A7 DATA 512
Marc Mascarenhas
December 14, 2021

1. Introduction

Why is this analysis interesting or important (to people besides you)? Does it solve a real problem or tackle an unresolved research question?

Jefferson County, Kentucky, like most counties in the U.S., has been badly affected by the COVID -19 pandemic. With a vaccination rate that is below the US average, it's been struggling with COVID cases. On June 11th 2021, the mask mandate that was in effect in the county was lifted - this was accompanied by an uptick in cases soon after the removal of the mandate.

The below visualization shows the 7day rolling average of COVID case counts for Jefferson County, Kentucky from Feb 2020 - Oct 2021. The case counts are color coded by whether or not, on that day a mask mandate was in place.

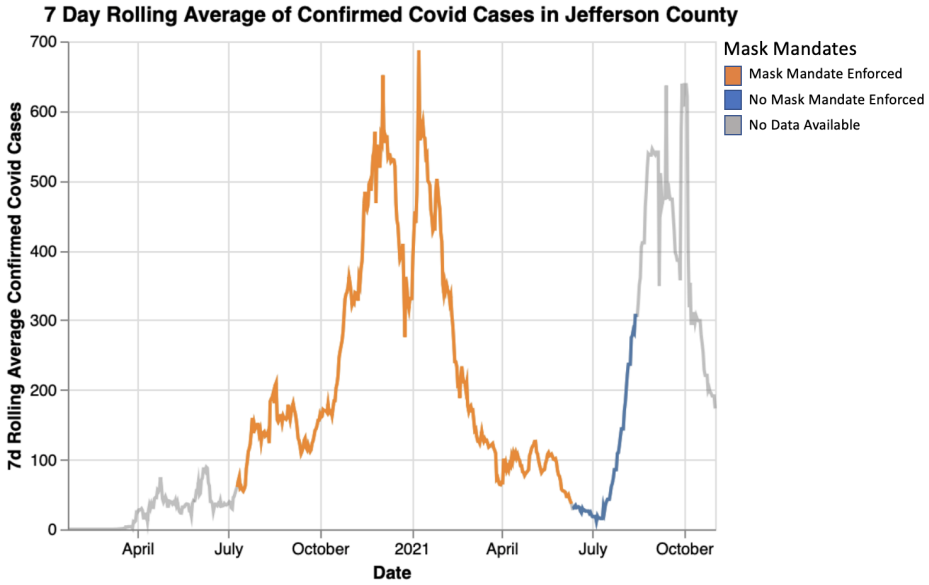

Figure 1: 7day rolling average of COVID case counts for Jefferson County, Kentucky from Feb 2020 - Oct 2021

In this analysis, I decided to dig deeper into this uptick, to better understand the human behaviour associated w.r.t. the pandemic - specifically was there a psychological change when the mask mandate was removed? Did people feel more comfortable going out to public spaces (parks) ?

I think this is interesting because it would give us more insight into people's perception of risk and openness during the course of the pandemic. This could be

helpful in terms of behavioural economics - once governing bodies have a better understanding of human behavioural response towards the pandemic, it could help structuring policies to nudge behavior towards social good without the need for harsh restrictions.

If this indeed was the case, it might do good to re- look into the wording associated with the relaxation of the mask mandate - something that wouldn't give people the false sense that it's ok to go back to normal.

Being able to quantify this would also help people understand their actions at a macro level and help in better personal decision making.

A7 DATA 512
Marc Mascarenhas
December 14, 2021

2. Background/Related Work

What other research has been done in this area? How does this research inform your hypotheses, your analysis, or your system design? What are your hypotheses or research questions?

For these COVID related questions there may not be peer-reviewed publications that are directly related to your hypothesis. There may be anecdotal claims in the popular press (blogs, newspapers) related to your analysis.

Over the last year, there has been a lot of research in terms of public mobility in the wake of COVID. However most of the analysis that I came across was at a macro level comparing different countries /states.

As noted in the research article, "Human Mobility to Parks Under the COVID-19 Pandemic and Wildfire Seasons in the Western and Central United States", different areas have had vastly different reactions to the Pandemic. In their qualitative analysis the authors observe how park mobility has decreased in more risk averse populations of society whereas it has increased in areas where people have used parks as a substitute for fitness and recreation.

In this analysis, I'm interested in getting a better understanding of how the people of Jefferson County, KY responded to this mask mandate relaxation on June 11th.

To be able to answer this question, I will supplement the data from A4 with Google Mobility data. The dataset provides us with percent increase (vis -a -vis a baseline) in mobility in public places at the county level. This data is exactly what we need to answer our question.

Over the last 2 years, they have been collecting geo location data and they have recently anonymized it and made it public. Since then, there have been a lot of interesting questions and blogs posted that leverage this data. Reading through the analysis done here (that looked into trends across different countries in the world) I decided to zoom into Jefferson county, KY and do something similar but at a specific point in time wrt park mobility.

There are a couple of caveats to keep in mind though -

- The baseline for this dataset is mobility in the county in Jan 2020.
- There are a number of other factors that also affect mobility in parks - seasonality being a big one
- This only records data from people whose location services are on and have consented to share data

The link to the data set is:
https://www.google.com/covid19/mobility/index.html?hl=en

To ensure privacy, Google has used differential privacy to try to better anonymize the data.

In A4, we looked at how mask mandates affect spread of covid. However, there are a number of intermediate factors between mask mandates and spread of covid - one of which is mobility in public spaces. This data extends the analysis to address this question.

Research Question:

How has people's mobility wrt parks changed since the onset of COVID in Jefferson County? Specifically, what effect did the mask mandate have?

Hypothesis:

With the relaxation of the mask mandate, people started frequenting public places like parks more, which in turn led to the increase in COVID cases as seen in Fig 1.

A7 DATA 512
Marc Mascarenhas
December 14, 2021

3. Methodology

Not just your analytical methods, but also, why you chose them, and how human-centered considerations such as ethics informed the way you designed your study.

To be able to test out the hypothesis, I broke down the process into 3 steps:

Step 1: Understand the data and the peculiarities associated with the data

Step 2: If the mask mandate wasn't lifted on June 11th, what would the park mobility be? Can we create a distribution of what would be expected?

Step 3: How extreme is the park mobility data for june 11th - june 18th compared to what was expected?

Step1:

The google mobility data is interesting in the fact that they don't provide actual mobility data values, but rather they show the percent increase w.r.t. the median value for a given day of the week before the start of the pandemic in January - February 2020. So a value of 61 on a monday means that on that day the park mobily was at 61% of the median value of the mondays in the baseline period.

	sub_region_2	date	parks_percent_change_from_baseline
272071	Jefferson County	2020-02-15	61.0
272072	Jefferson County	2020-02-16	30.0
272073	Jefferson County	2020-02-17	43.0
272074	Jefferson County	2020-02-18	-8.0
272075	Jefferson County	2020-02-19	21.0
...	...	...	...
271763	Jefferson County	2021-11-02	16.0
271764	Jefferson County	2021-11-03	8.0

Figure 2: Google Mobility data format

Since we were comparing the days of the week to their corresponding baseline, I expected the increases to be proportional - however this was not the case (fig 3). Saturdays (day 6) had much higher increases compared to the other days of the week. However, in the experimental setup, since I looked at the mean in a 7 day period, I decided to not investigate this further.

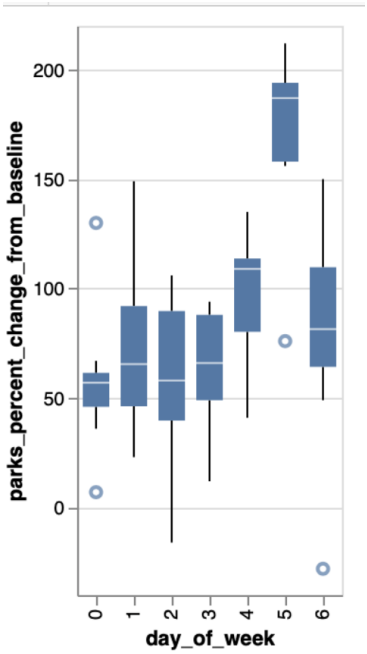

Figure 3: Park Mobility in the 1.5 months before the relaxation of the mask mandate (0: Monday, 1: Tuesday...)

Step 2:

For the analysis, I decided to use a non parametric t-test. This was done as it frees us from the assumption of normality of the underlying distribution of the data as we create the distribution through bootstrapping.

Inorder to create the expected population distribution, I decided to use data from 1.5 months before june 11th. To create the bootstrapped the distribution, I sampled with replacement 7 days and then calculated the mean:

- Repeat 1000 times:
    - Sample 7 days with replacement
    - Calculate mean park mobility of these 7 days

A7 DATA 512
Marc Mascarenhas
December 14, 2021

```
population = []
temp_data = data[(data["date"]>"2021-05-01") & (data["date"]<"2021-06-11")]
for i in range(1000):
    sampled = temp_data.sample(7,replace = True)
    population.append(sampled["parks_percent_change_from_baseline"].mean())

```

Figure 4: Code for creating population distribution

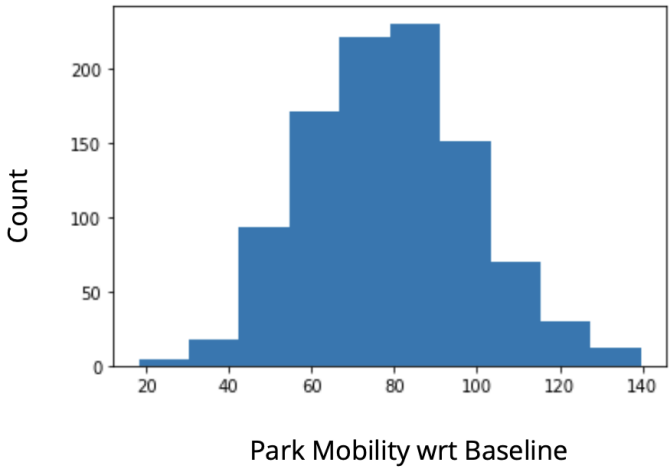

Figure 5: Population distribution of park mobility in Jefferson County, KY

Step 3:

The mean park mobility for the 7 day period from 11th June - 18th June was 100.57% with a sd of 30.58

A7 DATA 512
Marc Mascarenhas
December 14, 2021

4. Findings

What did you find? Use words and figures, don't just point to code.

In the week following the relaxation of the mask mandate, the mean park mobility was 100.57. This corresponds to a p value of 0.14. Thus while this is greater than the population average, it's not enough to reject the null hypothesis.

Hence we cannot conclude that there was any significant unexpected change in park mobility in Jefferson County, KY following the relaxation of the mask mandate.

Figure 6 visually displays the extremeness of the week of June 11- June 18.

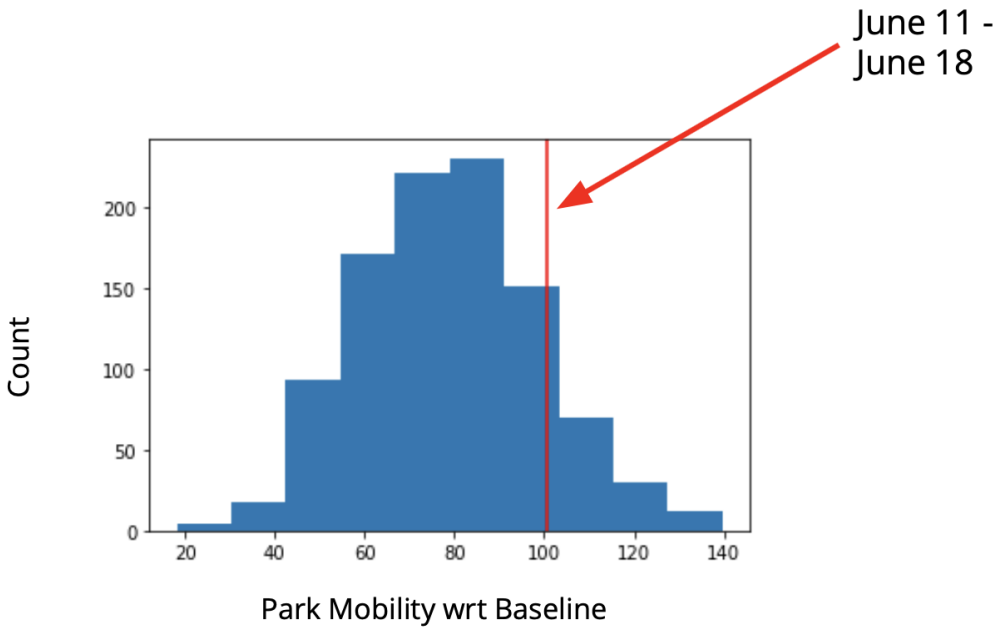

Figure 6

A7 DATA 512
Marc Mascarenhas
December 14, 2021

5. Discussion/Implications

Why are your findings important or interesting; How could future research build on this study?

Through this experimental setup we found that we do not have enough evidence to reject the null hypothesis. Based on the findings there are a couple of hypotheses that would make sense to look into next.

The underpinnings of the analysis was that the mask mandate relaxation led to an uptick in cases and that this uptick was due to a behavioural change that made people more relaxed and comfortable in going out to public spaces.

However other possibilities could be

- People's perception of going out to public spaces changed with the vaccine
- The change in perception was gradual and there wasn't a single event that caused the shift.
- The county is not granular enough and while some areas did experience upticks with the mask mandate relaxation, it wasn't uniform

I think the setup in this analysis is easily reproducible across other types of mobility and it would be interesting to see

a) What other type of mobility were affected with the mask mandate
b) What other events caused upticks in mobility as well as cases

A7 DATA 512
Marc Mascarenhas
December 14, 2021

6. Limitations

This is a required section for your report.There are often many, many limitations for any study. If you honestly tried to list them all, this might end up being the longest section. You should prioritize and list the ones that are most likely to have a significant impact on your results. Specific license issues could be a limitation, depending on what data sources you used. Flaws in the data, data cleaning techniques, potential assumptions and/or how you handled missing values could be a limitation. Statistical techniques often have specific assumptions and preconditions; if you're not certain all of the data meets those requirements - this is a good place to make that clear.

The experimental setup that was used was overly conservative and doesn't have as much power. This is due to the fact that I'm calculating the mean mobility across a 7 day period. Mobility during covid has affected different days of the week differently and even though google calculates the % mobility on a day of the week basis, the way we calculate the population sample - there is the possibility of sampling days with the highest % mobilities which depending on the sample, may result in a lower powered test.

Another thing that wasn't accounted for was the seasonality in the population sample. I used the data from May 1st - June 10th to create the sampling distribution. However, earlier in the summer, it most likely was colder and had lower park mobility.

There's also the constraints of the google mobility dataset - to ensure anonymity they add a little noise to the data - I haven't looked too much into how that would change the results of the analysis. My assumption is that it would average out , however, I would need to dig deeper.

Another aspect of the google mobility data is that it uses geolocation information that it obtains from google phones. An assumption that I made was that google users were a representative sample of the people that go to parks/ public places.

Another assumption that was made was with regards to covid case count data. The way Jefferson county dealt with reporting of cases changed over the course of the pandemic.Plotting the data for Jefferson County, Kentucky, there were a lot of days where no data was logged. Digging a little deeper, uptill July 2021, these days,

mostly corresponded to holidays. However, since then, Saturdays and Sundays were also days with no data logged.

To smoothen and address this seasonality, I decided to take a 7 day average.The issue with the 7 day average is that it creates a lag in the noticeability of a change in behaviour. This added to the fact that cases have an inherent lag (from point of getting covid to the point of it being reported), causes the graph to be shifted slightly to the right. I decided to not correct for that at the moment, but rather be cognizant of it and mention it in the analysis.

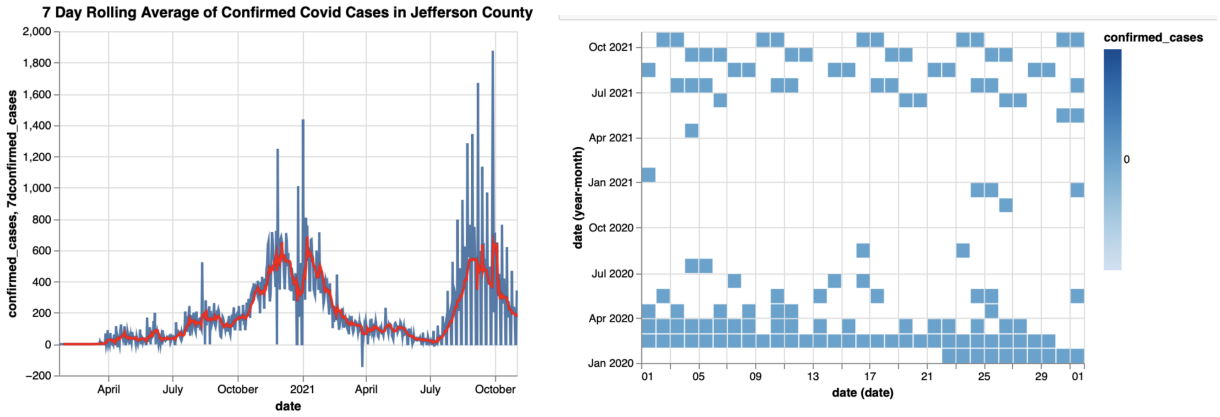

Figure 7 (left): 7d rolling average of confirmed covid cases in Jefferson County, KY; (right) days for which case count was 0 in Jefferson County, KY.

A7 DATA 512
Marc Mascarenhas
December 14, 2021

7. Conclusion

Restate your research questions/hypotheses and summarize your findings. Explain to the reader how this study informs their understanding of human centered data science.

This analysis attempted to examine the effect of COVID -19 & mask mandate relaxation on human behaviour. Specifically I looked into whether the relaxation of the mask mandate resulted in an increase in park mobility in Jefferson County, KY. Through the experimental setup outlined in the analysis, the park mobility in the week following the relaxation of the mask mandate wasn't extreme enough to suggest that the relaxation changed the amount of park mobility.

8. References

- https://www.youtube.com/watch?v=Iq9DzN6mvYA
- https://www.youtube.com/watch?v=TpgiFIGXcT4
- https://www.wdrb.com/in-depth/beshear-to-lift-kentuckys-mask-mandate-and-capacity-restrictions-june-11/article_c1343c64-b4c2-11eb-92b2-e727b89b76f7.html
- https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021GH000494

9. Data Sources

- https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv
- https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i
- https://www.google.com/covid19/mobility/