Final Project - COVID

Joe Harter

2024-04-27

Pre-requisites   Be sure to have the tidyverse package installed before knitting this document.

Question

How much did vaccinations help reduce deaths globally?

The data

```
base_url <- 'https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_cc
files <-c("time_series_covid19_confirmed_US.csv", "time_series_covid19_confirmed_global.csv", "time_ser
urls <- str_c(base_url, files)
```

We were given four different datasets from Johns Hopkins University that contained daily confirmed cases and deaths. For my analysis I chose to focus on the global deaths dataset (https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv)

Global Deaths Dataset   The global deaths dataset was pretty straightforward. Every row is a country or a province/state within a country. There are columns for lat, long, and then one column for every date from 1/22/20 to 3/9/23 when they stopped collecting data.

I needed to get a summary of all of this data to get the actual global numbers.

Vaccine Data   I found that the WHO makes datasets available through Github at https://github.com/owid/covid-19-data/. In that repository they have vaccination data for 233 locations, along with a set of rows for the entire world, and even for the 2022 Winter Olympics. There were 16 different variables, 13 of which were different ways to count vaccinations. There were total vaccinations, number of people vaccinated, and number of people fully vaccinated among others. I decided that "number of people fully vaccinated" was the variable I wanted to pay attention to as it was touted by the CDC and WHO as the best way to protect ourselves from hospitalization and death.

Here are the top 10 countries by number of fully vaccinated people (more populated countries will appear higher). Africa was listed as a whole continent but I decided to keep it in as we're not really comparing countries here
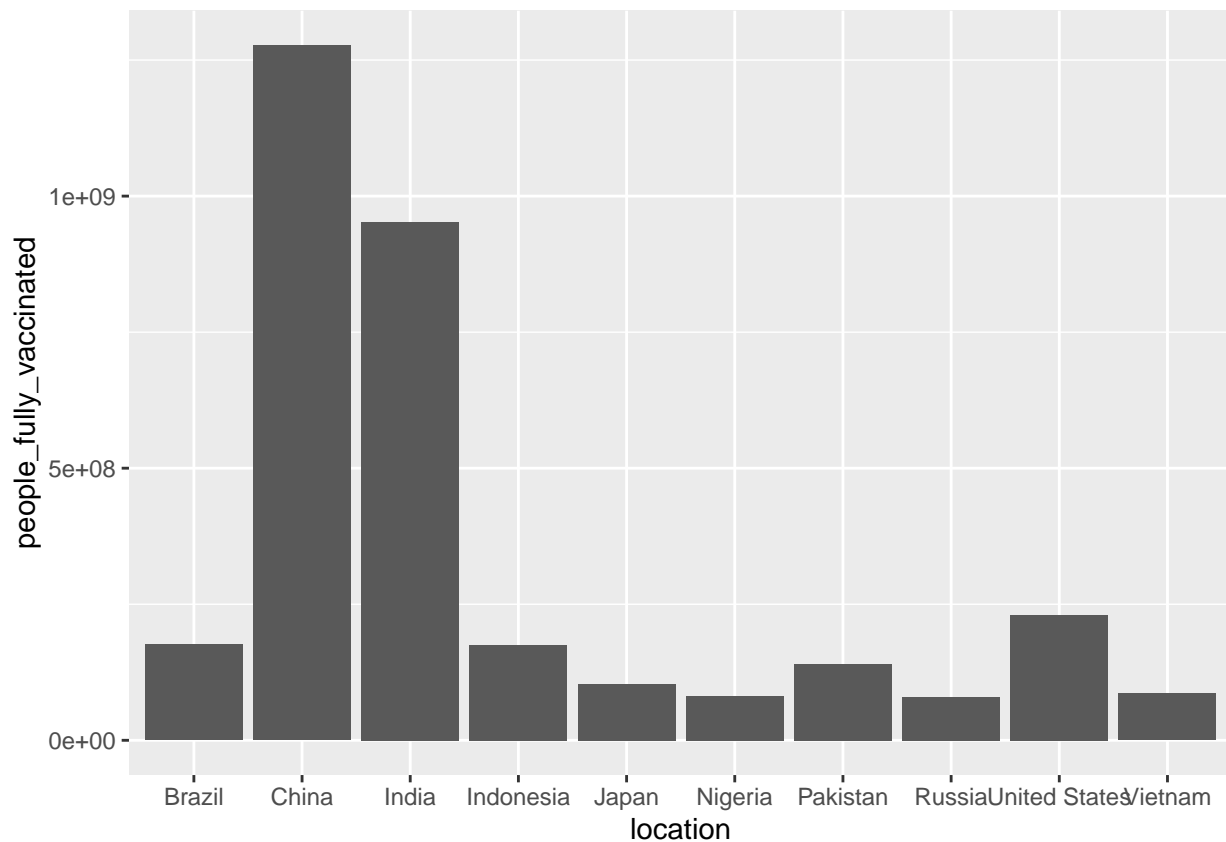
```
base_url <- "https://github.com/owid/covid-19-data/"
csv_route <- "blob/master/public/data/vaccinations/vaccinations.csv?raw=true"
```

```
global_vaccines <- read_csv(paste(base_url, csv_route, sep=''))

# Get the last row of each location so we can get the final count. Omit rows with any NA
global_vaccine_w_total <- na.omit(
  global_vaccines %>%
    select(c("location", "date", "people_fully_vaccinated")) %>%
    group_by(location) %>% arrange(date) %>% slice(n()))
top_10_countries_vaccinated <- top_n(
  global_vaccine_w_total %>% ungroup() %>%
    filter(!location %in% c("Low income", "South America", "Africa", "Oceania",
                            "Asia", "North America", "Europe", "World",
                            "High income", "European Union",
                            "Upper middle income", "Lower middle income")),
    10, people_fully_vaccinated)
ggplot(data = top_10_countries_vaccinated,
       aes(x=location, y=people_fully_vaccinated)) +
  geom_bar(stat="identity")
```



```
show_col_types = FALSE
global_deaths <- read_csv(urls[4])
vaccine_url <- "https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/vaccinations
global_vaccines <- read_csv(vaccine_url)

# We stopped getting covid death/cases data on March 9 2023 so we can filter dates after that
global_vaccines <- global_vaccines %>%
```

```
  filter(location == "World", date <= "2023-03-09") %>%
  select(c("date", "people_fully_vaccinated"))
total_of_global_deaths <- data.frame(colSums(Filter(is.numeric, global_deaths)))
total_of_global_deaths$date <- row.names(total_of_global_deaths)
total_of_global_deaths <- total_of_global_deaths[-1:-2,]
total_of_global_deaths <- total_of_global_deaths %>% mutate(date=mdy(date))

merged_covid_data <- merge(
  total_of_global_deaths, global_vaccines, all=TRUE) %>%
  rename(global_deaths = colSums.Filter.is.numeric..global_deaths..)
```
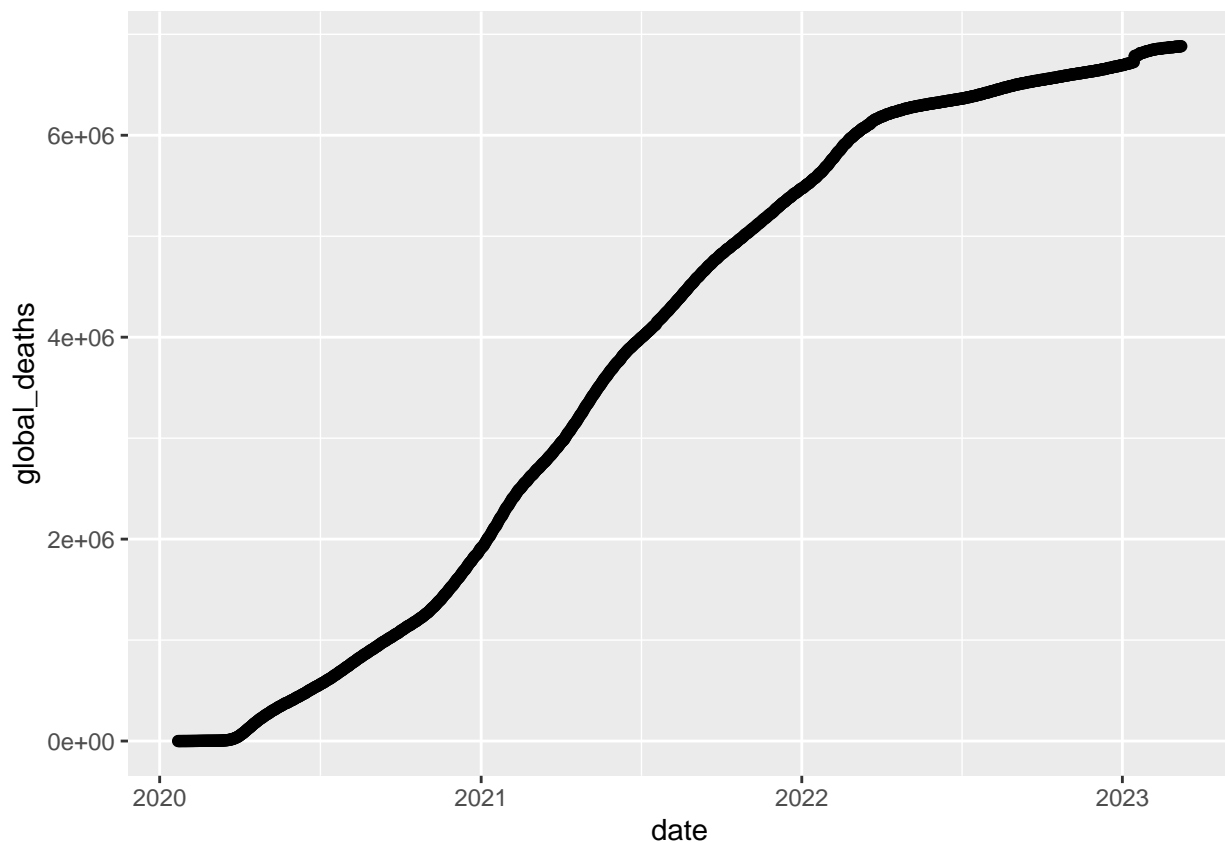
Analysis

How did COVID affect the world?

   Anyone reading this around the time of publication will be very familiar with life during COVID, but sometimes it can be sobering to put the number of lives lost into perspective globally.

```
ggplot(merged_covid_data, aes(x = date, y = global_deaths)) +
    geom_point()
```


In just 3 years nearly 7 million people lost their lives to this new virus.

How did vaccinations effect the deaths world wide?   I'm afraid I don't have the expertise yet to properly show how the rate of global deaths changed as the vaccines rolled out. For now we will have to "eyeball" it given this following plot.

```
merged_covid_data_w_lag <- merged_covid_data %>%
  mutate(diff_deaths = global_deaths - lag(global_deaths,
                                           default=first(global_deaths)))

 ggplot(merged_covid_data,
        aes(x = people_fully_vaccinated, y = lag(global_deaths))) +
      geom_point() +
      stat_smooth(method = "lm")
```
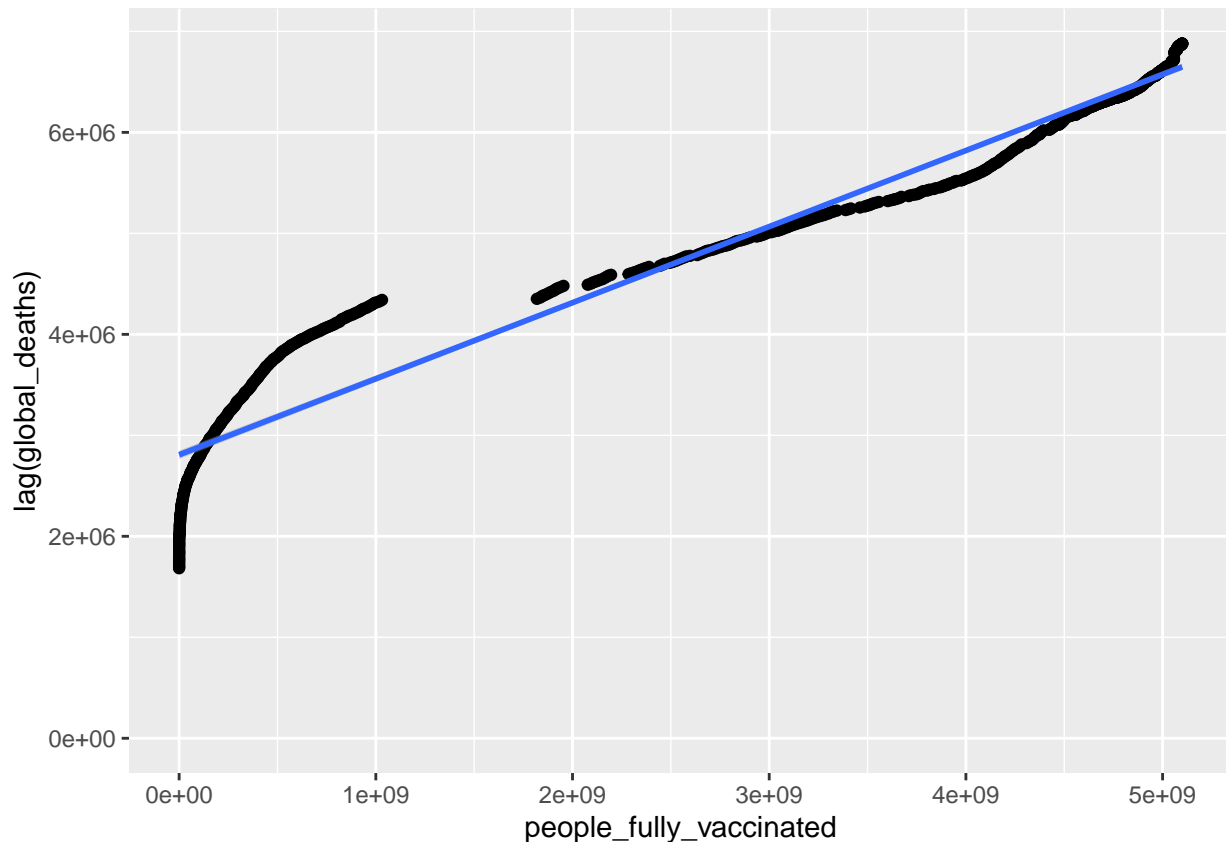

As you can see the slope is very steep early on when fewer people were vaccinated and it quickly become more shallow and almost appears to hit an equilibrium. I wish I had the statistical chops to better understand how these two variables should work together. Perhaps I'll get there throughout this degree program.

Bias

The datasources were from reputable sources that had rigorous standards for reporting. However, we must recognize that there is a serious imbalance in the world among who can test and report accurately and who can't. Poorer countries may not have the infrastructure or funds to make tests available or tabulate data in a reliable way. We should also consider poorer areas in each country would struggle with the same thing.

Although the model used above was mostly due to the limitations of the author we should also consider that modeling the data in the way I chose may not have been the one most likely to show what we are looking for.

Conclusion

While it is common-sense that vaccinations would reduce deaths and more educated people have done the proper analysis to prove so, I sought out to see what I could find with the data we have.

Based on this analysis can we say much about how the vaccine effected deaths? It certainly looks like when the vaccine first came out it quickly flattened the curve, but then had less of an effect over time. I found this surprising. On one hand I may just not be applying the right model. However, there are also more factors that would affect deaths globally than simply what the total vaccination count is. Perhaps this isn't due to the vaccine not being effective, but instead it could signify that there were large areas of the world that never took up the vaccine. For example, the United States and Europe became widely innoculated very early. Australia was heavily locked down. Their peak of deaths were behind them. But poorer countries may never have hit the threshold necessary to meaningfully reduce deaths. More analysis will be needed to draw any strong conclusions that would run counter to the mainstream narrative.