Visualization of vaccination process of COVID-19 and related comparisons

Process Book

Group F

Part I Charts with ggplot(ly) and highcharts

Data Source
Under the first section ggplot(), ggplotly(), highchart(). All data sources are from the github repository under owid/covid19. 3 csv files are included.

Packages Used
readr/ lubridate/dplyr/ggplot2/plotly/highcharter/dplyr/ggthemes/ggrepel

Descriptions:
The first graph (page 2) simply shows how total cases cluster overtime with some states of interests according to CDC's Vaccine Tracking System. In this graph, period starting from 1/13 to 4/16 and 6 states are included. Maine did pretty well in these three months.

In the second graph (page 3), cross-sectional data as of April 15 this year were used to create the highchart where x axis represents total cases per million while y axis represents people vaccinated per hundred, and by continents using color. Notice that both the x and y axis have the original scale, with many countries concentrated at the lower left corner.

The third plot (page 4) replicates the former but goem_smooth and labels are added, and y is log transformed to make each observation as spread out as possible. The red labels represent high-risk countries where the number of total cases is larger than their median and people vaccinated are less than the median. Although the United states have the most cases, they are excluded from the high-risk target owing to high vaccination rate. There are two countries dropped out of the graph by stat_smooth function since the number of people vaccinated are too small to be log scaled.

The fourth plot (page 5) is the same thing but interactive version with ggplotly.

The fifth graph (page 6), heatmap was introduced to better represent time series data across all states in the US, and notice that the introduction of vaccine has a lagged and minor effect on the recovery from the pandemic. In addition, the color was manually rescaled due to abnormally distributed cases over time. Geom_segment is used trying to mark the boundary of vaccination effects. Tile color represents new cases every week starting from Jan. 2020. (this graph can’t be made with ggplotly() which is incompatible with horizontal legend scale.)

Part 2 Maps (U.S.)

Data Source
I did 4 graphs to show the Vaccination data in the United States. The data I use is state-by-state data on United States COVID-19 vaccinations. The data is updated daily by the United States Centers for Disease Control and Prevention.

Package Used
dplyr/plotly

Descriptions:
The first two graphs show the total number of fully vaccinated people on 1/13 and on 4/16. I first summary the total number of fully vaccinated people by the state on these two dates. Then, I use built-in data to generate location information for all states. To merge these two data frames together, I lowercase the location name. Then, I use plot_ly to make these two graphs.

The 3rd graph shows changing daily vaccinations in 3 months. Daily_vaccinations measures new doses administered per day. For countries that don't report data on a daily basis, we assume that doses changed equally on a daily basis over any periods in which no data was reported. I use the number of daily vaccinations on 4/15/2021 divided by the number of daily vaccinations on 1/15/2021 to analyze the changes in the number of daily vaccinations.

The 4th graph shows total vaccinations per hundred on 4/16/2021. People_vaccinated measures the total number of people who received at least one vaccine dose. If a person receives the first dose of a 2-dose vaccine, this metric goes up by 1. If they receive the second dose, the metric stays the same. People_vaccinated_per_hundred measures people_vaccinated per 100 people in the total population of the state. We might predict the people fully vaccinated per hundred in the future by analyzing this graph.

Part 3 Maps (World)

Data Source:
https://github.com/owid/covid-19-data/blob/master/public/data

Package Used:
ggplot2/ggthemes/rgdal/plyr/dplyr/maps/maptools/plotly/threejs/readr/dplyr/scales/readxl

Descriptions:
Despite the previous introduction of my teammates, we also would like to explore the process of global vaccination.

We can see from the map in Page 11 that Britain and the United States have relative high vaccination rates compared to other countries in the world while countries in Africa have relative low vaccination rates.

Next, I made an interactive global map from which we could see the vaccination rates in different places clearly on a globe in Page 12.

And, we could see more detailed data on this different interactive map in Page 13.

Also, for the amounts of vaccinations each country has, I made another map indicating the absolute amount comparisons between countries. Here we can see that the United States, China and India has the largest amounts of vaccinations in Page 14.

And, if you would like to study on the specific amounts, here we prepared an interactive map for you in Page 15.

Part 4 Sentiment Analysis

Data Source:
The text analysis part used the New York Times API to collect data from March 2020 to August 2020, which is the time from the initial outbreak of the Covid-19 to the time before the president election.

Package Used:
library(magrittr)
library(wordcloud2)
library(tidytext)
library(topicmodels)
library(plyr)
Also, we used python to crawl all the data from New York Times and clean up all the words.

Descriptions:
The first graph is a word cloud of words that appeared most frequently in the news reports in the past six months since the outbreak of the Covid-19.

The second graph used LDA models to show the topic words in the all the lead paragraphs of the news reports.

In the last graph, we counted the top five words according to word frequency among the articles published by New York Times in each month, and also the specific number of times that they appeared.