SEARCHING FOR NEARBY DRIVERS – A SPATIAL ANALYSIS OF CHICAGO RIDESHARE DATA

A Thesis

Presented to:

The Faculty of the Quantitative Methods in the Social Sciences Program

Graduate School of Arts and Sciences

Columbia University

In Partial Fulfilment

of the Requirements for the Degree of

Master of Arts

---

by

Julia Tache

December 2020

Advisor: Sharon Di

Table of Contents

Abstract

Transportation Network Providers (TNPs) like Uber, Lyft, and Via have exploded in popularity and usage in recent years. Urban citizens previously served by taxis now rely on ride-hailing companies to get them to where they need to go. Besides the fact that rideshare companies have crowded the markets once dominated by cabs through sometimes predatory pricing, the onset of these services has left many more lingering questions about equity and sustainability. Rideshare applications have the potential of providing cities with a convenient transportation option that may lead to less traffic congestion and greater access to travel and mobility. However, like taxi services and public transit, there is potential for Lyfts and Ubers to underserve certain disadvantaged communities due to income-level, racial and demographic make-up, distance from the city center, and other factors. Using 2019 open-source data on TNPs from the City of Chicago and census data, spatial and regression analysis finds that wealthier and whiter areas do receive slightly better performance by these companies than areas with higher poverty rates (lower fares, shorter trips, more options for pooling, and less cancelations). Maps, tables, and graphs of the data reveal how Chicago residents use these platforms, the most common ride patterns, and where and when drivers tend to do most of their drop-offs. The dataset provides several interesting directions for future research from understanding rideshare pricing to determining consumer demand based on neighborhood, date, and time of day to finding the ways pick-ups and drop-offs in different areas influence ride characteristics.

Introduction

It is a busy evening in San Francisco; I am traveling with a friend to a burger place on the other side of town. The Travelodge we are staying in is in the neighborhood Russian Hill, quite a few literal hills away from the nearest BART station. Our Lyft driver is talkative and enthusiastic about his work— after chattering to us about his love of driving, he asks us about our own traveling patterns and our opinions on transit. I let him know that I prefer using public transit but appreciate the ease of services like Lyft and Uber, and he asserts that ridesharing, in particular ride-pooling, is the future: a sustainable compromise that includes the convenience of being driven to a destination with the opportunity for fewer cars to be on the road. He is not alone in his thinking: many scholars have written about ridesharing as a viable option for mainstream travel, citing the opportunity to reduce traffic congestion and pollution while optimizing travel times.[1][2] The use of rideshares, or transportation network providers (TNPs), like Uber, Lyft, and Via has proliferated greatly in urban areas due to the ease of use of their platform and, in Uber's case, the ability to undermine competition with other transportation options through deliberately lower prices.[3] In a country enamored with cars and starved of funding for infrastructure and public transport, taking advantage of rideshares and incentivizing activities like pooling seem to be viable courses of action in terms of transportation policy.

However, quite a few questions have surfaced due to the fact that these aforementioned responsibilities are now outsourced to private companies: how effective are rideshares at

---

[1] Bajpai, Jitendra N. "Emerging Vehicle Technologies & the Search for Urban Mobility Solutions." *Urban, Planning and Transport Research* 4.1 (2016): 83-100.
[2] Hou, Yi, et al. "Factors Influencing Willingness to Pool in Ride-Hailing Trips." *Transportation Research Record* (2020): 0361198120915886.
[3] Hill, Steven. "Rideshaning Versus Public Transit: How Uber and Lyft Tend to Widen Disparities of Race and Class in Urban Transportation Systems." *The American Prospect*, Vol. 29, No. 2, Spring 2018, p. 46+. *Gale Academic OneFile*, https://link.gale.com/apps/doc/A537718733/AONE?u=columbiau&sid=AONE&xid=fb609dda. Accessed 6 Oct. 2020.

providing quality service? Do these companies benefit some groups more than others, and do they reinforce disparities between the wealthy and the poor, the advantaged and disadvantaged? What kind of places or neighborhoods do these companies typically serve? What are the implications for workers, users, and the economy at large if we continue to rely heavily on decentralized "gig" work, and can we trust these platforms to provide equitable service?

Literature Review and Concept Selection

Data on rideshares have been used in a variety of different ways in existing research. Up until 2020, the number of Uber trips were increasing year-by-year, showcasing the growing popularity of TNPs as a transportation option.[4] As the COVID-19 pandemic surged, the number of trips taken by Ubers and Lyfts went down, but the companies have tried to find creative ways to stay afloat and even provide philanthropic services. Lyft, for example, has expanded programs to deliver supplies and access to travel. The pandemic has revealed a great deal of inequities amongst communities in the United States, and Lyft has recognized that by making a promise to reach out to those communities and provide social goods. When all aspects of life, including transportation and access to hospitals, resources, food, water, etc., go back to being "normal," will these private companies be able to continue making a commitment to equity and quality service as more people will once again use their product?

The COVID-19 era is a very specific context, but before this time researchers analyzed how the presence of Lyft and Uber in urban areas affect characteristically unlike communities differently. Several critical publications have found that the services may in fact promote negative effects: ridesharing has shown to widen disparities of race and class in urban

[4] Iqbal, Mansoor. "Uber Revenue and Usage Statistics." *Business of Apps*. October 5th, 2020. https://www.businessofapps.com/data/uber-statistics/.

transportation systems in comparison to public transit;[5] a randomized study showed that drivers in Boston exhibit racial discrimination against black drivers;[6] and rideshare supply tends to be disproportionately concentrated in densely populated areas, leading to geographical supply distortions.[7] Theoretically, rideshares could fill in some of the "gaps" not included in other forms of transportation (public transport, taxis, etc.) by providing services to those in underserved communities that have historically had poor access to transportation. For instance, across the globe ridesharing apps specifically made for women have been created to provide a safe travel option while filling a market demand that had not been met previously.[8] Access to transportation means access to the city center where business and commerce are concentrated- areas of greater economic opportunity. For households who need to purchase everyday supplies but exist in food deserts or lack access to clean water, having a car or rideshare service might be their current best option.

An important factor that could influence where transportation network providers go and why is the structure of the city itself. If a city is extremely segregated, for instance, and the city "center" which typically includes commercial areas, many places of employment, financial centers, nightlife, and other attractions represents the wealthier part of town, then rideshare activity could be concentrated in these more popular areas. Those in poorer neighborhoods may experience negative impacts like longer wait times for a car, higher fares due to longer distances from the city hubs, and discrimination by Uber and Lyft drivers who refuse to enter certain neighborhoods. In New York City, a crude dated "joke" exists about the cab driver who never

[5] Hill, Steven, ibid.
[6] Ge, Yanbo, et al. "Racial and Gender Discrimination in Transportation Network Companies." No. w22776. National Bureau of Economic Research, 2016.
[7] Ghili, Soheil, and Vineet Kumar. "Spatial Distribution of Supply and the Role of Market Thickness: Theory and Evidence from Ride Sharing." 2020.
[8] RideGuru Team. "Rideshare Services for Women / Females." *RideGuru*, 24 Mar. 2017, http://ride.guru/content/newsroom/rideshare-services-for-women.

goes above 96th street (the arbitrary barrier between the predominately wealthy Upper West and East Sides and the historically black and Latinx neighborhood of Harlem). There is some truth to this unfortunately: taxi services outside of yellow cabs have been sustained in areas like the Bronx to provide service for those without access to official transportation channels.[9] Anyone who has walked down the streets of midtown, though, can attest to the dozens of bright golden cabs as well as countless Ubers, Lyfts, and other vehicles on a given block. How does the structure and demographic make-up of a city impact fare, trip length, and other ride measures, and are people in lower socioeconomic statuses more negatively affected by these trends?

Perhaps data will reveal these troubling observed realities to be true. For my own research, I used spatial and rideshare data from the year 2019 from Chicago's open data portal connected with socioeconomic and demographic data from the most recent U.S. Census. I will begin by looking at the areas most serviced by transportation network providers and use census data to compare areas with different levels of service via regressions and will illustrate these concentrations using maps. The data also include trips with fares of $0.00, which can be read as trips that were canceled by drivers, which poses the question: what are the reasons why drivers are canceling trips? Is it distance from where customers are, or are they purposely avoiding certain areas because of bad reputations associated with certain neighborhoods?[10]

Rideshares are often advertised positively as a convenient way to get from here-to-there for all and even as an eco-friendly alternative to driving alone/having your own car. Research on rideshare technology, network analysis, and user equilibrium hypothesize that car-pooling via

---

[9] Barron, James. "Where Yellow Cabs Didn't Go, Green Cabs Were Supposed to Thrive. Then Came Uber." *The New York Times*, The New York Times, 3 Sept. 2018, www.nytimes.com/2018/09/03/nyregion/green-cabs-yellow-uber.html.

[10] Rides may also be canceled because drivers are tired of waiting for passengers that take a long time coming to the car. However, the data below show that there does seem to be a bias in *where* riders tend to cancel rides. Negative ideas of certain areas may also feed into how they view passengers from these areas as well, possibly leading them to cancel more rides.

these on-demand services will become more popular as time goes on.[11][12] Taking a look at how people in Chicago are using transportation network providers and which groups of people are more likely to use the sharing/carpool options will get a sense of user patterns in the city. On-demand services offer the environmental benefits of fewer cars and less traffic congestion while also placing the onus on private companies to develop and enhance these services. This may unfortunately come at the expense of existing taxi drivers, who have already experienced a great deal of hardship with the advent of TNPs, and in favor over public investment in infrastructure and public transit. Urban studies-focused researchers conclude overall that ridesharing will become more and more common in the U.S., and along with considering environmental factors I believe it will also be crucial to investigate how underserved communities are being affected by ride share usage. Urban and suburban neighborhoods that have historically been disenfranchised and lacking in necessary transportation resources should be prioritized when figuring out how to incentivize increasing transportation services in certain areas.

Though some studies have concluded that TNPs are responsible for mostly negative consequences, other researchers have found less convincing evidence that rideshare service in cities differs disproportionately based on neighborhood. A quasi-randomized study completed in Seattle found that TNPs offer adequate performances in dense urban areas and that access was not limited to "wealthy white areas."[13] Using local regression and a researcher-developed regression model, Hughes et. al studied spatial and temporal patterns and found lower wait times in densely populated areas of Seattle and the lowest wait time around the midday hours.

---

[11] Bajpai, Jitendra N., ibid.
[12] Di, Xuan, et al. "A Link-Node Reformulation of Ridesharing User Equilibrium with Network Design." *Transportation Research Part B: Methodological* 112 (2018): 230-255.
[13] Hughes, Ryan, and Don MacKenzie. "Transportation Network Company Wait Times in Greater Seattle, and Relationship to Socioeconomic Indicators." *Journal of Transport Geography* 56 (2016): 36-44.

Expected waiting times were longer in census tracts with higher average incomes, but only by a small amount. One article analyzing a similar dataset I am using and found huge increases in rideshare service and scope of transportation options in Chicago over the past five years as well as popularity of the pooling option amongst those from lower income areas.[14] Below is an example of one of the maps the researchers created to illustrate their findings:

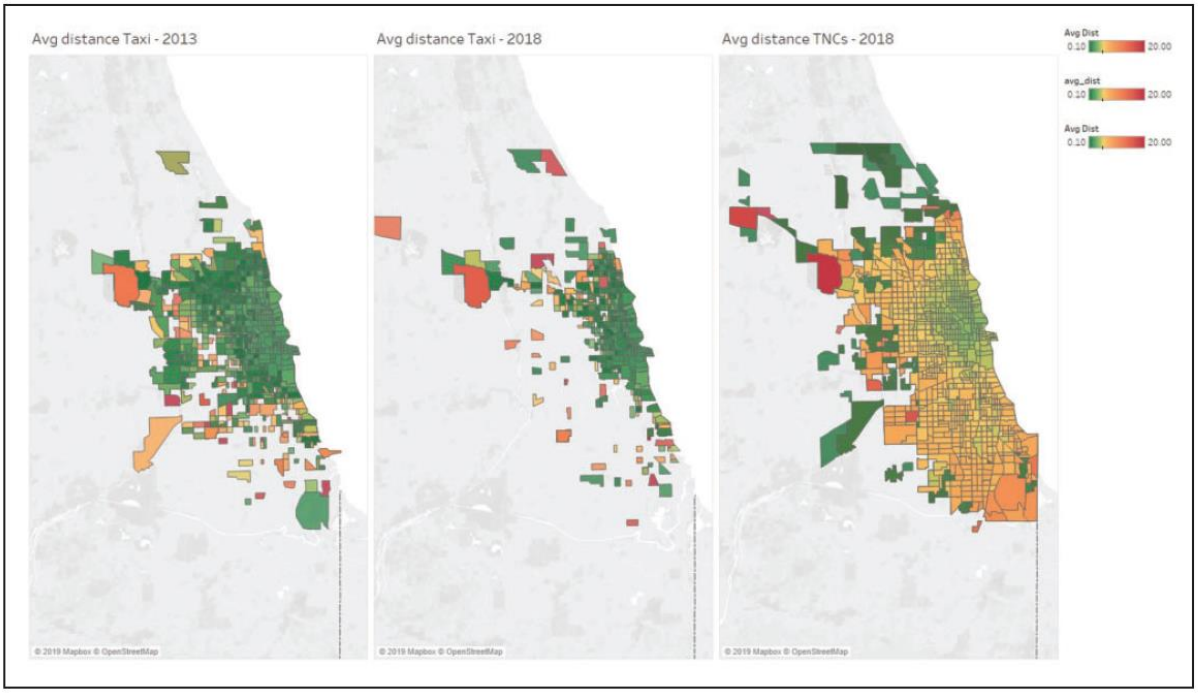

Figure 1: Average Distance Traveled by Taxis vs. TNCs (Roy et. al)*

**in their paper their refer to TNPs as TNCs, transportation network companies*

Chicago is a city whose population has stayed fairly constant in the past decade. Roy et. al found no major changes in the population, households, and employment characteristics of the city while the amount of ride-hailing trips quadrupled over the course of 5 years (2013-2018). Even though taxi service went down tremendously, the magnitude of trips increased when TNPs exploded onto the scene even though demographic characteristics stayed mostly constant. Their

[14] Roy, Sneha, Anurag Komanduri, and Kimon Proussaloglou. "Evolution of Transportation Network Companies and Taxis through 2013–2018 in Chicago." *Transportation Research Record* 2674.7 (2020): 385-397.

main strategy of their paper was to use t-statistic tests to determine the differences in fare, distance, and other ride characteristics between taxis and rideshare apps throughout Chicago. While their work is rich with detail, they list future avenues for their dataset including "studying average fare prices (per mile) for different zones and by different times of the day to understand the dynamic pricing structure of TNCs," which will be particularly important with regards to equity questions (Roy et. al, 395). If fares are disproportionately expensive in lower income areas, for instance, then there could be evidence of bias.

Taking a look at Uber/Lyft/Via driver biases that lead them to avoid certain areas due to reputation, crime rate, etc. will also be important, especially if existing data will be used to program routes and services. Unlike taxis which charge based on time and distance, pricing schemes for rideshare apps fluctuate based on pooling options, extra seating, area, and time of day ("surge" pricing). A 2019 analysis found that the price of shared rides in Chicago went up, leading to less people, especially those in lower income areas, to use the more environmentally friendly pooling option. When ride-hailing companies are taken to task over "congestion, driver treatment, and passenger rights," the bottom line may end up becoming more important than actually fairly serving communities, especially with increasing pressure for Uber and similar companies to finally turn a profit.[15]

Another problem that increasing reliance on rideshares poses is the issue of the structure of the platform itself and how it may be sensitive to algorithmic bias. Algorithmic bias is often due to training data which already include bias or preference used for modeling and software, and these training data may be gathered from user interactions. The gig economy is a huge area

---

[15] Bellon, Tina. "A New Chicago Ride-Hailing Law Reveals for the First Time What Uber and Lyft Really Charge." *Business Insider*, Business Insider, 26 Nov. 2019, www.businessinsider.com/ubers-carpool-pricing-strategy-revealed-by-chicago-fare-data-2019-11.

of interest within this topic and is full of opportunities to be affected by these biases. The implication of these biases is that they may cause societal inequities to be reflected in freelance technologies and platforms. Gig economy platforms and their users often rely on rating systems, such as the stars on Uber or Lyft, or longer reviews on Fiverr or TaskRabbit. This could lead in some cases to drivers seeking out specific clients in certain neighborhoods for better ratings and tips since they often also exist in a precariously employed state, leaving out groups of riders. If drivers are constantly, disproportionately, and unfairly avoiding certain areas, then the algorithm/gamified software that will lead them to their next ride will exacerbate these problems.

Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination[16] reveals that despite the fact that companies are prohibited from making employment decisions based on protected characteristics, reliance on potentially biased consumer ratings which feed into algorithms could lead to further inequalities. The writers of this paper bring up the existing legal protections for employment, how they fall short in the face of the ever-changing gig economy, suggestions on how to maintain fair legal practices, and the admission of the unfortunate lack of data in this area of research. Drivers with poor ratings for example are at risk of being banned from doing their jobs, affecting how they will complete their daily tasks for feedback. Just as drivers depend on positive user ratings which could lead them to shell out extra money and resources for upgraded rider experiences (water, phone chargers, etc.), rating systems also impact users as users with lower ratings are regularly denied service. Mix in these technological constraints with social determinants of mobility and accessibility and problematic results may arise.

[16] Rosenblat, Alex, et al. "Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination." *Policy & Internet* 9.3 (2017): 256-279.

Some articles I found provide their own literature reviews/overviews of existing findings of how socioeconomic characteristics interact with digital platforms,[17] while other sources break down helpful statistics on the platforms of interest.[18] Many articles describe the opportunities that rideshares provide as a valid transportation option that, with some modifications such as High Occupancy Toll (HOT) lanes and congestion taxes, can provide a sustainable alternative to single drivers without completely uprooting the existing systems.[19] Others are more weary of the effect that TNPs have had on existing disparities within cities as well as the problems with how rideshare companies operate. The earlier that some of these issues can be identified, the better that future technology can be crafted to make sure that those utilizing these tools are not impacted by or adding to existing inequalities.

Data Description and Methodology

I am curious to see how consumers of different socioeconomic levels are affected by this particular faction of the gig economy and whether the same inequities present in the "traditional" economy are replicated with regard to excluding marginalized populations in the delivery of services. With the gig economy continuously expanding, more individuals are relying on online platforms to seek services and employment. Although much of the language around the gig economy is about "flexibility," "freedom," "convenience," and "working on your own terms," do these seemingly silver-bullet services come at an expense? Could there be more "equality" in these open platforms, or do societal trends continue to perpetuate themselves, affecting those

[17] Muntaner, Carles. "Digital Platforms, Gig Economy, Precarious Employment, and the Invisible Hand of Social Class." *International Journal of Health Services* 48.4 (2018): 597-600.
[18] Clifford, Catherine. "Who Exactly Are Uber's Drivers?." *Entrepreneur*, January 22 (2015).
[19] Di, Xuan, et al, ibid.

looking for work and services? Is it possible that skewed training data provided by these services and used in their algorithms are themselves biased by users and affect how the app operates?

The Transportation Network Providers dataset provided by the City of Chicago includes a vast amount of information on this sector of the gig economy.[20] This dataset is collected as part of routine reporting by the city's local government and has data ranging from November 2018 to July 2020. The variables included are trip starts and end times, trip seconds, pick-up and drop-off census tracts, fare, tip, additional charges, trip seconds, and trip miles. There are many directions to go with this dataset from spatial analysis of rides to the relationship between drop-off location and fare. To begin, I look at the frequency of trips made to different neighborhoods of Chicago and the characteristics of the neighborhoods that rideshares are more likely to service. The units of my analysis are the rides themselves across the city. My variables of interest are the locations of rides based on drop-off data and demographic characteristics of these neighborhoods (found by using Census data through the API). My estimand is represented by:

$$\hat{E}(F \mid D = x) - \hat{E}(F \mid D = y)$$

Where "F" is the frequency of rideshares and D represents the demographic variables that account for higher and lower frequency of rideshares in certain areas. I employ ordinary least squares (OLS), logit, and local regression to test my hypotheses that areas disadvantaged by demographic factors and urban segregation will have worse performances from Uber and Lyft. To simulate how an app algorithm may lead to bias, I use data mining techniques to find the demographic predictors that are most significant to whether or not neighborhoods are being properly served by these companies. This is an especially important topic because a large body

[20] City of Chicago. "Transportation Network Providers (Ride-Hail Companies)." *City of Chicago: Transportation Network Providers (Ride-Hail Companies)*, 30 Oct. 2020, www.chicago.gov/city/en/depts/bacp/supp_info/transportation-network-providers.html.

of research suggests that people in less affluent areas are less likely to have adequate public transport or be serviced by taxis; could rideshares fill in these gaps, or are they simply replicating the same patterns?

For my methodology, I plan on first fitting a series of linear and non-linear regressions using *Tidyverse* tools on R.[21] The variables I want to look at are fare price and amount of rides as a function of geography, neighborhood demographics like race and median income, and poverty rates across the city of Chicago. I also, as mentioned earlier, am curious to see the characteristics of rides that are canceled (fare of $0.00) and use machine learning algorithms like logistic regression to try and model the frequency of rides in certain areas and to predict how inequalities in service may be replicated. Similar to Roy and Komanduri's paper, I will illustrate my data using choropleth maps through *Leaflet* to easily visualize the available data.[22]

The data and methods contain shortfalls that will not fully capture the reality of transportation networks in Chicago reliably or validly. First off, the original dataset itself is just under 112 million rows, so only the first 10 million rows were downloaded and from there one million rows were randomly sampled from the first subset. The data represent only a small portion of information from 2019 and may not be representative of rideshare activity in the city of Chicago or a large urban center in general. Secondly, the data exclude many direct pieces of information that could be very helpful in investigating bias, such as race, gender, income level, etc. of the user and driver. These factors are obviously protected and cannot be reported due to privacy concerns, if they are even recorded in the first place. The names of drivers and users for each ride and ratings are also not reported. Therefore, the analysis is mainly done through proxy

[21] Wickham et al. "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43), 1686, 2019, https://doi.org/10.21105/joss.01686.
[22] Agafonkin, Vladimir. *Leaflet Documentation*, OpenStreetMap, 2019, https://leafletjs.com/reference-versions.html.

by looking at the general characteristics of drop-off areas to see how ride variables like fare, trip seconds, pooling, etc. are impacted. The logic for this methodology is that through discovering variables with the largest relationships through regression, I can focus on those variables to build robust prediction models using the data and to create helpful maps and charts that will better highlight important trends in the data.

Results

Each row of the dataset compiled through joining census data with 2019 Chicago TNP data is represented by a ride and includes date and time stamps, drop-off and pick-up areas, and other unique identifiers. I am curious to see how different factors of rides (trip fare, cancellation, pooled rides, and trip length in seconds and minutes) are possibly affected by the demographics of the drop-off areas of each ride. My hypothesis is that poorer and less white areas will be disadvantaged by rideshare services by having higher fares, higher chances of rides being canceled (if the fare total is $0), more pooled rides to save money, and longer trips mainly due to the fact that poorer areas tend to be further away from where commercial behavior is concentrated.

Below are tables of descriptive statistics from the dataset. I divided up mean median income of drop-off locations from census data into five equal categories as factors: Low Income (1), Low-Medium Income (2), Medium Income (3), Medium-High Income (4), and High Income (5), with the mean amounts of each income level delineated below. Using these five categories, I created tables and graphs which present average fare, trip time in seconds, average pooled trips, and the average number of authorized shared trips within drop-off areas with these income levels. I also looked at the most common time of day for rides and the most and least popular dates in the data:

Table 1:

Income_Level	Mean Income
1	18449.8249027237
2	29410.9127182045
3	50233.1172413793
4	66843.2413793103
5	85697.90625

Figure 2: Income Level and Mean Income

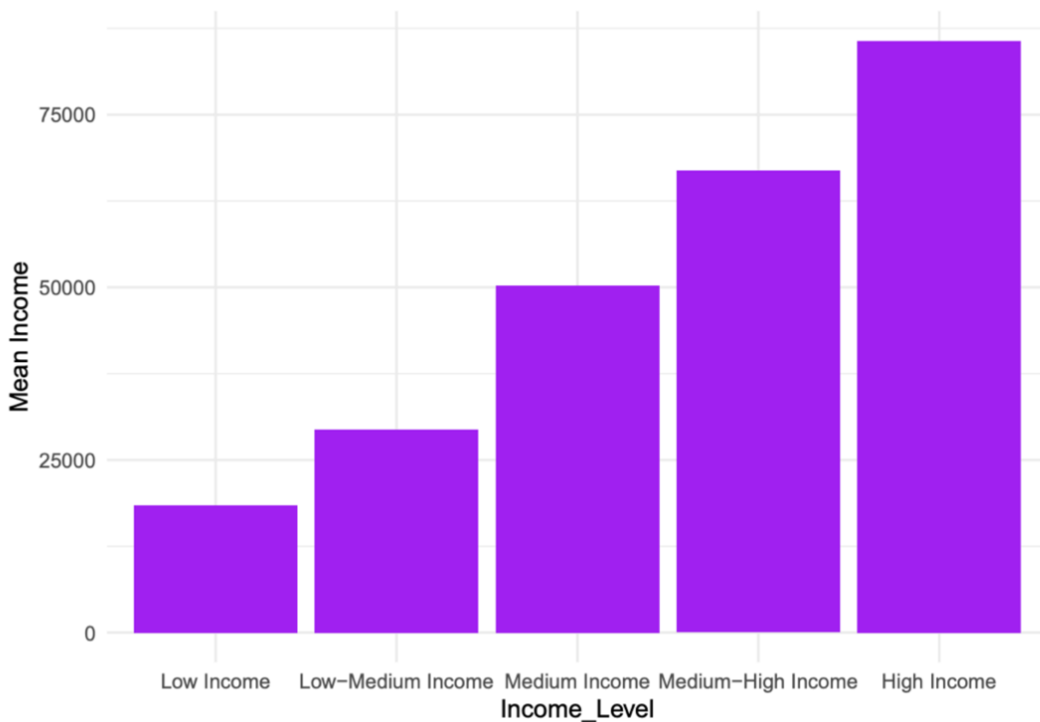

Figure 3: Income Level vs. Mean Income

Table 2:

Income_Level	Mean Fare
1	11.0702645028165
2	12.8796241461509
3	13.3824204274335
4	11.645567941716
5	13.3195139055362

Figure 4: Mean Fare by Income Level

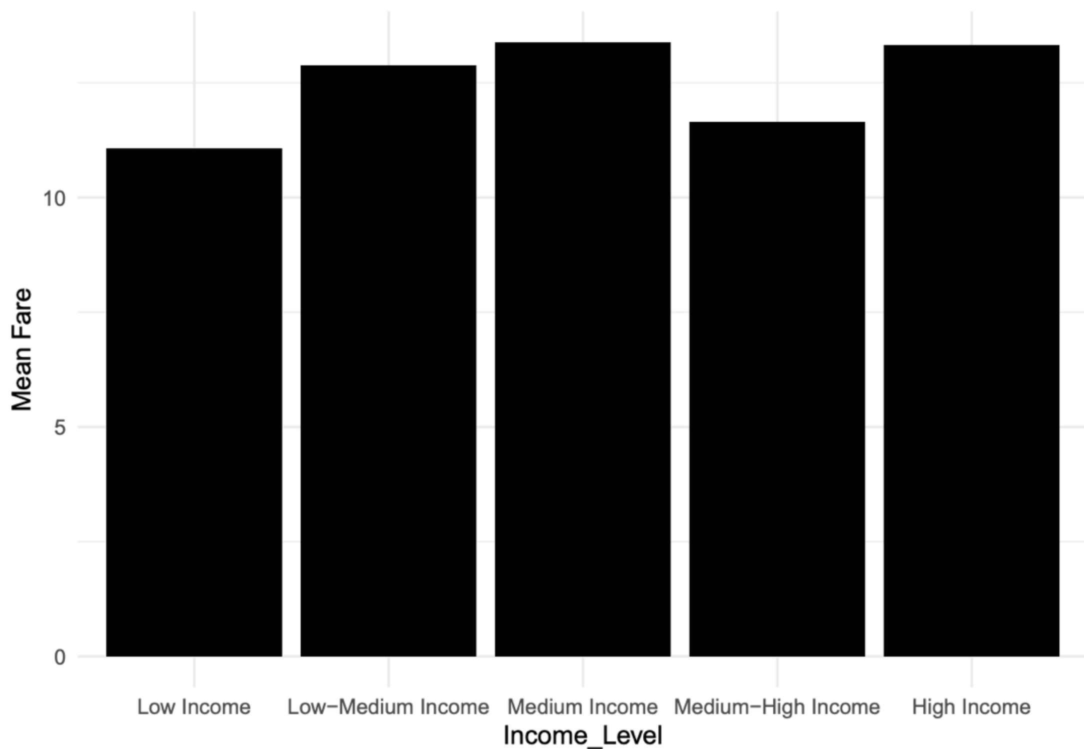

Figure 5: Income Level vs. Mean Fare

Table 3:

Income_Level	Avg. Trip Time in Seconds
1	1142.64894424526
2	1315.16442860796
3	1376.58117447556
4	1284.71829484234
5	1977.4375

Figure 6: Average Trip Time in Seconds

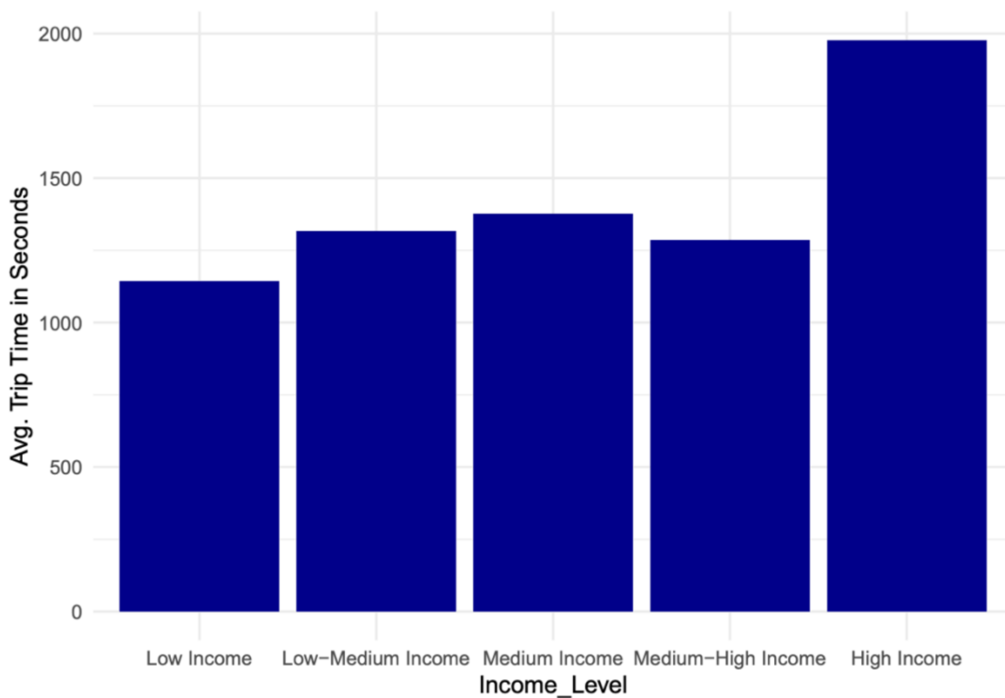

Figure 7: Income Level vs. Trip Time (in seconds)

Table 4:

Trip_Start	Number of Trips
Evening	247237
Night	210225
Early Afternoon	186464
Morning	185753
Late Night/Early Morning	85248
Late Morning	85073

Figure 8:Trips by Time of the Day

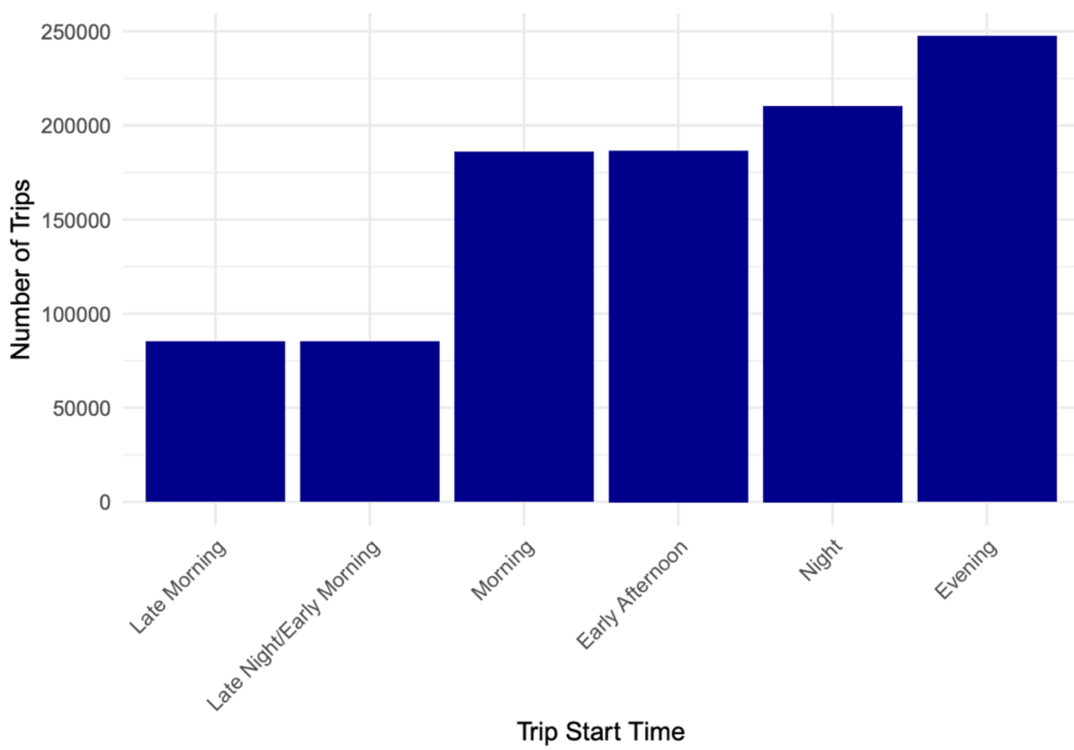

Figure 9: Trip Start Time vs. Number of Trips

Table 5:

date	Amount of Trips per Day
2019-03-16	4613
2019-10-26	4123
2019-12-14	3805
2019-12-07	3782
2019-04-27	3754
2019-02-23	3751
2019-03-09	3727
2019-05-18	3710
2019-12-13	3699
2019-07-20	3697

Figure 10: Most popular days in 2019

Table 6:

date	Amount of Trips per Day
2019-05-27	2017
2019-11-28	2006
2019-09-02	1984
2019-01-08	1977
2019-07-01	1970
2019-01-02	1883
2019-12-26	1696
2019-01-30	1596
2019-12-24	1528
2019-12-25	1009

Figure 11: Least popular days in 2018

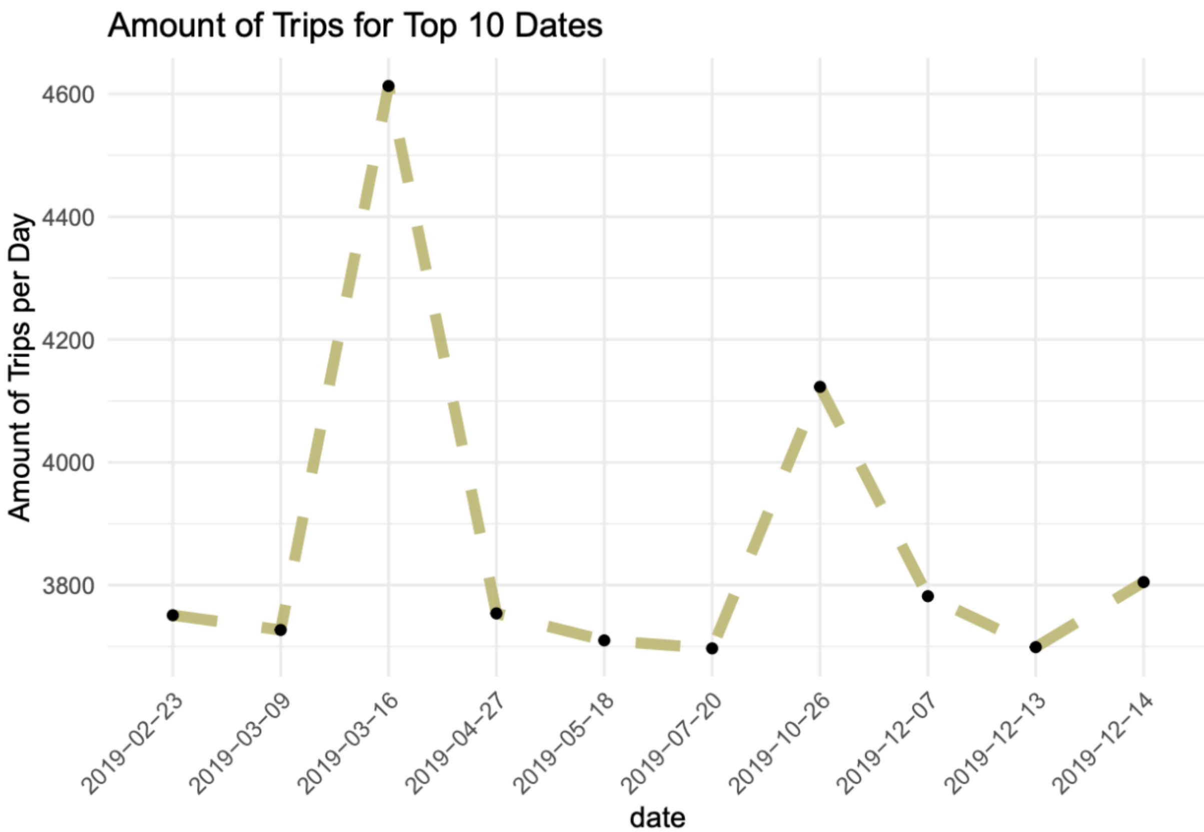

Figure 12: Trips on Top 10 Dates

Table 7:

Income_Level	Avg. Pooled Trips
1	50.8871595330739
2	44.2743142144638
3	88.1379310344828
4	252.965517241379
5	551.8125

Figure 13: Average Pooled Trips by Income Level

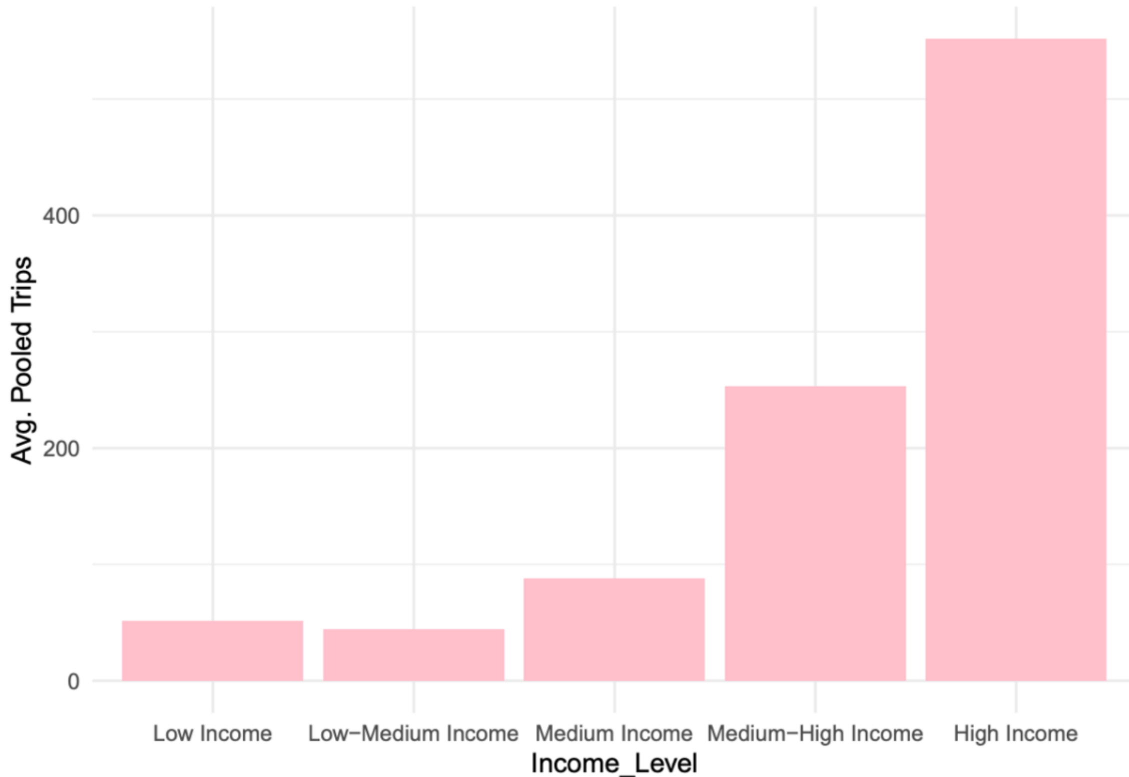

Figure 14: Avg. Pooled Trips

Table 8:

Income_Level	Avg. Num of Authorized Shared Trips
1	178.307392996109
2	168.483790523691
3	574.186206896552
4	2183.89655172414
5	5261.875

Figure 15: Shared Trips Authorized by Income Level

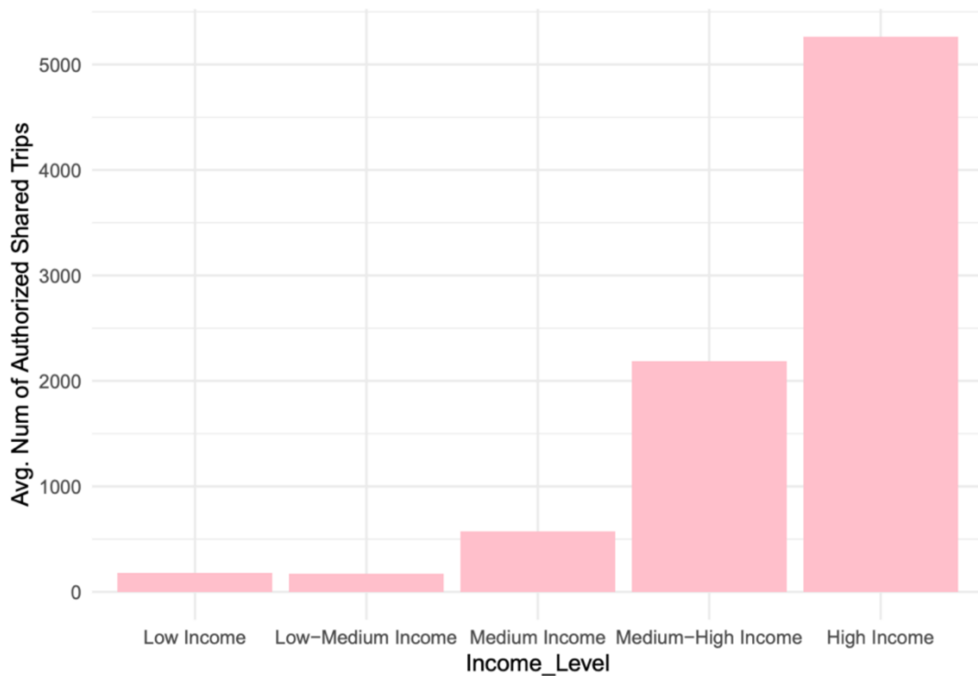

Figure 16: Avg. Shared Trips

Table 9:

Income_Level	Avg. Trip Miles
1	6.4561392544756
2	7.3816242280323
3	7.74663906088419
4	5.82178745657061
5	7.38516295041501

Figure 17: Average Trip Distances in Miles

Figure 18: Average Trip Miles by Income Level

Right away we can begin to see potential relationships and trends within the data, though some are stronger than others. The average fare between low-medium income and high income drop-off areas stays fairly consistent, with medium and high income areas having the highest fares on average (~$13.30) and low and medium-high income areas having lower fares (~$11) (Figures 4 and 5). Trips to higher income areas seem to take longer as shown in Figures 6 and 7, but there's a similar distribution of trip seconds as there were with fares where trip seconds goes *down* for medium-high areas then shoots back up again for high income areas. Despite shared trips and pooling being cheaper options, those going to medium-high and high income areas are more likely to have shared trips authorized and have more pooled trips (however, it is possible that many of these users are being picked-up from or reside in lower income areas). Most rides in the dataset occurred during the evenings and nights, which could also impact where people are going based on where they live, where nightlife is located, or whether certain neighborhoods are deemed as “unsafe” during certain hours of the day. Interestingly, the least popular days of the year for transportation network providers are winter holidays like Christmas and Thanksgiving. Most people, both riders and drivers, might be busy with family or friends during these festivities, so it may be worthy later investigating how this impacts fares during these “off” days. The most popular days seem to range across the year but mostly occur during the winter; the most popular day, March 16th, 2019, was the day right before St. Patrick’s Day, although St. Patrick’s Day itself did not make it to the top 10. The number of trips per day ranges from around 1,000 on the slowest days and above 4,000 on the busiest. Similar trends were also observed when looking at pick-up locations.

With the joined dataset, I ran some regressions to explore whether demographic elements of certain areas such as poverty rate, the percent of the population who is black, the percent of the population who is white, and an interaction term between poverty rate and percentage of black residents affect ride variables. For fare, trip length, trip length in seconds, and number of pooled trips, I used OLS regression on the given observations. For whether or not rides were canceled I used logistic regression because the data were presented as binary outcomes (coded as 1 for canceled, 0 for not canceled). The original hypotheses driving these regressions are that fares will be higher, trips will be longer, more trips will be pooled, and more rides would be canceled for lower income drop-off areas and neighborhoods with higher proportions of black and/or impoverished residents. As shown in the descriptive statistics, the trends presented in the data at face value do not particularly confirm and even contradict many of these claims, but the data have not been broken down into racial categories or poverty rate. Overall, the effects of the predictive independent variables on the dependent variables seem to be fairly small, but not insignificant (the strength of relationships are determined by reported p-values):

Table 10: Fare

	Dependent variable:	
	fare	
	(1)	(2)
‘Poverty Rate‘	0.033***	
	(0.002)	
‘Percent Black‘	0.020***	
	(0.001)	
‘Poverty Rate‘:‘Percent Black‘	-0.001***	
	(0.00004)	
‘Percent White‘		-0.016***
		(0.0005)
Constant	9.292***	10.859***
	(0.020)	(0.032)
Observations	666,644	666,644
R^2	0.002	0.002
Adjusted R^2	0.002	0.002
Residual Std. Error	7.273 (df = 666640)	7.272 (df = 666642)
F Statistic	368.505*** (df = 3; 666640)	1,199.406*** (df = 1; 666642)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 19: OLS Regression on Fare and Poverty Rate, Percent of Population who are Black, Percent of Population who are White, and an interaction term between Poverty Rate and Black Population Rate

Table 11: Trip Seconds

	Dependent variable:	
	trip_seconds	
	(1)	(2)
‘Poverty Rate‘	4.908***	
	(0.156)	
‘Percent Black‘	2.041***	
	(0.098)	
‘Poverty Rate‘:‘Percent Black‘	-0.059***	
	(0.003)	
‘Percent White‘		-2.189***
		(0.042)
Constant	890.520***	1,107.703***
	(1.866)	(2.971)
Observations	665,180	665,180
R^2	0.004	0.004
Adjusted R^2	0.004	0.004
Residual Std. Error	666.347 (df = 665176)	666.164 (df = 665178)
F Statistic	797.135*** (df = 3; 665176)	2,756.945*** (df = 1; 665178)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 20: OLS Regression on Trip Length (in seconds) and Poverty Rate, Percent of Population who are Black, Percent of Population who are White, and an interaction term between Poverty Rate and Black Population Rate

Table 12: Trips Pooled

	Dependent variable:	
	trips_pooled	
	(1)	(2)
‘Poverty Rate‘	0.006***	
	(0.0002)	
‘Percent Black‘	0.003***	
	(0.0001)	
‘Poverty Rate‘:‘Percent Black‘	-0.0001***	
	(0.00000)	
‘Percent White‘		-0.003***
		(0.00004)
Constant	1.131***	1.459***
	(0.002)	(0.003)
Observations	666,555	666,555
R^2	0.009	0.010
Adjusted R^2	0.009	0.010
Residual Std. Error	0.685 (df = 666551)	0.685 (df = 666553)
F Statistic	2,120.655*** (df = 3; 666551)	6,492.134*** (df = 1; 666553)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 21: OLS Regression on Pooled Trips and Poverty Rate, Percent of Population who are Black, Percent of Population who are White, and an interaction term between Poverty Rate and Black Population Rate

Table 13: Canceled Rides

	Dependent variable:	
	canceled_bool	
	(1)	(2)
‘Poverty Rate‘	0.006**	
	(0.002)	
‘Percent Black‘	0.005***	
	(0.001)	
‘Poverty Rate‘:‘Percent Black‘	-0.0001	
	(0.0001)	
‘Percent White‘		-0.004***
		(0.001)
Constant	-4.825***	-4.409***
	(0.030)	(0.043)
Observations	666,644	666,644
Log Likelihood	-33,906.934	-33,904.863
Akaike Inf. Crit.	67,821.868	67,813.727
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 22: Logistic Regression on Canceled Rides and Poverty Rate, Percent of Population who are Black, Percent of Population who are White, and an interaction term between Poverty Rate and Black Population Rate

Table 14: Mean Income vs. Mean Fare

	Dependent variable:	
	‘Mean Census Tract Income‘	
	(1)	(2)
‘Mean Trip Fare‘	592.226***	-504.139
	(103.231)	(382.346)
‘Number of Trips Pooled‘	28.404***	
	(2.353)	
I(‘Mean Trip Fare‘^2)		20.150**
		(8.323)
Constant	25,410.464***	37,851.556***
	(1,452.211)	(3,359.588)
Observations	922	922
R^2	0.148	0.019
Adjusted R^2	0.146	0.017
Residual Std. Error (df = 919)	17,256.764	18,515.570
F Statistic (df = 2; 919)	79.935***	9.080***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 23: Polynomial Regression on Mean Census Tract Income and Mean Trip Fare and Number of Trips Pooled

When broken down, the variables generally move in hypothesized directions: as shown in Table 10, fares tend to be higher in areas with greater poverty rates and increase with the proportion of the population who are black while they tend to be lower in areas with a greater portion of whites (oddly enough, the interaction term reveals a negative relationship, with a 1% increase in the interaction term between the black population and poverty rate leading to a -0.01% decrease in fare- the effect is small enough to be nontrivial). The largest impact on fare is poverty rate where a 1% increase leads to a roughly 3% increase in fare. The results in Table 14 are from a polynomial regression that shows that mean trip fare actually decreases at a certain point as income increases, but then increases past a certain level (Figure 27 maps out this trend). Regression analysis in Table 12 reveals that rates of pooled trips increase with poverty rate and the percentage of African-Americans in a given area while decreasing with the number of whites, despite more pooled trips happening in drop-off areas with higher incomes. Trying to predict canceled rides based on basic regression revealed small effects, with the most significant effect being a 1% increase in the poverty rate of the drop-off area leading to an estimated 0.6% increase in the chance that the ride will be canceled. Trips to areas with higher rates of black residents and higher rates of poverty also seem to take longer than trips to places with higher rates of white residents, as presented in Table 11.

I used *ggplot* in R to graph the relationships between trip fare, trip length in seconds, trips pooled, whether or not trips were canceled, and the poverty rate of the different areas serviced using box plots and graphs. Again, though the trends move in expected directions, they are still pretty small. I also did the same using aggregated data that looked at average values over income levels:

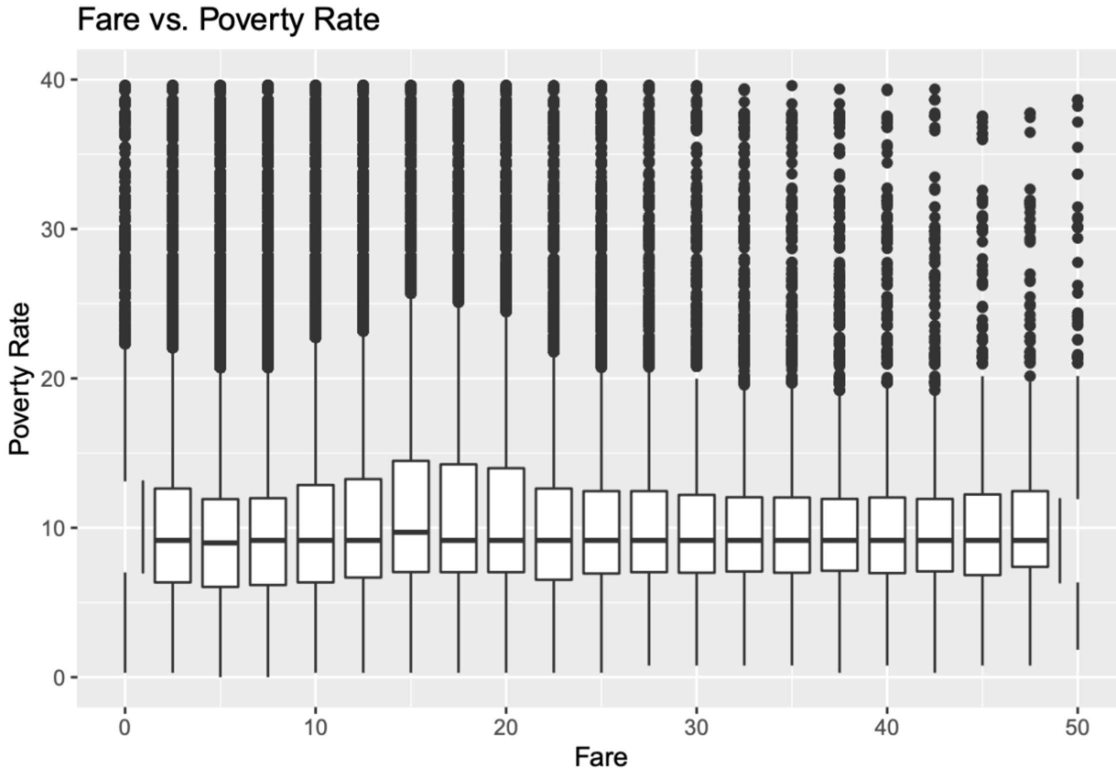

Figure 24: Boxplot of Fare vs. Poverty Rate. This figure shows a steady increase in fare as poverty rate increases up until around the $15 fare amount, but overall fewer rides are made to areas with high poverty rates.

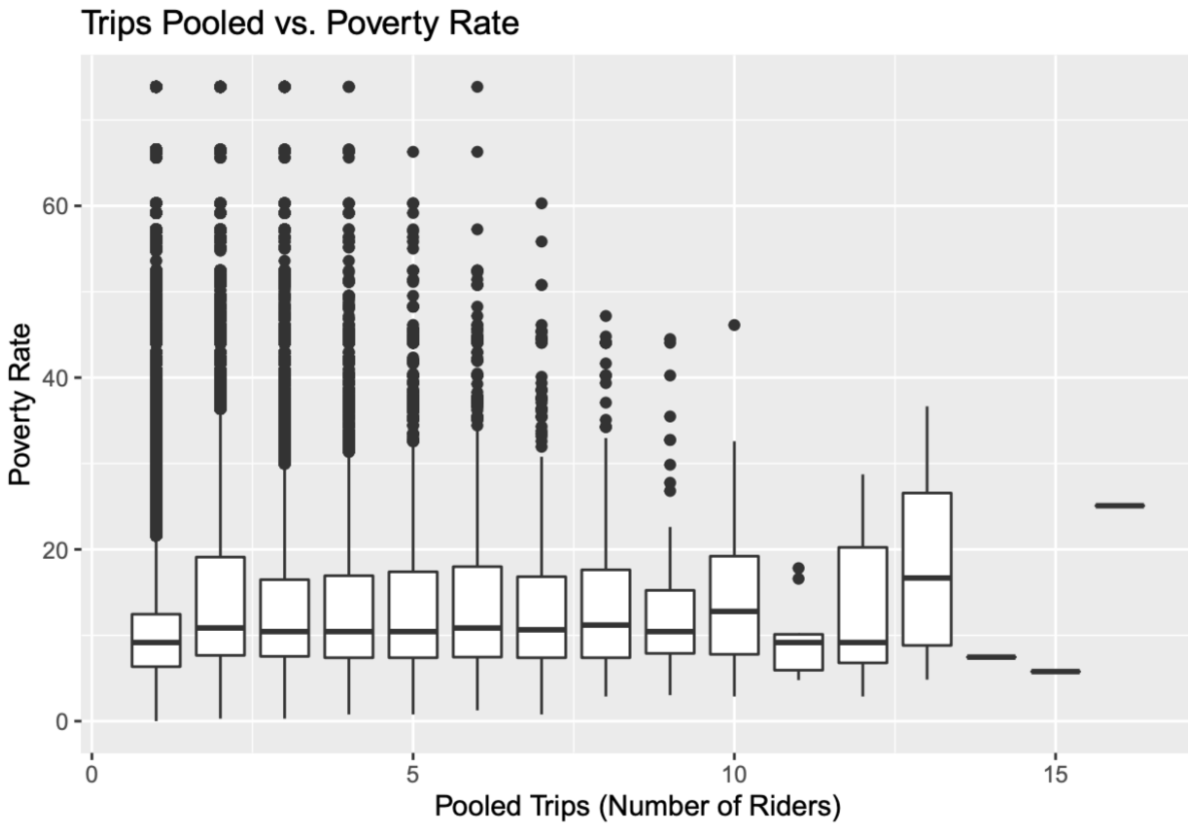

Figure 25: Box plot of Pooled Trips (Number of Riders) vs. Poverty Rate. The graph reveals most rides have fewer pooled riders and a slight increase in poverty rate as pooled trips increases.

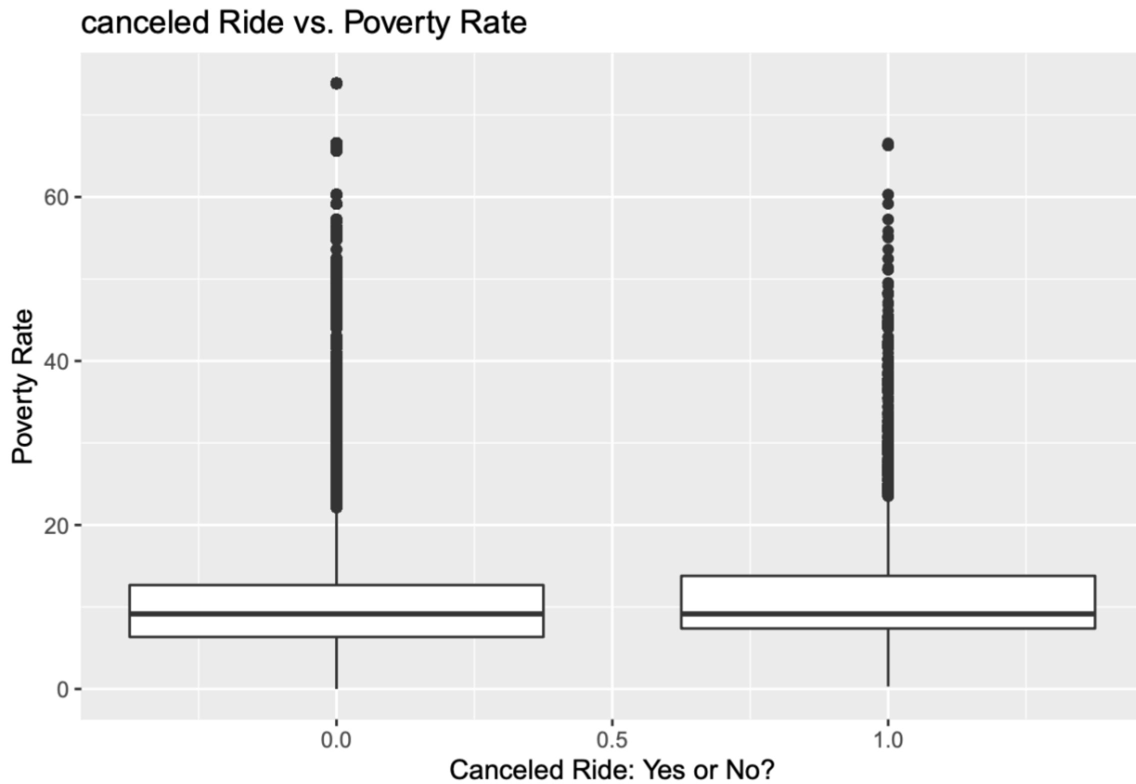

Figure 26: Boxplot of Canceled Rides vs. Poverty Rate. Canceled rides are set up as a binary value: if the trip fare = $0, then the ride is labeled as "canceled." Slightly more trips are canceled in areas with higher poverty rates.

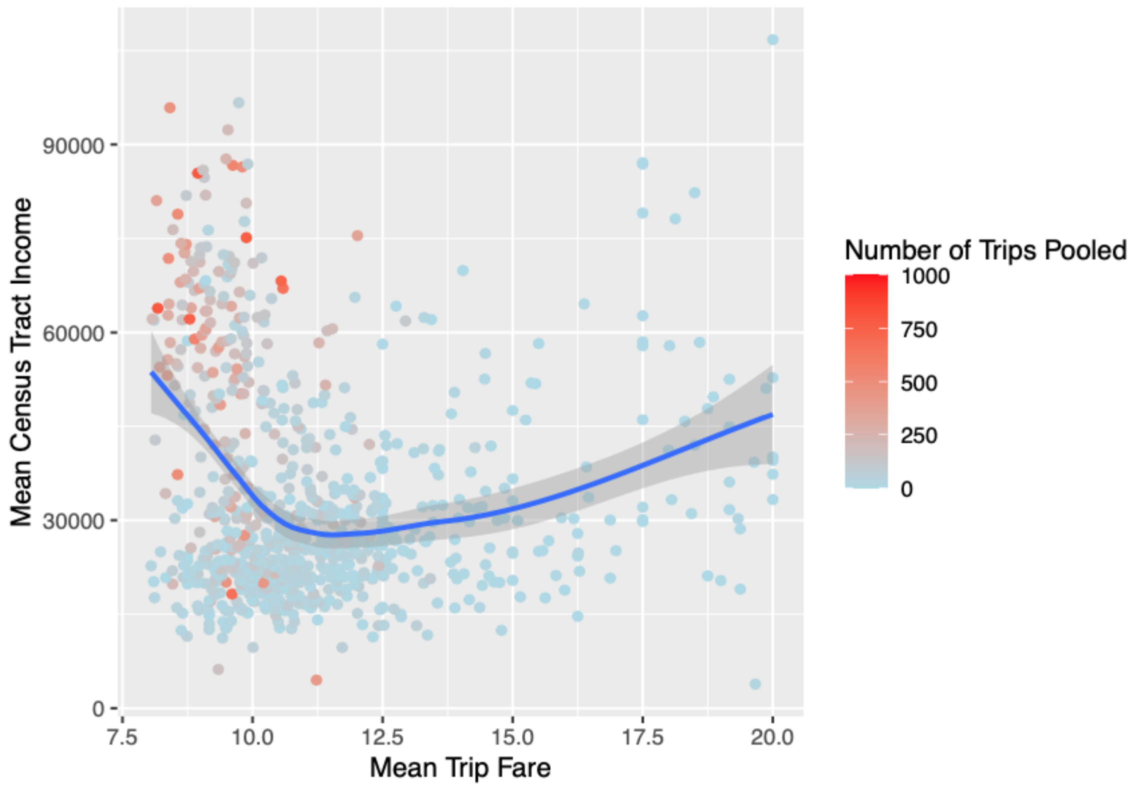

Figure 27: Line graph of Mean Trip Fare vs. Mean Census Tract Income with a U-shaped trend fit with local (LOESS) regression (decreasing census tract income means higher fares on average up until around the $11 mark, where trip fare begins to steadily increase with income). Points represent census tracts and are color-coded based on the number of pooled trips in the area. This graph also reveals that most trips fall within the $7.50-$12.50 range of fares.

What could lead to these trends? The fact that many rides are collected share similar values in terms of fare, income rate, etc. could reveal more about how people are using these rideshare apps and where they are going. Residents of Chicago seem to prefer taking short, cheap trips at later hours likely out of convenience to get to nightly activities rather than using the apps for longer journeys or to travel to work. Observed trends seem to be the strongest for trips within the $7.50 and $20 price points, which represents most of the trips in the data. As shown by previously reported means by income levels, trip values overall do not vary that greatly for the most part but may be sensitive to outliers.

The higher rates of drop-offs in higher income areas as well as longer trips and larger fares in areas with lower incomes could be mostly due to the structure of the city. The maps below illustrate how income rates, poverty rates, and population rates of black and white residents are spread across Chicago:

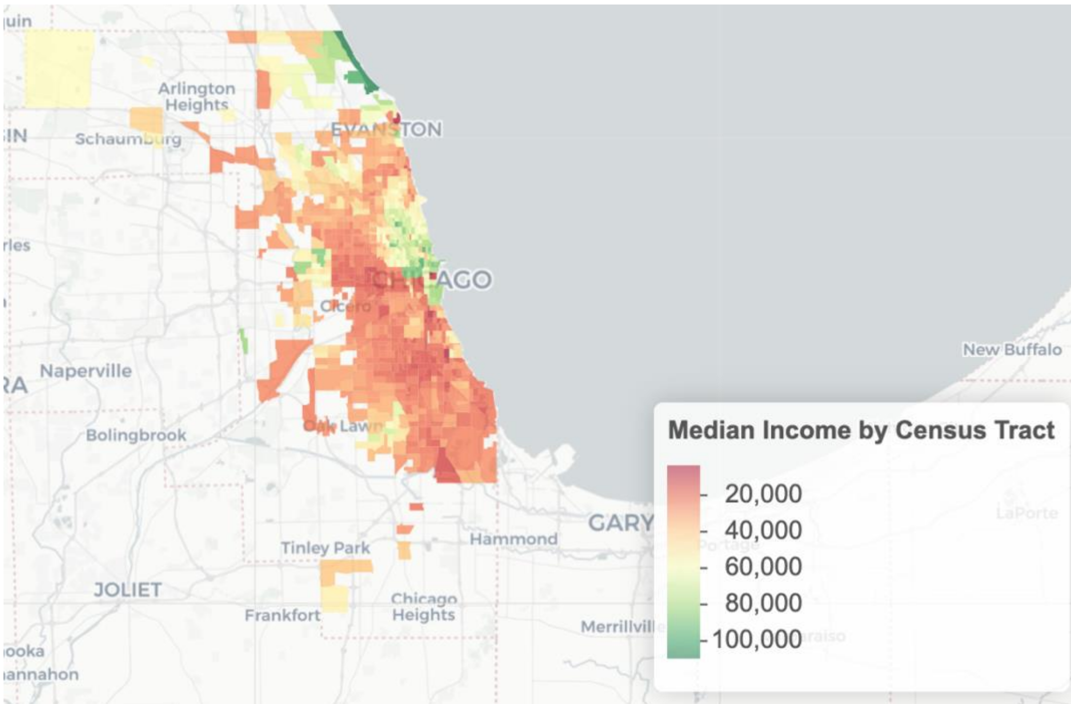

Figure 28: Median Income in the City of Chicago

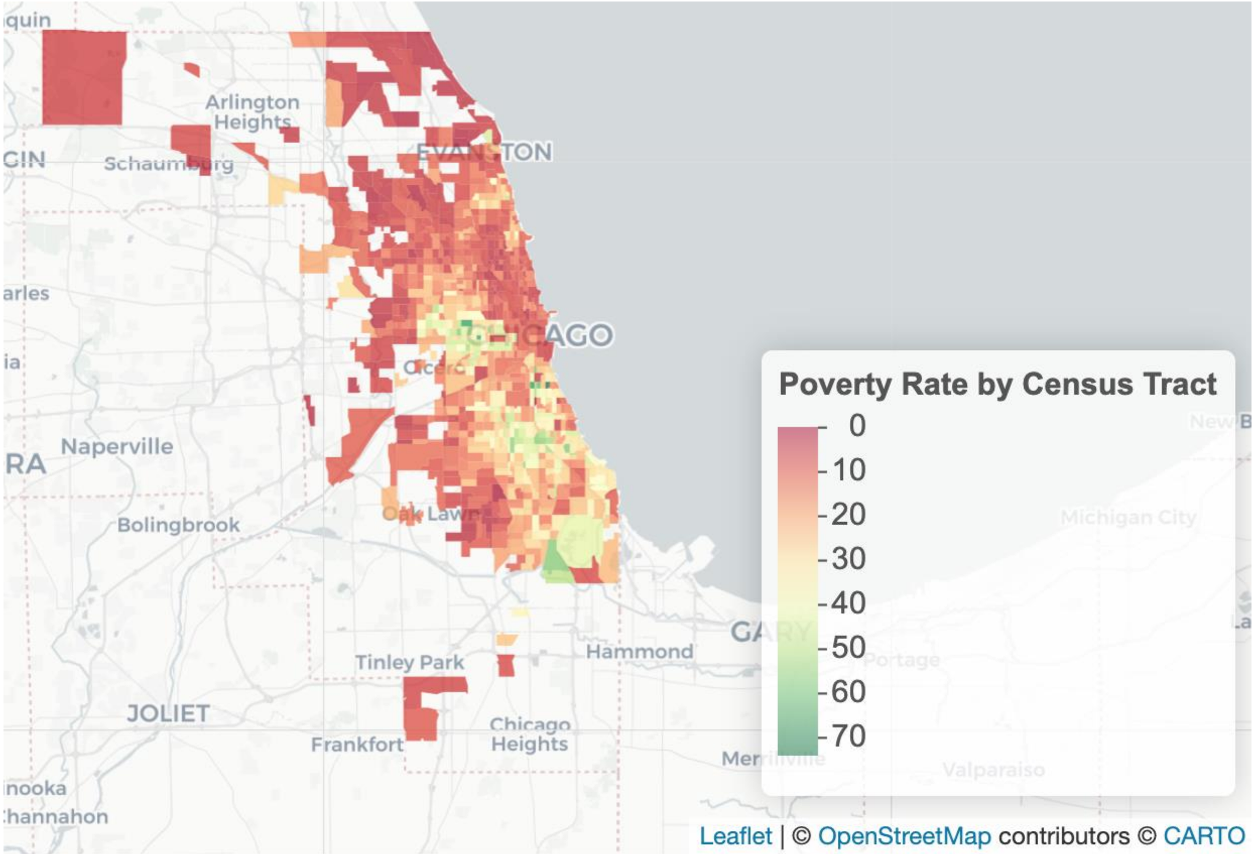

Figure 29: Poverty Rate in the City of Chicago

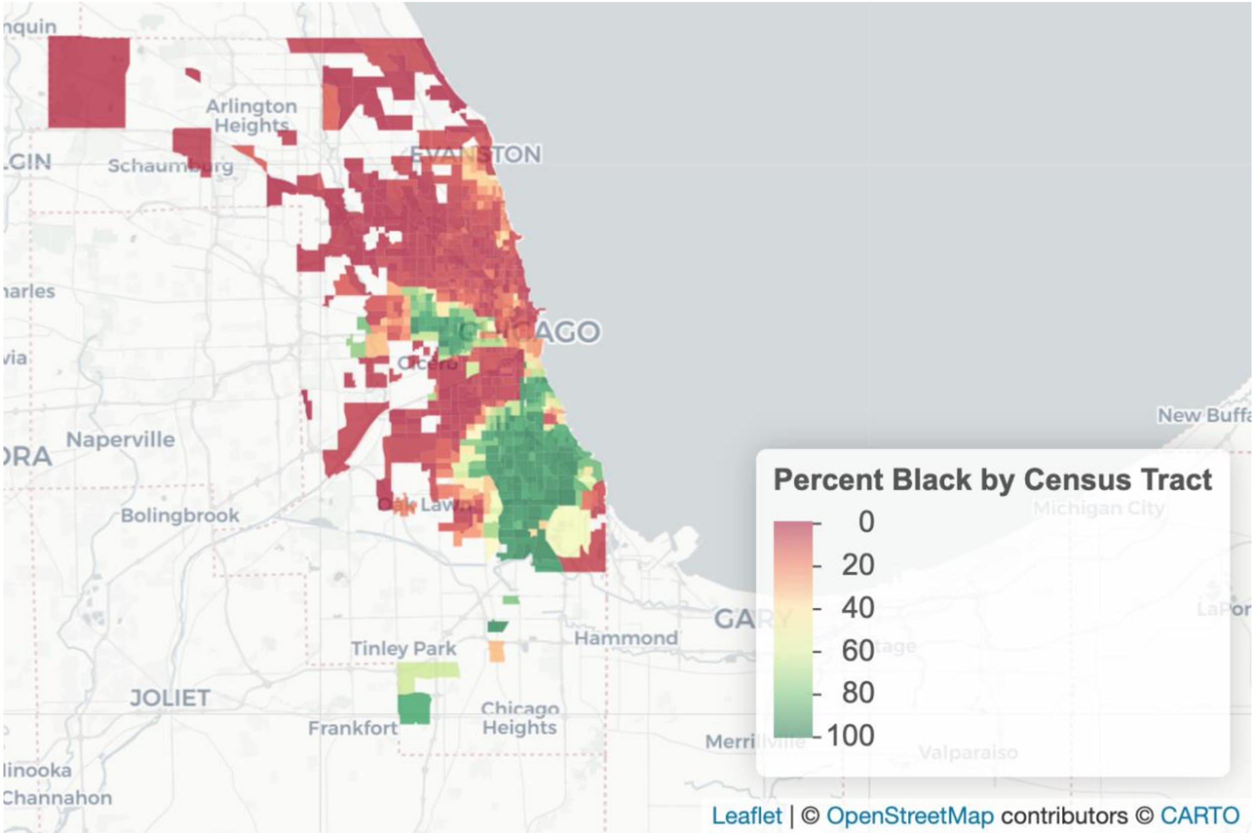

Figure 30: Percent of black residents by census tract in the City of Chicago

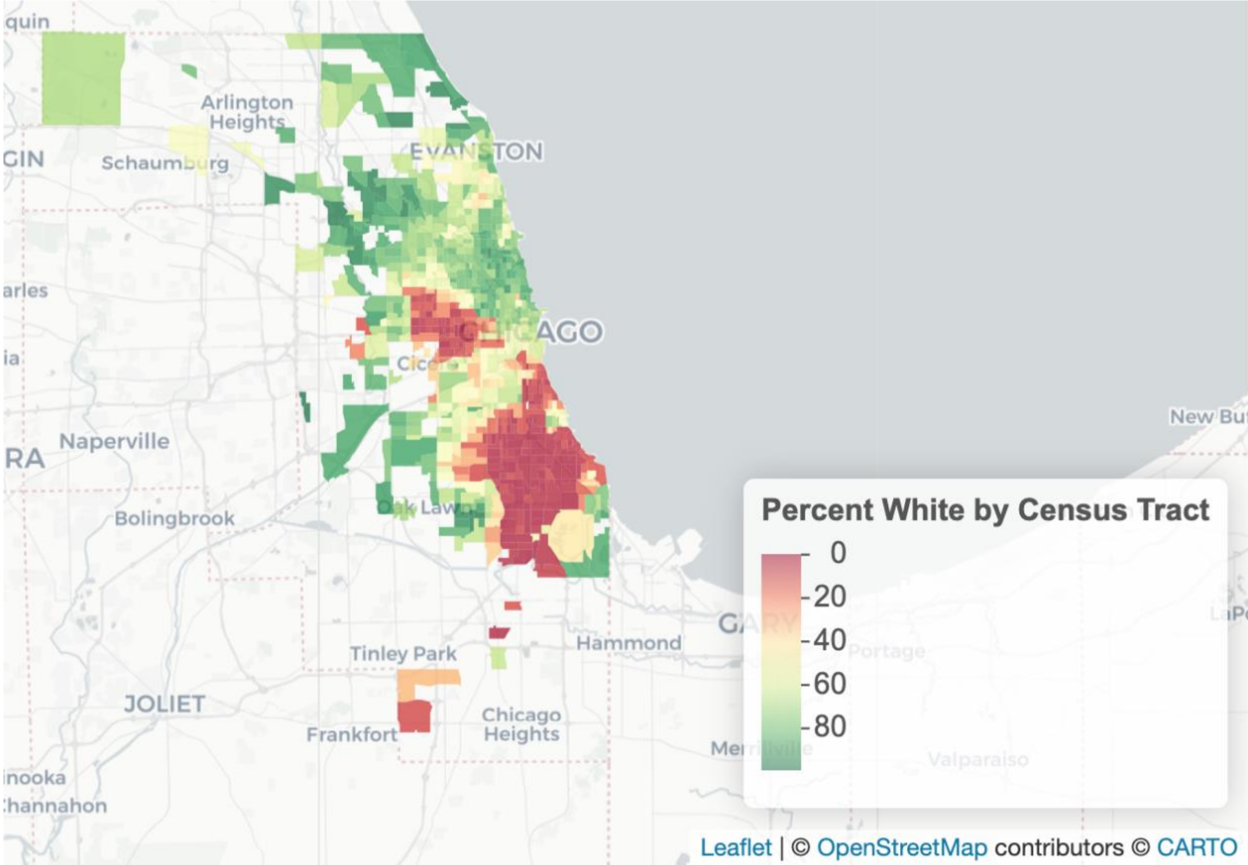

Figure 31: Percent of white residents by census tract in the City of Chicago

Like many cities, income and racial disparity can be obviously and visibly seen once mapped out. While south Chicago, except for a few neighborhoods, has median household incomes around $10,000-$30,000, a collection of affluent neighborhoods exists on the north-east side of the city right by Lake Michigan with median incomes exceeding $60,000 a year. This area of concentrated wealth contains neighborhoods such as Lincoln Park, Old Town, Logan Square, the "Loop" – the downtown area named literally for the fact that it literally exists as a bustling transportation crossroad, and attractions like Millennium Park, theatres, etc. I had hypothesized that the maps below of rideshare data would show a clear "heart" of activity, and while the "center" of activity is a bit ambiguous, the maps reveal that rideshare service is mostly clustered within higher income, whiter areas:

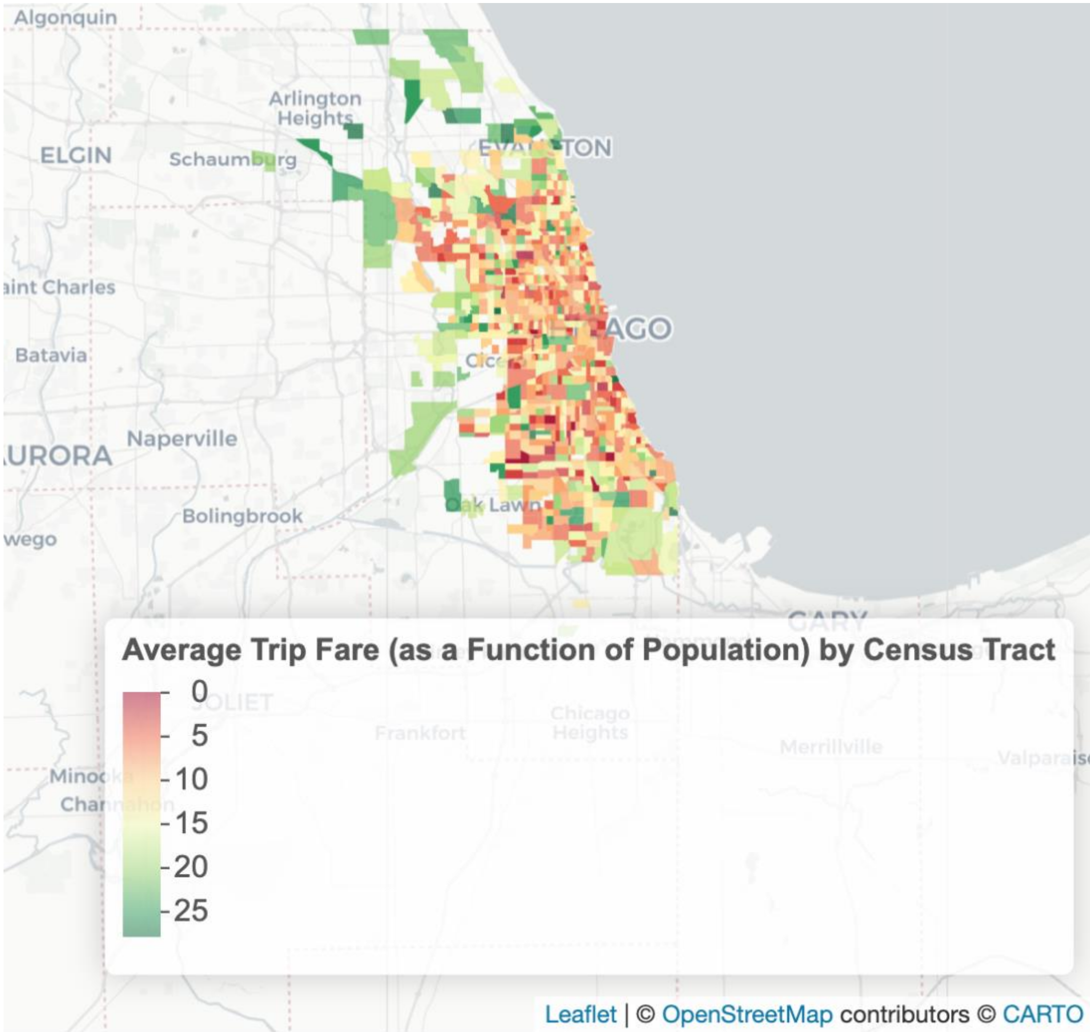

Figure 32: Average Trip Fare in the City of Chicago weighted by population. Each colored square represents the destination of the trip by census tract.

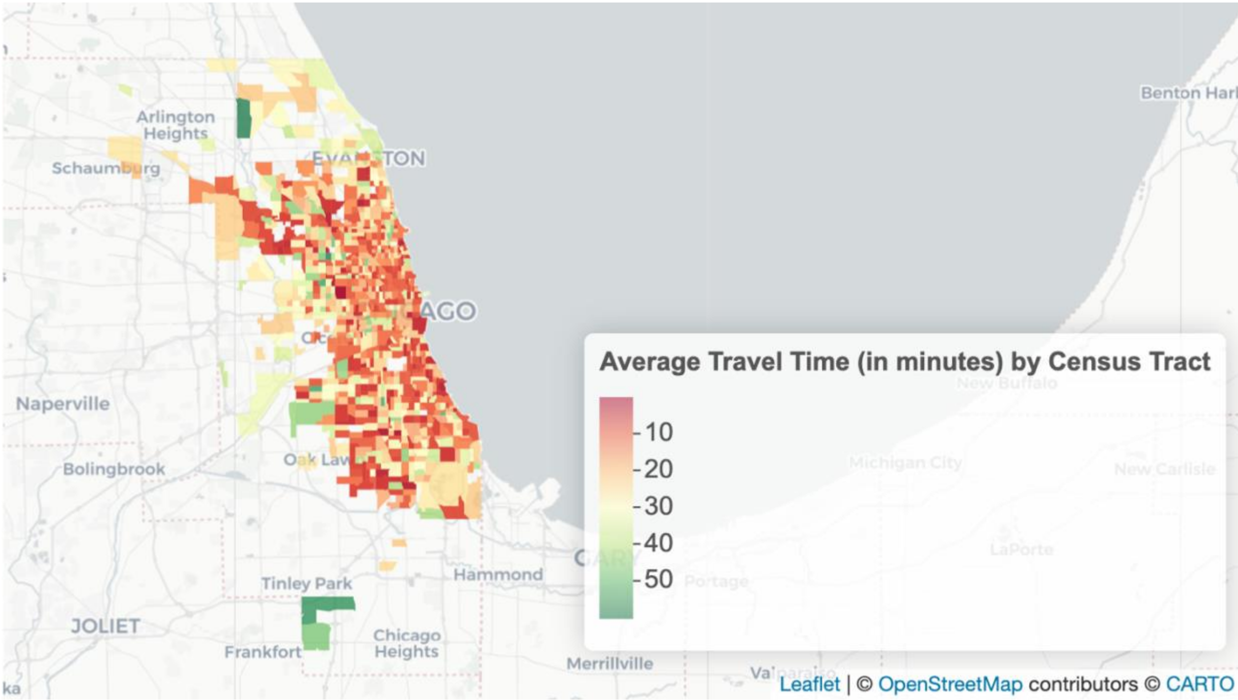

Figure 33: Travel Time of Ride by destination census tract (in minutes)

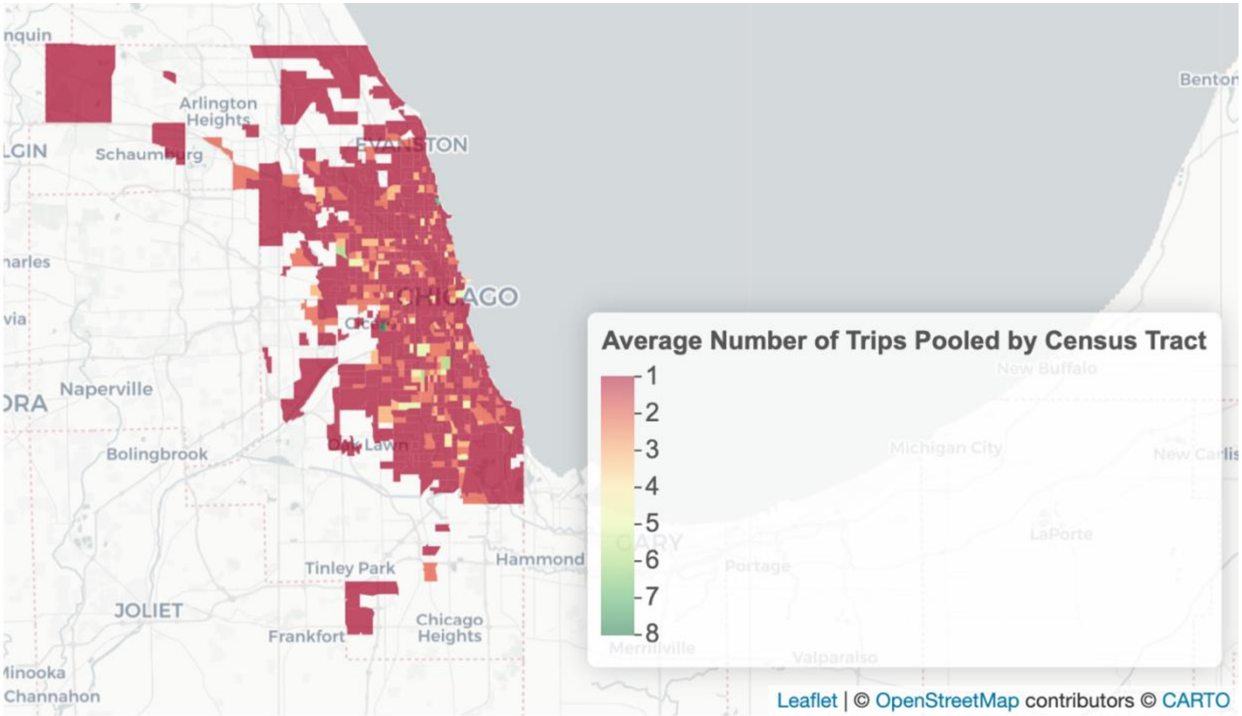

Figure 34: Number of Pooled Trips by Census Tract[23]

[23] Note: The bright red area outside the city limits on the north-west side of the map represents the area where the Chicago O'Hare airport is located.

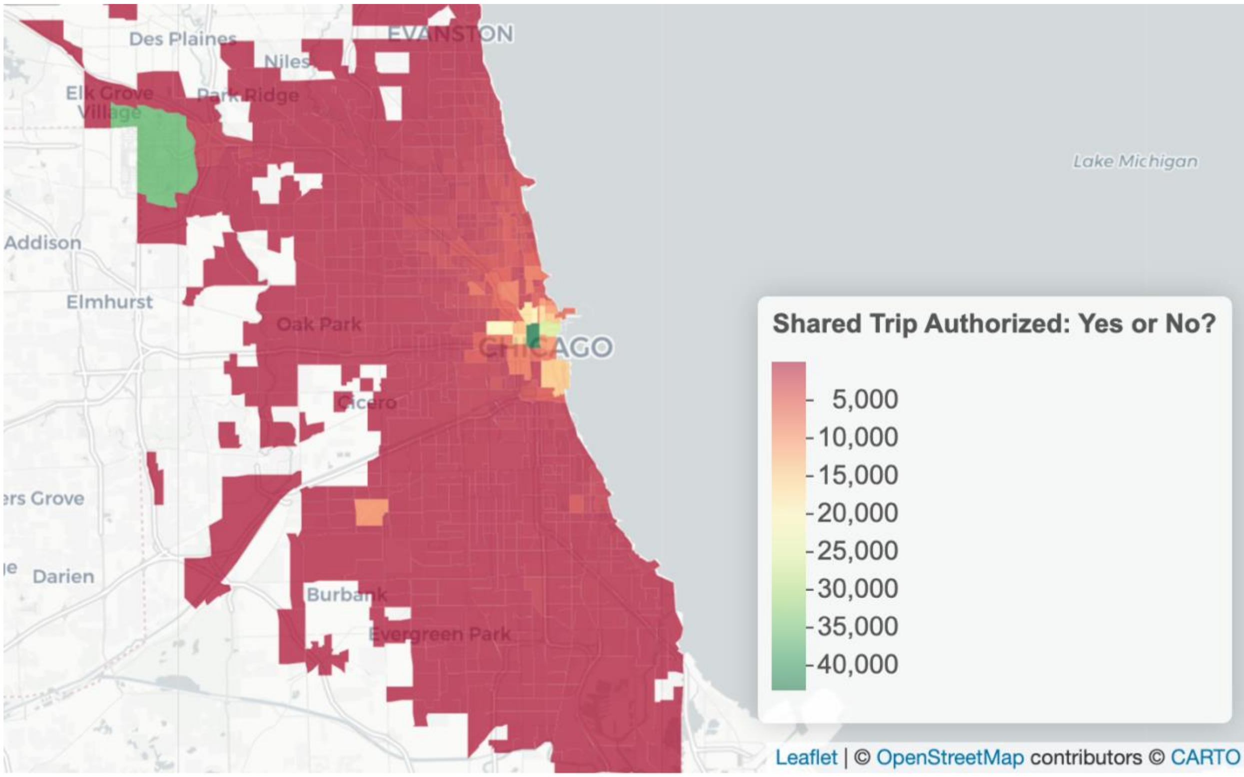

Figure 35: Number of Shared Trips Authorized by Census Tract

Figure 33, the most revelatory of the maps, shows that fares tend to be higher as one reaches the outskirts of the city limits. The maps on pooled trips and whether or not shared trips were authorized tell us that most trips as a whole were not pooled, and that shared trips were mostly authorized in busy neighborhoods, probably due to high demand. With regards to these neighborhoods, these richer, more popular areas have more attractions for locals and visitors alike such as tourist sites, restaurants, and businesses. In other neighborhoods, drivers may be less willing to authorize shared trips if there is lower demand or because of negative perceptions of people living in certain places. Areas with lower median incomes and higher poverty rates, for example, tend to have higher rates of crime reported due to a myriad of factors (higher rates of policing, lower resources, lack of community support and investment, etc.). Chicago is infamously, perhaps exaggeratedly considering the city is rated as safer on average[24] compared to other American cities, associated with a great deal of gun violence.[25] Community segregation and poverty are deeply rooted problems that go hand-in-hand with issues of violence and crime, and unfortunately discussing these dynamics in full is beyond the scope of this paper. It is relevant to mention however that perceived negative associations of certain neighborhoods based on racialized fears could cause less people to visit those areas via any means of transportation. Furthermore, Uber and Lyft drivers could go out of their way to avoid these areas out of fear and negative bias.

More likely, the inequitable service to these areas is due to de facto problems with segregation: the wealthier areas will have more to offer consumers because they are the center of investment and more gig workers will congregate in these areas to offer their services. Areas with poorer residents and higher rates of poverty, crime, etc. are further away from city centers, leading to higher fare rates and longer trips. Some of the most noteworthy observations from these maps are how many more rides are shared within higher income areas. While riders have the option to pick shared rides, drivers have the option to make their cars available for pooled trips and can even set up an option to have riders automatically matched and added to queues in the case of Lyft shared rides.[26] Allowing shared rides could be a way to keep up with consumer demands in higher-trafficked areas, but since the option is up to the discretion of the Uber or Lyft driver, they could opt to use this service less often in lower income areas.

---

[24] Neighborhood Scout. "Chicago, IL Crime Rates." *NeighborhoodScout*, Location Inc., 2020, www.neighborhoodscout.com/il/chicago/crime.

[25] Stef W. Kight, Michael Sykes. "The Deadliest City: Behind Chicago's Segregated Shooting Sprees." *Axios*, 14 Aug. 2018, www.axios.com/chicago-gun-violence-murder-rate-statistics-4addeeec-d8d8-4ce7-a26b-81d428c14836.html.

[26] Lyft, Inc. "How to Use Pre-Matching for Lyft and Shared Rides." *Lyft Help*, 2018, http://help.lyft.com/hc/en-us/articles/115013080268-How-to-use-pre-matching-for-Lyft-and-Shared-rides.

Further standardization of the data could help bring down some variances, making the relationships observed earlier more reliable and valid. Unsupervised learning techniques can help reveal broad patterns of similarity in the data by using clustering, which finds the x-values that are "close" to each other. This optimization problem minimizes the squared Euclidean distances between vectors in order to figure out how x-values are related to one another. In theory, x-components like tip and distance could have more of an impact on y-values of interest like fare, for example. For the clusters below, I compared fare and mean income as points to plot:

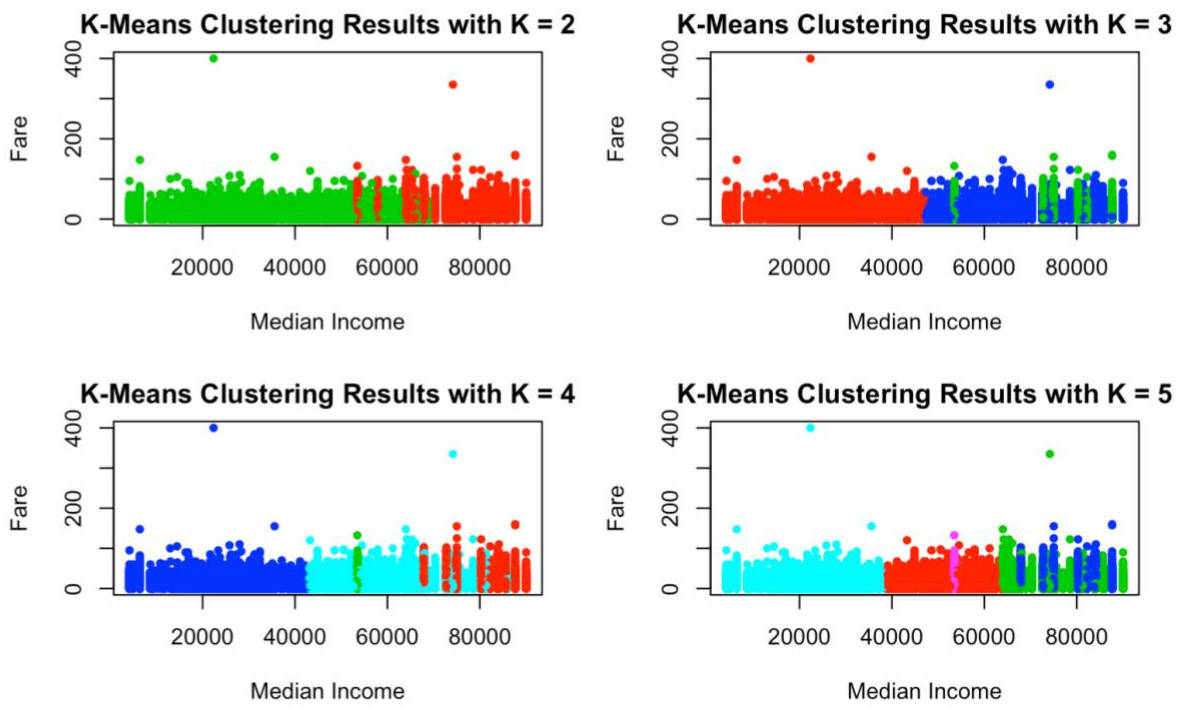

Figure 36: K-Means Clustering

Interestingly, the clusters seem to replicate the income levels established somewhat arbitrarily above. Areas with medium to high income tend to have more cross-over, while outliers with especially high fares existing mainly in lower income areas with higher-distances from the metropolis.

Uber and Lyft drivers are given GPS and drop-off information before they decide to pursue a ride and can decline ride requests that come along. Perhaps after a long day, drivers, who are most likely already overworked, are less willing to go long distances to the outskirts of town to drop off passengers. Because of the relatively small amount of data on canceled rides (represented by rides that end up with $0), I decided to use logistic classification for a preliminary supervised learning modeling:

```
## Accuracy    Kappa
##   0.9895   0.0000
```

```
confusionMatrix(z, reference = testing$y)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  yes   no
##        yes 1979   21
##        no     0    0
##
##                Accuracy : 0.9895
##                  95% CI : (0.9839942, 0.9934889)
##     No Information Rate : 0.9895
##     P-Value [Acc > NIR] : 0.5576868
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : 1.274967e-05
##
##             Sensitivity : 1.0000
##             Specificity : 0.0000
##          Pos Pred Value : 0.9895
##          Neg Pred Value :    NaN
##              Prevalence : 0.9895
##          Detection Rate : 0.9895
##    Detection Prevalence : 1.0000
##       Balanced Accuracy : 0.5000
##
##        'Positive' Class : yes
##
```

Figure 37: OLS Predictions

The test data predicts that around 1.3% of rides may be canceled given the predictions from training data. Canceled rides do not seem to the be the biggest worry within the dataset, especially when larger trends point to the existing problems in how riders and drivers are affected by the different areas they go to or come from. Further research can investigate and predict how real time information fed into the Lyft/Uber algorithms may affect where riders go, what kind of riders they pick up, and ways the app is learning to perhaps prioritize some neighborhoods or people over others. Studying the error rates of classification models of canceled rides and other driver activity compared to actual data may help give insight to the potential problems with application algorithms.

Discussion

While the analysis above does not strongly confirm or reject the listed hypotheses, there are some observed patterns of imbalanced service across different areas in Chicago. Objectively, the data show that rideshare service do indeed differ from neighborhood to neighborhood based on demographic variables, even if by a small amount. What the data cannot tell us directly at this moment are how individuals themselves actively make decisions that affect service. At best, one can assume that these differences are not directly out of malice: the economic structure of the city of Chicago does not lend itself to an equitable spread of services, which in turn affects the gig economy just like other kinds of industries. Rides with drop-off points in poorer neighborhoods or neighborhoods with higher proportions of black residents are likely to be more expensive. This is also true for rides with pick-up spots in these locations. Rides with drop-off locations in areas with higher rates of poverty are more likely to be canceled. All of this could lead to a vicious cycle of Chicago citizens already disadvantaged by poverty, race, and other socioeconomic factors to be further barred from these services due to expense and de facto discrimination. As shown in the subway map below, the transportation network of Chicago is centered around its commercially busiest area, and in turn prioritizes those with access to the opportunity and opulence of the city-center:

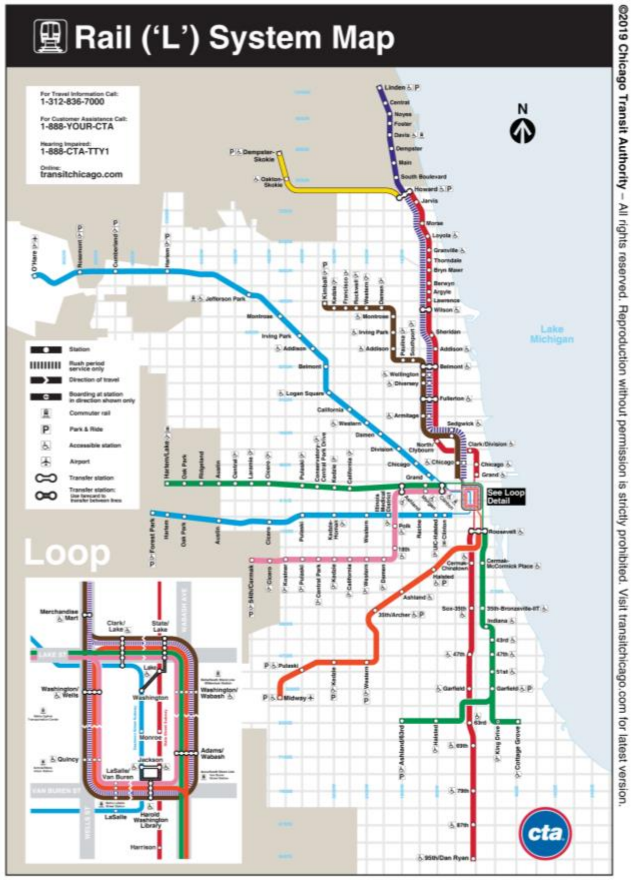

Figure 38: "All roads lead to the Loop" - Map of Chicago's Public Transit System[27]

Regardless of their lack of usage during the pandemic, Uber, Lyft, Via, and other rideshare companies that follow do not seem to be going away anytime soon. They will continue shaping transportation landscapes and policy as much as they are influenced by existing forces discussed by this paper. Therefore, in order to imagine a future that provides opportunities and services for disenfranchised individuals while also prioritizing environmental needs and

[27] Chicago Transit Authority.

pollution reduction, we cannot ignore rideshare companies from any discussion on transit plans. As mentioned above, shared ride services do offer a promising future in affordable and sustainable travel, especially in a country reliant on cars for transportation. However, the same problems related to lack of service in lower income areas also impact the frequency and locations of pooled and shared rides. Policy proposals related to rideshare companies and drivers could include further incentives for shared and pooled rides, especially those who travel long distances to reduce the need for multiple trips, more equitable spread of drivers across cities, and incentives to travel to underserved communities.

Still, these proposed changes are only the tip of the urban mobility iceberg. Ultimately, changes in how Uber and Lyft operate will not serve as panaceas for the intertwined and intersectional problems of transportation and inequity in our cities. “Transportation deserts” will still exist if underserved communities are not invested in through public transportation, business, and affordable housing initiatives. The implications of this are not to turn every lower-income neighborhood into a gentrified “hub” to attract tourists, but rather to work within and centering these communities in rebuilding and reshaping the area in a way that most benefits them. As far as TNPs go, their services can be used as tools to aid communities out in the margins, similar to Lyft’s efforts mentioned before to provide supplies to people in need during the pandemic. Philanthropic enterprises can be encouraged through private/public partnerships (ex. government grants and contracts to companies that are using their services for social good and responsibility). Anti-trust policy should also be considered to ensure that competition within this field can be encouraged instead of relying on the duopoly between Uber and Lyft we essentially have today. More options could lead to fairer pricing, expansions to different markets, and broader focuses on what cities need as a whole instead of extracting maximum profitability from and for the richest citizens.

Lyft and Uber drivers are only one small part of a larger system: further studies could explore how riders may replicate discriminatory behaviors based on the race, gender, presentation, etc. of their drivers, and vice-versa. A simulation of some kind could test the factors that may lead a user to choose one driver over another. Uber, Lyft, and similar companies are rightfully very private with regards to user and specific trip information, so many of these tests may have to be done in a controlled experiment setting. Besides researching the aftereffects of these apps and services, app designers, engineers, and policy makers could look into how these services could be more equally accessible and distributed across cities on the front end and how to help those in need, similar to Lyft and their COVID-19 programs. Ride-hail drivers need to have paths to financial stability and job security, an investment in both well-being and to ensure that drivers are providing quality service to all without being exclusively worried about their paychecks. In addition to rideshare data, city-wide surveys and other forms of feedback on transit service overall could be used by policymakers to decide the best course of action in order to provide transportation options for those on the margins.

Conclusion

The problems of segregation and access to resources within a city become glaringly obvious when mapped out. Although TNPs are not the *cause* of unbalanced services across the city of Chicago, existing economic realities end up designing the trajectories these companies take, further entrenching existing disparities. As mentioned, gig economy workers like Uber and Lyft drivers are not all a source of blame: they too are disadvantaged within this line of work as they trudge through the streets at night and set their apps to automatically give them more and more rides in order to make ends meet. Rationally, it makes *sense* for drivers to focus on areas with more potential clients in order to optimize their chances for profitability. The main problems therefore lie with the structures of the companies, cities, and the focus on individuals rather than systemic problems that are replicated through these practices. Uber, Lyft, and companies of the like should be motivated to provide their workers with fair pay, benefits, and basic necessities while also focusing on optimizing their social impact as companies, which provide an admittedly great service for many. Refocusing company goals on sustainable development, environmental protection, and consumer and labor well-being are the best ways for these companies to move forward, but it will take smart policy to galvanize them into action. In a city as deeply divided as Chicago, the effects of segregation are felt in all facets of life: transportation, education, professional opportunities, safety, security, etc. Increasing urban mobility through rideshares and using these services in the most environmentally friendly way possible are potential solutions to issues like unequal service and congestion. Cities like Chicago can be benefited and reinvigorated by helping those most vulnerable, revising transportation focus, ensuring workers and riders are being treated fairly, and *all that jazz*.

Appendix

```
 trip_seconds    trip_miles           fare              tip
Min.   :  41   Min.   : 0.00059   Min.   :  0.000   Min.   : 0.0000
1st Qu.: 497   1st Qu.: 1.50401   1st Qu.:  5.000   1st Qu.: 0.0000
Median : 767   Median : 2.82584   Median :  7.500   Median : 0.0000
Mean   : 920   Mean   : 4.10919   Mean   :  9.328   Mean   : 0.5303
3rd Qu.:1166   3rd Qu.: 5.17631   3rd Qu.: 10.000   3rd Qu.: 0.0000
Max.   :5166   Max.   :74.01092   Max.   :175.000   Max.   :20.0000
additional_charges   trip_total      trips_pooled   pickup_centroid_latitude
Min.   : 0.000     Min.   :  0.67   Min.   :1.000   Min.   :41.67
1st Qu.: 2.550     1st Qu.:  7.55   1st Qu.:1.000   1st Qu.:41.88
Median : 2.550     Median : 10.05   Median :1.000   Median :41.90
Mean   : 2.622     Mean   : 12.48   Mean   :1.225   Mean   :41.90
3rd Qu.: 2.550     3rd Qu.: 15.05   3rd Qu.:1.000   3rd Qu.:41.93
Max.   :15.850     Max.   :180.85   Max.   :7.000   Max.   :42.02
pickup_centroid_longitude dropoff_centroid_latitude dropoff_centroid_longitude
Min.   :-87.90            Min.   :41.67             Min.   :-87.90


1st Qu.:-87.67            1st Qu.:41.88             1st Qu.:-87.66
Median :-87.65            Median :41.89             Median :-87.64
Mean   :-87.66            Mean   :41.90             Mean   :-87.65
3rd Qu.:-87.63            3rd Qu.:41.92             3rd Qu.:-87.63
Max.   :-87.55            Max.   :42.02             Max.   :-87.55
   Poverty          Population     Median Income       Black
Min.   :  17.0   Min.   :  786   Min.   : 4494   Min.   :   0.0
1st Qu.: 274.0   1st Qu.: 3150   1st Qu.:44229   1st Qu.: 123.0
Median : 422.0   Median : 4480   Median :63431   Median : 239.0
Mean   : 558.3   Mean   : 5273   Mean   :60571   Mean   : 562.8
3rd Qu.: 696.0   3rd Qu.: 6603   3rd Qu.:79064   3rd Qu.: 600.0
Max.   :2770.0   Max.   :20087   Max.   :96667   Max.   :6448.0
    White         Poverty Rate     Percent Black    Percent White
Min.   :    0   Min.   : 0.7809   Min.   : 0.000   Min.   : 0.00
1st Qu.: 2138   1st Qu.: 6.3468   1st Qu.: 2.679   1st Qu.:59.41
Median : 3172   Median : 9.1584   Median : 4.975   Median :74.24
Mean   : 3620   Mean   :11.7126   Mean   :12.066   Mean   :68.70
3rd Qu.: 4480   3rd Qu.:13.1083   3rd Qu.:11.429   3rd Qu.:82.04
Max.   :11933   Max.   :73.8627   Max.   :99.663   Max.   :96.18
Number of trips/GEOID Number of Shared Trips Auth by GEOID  trip_minutes
Min.   :    6         Min.   :    6                        Min.   : 0.6833
1st Qu.: 1346         1st Qu.: 1346                        1st Qu.: 8.2833
Median : 3886         Median : 3886                        Median :12.7833
Mean   : 9156         Mean   : 9156                        Mean   :15.3326
3rd Qu.:14198         3rd Qu.:14198                        3rd Qu.:19.4333
Max.   :42768         Max.   :42768                        Max.   :86.1000
weighted_trip_min     y
Min.   :   0.032   yes:7921
1st Qu.:  15.858   no :  80
Median :  50.024
Mean   : 140.566
3rd Qu.: 165.359
Max.   :3602.491
```

Figure 39: Summary Statistics

Citations:

Agafonkin, Vladimir. *Leaflet Documentation*, OpenStreetMap, 2019, https://leafletjs.com/reference-versions.html.

Ajimi, Inès. "A Ride Sharing App Plays the Political Game (and Wins)." *The Economics Review at NYU*, 21 Feb. 2018, http://theeconreview.com/2017/11/03/a-ride-sharing-app-plays-the-political-game-and-wins/.

Bajpai, Jitendra N. "Emerging Vehicle Technologies & the Search for Urban Mobility Solutions." *Urban, Planning and Transport Research* 4.1 (2016): 83-100.

Barron, James. "Where Yellow Cabs Didn't Go, Green Cabs Were Supposed to Thrive. Then Came Uber." *The New York Times*, The New York Times, 3 Sept. 2018, www.nytimes.com/2018/09/03/nyregion/green-cabs-yellow-uber.html.

Bellon, Tina. "A New Chicago Ride-Hailing Law Reveals for the First Time What Uber and Lyft Really Charge." *Business Insider*, Business Insider, 26 Nov. 2019, www.businessinsider.com/ubers-carpool-pricing-strategy-revealed-by-chicago-fare-data-2019-11.

City of Chicago. "Transportation Network Providers (Ride-Hail Companies)." *City of Chicago: Transportation Network Providers (Ride-Hail Companies)*, 30 Oct. 2020, www.chicago.gov/city/en/depts/bacp/supp_info/transportation-network-providers.html.

Clewlow, Regina R., and Gouri S. Mishra. "Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States." (2017).

Clifford, Catherine. "Who Exactly Are Uber's Drivers?" *Entrepreneur*, 22 Jan. 2015, http://www.entrepreneur.com/article/242096.

Cognetta, Stephen. "I Analyzed My Lyft Driver Tips, Here's What I Found." *Hackernoon*, 3 Apr.

2018, http://hackernoon.com/i-analyzed-my-lyft-driver-tips-heres-what-i-found-94c890a36c0b.

Di, Xuan, et al. "A Link-Node Reformulation of Ridesharing User Equilibrium with Network Design." *Transportation Research Part B: Methodological* 112 (2018): 230-255.

Fingas, Jon. "System Helps Spot Bias in Algorithms." *Engadget*, New York: AOL Inc. May 26, 2016.

Ge, Yanbo, et al. "Racial and Gender Discrimination in Transportation Network Companies." No. w22776. National Bureau of Economic Research, 2016.

Ghili, Soheil, and Vineet Kumar. "Spatial Distribution of Supply and the Role of Market Thickness: Theory and Evidence from Ride Sharing." (2020).

Hannák, Anikó, et al. "Bias in Online Freelance Marketplaces: Evidence from Taskrabbit and Fiverr." *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2017.

Hill, Steven. "Rideshaning Versus Public Transit: How Uber and Lyft Tend to Widen Disparities of Race and Class in Urban Transportation Systems." *The American Prospect*, vol. 29, no. 2, Spring 2018, p. 46+. *Gale Academic OneFile*, https://link.gale.com/apps/doc/A537718733/AONE?u=columbiau&sid=AONE&xid=fb609dda. Accessed 6 Oct. 2020.

Hou, Yi, et al. "Factors Influencing Willingness to Pool in Ride-Hailing Trips." *Transportation Research Record* (2020): 0361198120915886.

Hughes, Ryan, and Don MacKenzie. "Transportation Network Company Wait Times in Greater Seattle, and Relationship to Socioeconomic Indicators." *Journal of Transport Geography* 56 (2016): 36-44.

Iqbal, Mansoor. “Uber Revenue and Usage Statistics.” *Business of Apps*. October 5th, 2020. https://www.businessofapps.com/data/uber-statistics/.

Leong, Nancy, and Aaron Belzer. “The New Public Accommodations: Race Discrimination in the Platform Economy.” *Georgetown Law Journal* June 2017: 1271+. *Business Insights: Essentials*. Web. 16 Sept. 2019.

Lyft, Inc. “How to Use Pre-Matching for Lyft and Shared Rides.” *Lyft Help*, 2018, http://help.lyft.com/hc/en-us/articles/115013080268-How-to-use-pre-matching-for-Lyft-and-Shared-rides.

Madell, Robin. “There's a Gender Pay Gap Even in the Gig Economy-But It's Not as Big.” *FlexJobs Job Search Tips and Blog*, FlexJobs.com, 7 Mar. 2019, www.flexjobs.com/blog/post/gender-earnings-gap-gig-economy/.

Muntaner, Carles. “Digital Platforms, Gig Economy, Precarious Employment, and the Invisible Hand of Social Class.” *International Journal of Health Services* 48.4 (2018): 597-600.

Neighborhood Scout. “Chicago, IL Crime Rates.” *NeighborhoodScout*, Location Inc., 2020, www.neighborhoodscout.com/il/chicago/crime.

RideGuru Team. “Rideshare Services for Women / Females.” *RideGuru*, 24 Mar. 2017, http://ride.guru/content/newsroom/rideshare-services-for-women.

Rosenblat, Alex, et al. “Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination.” *Policy & Internet* 9.3 (2017): 256-279.

Roy, Sneha, Anurag Komanduri, and Kimon Proussaloglou. “Evolution of Transportation Network Companies and Taxis through 2013–2018 in Chicago.” *Transportation Research Record* 2674.7 (2020): 385-397.

Shou, Zhenyu, et al. “Optimal Passenger-Seeking Policies on E-hailing Platforms using Markov Decision Process and Imitation Learning.” *Transportation Research Part C: Emerging Technologies* 111 (2020): 91-113.

Simon, Jean Paul. “Artificial Intelligence: Scope, Players, Markets and Geography.” *Digital Policy, Regulation and Governance* (2019).

Stef W. Kight, Michael Sykes. “The Deadliest City: Behind Chicago's Segregated Shooting Sprees.” *Axios*, 14 Aug. 2018, www.axios.com/chicago-gun-violence-murder-rate-statistics-4addeeec-d8d8-4ce7-a26b-81d428c14836.html.

Thebault-Spieker, Jacob, et al. “Simulation Experiments on (the Absence of) Ratings Bias in Reputation Systems.” *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW (2017): 101.

Turner Lee, Nicol. “Detecting Racial Bias in Algorithms and Machine Learning.” *Journal of Information, Communication and Ethics in Society* 16.3 (2018): 252-260.

Wickham et al. “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686, 2019, https://doi.org/10.21105/joss.01686.

Zakrzewski, Cat. “The Technology 202: Alexandria Ocasio-Cortez is Using her Social Media Clout to Tackle Bias in Algorithms.” *Washington Post – Blogs*, Washington: WP Company LLC. Jan 28, 2019.