Assignment 5 (Mini project) -  
Basic Information

in lecture “Interactive Data Visualization” (SS 2020)

## 

# Topic: COVID-19

### **Group 02**

* Julia Bresien
* Christopher Schrewing
* Soundarya Krishnan
* Tejas Ravindra Dhawale
* Rigers Sulku

## 

### **Data**

The data we visualize in our main view is:

* **Number of currently active cases** **in relation to the total population of a specific country** (#active\_cases = #total\_cases - #recovered\_cases - #death\_cases)

→ All this data can be found in the datasets provided by Johns Hopkins University and Oxford University. Hence, we used and combined these datasets.

*Reasons for selecting*:

* **University of Oxford (UoX)**: Country-specific population according to specific dates needed for our use case. This population data was not provided by Johns Hopkins University. Since the other data (cases, death, …) are not updated daily, and we found gaps in between, we will only take the population data information from this dataset.  
  <https://github.com/owid/covid-19-data/tree/master/public/data>
* **Johns Hopkins University (JHU)**: It is daily updated and always up to the current date. Time-series data of recovered, death, and total cases for specific countries is also available.  
  <https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series>

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset Title** | **Source** | **Columns interesting for our use case** | **Data characteristics** |
| time\_series\_covid19\_confirmed\_global.csv | Novel Coronavirus (COVID-19) cases, provided by Johns Hopkins University CSSE | **Country/Region**, all date columns of form **DD/MM/YY** | The **country** column will have **nominal** data with the names of all the countries. There are different **Date** columns, **one for each day**. They consist of **quantitative** data on the number of **confirmed cases** per country. |
| time\_series\_covid19\_deaths\_global.csv | Novel Coronavirus (COVID-19) deaths, provided by Johns Hopkins University CSSE | **Country/Region**, all date columns of form **DD/MM/YY** | The **country** column will have **nominal** data with the names of all the countries. There are different **Date** columns, **one for each day**. They consist of **quantitative** data on the number of **deaths** per country. |
| time\_series\_covid19\_recovered\_global.csv | Novel Coronavirus (COVID-19) recoveries, provided by Johns Hopkins University CSSE | **Country/Region**, all date columns of form **DD/MM/YY** | The **country** column will have **nominal** data with the names of all the countries. There are different **Date** columns, **one for each day**. They consist of **quantitative** data on the number of **recovered cases** per country. |
| owid-covid-data.csv | Coronavirus Pandemic. Our World in Data, provided by University of Oxford | **location, population,**  **iso\_code** | The **country** column will have **nominal** data with the names of all the countries. For every country, the population will be specified for each specific date (but it is always the same). **Population** data is **quantitative**.  The **iso\_code** is needed for the mapping of data into countries in our visualization. The code is **nominal data**. |

*Additional information about missing data:*

We found out about **different irregularities in the two data sets** we would like to use, which are briefly described here:

* Some countries are only mentioned in JHU and not UoX (or vice versa), e.g. Cabo Verde

→ we will only show data for countries which exist in all datasets, the rest will be ignored (belong to group “no data”)

* There are examples where JHU and UoX do not use the same name for a country, e.g. Czech Republic (UoX) vs. Czechia (JHU)

→ the naming of countries must match, so we manually checked and renamed them

* JHU sometimes splits up countries into different regions/territories, e.g. Canada, China

→ we will merge different territories to one country and sum up the numbers

### **User and Tasks**

The visualization is aimed to be used on the **dashboard of a travel agency’s website**. These days, with summer holidays just in front of us, many people are unsure whether or not to travel and especially where to go.

While planning the next holiday trip, **travelers** would want to know the latest updates and history about the coronavirus situation in different parts of the world.  
Thus, it is important to consider the **currently active cases** in a county. But since a number alone is not meaningful enough (2.000 might be better than 1.500, if the country is a large one with many residents compared to a smaller one), we would like to visualize this number **in relation to the population** of a country.

A dashboard showing the most needed information like the total number of confirmed cases, cured cases, and deaths would help the traveler to understand the severity of the situation in different countries. This can help the traveler to choose the prospective destinations. It would also help the agency to avoid having to deal with a myriad of phone or direct personal inquiries which might become overwhelming for the agency staff to handle.

Some concrete questions the users from our visualization could have, are:

* How is the current situation of active cases in a country considering its total population?
* What was the course of active cases in a country?
* In which country is it most likely to get in contact with someone who currently has the coronavirus?

### **Visualization technique**

Our visualization has a **world map with countries color-coded based on the severity of coronavirus cases**. Each country’s **confirmed cases, currently active cases, the death count, cured cases, and total population number along with its affected population ratio** can be seen in a detailed view. Users will also be able to view these statistics for any **particular date starting from 1/22/20**.

The data available in the considered datasets have Spatio-temporal attributes. Hence, the **Geospatial Choropleth map** is used as the visualization technique. The 2D view of the world map is **preattentive** and helps display the coronavirus related information for each country in the world. A color bar that's based on the **count of currently affected population with regard to total population** helps understand the different levels of severity across the globe. The **countries are color-coded** based on this scale. Additionally, by altering the selected date by using a **time slider**, the view of **coronavirus information** across different countries can be updated **for the particular date selected**. In addition to that, the time can automatically be altered by starting an **animation** which chronologically goes through the days.

We used the **library plotly** for the visualization. Precisely, we used plotly.express, more information can be found here: <https://plotly.com/python-api-reference/generated/plotly.express.choropleth.html>

### **Interactions**

***Interaction Operators:***

1. **Navigation:** 
   1. When a user hovers over a county, the following coronavirus parameters are displayed: the total number of confirmed cases, the total number of recovered cases, the total number of current active cases, the total number of deaths, total population, and finally the percentage of currently active cases.
   2. The user can pan across the map.
   3. For better visual scales, the user can change the level of detail and can zoom in or out of the map.
2. **Selection:** 
   1. The user can isolate a subset of the graph using either box selection or lasso selection.
   2. With the reset button, the user can reset the view to the original view.
3. **Filtering**
   1. The user can use the slider to choose different dates for which the data values are filtered and appropriate coronavirus parameters would be updated.
   2. Also, the user can use the animation play button which will filter the values based on the date and show a simulation of how the pandemic evolved.

***Interaction operands/Interaction Spaces:***

1. **Screen Space:**
   1. The visualization consists of a global map, in which the regions get differentiated based on the county’s territory. When a user navigates through the map or when he zooms a particular portion, it is the screen space that is manipulated and not the data.
2. **Data value Space:**
   1. Based on the date selected, the data values displayed on the screen will change and the corresponding map will be updated accordingly.
   2. Also, based on the selected date, the color bar scale will be dynamically updated depending on the maximum and the minimum value of percentage of active cases with respect to population for the selected date.

***Display***

It will be intended for Tablet/PC/Laptop users. We have not optimized the visualization for mobile devices.

### **Concept**

In comparison to our concept, we made the following changes:

* We needed more information from the datasets than originally assumed. To connect the countries to the right position on the world map (using a choropleth map in plotly) we needed the official ISO code, like DEU for Germany, which was not stored in the Johns Hopkins datasets, but in the Oxford dataset, so we took it directly from there.
* We added a little more information to the detailed view when hovering over a country. Since all the percentage numbers are obviously tiny, it makes sense to explicitly display them when hovering, together with the total population number which is one important input to calculate this number. So, we added these two attributes to the details.
* We originally had planned a Selection operation with the color legend. The idea was to fade out countries whose color does not match to a specific range the user has selected. But since many countries are very close together, this does not really seem to be usable. In addition, we use a continuous color space which nicely shows all the differences, so our idea to emphasize some countries did not really seem to be useful.

### **Division of responsibilities**

We as a group worked strongly together and had many online meetings. Overall, everyone contributed something to every subpart of this project.

We distributed some “main responsibilities” to all different group members, which we will explain in the following table:

|  |  |
| --- | --- |
| **Group member** | **Main responsibilities** |
| Soundarya | Concept |
| Tejas | Research on different visualizations and interactions |
| Christopher | Preprocessing data, merging different datasheets |
| Rigers | Time slider in visualization |
| Julia | Basic information, presentation |