Authors: Eamon Worden

Assignment: Individual Project

Due date: 3/14/2022

Online Sources: See below

Help obtained: Prof Jimenez

I confirm that the above list of sources is complete AND that I have not talked to anyone else about the solution to this problem

Project documentation:

What is its purpose?

The purpose of my project is to provide a visualization about covid, the number of cases and deaths, how different countries were effected differently by and what government measures were used and their effectiveness in each country.

Data description:

I used 2 data sets for my visualizations. First was all the information you would ever want to know about covid. This data, collected by John’s Hopkins, includes information such as covid cases and deaths, vaccination rates, and infinitely more data such as population density, age of a country’s population, diabetes prevalence and more. I primarily focused on cases, deaths and vaccination rates in my project.

The other data set I used was about all the government measures used because of covid. This is information such as border closures, mask mandates, and surveillance. This would include which country used these measures, and when they were phased in and phased out.

How was the data collected?

The covid data comes from WHO, who kept very good track of lots of covid data over the last 2 years. The government measures data was collected by The Oxford Covid-19 Government Response Tracker who got information from any source available (laws, news and even Facebook).

Who are the users that this visualization was made for?

The visualizations are made for decision makers who want to know how effective different measures were, people who want to see how covid worked over time, and anyone who was generally curious about how different countries handled covid during different intensities of covid.

What questions are you trying to answer? What works?

I am trying to answer 3 questions: How did covid trend in different countries over time? When were different government measures put into place? How effective were different government measures?

I also feel my data is good for answering other questions such as: How did different events effect covid? Did noncompliance effect government decisions? This is not encoded in my visualizations, however, being able to look up certain events, such as a super spreader concert on August 14, 2021, which has visual repercussions in my visualizations when you select dates following August 14, 2021.

What insights did you get from your data?

What I learned is that it is difficult to determine the effectiveness of government measures. Many measures were put in in response to a big uptick in cases, and sometimes covid leveled off after those measures and sometimes it kept rising. Most measures had varied effects as well, even within countries. For example, closing borders in Italy had no effect the first time it was done but did seem to lower the number of new cases the second time it was done about a year later.

One thing I did see though is that many government measures tended to be implemented after the first big covid spike but were rarely implemented after a second or third spike even if those spikes had more cases and more deaths. Whether this is due to expected noncompliance, political issues, people being tired of government measures or other reasons there are noticeable trends about when governments implement different measures.

What needs improvement?

One problem with the project is none of the data is perfect. This was most apparent in certain countries underreporting deaths and cases, which I think was especially prevalent in Africa, and there were occasional apparent issues in government measures as many measures were never recorded as phasing out and others were phased out before any were phased in.

Besides the data, the visualizations are not quite perfect and if I had more time, I would have improved a few things. First, the measures being phased in and out do not tell the whole picture. I would want to be able to click on and view more valuable information such as whether measures were global or local, penalty for noncompliance (i.e. are you fined or jailed if you don’t follow a certain measure?) and comments about certain events (certain awareness campaigns had comments such as “advertising health services” and “avoid kissing, weddings and funerals”).

How to reproduce my work:

First you want to collect covid data from WHO. After that I collected data regarding government measures. Then I spent a lot of time choosing what data was the most important, i.e. what covid data did I consider the most important and how specific did I want to make government measures information.

Following that I had to format my data. This included renaming columns to be consistent between my data sets, merging data sets (specifically covid data and the latitude and longitude of each country), and selecting the specific columns I wanted.

After that I wanted a proof of concept for being able to select the date and display it in a useful visual manner. That started off as visualizing the number of covid cases per million in each continent and overlaying it on a world map. Then, I visualized the number of covid cases per million in each continent and added a slider so that users could select a certain date to visualize covid cases per million. Following that I used the covid data merged with the latitude and longitude of each country to visualize covid cases per million in each country of the world.

This left me with a graph with too much information and a histogram with very little information. Having every country on the map was a bit too much information, so I added the ability to zoom in on that map and visualize just a selected portion side by side with the world map.

Next, I wanted to be able to click on that zoomed in map so you could select a country and view more information about it. First, I added just a text display of covid deaths, cases, new deaths, new cases, and vaccine information about a given day for the selected country. Then, I added a graph of total cases and total deaths over time for the entire timeframe of data being recorded. Finally, I added vertical lines with government measures (each with their own color), when they were phased in and when they were phased out. This, however, was messy so I instead added a box in the side bar, below the date slider, where users could select the specific government measure they wanted overlayed with covid cases and deaths.

What-why-how:

World map:

What: A map of the world with information for the number of new cases per million in each country in the world on a given day.

Why: To visualize how covid is doing on a specific day across the world. It makes it easy to view how covid is affecting different countries differently, visualize how regions are doing as a whole, and compare trends over time.

How: This is visualized as a map of the world with a dot at the center of each country with a different color for convenience with the dot being scaled to the number of new cases per million on a given day in that country. There is a slider which lets you change the date that you want the information visualized for.

Reactive world map:

What: A map of the world with information for the number of new cases per million in each country in the world on a given day, except that it is zoomed in according to a brush on the world map.

Why: To visualize how covid is doing on a specific day across the world. It makes it easy to view how covid is affecting different countries differently, visualize how regions are doing as a whole, and compare trends over time. This map specifically lets you zoom in on a region to compare how certain regions are doing (i.e. you can zoom in on Europe, North Africa or the middle east easily).

How: This is visualized as a map of the world with a dot at the center of each country with a different color for convenience with the dot being scaled to the number of new cases per million on a given day in that country. There is a slider which lets you change the date that you want the information visualized for. This also reacts to a brush on the world map which zooms in.

Clicked country information:

What: Text output of the number of new cases, new deaths, total deaths, total cases, and total vaccinations of a given country on a certain day.

Why: To easily read information about how a country on a given day.

How: Once you click on the reactive world map it reads the nearest country and then displays the total number of cases and deaths up to that day, the new cases and deaths per million on the selected day, and the total number of vaccinations in that country up until that day.

Graph of total cases and deaths over time with government measures:

What: A line graph showing the total number of cases and deaths of the selected country over time, with government measures overlayed.

Why: To view how covid trended over time, and how this impacted decision making about what measurements governments wanted to implement, and how those measurements affected covid cases and deaths.

How: A country is read off the reactive plot when the user clicks, then it graphs total cases and total deaths over time. It also reads a selected measure from a textbox and adds a black vertical line to the graphs when a measure of the selected type was phased in, and a blue vertical line when a measure of the selected type was phased out.

Histogram of new cases per million in each continent:

What: A histogram showing the new cases per million in each continent.

Why: Covid affected different continents different at different times, and this is meant to help understand how to spread from continent to continent over time.

How: It displays a histogram showing the number of new cases per million across the continent on the selected day and makes it easy to compare continent by continent on different days.

Sources and References (with appropriate credits)

R4DS (general R things)

<https://www.r-graph-gallery.com/279-plotting-time-series-with-ggplot2.html> (ggplot time graph)

<https://campus.datacamp.com/courses/free-introduction-to-r/chapter-6-lists?ex=6> (working with lists)

https://mastering-shiny.org/ (lots of help for all things shiny)

<https://stackoverflow.com/questions/40908808/how-to-sliderinput-for-dates> (slider for dates in Shiny)

<https://www.datanovia.com/en/blog/ggplot-legend-title-position-and-labels/#change-legend-title> (formatting graph)

Most of my Covid data from: https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv

Government measures data from <https://github.com/OxCGRT/covid-policy-tracker>

<https://statisticsglobe.com/change-formatting-of-numbers-of-ggplot2-plot-axis-in-r> (changing axis of plots)

<https://www.datanovia.com/en/lessons/rename-data-frame-columns-in-r/> (renaming columns)

<https://bookdown.org/dli/rguide/scatterplots-and-best-fit-lines-two-sets.html> (working on overlaying lines)