Interactive Data Visualization - Project Report

Jimi Hytönen

4. toukokuuta 2020

1   Abstract

This project is part of interactive data visualization course. The purpose of this project was to create an interactive visualization from real-world data. I chose Covid-19 pandemic dataset as my datasource because it is currently a topical event and I immediately had a vision what my visualization could look like. Also, I wanted to.. TBD-solve some problem?

In this report I will cover the used data sources, explain the technical aspects of the visual representation, tell about the design processes and ending with lessons learned. The project can be found in GitHub `https://github.com/Jimmeeee/idv-project`.

2   Task abstraction

Instead of jumping into the visualization tool with the existing dataset and trying stuff out until I get something to work, I broke the project process into smaller tasks. The following list is meant to give an overview of the process I went through while doing the project.

1. Cleaning the data and getting the data into easy-to-use form.

2. Getting familiar with the visualization software and getting to know its limits.

3. Doing the basic visualization and step by step pushing myself to try things differently.

3   Data

I used dataset provided by John Hopkins University from `https://github.com/CSSEGISandData/COVID-19`. The data was collected from multiple different sources such as World Health Organization(WHO) and cleaned into daily reports that were in csv-format. In addition to JHU-Covid19 dataset I used timeline events of coronavirus outbreak from February and March (`https://en.wikipedia.org/wiki/Timeline_of_the_2019%E2%80%9320_coronavirus_pandemic_in_February_2020` and `https://en.wikipedia.org/wiki/Timeline_of_the_2019%E2%80%9320_coronavirus_pandemic_in_March_2020`).

I used Python with pandas for data processing as I needed to get the data into a certain form for further processing and visualization. For JHU dataset I read the daily reports with pandas and made the reports to have uniform column names and country names. Finally, when the daily reports were cleaned I created a single csv file containing the information about confirmed, recovered and death cases in reported countries, which I used in the Tableau public to maek the visualization.

For the other dataset (timeline events from wikipedia), I used web scraping to extract useful information from the site. For web scaraping I used Python's BeautifulSoup library and Pandas. I matched countries with events if the event contained the country. This allowed me to have events that were more or less connected to the country showed when the country was selected for further examination in the visualization.

4   Design rationale and Visual representation

I wanted to make an interactive visualization that would provide fast and effortless information about the Covid-19 pandemic in the world. My visualization contains four elements that are:

1. Summary of the current situation

2. Interactive map of the world

3. Line plot containing Confirmed, Death and Recovered cases

4. Events textbox

The first element shows the total confirmed, death and recovered cases in the world. However, if a country/countries are selected from the map then the summary shows the total cases in the selected countries.

The second element contains three interactive maps on top of each others, one for each of the following cases; confirmed, death and recovered. Each map shows the total amount of corresponding cases in the country with a colored circle. The size of the circle is directly proportional to the sum of the cases in the country. Also, each country has hover-over tooltip which tells the total cases.

The third element consists of three line-plots, one for each of the following cases; confirmed, death and recovered. Each line-plot shows the total number of the corresponding cases as a function of date. One can choose a certain day or month from the plot and it will change the summary to show all of the cases on that day/month. Also, the events textbox would only show the events from that single day/month.

The fourth element shows the events of the selected country/countries. Also, one can limit the events by selecting certain country from the map or dates from the line plot.

Overall, all the elements are stacked on top of each other so that it is intuitive which plot will change and what changes with it.

5 Lessons learned

6 References