Remote Work During COVID

Joseph Lennan

2022-09-27

Methodology

The initial target of study of this project was looking into possible relationships between remote work and COVID. Specifically, how does the danger COVID possesses as shown in data and graph relate to the level of remote work within the United States?

This project was also aimed at practicing working with data analysis tools and the skills themselves. In order to practice and learn how to communicate data between RStudio, SQL, and Python, I used DBI to fetch data from SQL which I had sent over to it from Python. I also had to format and work with the data using pandas in Python. Later in the project, I decided to incorporate vaccination data which led me to use httr to directly import the data through an API into RStudio.

Analyzing and managing the data once it was in RStudio was done through the use of libraries, namely dplyr, tidyr, lubridate, janitor, and skimr. Some tools were steadily added as I discovered more tools during my analysis and discovered better practices. Notably, had I known to lay more groundwork and better established what data columns to align for the sake of analysis, I would have done a better job setting up the dataframes.

In order to test my hypotheses and create graphs to give insight into the relationship between COVID and remote work, I used ggplot2 and aligned the graphs using patchwork such that the dates would line up and more clear.

Data Cleaning and Preparation

The first step after importing the data is making sure it is intact and able to be utilized. Manually observing the imported data and comparing it to the initial, raw data gives a brief understanding of whether or not the data was correctly transferred.

One problem that initially occurred and was observed during this phase was that not all of the data was initially being imported, since the API standard setting for limiting the data led to not receiving all observations. This was fixed after reading up on how to send the desired limit of data to the API.

After verifying the data, the next process was preparing, standardizing and classifying the data such that the different numerical, date, and character data can be compared and utilized.

Column names were set if necessary, janitor was used to format the column names, then columns were established as Numeric, Date, or Character type.

Within the COVID data is some negative values that were removed, namely new cases and new deaths, but kept in the case of total cases and deaths. The decision I made for this is that there is uncertainty whether these negative numbers are corrections or mistakes, leading to the choice of setting those specific columns to zero, rather than removing their rows entirely.

```r
## T1 Editing ----------------------------------------------------------------------

#Column Names

colnames(T1) = c('characteristic',
                 'total_employed',
                 'total',
                 'percent_of_total_employed',
                 'total_employed_Percent',
                 'teleworked_because_coronavirus',
                 'date')

T1 = clean_names(T1)

#Date Conversion
T1$date = my(T1$date)

#Numeric Conversion
T1[,c(2,3,4,5,6)] = sapply(T1[,c(2,3,4,5,6)],as.numeric)

## Covid Editing -----------------------------------------------------------

#Column Names
covid = clean_names(covid)

#Date Conversion
covid$submission_date = as.Date(covid$submission_date,"%Y-%m-%d")

#Numeric Conversion
covid[,c(3,6,8,9)] <- sapply(covid[,c(3,6,8,9)],as.numeric)

#Clean negative values from new deaths and cases - leaving their effects on total_case and total_death
covid[covid$new_case<0,6] = 0
covid[covid$new_death<0,9] = 0

#Year-Month For Grouping and Joining
covid = mutate(covid,submission_date_ym = format(submission_date, "%Y-%m"))

## BRS Editing -------------------------------------------------------------

#BRS2020


#Remove Table Name Column
brs2020 = brs2020[1:34,2:6]

brs2020 = clean_names(brs2020)



#BRS2021
#BRS Columns - Specify Standard Errors
brs2021 = clean_names(brs2021)
```

```r
brs2021 = rename(brs2021,standard_error_PoE = unnamed_5,standard_error_NoE = unnamed_7,standard_error_Pe

#Break into subtables
brs2021_subtables = list(
  Telework = brs2021[c(2:6),],
  Workplace_Flexibilities = brs2021[c(7:12),],
  Changes_in_pay = brs2021[c(13:18),],
  COVID_19_workplace_requirements = brs2021[c(19:26),],
  Establishment_space_size = brs2021[c(27:32),],
  Relocation = brs2021[c(33:40),],
  Supplementing_workforce = brs2021[c(41:52),],
  Automation = brs2021[c(53:63),],
  Drug_Testing = brs2021[c(64:66),],
  COVID_19_loans_grants = brs2021[c(67.68),]
)

## Vaccine Editing ------------------------------------------------------------


#Filter US Only Data

Vacc_Data_US = vacc_parsed%>%filter(location=="US")
Vacc_Data_US = clean_names(Vacc_Data_US)

Vacc_Data_States_Territories = vacc_parsed%>%filter(location!="US")

#Convert Date from POSIXct to Date for Graphing
Vacc_Data_US$date = as.Date(Vacc_Data_US$date)

#Data Prep - Add columns for analysis
Vacc_Data_US$variant = "Alpha_Gamma_Beta"
Vacc_Data_US$variant = ifelse(Vacc_Data_US$date>=as.Date("2021-02-23"),"Delta",Vacc_Data_US$variant)
Vacc_Data_US$variant = ifelse(Vacc_Data_US$date>=as.Date("2021-12-01"),"Omicron",Vacc_Data_US$variant)


#Month Groupings

Vacc_Data_US$month_group = "january_april"
Vacc_Data_US$month_group = ifelse(month(Vacc_Data_US$date) %in% c(5,6,7,8),"may_august",Vacc_Data_US$mon
Vacc_Data_US$month_group = ifelse(month(Vacc_Data_US$date) %in% c(9,10,11,12),"september_december",Vacc_
```

Data Filtering and Joining

After cleaning and preparing the data, the data can be used such as adding together numerical data which leads to focusing on 16 years and over data in the current population survey (T1), with the intent to compare the general population against our COVID data.

By using skimr to glance at the data, we can see rows, columns, means, mins, max, medians, unique values, deviations, completion rates and missing values. This is both useful for further verification of the Data as these can help catch when data is cause for concern.

```r
#Only summarizing Covid Confirmed Totals/Cases/Deaths, not probable ones
Country_Totals = summarise(group_by(covid,submission_date),
                           country_new_cases = sum(new_case),
                           country_new_death = sum(new_death),
                           country_tot_deaths = sum(tot_death),
                           country_tot_cases = sum(tot_cases),
)

#Group COVID Data by Month, to join Covid Data With Remote Work on date
Country_Totals_Month =
  Country_Totals %>%
  group_by(date = floor_date(submission_date,"month")) %>%
  summarize(country_new_cases_month = sum(country_new_cases),
            country_new_deaths_month = sum(country_new_death),
            )

#Summarize T1 Total, (16+)
T1_Total = T1%>%filter(characteristic == "Total, 16 years and over")
```

```r
#Join Covid Data With Remote Work on Date - Country_Totals_Month + T1_Total
T1_Covid_Table = left_join(T1_Total,Country_Totals_Month,by=c("date"))
```

```r
#Mutate new columns for analysis - Notable Variants First U.S. Infection
  #From Start to 09-Mar-2022 (Alpha-Gamma-Beta)
  #From 09-Mar-2022 to "07-Jun-2022" (Delta)
  #From "07-Jun-2022" to Current (Omicron)
  #Sources:
    #https://www.cdc.gov/museum/timeline/covid19.html
    #Delta First Infection Date - https://en.wikipedia.org/wiki/SARS-CoV-2_Delta_variant
    #https://covid.cdc.gov/covid-data-tracker/#variant-proportions
    #https://dhhs.ne.gov/Documents/COVID-19-Genomics-Data.pdf

Country_Totals$variant = "Alpha_Gamma_Beta"
Country_Totals$variant = ifelse(Country_Totals$submission_date>=as.Date("2021-02-23"),"Delta",Country_T
Country_Totals$variant = ifelse(Country_Totals$submission_date>=as.Date("2021-12-01"),"Omicron",Country_

T1_Total$variant = "Alpha_Gamma_Beta"
T1_Total$variant = ifelse(T1_Total$date>=as.Date("2021-02-23"),"Delta",T1_Total$variant)
T1_Total$variant = ifelse(T1_Total$date>=as.Date("2021-12-01"),"Omicron",T1_Total$variant)


#Month Groupings

Country_Totals$month_group = "january_april"
Country_Totals$month_group = ifelse(month(Country_Totals$submission_date) %in% c(5,6,7,8),"may_august",C
Country_Totals$month_group = ifelse(month(Country_Totals$submission_date) %in% c(9,10,11,12),"september_

T1_Total$month_group = "january_april"
T1_Total$month_group = ifelse(month(T1_Total$date) %in% c(5,6,7,8),"may_august",T1_Total$month_group)
T1_Total$month_group = ifelse(month(T1_Total$date) %in% c(9,10,11,12),"september_december",T1_Total$mon
```

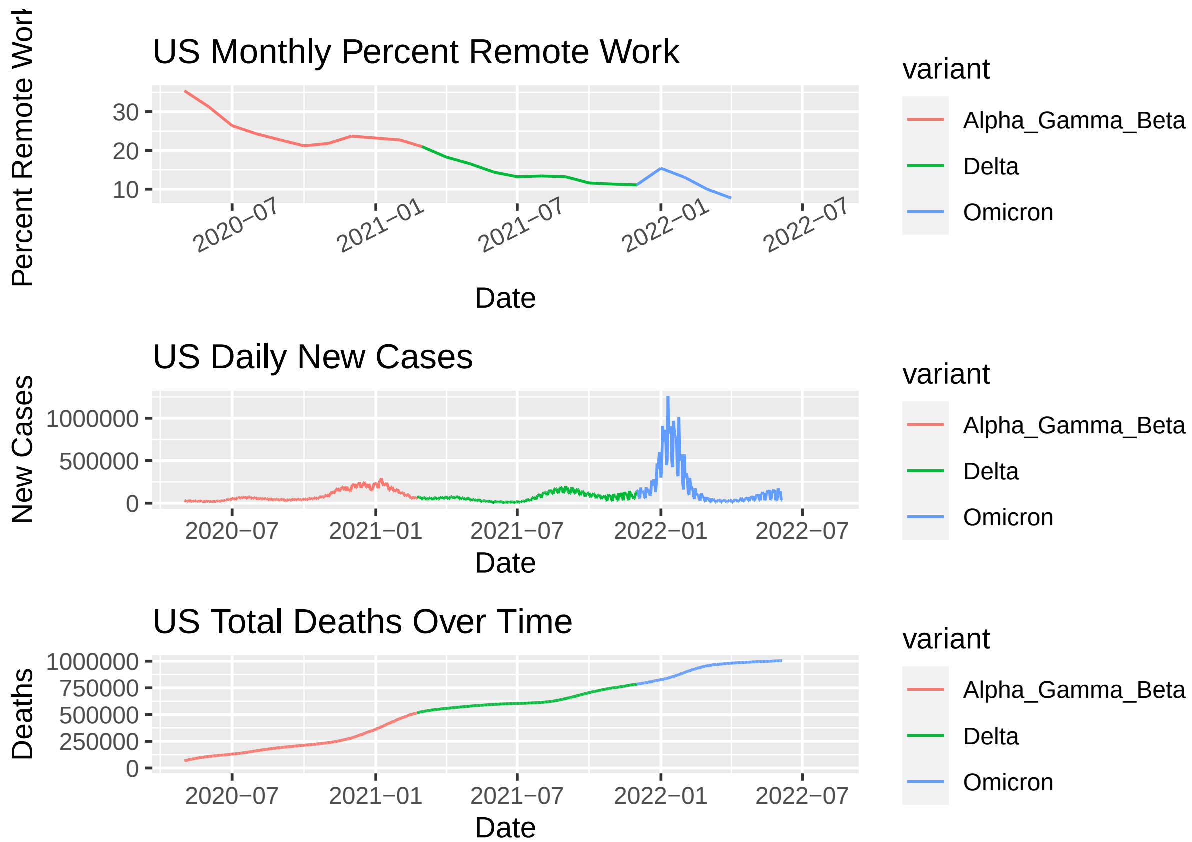
Graphs

The data was examined with help from skimr and prepared for filtering such that it was possible to focus on the most important columns.

We also can add additional information based on column data, such as assigning assigning what was the latest first infection of among the variants of Alpha, Delta, or Omicron. I also grouped months together to graph in an effort to see if the holiday seasons might have an impact on either remote work or COVID statistics.

With that work done, graphs are the next step in discovering and illustrating relationships. I created a variable, date_range, with the potential to limit the range of dates displayed on the graphs, although I ended up not having a large purpose to focus or zoom in on certain sections.

```
## Warning: Removed 100 row(s) containing missing values (geom_path).
## Removed 100 row(s) containing missing values (geom_path).
```



From these date aligned graphs we can see a steady decrease in remote work despite total deaths from COVID rising at the same rate, the remote work rate only seems to react in alignment with a surge in cases from a new variant.

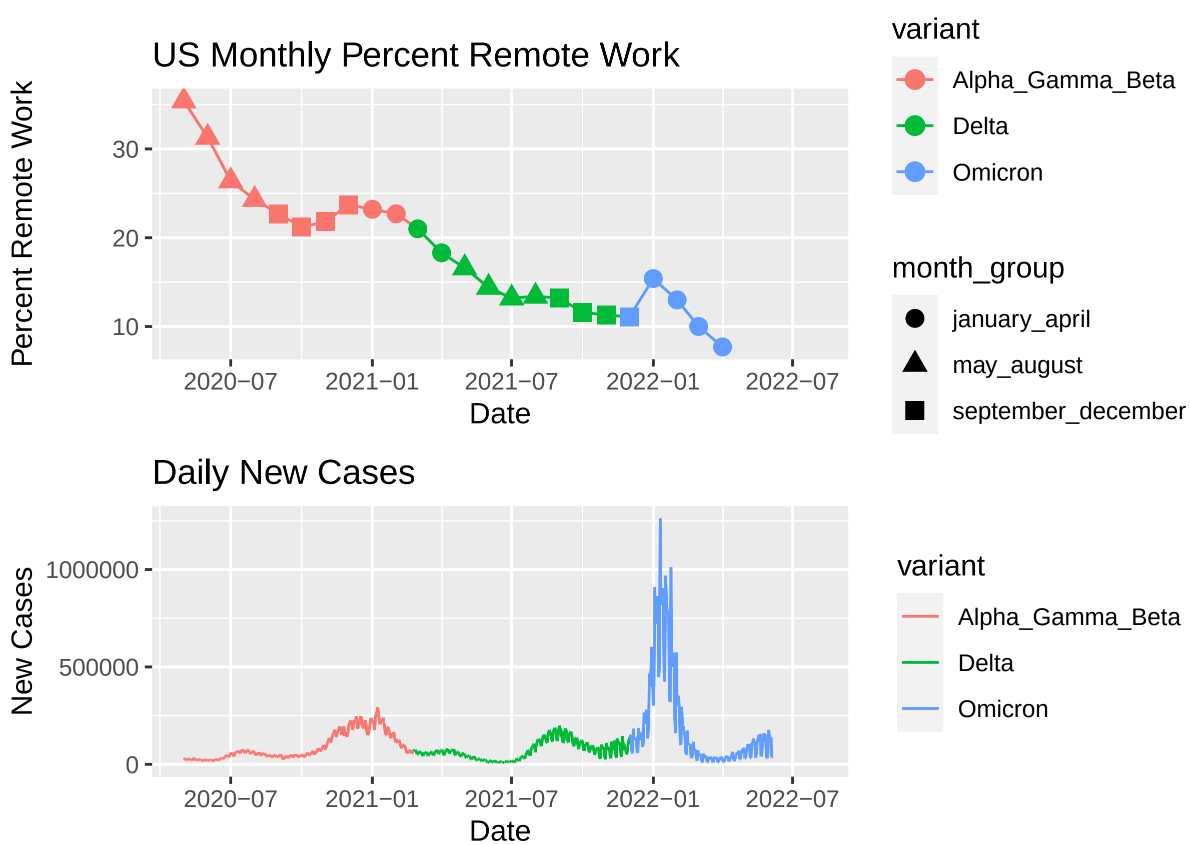
This steady decline of remote work in the face of fairly consistently increasing total deaths to COVID seems to speak a great deal about how fatigue about the COVID situation has led to a lack of protection and caution, especially when paired alongside the daily new cases graph and the knowledge of how variants like Omicron have an incredibly large potential to infect new cases.

Future studies could look at what factors lead to this surge in infections. Was this just a combination of holiday season and emergence of a new variant? Are we seeing higher rates due to the COVID benefiting from certain temperatures or other factors?

```r
T1_Graph_Month_Variant =
  ggplot(T1_Total, aes(x = date, y = percent_of_total_employed,group=1,color=variant)) +
  geom_line(size = 0.5) +
  scale_x_date(limits=date_range) +
  geom_point(aes(x=date,y=percent_of_total_employed,group=1,shape = month_group), size = 3) +
  #Angling X Test for Readability
  xlab('Date') +
  ylab('Percent Remote Work') +
  ggtitle('US Monthly Percent Remote Work')

Country_New_Cases_Graph_Month_Variant =
  ggplot(Country_Totals, aes(x = submission_date, y = country_new_cases,group=1,color=variant)) +
  geom_line(size = 0.5) +
  scale_x_date(limits=date_range) +
  #Angling X Test for Readability
  xlab('Date') +
  ylab('New Cases') +
  ggtitle('Daily New Cases')


T1_Graph_Month_Variant / Country_New_Cases_Graph_Month_Variant
```



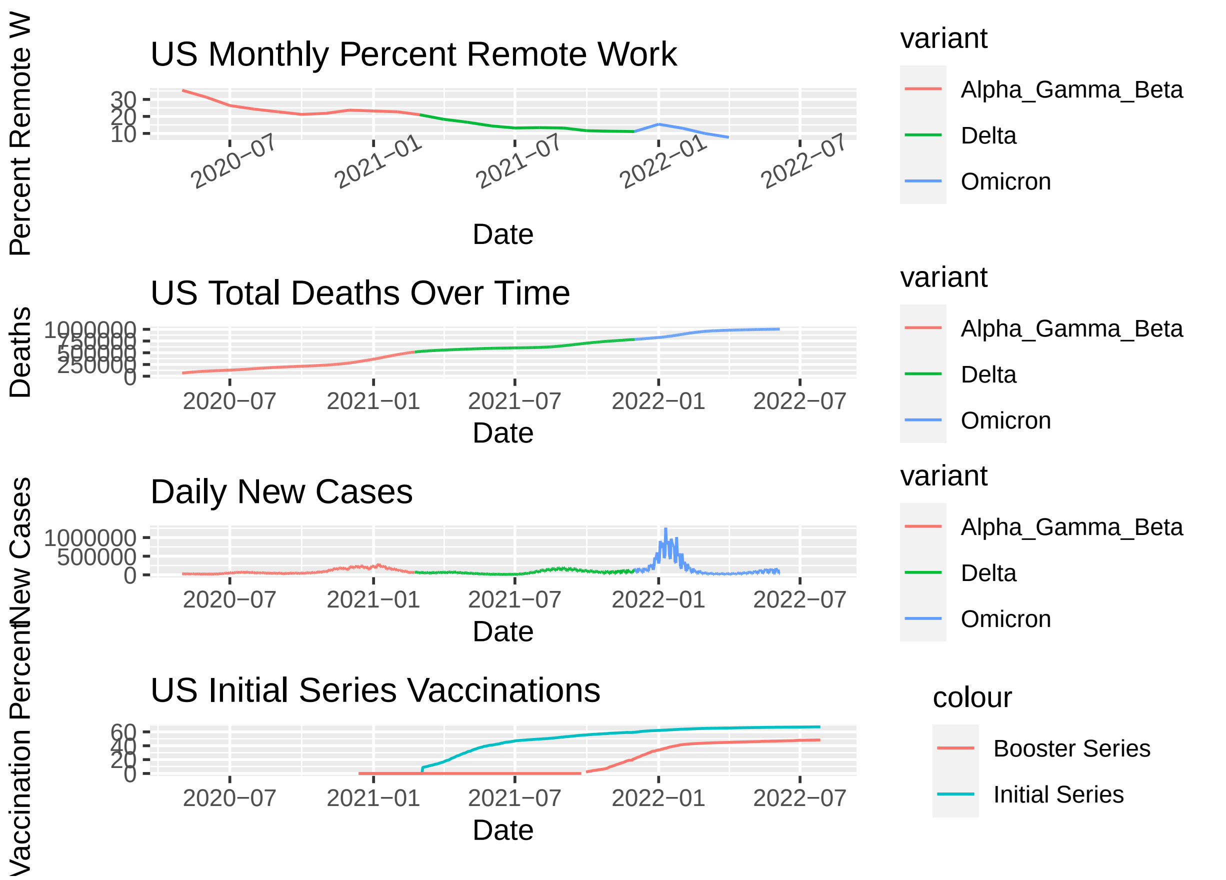
The above graph displays both variants alongside groupings of months for easier understanding, attempting to understand if there is a summer or holiday reaction by either remote work or COVID infections

At 2022-01 we can see there was a large spike, possibly the result of Omicron hitting towards the end of the holiday season, stronger compared to the Alpha/Gamma/Beta holiday spike, possibly due to the timing of omicron and holidays coinciding.

```
T1_Total_Graph / Country_Deaths_Graph / Country_New_Cases_Graph_Month_Variant / Vacc_US_Pct
```

```
## Warning: Removed 100 row(s) containing missing values (geom_path).
## Removed 100 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 9 row(s) containing missing values (geom_path).
## Removed 9 row(s) containing missing values (geom_path).
```



By looking at vaccination data and if we assume that holiday gatherings have gotten less restrictive and careful about COVID, we can begin to understand part of the reasons behind the surge and following control of the Omicron variant at around 2022-01.

As the Omicron variant hit the U.S., people began to come back from holiday travel and socialization, while they were also slowly beginning to get booster vaccination shots. From a simple Google search, COVID takes on average 5.6 days to show symptoms after contact. Thus, as the holidays faded and people began falling sick, they likely began to be tested and added to the data, causing the surge leading up to and after the new year.

In future projects, it may be valuable to investigate how the Omicron spike was brought under control, if it was simply the effect of the booster vaccinations, laws, or information and warnings to the public.

Conclusions

My conclusion is that despite surges in cases that may be a result of infectious variants, remote work appears to react slowly and only in response to large stimulus, possibly signifying a reliance on governmental actions and pressures compared to business or individual actions in the face of general fatigue about COVID and lock downs.

The total deaths over time graph shows us a steadily increasing amount of deaths in the face of varying new cases that doesn't stop the steady decrease of remote work. However, this data is with the lock downs, masks, vaccinations, and stressful conditions faced by our front line heroes, whose only reward was burnout, danger, and brief recognition. Future studies might aim to break down where critical steps could have been made to contain the pandemic in advance, where points of failure were, and how pandemic relief could have been improved.

There was a lot of disinformation during COVID, such as the belief it would be as simply easy to get over as a minor flu, as well as frustration, as the lock down and pandemic has dragged on and on. One key idea that resonated a lot while working on this project during COVID was the idea of the preparedness paradox - the idea that the more prepared we are to handle something, the more likely we are to believe the danger wasn't that significant. This is because it's easy to get tired of staying alert or prepared, to blame others for the lock downs taking so long, or to get complacent about the dangers.

Personal Thoughts

We didn't have the swiftest of lock downs across the United States, and a lot of the reactions were left to States. People reacted slowly to the government's relaying of information, shifting to a state of alarm and caution, following the social advice provided to them. However, without reinforcement and with time both reducing the fear felt and the appearance of danger to people, remote work steadily declines, only receiving small bumps seemingly in reaction to new, dangerous variants.

In hindsight and an ideal world, we should have reacted quickly and contained the infection before it became so hard to control. However, the COVID pandemic is apparently over, at least according to President Biden. Instead COVID seems to be becoming endemic - a regular disease that will stay around.

COVID seems to reflect the preparedness paradox, that despite our hard-fought victories over COVID to keep infections down, we now believe that the danger isn't that significant. Masks are less frequent and less people got the booster vaccination than the initial series. People are getting tired of this dragging on and on when we had our hopes that it'd be eradicated much, much easier.

Some people experience what is currently known as Long COVID after recovering, even if vaccinated. Long COVID currently is being studied but it often reoccurs four or more weeks after infection. People with Long COVID may feel fatigued, fever, respiratory issues, and neurological symptoms like brain fog, headache, or anxiety. For more information, please check out https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html.

Yet despite the fatigue, I believe that most people have to take this threat seriously day by day - people are still dying and getting infected. I am not a disease expert, but it certainly still seems worthwhile to take what remaining caution you can still spare towards COVID. It's hard to expect people to stay isolated forever unless you are blessed to be an extreme introvert with a well-off work from home job, but it's important to take the easy, daily cautions. Simple, everyday stuff that you can still continue to do like wearing a mask, social distancing as much as possible and comfortable, washing your hands, and avoid touching your face. At the very least, it seems important to get both vaccines both old and new as they come out to help protect yourself against the yet to be fully understood threat of Long COVID.

Workflow Improvements

There was a lot of revisions and ways to work with the data better, notably due to finding out new things and coming up with new ideas. Brainstorming longer for what data to work with one another use prior to beginning, and then identifying what common lines to draw between and connect the data with it.

Making a system or template to input searches and filters would have been interesting rather than hard coding each part of the graph. Being easily able to modify Then would have been able to make an easy system to modify graphs (i.e. replace variant with "Modifier" and set modifiers as needed in the graphs), although perhaps hard to implement in an easy and thorough way.

Sources

Covid Data: https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36

Current Population Survey (T1): https://www.bls.gov/cps/effects-of-the-coronavirus-covid-19-pandemic.htm

Business Response Survey (BRS): https://www.bls.gov/brs/data/tables/2020/home.htm https://www.bls.gov/brs/data/tables/2021/

COVID-19 Vaccinations in the United States,Jurisdiction: https://dev.socrata.com/foundry/data.cdc.gov/unsk-b7fc