chart.



(a) Before (b) After

Figure 12: Before and After Heuristic Evaluation

3. Also in the second subsection, many users complained that it was not possible to see the percentage of number of deaths that a disease had over the total, both in the bubble chart and in the donut chart. For this reason, we add a calculated field in the details called “percentage over the total”.

4. Lastly, the users found some problem with the comments in the last subsection. It was in fact difficult to locate precisely the events that we described on the year filter. To resolve this inconvenience, we use a different shape to enclose the comments. We decide to use a dialog shape that has a slim tip, in this way every comment is precisely positioned on the time line.

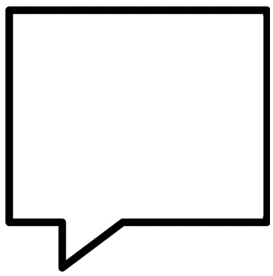


Figure 13: Comment Shape utilized in the last section

5.2 User Test

In a user test, designers observe the users perform a task requiring to interact with their data viz. We plan the task in advance on the basis of our objectives, for which we have designed our data viz. We create seven task to submit to seven users. The test involves the use of quantitative measures, like error rate and execution time. These are used in order to evaluate the efficiency of our infographics.

5.2.1 Map Chart

In the first task, we ask users to interact with the map chart that represent the geographical distribution of different causes of death in the world. The goal is to evaluate if they are able to use the filter related to the choice of the disease they would like to explore. In particular the task is associated with the following questions: "How are Neoplasms spread over the world?" (task 1) and "What is their average percentage of deaths through the 30 year span in Italy?" (task 2). We use Italy to avoid possible difficulties related to finding the country on the map: since all the users are Italians this difficulty will be avoided.

5.2.2 Bubble & Donut Chart

For what regards the second section of our dashboard, we want to evaluate the users' capabilities of navigating through time and using the legend for selecting diseases for which the text does not appear in the bubble. To evaluate their abilities, we pose them the following objectives: "tell us the number of deaths for meningitis in 1995"(task 3) and "tell us the percentage of deaths caused by meningitis in Africa in 1995"(task 4).

5.2.3 Stacked Area Chart

In the stacked area chart we want to focus on the interpretability of the graph, and on the user ability to understand the possible developments of illnesses. To do this, we ask to the user "when was the peak of HIV in Africa?" (task 5) and "What was the maximum percentage of deaths?"(task 6).

5.2.4 Forces of Nature Map Chart

For the last section of our dashboard we want to evaluate the ability of the user to understand the year to which a comment refers to and to individuate the corresponding country on the map. Hence, we don't want to say to the user a specific year, but we require him to move to the exact time based on the corresponding comment. Furthermore, we do not use anymore an easily identifiable country like Italy like in the first task, but the user have to find the country based on the intensity of its color. Therefore, we ask him to "tell us the percentage of deaths caused by the earthquake and seaquake in Indonesia"(task 7).

5.2.5 Results

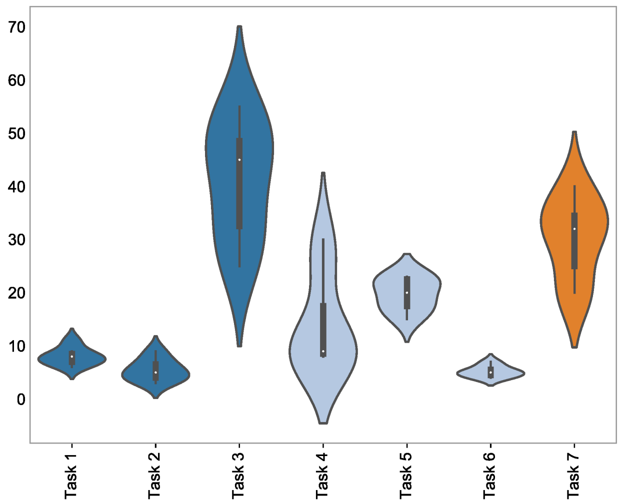


Figure 14: Time needed for every task

Task	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Mean	8"	5.43"	41"	14.14"	19.71"	5.14"	30.14"
Median	8"	5"	45"	9"	20"	5"	32"

Table 1: Mean and Median of task results

From the results we notice that the Task 3 is the one that presents the highest values in terms of time spent (Table 1). Furthermore, four users need our help in the execution of their task (figure 15). The major problems involve the legend: many users don't understand how to use it. Initially, many of them only use the bubblechart and try to search for the disease inside the graph. Consequently, they are not able to detect the disease we ask them.
In order to avoid these problems, we add a suggestion on the dashboard near the legend, encouraging the use of the legend as a tool to select whichever cause of death. In this way, we help users in their interactive experience.

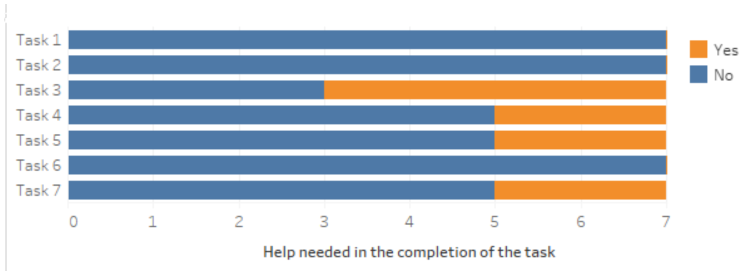


Figure 15: Help needed for every task

Another problem comes out in task 4, where 2 users needed our help to perform a seemingly simple operation. This problem is related to Tableau's behaviour: when selecting a specific disease from the legend, all the donut chart transform back into pie chart and hide the information regarding the continent. This makes it hard, once a cause of death is highlighted from the legend, to understand to which continent every chart corresponds. Given that previously, in order to avoid information repetition, we removed the continent information from the chart, we now have to reinsert the information. In this way, even though Tableau hides the text inside the chart when a disease is selected, the users are still able to understand to which continent every chart corresponds just by hovering on it.

5.3 Psychometric Questionnaire

In this method, we administer a validated questionnaire to a sample of users to evaluate some quality dimensions of the interaction of the users with the data. More precisely, we use the " Cabitza - Locoro " scale, that measures ow much every infographic is:

1. Useful
2. Clear
3. Informative
4. Good Looking
5. Intuitive
6. An overall judgement

Every quality is evaluated on a scale from 1 (very bad) to 6 (very good). The questionaire is submitted to 23 users.

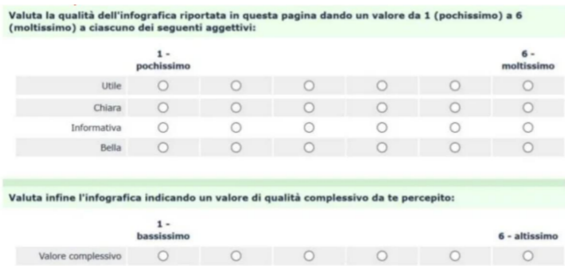


Figure 16: Cabitza- Locoro scale

We ask the users to evaluate three plots, one for each of the first three sections of our dashboard. A big problem in building the questionnaire comes from the fact that we have to show a static image of a graph that has been designed having in mind interactivity. We decide to ask the users an opinion on this three plots:

1. The map chart of section one, related to "Alcohol use disorders". We believe that this choice of the cause of death can be the most informative and interesting.

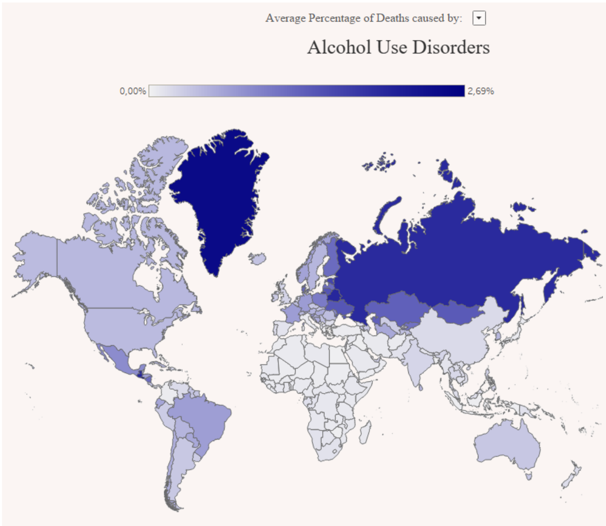


Figure 17: Map Chart utilized in the questionnaire

2. The bubble chart of section two, filtered on the last year available (2019). We choose the last year in order to give the users a more useful visualization.

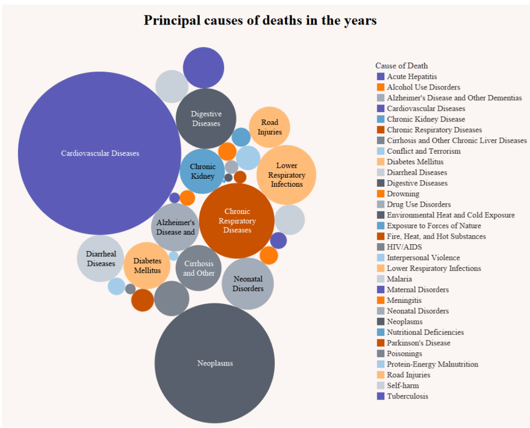


Figure 18: Bubble Chart utilized in the questionnaire

3. The stacked area chart of section three, related to the disease: “HIV/AIDS”. We choose this cause of death among the other to demonstrate the plot’s usefulness in showing developments like peaks.

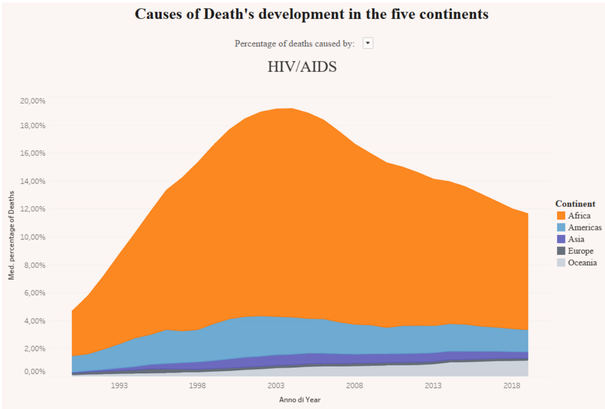


Figure 19: Stacked Area Chart utilized in the questionnaire

5.3.1 Results

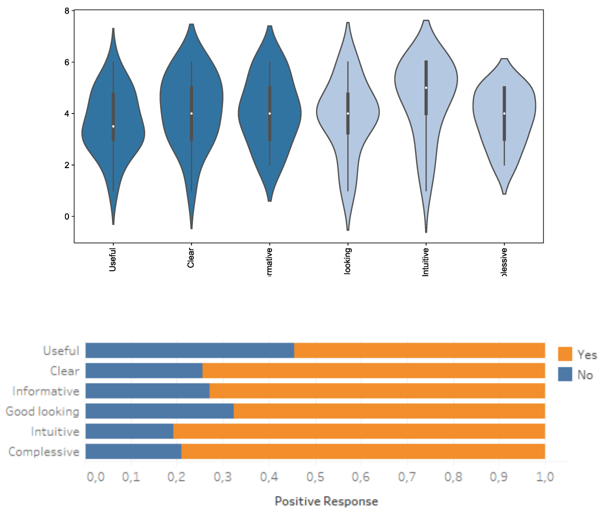


Figure 20: Map Chart results

The map chart achieves good results in the majority of measures. However, users find the graph not very useful. It is difficult to understand why, maybe this is due to the fact that, they can’t investigate the year by year evolution of the disease. In our opinion, though, this type of temporal analysis can be done more effectively on other types of graphs, like the ones below.

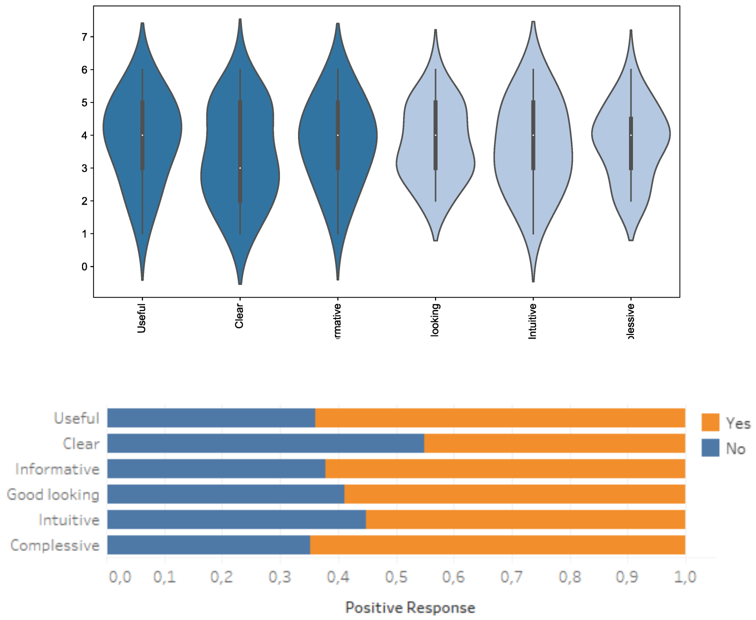


Figure 21: Bubble Chart results

For what regards the bubble chart, it seems that users find it to be not very clear. This is probably caused by the fact that, when the graph is static, it is difficult to understand to which disease every bubble refers to, especially for those that are too small for the text to appear. This is due to the fact that the legend, given the high number of causes of death, has to use repeat the colors. However, we believe this is not as much of an issue when the infographic is experienced in the way it is meant to be: interactively. In this environment, indeed, the user is able to get that information by hovering on a specific bubble, or selecting a specific illness by the legend.

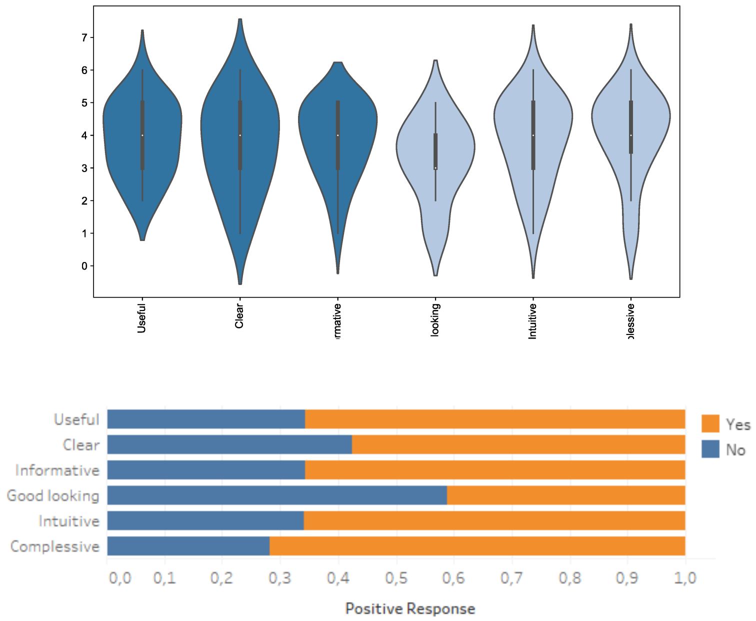


Figure 22: Stacked Area Chart results

Lastly, the area chart shows good results across all the measures, except for it not being regarded as "good looking". The cause of this judgement might be founded on the fact that, in order to demonstrate a type of phenomenon that showed a particular development over the years, we utilized HIV as the relative cause of death. Doing this, the African continent takes up most of the space, and it might be that users do not appreciate this. However, this is a problem that is particular to this disease.

6 Conclusion and Future Developments

We started by analysing the different diseases geographically. This first look at the data pointed out a deep problem of inequality in the way in which causes of deaths are distributed through the world. While industrialized countries fight with senile diseases like Cardiovascular diseases, Neoplasms and Alzheimer, the majority of developing countries still struggles with many curable illnesses like Diarrheal and Neonatal diseases, Maternal disorders and Tubercolosis. Non industrialized countries also have to face phenomenon like terrorism, interpersonal violence and malnutrition. Indeed, drug and alcohol disorders are mainly related to wealthier countries.
We moved on to analyse the principal causes of death in the world through time. We saw how, during the years, Cardiovascular Diseases and Neoplasms remained the principal causes of deaths in the world. This phenomenon is very similar between all the continents. Other diseases follow, between different continents, the same pattern discovered in the previous section. This infographic has been very useful to evaluate the various advancements in the medical field. Diseases like HIV, for example, have risen into the most deadly causes in the mid 90s and then slowly faded away 20 years later thanks to them. These same advancements that lowered the mortality of curable diseases, came at the price of an higher mortality for senile diseases like Cardiovascular diseases and Neoplasms. These two in 2019 accounted for respectively four times and double the number of deaths of the third cause of death: Chronic Respiratory diseases.
Lastly, thanks to the stacked area chart we analysed how every disease developed through time in each continent. This provided us with an overview on each continent's advancements and struggles with every cause of death. In general, in all the five continents, we have been able to identify four different types of development. Diseases like Chronic Respiratory disease that remain constant through time. Disorders like those related to Drug Use have constantly grown during these 30 years. Other causes like Neonatal Disorders have decreased thanks to advancements in the quality of life. Lastly, illnesses like HIV have shown a peak of cases and then got under control.

For what regards future developments, we think that it would be very interesting to include inside our sample of causes of deaths also Covid-19 and, thanks to updated data, investigate how it affected the world in this last years.

In conclusion, this research made us understand that death can be frightening, and regarded as a depressing subject to analyse, but studies around the matter can create a better world. Monitoring its development, its causes and the territory it affects can help us to act in the right time, in the right manner and at the right place. Analysing the subject in a temporal and geographical way is the first step for bringing a better future for everybody.
We think that this is the main lesson that this type of analysis can give.

References

[1] Max Roser, Esteban Ortiz-Ospina, and Hannah Ritchie. "Life expectancy". In: *Our world in data* (2013).

[2] Max Roser, Hannah Ritchie, and Fiona Spooner. "Burden of disease". In: *Our world in data* (2021).