A7: Project Report

Patrick Peng (ID: 2029888)
DATA 512, Autumn 2021

Introduction

This project explores the COVID-19 pandemic in Milwaukee County, Wisconsin and its effects. It is split into two loosely related parts: a common analysis (A4) that analyzes the effect of mask mandates on COVID-19 infection rates, and an extension (A5) that studies the relationship between infection rates and traffic accidents in Milwaukee City.

The common analysis explores an interesting (and polarizing) question: do local masking mandates have an effect on the COVID-19 infection rate? This question is not particularly novel - due to the political salience of the mask debate, there have been numerous studies studying it. Overall, the scientific evidence generally supports the use of face masks. Our study attempts to recreate these results using a much more limited set of data and practically no domain expertise, while hoping that the results of this amateur analysis are not weaponized by bad actors to the detriment of public health.

The extension analysis shies away from amateur pandemic modeling and instead explores a more human-centered issue in the context of the ongoing pandemic: that of traffic violence. Despite pandemic-related declines in vehicle miles traveled, US traffic deaths saw their biggest ever single-year increase in 2020[1]. In other words, each mile driven on American roads and streets got significantly more hazardous. This trend in higher traffic deaths extends into 2021, even as traffic volumes return to near pre-pandemic levels. This extension analysis attempts to determine whether there is any correlation between COVID-19 and the number of traffic accidents in Milwaukee.

Background/Related Work

Common Analysis

The effectiveness of face mask usage in preventing the transmission of COVID-19 has been the subject of numerous studies. Due to the nature of the pandemic, it has been difficult to perform randomized controlled trials (RCTs), putting aside the ethical risks of assigning a no-mask control group. Consequently, many studies involve comparisons of infection data between states or countries with differing masking rules. For example, Lyu and Wehby[2] compared case growth in US states that adopted mask mandates and those that did not and found that mask mandates

---

[1] Bolotnikova 2021.
[2] Lyu and Wehby 2020.

were associated with lower COVID-19 growth rates in the time period immediately following the implementation of the orders. Howard et al.[3] summarize the available evidence on the effectiveness of face masks, including regional comparisons, epidemiological modeling, and evaluations of their efficacy in reducing exposure to/spread of infectious particles, and reach the conclusion that widespread face mask usage is effective in reducing community transmission.

Crucially, the prescribed scope of the A4 common analysis makes it difficult to repeat the existing studies here. By limiting the analysis to a single US county, we are unable to compare against other locales that did/did not impose mask mandates, and can only compare within the same time series for the same county. However, this introduces numerous external factors (e.g. vaccinations, variants) that are not accounted for in the analysis.

Extension Analysis

Much of the available literature on traffic collisions, injuries, and deaths in the COVID-19 pandemic is concerned with the effect of lockdown and stay-at-home orders, rather than with the COVID-19 infection rate specifically. A review of pandemic-related traffic studies from around the world by Yasin et al.[4] found that while traffic volume fell everywhere, even in countries that did not institute a lockdown, the impact on road deaths was mixed. Of the 42 countries studied, 33 saw fewer road deaths in 2020 than in 2019, while the remainder (including the United States) saw an increase. Further, the changes were not evenly distributed among all road users - some studies found that the death rate for vulnerable road users (e.g. pedestrians, motorcyclists) decreased less than that for drivers or even increased. Incidences of speeding, seatbelt noncompliance, and driving under the influence of drugs or alcohol were also reported to have risen in many countries. In some areas, the overall number of collisions declined while the number of serious collisions remained steady or increased. Lin et al.[5] observed that in Los Angeles and New York City, pandemic-era reductions in traffic volume led to fewer collisions, but that the collisions tended to be more more concentrated among Hispanic and male drivers and in lower-income minority areas compared to the pre-pandemic baseline.

Methodology

Common Analysis

The common analysis was focused on producing two deliverables: a data visualization showing the time series of the infection rate and its derivative, with the duration of the mask mandate in effect, and answering the question of whether the mask mandate had an effect on the progression of the pandemic. The following datasets were provided for this purpose:

[3] Howard et al. 2021.
[4] Yasin et al. 2021.
[5] Lin et al. 2021.

- The RAW_us_confirmed_cases.csv file from the Kaggle repository of John Hopkins University COVID-19 data.
- The CDC dataset of masking mandates by county.
- The New York Times mask compliance survey data.

I determined that the mask compliance survey data would not be particularly useful in this analysis, as it is provided at a single time point, and discarded it. Producing the deliverables first required the calculation of the daily infection rate, based on the daily new case counts from the JHU dataset. Infection rate is a parameter that describes the frequency of new infections in a population during a specific time period and is calculated, per Wikipedia, as Equation 1.

$$Rate\ of\ infection\ = \frac{number\ of\ infections}{number\ of\ people\ at\ risk\ of\ infection}$$

Equation 1: Calculation of infection rate

To calculate the infection rate from the case count data, I made the following simplifying assumptions:

- All infections are asymptomatic for the first 7 days (value chosen midway between the generally acknowledged range of 2-14 days).
- Upon the appearance of symptoms on day 7, every infection is reported and added to the confirmed case counts dataset. In other words, no infections go uncounted.
- Every infected patient recovers 14 days after symptoms appear (21 days after first infection). Nobody dies.
- Patients are contagious for the full 21 days of the infection and are not contagious after recovery.
- Recovered patients cannot be reinfected.

I first applied a 7-day moving average to the daily new case counts data to smooth out day-to-day volatility. Next, under the assumptions outlined above, I calculated the number of active infections on each day (the sum of that day's new infections and ongoing infections from prior days) and population at risk (the 2020 Census count for Milwaukee County less the cumulative number of recovered patients), then divided them to obtain the daily infection rate. The daily change in infection rate was obtained by calculating the day-to-day difference in infection rate.

Performing a database-style join of the infection rate table and the mask mandate table using the date as the join predicate allowed the data to be consolidated into a single table for ease of plotting. The resulting visualization is shown in Figure 1.

The second deliverable, evaluating the effect of the mask mandate on the course of the pandemic, was somewhat trickier to answer. Using only the infection rate and the general knowledge that pandemics exhibit exponential growth, I tried to fit the exponential function y=A*exp(B*x) to the infection rate curve at various points to see whether the coefficient B (which influences how rapidly the function increases) changes after the mask mandate is put in place. If mask mandates decrease the rate of disease transmission, we would expect the value of the B coefficient to be smaller after the mandate takes effect.

I selected time intervals corresponding to the build-up of each pandemic “wave,” believing that this would be the most appropriate application of the exponential growth model. As seen in Figure 1, Milwaukee County experienced two distinct “waves” of infection starting in April and June 2020, prior to the mask mandate, and two waves while the mask mandate was in effect, in September 2020 and March 2021. I excluded data after about May 2021 as the combined effect of vaccines and the Delta variant rendered any comparison to earlier data inappropriate.

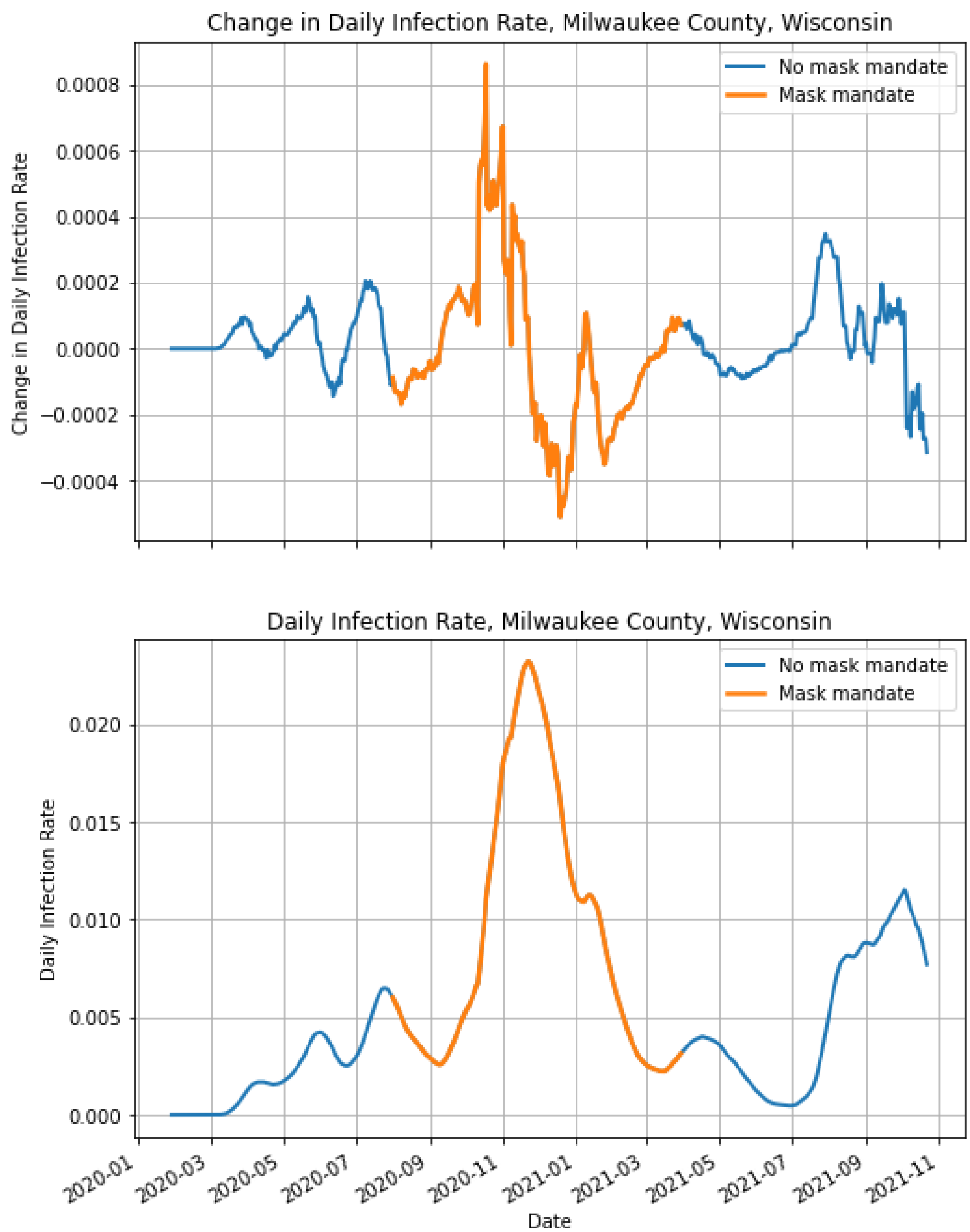

Figure 1: Visualization for the common analysis, showing the COVID-19 infection rate and its derivative between January 2020 and the present. The period in which a mask mandate was in effect is plotted in a different color.

Extension Analysis

The extension analysis sought to determine whether there was a correlation between COVID-19 infection rates and the number of traffic collisions. In addition to the COVID-19 data used in the common analysis, two additional datasets were introduced for this exercise:

- The City of Milwaukee's traffic accidents dataset, cataloguing the date, time, and location of all traffic accidents in Milwaukee City recorded by police, dating back to about 2015. Data is provided by the City via its data portal and is updated daily.
- The average daily bidirectional traffic at Wisconsin DOT continuous traffic counter #400003, located on Interstate 94 at 26th Street, west of Downtown Milwaukee. Averages for each weekday are computed at the monthly level. Data provided by the Wisconsin Traffic Operations and Safety Laboratory at the University of Wisconsin, Madison.

After applying a 7-day moving average to the traffic accidents series to smooth out day-to-day volatility (shown in Figure 2), I computed a normalized accident rate using the traffic counter data. The rationale for this normalization is that traffic volume and the number of traffic accidents are known to be correlated. As changes in the accident count could be driven by pandemic-induced changes in traffic volume, using a normalized value controls for that. Typically, accident or death rates are reported in terms of collisions (or deaths) per mile driven, as it is easily interpretable - what is the human toll of each mile? However, vehicle miles traveled (VMT) data is generally not available at a daily or monthly granularity, which would be necessary to calculate this normalized rate over time. In the absence of VMT data, I assumed that the traffic counts at this counter would follow the same trends as regional VMT. Indeed, the traffic count data does show a sharp drop in April 2020 when stay-at-home orders were instituted in Milwaukee, recovering to normal levels later in the year, as shown in Figure 3. The resulting normalized accident rate is shown in Figure 4.

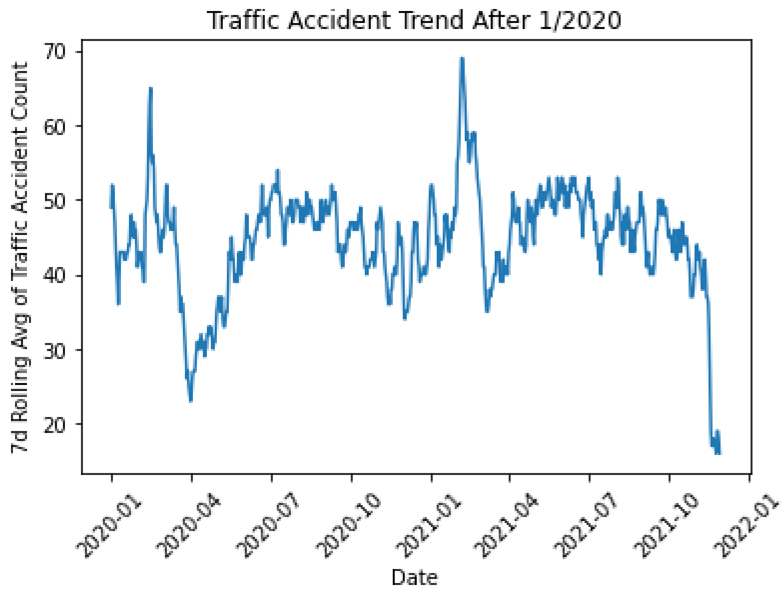

Figure 2: 7-day rolling average of traffic collisions in the City of Milwaukee

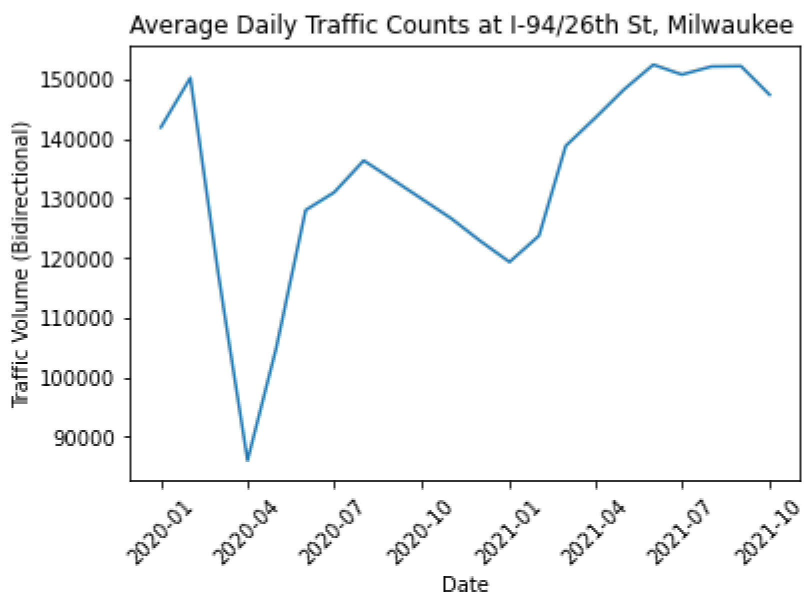

Figure 3: Daily bidirectional traffic counts at Wisconsin DOT counter #400003 at I-94/26th Street, west of downtown Milwaukee

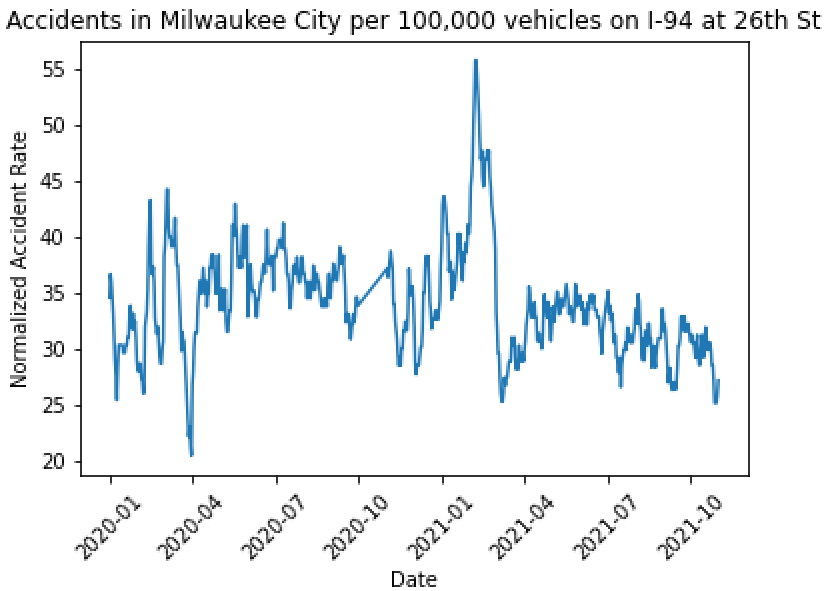

Figure 4: Normalized Accident Rate: Accidents in Milwaukee City per 100,000 vehicles on I-94 at 26th Street

To test for a correlation between the COVID-19 infection rate and the normalized accident rate, or a correlation between the change in COVID-19 infection rate and the normalized accident rate, I performed a linear regression using the Ordinary Least Squares (OLS) function in the Statsmodels[6] Python package and assessed statistical significance using the p-value.

Findings

Common Analysis

The exponential model fits to different intervals of infection rate data as discussed in the Method section had the following coefficients. The B coefficient governs the rate at which the exponential function increases.

Table 1: Coefficients of exponential model fits to the Milwaukee Co. infection rate curve

Wave	Mandate?	B coefficient value
1 (April 2020)	No	0.032
2 (June 2020)	No	0.038
3 (Sept 2020)	Yes	0.038
4 (March 2021)	Yes	0.013

[6] https://www.statsmodels.org/stable/index.html

We find that post-mandate Wave 3's coefficient is practically the same as the pre-mandate Waves 1 and 2, but Wave 4's coefficient is significantly lower than the rest. However, with only two waves before and after the mask mandate, and with one of the post-mandate waves (3) almost certainly driven by non-mask factors, it is hard to draw any conclusions about the effect of mask mandates. This method of analysis is also somewhat sensitive to the exact interval boundaries used for selecting points for fitting, as that can affect the resulting B coefficient value significantly.

Extension Analysis

The OLS regression did not find a statistically significant correlation between the infection rate and the normalized accident rate. The interpretation of this coefficient is that given two separate days that have infection rates that differ by 1 percentage point (about 10,000 people, based on Milwaukee County pop ~ 1M) the day with the higher infection rate will see about 0.5 fewer car accidents in Milwaukee City per 100,000 vehicles counted on I-94. However, within the 95% confidence interval, this estimate could vary from 1.2 fewer car accidents to 0.2 *more* car accidents, hence the lack of statistical significance. As the p-value exceeds 0.05, we are unable to reject the null hypothesis that these variables are uncorrelated.

Table 2: Coefficient of linear model fit to the infection rate - normalized accident rate relation

coef	Std err	t	P>|t|	95% CI (low)	95% CI (high)
-49.6335	35.716	-1.390	0.165	-119.776	20.509

Figure 5 shows a scatter plot of the infection rate against the normalized accident rate, with the regression trend line superimposed on the plot in red.

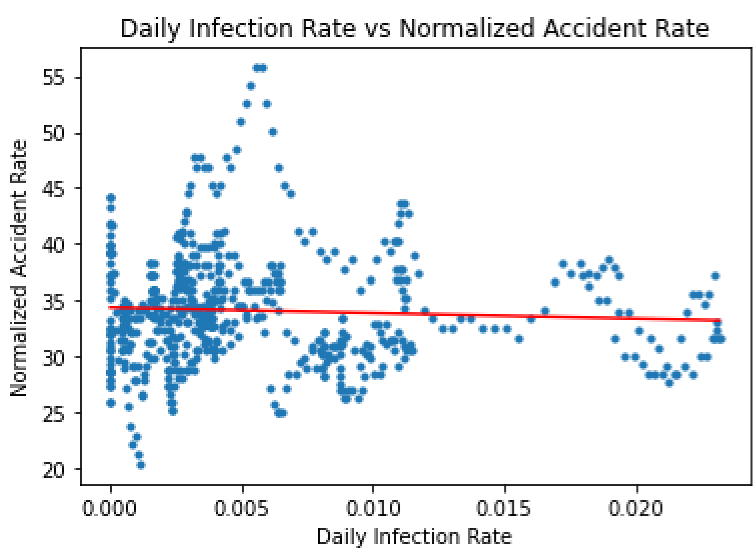

Figure 5: Plot of Infection Rate against Normalized Accident Rate. Each data point represents a day between January 2020 and October 2021. The regression trend line is shown in red.

The OLS regression found a statistically significant correlation between the daily change in infection rate and the normalized accident rate. The coefficient can be interpreted as a day to day increase in the infection rate of 0.01 percentage points (~100 new cases) being associated with 0.7 fewer car accidents in Milwaukee City per 100,000 cars on I-94. Within the 95% confidence interval, this could range from 0.9 to 0.4 fewer accidents. With a p-value below 0.05, we can reject the null hypothesis and say that these variables are correlated.

Table 3: Coefficient of linear model fit to the Δ infection rate - normalized accident rate relation

coef	Std err	t	P>|t|	95% CI (low)	95% CI (high)
-6730.8037	1289.138	-5.221	<0.001	-9262.499	-4199.108

Figure 6 shows a scatter plot of the daily change in infection rate against the normalized accident rate, with the regression trend line superimposed on the plot in red.

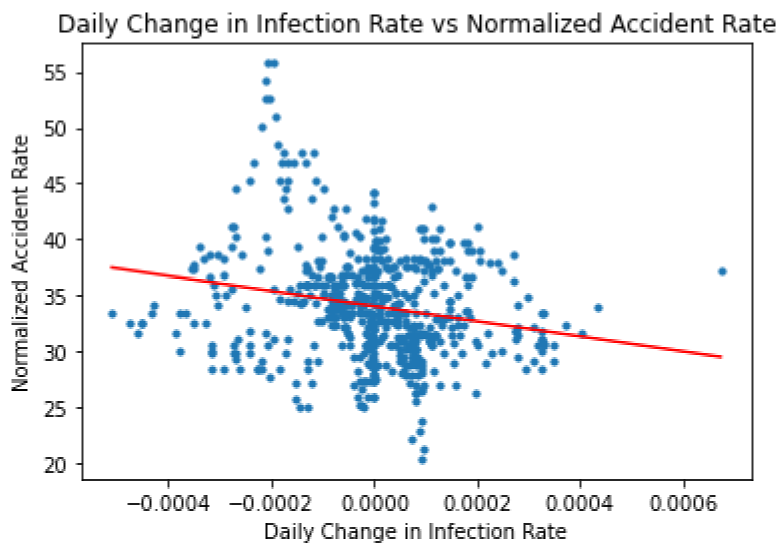

Figure 5: Plot of Daily Change in Infection Rate against Normalized Accident Rate. Each data point represents a day between January 2020 and October 2021. The regression trend line is shown in red.

Discussion

Common Analysis

The result of the common analysis portion of this project was inconclusive. Due to the limited data available and my own lack of experience with epidemiological modeling, I was unable to take advantage of higher-fidelity modeling methods, such as time series predictions, that might have been better suited for answering the question of mask mandate efficacy. The method I ultimately chose, fitting exponential models to the data, is only loosely grounded in epidemiological theory and did not provide enough information to comment on the effect of the mask mandate either way.

Ironically, this part of the project actually runs contrary to the idea of human centered data science in some respects. With attitudes towards public health measures like mask mandates increasingly a question of partisanship, a more thorough accounting of the potential harms of asking students with no background in epidemiology to answer a crucial question like “do mask mandates work” should have been performed prior to designing and assigning this project. We have real world examples of amateur efforts to model the pandemic going poorly and causing harm to public health: early in the pandemic, former Council of Economic Advisers chair Kevin Hassett infamously predicted that the pandemic would end by May 2020 after fitting a stock Microsoft Excel cubic polynomial function to the available data. There is no reason to believe that the pandemic would exhibit cubic polynomial behavior. Based on this model, the Trump administration sidelined advice from public health experts and its own task force, who warned that lifting of pandemic restrictions would cause a substantial increase in the death toll[7]. Granted, no one in this class is advising the President, but it would be trivial for bad actors to point to analysis performed in this class to claim that “research by University of Washington graduate students shows that mask mandates are ineffective,” further adding to the abundant misinformation related to the pandemic.

Future work in this domain could explore trends in infection rates across different counties that did or did not implement mask mandates. As noted in the existing literature, the patchwork of state and local level mask mandates creates an ideal natural experiment for evaluating the efficacy of these policies. This analysis would ideally incorporate input from subject matter experts, who could advise on the best way to accurately model the course of a pandemic.

Extension Analysis

The analysis found no statistically significant correlation between the absolute infection rate and the normalized accident rate, and a slight but statistically significant negative correlation between the change in infection rate and the normalized accident rate. It is difficult to explain the existence of this correlation. Saying that there is a direct causal relationship would be irresponsible, but it is possible that the trend in infection rates indirectly affects driving ability by, for example, increasing stress during periods when cases are rising.

The traffic collision data used for this analysis was semi-anonymized - each record listed the timestamp, cross streets, and an incident number. While the incident number can be used to get further detail about the collision such as the police report, there is a small fee for each request. I discard the incident number in the process of cleaning and aggregating the data, so there are no concerns about protecting privacy. Further, the fee for pulling police reports represents a barrier to any attackers. At the same time, lacking demographic details on the people involved in

[7] Yglesias 2020.

each accident and whether any injuries or deaths occurred prevents us from carrying out a more detailed analysis that takes such factors into account.

There are numerous paths forward for future research in this domain. Milwaukee specifically has many attributes that are of interest for human centered data science, as it is one of the most segregated metropolitan areas in the country. Milwaukee's African-American residents predominantly reside in the city's North Side[8]. During the COVID-19 pandemic, this section of the city had the highest hospitalization rates[9]. It also appears to have a higher concentration of traffic deaths compared to other parts of the city[10]. The difference between the North Side and adjoining predominantly white areas such as suburban Wauwatosa and the East Side neighborhoods is striking. A slide from my PetchaKutcha presentation is shown in Figure 6, comparing maps that show the city's racial demographics, COVID-19 hospitalization rates, and traffic fatalities. Unsurprisingly, they look similar.

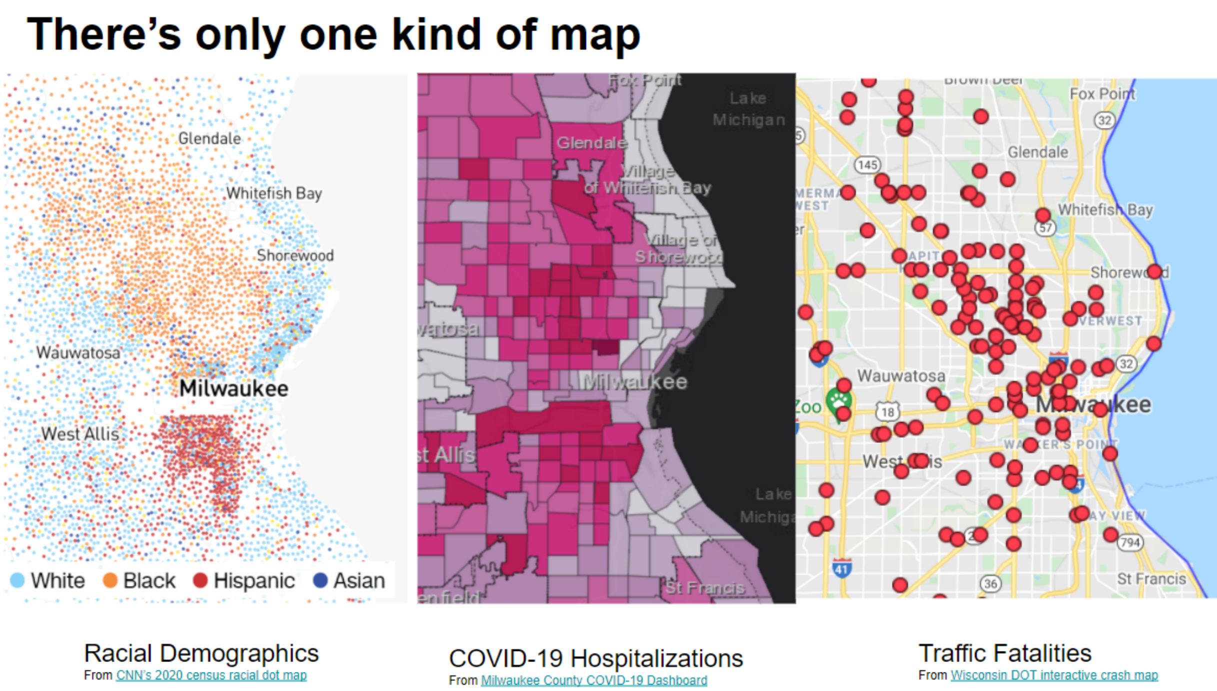

Figure 6: Comparison of Milwaukee's racial demographics, COVID-19 hospitalizations, and locations of traffic fatalities.

Were I more experienced with Geographic Information Systems (GIS) tools, it would be interesting to reproduce the work done by Lin et al. in studying changes to the geographic distribution of traffic collisions before and after the pandemic using the same dataset. The cross streets provided for each incident could be converted to coordinates and then assigned to a census tract, ZIP code, or aldermanic district to see whether the pandemic had the effect of shifting the burden of traffic collisions more towards minority districts, as Lin et al.[11] found in New

[8] Keefe et al 2021.
[9] Milwaukee County COVID-19 Dashboard. 2021.
[10] Community Maps - Wisconsin County TSC Crash Mapping. 2021.
[11] Lin et al. 2021.

York and Los Angeles. Similar work could be done if demographic information for the collision participants/victims were made available.

Limitations

Common Analysis

The common analysis fails to take into account other factors other than infection rate and mask mandates, such as lockdown/stay-at-home orders, superspreader events, vaccination, and the emergence of more contagious variants of the COVID-19 virus such as Delta and Omicron. By relying solely on the presence or absence of mask mandates, it ascribes too much influence to that single factor.

The assumptions regarding infection duration and contagiousness outlined in the Methods section are also a major limitation of the study. Although they greatly simplified the conversion from daily reported cases to infection rate, they fail to take into account person-to-person variability in infection effects and underreporting of cases.

Further, the analysis method selected that of exponential model fitting over selected intervals, was a crude and low-fidelity model chosen out of the necessity of not having sufficient domain expertise to select a more suitable model.

Extension Analysis

The lack of availability of desired data drove several close-enough substitutions. First, the traffic accident dataset does not provide any details on injuries or deaths. This is not in and of itself a problem, but the existing literature has shown that collisions, injuries, and deaths are not necessarily correlated. Focusing solely on collisions does not allow us to draw any conclusions about traffic injuries and/or deaths. The dataset also covers only Milwaukee City, which accounts for slightly more than 60% of the population of Milwaukee County. Next, the lack of sufficiently granular VMT data required me to substitute data from a single traffic counter, which also only issues reports on a monthly basis. While the trends in detected traffic volume and VMT are likely similar, using the traffic counter to normalize the accident count results in a much less interpretable parameter of “Accidents in Milwaukee City per 100,000 cars on I-94 at 26th St.”

I also encountered some issues with missing values. For example, the traffic counter report for October 2020 was missing. As a result, I simply discarded the accident data from that month because I did not have a strong basis for interpolating it. The traffic counter report for November 2020 did not compute a daily average for Sundays that month, and consequently did not compute an overall daily average. Not wanting to discard the entire month, I edited the spreadsheet containing the data to manually compute an average using the weekdays and double-counting Saturday.

Conclusion

The common analysis sought to study whether the implementation of a mask mandate in Milwaukee County affected the trajectory of COVID-19 infections there. Its results were inconclusive and did not make a strong case either way. Frankly, even if it did produce conclusive results, I would be somewhat uncomfortable about sharing them, given my lack of background in the epidemiology domain. Half-baked results using potentially sketchy methods have the potential to cause harm, especially on a needlessly controversial issue like this one.

The extension analysis sought to determine if a correlation existed between COVID-19 and traffic accidents in Milwaukee City. It found no correlation between the absolute infection rate and the normalized traffic accident rate, but found a slight negative correlation between the *rate of change* of the infection rate and the normalized traffic accident rate. Lacking a strong argument for the causal link between them beyond theories about the effect of stress on driving, it is possible that the correlation is entirely spurious. Given the observation that minorities in Milwaukee are both more likely to be hospitalized from COVID-19 and also live in neighborhoods with more traffic deaths, the extension analysis attempts to relate these issues.

References

Bolotnikova M. 2021. America's car crash epidemic. Vox. [accessed 2021 Dec 13]. https://www.vox.com/22675358/us-car-deaths-year-traffic-covid-pandemic.

Community Maps - Wisconsin County TSC Crash Mapping. 2021. [Accessed 2021 Dec 13]. https://transportal.cee.wisc.edu/partners/community-maps/crash/

Howard J, Huang A, Li Z, Tufekci Z, Zdimal V, van der Westhuizen H-M, von Delft A, Price A, Fridman L, Tang L-H, et al. 2021. An evidence review of face masks against COVID-19. Proceedings of the National Academy of Sciences - PNAS. 118(4):1. doi:10.1073/pnas.2014564118.

Keefe J, Wolfe D, Hernandez S. 2021. Race and ethnicity across the nation. CNN. [accessed 2021 Dec 13]. https://www.cnn.com/interactive/2021/us/census-race-ethnicity-map/

Lin L, Shi F, Li W. 2021. Assessing inequality, irregularity, and severity regarding road traffic safety during COVID-19. Scientific reports. 11(1):13147–13147. doi:10.1038/s41598-021-91392-z.

Lyu W, Wehby GL. 2020. Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US. Health affairs Web exclusive. 39(8):1419–1425. doi:10.1377/hlthaff.2020.00818.

Milwaukee County COVID-19 Dashboard. 2021. [accessed 2021 Dec 13]. https://mcoem.maps.arcgis.com/apps/dashboards/d1ed60c368b54cb7a20d22f3f4a4fd31

Yasin YJ, Grivna M, Abu-Zidan FM. 2021. Global impact of COVID-19 pandemic on road traffic collisions. World journal of emergency surgery. 16(1):1–51. doi:10.1186/s13017-021-00395-8.

Yglesias M. 2020 May 8. The Trump administration's "cubic model" of coronavirus deaths, explained. Vox. [accessed 2021 Dec 13]. https://www.vox.com/2020/5/8/21250641/kevin-hassett-cubic-model-smoothing.

Data Sources

Dataset	Source	License	Used In
COVID-19 Data Repository	Kaggle via Johns Hopkins University CSSE	CC BY 4.0	Common Extension
Masking Mandates	US Centers for Disease Control	US Public Domain	Common
Mask Compliance Survey	The New York Times/Dynata	details	Common
Traffic Accidents	City of Milwaukee	CC BY 4.0	Extension
Hourly Traffic Data	University of Wisconsin - Madison TOPS Lab	details	Extension