Report On Proportion Of Positive Covid-19 Tests (2020-2022)

Analysis of data wrangled from CDC data

Manuel Roosevelt Lamptey, Rabeya Illyas Noon, and Tarana Ferdous

2022/07/29

Table of Contents

# Introduction

This report is aimed to provide an insight into the recent global epidemic, Covid-19 data from CDC, dated between December 17th, 2020 to July 21st 2022. CDC has been collecting different Covid-19 related data across the US since the pandemic began and published most of these data for public access. In this report, we showed the process of curating the raw data from CDC to focus on a particular subset of covid-19 related data, for example, the Covid-19 positive test results data to create a growth chart. The growth chart depicts the proportion of Covid-19 positive test results for past seven days and how it changes per month over the two years. We have provided detail of the graph to explain what it depicts.

# Curating the CDC Data

To create this report, we have used 389 raw data sets from the CDC website <https://healthdata.gov/Health/COVID-19-Community-Profile-Report/gqxm-d9w9> to explore the proportion of positive Covid-19 tests for past seven days and showed the changes as per months over two years. To execute the curating process, we followed steps including importing raw data by selecting specific portion of the data as per our interest, testing and creating final data set, converting and checking class of data set, performing ‘sanity checks’, and finally saving a new data file as a .csv file. The detail of the procedure is provided below.

## Importing raw data

After loading the required libraries, we imported 389 raw data sets to create a new data set named data\_df to execute our expected graph.

## Testing data outcome

After importing, we first tested with only one data set to check our final data set.

## Creating final data set

We applied the function we prepared above to create the final data set, data\_df with 389 data files.

## Converting and checking class of data set

Our date variable was showing as character, which we converted to date format.

## Sanity check

We first checked the number of total missing value in the newly created file. Next, we checked where the missing values are, and finally showed the file where we had three missing values.

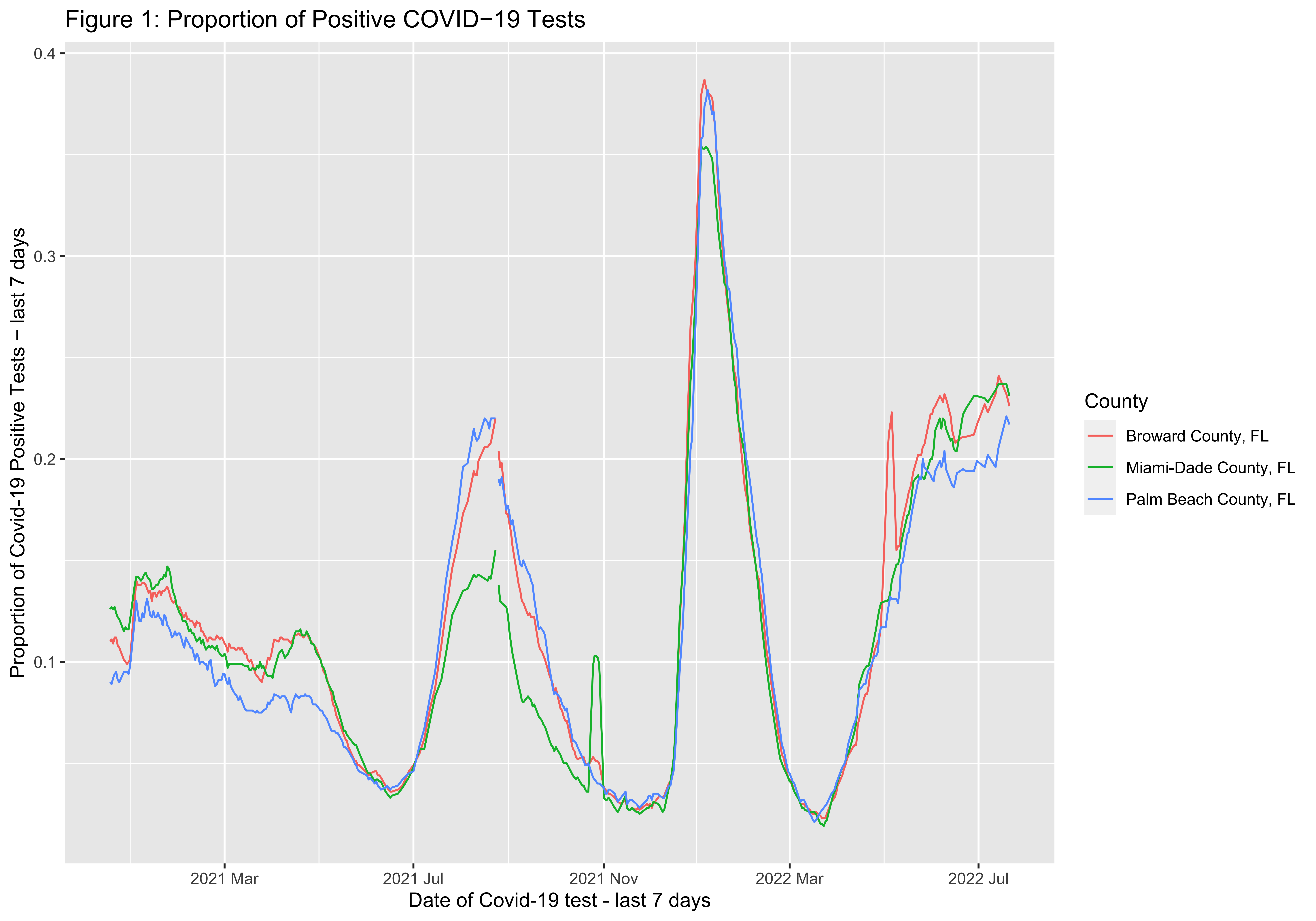
## Saving csv file

We have saved the file in .csv format.

# Creating a growth chart using ggplot

We are using growth chart to illustrate the growth of our expected variable, positivity rate of Covid-19 test, over time. It shows us the direction of positivity rate changes over the two years of pandemic from the beginning (December 2020) to until recent time (July 2022).

**Hypothesis:** We hypothesize that over time, positivity rate curve will be changing and showing high peaks during the new variant of Covid-19 virus exposure or outbreak.



**Overall comment:** In figure 1, the horizontal axis is Date of Covid-19 test-last 7 days, which is broken down into 4-months interval to show in the graph, and the vertical axis is Propotion of Covid-19 positive test - last 7 days. In addition, the color of the lines represents county for three selected countries in Florida state.

Figure 1 shows that we have data for Covid-19 positivity rate from December 2020, when the rate was around 0.10% in all three counties, but started to slightly increase to around 0.15% in January 2021. One reason could be because people were traveling during the Christmas break and exposed to the virus, and eventually tested positive after two or three weeks of the vacation. As the positivity rate increased, people were more aware, and the rate declined for the next three months. Except Palm Beach country, other two again showed a slight increase, because the population is less in the former one (1.5 million) compared to other two counties (more than 2 million), and most of the residents were older aged in Palm beach county. **Ref 1:** <https://worldpopulationreview.com/us-counties/fl/miami--dade-county-population>

The positivity rate declined to less than 0.05% by June 2021 and started to sharply increase after July 2021. From July to September 2021, the rate showed a sharp high peak, exceeding the rate more than 0.2%. This is because in this time range, the more transmissible Delta variant of Covid-19 virus started to spread, and later, despite having widespread vaccination facilities in Florida, the vaccination rate decreased by the end of August 2021. In addition, the Governor of Florida announced no mandates requiring vaccinations, masks, or social distancing, which can be considered as reasons for this sharp increase. Furthermore, the increase in rate was not same for all three counties, as it was 0.15% for Miami-Dade, and around 2.2% for other two counties. One reason could be that people in Miami were less likely to do the Covid-test compared to other two counties. **Ref 2:** <https://www.cbsnews.com/miami/news/florida-covid-19/>

In November 2021, the rate declined to less than 0.05%, and started a sharp peak in December 2021, and reached to almost 0.4%, doubling the rate than the Delta variant peak. This is the time when the new variant, Omicron strain was dominant in the US, which was more infectious than the previous strain. One good reason could be it was Christmas and winter vacation, and people were meeting families or traveling, specially from different states to Florida. Importantly, this increase was same among all three counties in Florida state. **Ref 3:** <https://www.epi.ufl.edu/covid-19-resources/covid-19-models/florida-covid-19-omicron-wave-projections-updated-0105.html>

Finally, the positivity rate started to decline after the end of January 2022, and again came down to less than 0.05% in April 2022. By this time, most of the people were vaccinated and were not following any Covid-related mandates in Florida or even in the US. However, in March 2022, the rate started to increase and as of July 2022, the Covid-19 test positivity rate is more than 0.2% in the three counties in Florida state. This is because, the new variant, BA.5 has started to dominant, which is more contagious than the Omicron variant. **Ref 4:** <https://www.yalemedicine.org/news/5-things-to-know-omicron>

**Comment on missing values:** One important issue here is Covid-19 data by CDC were missing before December 2020. During that time, the health department in the US government was politically controlled, and CDC were not allowed to publish or collect data on Covid-19 status update. It was also the election time, and politicians were concerned that people would be discouraged to go to vote if more updated information on Covid-19 were published. However, once the new government took over the lead, CDC started to update Covid-19 related data in their website in the end of December 2020. **Ref 5:** <https://www.politico.com/news/2020/09/11/exclusive-trump-officials-interfered-with-cdc-reports-on-covid-19-412809> **Ref 6:** <https://www.nytimes.com/2020/12/16/world/heres-how-the-trump-administration-crushed-the-cdc-according-to-two-who-were-there.html>

**Public health implication:** This data has significant importance in understanding the progression of Covid-19 over the two years in three different counties in Florida state. Such data can help in prioritizing the important strategies to implement suitable and timely bound public health interventions. This also helps to understand where to distribute the required resources, ultimately using the healthcare facilities precisely and effectively and facilitate healthcare manpower. Moreover, it can help in health budgeting, to consider which area to prioritize. Furthermore, looking at the different Covid-19 positive rates in different time points of the months, CDC were able to improvise their awareness strategies, announcing effective mandates and eventually helping the national health department to control the pandemic.

# Conclusion

Overall, this growth line graph shows the Covid-19 positive rate over the two years (by date) for three different counties in Florida state. The raw data were imported in R studio by using GitHub and then we imported all 389 data sets to create our required data set. During importing and creating the final data, we had to create one new column of data (positivity\_rate) from the raw data where this variable were given three different variable names, which is a limitation of using CDC data. However, using this project learnings, we were able to fix this issue and succesfully imported all data files. Finally, we created a growth chart to show the progression of Covid-19 positive tests over time for three individual counties in Florida state in the US.

# Project group members and Acknowledgment:

In this project 3, we worked in a group of three members. We are, Manuel Roosevelt Lamptey, Rabeya Illyas Noon, and Tarana Ferdous. We are from Statistics, Biostatistics and Epidemiology departments, respectively. We would like to acknowledge our TA, Catalina and Anny for their explanation on our queries and Dr. Odom for this amazing opportunity to learn through this project.