

# **A COVID-19 Data Set Design, Modelling and Analysis on Selected Thirty Counties from Six States of the United States**

**Raissa Ji**

**Email: raissa.ji.me@gmail.com**

**Github: <https://github.com/coronavirus-2020-data/coronavirus-data>**

## **Abstract**

In this four-month-long study (from April 1, 2020 – August 1, 2020), we collected, modeled, and analyzed COVID-19 data from the top five most infected counties per top six most infected states in the United States (30 counties total). More specifically, we collected data on each state's total COVID-19 cases, deaths, tests conducted, and their counties' population, density, percentage of seniors, number of hospitals, total COVID-19 cases, and total COVID-19 related deaths. In this study, we have models illustrating the growth of COVID-19 cases and deaths per county, growth of COVID-19 cases and deaths per state (which is really the sum of our chosen five counties), and growth of COVID-19 tests conducted per state. In addition, our study also contains models illustrating the statistics of several variables that might have affected a county's COVID-19 data, which has been mentioned above: population, density, percentage of seniors, and number of hospitals. An interesting finding we have noticed upon modeling the 30 counties' density and total COVID-19 cases as an xy scatter plot is that there is a considerably strong relationship between the two variables. Los Angeles County (which was an extreme outlier), in particular, supports the idea that a county's most populous city can greatly affect its entire county's COVID-19 cases; if the largest city is extremely dense, it appears that the entire county has a greater total COVID-19 case count.

**Keywords:** COVID-19, Coronavirus, Population, Density, Hospital, Pandemic.

## **1. Motivation**

Seventeen years since the last global pandemic, on January 30, 2020, the novel coronavirus, or COVID-19, has been declared by the World Health Organization (WHO) as a global health emergency [1]. The supposed first human cases of COVID-19 have been identified in December 2019, in Wuhan city, Hubei Province, China—many of these cases have all been linked to visiting the Huanan Seafood Wholesale Market in Wuhan, which sold live animals such as poultry, bats, marmots, and snakes [2]. However, earlier cases of COVID-19 have later been discovered, starting from November 17, 2019, with no relation to the seafood market [3]. Following that day, several new cases have been reported daily, leading up to many more cases than the suspected first case that December. Though China imposed a lockdown and travel ban in Wuhan and other cities in Hubei to mitigate outbreak of the virus, the virus spread globally at an alarming rate, in which the WHO declared COVID-19 a pandemic on March 11, 2020 [4]. On March 13, 2020, US President Donald Trump declared a national emergency [5], with the White House issuing social-distancing guidelines on March 16 [6]. Shortly after, New York, New Jersey, Massachusetts, Michigan, Pennsylvania, and California, the six focus states of this study, all entered state-wide lockdown [7]. The purpose of this study is to track and analyze each focus state's top five infected counties' COVID-19 related data. We want to then compare several variables which may have affected those data, form a conclusion, and suggest a solution to possibly lessen the growth of those numbers.

## **2. Data Set Coverage**

As of April 1, 2020, our study has been collecting data on the top six most COVID-19 infected states in the United States, along with their respective top five most infected countries: New York, with counties Bronx, Kings, Nassau, Queens, and Suffolk; New Jersey, with counties Bergen, Essex, Hudson, Passaic, and Union; Massachusetts, with counties Essex, Middlesex, Norfolk, Suffolk, and Worcester; Michigan, with counties Genesee, Macomb, Oakland, Washtenaw, and Wayne (including Detroit); and California, with counties Los Angeles, Orange, Riverside, San Diego, and Santa Clara.

Since May 27, 2020, we have been publishing our data, along with various graphs illustrating the data collected, on GitHub: <https://github.com/coronavirus-2020-data/coronavirus-data>.

### **3. Data Modelling Schema**

Using data from the United States Census Bureau, our study has collected data on each county's population, population of people over 65 years old (elderly population), and density. From the Homeland Infrastructure Foundation-Level Data, we've collected data on each county's number of hospitals. These hospital and population statistics have all been modeled as bar graphs. Additionally, our study has tracked and modeled various data regarding COVID-19 cases, deaths, and tests, all with respect to time (days).

To start, we've modeled five line graphs (one graph per state) to visualize the daily increase of COVID-19 cases per state's top 5 infected counties. We summed the daily cases per county to create another line graph that compares the increase of COVID-19 cases per state (which only includes the top five infected counties). Following that, we summed the daily cases per "state" and graphed the values as a bar graph. For the daily increase of COVID-19 deaths, we created the same 7 graphs. Our study modeled the increase of COVID-19 tests per state as line graphs, and summed those values to create a bar graph showing the total daily tests conducted by those states.

### **4. Hospital and Population Statistics of Selected Counties**

This section contains data on each county's population, number of hospitals, population of the 65 years and older demographic (seniors), and density (population per square mile) [8, 9]. There are individual graphs illustrating these data which will later be analyzed to see if they have any strong correlation with the county's COVID-19 data.

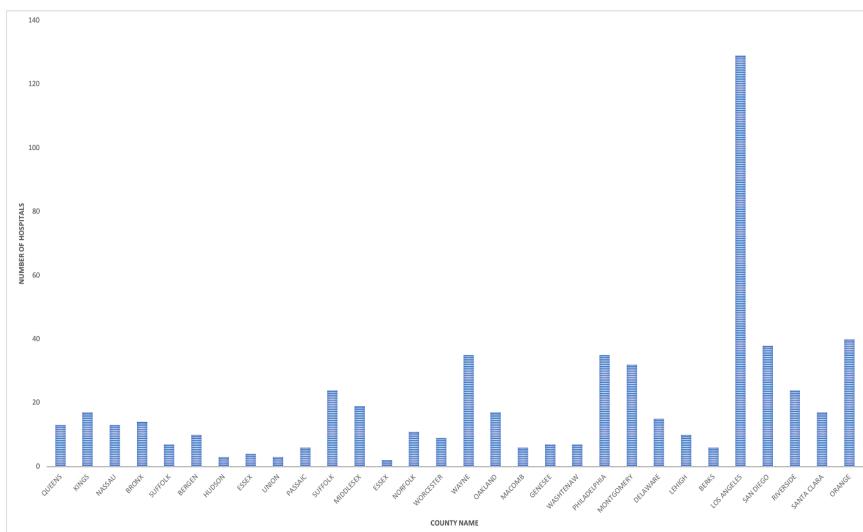
County	Population*	Density***	Percentage of Seniors*	Number of Hospitals**
Queens County, NY	2253858	20553.6	16.3	13
Kings County, NY	2559903	35369.1	14.4	17
Nassau County, NY	1356924	4704.8	18.2	13
Bronx County, NY	1418207	32903.6	13.3	14

Suffolk County, NY	1476601	1637.4	17.3	7
Bergen County, NJ	932202	3884.5	17.7	10
Hudson County, NJ	672391	13731.4	12.2	3
Essex County, NJ	798975	6211.5	13.9	4
Union County, NJ	556341	5216.1	10.8	3
Passaic County, NJ	501826	2715.3	15	6
Suffolk County, MA	803907	12415.7	12.3	24
Middlesex County, MA	1611699	1837.9	15.7	19
Essex County, MA	789034	1508.8	17.6	2
Norfolk County, MA	706775	1693.6	17.2	11
Worcester County, MA	830622	528.6	13.2	9
Wayne County, MI	1749343	2974.4	14.2	35
Oakland County, MI	1257584	1385.7	17.3	17
Macomb County, MI	873972	1754.9	17.4	6
Genesee County, MI	405813	668.5	18	7
Washtenaw County, MI	367601	488.4	14.5	7
Philadelphia County, PA	1584064	11379.5	14	35
Montgomery County, PA	830915	1655.9	18.2	32
Delaware County, PA	566747	3040.5	16.8	15

Lehigh County, PA	369318	1012.5	17.1	10
Berks County, PA	421164	480.4	17.6	6
Los Angeles County, CA	10039107	2419.6	14.1	129
San Diego County, CA	3338330	735.8	14.5	38
Riverside County, CA	2470546	3745	14.8	24
Santa Clara County, CA	1927852	1381	13.9	17
Orange County, CA	3175692	3807.7	15.3	40

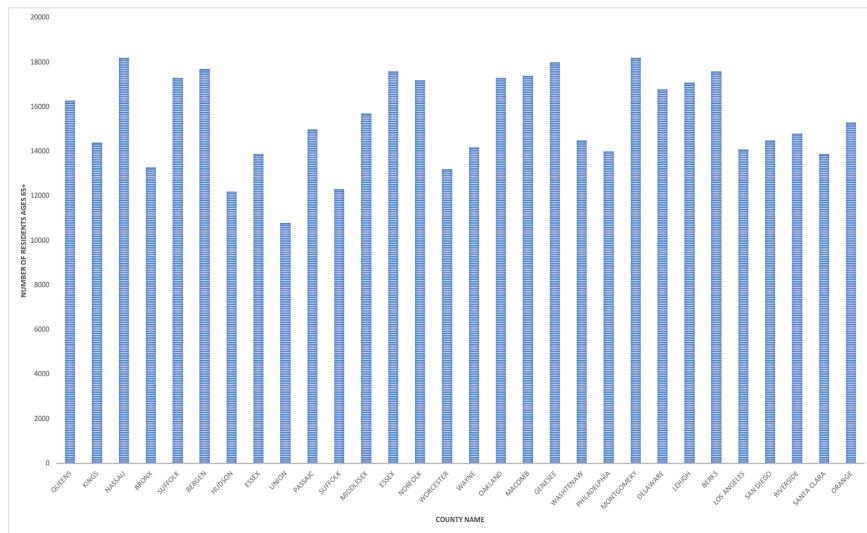
\*data as of July, 2019 \*\*data as of June, 2020 \*\*\*data as of 2010

As illustrated by Figure 1 below, the top nine counties with the most hospitals are Los Angeles County, Orange County, San Diego County, Philadelphia County, Wayne County, Montgomery County, Suffolk County (MA), Riverside County, and Middlesex County (MA).



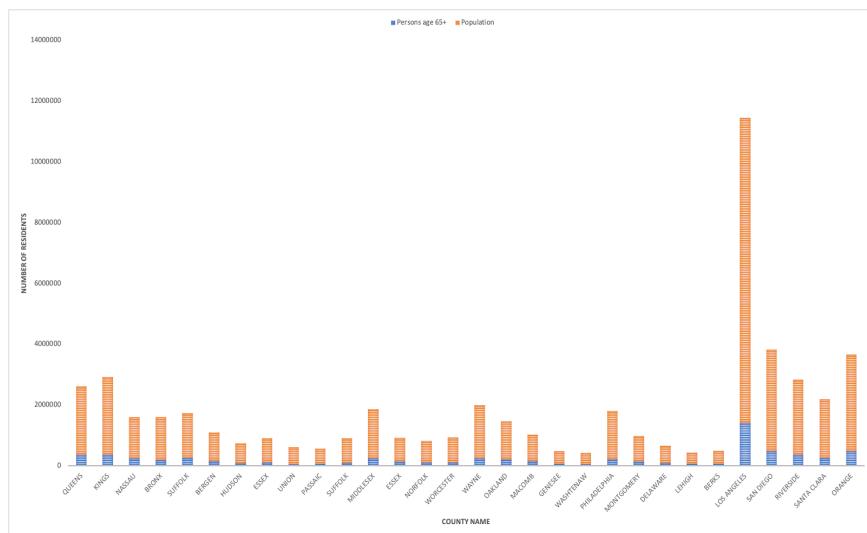
**Figure 1.** Number of Hospitals per County

A county may have more COVID-19 cases or deaths because of its high elderly demographic (people ages 65 years and older). As Figure 2 shows, The top ten counties with the highest percentage of ages 65 and older are Montgomery County, Nassau County, Genesee County, Bergen County, Berks County, Essex County (MA), Macomb County, Oakland County, Suffolk County (NY), and Norfolk County.



**Figure 2.** People Ages 65 and Older per 100.00 Residents per County

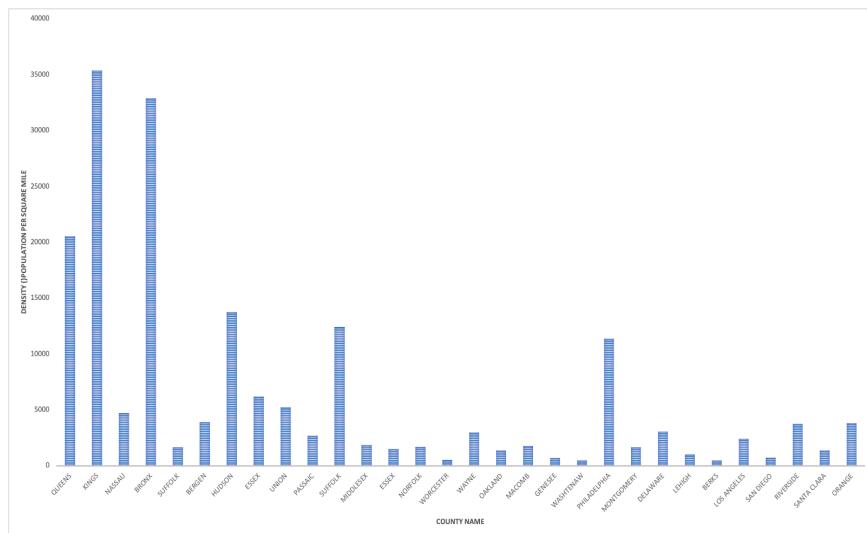
The figure 3 below is a model illustrating the elderly demographic in comparison to the county's total population.



**Figure 3.** Elderly Demographic Among the Entire County's Population

The Figure 4 below is illustrating the density per county, which is in population per square mile. We suspect that a county's COVID-19 cases and deaths may be related to its density, since we predict that COVID-19 is more easily transmitted in a dense area, where residents may not be able to social distance as properly as residents of a less dense county.

Additionally, we've noticed that Los Angeles, the largest city in Los Angeles County, has a population of 3,979,576 (39.64% of the entire county), but has a density of 8092.3 (334.45% greater than its entire county). Thus, later in Section 6 we will further discuss this finding and see if a highly dense city has a direct relationship with COVID-19 cases and deaths.



**Figure 4.** Population Density per County

## 5. COVID-19 Cases and Deaths Statistics

This section contains data on each county's COVID-19 cases and deaths, including their total cases, total deaths, daily cases, daily deaths, and each state's total and daily tests conducted [10 – 17]. These data have been individually graphed, and each county's data have been summed and graphed to compare between states.

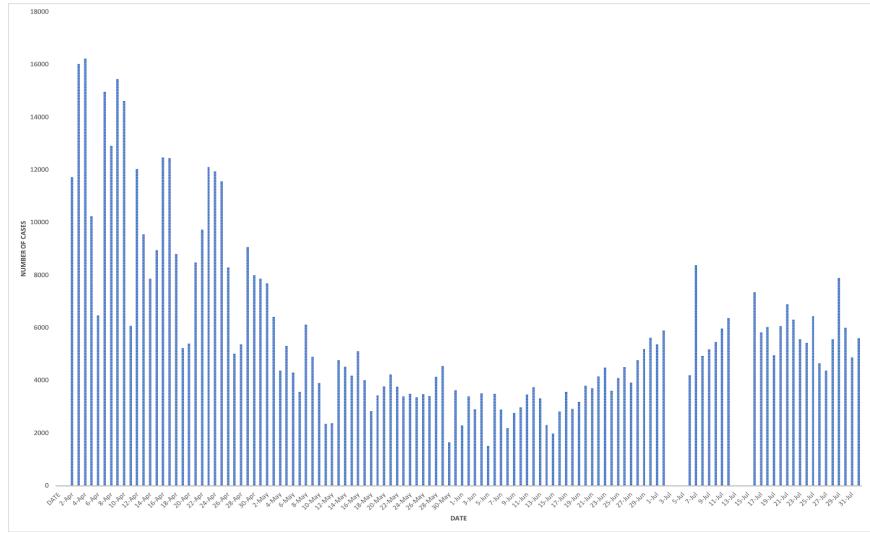
County	Total Cases*	Total Deaths*
Queens County, NY	68041	4059

Kings County, NY	62494	4551
Nassau County, NY	43271	2706
Bronx County, NY	49626	3075
Suffolk County, NY	43300	2044
Bergen County, NJ	20494	1785
Hudson County, NJ	19497	1332
Essex County, NJ	19505	1859
Union County, NJ	16545	1176
Passaic County, NJ	17465	1091
Suffolk County, MA	21279	1057
Middlesex County, MA	25801	1983
Essex County, MA	17305	1182
Norfolk County, MA	10305	986
Worcester County, MA	13376	991
Wayne County, MI	25632	2678
Oakland County, MI	11470	1084
Macomb County, MI	9225	899
Genesee County, MI	2796	269
Washtenaw County, MI	2084	112
Philadelphia County, PA	25831	1690

Montgomery County, PA	9813	850
Delaware County, PA	8770	730
Lehigh County, PA	4828	335
Berks County, PA	5168	365
Los Angeles County, CA	191023	4669
San Diego County, CA	29579	565
Riverside County, CA	38216	695
Santa Clara County, CA	10281	191
Orange County, CA	36703	649

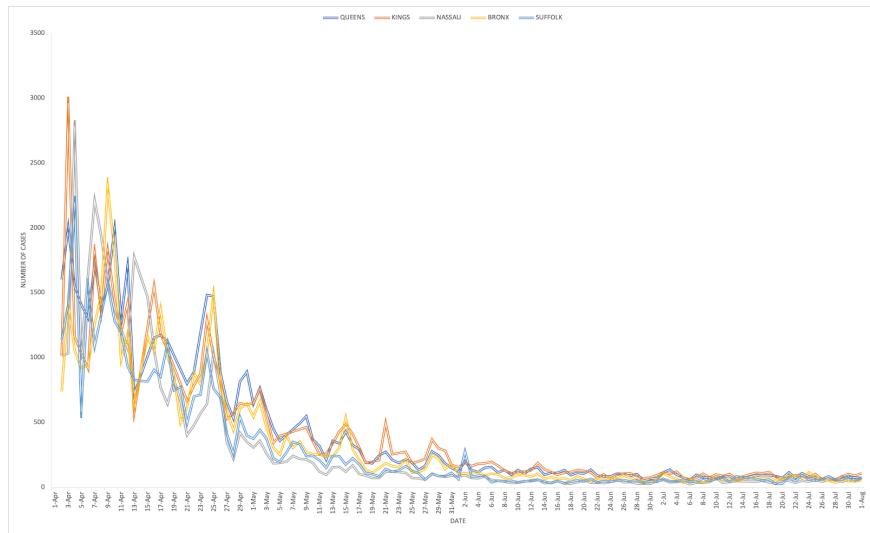
\*data from April 1, 2020 – August 1, 2020

Figure 5 illustrates the total daily growth in COVID-19 cases among our data set (the top 5 most infected counties per top 6 most infected states in the United States). Cases steadily decreased from early April to late May, slightly plateaued in early June, rose until mid July, and started decreasing again.



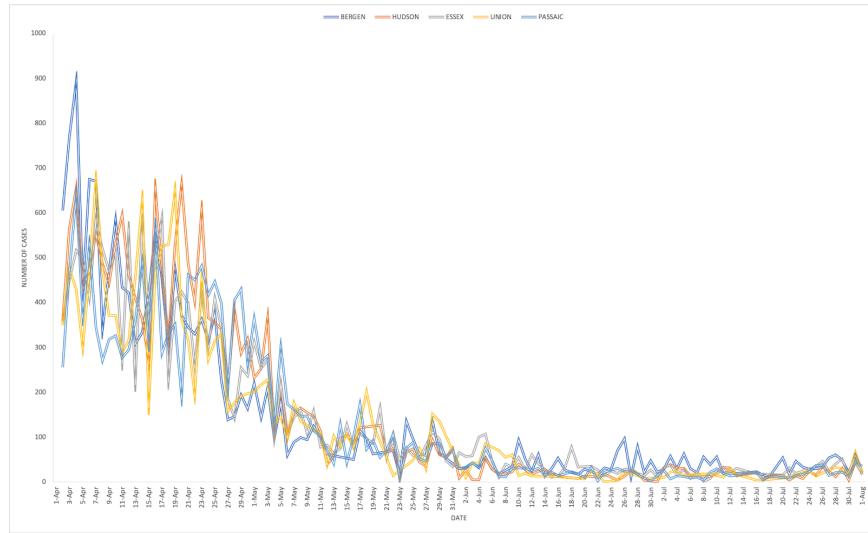
**Figure 5.** Total Daily COVID-19 Cases Among the Top 5 Most Infected Counties of the Top 6 Most Infected States in the United States

Figure 6 graphs the daily COVID-19 cases per top five infected counties of New York. Though New York has been the epicenter of COVID-19 initially, since April 1, 2020, the start of this study, New York's counties' cases have been steadily decreasing, and seems to plateau near the beginning of June, 2020.



**Figure 6.** New COVID-19 Cases in the Top Five Most Infected Counties of New York

Figure 7 graphs the daily COVID-19 cases per top five infected counties of New Jersey. New Jersey's counties' cases have been steadily decreasing since mid April, 2020 (slightly later than New York), and seems to plateau near mid June, 2020.



**Figure 7.** New COVID-19 Cases in the Top Five Most Infected Counties of New Jersey

Figure 8 graphs the daily COVID-19 cases per top five infected counties of Massachusetts. Unlike New York and New Jersey, Massachusetts's counties' cases have been rising until mid April, and then finally started decreasing in late April, followed by plateauing around early June. However, since early July, the cases seem to be slowly increasing again.

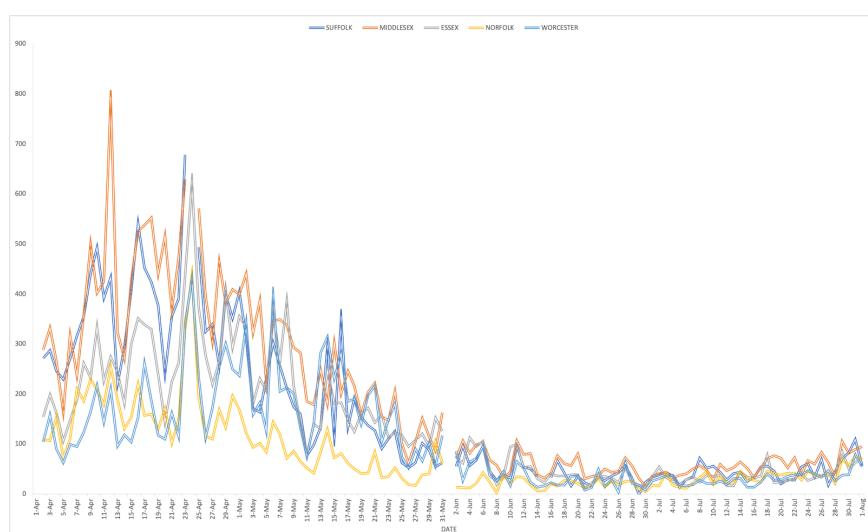
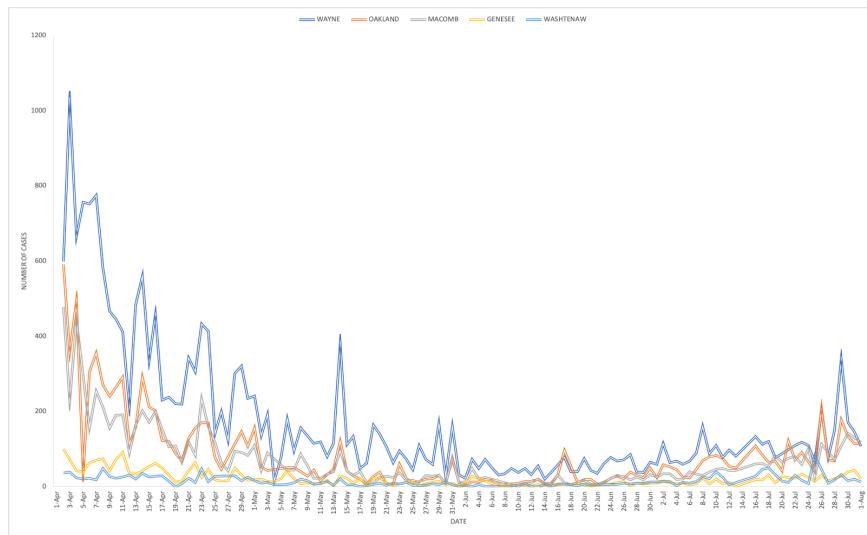
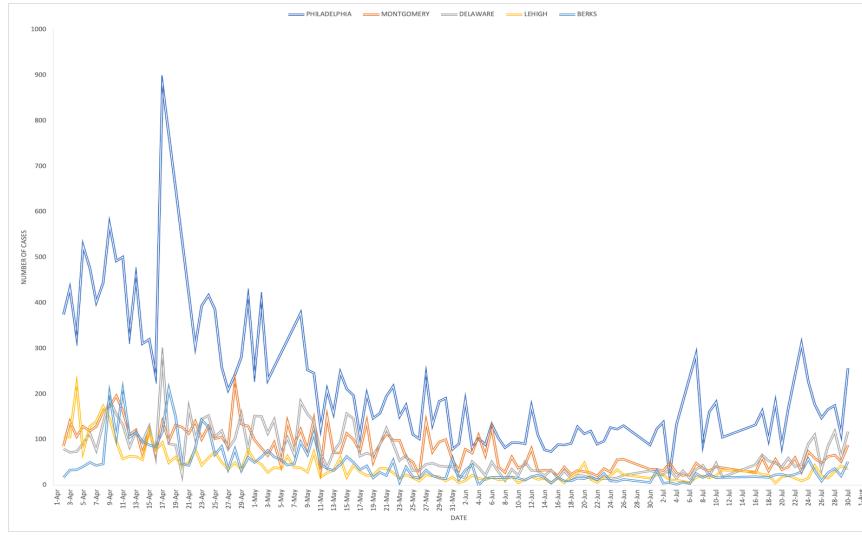


Figure 9 graphs the daily COVID-19 cases per top five infected counties of Michigan. As observed from the figure, Genesee and Washtenaw have remained steady in the increase of cases since the beginning of this study, with a nearly horizontal trend. However, the other three counties' cases steadily decreased from early April to late May, plateaued from early June to early July, and then started increasing again since mid July.



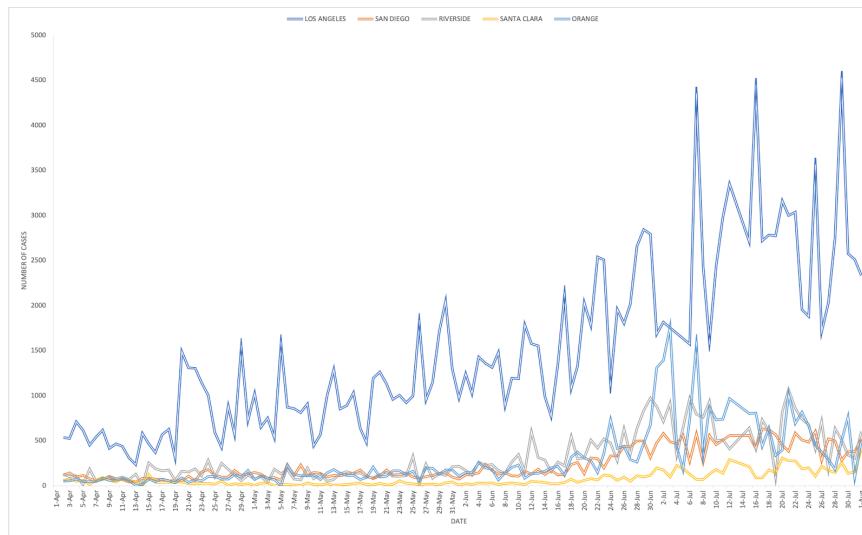
**Figure 9.** New COVID-19 Cases in the Top Five Most Infected Counties of Michigan

Figure 10 graphs the daily COVID-19 cases per top five infected counties of Pennsylvania. Besides Philadelphia County, Pennsylvania's counties' cases have remained constant with a nearly horizontal trend. However, Philadelphia, the most infected county, decreased from early April to late May, plateaued from early June to early July, and seems to be growing again. Its cases have never reached as low as the other counties, too.



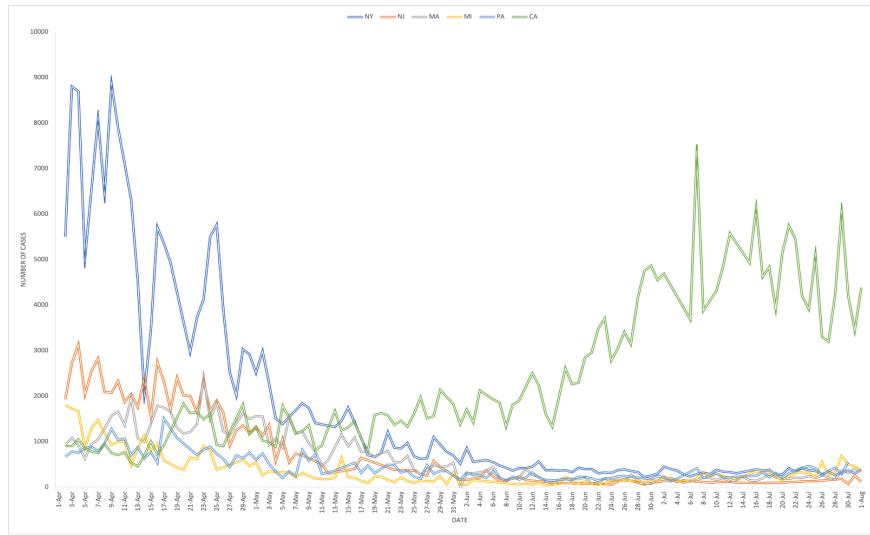
**Figure 10.** New COVID-19 Cases in the Top Five Most Infected Counties of Pennsylvania

Figure 11 graphs the daily COVID-19 cases per top five infected counties of California. Besides Los Angeles County, California's counties' cases have remained fairly flat from early April to mid June, starting rising until late July, and seem to be decreasing now. Los Angeles County, which is California's most infected county, has been steadily increasing since early April, and has not shown clear signs of plateauing, making it the most infected county in the United States, as of August 1, 2020.



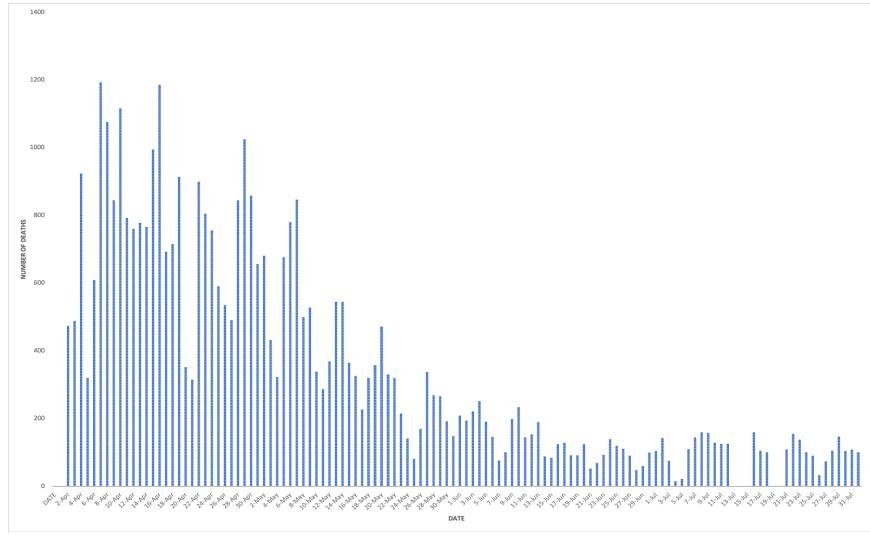
**Figure 11.** New COVID-19 Cases in the Top Five Most Infected Counties of California

Figure 12 is a visual to compare the sum of the states' counties' daily COVID-19 cases. As shown below, New York, New Jersey, Massachusetts, Michigan, and Pennsylvania have all decreased drastically, and started plateauing around early June, while California has been rapidly growing, but started showing decreasing near mid July.



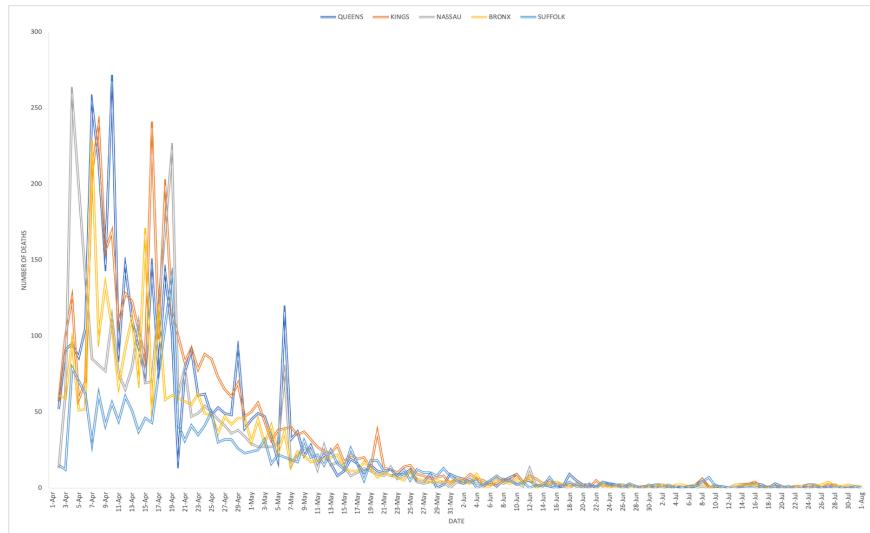
**Figure 12.** Daily Cases of COVID-19 Among the Top Five Most Infected Counties of the Top Six Most Infected States in the United States

Figure 13 illustrates the total daily growth in COVID-19 deaths among our data set (the top 5 most infected counties per top 6 most infected states in the United States). The death have been steadily decreasing since the start of our study, and plateaued around mid June.



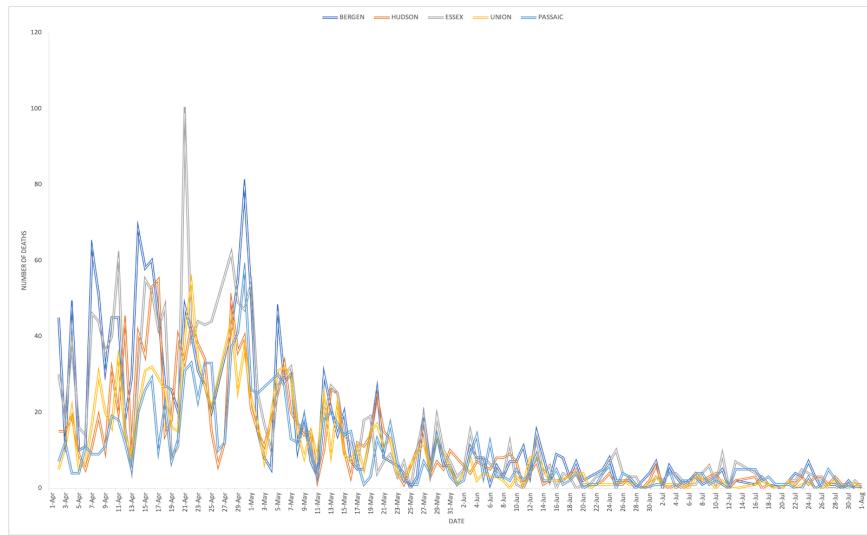
**Figure 13.** Total Daily COVID-19 Deaths Among the Top 5 Most Infected Counties of the Top 6 Most Infected States in the United States

Figure 14 graphs the daily COVID-19 deaths per top five infected counties of New York. The county rose until mid April, started decreasing, and eventually plateaued around early June (around the same time its daily cases plateaued).



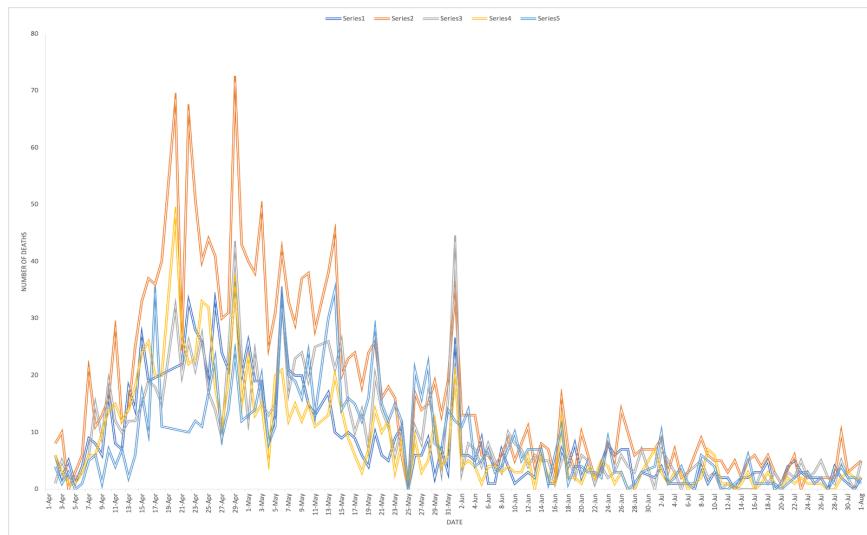
**Figure 14.** New COVID-19 Related Deaths in the Top Five Most Infected Counties of New York

Figure 15 graphs the daily COVID-19 deaths per top five infected counties of New Jersey. The county rose until mid April, started decreasing, and finally plateaued around mid June.



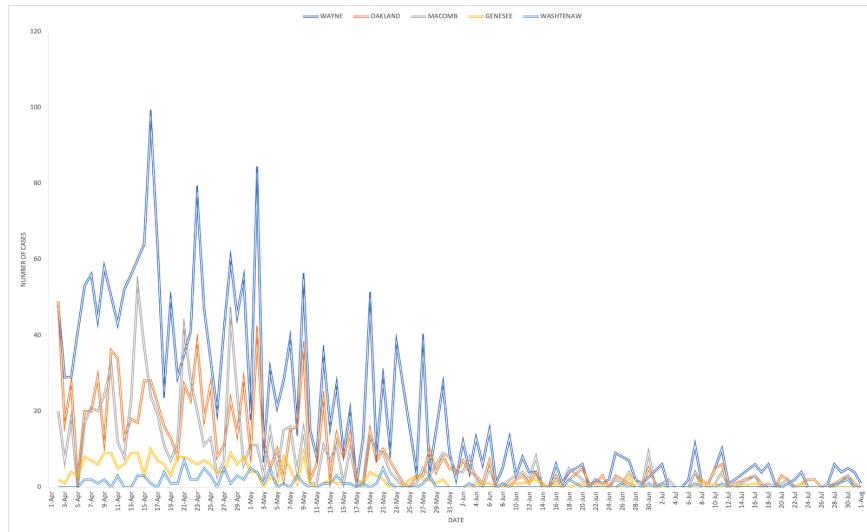
**Figure 15.** New COVID-19 Related Deaths in the Top Five Most Infected Counties of New Jersey

Figure 16 graphs the daily COVID-19 deaths per top five infected counties of Massachusetts. The county rose until late April, started decreasing, and plateaued around mid June.



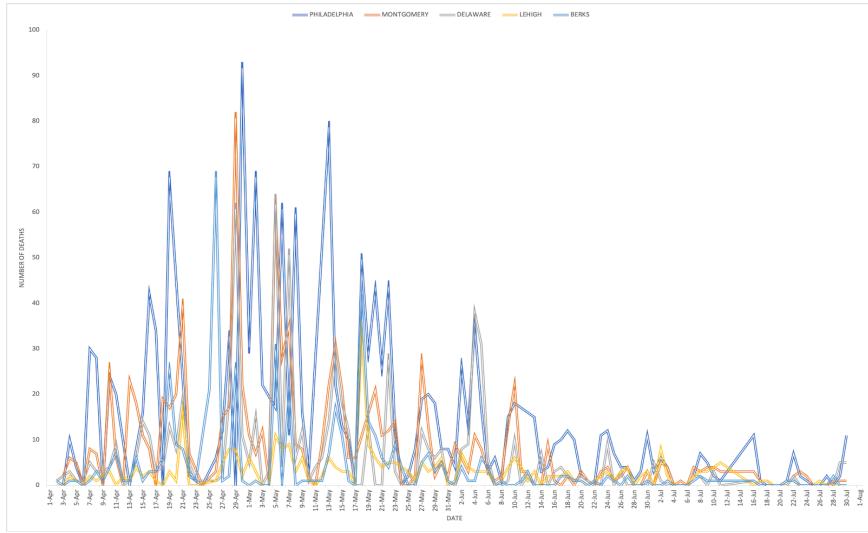
**Figure 16.** New COVID-19 Related Deaths in the Top Five Most Infected Counties of Massachusetts

Figure 17 graphs the daily COVID-19 deaths per top five infected counties of Michigan. The count decreased until late May and plateaued around early June. Throughout this study, though Michigan's COVID-19 case count hasn't been as high as those of other states, Michigan's death county, particularly Wayne County, has been extremely high.



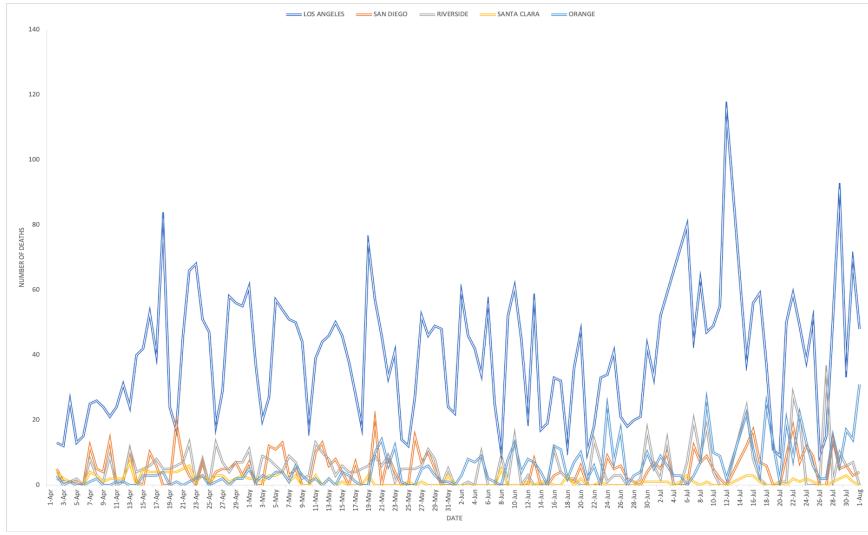
**Figure 17.** New COVID-19 Related Deaths in the Top Five Most Infected Counties of Michigan

Figure 18 graphs the daily COVID-19 deaths per top five infected counties of Pennsylvania. The county increased until early May, starting decreasing, and eventually plateaued at around mid June.



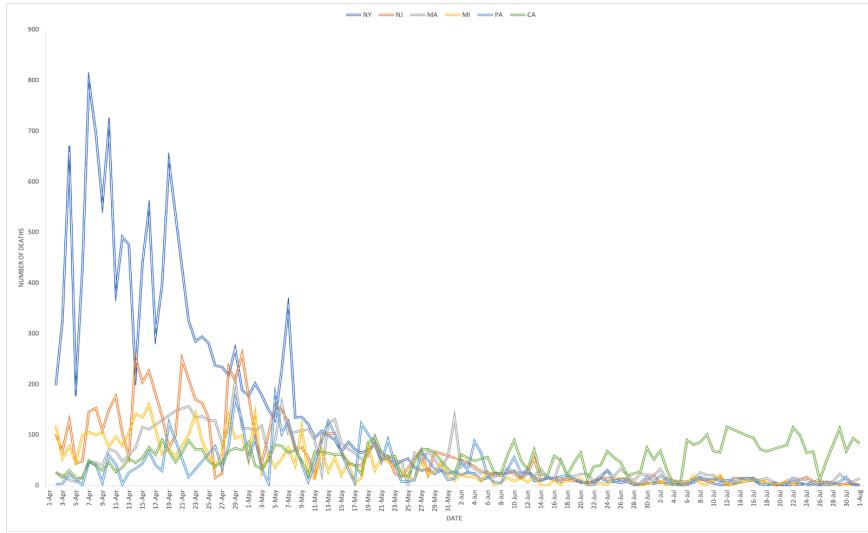
**Figure 18.** New COVID-19 Related Deaths in the Top Five Most Infected Counties of Pennsylvania

Figure 19 graphs the daily COVID-19 deaths per top five infected counties of California. The count trend was fairly horizontal in the sense that the amount increased one day was equally decreased the next day, leading to a fairly straight line, just with oscillations. Los Angeles County's daily deaths decreased from early to late June, but started rising again from early July. Additionally, though all counties have not significantly fluctuated up or down, there is still a large gap between Los Angeles County's numbers and those of the rest of the concerned counties in California.



**Figure 19.** New COVID-19 Related Deaths in the Top Five Most Infected Counties of California

Figure 20 is a visual to compare the sum of the states' counties' daily COVID-19 deaths. Interestingly enough, this figure shows something that could not be visualized from the previous five graphs (figures 14, 15, 16, 17, and 18). Though New York's counties' daily death count was steadily decreasing since early April, the other four states' counties' death count started rising until early May, and then reached the same numbers as New York and plateaued at around early June. Also, though California's counties' death count plateaued, as of early July it has started increasing again.



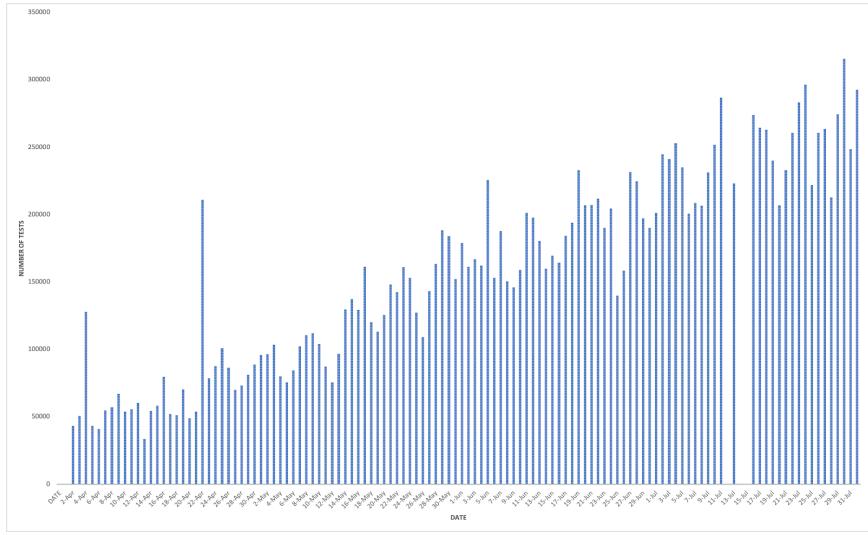
**Figure 20.** Daily COVID-19 Related Deaths Among the Top Five Most Infected Counties of the Top Six Most Infected States in the United States

## 6. COVID-19 Tests Conducted Statistics

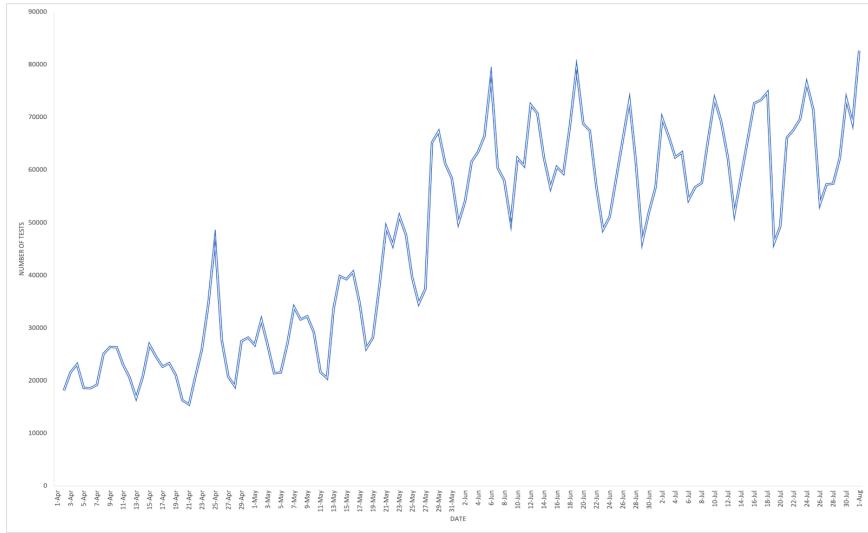
This section contains data on the daily and total tests conducted per state, from April 1, 2020 – August 1, 2020 [10–17]. We believe that an increase in testing can help a state's COVID-19 cases go down because more testing means a higher chance of discovering people who may have been infected with the virus. Thus, they can isolate themselves earlier and prevent the spread to other people.

State	Total Tests Conducted*
New York	5971974
New Jersey	2096329
Massachusetts	1193200
Michigan	1876830
Pennsylvania	1232322
California	8035975

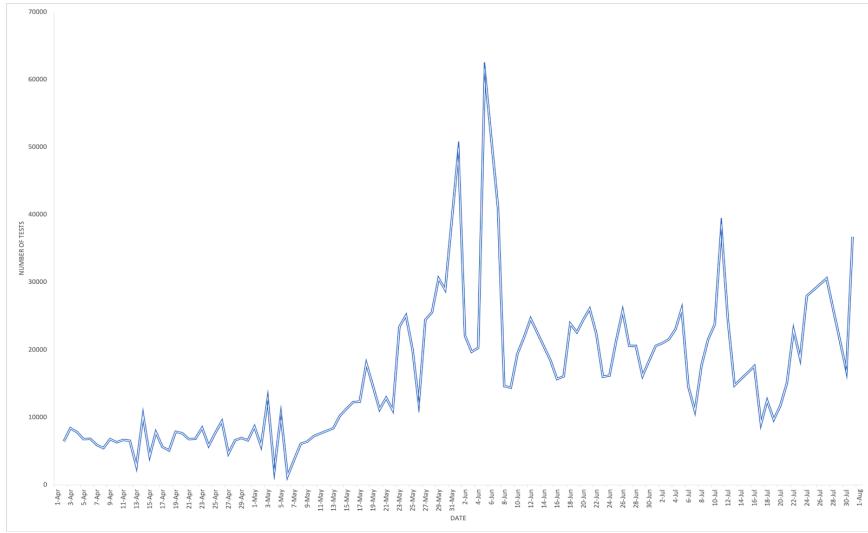
\*data from April 1, 2020 – August 1, 2020



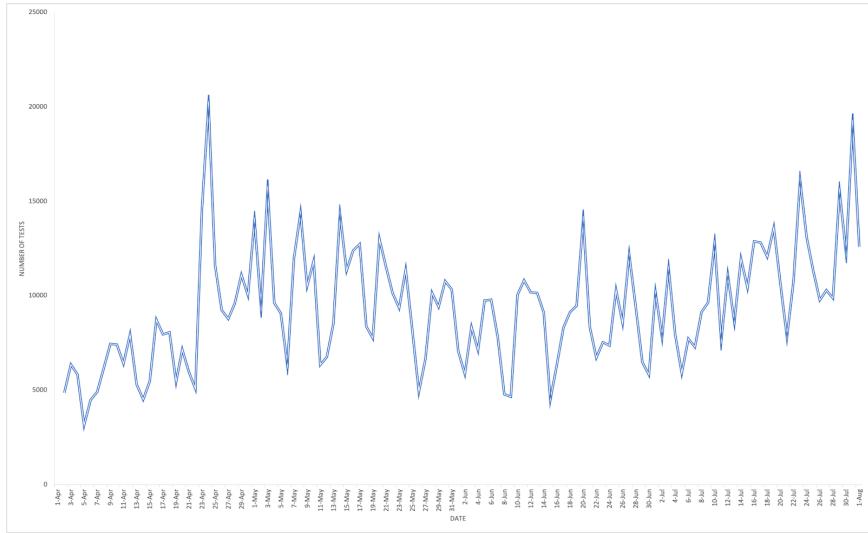
**Figure 21.** Total COVID-19 Tests Conducted by the Top Six Most Infected States



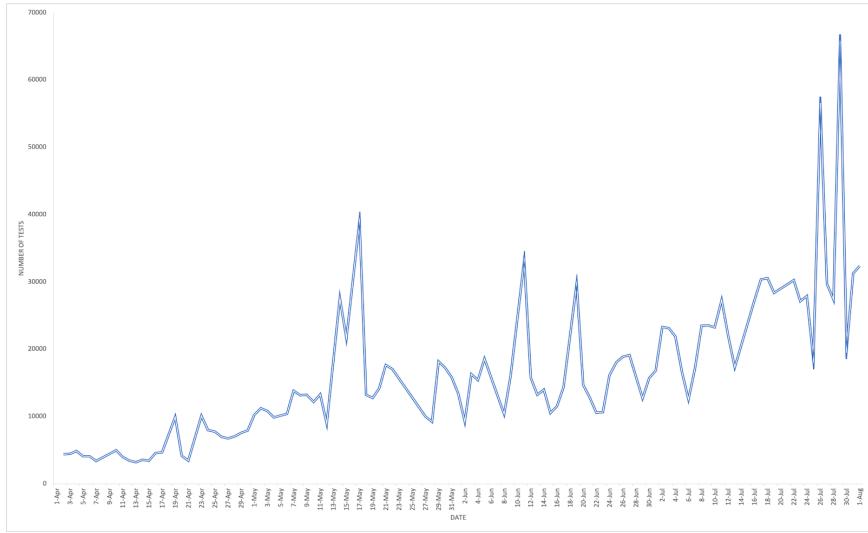
**Figure 22.** New COVID-19 Tests Conducted by New York



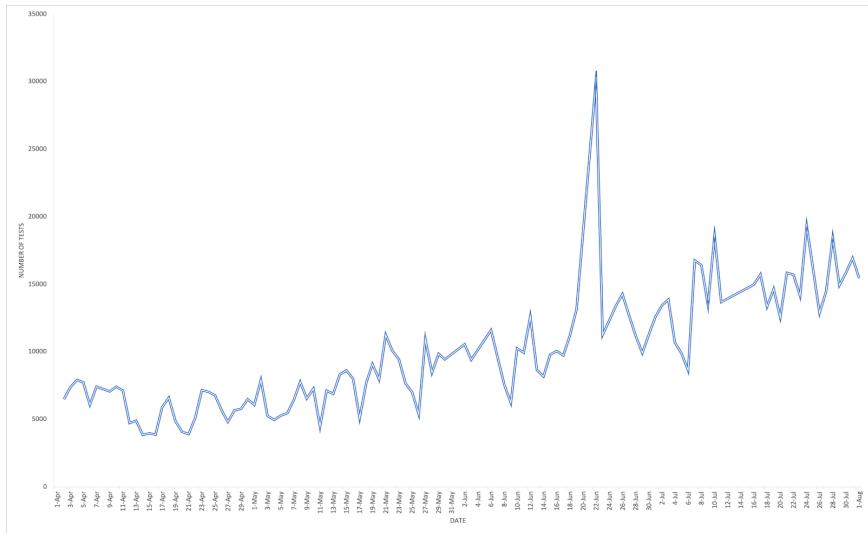
**Figure 23.** New Tests Conducted by New Jersey



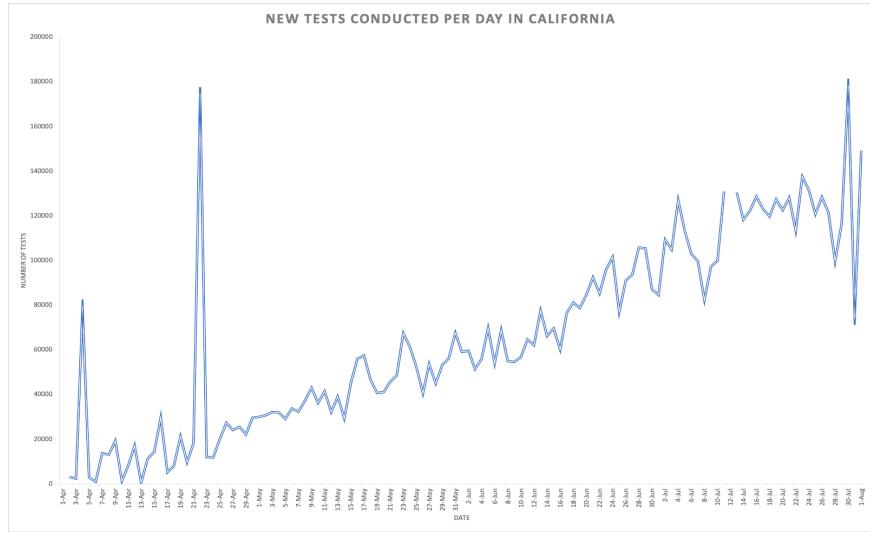
**Figure 24.** New Tests Conducted by Massachusetts



**Figure 25.** New Tests conducted by Michigan



**Figure 26.** New Test conducted by Pennsylvania



**Figure 27.** New Test conducted by California

## 7. Correlation between COVID-19 Cases and Population Density

Section 7 is where our study dives deeper into the relationship between a county's density and its total cases. The chart below lists the 30 counties concerning our study listed in order of increasing density. Our decision to compare the possible effects of a county's density on it's COVID-19 cases comes from several factors.

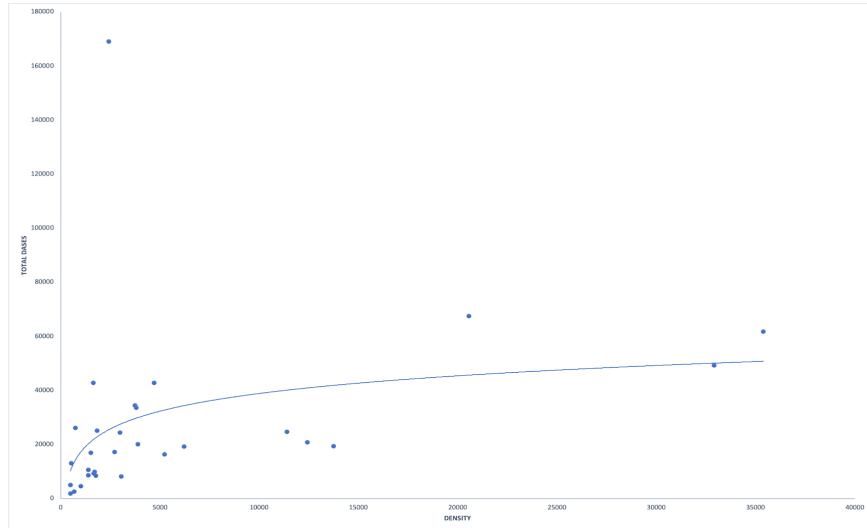
Firstly, we've noticed that the number of hospitals per county does not seem to have a strong impact on a county's cases. When the number of hospitals per person (hospitals per county population number) is graphed alongside its county's case number (figure not shown in the paper), there is no strong association between the coordinates (the trend is practically horizontal). There was also a low correlation between a county's percentage of elders and its COVID-19 cases (figure not shown in the paper). We suspect that is the case because the mean and variance of the percentages is 15 and 4 respectively, showing that the values are very close to one another.

Finally, we looked at our last factor (besides density), the number of tests conducted. Since the start of our study, four months since the start of our study (April 1), as shown by Figure 21, there has only been an increase in tests conducted in the United States. New York, New Jersey, Massachusetts, Michigan, Pennsylvania, and California all had increases in testing (new regulations and goals were set regarding upping the tests). Thus, we have seen decreases in each state's COVID-19 cases, except one county: Los Angeles County. Even though California is conducting thousands of tests per day, much more than the other states, unlike other counties, Los Angeles County has not shown obvious signs of decreasing. Thus, this is the reason we decided to focus this section of the study towards the relationship between a county's density and its total cases, among other factors.

County	Density*	Total Cases**
Berks County, PA	480.4	5168
Washtenaw County, MI	488.4	2084
Worcester County, MA	528.6	13376
Genesee County, MI	668.5	2796
San Diego County, CA	735.8	29579
Lehigh County, PA	1012.5	4828
Santa Clara County, CA	1381	10281
Oakland County, MI	1385.7	11470
Essex County, MA	1508.8	17305
Suffolk County, NY	1637.4	43300
Montgomery County, PA	1655.9	9813
Norfolk County, MA	1693.6	10305
Macomb County, MI	1754.9	9225
Middlesex County, MA	1837.9	25801
Los Angeles County, CA	2419.6	191023
Passaic County, NJ	2715.3	17465
Wayne County, MI	2974.4	25632
Delaware County, PA	3040.5	8770
Riverside County, CA	3745	38216
Orange County, CA	3807.7	36703
Bergen County, NJ	3884.5	20494
Nassau County, NY	4704.8	43271
Union County, NJ	5216.1	16545
Essex County, NJ	6211.5	19505
Philadelphia County, PA	11379.5	25831
Suffolk County, MA	12415.7	21279
Hudson County, NJ	13731.4	19497
Queens County, NY	20553.6	68041

Bronx County, NY	32903.6	49626
Kings County, NY	35369.1	62494

\*data as of 2010 \*\*data from April 1, 2020 – August 1, 2020



**Figure 28.** Relationship Between Counties' Density and Their Total Cases



**Figure 29.** Map of LA County [18]

Upon analyzing Figure 28, we've noticed 2 details. Firstly, there seems to be a relationship between a county's density and its COVID-19 cases, as the data points fall nicely on a logarithmic trend line (this trend supports our idea that COVID-19 is possibly more easily

transmitted in a dense area, since residents may not be able to social distance as properly as residents of a less dense area). Second, we observed an extreme outlier, the Los Angeles County data point. With a density of 2,419.6 residents per square mile (the fifteenth least dense out of the 30 studied counties), but a total case count of 191,023 (the most in the United States as of August 1, 2020), we found this very strange. We've come to the suspicion that Los Angeles City, the largest city in Los Angeles County and California, may have something to do with this situation.

Los Angeles County has a population of 10,039,107 and a density of 2419.6 residents per square mile. Los Angeles City (Los Angeles County and California's largest city) has a population of 3,979,576 (39.64% of its county) and a density of 8,092.3 residents per square mile (234.45% greater than its county). We then chose Wayne County, Michigan, to compare with these statistics. Wayne County has a population of 1,749,343, a density of 2974.4 residents per square mile, and a predictable case count of 25,632 (according to Figure 28). Detroit City (Wayne County and Michigan's largest city) has a population of 670,031 (38.3% of its county) and a density of 5,144.3 residents per square mile (72.95% greater than its county). Wayne County is 22.93% more dense than Los Angeles County, but its largest city, Detroit City, is 36.43% less dense than Los Angeles County's largest city, Los Angeles City. Thus, we believe that the counties' most populated city has an impact on its state's overall COVID-19 cases. Although Wayne County is more dense than Los Angeles County, it may be implied that Wayne County's most populated city, Detroit, is equally more dense than Los Angeles County's most populated city, Los Angeles. However, this is not the case, so we think that among multiple other factors that could have affected each county's case count, this difference is a factor of what makes Los Angeles County have soaring COVID-19 cases. We want to acknowledge the fact that this theory is not completely applicable towards all of the counties in our study. For example, Bronx County and King County have the highest population density, with 32,903.6 and 35,369.1 residents per square mile respectively. However, their case counts are 49,626 and 62,494 respectively (74.02% and 67.28% lower than Los Angeles County's case count respectively). Density, obviously, is not the only factor that affects a county's COVID-19 cases. However, since Los Angeles County's data point is such an extreme outlier, we believe that its most populated city's density can address why it's an outlier among the data points near its own (smaller counties and smaller cities), rather than in the grand scheme of things, with larger counties like Queens County, Bronx County, and Kings County.

## 8. Conclusion and Future Work

We are inspired to analyze COVID-19 among counties because we've noticed a lack of data in that aspect. We saw easily accessible COVID-19 data for the US as a country, and for its states, but the county based data study was scattered and not easily discoverable. Thus, we wanted to focus on our study to specifically analyze inter-county data, and chose the top 5 most infected countries of the top 6 most infected states to work with.

As of April 1, 2020 (the start of our study), we have been tracking each state and their counties' total COVID-19 cases, related deaths, and tests conducted (value is for the entire state). We have been graphing and publishing these data on GitHub:

<https://github.com/coronavirus-2020-data/coronavirus-data>. Alongside COVID-19 data, we've collected data on each county's population, density, percentage of seniors (elder demographic), and number of hospitals; these all were factors we wanted to analyze to see if they had any affect on a county's COVID-19 data. Through comparison, we've deemed the elder demographic and number of hospital values as not an important influence on a county's cases, since upon graphing those variables with each county's total cases as the dependent variable, there has not been a strong trend and relationship (horizontal line of best fit). However, after graphing an xy scatter plot on a county's density vs. its total COVID-19 cases, we've noticed that the data points fell on a logarithmic line of best fit. We believe this supports the idea that density affects a county's COVID-19 cases; a possible explanation for this is that because an area has a high population density, residents might not be able to social distance as properly as those from a less dense area, thus COVID-19 could be more easily transmitted in the denser location. We've also noticed that Los Angeles County's data point was an extreme outlier, having the 15 less dense area of the study but the highest COVID-19 case count in the entire nation. We found out that although Los Angeles County has a smaller density than that of Wayne County, a data point near Los Angeles County's, Los Angeles County's most populous city Los Angeles has a much higher density than Wayne County's most populous city Detroit. We believe this shows that in addition to a county's density, a county's largest city's density is also an important factor on the county's total cases. Of course, we've acknowledged that this finding only applied to the left edged cluster of data points. The most dense counties, like Bronx County and Kings County, did not have as much cases as Los Angeles County.

Although the growth of COVID-19 cases in the US has slowed down considerably, it is for certain that society and our practices (like social distancing and wearing masks in public) won't revert back to normal for a long time, if it ever will. In fact, scientists have raised the possibility that COVID-19 will become like the influenza, a seasonal disease [19].

Additionally, Bill Gates predicted that the rich world will get a vaccine by the end of 2021 [20]. Given this, we aim to continually collect COVID-19 data regarding the 6 states and their 5 counties in this study, and publish the numbers on GitHub. Overall, conducting this four month study and analysis has been an interesting, rewarding experience to keep up with the latest COVID-19 statistics and hopefully contribute towards the community.

## References

1. <https://www.statnews.com/2020/01/30/who-declares-coronavirus-outbreak-a-global-health-emergency/>
2. <https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.25678>
3. <https://www.livescience.com/first-case-coronavirus-found.html>
4. <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>
5. <https://www.ncsl.org/ncsl-in-dc/publications-and-resources/president-trump-declares-state-of-emergency-for-covid-19.aspx>
6. <https://www.sciencealert.com/starting-isolation-two-weeks-sooner-could-have-saved-9-out-of-10-us-coronavirus-deaths>
7. <https://www.kff.org/other/slide/when-state-stay-at-home-orders-due-to-coronavirus-went-into-effect/>
8. <https://catalog.data.gov/dataset/hospitals-dcdfc>
9. <https://www.census.gov/quickfacts/fact/table/US/PST045219>
10. <https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Map?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n>
11. <https://covid19.nj.gov/>
12. <https://www.mass.gov/info-details/covid-19-updates-and-information>
13. <https://www.michigan.gov/coronavirus/>
14. <https://covidtracking.com/data>
15. <https://www.health.pa.gov/topics/disease/coronavirus/Pages/Coronavirus.aspx>
16. <https://www.nytimes.com/interactive/2020/us/pennsylvania-coronavirus-cases.html>
17. <https://covid19.ca.gov/>
18. [https://en.wikipedia.org/wiki/Los\\_Angeles\\_County,\\_California#/map/0](https://en.wikipedia.org/wiki/Los_Angeles_County,_California#/map/0)
19. <https://www.covid-19facts.com/?p=84419>
20. <https://www.cnet.com/how-to/when-will-covid-19-end-the-situation-now-as-the-coronavirus-rages-on/>