<u>CMSC 11900 : Final Report</u>

An analysis of the relationship between Covid-19 cases on ridership of different public transportations in Chicago.

Rhedintza Audryna

1 | Overview and Motivation

Due to being highly infectious, the COVID-19 pandemic has reduced transport and mobility all around the world as multiple countries issued lockdowns to curb the spread of the virus. This reduction in mobility will inevitably reduce the use of public transportations significantly. Furthermore, according to Mussellwhite et al, public transportations are high-risk environments for COVID-19 transmission due to the lack of physical distancing resulting from a high occupation density and enclosed spaces[1]. The danger of transmission is further exacerbated by the presence of multiple surfaces that can aid in the spread of the virus, given that the virus could remain active in plastics and metals for long periods of time[2].

In contrast, bicycles enable social distancing. Being a personal mode of transport, it allows people to travel while being 6 feet apart. With significantly lower risks of contagion, bicycles are safer alternatives to serve residents' needs for mobility in the current pandemic. Bikes as a mode of transport have seen rapid growth in the past couple of years, especially in urban areas around China, Europe, and North America[3]. They are typically deployed through a bike-sharing system, consisting of docking stations where people can rent bicycles for a short amount of time.

In Chicago, the first COVID-19 case was confirmed on March 5, 2020. This was superseded by a shelter-in-place order on March 18, 2020, followed by a stricter Stay-at-Home executive order on March 26, 2020[4]. These restrictions greatly reduced mobility in the city as offices start to implement work-from-home initiatives, multiple public areas got closed down, and residents start to get worried about the prospect of catching the virus. This reduction in mobility could potentially impact ridership trends for Chicago's public transportation and bike sharing options in different ways. Given that bikes are positioned as a safer transit option, I hypothesize that there would be a decreasing trend in public transportation ridership and an increasing trend in bike ridership. I sought to explore this hypothesis in this analysis.

2 | Background Information

A paper published in May 2020 by Teixiera and Lopes[5] discussed the impact of COVID-19 on New York City's transport system, specifically how the virus changed the use patterns of the city's subway and bike-sharing system. Just like Chicago, NYC has both a large public transport and an extensive bike-sharing network.

The paper ultimately concluded that NYC's bike sharing system suffered a significantly smaller ridership drop (71% vs 90%) compared to the subway and that some subway users shifted to utilize the bike-sharing system. This analysis was only done for the month of March, the start of the spread of the coronavirus in America, and it focused on one of the hardest-hit cities, NYC.

Thus, I am curious to see if the same trend occurred in Chicago, given that the city had a lower COVID-19 case count, and whether this trend persisted throughout the following months as the coronavirus continued to spread. It would also be more interesting to analyze this trend in Chicago because unlike the MTA, the CTA kept busses and trains running on a normal schedule throughout the pandemic to help prevent overcrowding[6]. Thus, supply remained mostly unchanged and reductions in ridership can be attributed to travelers' behavioral response to the pandemic.

3 | Research Question

This paper aims to identify the effect of COVID-19 on ridership trends for Chicago's main form of public transport, the CTA busses and L trains, and compare it to trends in the city's relatively large bike-sharing network, the Divvy

[1] https://www.sciencedirect.com/science/article/abs/pii/S2214140520300578?via%3Dihub
[2] https://www.nejm.org/doi/10.1056/NEJMc2004973
[3] https://www.tandfonline.com/doi/full/10.1080/01441647.2015.1033036
[4] https://www.chicago.gov/city/en/sites/covid-19/home/health-orders.html
[5] https://www.sciencedirect.com/science/article/pii/S2590198220300774
[6] https://www.roadsbridges.com/public-transit-qa-how-cta-has-operated-during-covid-19

bikes. In particular, I seek to answer the question of whether COVID-19 impacted ridership trends for the three modes of transit in Chicago differently, with a focus on the two big groups: public transport and bikes.

If differences exist between ridership trends, I would also like to explore if there occurred a shift in transit preference from crowded public transportations like busses and trains to more social distancing options like bikes, as a result of the need for safer transit options.

4 | Data Handling

The datasets that I used in this analysis include daily COVID-19 case data, as well as CTA bus and L ridership data from the City of Chicago data portal (https://data.cityofchicago.org/), Divvy ridership data from the Divvy system data (https://www.divvybikes.com/system-data), and average temperature data from the NCDC-NOAA (https://www.ncdc.noaa.gov/).

I gathered data for both 2019 and 2020 to compare differences in ridership trends across the two years, with 2019 as my control year. Furthermore, since the bus and L ridership data is only available until the end of September 2020, I decided to focus on the period between February – September for consistency. February is chosen as the starting month so that I could also observe ridership data slightly before the pandemic. Thus, I standardized and merged my datasets such that I had DataFrames that contain daily L, bus, and Divvy ridership data focusing only on the period spanning February to September in both 2019 and 2020 (Table 1).

date	rides_l	rides_bus	rides_divvy
2019-02-01	531924	677622	1790
2019-02-02	280328	432560	1905

Table 1 – Merged DataFrame for 2019

Since my datasets convey daily data, I also created DataFrames that groups daily data into weekly data for each year. This was to allow for more accurate results when I want to compare relative changes in ridership between the two years by preventing comparisons between a weekend and a weekday, which tends to have vastly different ridership.

5 | Exploratory Analysis

The first step in my analysis was to create a statistic that reflects relative ridership between the three modes of transport. Given that my focus is to compare ridership trends in conventional public transportations like busses and trains with the bike-sharing system, I decided to group bus and train rides together and created a ridership ratio between non-bike transportations and bicycles (Eq. 1).

$$Ridership\ Ratio = \frac{Total\ non\ bike\ (bus + train)\ rides}{Divvy\ bike\ rides}$$

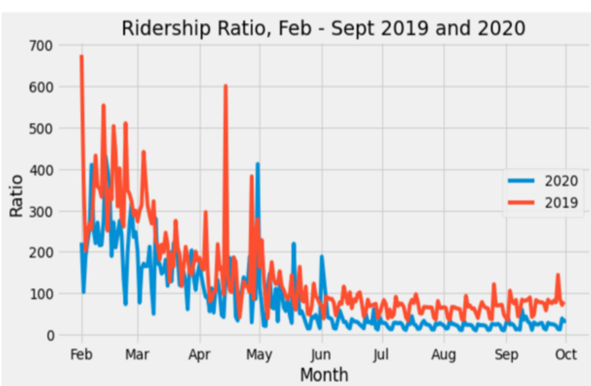

Figure 1 – Ridership Ratio in 2019 and 2020

A higher ridership ratio indicates a larger gap between bike and non-bike ridership and a lower ratio indicates a smaller gap. I compared the ridership ratio in 2019 and 2020 and saw that for the later months, the ridership ratio in 2020 is consistently lower than in 2019 (Figures 1 and 2).

As seen in Figure 1, the daily ridership ratio in February was relatively similar between the two years as indicated by the overlaps between the red and blue lines. However, in March, which was when COVID started to spread in the city, we can see a huge gap in the blue and red lines, followed by the blue line being consistently lower than the red, with the gap being more pronounced in later months.

This lower 2020 ratio can be seen more clearly in the boxplot (Figure 2) which shows that ratios in 2020 have a lower minimum, smaller range, and significantly lower median (by 51) compared to ratios in 2019. This decrease in ridership ratio means that the difference between ridership in non-bike and bike transportations was narrower in 2020.

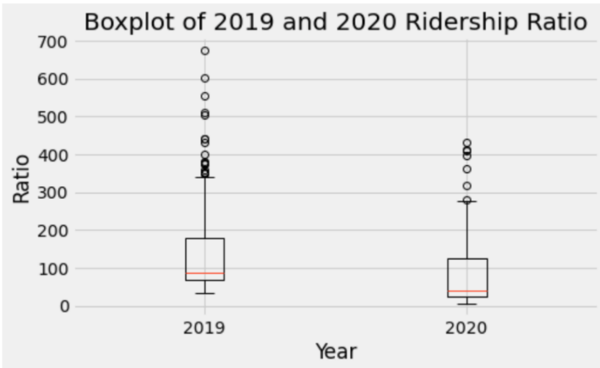

Figure 2 - Boxplot of RR in 2019 and 2020

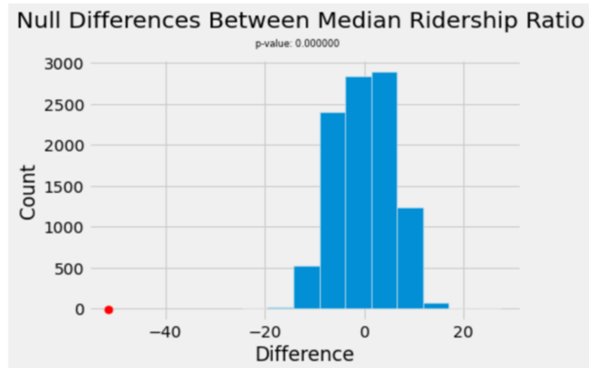

Figure 3 – Null difference between median RR in 2020 and 2019

To identify if this lower 2020 ridership ratio is statistically significant and not due to chance, I performed an A/B test. My null hypothesis is that the median ridership ratio in 2019 and 2020 are equal, and my alternative hypothesis is that the median ridership ratio in 2020 is *less than* in 2019. The test statistic used is the difference between the medians. The median is used instead of the mean given that it is a more robust statistic when we have outliers, which my dataset has in the form of really high ratios during the 2019 polar vortex period when biking was impossible. I simulated 10,000 permutations and the results were displayed as a histogram (Figure 3), providing visualization for the observed difference compared to the null differences. Here, the red dot represents the observed difference.

The result indicates that the p-value of having a median ridership ratio that is lower by 51 in 2020 compared to 2019 is close to 0, meaning that this result is statistically significant at more than a 99% confidence level. Assuming that the only fundamental difference between the two years is the presence of COVID-19, the results suggest that this narrowing of ridership gap between non-bike and bike transportations in Chicago is not due to chance and could potentially be associated with the pandemic. However, this assumption might not be entirely accurate since factors such as weather and temperature might be *slightly* different between the two years.

Thus, to evaluate whether and how COVID-19 is associated with the lower 2020 ridership ratio, I decided to perform an Ordinary Least Squares regression. In the regression, the daily ridership ratio was considered the dependent variable while daily COVID-19 cases and average temperature data were considered the independent variables.

	coef	std err	t	P>|t|	[0.025	0.975]
const	331.4087	11.446	28.953	0.000	308.860	353.958
daily_cases	−0.0283	0.011	−2.597	0.010	−0.050	−0.007
TAVG	−3.9735	0.180	−22.129	0.000	−4.327	−3.620

Table 2 – Regression Table Between Daily Ridership Ratio vs COVID-19 cases and temperature

Based on Table 2, controlling for temperature, the coefficient of daily COVID-19 cases is -0.0283, with a p-value of 0.01. This coefficient indicates that a 1 case increase in daily COVID-19 cases is associated with a 0.0283 decrease in ridership ratio. This result is significant at a 99% confidence level and thus is not attributed to chance. I have also controlled for potentially major confounding effects by including temperature in the regression. Although there might be other confounding variables, given that people's choice of transport is mostly dependent on weather, doing this will lead to a significantly more accurate representation of COVID-19's association with ridership ratio.

The analysis above shows that COVID-19 did impact ridership trends in non-bike and bike transportations differently. This was reflected through a reduction in the ridership ratio, which indicates that the ridership gap between the two groups of transportation has been reduced. If the impact of COVID-19 were homogenous across the different transportations, which happens when we have the same magnitude of ridership drop across the different modes of transit, we would not be seeing this reduction in ridership ratio. Thus, it would be interesting to identify the cause of this reduction. Based on the equation, a lower ridership ratio could be attributed to two things.

First, assuming that both bike and non-bike ridership decreases in 2020 compared to 2019, a lower ratio indicates that bike ridership decreased less than non-bike ridership in the pandemic year. Another scenario would be that bike ridership increased while non-bike ridership decreased. Thus, to evaluate which of these scenarios apply in this situation, I decided to look at the relative change in ridership for the three modes of transportations between the two years. To achieve consistent comparisons, I used the weekly ridership data to prevent dividing weekend ridership data with weekday ridership data across the two years. Here, relative change refers to the percent change in weekly rides between the years 2019 and 2020 for each mode of transport.

In Figure 4[7], all three transportations experienced a decrease in ridership in mid-March, consistent with the timeline of the coronavirus. However, although all three transportations experienced a decrease in 2020 rides, the degree of ridership drop varied. Bus and train experienced a more severe drop, with a maximum drop of 82% and 91% respectively, and remained in the negative region even until October. Meanwhile, the drop in Divvy rides was not as drastic with a maximum drop of 72% and this trend was reversed by July where we can see a steady increase in ridership all the way to October. Over the period, week-by-week, Divvy bike's ridership reductions were mostly less severe than the other modes of transportation.

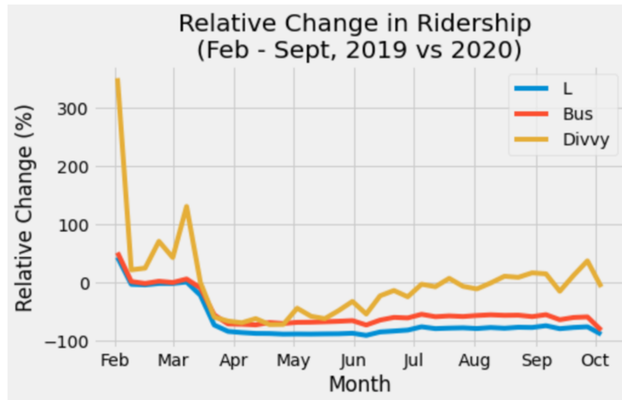

Figure 4 – Relative change in ridership for all three transportations

In other words, compared to busses or trains, bikes were more resilient in facing the mobility reduction caused by the pandemic. Moreover, Divvy bike ridership actually increased compared to 2019 from July to October. This provides slight evidence that some riders might have shifted from using busses and trains to Divvy bikes. Thus, the reduction in ridership ratio is attributed to *both* a smaller decrease as well as an increase in 2020 bike ridership.

6 | Discussion

This paper examined the effects of COVID-19 on the ridership of three modes of transit in Chicago and concludes that ridership trends were impacted differently by the pandemic. This is reflected by a reduction in ridership ratio, which indicates a narrowing of the non-bike and bike ridership gap. Through the A/B test, it was concluded that this reduction in ridership ratio between 2019 and 2020 was statistically significant. Through the regression, it can be inferred that COVID-19 is associated with this reduction and this result was also statistically significant.

The paper also explored the reasons for this decrease in ridership ratio and found that it was because the pandemic caused varying magnitudes of ridership drop for the different transportations, with the largest decrease in the L trains, followed by busses, and less severely, bikes. While both non-bike and bike transportations experienced plummeting ridership, the results show that bikes were more resilient in this era of social distancing. There is also some evidence of an outflow of riders from non-bike to bike transportations, especially in the later months.

These results demonstrate how Chicagoans adjusted their mobility behavior in the face of this pandemic. In particular, it highlights the importance of bikes as an alternative mode of transport especially in special situations such as the one we are experiencing right now. This information might be useful for the government of Chicago to develop policies and invest in certain infrastructures going forward, such as potentially increasing the number of bikes in the city and offering it at lower costs to make this safer transit option more affordable and accessible.

A future extension of this study could be to perform a meta-analysis of how COVID-19 has caused changes in bike ridership trends across other cities and countries, as this would give world governments some insights into the growing importance of bikes as a mode of transport going forward. Furthermore, identifying and including more confounding variables beyond temperature might also improve the accuracy of the OLS regression performed in this analysis regarding COVID-19's association with ridership ratio.

[7] Note that the huge positive difference in ridership at the beginning of February 2020 was due to the low bike ridership in early February 2019 caused by the polar vortex in Chicago