A7 - Final Report

The Impact of COVID-19 on the Job Market in Maricopa County, Arizona

Tommy Chen

2021.12.13

Introduction

In this analysis, I plan to explore how the Covid-19 pandemic and the mask mandate impacted the job market (unemployment rate) of Maricopa County, Arizona. I also want to see how another county with a similar population - Cook County in Illinois, is doing compared to Maricopa County. (Maricopa, AZ has a population of 4.485 million while Cook, IL has a population of 5.15 million.) Furthermore, I want to compare the situation in Maricopa County and that in the whole Arizona State.

From the common analysis assignment, I learnt that there was never a mask mandate in Arizona. In fact, the Governor rescinded all local mask mandates, and banned mask mandates implementation in schools. One major impact of the pandemic is the change in the job market. In some other states with mask mandates (like WA), the government was able to control the pandemic and make some economic recovery. I want to explore how bad the pandemic has influenced the job market in Maricopa, Arizona since this is a state with no mandates at all. Furthermore, I chose Cook County, Illinois as the county for comparison since it has a mask mandate during 2020/05/01 and 2021/06/10. In addition, I want to see if there is a potential difference in the metropolitan area and the rural area since the city population could have better awareness of the pandemic. This project is important and human-centered because it could potentially provide some insight on the importance of pandemic awareness, and show how different policies could impact the pandemic situation and local economy.

Background/Related Work

From the start of this project, I've noticed the unusual, strong anti-mask sentiment of politicians in Arizona State. I did a little search online after completing the common analysis assignment, and it was quite shocking - Not only did the governor rescind all mask mandates in the state, he went as far as to ban the mask implementation in schools. According to one report from the Associated Press, the governor "blocked cash from schools mandating masks". This motivated me to choose a county with a similar population with mask mandates and compare the pandemic situation. I was also interested in how the pandemic would affect the unemployment rate differently in those two counties. I read from this blog that the hardest-hit areas by the pandemic could have jobless claims "double or triple that of the least affected areas." Since

Arizona is one of the states that was hit hard by the pandemic, it would be interesting to see if that is the case. Last but not least, I would like to compare the situation in Maricopa County vs the rest of Arizona since one major metropolitan area - Phoenix is within Maricopa County. I was inspired by this news article to compare the pandemic situation in the two areas.
Therefore I derive the following three research questions from my background research.

- Research Question #1
    - What does the pandemic situation look like in Maricopa County vs Cook County? (Infection rate, daily cases) What about the estimated mask prevalence in those counties?
    - Hypothesis: Maricopa County has a higher infection rate and significantly more cases than Cook County. The estimated mask prevalence in Maricopa County is lower than that in Cook County.
- Research Question #2
    - How was the unemployment rate in Maricopa County affected by the pandemic? What about Cook County?
    - Hypothesis: Every 10% increase in infection rate would result in a 1% increase in unemployment in Maricopa County, Arizona. The number would be lower for Cook County, Illinois.
- Research Question #3
    - One of the major metropolitan areas - Phoenix, is in Maricopa County, Arizona. What does the infection rate look like in Maricopa County compared to the rest of Arizona? What about the unemployment rate? Which area's unemployment rate is more sensitive to the pandemic?
    - Hypothesis: The infection rate is lower in Maricopa County compared to the rest of Arizona State. The unemployment rate is lower in Maricopa County compared to that of Arizona State. The unemployment rate in Maricopa County is less sensitive to the pandemic.

Methodology

The first research question is answered through exploratory data analysis, so there are no fancy statistical methods/models involved. I chose to mainly use data visualizations to present my results. For those plots I made sure that the graphs are properly scaled and titled. I also chose the appropriate colors and legends. Last but not least, I specifically watched out for lie factors and potentially misleading factors because I learnt in this course that graphs could unintentionally tell lies. (the forest fire plot)

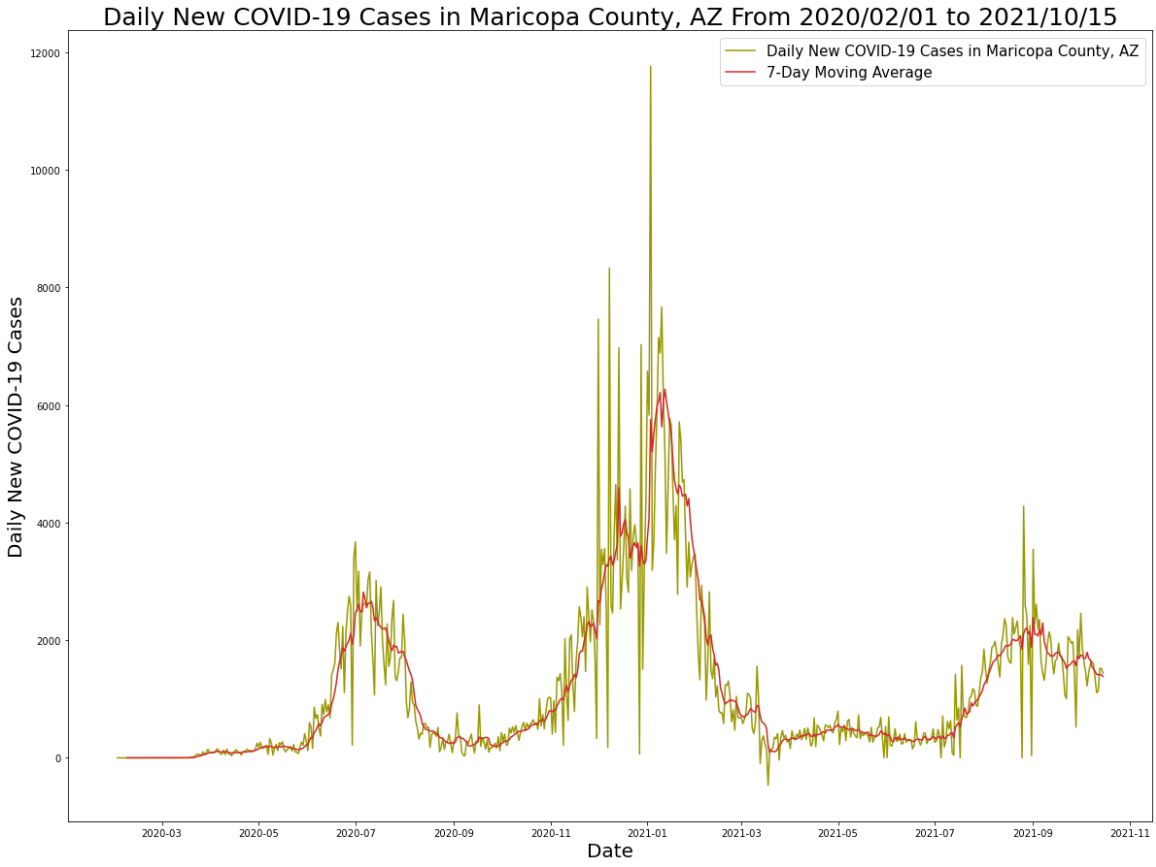
The second and third research question involves exploring the relationship between the unemployment rate and the Covid-19 infection rate. Since I am also interested in how the rate of change of one impacts the other, I built an ordinary least squares linear regression model and tried to interpret the coefficient. In this model, I used the infection rate by month as the predictor variable and the unemployment rate by month as the response variable, both in Maricopa County. However, the initial model's coefficient is insignificant and negative - meaning an increase in Covid-19 infection rate would result in a decrease in unemployment rate. After

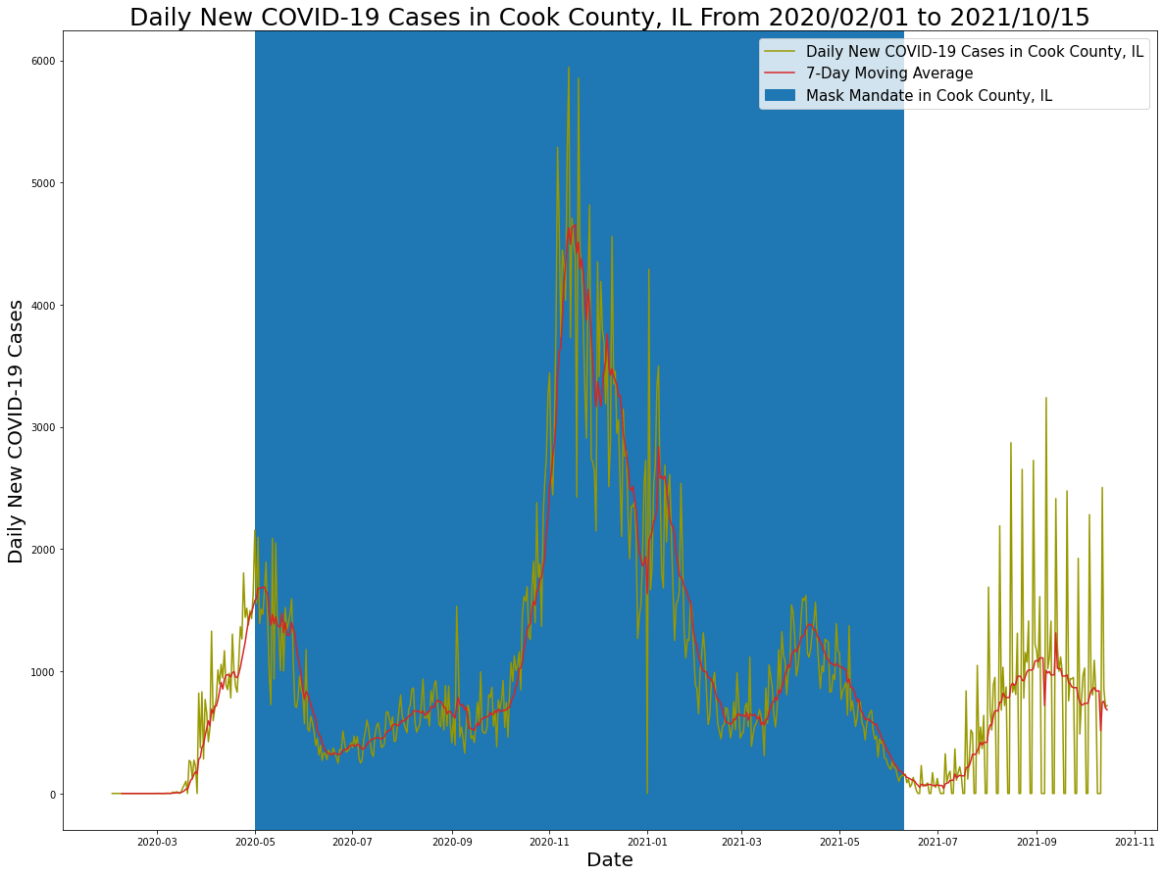
examination, I realized that both the seasonality trend of the virus and the government efforts to boost the economy are potential reasons for this outcome. Therefore I switched my method - instead of modeling unemployment rate against infection rate, I decided to model unemployment rate against the percentage change in infection rate (with respect to the previous month). This makes sense to me since the job market would be mostly impacted by sudden changes brought by the pandemic and stabilize afterwards. This time, the coefficients make much more sense and I have a good interpretation of the models.

Findings

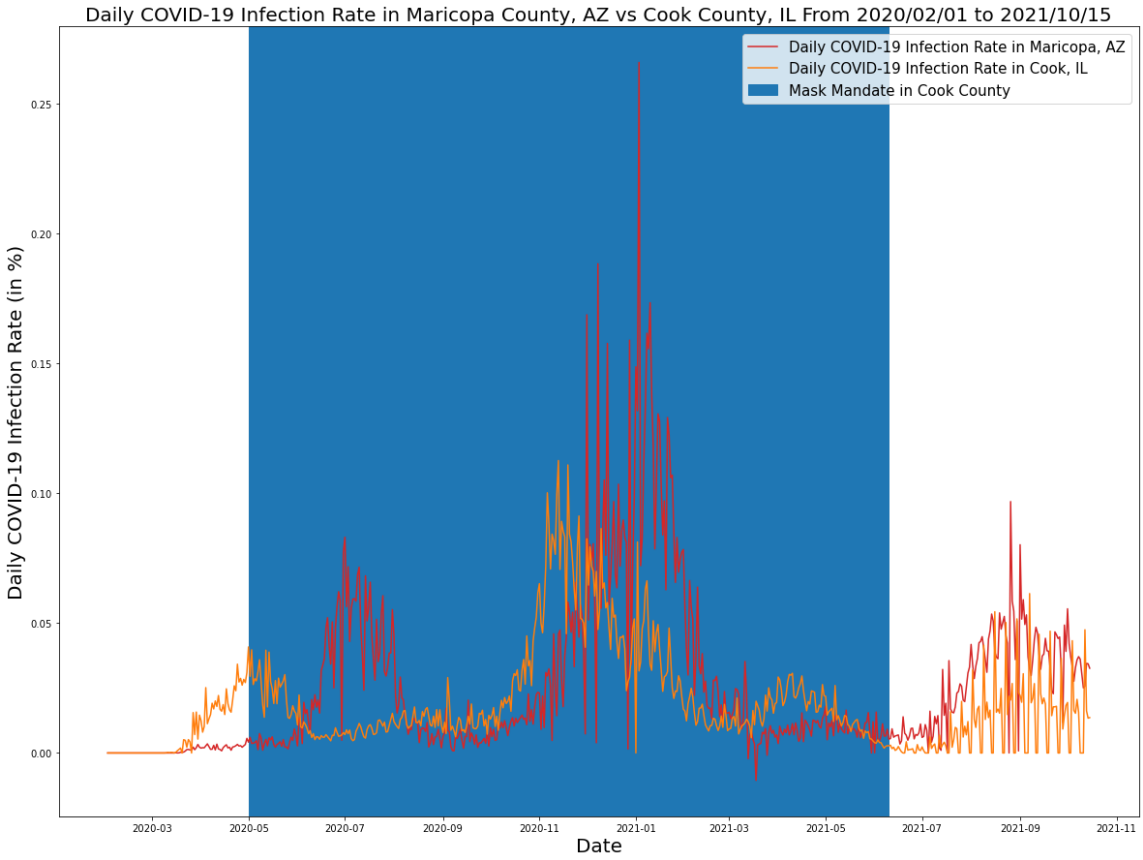
My code and data for this project could be found in [this Github repository].

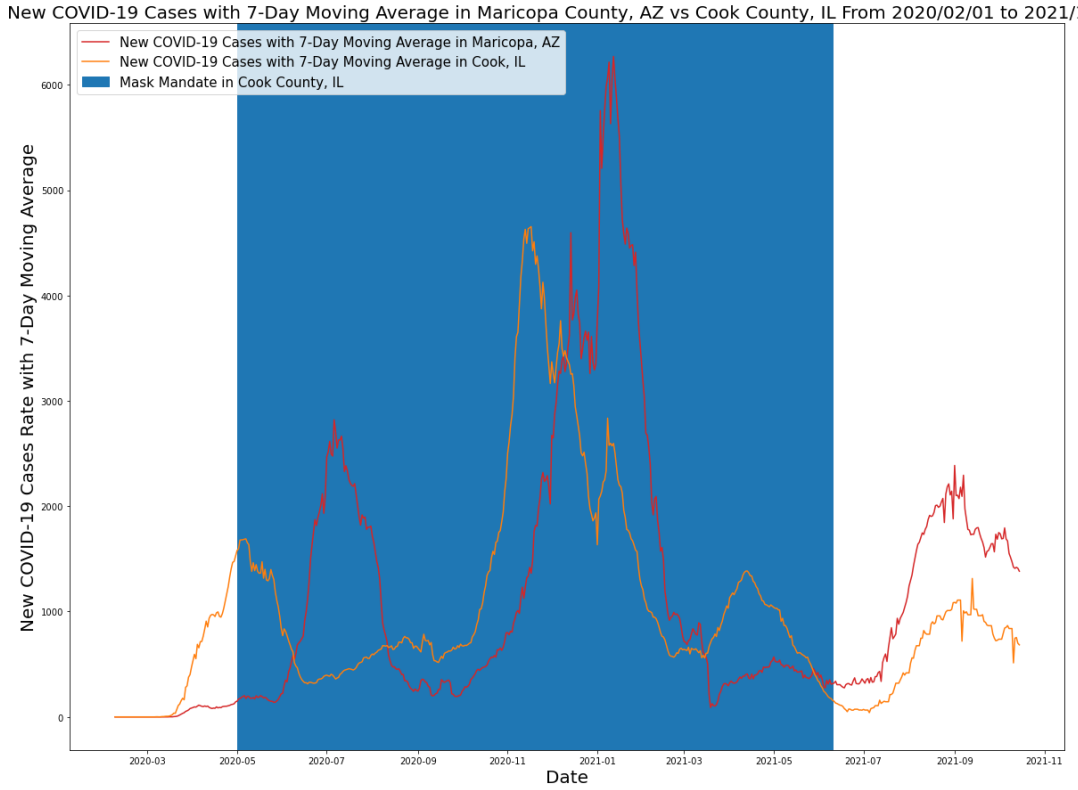
Findings Related to Research Question 1



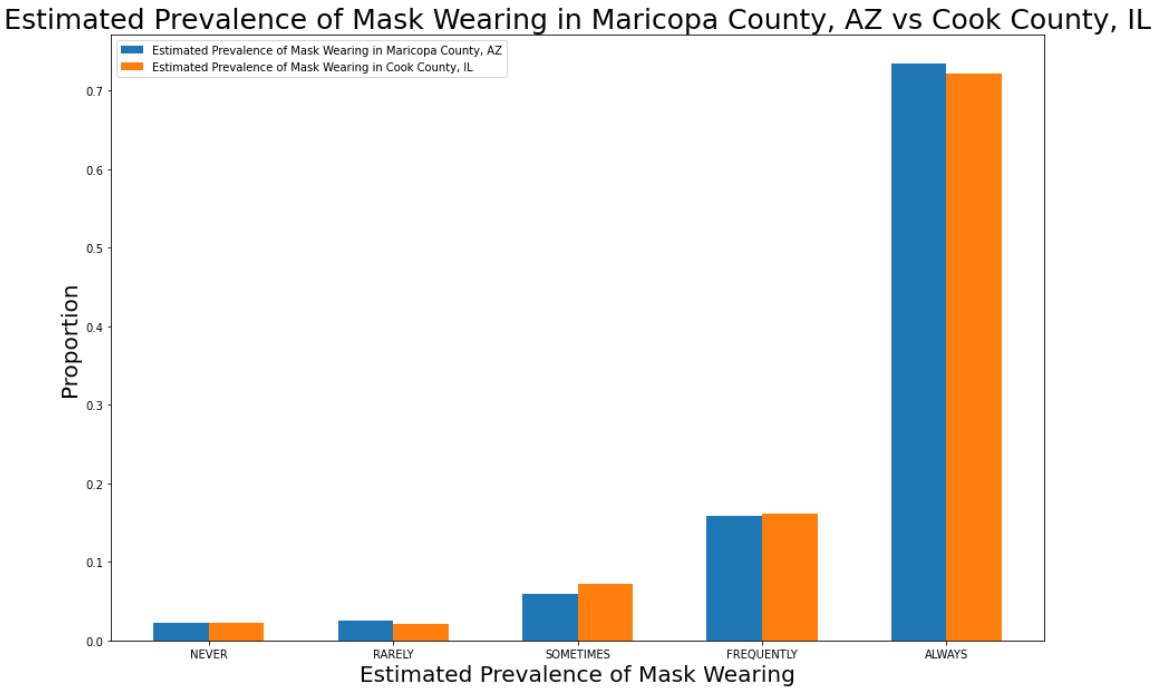


The above two graphs show the daily new cases and the 7-day moving average of daily new cases in both Maricopa County and Cook County from 2020/02/01 to 2021/10/15. With similar numbers in population and estimated mask prevalence, we can see that Maricopa County, Arizona had more daily cases than Cook County, Illinois during 2020/05/01 and 2021/06/10 (marked by the blue area in the second graph), when Cook County had a mask mandate. We can also notice the significant difference in peak cases during this period of time - Maricopa County reached a peak daily new case of 11761 in 2021/01/03 while Cook County reached a peak daily new cases of 5945 in 2020/11/13. However, after the mask mandate was lifted in Cook County, the Covid cases started to tick up and almost matched the situation in Maricopa County. For better comparison, I also plotted the daily infection rates and the 7-day moving average of daily new cases for the two counties in the two plots below.


Daily COVID-19 Infection Rate in Maricopa County, AZ vs Cook County, IL From 2020/02/01 to 2021/10/15
Daily COVID-19 Infection Rate in Maricopa, AZ
Daily COVID-19 Infection Rate in Cook, IL
Mask Mandate in Cook County
Daily COVID-19 Infection Rate (in %)
0.25
0.20
0.15
0.10
0.05
0.00
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
Date


New COVID-19 Cases with 7-Day Moving Average in Maricopa County, AZ vs Cook County, IL From 2020/02/01 to 2021/10/15
New COVID-19 Cases with 7-Day Moving Average in Maricopa, AZ
New COVID-19 Cases with 7-Day Moving Average in Cook, IL
Mask Mandate in Cook County, IL
New COVID-19 Cases Rate with 7-Day Moving Average
6000
5000
4000
3000
2000
1000
0
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
2021-03
2021-05
2021-07
2021-09
2021-11
Date

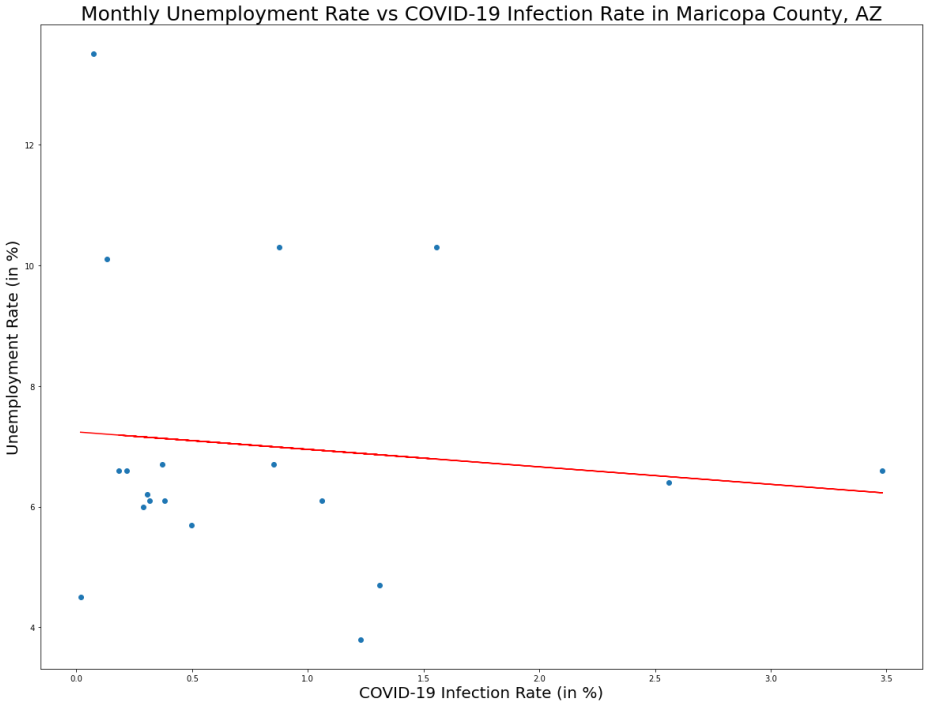
These two plots further confirm the conclusions I drew from the first two plots. From the infection rate plot, it's easy to spot that Maricopa County, AZ (without a mask mandate) had a much worse 'winter peak' than Cook County, IL (with a mask mandate, at the time). However, after the mask mandate was lifted in Cook County, the infection rate and 7-day moving average started to tick up, but the pandemic situation in Cook County was still better as of 2021/10/15.



I was also interested in the mask prevalence in both counties, so I plotted the above bar chart. I hypothesized that with no mask mandate, Maricopa County would have a much lower mask prevalence than Cook County. To my surprise, the two counties had very similar estimated prevalence of mask wearing. After much consideration, I developed two possible explanations for this, which I included in the Discussion/Implication section.

Findings Related to Research Question 2

For the second research question, I initially tried to model the monthly unemployment rate vs the monthly Covid-19 infection rate using ordinary least squares linear regression. The initial regression line with scatter plot and model diagnostic information are the following:



OLS Regression Results

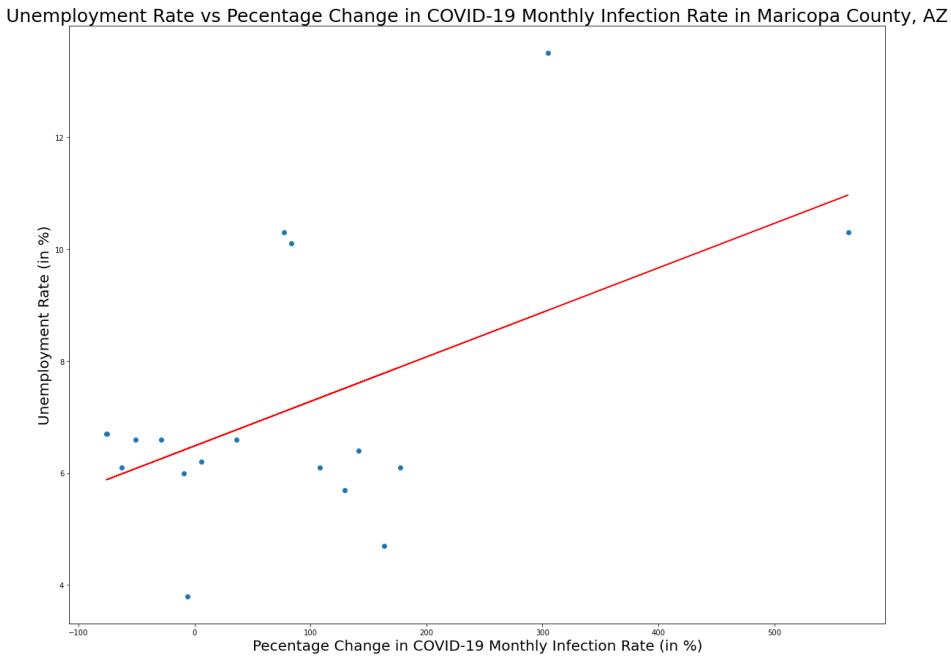
Dep. Variable:	y	R-squared:	0.012
Model:	OLS	Adj. R-squared:	-0.046
Method:	Least Squares	F-statistic:	0.2100
Date:	Sat, 11 Dec 2021	Prob (F-statistic):	0.653
Time:	17:51:33	Log-Likelihood:	-42.827
No. Observations:	19	AIC:	89.65
Df Residuals:	17	BIC:	91.54
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.2395	0.765	9.460	0.000	5.625	8.854
x1	-0.2897	0.632	-0.458	0.653	-1.623	1.044

Omnibus:	7.902	Durbin-Watson:	1.117
Prob(Omnibus):	0.019	Jarque-Bera (JB):	5.309
Skew:	1.219	Prob(JB):	0.0703
Kurtosis:	3.873	Cond. No.	2.36

It's not hard to see that this model is not exactly working. First of all, we got a negative coefficient of -0.2896. The model interpretation would be the following: Every 1% decrease in infection rate would result in a 0.2896% increase in unemployment in Maricopa, Arizona. So this naive model is telling us that an increase in Covid-19 infection rate helps decrease

unemployment rate, which is absurd. In addition, we got a p-value of 0.653, way above the traditional 0.05 alpha value. It means that we cannot reject the null hypothesis that 'the model coefficient is equal to 0'. After much exploration, I decided to use the percentage change in infection rate as my predictor variable, and the results of which are shown below.



OLS Regression Results

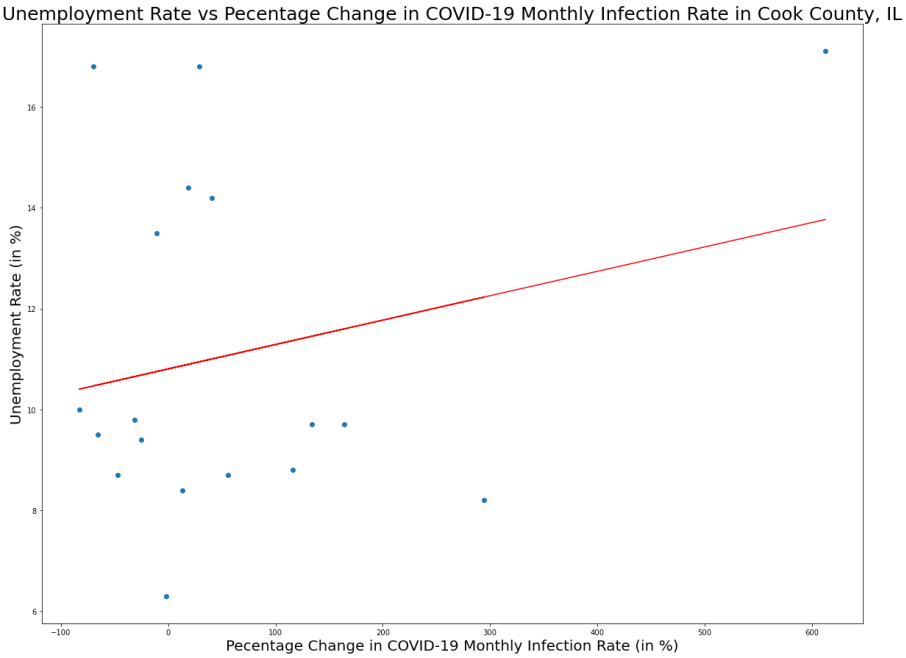
Dep. Variable:	y	R-squared:	0.283
Model:	OLS	Adj. R-squared:	0.238
Method:	Least Squares	F-statistic:	6.321
Date:	Sat, 11 Dec 2021	Prob (F-statistic):	0.0230
Time:	17:51:33	Log-Likelihood:	-37.573
No. Observations:	18	AIC:	79.15
Df Residuals:	16	BIC:	80.93
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.4847	0.553	11.729	0.000	5.313	7.657
x1	0.0080	0.003	2.514	0.023	0.001	0.015

Omnibus:	2.560	Durbin-Watson:	0.730
Prob(Omnibus):	0.278	Jarque-Bera (JB):	1.461
Skew:	0.698	Prob(JB):	0.482
Kurtosis:	3.036	Cond. No.	198.

This time we have much more reasonable results. The model coefficient is now 0.008, meaning that every 1% increase in the percentage change of infection rate would result in a 0.008%

increase in unemployment rate in Maricopa, Arizona. The p-value is now 0.023, smaller than the 0.05 alpha threshold. It means that we can reject the null hypothesis that the model coefficient is equal to 0. Using the same variables, I produced another model for Cook County, IL.



OLS Regression Results

Dep. Variable:	y	R-squared:	0.057
Model:	OLS	Adj. R-squared:	-0.002
Method:	Least Squares	F-statistic:	0.9636
Date:	Sat, 11 Dec 2021	Prob (F-statistic):	0.341
Time:	17:51:34	Log-Likelihood:	-46.461
No. Observations:	18	AIC:	96.92
Df Residuals:	16	BIC:	98.70
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	10.8034	0.859	12.584	0.000	8.983	12.623
x1	0.0048	0.005	0.982	0.341	-0.006	0.015

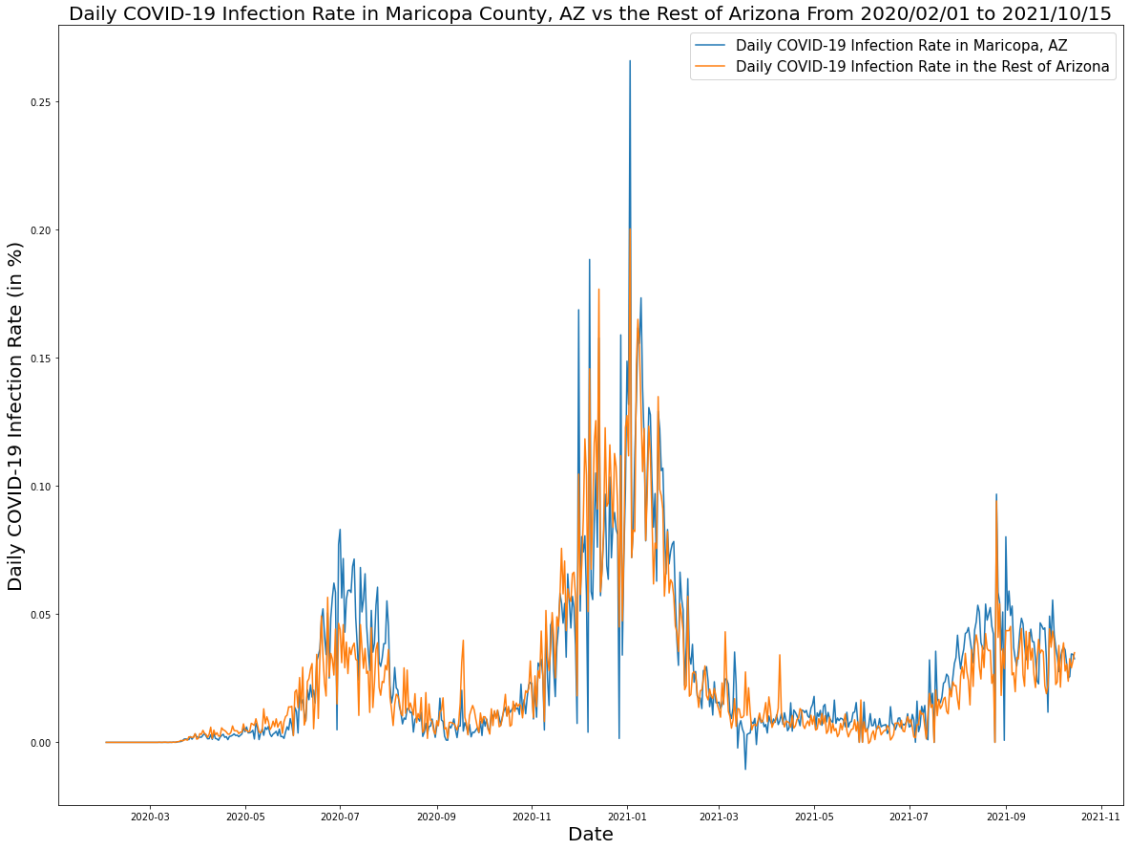
Omnibus:	2.151	Durbin-Watson:	0.257
Prob(Omnibus):	0.341	Jarque-Bera (JB):	1.693
Skew:	0.619	Prob(JB):	0.429
Kurtosis:	2.150	Cond. No.	187.

Compared to 0.008 for Maricopa County, the model coefficient for Cook County is 0.0048. It means that every 1% increase in percentage change of infection rate would result in a 0.0048% increase in unemployment rate in Cook County, Illinois, compared to the 0.008% increase in

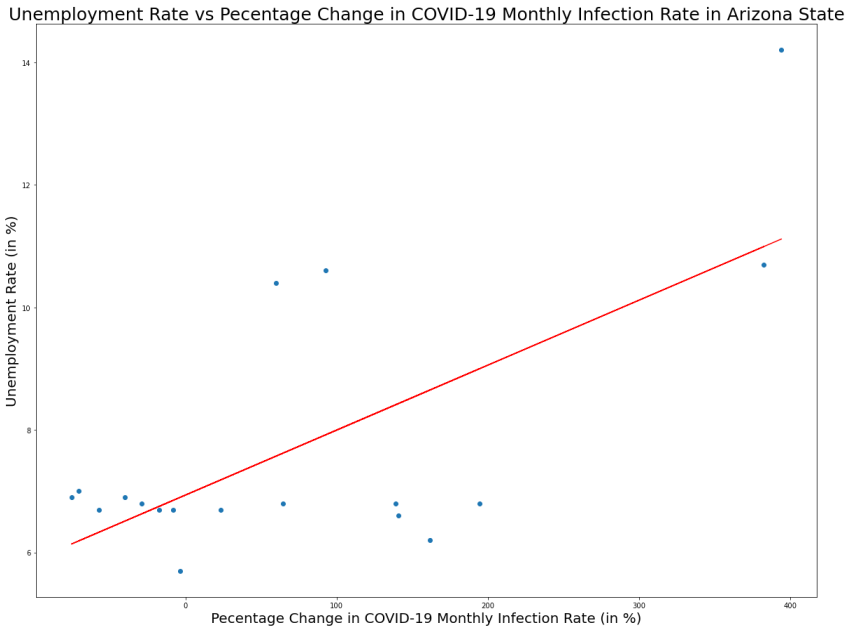
unemployment rate in Maricopa County, Arizona. However, I have to recognize that the p-value is 0.341, which is most likely due to the simplicity of my models.

Findings Related to Research Question 3

In the third research question, I looked into the pandemic situation in Maricopa County vs the rest of Arizona. I hypothesized that the situation in Maricopa County would be better since one major metropolitan area - Phoenix, is within Maricopa County. However, the results shown below displays a similar daily Covid-19 infection rate in Maricopa County and the rest of Arizona, with Maricopa County having a slightly higher peak infection rate. I realized that I only considered the awareness factor in my original hypothesis and ignored the population density factor. The city areas are much more populated and could result in a higher infection rate.



Last but not least, I modeled the monthly unemployment rate vs the percentage change in Covid-19 infection rate for the entire Arizona State and compared it with the Maricopa model.



OLS Regression Results

Dep. Variable:	y	R-squared:	0.454
Model:	OLS	Adj. R-squared:	0.420
Method:	Least Squares	F-statistic:	13.32
Date:	Sat, 11 Dec 2021	Prob (F-statistic):	0.00216
Time:	17:51:35	Log-Likelihood:	-33.888
No. Observations:	18	AIC:	71.78
Df Residuals:	16	BIC:	73.56
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.9371	0.453	15.300	0.000	5.976	7.898
x1	0.0106	0.003	3.650	0.002	0.004	0.017

Omnibus:	1.007	Durbin-Watson:	0.790
Prob(Omnibus):	0.605	Jarque-Bera (JB):	0.860
Skew:	0.472	Prob(JB):	0.650
Kurtosis:	2.494	Cond. No.	178.

The coefficient for the Arizona model is 0.0106, meaning that every 1% increase in percentage change of infection rate would result in a 0.0105% increase in unemployment in Arizona State. Compared to 0.008, the coefficient for the Maricopa model, we can see that Maricopa's unemployment rate is less impacted by the change in pandemic situation than that of Arizona State. It could imply that the metropolitan area's job market is less sensitive to sudden changes such as the pandemic and the rural areas are more affected. The p-value here is very small (0.002) so we can safely reject the null hypothesis that the coefficient of this model is equal to 0.

Discussion/Implications

During this project, I encountered two surprising results and the knowledge of human-centered DS aided me in resolving those surprises. One surprising aspect of my findings is that Maricopa County and Cook County display similar estimated mask prevalence, although the latter has a mask mandate and much fewer daily new cases. I have two potential explanations for this. The first explanation is that the estimated mask prevalence dataset is collected through surveys by the New York Times. I have learnt from this human-centered DS course that surveys could be biased and sometimes survey results don't represent the general population. For example, there could be survival bias involved - people who are willing to take the survey are oftentimes not a good representation of the general population. Another possible explanation is that under the assumption of similar mask wearing, mask mandate policies actually play a very important role in regulating people's behavior, and thus influencing the severity of the pandemic. It's not hard to imagine that people in Arizona could develop a false sense of security with no mask mandates and become more reckless, resulting in significantly more Covid-19 cases. This is interesting for future work because it could show the impact of government guidance on citizen's ordinary behavior.

Another surprise is the unideal coefficient and p-value of my initial modeling of unemployment rate vs infection rate. After getting those results, I attempted to think in the human-centered DS mindset and it helped me come up with some explanations. The virus has a strong seasonal trend, most notably the winter peak. During that time, the federal government has already taken actions to boost the economy and the job market. The unemployment rate stabilized, and the economy partially recovered. It soon occurred to me that human perceptions of the pandemic could play an important role here - If the news of a mysterious, new virus started to spread, people are going to panic, and that's when the job market and economy would be hit most severely. Although there are more daily new cases and higher infection rate during the winter peak, people had a good understanding of the virus and the economic situation, and thus were less impacted than the starting stage of the pandemic. Therefore I made the decision to use the percentage change of infection rate in my modelling process, and it produced much more reasonable results. One possible future research direction is that if we were able to find data quantifying panic to the pandemic, we can explore the relationship between the panic caused by the pandemic and the economic situation.

Limitations

Although I was able to resolve some challenges in this project, I recognize that there are many limitations. In my opinion, the following are the most important three limitations for the project - data limitation, model limitation and the limitation of unemployment rate as indicator for the overall job market.

Data

First of all, there are many data limitations. I have found limited datasets for this project and wished for more data details. For example, I was hoping that the unemployment dataset would

contain more details and I could possibly research more into how Covid-19 impacted different sections of the job market. I also recognize that there are limited data points available - for the monthly data I have only 19 data points from 2020/02/01 to 2021/09/01. The pandemic started in February 2020 and the current unemployment datasets only contain results up to 2021/09/01. In addition, I was hoping to explore the impact of the mask mandate for Maricopa and was disappointed to find out that there was never a mandate. Fortunately, I was able to tackle this limitation by introducing another county with a mask mandate for comparison.

Model

In addition to data limitations, I recognize that there are a few model limitations as well. First of all, the models in this project are one-variable, OLS linear regressions. Since I'm using the unemployment rate as the response variable, I'm almost certain that there are 'hidden variables' - other factors that contribute to the unemployment rate. However, it would be very challenging trying to find those variables and most importantly, find open-source data for them. (again, data limitation). Another limitation is that I'm not doing any seasonal adjustments to the time series data in my models. Maybe a more complicated model with adjustments would perform better. However, I have to remind myself that just because a model is more complicated doesn't mean that it is necessarily better. In human-centered DS I learnt that traditional accuracy metrics are only part of model evaluation - human interpretation is just as important!

Unemployment Rate vs Job Market

Although I modeled the relationship between the unemployment rate and the percentage change in infection rate, I believe there are other factors impacting the unemployment rate other than the pandemic itself. For example, the stimulus checks distributed by the government and the Federal Reserve's monetary policies could both significantly influence the unemployment rate. In addition, the unemployment rate is not representative of the situation of the job market. This involves some complicated economical background knowledge and I don't possess them.

Conclusion

Here are the research questions/hypotheses and the summary of my findings.

- Research Question #1
    - What does the pandemic situation look like in Maricopa County vs Cook County? (Infection rate, daily cases) What about the estimated mask prevalence in those counties?
    - Hypothesis: Maricopa County has a higher infection rate and significantly more cases than Cook County. The estimated mask prevalence in Maricopa County is lower than that in Cook County.
    - Finding: Within the time frame of the mask mandate in Cook County, there are fewer daily cases in Cook County compared to Maricopa County. However, after the mask mandate was lifted, the daily cases and infection rate in Cook County started to rise. The two counties have similar estimated mask prevalence.

- Research Question #2
    - How was the unemployment rate in Maricopa County affected by the pandemic? What about Cook County?
    - Hypothesis: Every 10% increase in infection rate would result in a 1% increase in unemployment in Maricopa County, Arizona. The number would be lower for Cook County, Illinois.
    - Finding: It's a bad idea to model unemployment rate against infection rate! Instead I modeled the unemployment rate against the percentage change in infection rate. Every 1% increase in the percentage change of infection rate would result in a 0.008% increase in the unemployment rate in Maricopa, Arizona. Every 1% increase in the percentage change of infection rate would result in a 0.0048% increase in the unemployment rate in Cook, Illinois. The unemployment rate in Maricopa County is more sensitive to the percentage change in infection rate than that in Cook County.
- Research Question #3
    - One of the major metropolitan areas - Phoenix, is in Maricopa County, Arizona. What does the infection rate look like in Maricopa County compared to the rest of Arizona? What about the unemployment rate? Which area's unemployment rate is more sensitive to the pandemic?
    - Hypothesis: The infection rate is lower in Maricopa County compared to the rest of Arizona State. The unemployment rate is lower in Maricopa County compared to that of Arizona State. The unemployment rate in Maricopa County is less sensitive to the pandemic.
    - Finding: The Phoenix metropolitan area (Maricopa County) in Arizona has a slightly higher infection rate but slightly lower unemployment rate compared to the rest of Arizona. Unemployment rate in rural areas of Arizona is more sensitive to sudden changes brought by the pandemic than metropolitan areas. Every 1% increase in the percentage change of infection rate would result in a 0.0105% increase in the unemployment rate in Arizona State (compared to 0.008% for Maricopa County).

Throughout this project, I was able to have a deeper understanding of human-centered data science. I was able to tackle the challenges met with a human-centered DS mindset and use that mindset for model interpretation (detailed in the Discussion/Implication section). I was also able to reaffirm the idea that traditional accuracy metrics are only part of model evaluation - human interpretation is just as important!

References

1. *Just Before Taking Effect, Arizona's School Mask Mandate Ban Ruled Unconstitutional*, 2021.09, NPR https://www.npr.org/2021/09/27/1041044436/just-before-taking-effect-arizonas-school-mask-mandate-ban-ruled-unconstitutiona

2. *Arizona governor blocks cash from schools mandating masks*, 2021.08, Associated Press
https://apnews.com/article/health-arizona-coronavirus-pandemic-f4c807150376d4a1dc1ea88b3aecce7c

3. *The unemployment impacts of COVID-19: lessons from the Great Recession*, 2020.04, UP Front
https://www.brookings.edu/blog/up-front/2020/04/15/the-unemployment-impacts-of-covid-19-lessons-from-the-great-recession/

4. Covid is killing rural Americans at twice the rate of people in urban areas, 2021.09, NBC
https://www.nbcnews.com/health/health-news/covid-killing-rural-americans-twice-rate-people-urban-areas-n1280369

Data Sources

1. The RAW_us_confirmed_cases.csv file from the Kaggle repository of John Hopkins University COVID-19 data.
https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv

2. The CDC dataset of masking mandates by county.
https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i

3. The New York Times mask compliance survey data.
https://github.com/nytimes/covid-19-data/tree/master/mask-use

4. Unemployment rate by month in Maricopa County & Cook County from the Federal Reserve Bank of St. Louis.
https://fred.stlouisfed.org/series/AZMARI3URN
https://fred.stlouisfed.org/series/ILCOOK1URN

5. Unemployment rate by month in Arizona State from the Bureau of Labor Statistics.
https://www.bls.gov/eag/eag.az.htm