

Exploring COVID-19 Cases & Spread

Harsh Modi
Pace University

Shreshth Rathod
Pace University

ABSTRACT

COVID-19 is a serious illness caused by a contagious virus. It can spread from a contaminated surface to person or from person to person. With the rising spread of COVID-19, new cases, hospitalization, and deaths have also increased on a significant level. There are various factors that may have contributed to the rise in the spread such as number of households, population, population density, metro classification. As models are being updated on a daily basis, CDC has changed the recommendations and advised the public to wear face masks in public settings because six-feet social distancing can be difficult in densely populated area¹.

Keywords

COVID-19, Coronavirus, Infectious Disease, Disease Modeling, CDC Guidelines

INTRODUCTION

Coronavirus, also known as COVID-19, is a serious illness, caused by a contagious virus. It quickly spreads via contaminated surface-to-person contact and person-to-person contact. It has been considered as a pandemic as it has spread over multiple countries or continents and has affected numerous people in multiple different ways.

This infectious disease falls under respiratory disease. Symptoms observed amongst the infected people are in wide range which include mild fever to severe illness. It has caused numerous hospitalizations and deaths globally and particularly in the U.S.

Researchers around the world have created models to predict the trajectory of COVID-19. Models are not always perfect in making predictions, but they may be substantially useful if applied correctly³. In case of COVID-19 responding to these models may be the difference between death tolls in the thousands or millions⁴. On the basis of existing prediction models for COVID-19, CDC recommends the public to wear face masks in places with high population density because it is difficult to maintain social distancing or 6 feet distance¹.

Population, population density, number of households, county classification are some explanatory variables that may have led CDC to update their recommendation on April 13th and that may play a significant role in increasing the number of cases and spread over the period of time.

DATA SELECTION

COVID-19 cases for county level was obtained from the New York Times and GitHub database.

1. State level variables include
 - Top 50 Populated U.S. city
 - Top 50 walkable U.S. city
 - Top 50 public transit U.S. city
2. County level explanatory variables were obtained from USDA-ERS.
 - Number of households
 - Population
 - Population density
 - County classification

DATA WRANGLING

1. USDA dataset was downloaded in the initial stage, after manipulating the dataset in excel, columns taken into consideration were:
 - Number of households
 - Population
 - Population density
 - County classification
2. Total population, number of household and population density data were present in single excel file.
3. County classification column was present in another excel file.
4. GitHub-County level cases data was present in another excel file.
5. These files were merged on **FIPS code**.
6. State level explanatory variables were taken into consideration for further exploration of data. Hence, they were merged on **State**.

GEOGRAPHIC EXCEPTIONS FOR COVID CASES

1. Cases for 5 NYC counties are reported as just NYC.
2. We changed explanatory variables accordingly:
 - NYC households = sum households of 5 NYC counties
 - NYC population = sum population of 5 NYC counties
 - NYC population density = sum population of 5 NYC counties / sum area of 5 NYC counties
3. Cases for Kansas City, MO were not reported at county-level.
4. Cases for Alameda County, CA include Berkeley and the Grand Princess cruise ship.

EXCLUSION CRITERIA

We removed 23 states that have less than 5000 total COVID-19 cases on 04/24/20. We removed counties with less than

5 COVID-19 cases. Thus, 77.3% (1540/1992) of counties were included in the analysis (Figure 1).

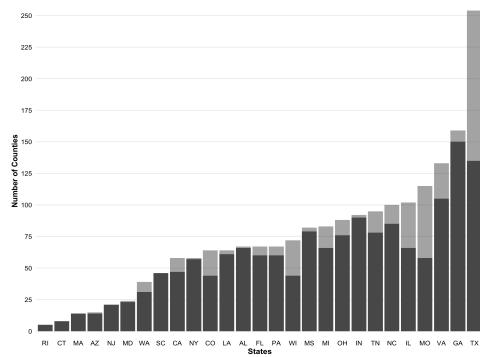


Figure 1. Bar plot displaying number of counties versus states. Dark bars represent counties with cases > 5 (1540 counties in total). Light bars represent counties with cases < 5 (335 counties) or unreported cases (117 counties). It is important to note that some counties (452) are not included in the exploratory analysis.

DESCRIPTIVE ANALYSIS

Firstly, three exploratory variables were taken into consideration to observe the impact of cases amongst the top 27 states. If any of the city falls under these categories, then that state has higher chances of getting a greater number of cases as compared to other states.

77% (21/27) of Top 27 states contain a Top 50 populated U.S. City (Figure 2).

Top 50 Populated U.S. city

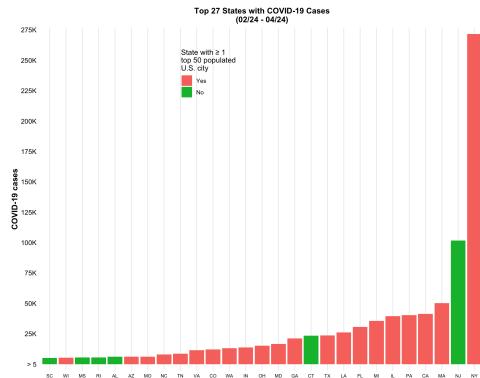


Figure 2. Bar plot of COVID-19 cases versus States.

77% (21/27) of Top 27 states contain a Top 50 walkable U.S. City (Figure 3).

Top 50 walkable U.S. city

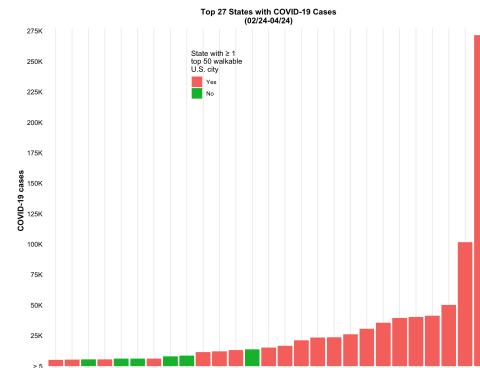


Figure 3. Bar plot of COVID-19 cases versus States.

85.1% (23/27) of Top 27 states contain a Top 50 public transit U.S. City (Figure 4).

Top 50 public transit U.S. city

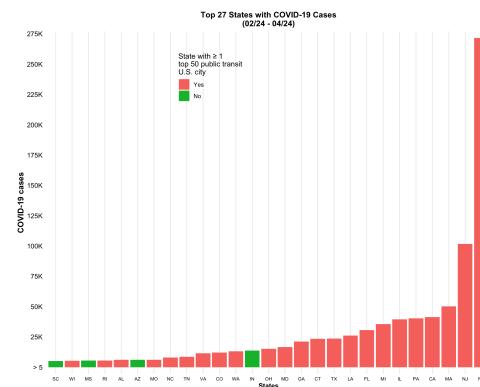


Figure 4. Bar plot of COVID-19 cases versus States.

EXPLORATORY ANALYSIS

COVID-19 Cases among Top 27 States.

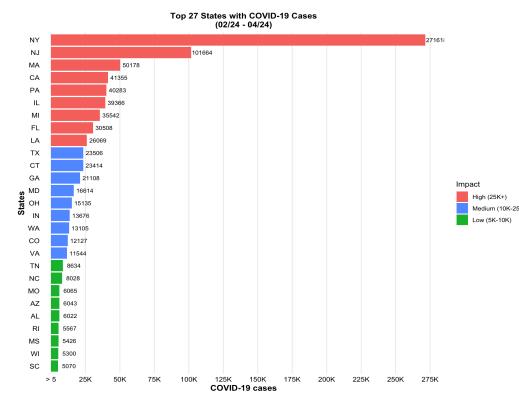


Figure 5. Case count for top 27 states

COVID-19 spread among top 27 states.

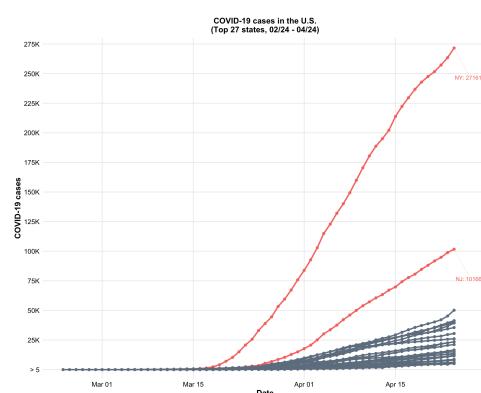


Figure 6. Here, NY & NJ are considered to be outliers and the spread for remaining 25 states is not easily interpretable in this figure.

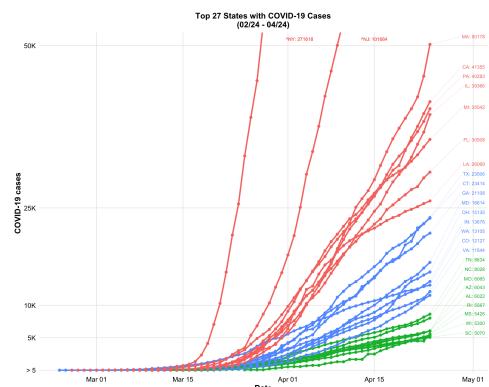


Figure 7. The main reason to plot zoomed graph was to examine the spread for the remaining 25 states. As there is an exponential growth, it is not possible to interpret what is the rate of change in cases for such large numbers. So, it is better to plot a logarithmic version of this graph.

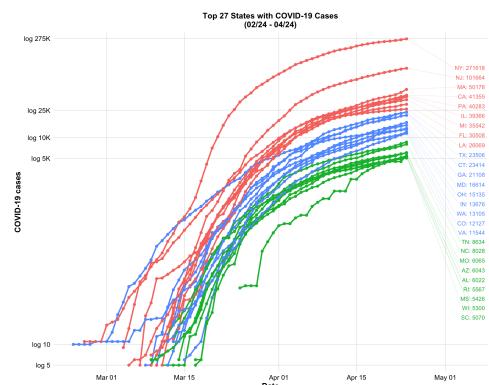


Figure 8. Log scale makes it possible to see when public health measures are starting to have the desired results. For instance, consider CDC's recommendation to wear face mask in public setting's where it is difficult to maintain social distancing or 6 feet distance from others².

EXPLANATORY ANALYSIS

- Does population affect COVID-19 case count and spread?
- Yes, COVID-19 cases and spread increase as the population increases.

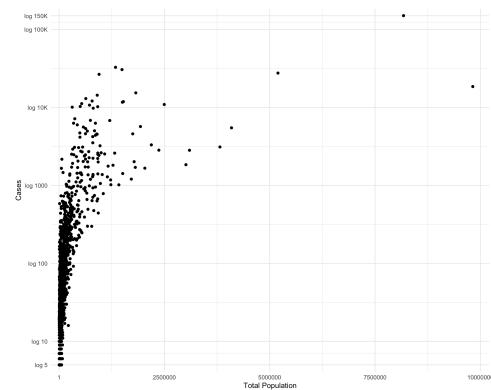


Figure 9. Scatterplot of COVID-19 cases versus population.

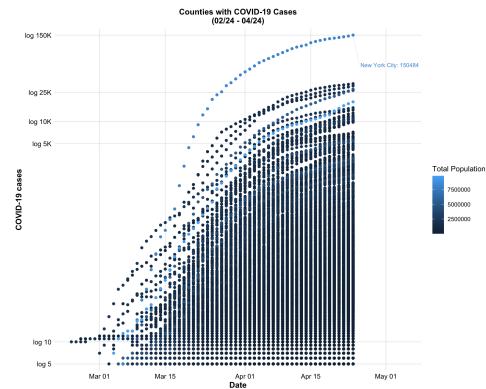


Figure 10. More populated counties are more likely to experience faster COVID-19 spread and greater number of COVID-19 cases.

- Does population density affect COVID-19 case count and spread?
- Yes, COVID-19 cases and spread increase as the population density increases.

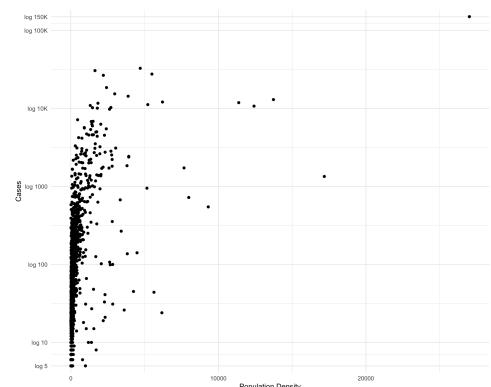


Figure 11. Scatterplot of COVID-19 cases versus population density.

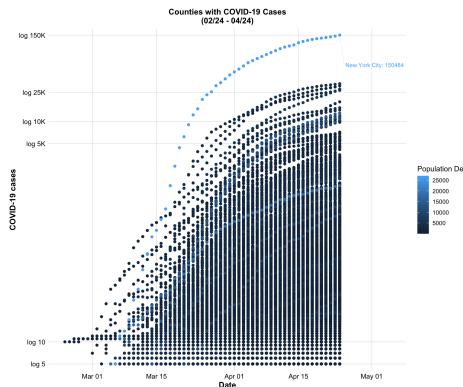


Figure 12. More densely populated counties are more likely to experience faster COVID-19 spread and greater number of COVID-19 cases.

3. Does number of households affect COVID-19 case count and spread?

Yes, COVID-19 cases and spread increase as the number of households increases.

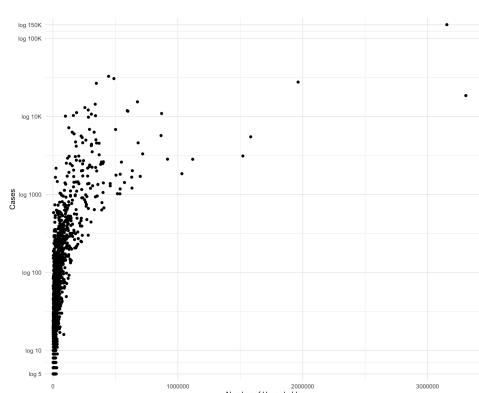


Figure 13. Scatterplot of COVID-19 cases versus number of households.

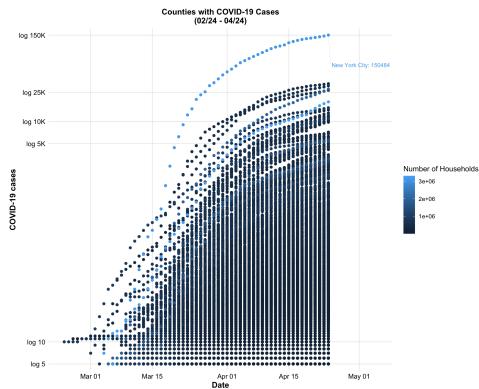


Figure 14. Counties with more households are more likely to experience faster COVID-19 spread and greater number of COVID-19 cases.

4. Does nearby public transportation affect COVID-19 case count and spread?

Yes, COVID-19 cases and spread increase as the public transportation increases.

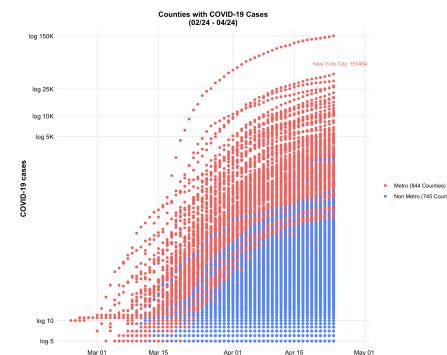


Figure 15. Metro counties are more likely to experience faster COVID-19 spread and greater number of COVID-19 cases.

5. Which state and county have the greatest number of cases and fastest spread for COVID-19?

As per the observations from the above graphs, New York state and New York city have the greatest number of cases and fastest spread for COVID-19. The variables that likely contributed to the enormous spread are:

- Population
- Population Density
- Number of households
- Metro Classification

6. Which explanatory variables could have led CDC to update their recommendations about wearing face mask in public place?

The explanatory variables that could have led CDC to update their recommendations are:

- Population
- Population Density
- Number of households
- Metro Classification

CONCLUSION

The results from this exploratory data analysis show that greater number of households, population size, population density leads to greater COVID-19 cases and faster COVID-19 spread. Metro counties have greater COVID-19 cases and faster COVID-19 spread than Nonmetro counties. NY state and NY county have the greatest cases and fastest spread for COVID-19. Therefore, the number of households, population size, population density and county classification contributed to the enormous COVID-19 cases spread. As population density increases, social distancing can be difficult. This may have led CDC to update recommendation about wearing face masks in public places on April 13th.

REFERENCES

1. Coronavirus Disease 2019 (COVID-19): Situation Summary. Centers for Disease Control and Prevention (CDC). 2020 April 19.
<https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/summary.html#background>
2. Fleming, S. What's a logarithmic graph and how does it help explain the spread of COVID-19? World Economic Forum. 2020 Apr 03
<https://www.weforum.org/agenda/2020/04/covid-19-spread-logarithmic-graph/>
3. Rogers, A. The Mathematics of Predicting the Course of the Coronavirus. Wired. 2020 Mar 03
<https://www.wired.com/story/the-mathematics-of-predicting-the-course-of-the-coronavirus/>
4. Cohut , M. Why social distancing is key in containing the new coronavirus Medical News Today. 2020 Mar 24
<https://www.medicalnewstoday.com/articles/why-social-distancing-is-key-in-containing-the-new-coronavirus>