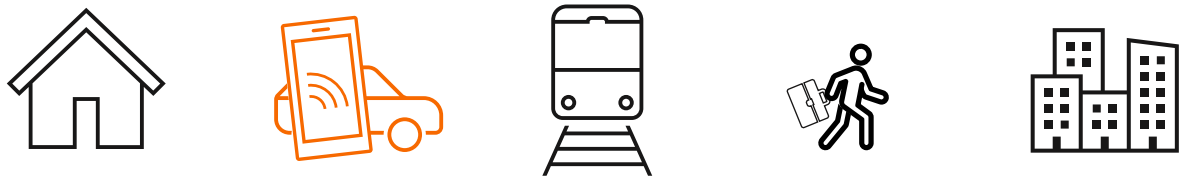
Modeling First-Mile Ride-Hailing Needs and Carpool Likelihood in Chicago, IL



CPLN 505 Final Project, April 2021
Jiamin Tan, Qi Si, and Ruqi Chen

Modified by Jiamin Tan, August 2022

Ride-Hailing and First/Last-Mile Problems

SEPTA + UBER

Uber rides will be discounted by 40 percent to-and-from 11 suburban Regional Rail stations … with a maximum discount of $10 per ride.

- SEPTA website 2016

Source: iseptaphilly.com

Ride-Hailing and First/Last-Mile Problems

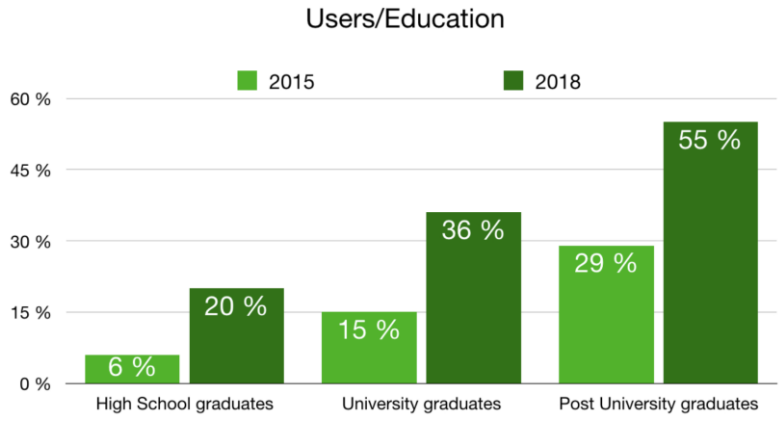
Via to Transit

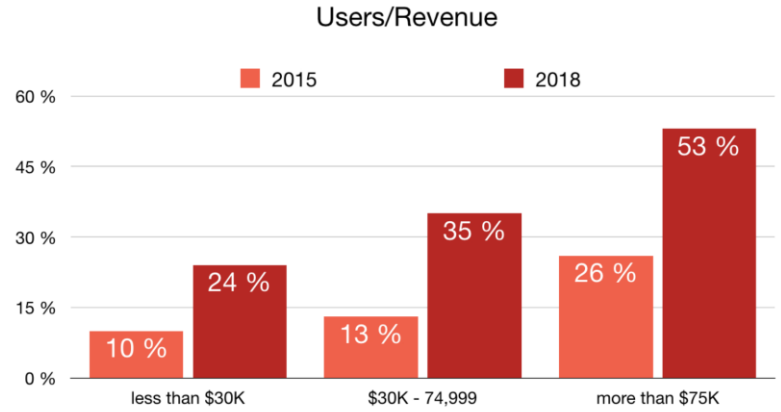
Via to Transit is a pilot, on-demand service… connect[ing] riders to and from three transit hubs in southeast Seattle and Tukwila. Rides will be shared with other Metro customers…

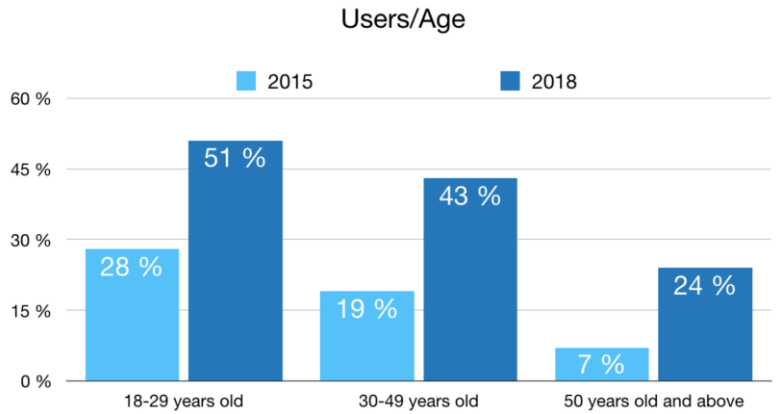
- King County website

Source: kingcounty.gov

Existing Researches on Ride-Hailing







Source: Pew Research Center, 2018

Research Questions

What socioeconomic factors are related to the first-mile ride-hailing needs of commuting trips?

What socioeconomic factors are related to people’s willingness to carpool in first-mile ride-hailing trips?

Ride-Hailing Data Available

Transportation Network Providers – Trips

- Nov. 2018 – Jul. 2022 (last updated)
- 263 million rows (trips) with 21 columns (trip attributes)
- Spatial resolution: census tract
- Temporal resolution: to the nearest 15-minute

Ride-Hailing Data Available

trip_start_timestamp	trip_end_timestamp	trip_seconds	trip_miles	pickup_census_tract
2018-12-03 07:45:00	2018-12-03 08:00:00	597	1.4106094	17031081403
2018-12-03 06:45:00	2018-12-03 07:00:00	1089	5.3045963	17031062400
2018-12-03 08:45:00	2018-12-03 09:00:00	776	1.5944449	17031320100
2018-12-03 08:00:00	2018-12-03 08:15:00	989	1.9820785	17031281900
2018-12-03 08:00:00	2018-12-03 08:45:00	2439	6.0767964	17031060200
2018-12-03 06:00:00	2018-12-03 07:00:00	3290	15.9732930	17031241500
2018-12-03 08:45:00	2018-12-03 09:00:00	1310	5.2808794	17031242500

Note: No Socioeconomics factors in this dataset.

Ride-Hailing Data Used in This Project

Transportation Network Providers – Trips FILTERED

- Weekdays in Apr. 2019
- Trips started between 6 and 9 am

More manageable data size (137K trips)

Mild weather

Riders represented by census data

Socioeconomics Data Used in This Project

American Community Survey (ACS) 2019 5-Year Estimate

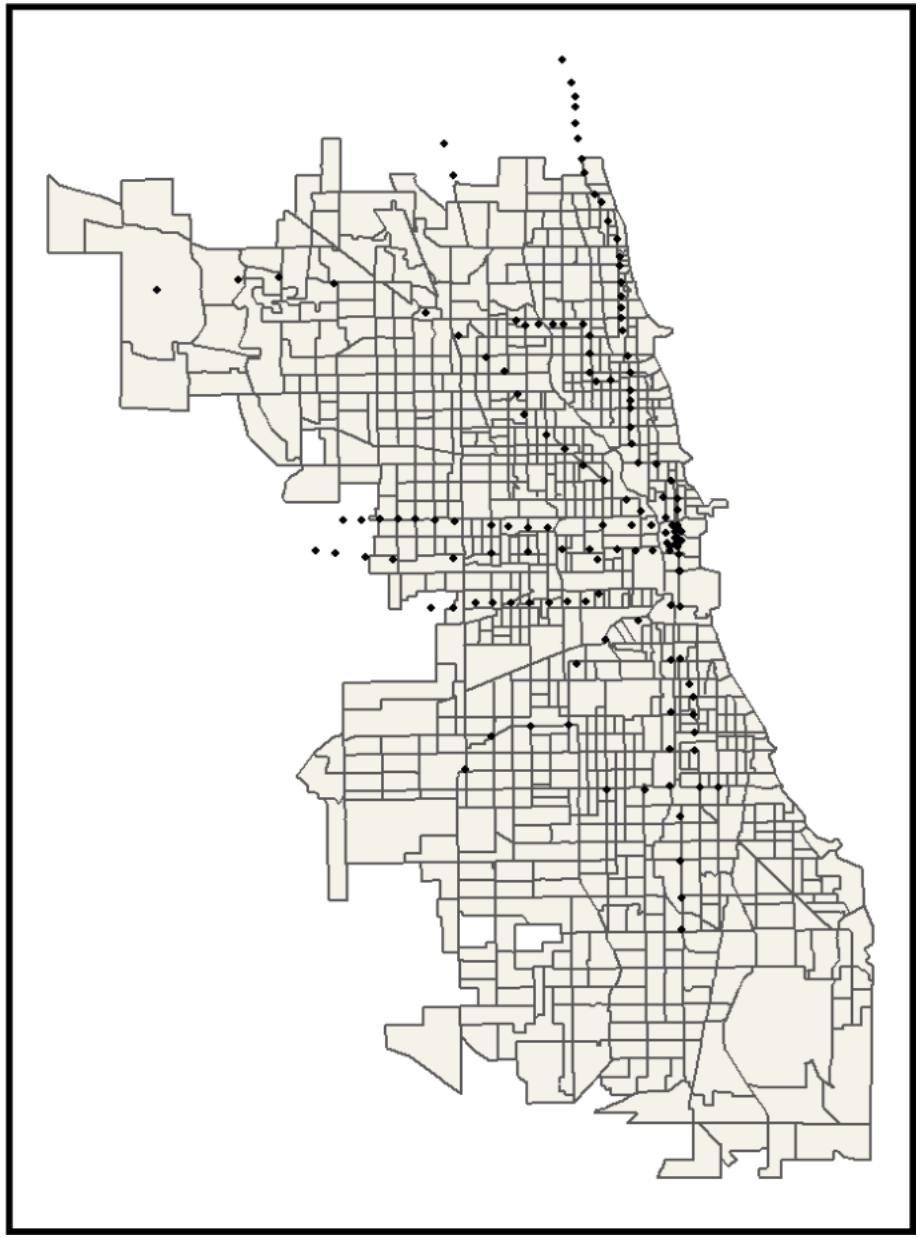
- ACS data in census tract level

Demographics	Housing	Commuting	Other
Age	Rent	Commuting Time	Educational Attainment
Gender	Ownership		Household Vehicle Availability
Income	Mortgage		Marital Status

Other Data Used in This Project

Spatial data of Chicago

Chicago Transit Authority (CTA) data



Define and Find First-Mile Trips

A trip is assumed as a first-mile trips if

the drop-off census tract is accessible to at least one CTA train station

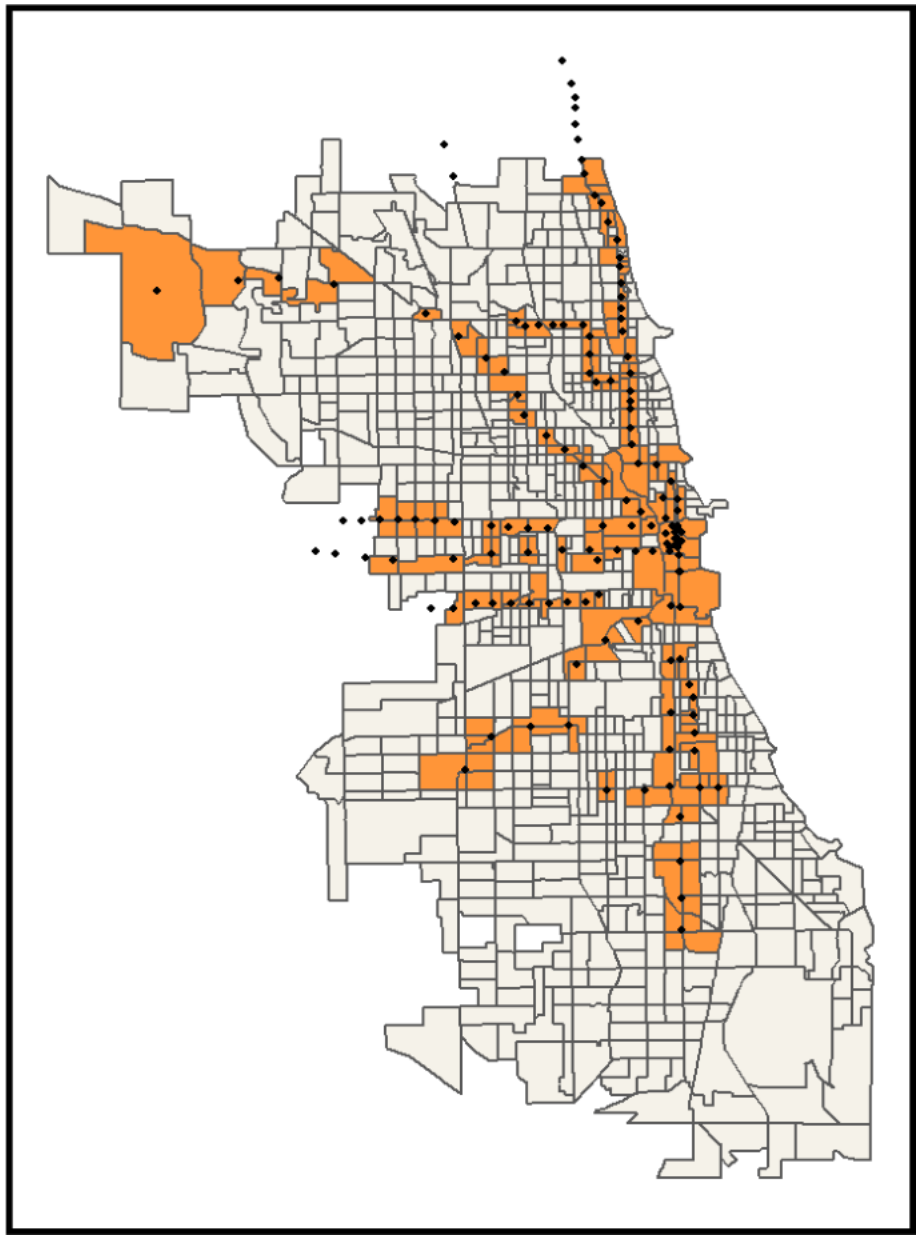
AND

the trip is shorter than 1.5 mile

Data Processing

STEP 1 – Find any census tracts accessible to a CTA trains stations

*Select any **census tract** touched by a 100-meter (328-ft) buffer from each station*

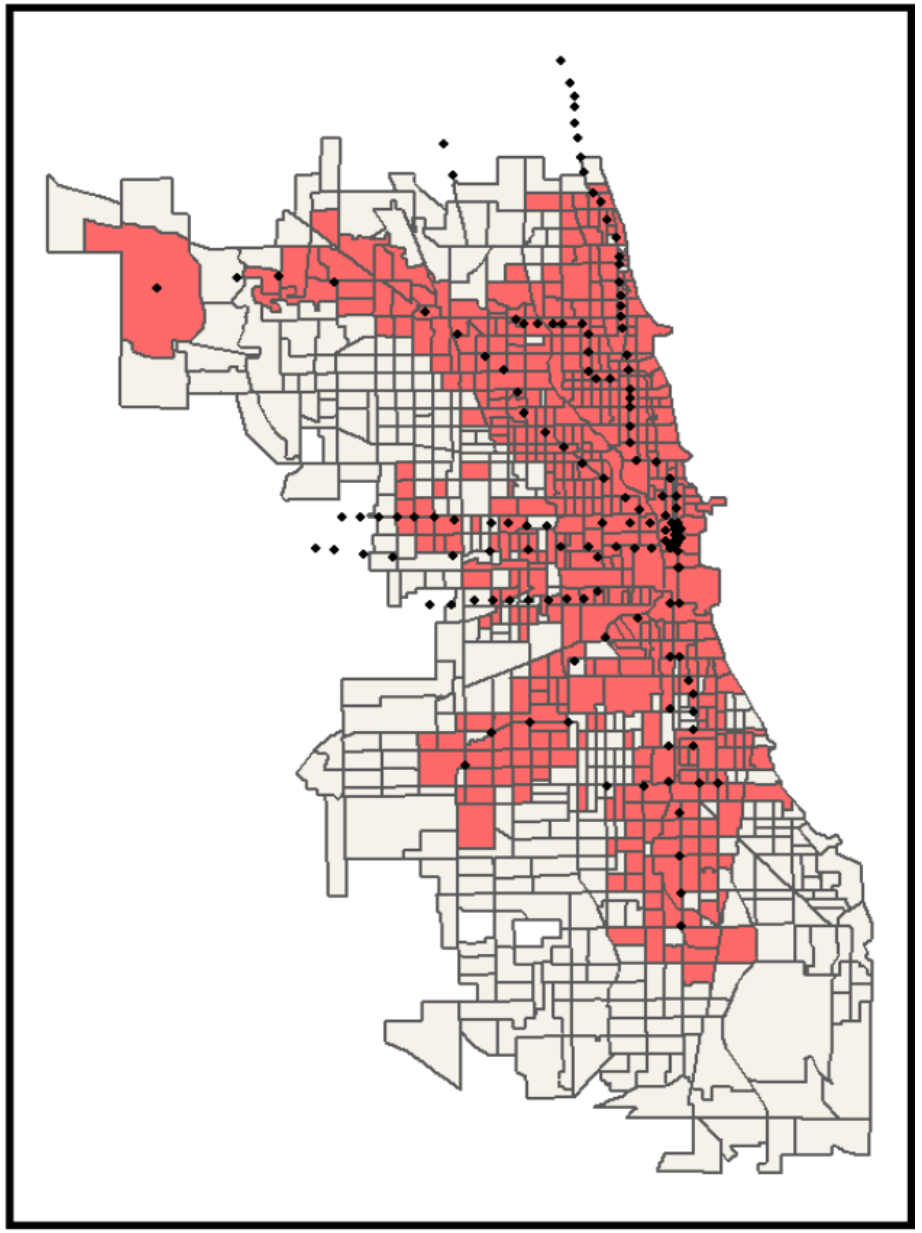


Data Processing

STEP 2 – Find pick-up census tracts of first-mile trips

Select trips shorter than 1.5 miles (network distance) and dropped off at station tracts

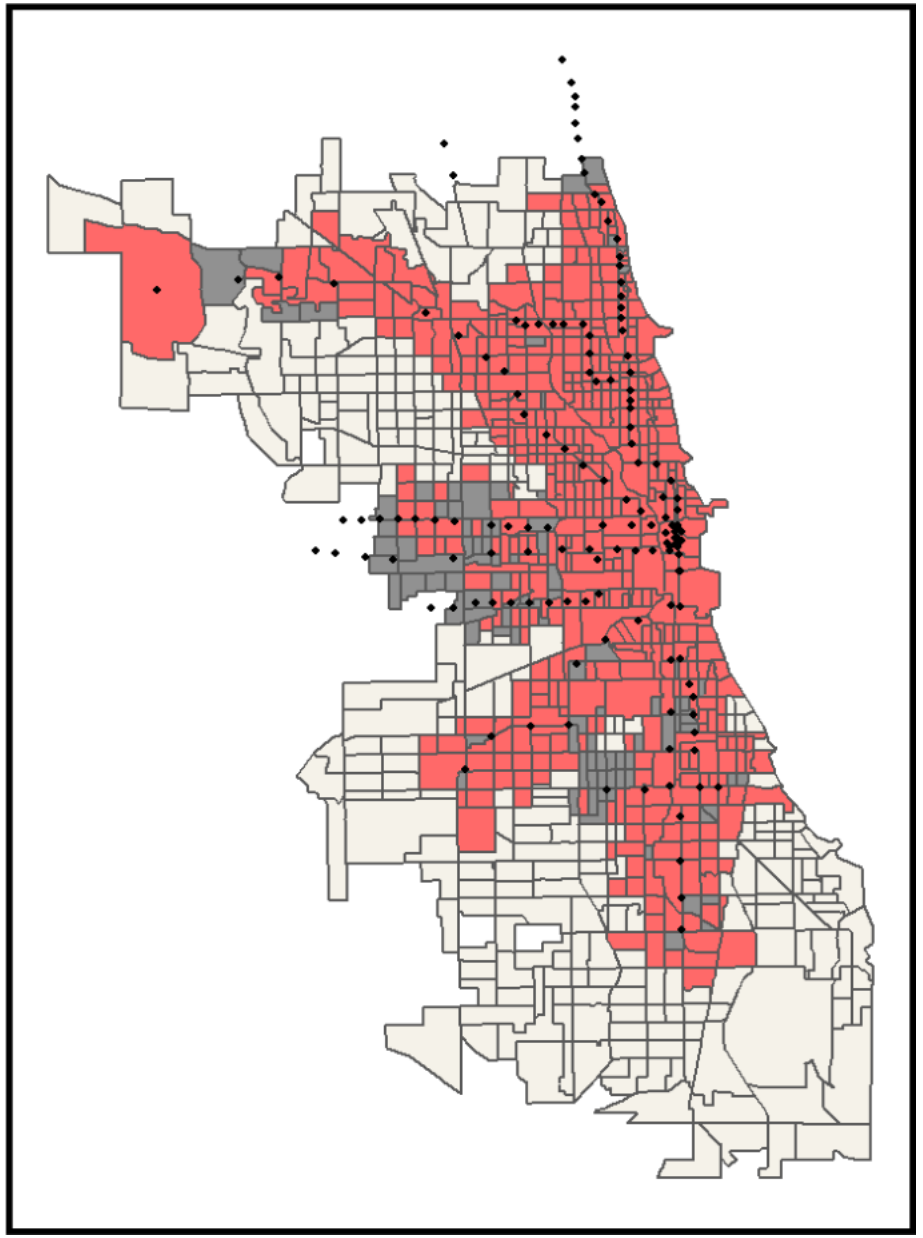
*Retrieve their **pick-up census tracts***



Data Processing

STEP 3 – Find the census tracts close to a train station but never generated any first-mile trips.

Select all census tracts whose centroids locates within 1-mile buffers from stations but never generated any first-mile trips.



Data Processing

census variables …

ride-hailing variables…

Each row is a census tract

	GEOID	TotalPop	MedRent	MedIncome	carpool	totalcost	pickup_num	carpool_ratio
1	17031010100	4522	874	62177	0	24.50	3	0.00000000
2	17031010201	7039	1023	47411	2	56.67	8	0.25000000
3	17031010202	2852	973	51719	0	12.50	2	0.00000000
4	17031010300	6650	976	66875	1	5.00	1	1.00000000
5	17031010400	5153	1010	59861	0	0.00	0	0.00000000

The total number of first-mile ride-hail trips

The ratio of carpooled first-mile ride-hail trips

STEP 4 – Merge first-mile trips count with census tract data.

A dataset with 529 rows (census tracts) was generated.

Research Question 1

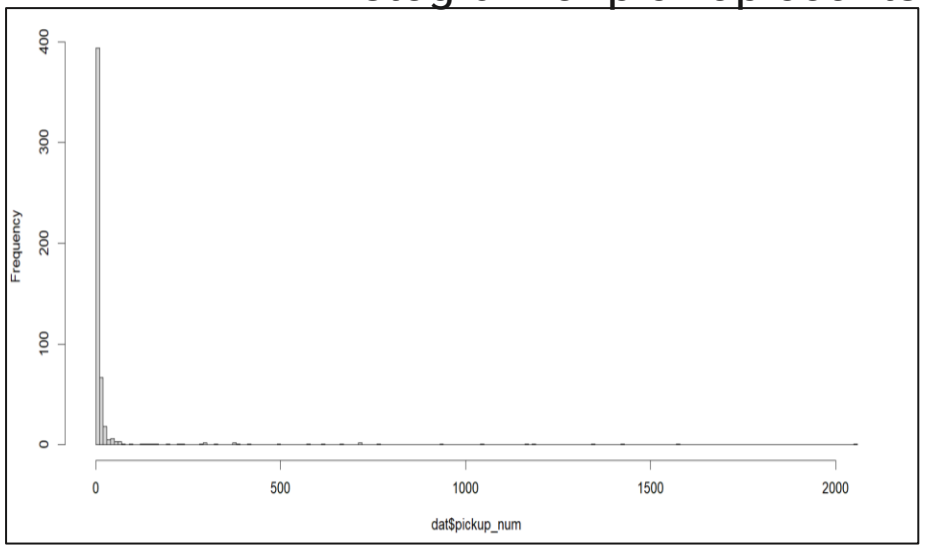
What socioeconomic factors are related to the first-mile ride-hailing needs of commuting trips?

Research Question 1
Exploratory Analysis

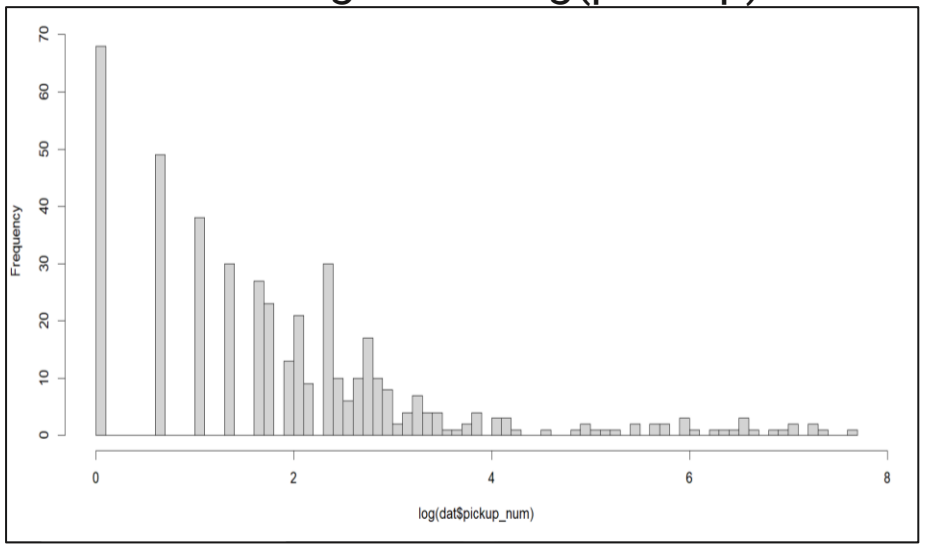
Distribution of the data is not normal

- *Linear regression not suitable*
- *Count data*
- *Negative binominal distribution*

Histogram of pick-up counts



Histogram of log(pick-up) counts



Research Question 1
Negative Binomial Models

Three models fit

- *The full model with all socioeconomic factors mentioned*
- *Two Trimmed models with significant variables only*

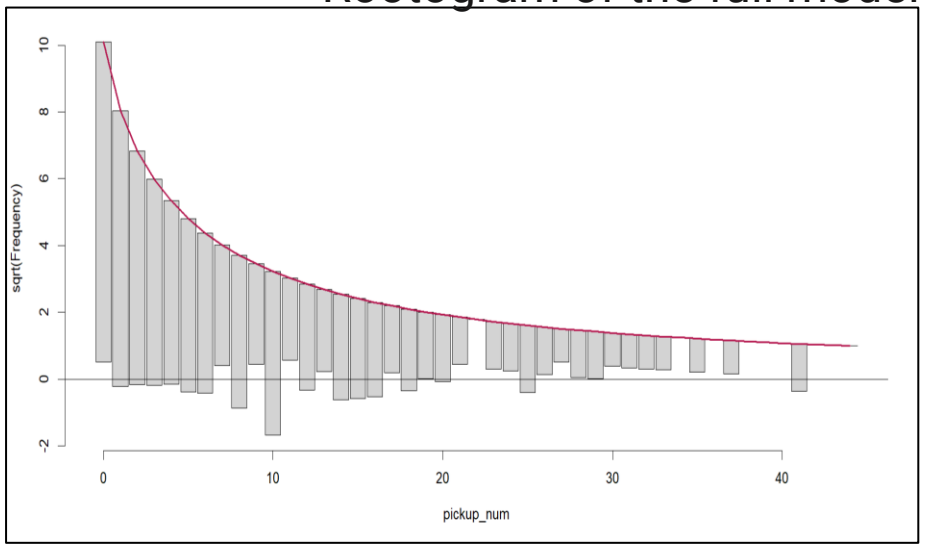
	Dependent variable:		
	pickup_num		
	(1)	(2)	(3)
MedRent	0.001***	0.001***	0.001***
	(0.0003)	(0.0002)	(0.0002)
MedIncome	-0.00000		
	(0.00000)		
MedAge	-0.047***	-0.054***	-0.051***
	(0.012)	(0.012)	(0.012)
CollegeRate	3.465***	4.089***	4.193***
	(0.884)	(0.658)	(0.653)
as.factor(maleTofemale)1	0.022		
	(0.107)		
Commute0_29mins_percent	5.185***	5.053***	5.101***
	(0.443)	(0.438)	(0.439)
HH_ownership_rate	-1.369		
	(0.901)		
NoVehicle_rate	4.007***	4.166***	3.661***
	(0.622)	(0.581)	(0.479)
Pop_Den	10.038		
	(7.106)		
single_Rate	-2.853***	-2.634***	-2.686***
	(0.708)	(0.692)	(0.685)
mortgage_Rate	2.574**	0.987	
	(1.192)	(0.676)	
Constant	-0.384	-0.326	-0.084
	(0.704)	(0.690)	(0.660)
Observations	529	529	529
Log Likelihood	-1,685.433	-1,687.034	-1,688.087
theta	0.815*** (0.058)	0.811*** (0.057)	0.809*** (0.057)
Akaike Inf. Crit.	3,394.867	3,390.068	3,390.174
Note:			*p<0.1; **p<0.05; ***p<0.01

Research Question 1
Model Selection

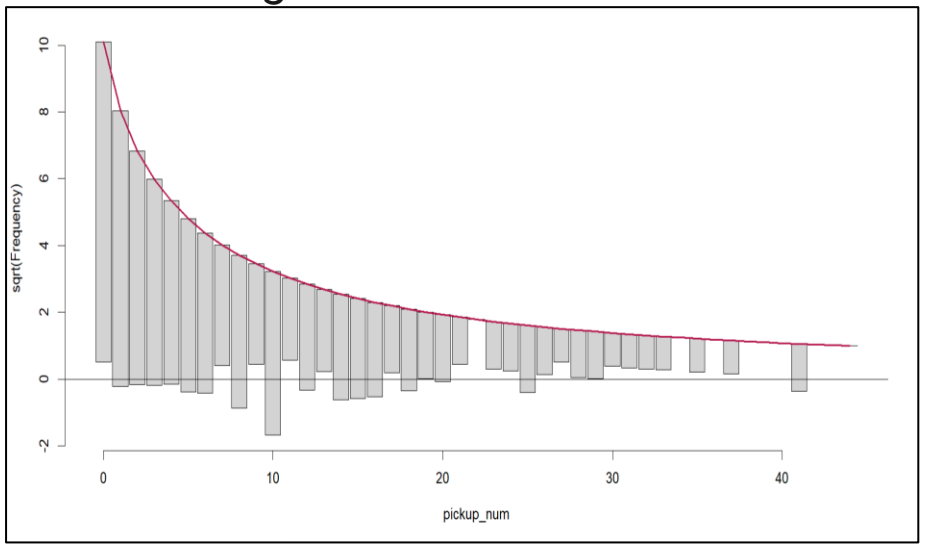
Full vs Trimmed

- *Likelihood ratio tests indicate there are no differences between the trimmed models and the full model*
- *Rootograms look the same*
- *The trimmed model with all variables significant selected*

Rootogram of the full model



Rootogram of the 2nd trimmed model



Research Question 1 Interpretation

Variables lead to increase trip counts

- *Median rent*
- *% Residents with bachelor's degrees*
- *% Commuters spending 0 - 29 min*
- *% Household without cars*

Variables lead to decrease trip counts

- *Median age*
- *% of single resident*

	Dependent variable:
	pickup_num
MedRent	0.001***
	(0.0002)
MedAge	-0.051***
	(0.012)
CollegeRate	4.193***
	(0.653)
Commute0_29mins_percent	5.101***
	(0.439)
NoVehicle_rate	3.661***
	(0.479)
single_Rate	-2.686***
	(0.685)
Constant	-0.084
	(0.660)
Observations	529
Log Likelihood	-1,688.087
theta	0.809*** (0.057)
Akaike Inf. Crit.	3,390.174
Note:	*p<0.1; **p<0.05; ***p<0.01

Research Question 2

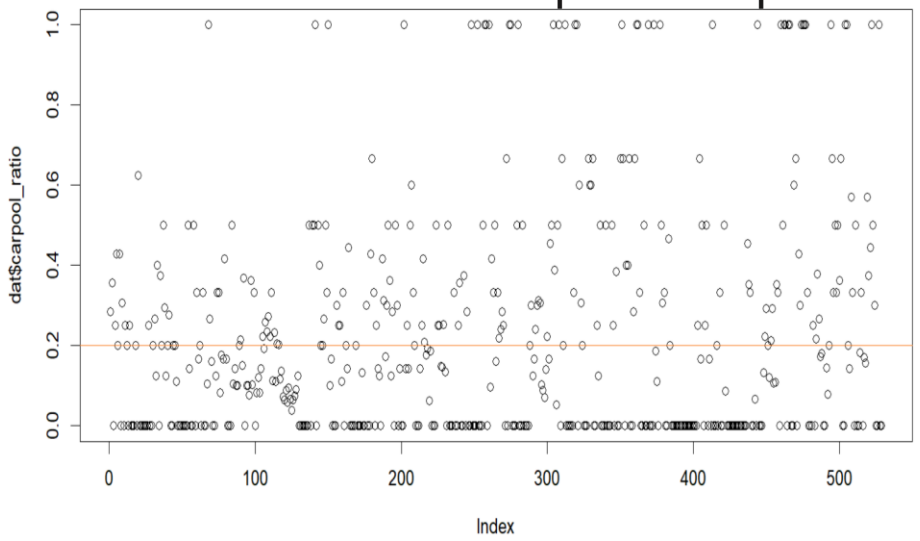
What socioeconomic factors are related to people’s willingness to carpool in first-mile ride-hailing trips?

Research Question 2
Exploratory Analysis

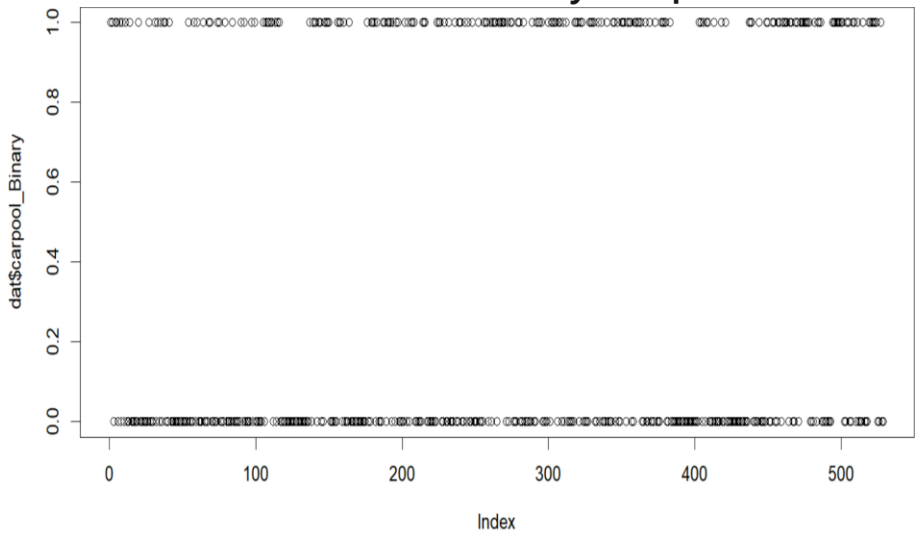
Census tract with carpool willingness

- *Based on carpooled trips ratio*
- *Binary levels assigned (threshold = 0.2)*

Carpooled trips ratio



Binary carpool choice



Research Question 2
Binomial Logistic Models

Two models fit

- *The full model with all socioeconomic factors mentioned*
- *The trimmed model is selected by backward stepwise AIC*

	Dependent variable:	
	carpool_Binary	
	(1)	(2)
TotalPop	-0.00000	
	(0.0001)	
Commute0_29mins_percent	1.587*	1.423
	(0.939)	(0.920)
CollegeRate	6.641***	5.541***
	(2.049)	(1.571)
child_rate	1.261	1.154*
	(0.805)	(0.663)
HH_ownership_rate	-1.289	-1.401*
	(1.117)	(0.785)
single_Rate	1.134	
	(1.660)	
MedRent	-0.002***	-0.002***
	(0.001)	(0.0005)
MedAge	-0.059*	-0.072***
	(0.033)	(0.025)
MedIncome	-0.00000	
	(0.00000)	
NoVehicle_rate	-0.713	
	(1.226)	
busstop_num	0.026*	0.025*
	(0.014)	(0.013)
Constant	1.698	2.723**
	(2.333)	(1.289)
Observations	423	423
Log Likelihood	-260.348	-261.065
Akaike Inf. Crit.	544.696	538.131
Note:	*p<0.1; **p<0.05; ***p<0.01	

Research Question 2 Interpretation

Variables lead to carpool likelihood increase

- *% Commuters spending 0 - 29 min*
- *% Residents with bachelor's degrees*
- *% Household with kids*

Variables lead to carpool likelihood decrease

- *% Household homeowner*
- *Median rent*
- *Median age*

	Dependent variable:	
	carpool_Binary	
	(1)	(2)
TotalPop	-0.00000	
	(0.0001)	
Commute0_29mins_percent	1.587*	1.423
	(0.939)	(0.920)
CollegeRate	6.641***	5.541***
	(2.049)	(1.571)
child_rate	1.261	1.154*
	(0.805)	(0.663)
HH_ownership_rate	-1.289	-1.401*
	(1.117)	(0.785)
single_Rate	1.134	
	(1.660)	
MedRent	-0.002***	-0.002***
	(0.001)	(0.0005)
MedAge	-0.059*	-0.072***
	(0.033)	(0.025)
MedIncome	-0.00000	
	(0.00000)	
NoVehicle_rate	-0.713	
	(1.226)	
busstop_num	0.026*	0.025*
	(0.014)	(0.013)
Constant	1.698	2.723**
	(2.333)	(1.289)
Observations	423	423
Log Likelihood	-260.348	-261.065
Akaike Inf. Crit.	544.696	538.131
Note:	*p<0.1; **p<0.05; ***p<0.01	

Research Question 2 Accuracy

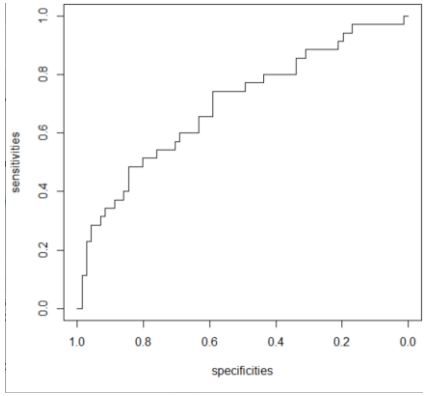
Trained on 423 observations (80%)

Tested on 106 observations (20%)

Actual

Predicted	0	1
0	60	20
1	11	15

ROC

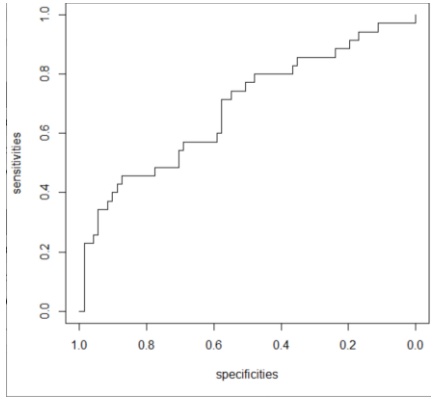


Full Model – Acc: 71%, AUC: 0.70

Actual

Predicted	0	1
0	58	19
1	13	16

ROC



Trimmed Model – Acc: 70%, AUC: 0.69

Takeaways

Socioeconomic data, such as rent, educational attainment, age, etc. are related to first-mile trips using ride hailing services.

In addition, they are related to whether the rider accepts carpooling as an option when using the service.

Transit agencies in other cities can use the analysis and conclusions for planning and allocating their own first-mile services.

Takeaways

Socioeconomics data alone are not enough to generate accurate trips predictions. Other factors, such as spatial/temporal lags and weather conditions can be considered in future research.

Data and modeling tools used in this project are from open sources. The analysis in R can be easily replicated and customized as needed.

Money saved from purchasing data and commercial software might be used for community engagements/surveys.

Thank You!