Charging fees on Uber of Lyft rides in the City of Chicago

Ziang Wang, Tao Ni, Alicia Zhou

April, 2020

Background Introduction

Content of the analysis

With the popularity of Uber and Lyfts that could help people save money and make transportation more convenient, the normal taxi industry and the parking fees collected by the government would be definitely affected, especially in big cities. In this project, we chose Chicago as the city to research, and we used the dataset containing 5000 records of ride from November to December in 2018 in Chicago from the government to analyze on the reasonable charge amount on rides going downtown in the City of Chicago that could be collected by the government and invested either in the taxi industry or as a compensation for the government's revenue.

We would mainly focused on rides dropping off in community area 8, 32, 33, 35, 38, 39, which are central community areas of Chicago. We would try to figure out a reasonable amount by looking at the relationship of fare, tip, and additional charges of each ride, together with information such as if the ride is pooled or shared. Since the situation might be different with different areas or different dates, we also provide analysis based on different regions and holiday(Chrismas) to see if there's a need to charge a different amount on those areas or on those dates.

We would generate the final suggestions by making related plots and tables, doing statistical tests, and conducting regression analysis.

Variables of the dataset

- `Trip ID`: ID of a ride record
- `Day`: Date of the ride, ranging from 11/01/2018 to 12/31/2018
- `Trip seconds`: Time using for a ride
- `Trip Miles`: Distance travelling for a ride
- `Pickup Community Area`: Community area where passenger is picked up
- `Dropff.Community.Area`: Community area where passenger is dropped off
- `Fare`: Fare amount of the ride
- `Tip`: Tip amount to the driver
- `Additional Charges`: Additional charges to this ride
- `Trip Total`: Total cost of this ride
- `Shared Trip Authorized`: If the trip is shared
- `Trips Pooled`: How many passengers are picked up during the pooled trip. If the number is 1, then it's not a pooled trip.

We made several variables as factors, as values of these variables should be discrete, and numeric values won't make sense. We also make the character "NA" of some variables as missing values for further analysis. In addition, we releveled factor `Shared Trip Authorized` to simplify it as "Yes"" or "No". We divided the dataset into three parts: training, query and test. Training set is mainly for EDA purpose. We used query part to build some models and see results. We used the test part to finalize our models and get the results.

Exploratory Data Analysis

Univariate plots

Since we concentrate on rides going downtown, we filter `Dropoff.Community.Area` such that it includes 8, 32, 33, 35, 38, 39, which are the central region and part of south region representing the downtown area. By filtering specific community areas, we could eliminate unuseful information and focus on rides we're interested in. For variables `Tip`,`Trip.Total` and `Trip.Miles`, we observed that most tip amounts are in between 0 and 6; most trip total costs are in between 0 and 50; most distances of trip are in between 0 and 30. We filtered these variables to make plots look better and more informative.
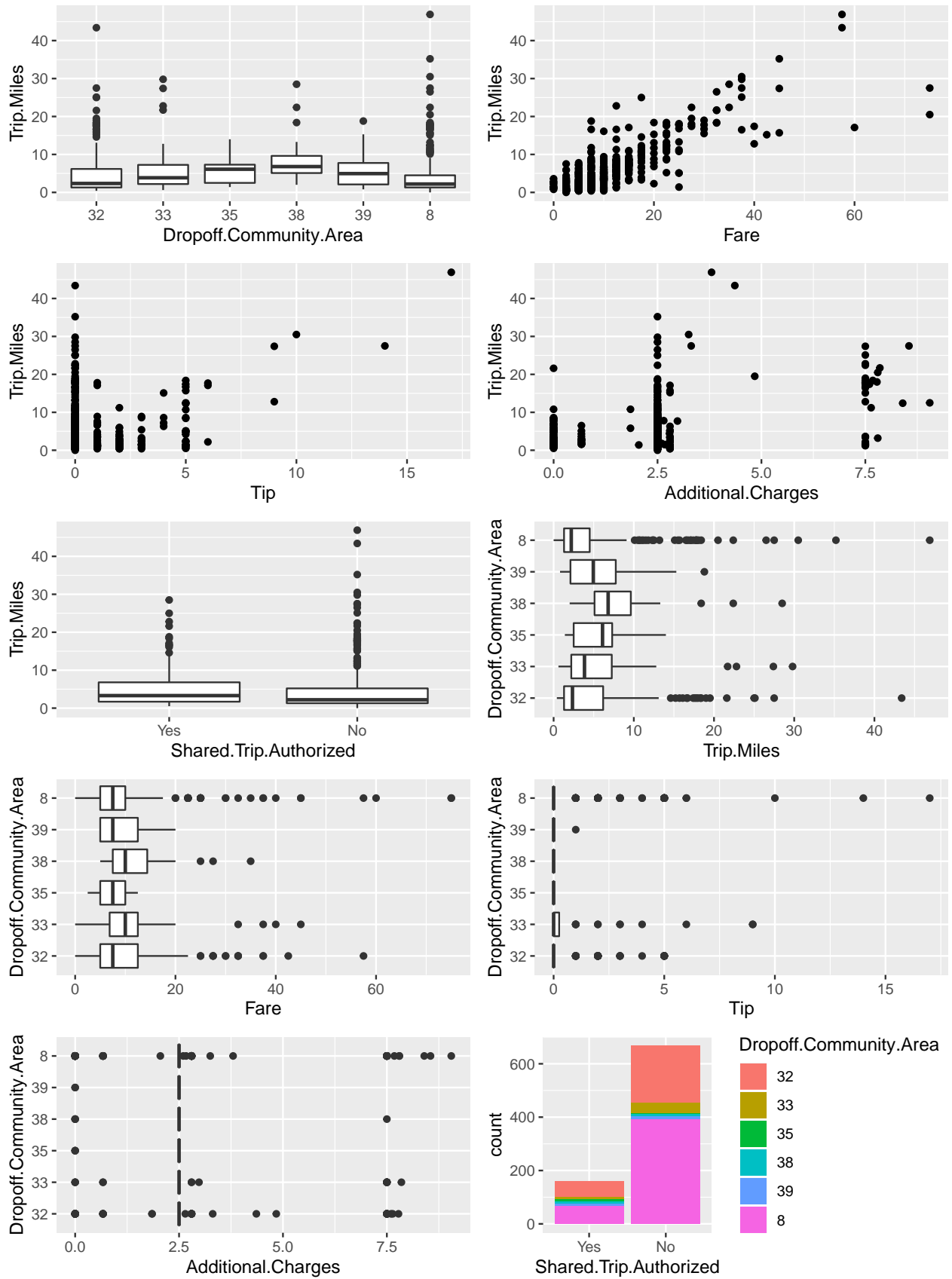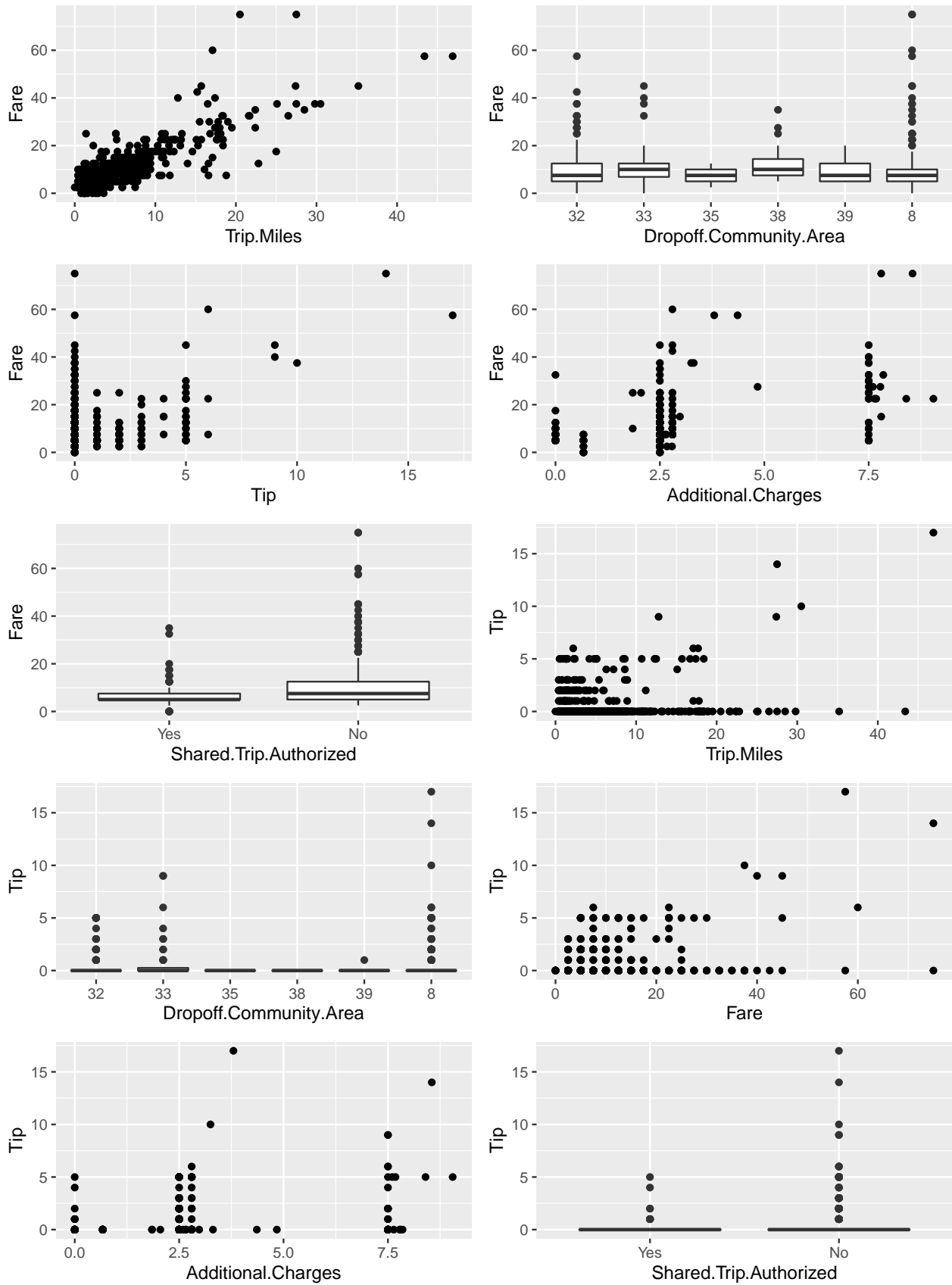
Univariate Graphics
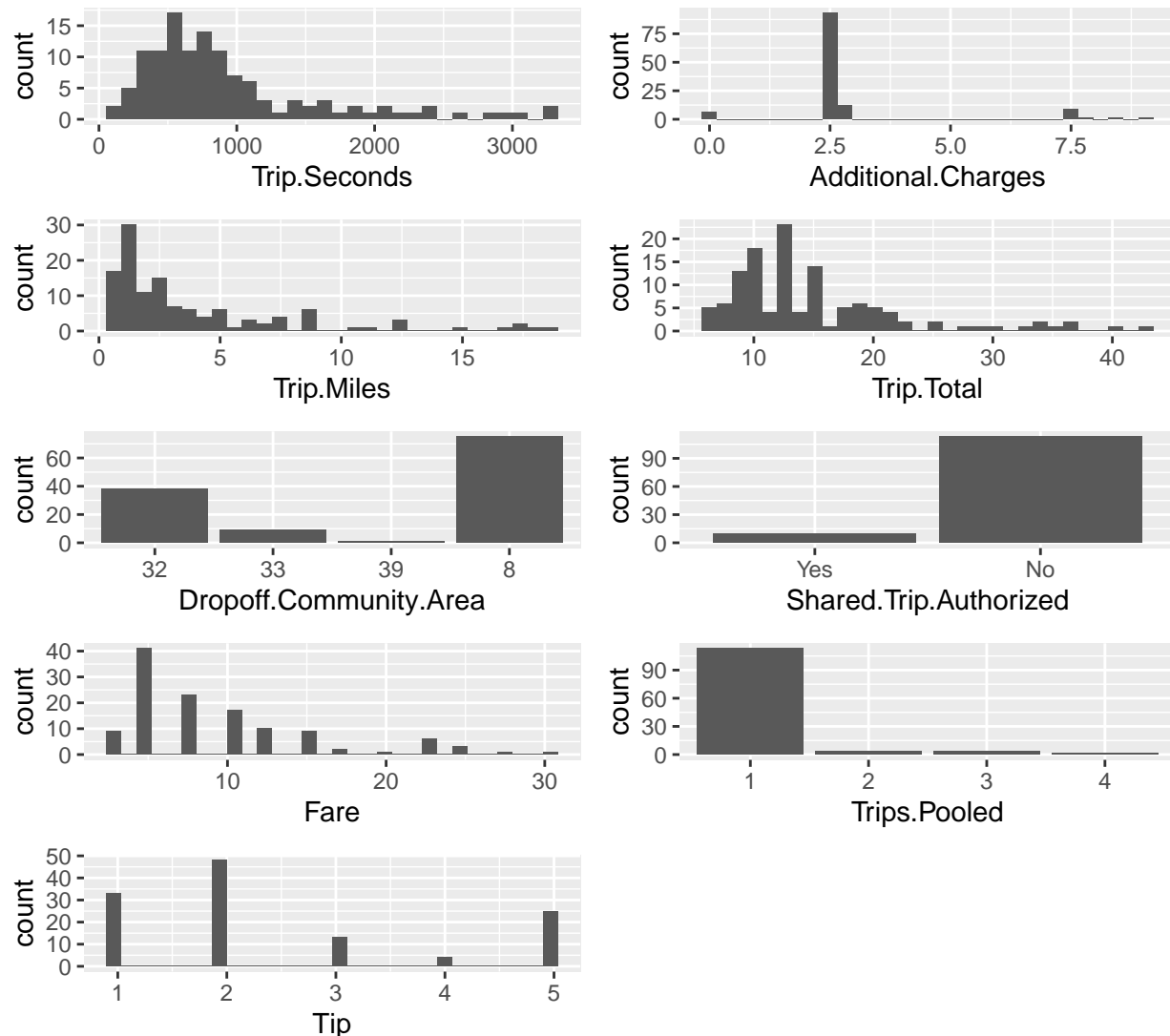
Bivariate plots

Again, we chose to filter `Dropoff.Community.Area` since we're interested in rides going to the downtown area.
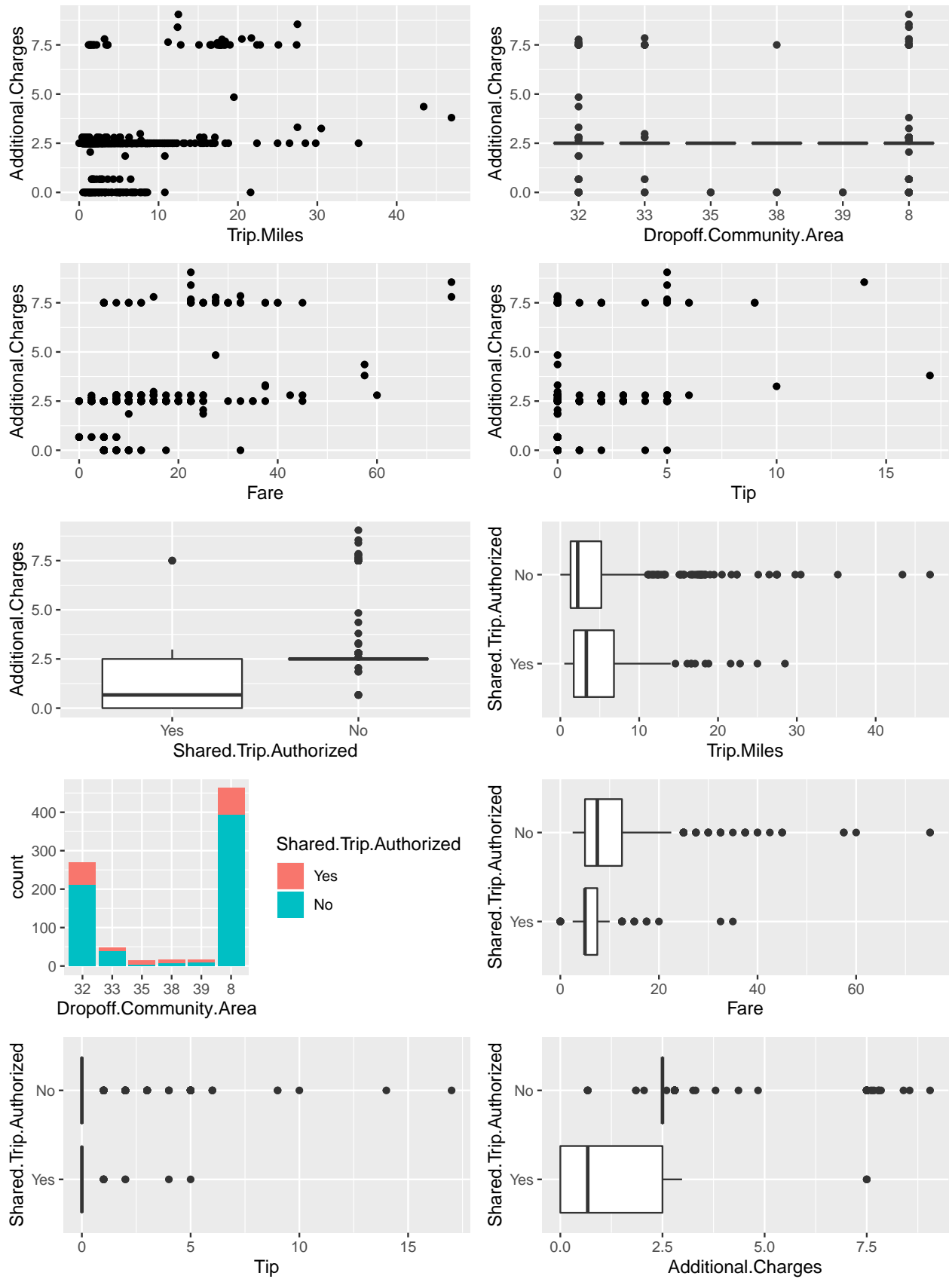
Bivariate Graphics

Bivariate Graphics

Bivariate Graphics

Tables

Table 1: Trip cost of ride to downtown community areas

Dropoff.Community.Area	Fare	Tip	Additional
32	9.32	0.37	2.49
33	11.46	0.88	3.03
35	7.83	0.00	2.00
38	12.97	0.00	2.34
39	8.75	0.06	2.19
8	9.06	0.54	2.51

Table 2: Trip cost of ride to downtown community areas on Chrismas

Dropoff.Community.Area	Fare	Tip	Additional
32	12.50	0.00	0.00
33	5.00	0.00	2.50
8	16.43	4.71	3.43

Hypotheisis Generation

Observations from EDA

From the EDA section, we could observe some patterns that worth further researching:

- The common additional charges are around $2.5, and people who would like to pay a tip usually pay less than 6 dollars. We also observe there might exist linear relationship between trip miles and tip amount. Since additional charges and tip are part of the "extra" cost that could be related with the "extra" charge we research on, we could generate some tests and models on `tip`,`Additional Charge`, `Trip Mile`, etc..

- We care more about the ride dropping off downtown, and we would like to charge on rides going downtown, indicating we should find out if rides going downtown have unique pattern that's different from the general trend. Thus, we will research on trip miles, trip seconds, pickup community areas with dropoff community areas. There could be significant differences for some community areas.

- It's observed that most rides going downtown would not authorize shared trips. There's a need to research if people would like to pay more if they don't authorize shared trips. In other words, we could charge more on unauthorized rides going downtown.

- The dataset includes information during Chrismas holidays. From the EDA table, we observe the fare is greater than the fare on other days, and people would like to go to specific downtown community areas. There's a need to research if there's new trend during Chrismas holiday.

Models and tests

We used the query dataset to build some models according to the information we get from EDA.

Case QQ: Trip.miles, Additional Charges, Tip, etc.

```
lm2<-lm(Trip.Miles~Tip+Additional.Charges, data=dat.query)
summary(lm2)
anova(lm2)

lm3<-lm(Tip~Fare+Additional.Charges+Trip.Miles, data=dat.query)
summary(lm3)
```

Case CQ/CC: downtown Area with other variables

```
glm1<-glm(downtown~Trip.Miles+Tip+Additional.Charges+Shared.Trip.Authorized, dat.query.downtown,
          family=binomial(link="logit"))
summary(glm1)

glm2<-glm(downtown~Shared.Trip.Authorized, dat.query.downtown, family=binomial(link="logit"))
summary(glm2)

lm4<-lm(Tip~Shared.Trip.Authorized+downtown, dat.query.downtown)
summary(lm4)

lm5<-lm(Additional.Charges~Shared.Trip.Authorized+downtown, dat.query.downtown)
summary(lm5)

chisq.test(dat.query.downtown$downtown, dat.query.downtown$Trips.Pooled)

glm3<-glm(downtown~Trips.Pooled, dat.query.downtown, family=binomial(link="logit"))
summary(glm3)
```
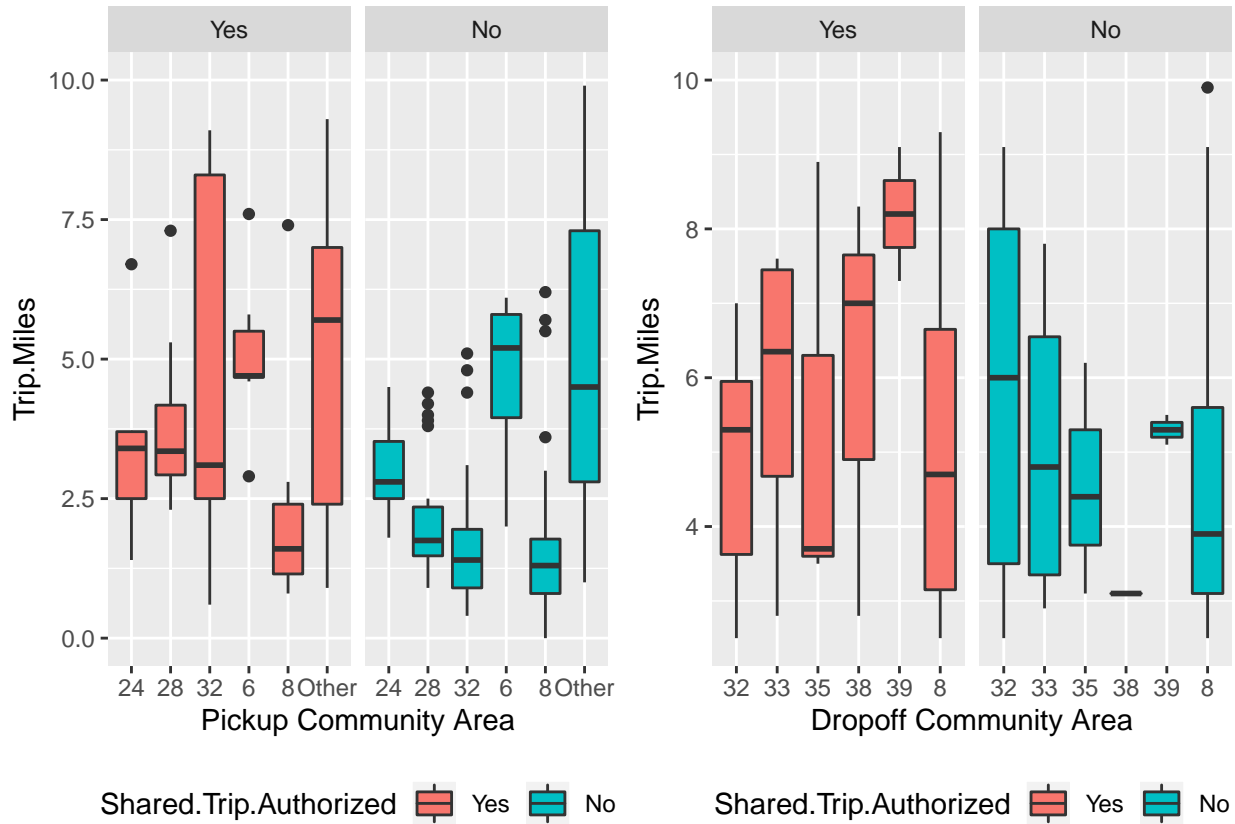
Case CQ/CC: Specify specific community Area

```
aov.test1<-aov(Tip~Dropoff.Community.Area, dat.query.community)
summary(aov.test1)

aov.test2<-aov(Additional.Charges~Dropoff.Community.Area, dat.query.community)
summary(aov.test2)
```

Case CQ/CC/QQ: Research on data during Chrismas Holiday

Table 3: Trip cost of ride to downtown community areas on Chrismas

Shared.Trip.Authorized	Fare	Tip	Additional
Yes	7.50	0.00	2.32
No	13.44	0.62	3.16

```
lm6<-lm(Trip.Miles~Tip+Additional.Charges, data=dat.query.chrismas)
summary(lm6)
```

Analysis Plan

After developing some models with statistical tests, the analysis plans are as follows:

- Use linear regression to test the relationship between `Tip` & `Additional.Charges` and `Trip Miles`. We need to know whether people would like to pay more for additional cost when travelling further, and what could be a reasonable average charge.

- Use logistic regression to test whether people authorizing shared trip are more likely to go downtown. We could apply price discrimination, and charge more on one group to earn more.

- Use the plot to show whether there are significant differences for previous analysis in specific community areas. We could apply price discrimination, and charge more on rides starting from and going to specific community areas.

- Use the table to show the situation of `Fare`, `Tip`, `Additional Charges`, and `Shared trip` during Chrismas holiday. We might need to apply different policies on those days, and we might charge more.

Hypothesis Confirmation Analysis

In this project, we used the dataset containing 5000 records of ride from November to December in 2018 in Chicago from the government to analyze on the reasonable charge amount on rides going downtown in the City of Chicago. Since we concentrate on rides going to downtown Chicago, I filter `Dropoff.Community.Area` such that it includes 8, 32, 33, 35, 38, 39, which are the central region and part of south region representing the downtown area. Our analysis mainly uses variables including `Day`, `Trip.Miles`, `Pickup.Community.Area`, `Dropoff.Community.Area`, `Tip`, `Additional.Charges`, `Shared.Trip.Authorized` to provide suggestions on average charge amount and additional charge with different situations, and we used the test dataset to fit models and analyze.

Suggested average charge amount

We used the linear regression to test the relationship between trip miles, tip amount, and additional charge amount to see how much additial fee people would like to pay, indicating how much fee on base fare rate on rides that we should charge.

Table 4: Regression Results of the suggested average charge amount

	Estimate	Lower bound	Upper bound	p-value
(Intercept)	1.2925	0.6445	1.9405	0.0001
Tip	0.2192	-0.0048	0.4431	0.0551
Additional.Charges	1.6801	1.4796	1.8807	0.0000

mean(Trip.Miles)
4.366429

The linear regression indicaes if people pay one dollar of tip and one dollar for Additional.Charges, the trip miles would be 1.2925+0.2192+1.6801=3.19 miles. We find the average distance of rides going downtown is about 4.37. Thus, people would like to pay 2*4.37/3.19=2.74 dollars for additional cost on average. Since people still need to pay for tip and other costs after we charge the small fee, it's suggested that the charge amount range from 1.5 dollars to 2.0 dollars on average.

Additional charge on non-shared trips

We categorized dropoff community areas of downtown as the downtown area and other areas. We used logistic regression to see if people who don't authorize shared trips are more likely to go to the downtown region.

Table 6: Regression Results for the Additional charge

	Estimate	Lower bound	Upper bound	p-value
(Intercept)	-1.1611	-1.454	-0.8825	0.0000
Shared.Trip.AuthorizedNo	0.5298	0.205	0.8646	0.0016

The logistic regression indicates if the passenger doesn't agree to share the trip with other people, then the

odds of going downtown would be exp(0.5298)= 1.699 times the odds of going downtown when the passenger allows for shared trips. In other words, if the trip is not shared, passengers are more likely to be dropped in downtown. Thus, we could charge more (could be 3 dollars since people would like to pay around 3 dollars more) on non-shared trips to make more revenue.

Additional charge on some community areas

The situation might be different in some community areas. We used the plot to show if we could charge differently on rides starting in some areas or stopping in some ares.
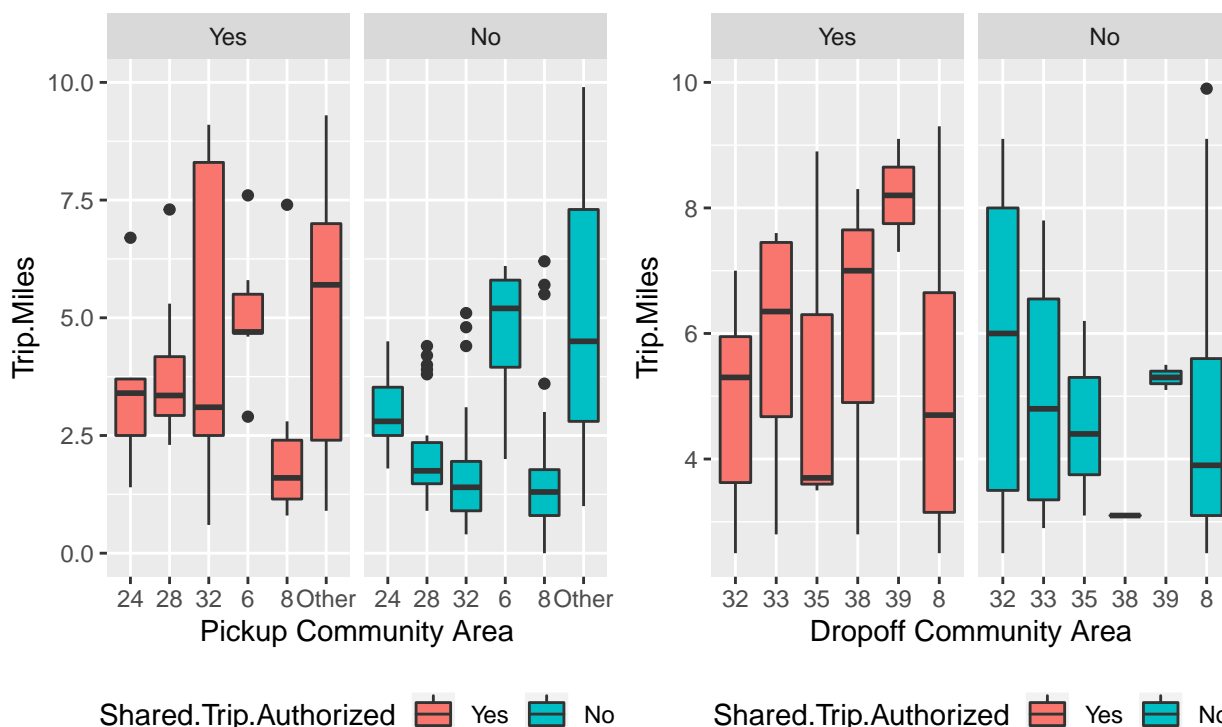
From the plot on the left, most rides start from the downtown area or areas not far from the downtown region, and there's no obvious difference if the trip is shared or not. For additional revenue, we could also charge a small fee (could be also 1.5 to 2.0 dollars) for rides starting in the downtown central area.

From the plot on the right, Community 32 and Community 33 are popular areas. It's not surprising since these two areas are central part of Chicago. Thus, we could charge more fees(could be also 3 dollars or even more) on rides dropped of in these two community areas.

Different charge on Chrismas (12.24 and 12.25)

We use the table to show if there's a difference of those variables from previous analysis on Chrismas.

Table 7: Trip cost of ride to downtown community areas on Chrismas

Shared.Trip.Authorized	Fare	Tip	Additional
Yes	15.00	0	2.50
No	12.25	1	3.58

From the table, we see people would like to pay much more on Chrismas when the shared trip is not authorized. One reason is that it's often hard to call for a taxi on holidays. It's suggested that we could charge about 4

dollars when the trip is not shared, and charge 2.5 to 3.0 dollars when the trip is shared.

Conclusion

Based on the analysis, a reasonable amount to charge is about 1.5 to 2.0 dollars. If passengers don't agree to share trips with others, the fee could be about 3.0 dollars. To increase the revenue, we could not only charge fees on rides going downtown, but also rides leaving downtown. Again, we could charge approximately 1.5 to 2.0 dollars. It's found that there's a difference within the downtown region. If the ride is going to community 32 and community 33, we could charge more to be 3.0 dollars. Finanly, people would like to pay more during holidays. During Chrismas or other holidays, we could charge 3.0 dollars for shared trips and 4.0 dollars on non-shared trips.

References

City of Chicago. 2019. "Community Areas with Numbers in Chicago," May. https://www.chicago.gov/content/dam/city/depts/doit/general/GIS/Chicago_Maps/Citywide_Maps/Community_Areas_W_ Numbers.pdf.