

Andrew Zhou

Professor McDonald

Human Centered Data Science

14 December 2021

Dallas, Texas – COVID-19, Mask mandates, Unemployment, and Uninsured health insurance rates

1. Introduction

For my analysis, I will be looking at COVID-19 cases, mask mandates, unemployment rate, and uninsured health insurance rates for Dallas, Texas. Dallas is known to be a republican state and relatively against health insurance. It is common to find Dallas among the highest uninsured health insurance rates in the United States, and it has held the highest uninsured health insurance rates quite consistently over the past few years. “Hospital leaders have said that these uninsured patients put off preventable care until something is wrong and end up in the emergency room with costly uncompensated care for hospitals to deal with.” This is a serious issue on itself, however, with COVID-19 the issue may be much more severe.

It is quite common for people to get their health insurance from their employment. So, it should not be without a doubt that health insurance rates and employment rate are highly correlated values. One of the most popular topics during the period of COVID-19 was unemployment. As COVID-19 pandemic became a serious problem in the United States, the unemployment rate shot up, which likely made a lot of people lose their health insurance, and in turn these people did not seek their needed medical attention. So for my analysis, I would like to understand how COVID-19, mask mandates, unemployment rate, and uninsured health insurance rates impacted each other and ultimately how did these impact death rate?

2. Background

Before starting my analysis, I had a hypothesis of how COVID-19, mask mandates, unemployment rate, and uninsured health insurance rates impacted each other and how they impacted death rates. My hypothesis was that mask mandates and COVID-19 cases are negatively correlated. That is, when a mask mandate law is put into place, the new COVID-19 cases are likely to drop. The reason is that masks prevent water droplets that may contain the virus from spreading in to the air and prevents people from catching the virus. My hypothesis also guesses that unemployment rate and uninsured health insurance rates are positively correlated with COVID-19 cases, and that when COVID-19 cases rise, then unemployment rate and uninsured health rate will likely rise as well. I suspected this because as COVID-19 cases rise, then more businesses are likely to close down and there are less jobs open for people. Unemployment rate and uninsured health insurance rates are probably positively correlated. A lot of people have their health insurance from their employment. Also, for people who do not have insurance from their employment, they are likely to buy insurance out of pocket, but if they lose their job, they are also likely to cancel their insurance. With an even higher uninsured health insurance rate, I suspect that many people are likely to refuse seeking medical attention and increase the death rate in Dallas.

In summary, my hypothesis for my analysis is:

- COVID-19 daily cases is positively correlated with unemployment.
- Mask mandates will slow down the spread of COVID-19.
- As COVID-19 cases increases, unemployment increases.
- As unemployment increases, uninsured health insurance rates increase.
- As uninsured health insurance rates increase, more people are likely to not to get their needed medical attention.

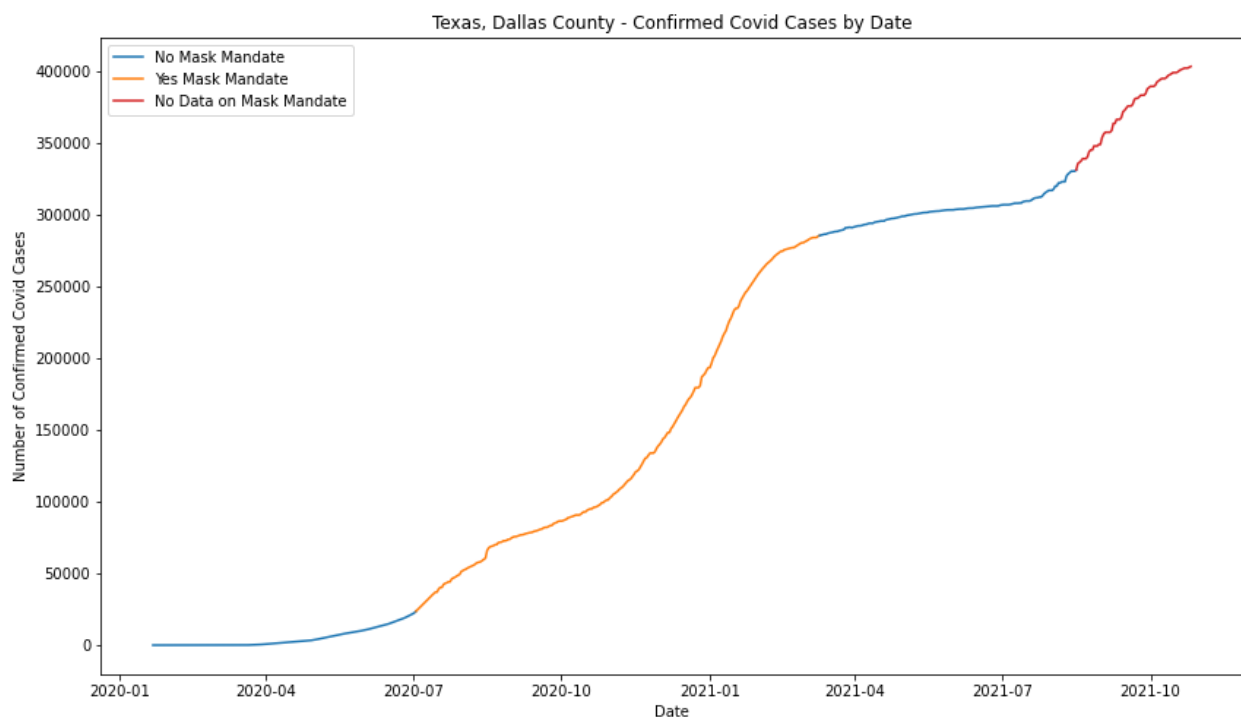
3. Methodology

For my analysis, the main methodology methods will be using data visualization and linear regression. As any data science project, I started off with exploratory data analysis, which mainly consisted of data visualization. As all of my data is time series data, I will be making many time series plots with unemployment rate and uninsured health insurance rates to see if there are any correlations and interesting findings. I will start off by plotting COVID-19 case count and daily cases along with mask mandates in a time series to get a better understanding of COVID-19 and mask mandates data in Dallas, Texas. Since unemployment and uninsured health insurance data is annual and monthly, I plan on pulling data before the pandemic and compare how the pandemic impacted the values in comparison. I will plot unemployment rate and uninsured health insurance rate in a time series on its own to see where the rate usually hovers before the pandemic hits. I will also plot unemployment rates and uninsured rates that are close to the pandemic, so I can include the new COVID-19 cases and mask mandate data into the visualizations. To not confuse the reader of the plots, I decided to make two plots for unemployment rate and uninsured health insurance rates for better visibility and clarity. However, these plots are all just for exploratory data analysis and visualizations, they are not good tools for determining correlation between variables.

For my second part of analysis, I will be using statistical methods such as regression and ANOVA to determine correlation between variables and coefficients to determine how a variable is correlated with another. Regressions are very nice for when the dependent and independent variables are continuous. Regression will be useful for determining the correlation between unemployment rate and new COVID-19 cases. ANOVAs are very nice for when dependent variable is continuous but the independent variable is categorical. ANOVA is good to determine

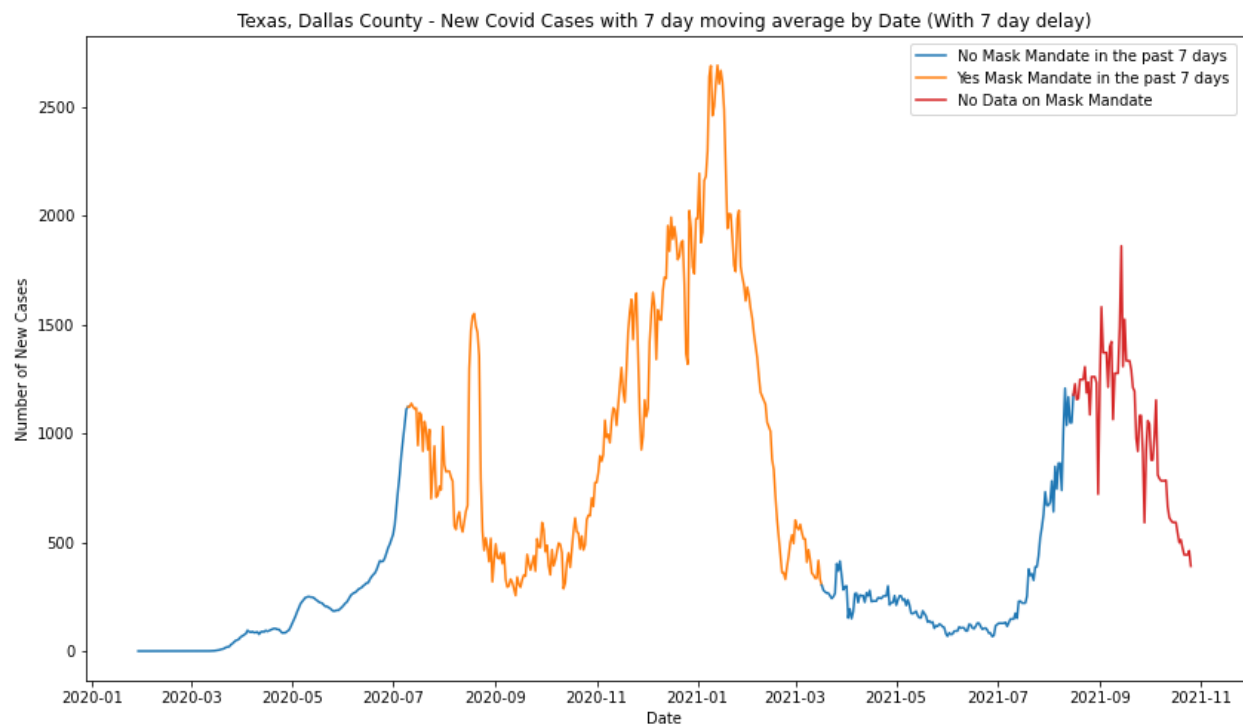
the correlation between mask mandates and the new COVID-19 cases. Due to very few data points overlapping for uninsured health insurance rates and COVID-19, correlations will have to be inferred from data visualizations and results from the other tests. I will also be comparing how Dallas, Texas compares in death rate compared to other counties in the United States. To ensure a fair comparison with other counties, I will be using death rate normalized over the population and multiplying by 100,000 for interpretability. The metric will turn out to be something like 100 deaths / 100,000 people.

4. Findings



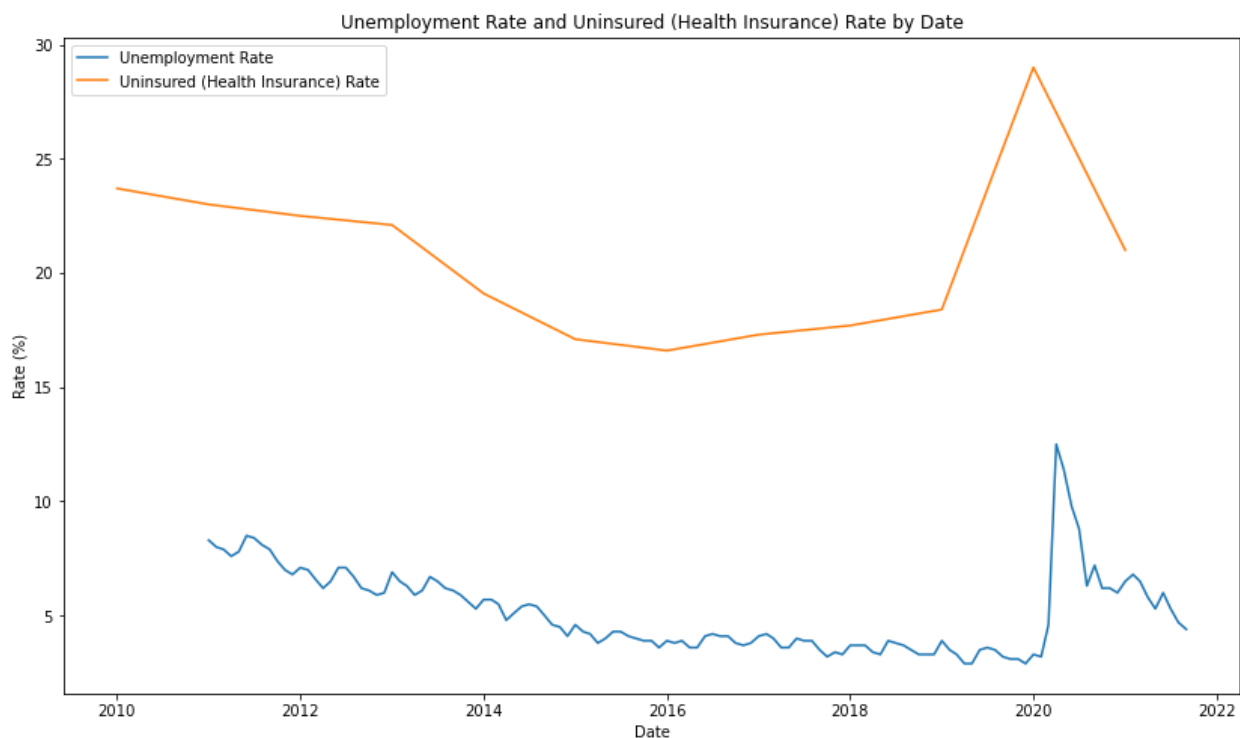
This plot displays the total confirmed cases per day in Dallas, Texas. The plot has three different sections: "No Mask Mandate", "Yes Mask Mandate", and "No Data on Mask Mandate." The "No Mask Mandate" means there was no mask mandate during that day, and The "Yes Mask Mandate" means there was mask mandate during that day. The "No Data on Mask Mandate" means there was no data during that time period, and I am unsure if there is a mask mandate or

not. This plot shows us that the mask mandates seem to be during the worst times of covid and there are no mask mandates during relatively low new cases. Since this plot shows the total confirmed cases, the plot is a little hard to read. I decided to plot new confirmed cases against date instead.



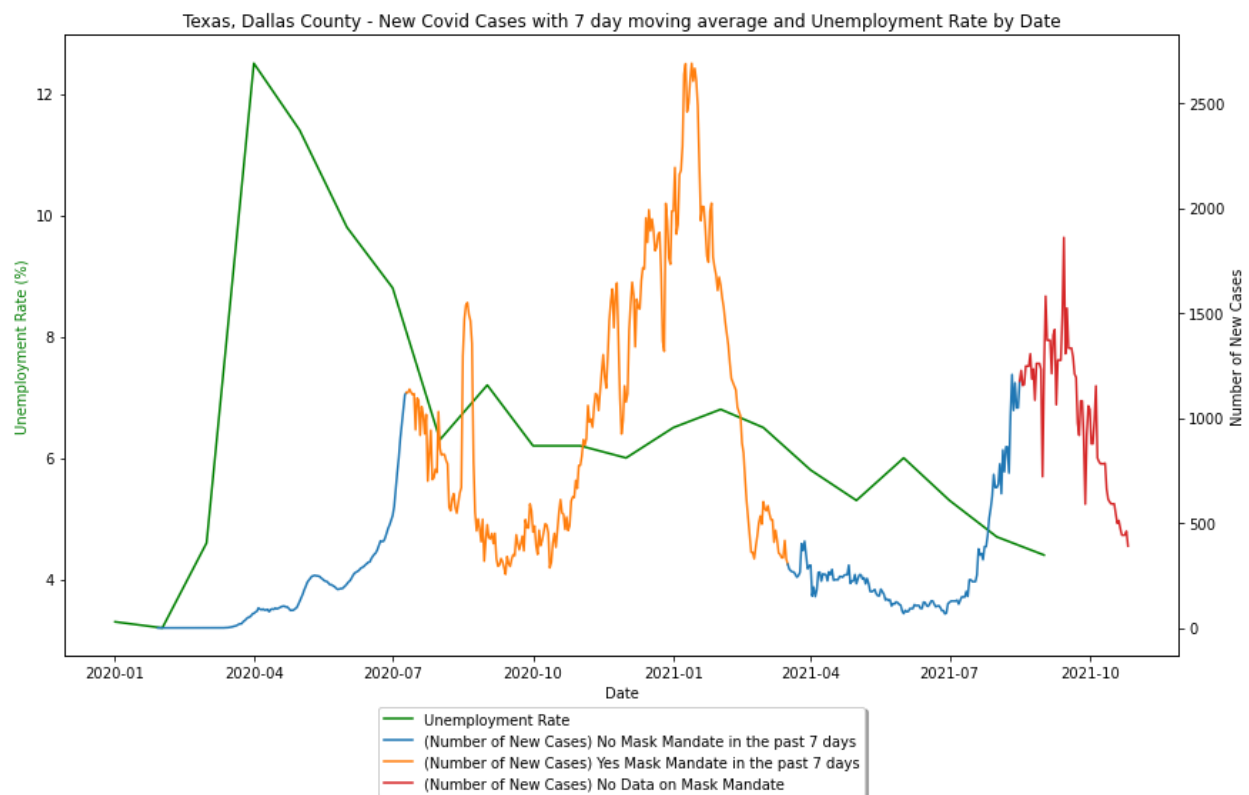
This plot displays the new confirmed cases per day in Dallas, Texas. The plot uses the same section segmentation schema as the previous plot. The only difference is that each section includes a 7-day lag, which means the sections are shifted to the right by 7 days. It is also worthy to note that this plot is using the 7-day moving average for new COVID-19 cases. There were a lot of days with 0 new cases and by investigating further, every week there is a two-day break in COVID-19 reporting. That break is typically Saturday and Sunday, but it can also be found on Monday and Friday. The day following the two-day break will have a very large value which consists of the previous two days new COVID-19 cases and to counter that the 7-day moving average was used instead. This smoothened out the plot a lot more.

From the plot, it looks like the mask mandate laws were put in place during the worst parts of Covid (high number of daily new cases). Perhaps, it would be worse without the mask mandate. We do see a spike in covid cases right before the mask mandate though, and that is probably why the mask mandate was put in place. It does look like soon after the mask mandate, there is a slight drop in covid cases before it spikes up really high around December 2020 and January 2021. However, the spike drops very soon. I imagine that drop is due to the vaccinations given to the population around that time period. I also choose not to comment on the dates after August 2021, as the mask mandates data set does not reach this far, and I'm not sure if there is a mask mandate there or not. I can only comment on the huge spike that has appeared around August 2021 time period.

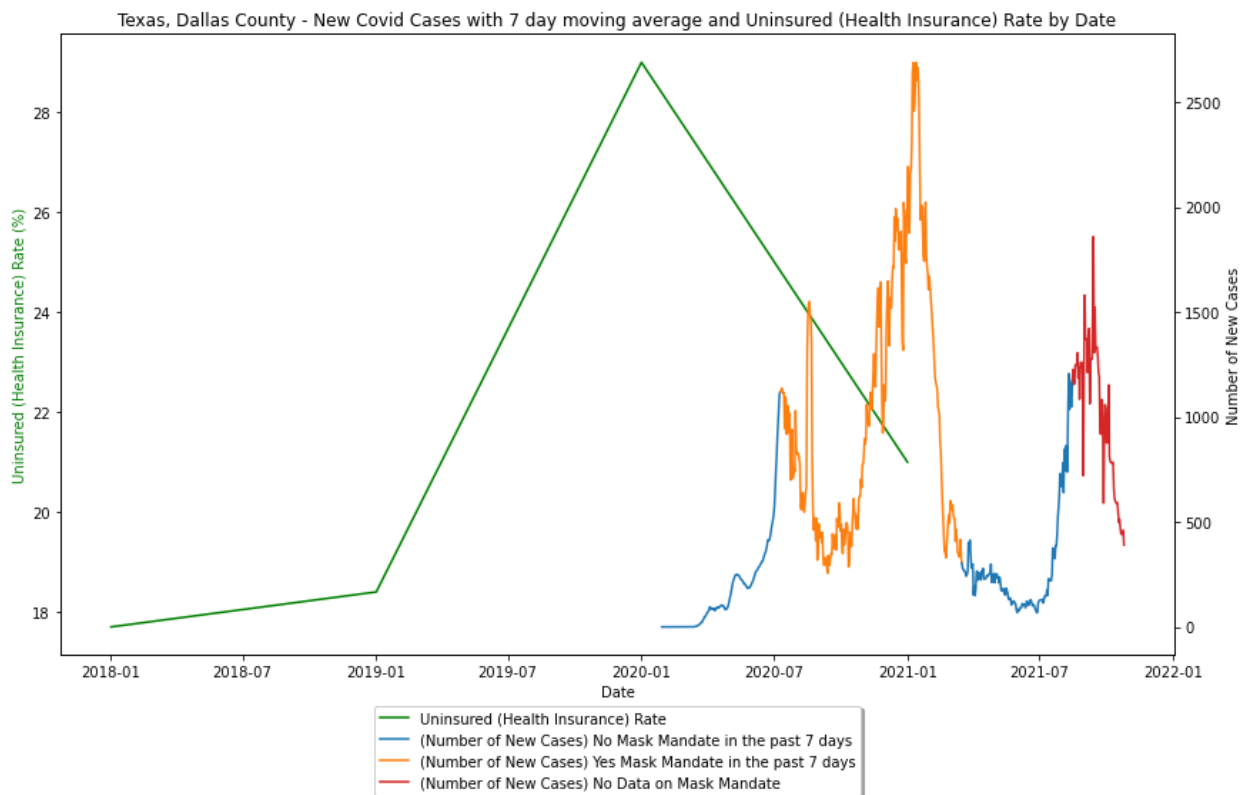


Now I decided to look into my unemployment and uninsured health insurance rates data. I start off by plotting only unemployment and uninsured health insurance rates into one plot that

range all the way back to year 2010. As you can see the unemployment in the plot is monthly and the uninsured health insurance rates is annually. There are very few data points that overlap over the pandemic, so it's worth looking at the history of the two rates and how they changed once the pandemic started. The first thing we can note is that both rates are steadily decreasing over time right before the pandemic. At the pandemic, we can see a huge spike in both rates. Dallas had a relatively low rate of unemployment, and uninsured health insurance rates, although were decreasing before the pandemic, had a relatively high rate even before the pandemic. Dallas has a history of having one of the highest uninsured health insurance rates in the United States and has held the highest uninsured health insurance rate a couple times over the years. The pandemic spike only makes the uninsured health insurance rate a lot worse, spiking up to near 30% at the start.



This is a plot of unemployment rate plotted over the pandemic. The COVID-19 cases plot is from the original plot earlier with unemployment rate added into the graph. Please note that these plots use a different scale: unemployment rate with the left y-axis and number of new cases on the right y-axis. Looking at this plot, we can see a huge spike in unemployment rate at the start of the pandemic, even before COVID-19 cases spike. After the initial spike, we can see that the unemployment slowly dips back down over time. It is actually not easy to tell if there is any correlation between the two, and this may need further studying to look at more closely.



This is a plot of uninsured health insurance rate plotted over the pandemic. Again, the COVID-19 cases plot is from the original plot earlier with uninsured health insurance rate added into the graph. Also note the axes of the plots. Looking at this plot, we can also see a huge spike in uninsured health insurance rate at the start of the pandemic, even before COVID-19 cases spike, then it is followed by an immediate dip afterward. Unfortunately, there are very few data points overlapping for uninsured health insurance rates, so very little about the correlation can be

inferred from a plot and statistical testing method. Instead, we can only infer uninsured health insurance correlation with COVID-19 cases from other variables such as unemployment rate.

Residuals:

Min	1Q	Median	3Q	Max
-5.5081	-0.7392	0.3505	1.6163	3.8356

Coefficients:

	Estimate	Std. Error	t value	p value
_intercept	7.035601	0.726765	9.6807	0.000000
New_Cases	-0.000632	0.000637	-0.9928	0.333299

R-squared: 0.02285, Adjusted R-squared: -0.03144

F-statistic: 0.42 on 1 features

The above is the output for a regression with unemployment rate as the dependent variable and new COVID-19 cases as the independent variable. The p-value for the new cases is 0.33, which is pretty high. It comes nowhere to 0.05, and we fail to reject the null hypothesis. In other words, I can safely conclude that there is not enough evidence for a correlation between COVID-19 cases and unemployment rate. This was actually shocking for me, as I hypothesized that higher COVID-19 cases would result in higher unemployment rate, which is actually not true.

Residuals:

Min	1Q	Median	3Q	Max
-5024.62	-124.62	198.3588	358.3588	1092.38

Coefficients:

	Estimate	Std. Error	t value	p value
_intercept	358.358779	35.349940	10.1375	0.0
Mask Mandate	691.021221	60.006186	11.5158	0.0

R-squared: 0.16181, Adjusted R-squared: 0.16051

F-statistic: 123.75 on 1 features

The above is the output for a regression for mask mandates and COVID-19. Since mask mandates is a categorical variable and COVID-19 cases is a continuous variable, this is actually equivalent to an ANOVA and two sample t-test for difference of means. The reason the regression is more useful is because it can tell me coefficient, so I can tell how the mask

mandates impacted COVID-19 cases if there is evidence of correlation. The p-value for the mask mandate is 0.0, which is very low. This means the results are significant and that there is strong evidence of correlation between mask mandate and COVID-19 cases. However, the next thing I notice immediately is the coefficient, which is 691. The value is a positive value, which indicates that with mask mandates, the increase of COVID-19 cases is higher than no mask mandates. This was also extremely shocking, but it also makes sense that mask mandates were put in place in times of high COVID-19 cases. It also makes more sense that COVID-19 mask mandates were loosened up when COVID-19 cases were not very severe. From this result, I can conclude that mask mandates did not help unemployment and uninsured health insurance rates in a way that is quantifiable or useful.

Other than data visualizations and statistical methods, I also ranked the death rate of Dallas, Texas in comparison with other counties in the United States. Dallas has a total of 5228 which is ranked 13th highest county in the United States. However, considering Dallas is a relatively populated city, I felt this was not a fair comparison. Therefore, death rate is normalized by population for a fairer comparison. Getting the rank for death rate normalized over the population yields a death rate of 200.6 deaths / 100,000 people, which is ranked 2004 out of the 3227 counties that were analyzed. This death rate by Dallas is actually not bad at all.

5. Discussions and Implications

Through my visualizations and statistical methods, I have found many interesting things. The first thing to note is that days with mask mandates had significantly higher COVID-19 cases than days without mask mandates. Through this result, I would like to infer that mask mandates did not really help out that much for unemployment and uninsured health insurance rates. Especially, since my other regression test found that there was not enough evidence to support a

correlation between unemployment rate and COVID-19 cases. Judging by the similarity of the plots and correlation between unemployment and uninsured health insurance rates, I believe it is also safe to say that COVID-19 cases are not correlated with uninsured health insurance rates either. On top of having no correlations between unemployment rate, uninsured health insurance rate, and COVID-19 cases, Dallas has a relatively low death rate in comparison with other counties. It seems that Dallas seems to be doing quite well with surviving COVID-19.

6. Limitations

This analysis has quite a few caveats and limitations attached to them, and that the reader should be aware of. One of the first things to note is the quantity of data for unemployment and uninsured health insurance rates. Unemployment rates contain monthly data, and uninsured health insurance rates contains annual data. That means for unemployment roughly 25 data points overlap the pandemic dates, and for the uninsured health insurance rates there are only 2-3 data points overlapping. Therefore, uninsured insurance data can only be inferred from results of other tests and visualizations.

The regression test for COVID-19 cases and unemployment rate violates the assumptions of linearity and normally distributed data. However, the data size of the unemployment rate seems reasonably large enough for central limit theorem to apply and ignore this assumption. The assumption of linearity among the data is quite there, which brings slightly a bit of an issue for the regression. The other regression for mask mandates and COVID-19 cases, violates a bit of normality, but it doesn't seem to be extremely out of the ordinary. However, just like the previous regression, the sample size is relatively large for both periods of times with and without mask mandates, so this assumption can be ignored by the central limit theorem.

7. Conclusion

In conclusion, my hypothesis was completely wrong. My hypothesis was that mask mandates would slow down COVID-19 cases, and in turn help unemployment and uninsured rates. It turns out mask mandates still had a very high COVID-19 case count and not much can be added to this topic. I originally thought COVID-19 cases would increase unemployment rate and uninsured health insurance rates, which would in turn also increase the death rate of COVID-19. It turns out that COVID-19 and unemployment and it is likely that COVID-19 is also uncorrelated with uninsured health insurance rates. Furthermore, even with the high uninsured health insurance rates of Dallas, the death rates compared to other counties was relatively low. So it seems that Dallas is doing well in terms of countering against COVID-19, however, I must say that the high uninsured health insurance rates is still a major concern regardless of COVID-19.

8. References

- Fernández, Stacy. "Texas Has the Most People without Health Insurance in the Nation - Again." *The Texas Tribune*, The Texas Tribune, 10 Sept. 2019, <https://www.texastribune.org/2019/09/10/texas-has-most-people-without-health-insurance-nation-again/>.
- Rosin Saez, Jacob Villalobos, et al. "Dallas Has the Worst Uninsured Rate in the Nation." *D Magazine*, 14 Oct. 2019, <https://www.dmagazine.com/healthcare-business/2019/10/dallas-has-the-worst-uninsured-rate-in-the-nation/>.

9. Data Sources

US confirmed cases COVID-19 data from John Hopkins University
https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv

US confirmed cases (death) COVID-19 data from John Hopkins University
https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_deaths.csv

U.S. State and Territorial Public Mask Mandates From April 10, 2020 through August 15, 2021
by County by Day CDC

<https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i>

Mask-Wearing Survey Data The New York Times (July 2 and July 14)

<https://github.com/nytimes/covid-19-data/tree/master/mask-use>

Unemployment Rates

https://data.bls.gov/timeseries/LAUMT4819100000000003?amp%253bdata_tool=XGtable&output_view=data&include_graphs=true

Uninsured Health Insurance Rates

<https://www.census.gov/>