Miles S. Marimbire

Master of Science in Economics University of Cincinnati

Last Modification: 4/23/2022

Project: Covid County EDA Report

"Covid 19 began on December 12, 2019, when a cluster of patients in Wuhan, Hubei Providence, China began to experience shortness of breath and fever." (Cited by the CDC) This was only the beginning of when the global pandemic would begin. The Coronavirus spreaded quick and swiftly and within a span of three months forced the world into a global shutdown as well as a global pandemic, The world officials enforced masked mandates and quarantines to make sure the Coronavirus did not spread. After two years following the outbreak, we have been able to collect vast amount of data to analyze the consequences. We have also been able to collect data and find the effects of the Coronavirus on numerous aspects of our life. In this project we will be looking at how different counties in different states were impacted by the Coronavirus. Using data from the CDC and Economic tracker as well as other reputable data sources we will also be providing visualizations and regression analysis to verify our research.

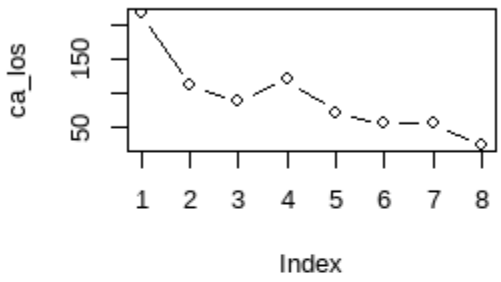
We always wanted to know the economic affects that Covid has had on counties in California, Texas, and New York. These states were chosen as they have the highest population among all states. In this paper we will focus around how covid has affected the amount of covid cases and population in the counties we selected and comparing the covid cases between different counties and the differences in hospitalization of covid patients. We will select a total of 3 counties across 3 states. We believe that these questions are important because we want to find the main contributors to what causes covid hospitalization, we want to figure out if population causes more covid cases per 100k and many more. We will be using regression analysis and different visualization techniques.

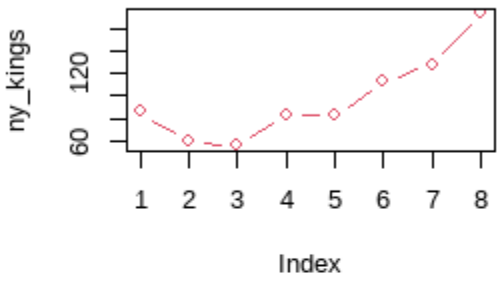
The data we used was collected from (https://data.cdc.gov/Public-Health-Surveillance/United-States-COVID-19-Community-Levels-by-County/3nnm-4jni). Initially were

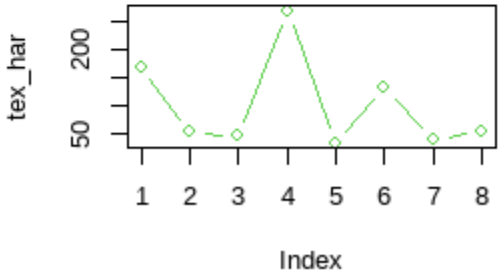
going to resort to web scraping as a form of data collection, but we were unable to find reputable websites that allowed for web scrapping. We collected the data, and the original size was 25,792 observations with 12 variables that consists of the county name, county flips, state, county population, health service area number, health service area, health service area population, covid inpatient bed utilization, covid hospital admissions per 100k, covid cases per 100k, covid 19 community level, and date updated. We removed all the data we no longer needed from the data set and kept the necessary ones for the counties we aimed to look at.

To begin, the first question we wanted to verify was how big of an affect population size had on the amount of covid cases per 100k. The method we used was to look for the states with the highest population and created visualizations to view the outcome and looking at the visualizations we can concluded that population in fact has no effect on the number of covid cases per 100k. We noticed that for a county such as Los Angeles County, California (Population 10,040,682) has fewer cases on average than a county such as Harrison County, Texas (Population 4,680,609) and Kings County, New York (Population 2,576,771); Apart from the visualizations to support our claims when comparing the means of Covid cases per 100k Los

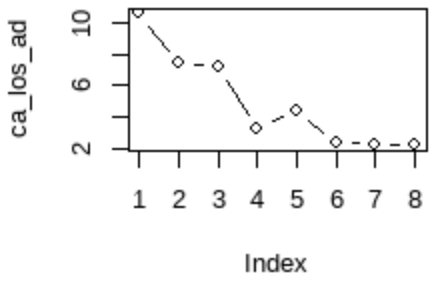
Angeles had a mean of 93.05 whilst Harrison had a mean of 99.90 and Kings 97.7.

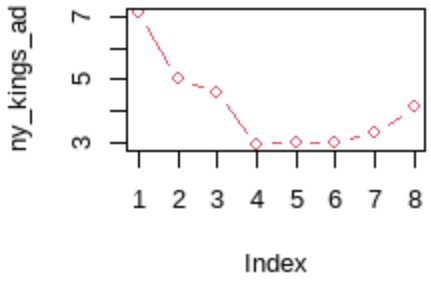


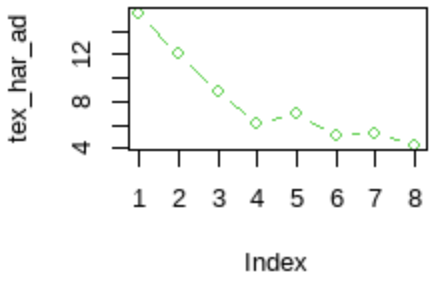




Keep in mind that the variables, “ca_los” is Los Angles, “tex_har” is Harrison, and “ny_kings” is Kings. As you can see from the plots that kings have the lowest population and has the highest covid cases per 100k and rising. Now that we have established that population has little to no effect on the amount of covid cases per 100k in our data without controls. We want to see if population has an effect on covid hospital admissions per 100k, what appears to happen is that for the hospital admissions it appears population has no effect on hospital admissions as well as New York has the highest admissions but the lowest population. Please refer to graph below.







New York also has the highest cases per 100k. After reviewing articles and looking over spatial data from https://www1.nyc.gov/site/planning/data-maps/open-data.page the reason they are the highest is because the Virus spreads much faster due to the close proximity of all of the housing units. After finding out that population has no effect on the amount of covid cases per 100k as well as hospital admissions per 100k we looked to see if hospital admissions and covid cases have any effect on each other. We correlated the results of the counties covid cases and hospitals admissions; for all counties and noticed a strong correlation between the Los Angeles County with a correlation of 0.84, but for the other counties it represented weak correlations with Harrison having 0.12 and Kings having a negative weak correlation relationship -0.24. This leads us to believe that neither population influences these counties.

Our next step was to create a panel model, we created a OLS model for all three variables to try and see what are the largest contributors as seen below ***

Covid Cases per 100k OLS Results

	Dependent variable:		
	Compensation Factors		
	(1)	(2)	(3)
covid_inpatient_bed_utilization	43.744	-75.866	-4.795
	(28.286)	(21.221)	(47.297)
	p = 0.183	p = 0.016	p = 0.924
covid_hospital_admissions_per_100k	-13.599	59.834	10.171
	(19.533)	(19.586)	(21.528)
	p = 0.518	p = 0.029	p = 0.657
Constant	6.336	52.879	58.551
	(21.952)	(33.438)	(32.342)
	p = 0.785	p = 0.175	p = 0.131
Observations	8	8	8
R2	0.807	0.736	0.553

```
F Statistic                          10.440**   6.980**     3.088
=====================================================================
Note:                                      *p<0.1; **p<0.05; ***p<0.01
                                          Standard errors in parentheses
```

For our Los Angeles results, covid inpatient bed utilization is insignificant on covid cases per 100k while for every covid hospital admissions per 100k added the covid cases per 100k will decrease by -14 percentage points and this is significant at the 0.1 level. For the Kings model we see that for every covid inpatient bed utilization covid cases 100k will decrease by -76 percentage points and this is significant at the 0.5 level and for covid hospital admissions per 100k with each unit change covid cases per 100k will increase by 60 percentage points and this is significant at the 0.5 level. For Texas both independent variables have very small effects and covid inpatient bed utilization is insignificant at all levels whilst for each unit change in covid hospital admissions per 100k covid cases per 100k will increase by 10 percentage points and is insignificant at all levels.

In essence, the density of the population tended to show a correlation according to the data models and the data that we ran through. We have seen countless and countless examples of how close proximities over the past two years of covid have led to more cases and in turn more hospitalizations. In our data, Los Angeles County showed a lower average of covid cases per 100k because homes are more spread out than the New York and Texas counties we investigated. Los Angeles had the strongest correlation between hospital visits and covid cases meaning there was a stronger chance that a person going to the hospital in California was going for covid complications than in Kings County, New York or in Harrison County, Texas. Our explanatory data analysis report concluded that population has no effect on the number of covid cases without other factors being controlled and no effect on the number of

hospitalizations as well. In conclusion of all our data and our variables, we can see how Covid-19 has been puzzling to many researchers and experts. In situations you think there would be correlation there isn't and vice versa. Covid-19 over the last two years has put us all through tough times and we hope that cases and hospitalizations continue to trend down

References

https://www.cdc.gov/museum/timeline/covid19.html#:~:text=December%2012%2C%202019%20A,of%20breath%20and%20fever.

https://browse.welch.jhmi.edu/datasets/Covid19

https://data.cdc.gov/browse?tags=covid-19&page=4

https://www1.nyc.gov/site/planning/data-maps/open-data.page

https://www.nytimes.com/2016/03/01/nyregion/overcrowding-worsens-in-new-york-as-working-families-double-up.html

https://comptroller.nyc.gov/wp-content/uploads/documents/Hidden_Households.pdf