Title Page: TBD

Team Name: CL08-T06

Team Member:
- BoRui Deng 104337579
- Oliver Bray 103070831

Tutorial day and time: Thursday Class 1-8 10:30 - 12:30

Year and Semester: The first semester in 2024ww

Link to Mercury-hosted website:

Word Count:

Table of Contents

1 Introduction

1.1 Background and Motivation

With the ever-growing importance of public health management and policy-making, it's imperative that we have access to powerful data visualisation. These visualisations not only help us to make informed decisions but also provide us with the ability to communicate complex information clearly and concisely. They empower public health officials, policymakers, medical researchers, epidemiologists, news media, data journalists and individuals with increased health awareness to analyse and compare data on preventable mortality across different regions and periods, assess the effectiveness of current health policies, and update strategies in a timely manner. These visualisations can assist in identifying trends, comparing different groups, and evaluating the impact of intervention measures.

1.2 Visualisation Purpose

The visualisation developed in this study aims to allow users to explore preventable mortality rates and answer multiple vital questions deeply. For example, users can analyse the trends in preventable mortality rates across various countries over the past years and investigate the potential reasons behind these trends. Additionally, through these visualisations, users can compare the differences in mortality rates between countries and explore potential social, economic, and policy factors influencing these disparities. More importantly, policymakers and public health experts can assess the effectiveness of specific health interventions by observing data changes before and after implementing these policies or measures. In addition, the visualisation will provide data support for public health policies. This will enhance the adaptability and effectiveness of these policies, enabling policymakers to respond more quickly to health crises and optimise resource allocation. Along with this, the visualisation will also present easy-to-understand health data visualisations, which will improve public awareness and understanding of significant health issues. This will promote changes in health behaviours and increase public health awareness and participation. Additionally, for researchers, detailed data visualisations will support scientific research and evidence-based medical practices, which will advance scientific development in the public health field.

2 Data

2.1 Data Source

Link: https://stats.oecd.org/Index.aspx?ThemeTreeId=9

Dataset 1: Health/Health Status/Avoidable Mortality
Dataset 2: Health/Healthcare Resources/Total health and social employment
Dataset 3: Health/Health Expenditure and Financing/Health Expenditure and Financing

All of our datasets are of the table type, meaning all of their data can be displayed in the form of a table with the attributes across the top, and data underneath each one. The attributes in these datasets consist of:

- Variable: The category defining what the value means for each record/row. The variable attribute includes categorical and nominal data, as the values in the column can't be ordered and can be grouped into categories instead of being measured numerically.
- Measure: Gives meaning to the values in the "Value" attribute. This data is also categorical and nominal, similar to the Variable attribute.
- Country: Defines which country the value belongs to. This data once again is categorical and nominal, as countries are just that, and can't be meaningfully ranked or ordered just by themselves.
- Value: Represents the exact value of the variable with respect to the Country. These values can be considered numerical and discrete.
- Year: The year in which the value was recorded. While year is represented by a number, in this case it is considered categorical as we are not doing any sort of calculation or transformation using this attribute, we are just displaying it as is.

There were a number of attributes and records that will not be used in the final dataset, as it isn't important to the questions we were going to ask. Most of the data not included was duplicate attributes that just used a different name, such as COU for Country, and YEA for Year.

2.2 Data Processing

Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented? Will you be deriving any variables?

Describe the clean up process that was implemented. Explanation and calculation of derived variables (if used).

We expect to perform quite substantial amounts of data cleanup, as these datasets come with a lot of data in the on table, too much to fit on one visualisation meaningfully. Most of the cleaning will involve removal of the duplicate attributes as mentioned earlier, but will likely also include simplification of the variables and measures. Since each database is so large, there are multiple variables and measures that provide way too much information for one chart, so cutting these down to only a few variables or a few measures makes the data a lot more clean looking when visualised.

We have processed and cleaned the data using a python script, as doing so manually would prove to be very laborious and time-consuming. We have made use of the pandas library for Python, which makes data cleaning very simple. For the most part, we have been filtering by the exact variable or measure we want to keep, and dropping the unwanted columns/attributes. With dataset 2 not all countries had the same amount of records, so to keep it consistent and to avoid problems with the year slider, we had to drop any records from countries that had data past 2020.

3 Visualisation Design

How will you display your data? Provide some general ideas that you have for the visualisation design. Include sketches of your design. Include at least 2-3 alternative ideas for your visualisation. Describe and justify your choice of visual encoding and chart types. Show the evolution of your design. How has it progressed? Justify the way you have chosen to represent your data.

Description (including screenshots) and explanation of final design.

[NOTE 1: You are encouraged to provide your own structure to this section (i.e., section headings etc).

NOTE 2: You MUST show evidence of iterative design (i.e., sketches of alternative and preliminary designs).]

Visualization Design Development and Iteration

Initial Design Concept

In public health data visualisation, presenting information is crucial for decision-makers, health officials, and the general public. The initial design of this study employed a line chart to display the number of deaths across different countries over various periods. In this chart, each country's data is represented as a separate line, clearly depicting each country's trends and facilitating a direct comparison between different countries. D3.js's ordinal colour scale was used to assign distinct colours to each line, making it easy for users to distinguish between the data series. This colour coding is an effective way to differentiate complex information. Additionally, the x-axis represents time, while the y-axis shows the number of death data. The design ensures that the time axis is clear and the data axis accurately represents the information. Furthermore, labels at the end of each line enhance the chart's informational value by marking the names of the countries. This approach was primarily chosen because it allows for the simultaneous display of multiple datasets and effectively tracks and compares changes in mortality rates across different countries over specified

period.

brrayzy 02/05/2024 15:27

Ok I have 3 datasets. The first one is preventable mortality which we talked about in class. It will be a line graph with 3 different lines representing 3 different countries, and will have the years on the x axis, and value on the y axis. The second one is health and social worker employment, which will be a classic bar chart. I've cleaned this one so that each country only has 1 value which is a 10 year average of how many people are employed as health or social workers. The final dataset is Perceived health status. This dataset has 3 variables, bad health, fair health, and good health. I was thinking, depending on how complex it will be to implement, the way we could display this data is using a world map. The user would have 3 options to filter the map by, which are the 3 variables. The map then displays a lighter or darker shade of a colour depending on the value of a certain country. I was thinking red for bad health, yellow for fair health, and green for good health. If you think this map will be too hard to implement I can easily find a simpler dataset.

```
Variable,Measure,Country,Year,Value
Preventable mortality,Number of total deaths,Australia,2010,22339
Preventable mortality,Number of total deaths,Australia,2011,22487
Preventable mortality,Number of total deaths,Australia,2012,22265
Preventable mortality,Number of total deaths,Australia,2013,22914
Preventable mortality,Number of total deaths,Australia,2014,23798
```

Expand preventable_mortality_cleaned.csv 3 KB

```
Variable,Measure,Country,Value
Total health and social employment,% of total civilian employment,Australia,12.4
Total health and social employment,% of total civilian employment,Austria,10.2
Total health and social employment,% of total civilian employment,Belgium,12.8
Total health and social employment,% of total civilian employment,Canada,11.3
Total health and social employment,% of total civilian employment,Chile,5.0
```

Expand health_social_employment_cleaned.csv 4 KB

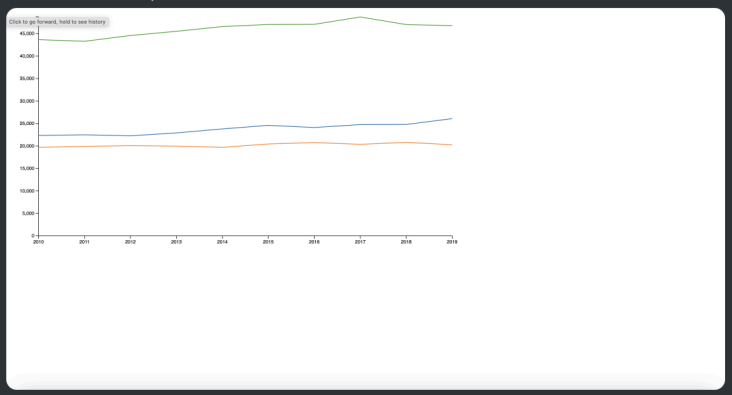
```
Variable,Measure,Country,Year,Value
"Bad/very bad health, total aged 15+",% of population (crude rate),Sweden,2010,5.4
"Bad/very bad health, total aged 15+",% of population (crude rate),Sweden,2011,5.4
"Bad/very bad health, total aged 15+",% of population (crude rate),Sweden,2012,4.9
"Bad/very bad health, total aged 15+",% of population (crude rate),Sweden,2013,4.6
"Bad/very bad health, total aged 15+",% of population (crude rate),Sweden,2014,4.6... (72 KB le
```

Expand health_status_cleaned.csv 122 KB

@brrayzy *Click to see attachment*

push Yesterday at 22:58

For first dataset, how does this draft visualisation looks like?



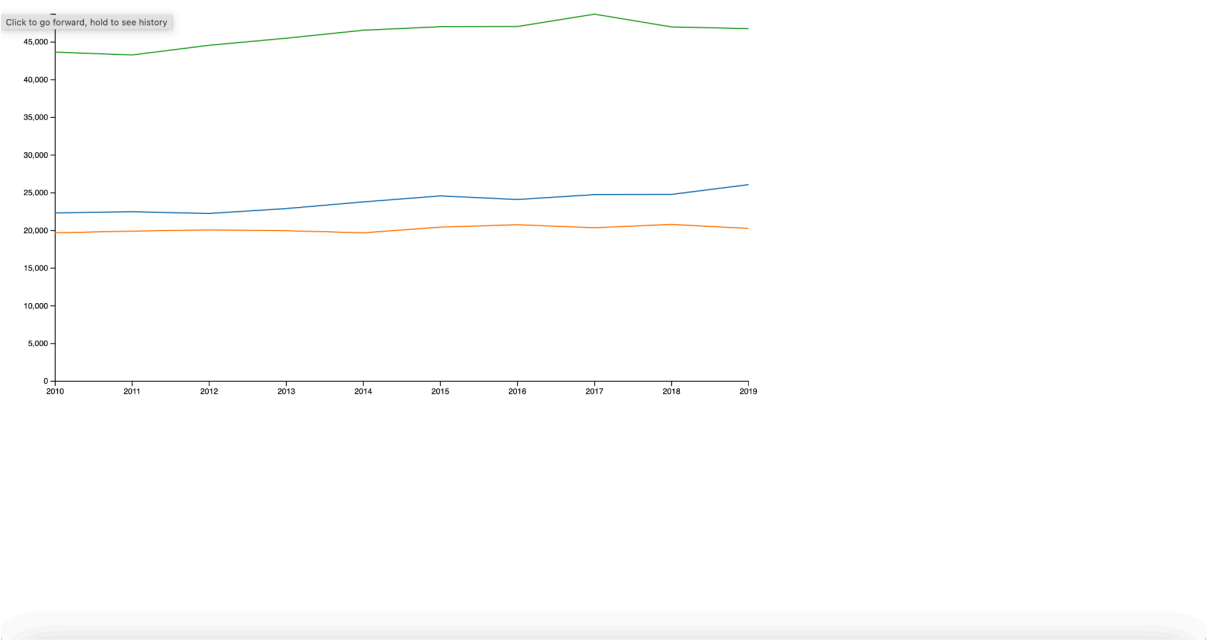
8 May 2024

brrayzy Today at 14:08

Hmmm, the graph looks great, but doesn't look as interesting as i was thinking. But that isn't a problem with the graph it's a problem with the data. There doesn't look to be any meaningful trends or anything. We could maybe try and think of a different visualisation type to use, or we'll have to try different data

push Today at 20:34

I see, let me think about that



In the second draft of

4 Validation

Design and conduct a usability evaluation on your visualisation. You must describe the method and results of your evaluation.

5 Conclusion

Provide a summary of the project and what you learnt from doing it.

References

References consulted (blogs, books, academic papers, discussion/help forums - for both design and programming)

https://d3js.org/d3-scale-chromatic/sequential

Appendices

Usability evaluation test materials (if used)

Notes/data collected in usability evaluation (if used)