In our project, the dataset we used is Taxi Trips from Chicago Data Portal with the link <https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew/data>. Since the data is too big, we filtered the Taxi trips with Trip Start Timestamp from last year, and then used Pandas notebook to read the file and drop the NaN value of the Taxi trips in 2021. This data set has 882,777 rows and 5 columns after removed, and we took some attributes such as Trip Start Timestamp, Trip Miles, Fare, Tips and Company. Then, we made a subdataset called TaxiData using groupby date to obtain the daily count of trips average distance, fare and tips of Taxi Trips in each day in Figure 1. Beside the TaxiData, we also filtered 4 subdatasets of Taxi trips in March with 18,524 rows, June with 62,813 rows, September with 117,561 rows and December with124,441rows of 2021, and those dataset contain attributes suach as date, distance, fare, tips and taxi company of each month as Figure2. After we cleaned and processed the data, we exported from panda notebook and imported them in D3.js to make data visulation and analysis.

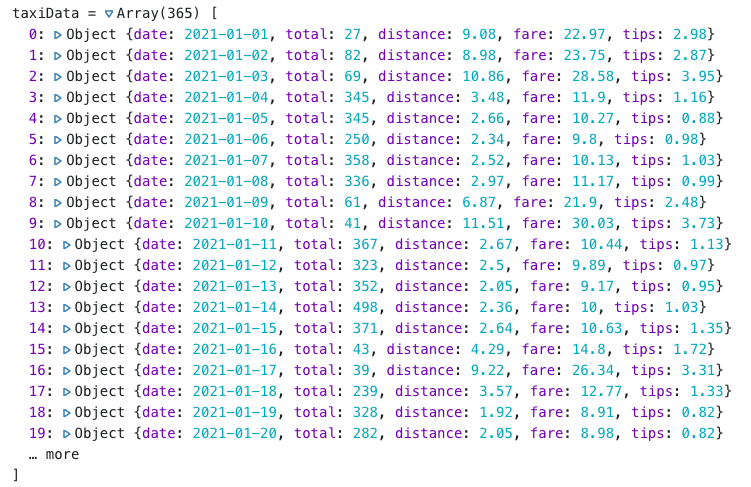


Figure 1. Dataset with date, monthly trips, average distance, fare and tips of taxi trips in 2021

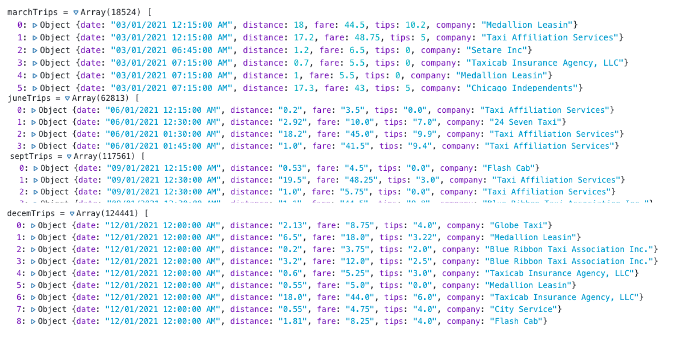


Figure 2: subdatasets represent the taxi trips info in March, June, September and December

Based on the attributes we used, we came up with three questions based on these data.

* Which month have larger number of trips and how the riders tipping does in 2021?
* Comparing the taxi trips in March, June, September and December, which taxi company has favorable fare?
* Comparing the taxi trips in March, June, September and December, which taxi company has more trips and better service?

Since we have the attributes of date, number of trips, average distance, fare and tips of trips, we used a line chart to gain the monthly distribution of taxi trips in each month in 2021 in Figure 3. The x-axis represents each month in 2021, and the y-axis represents the total number of trips. Through this line chart, we can see the number of taxi trips increasing from January to October and the winter seasons Octomber and December have more trips than other months because riders prefer to take taxis under colder weather. Besides the line chart, we also used a scotter plot to display the distribution of trips and tipping situations of trips in each month in Figure 4. We took three scales of tips, the orange points representing the tips are less than or equal to $2, the gray points represent the tips are greater than $2 and less than or equal to $4, and the red dots are tipping greater than $4. It is clear to see that riders are likely to pay more tips with longer distance and higher fare of the trips, and we can concluded that the average tipping of taxi trips is around $2 to $4 in 2021.

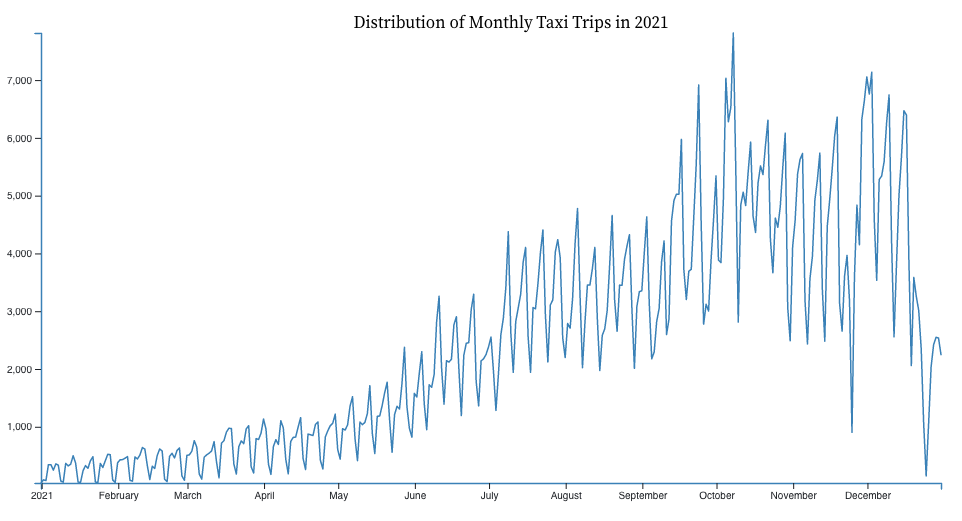


Figure 3. Distribution of Monthly taxi trips in 2021

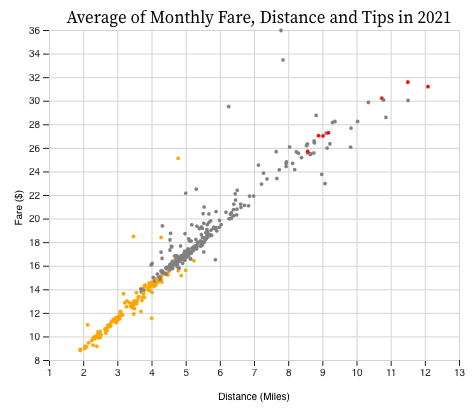


Figure 4. Average of distance, fare and tipping of trips from each month in 2021

In order to analysis the taxi companies with trips in 2021, we selected 4 months taxi trips in March, June, September and December to and use interactive scatterplots to estime the relationship between distance, fare and company. We took distance, fare and company as x-axis, y-axis and color scale to create brushable scatterplots with Taxi trips in 4 monthes. When we hovered the brushable scatterplots, we can retrieve the information of date, distance, fare, tips and company of those dots,. Then, we took the trips with longer distance and lower fare as estimation to access the taxi companies within more favorable fare. By looking at figure 5, figure 6, figure 7 anf figure 8, we noticed the companies have more favorable fare than other competitors are Taxicab Insurance Agency, LLC, Sun Taxi, Taxi Affiliation Service and Taxi Affiliation Service in March, June, September, and December.

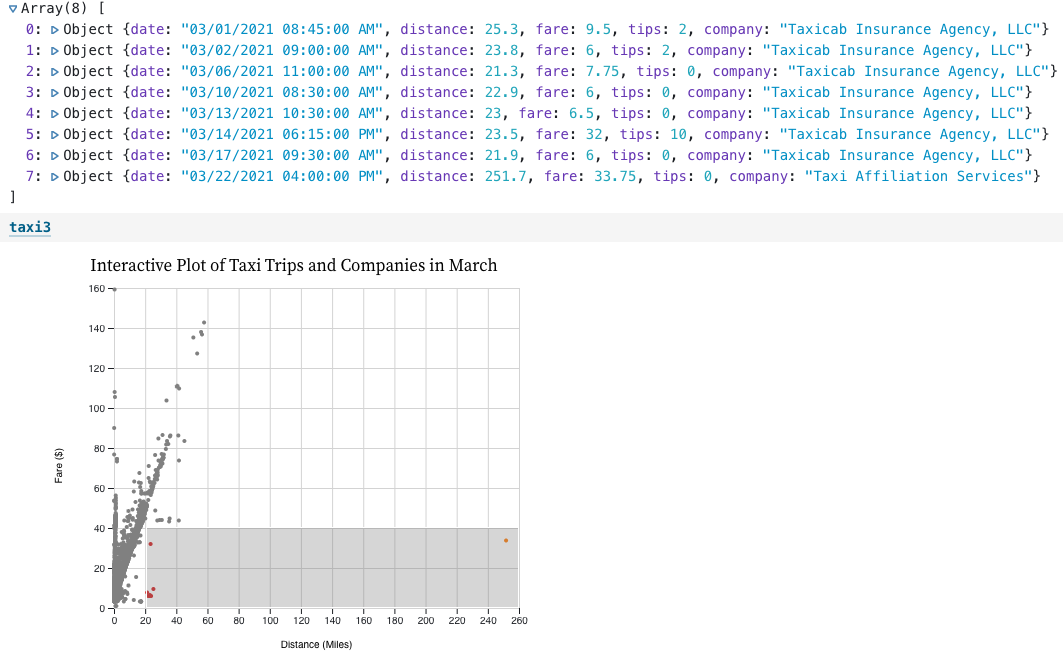


Figure 5. Interactive plot of distance, fare and company of each trips in March

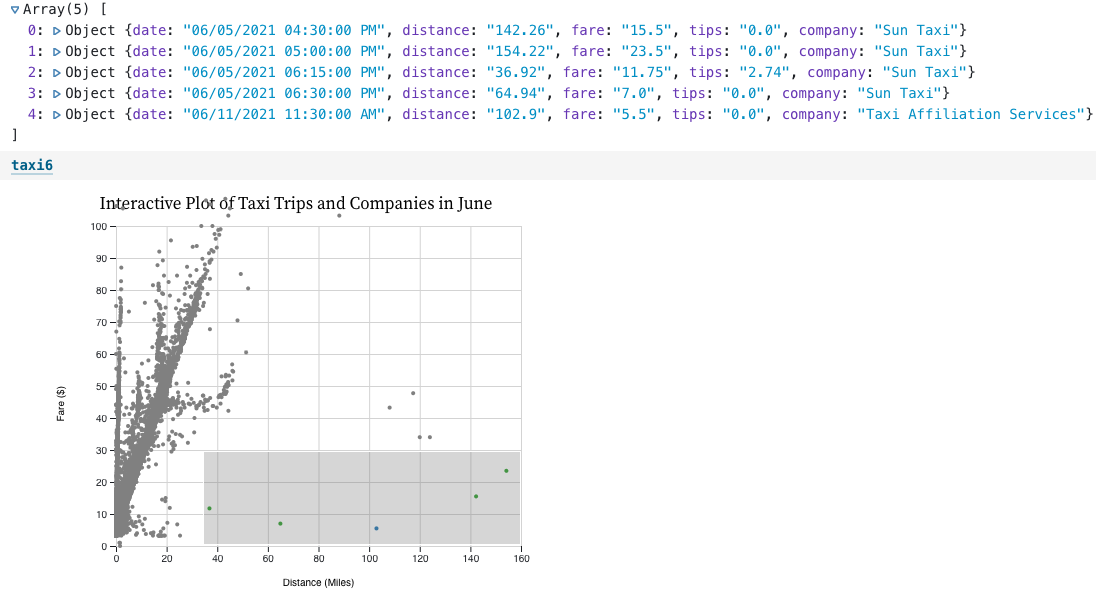


Figure 6. Interactive plot of distance, fare and company of each trips in June

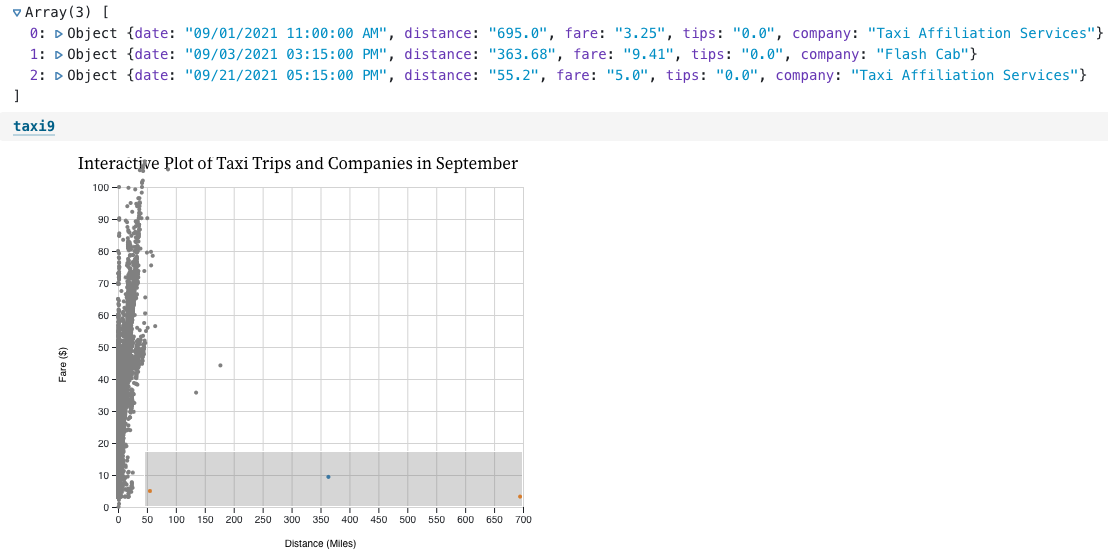
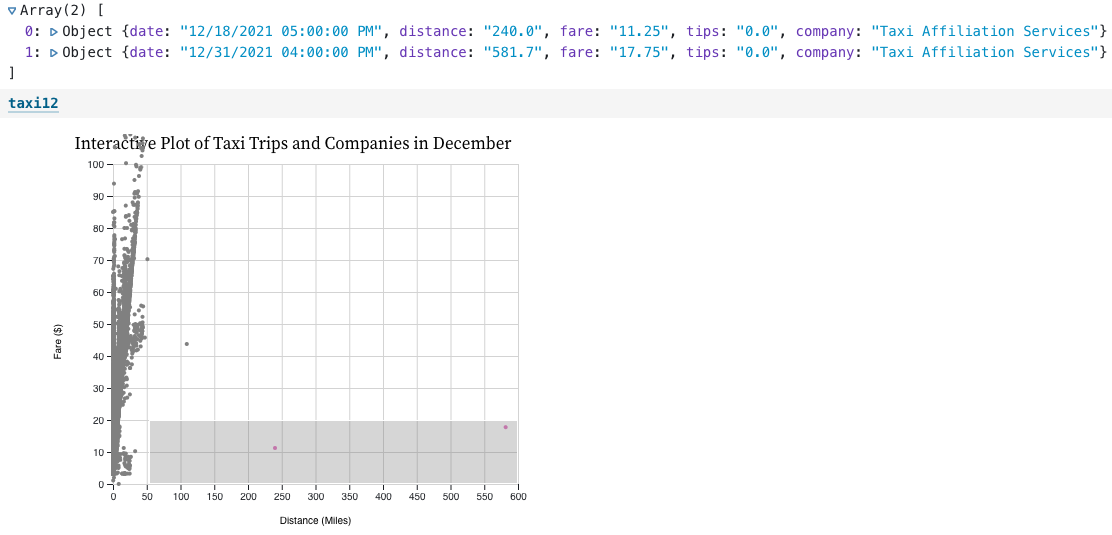


Figure 7. Interactive plot of distance, fare and company of each trips in September

Figure 8. Interactive plot of distance, fare and company of each trips in December

In addition, we also use the attributes fare, tips and company as x-axis, y-axis, and scale of dots to created interactive and brushable scatterplots of each trips in March, June, September and December in Figure 9, 10, 11, and 12. However, there is a large number of trips which caused some plots out of the svg. We tried changing the margins and scales, but it does not work. Since we will focus on the dots inside the svg to analysis the company with lower fare and higher tips, we would ignore the dots out of the svg.

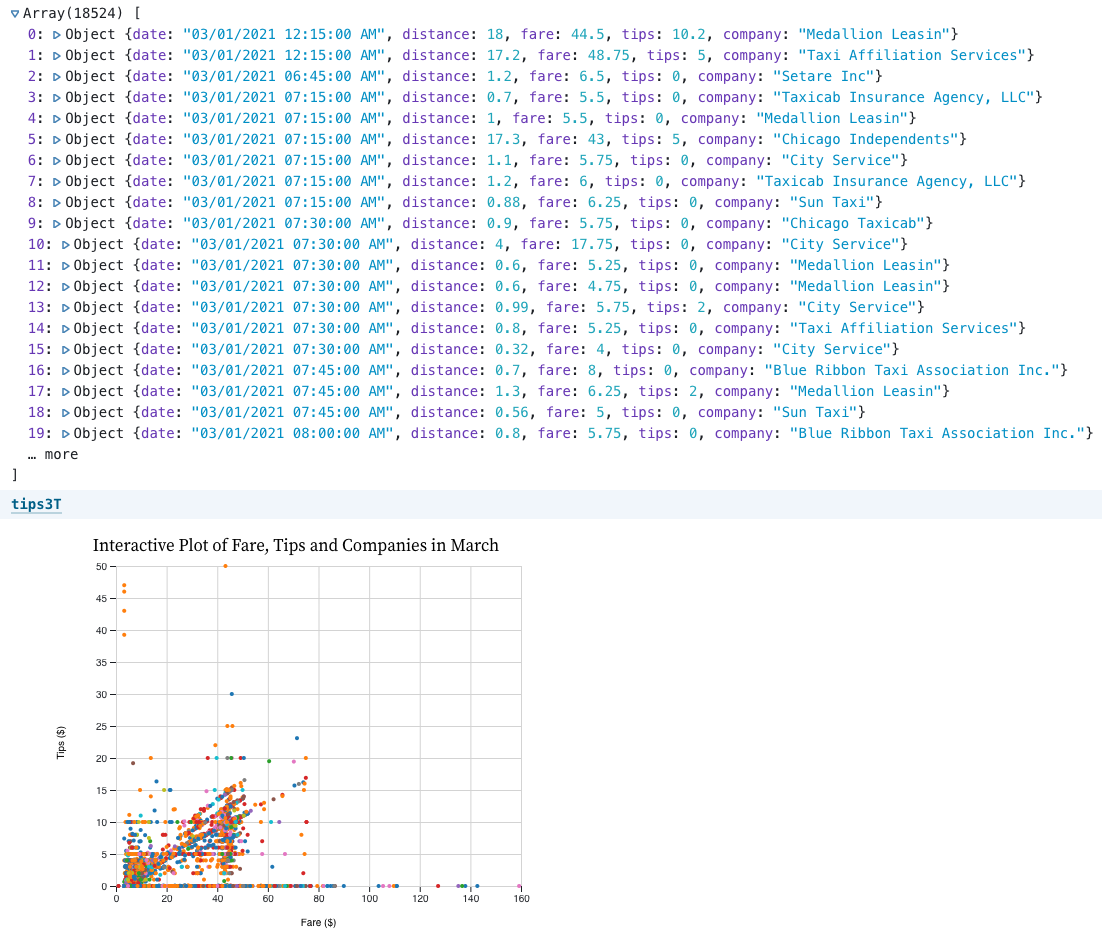


Figure 9. Brushable Scatterplot with fare, tips and taxi company in March

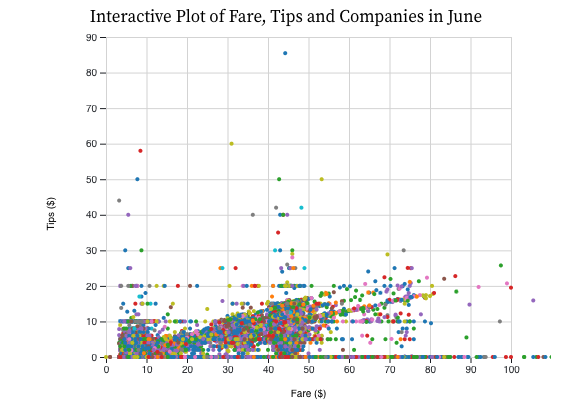


Figure 10. Brushable Scatterplot with fare, tips and taxi company in June

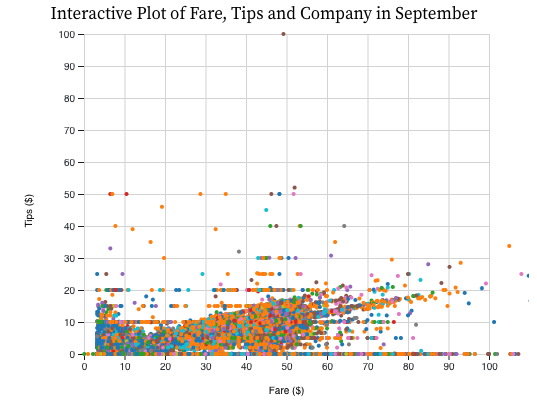


Figure 11. Brushable Scatterplot with fare, tips and taxi company in September

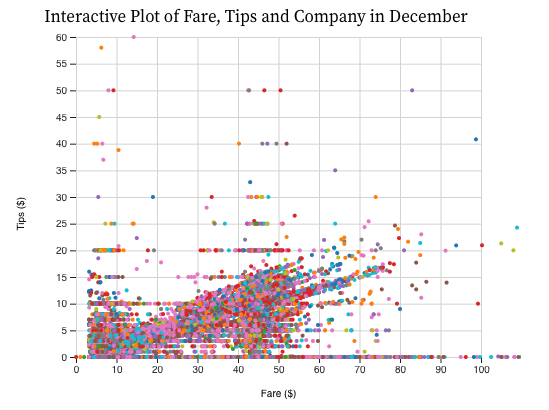
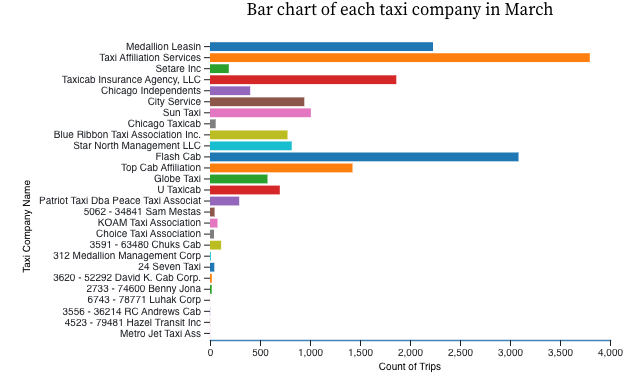
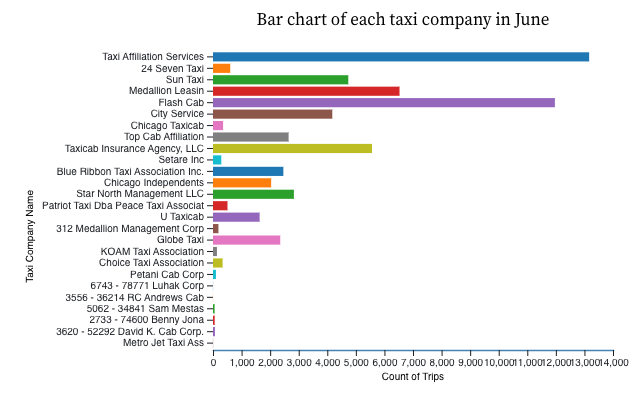


Figure 12. Brushable Scatterplot with fare, tips and taxi company in December

In order to analysis the service of taxi company, we also created four bar charts represented the total number of trips of each taxi company in March, June, September and December in Figure 13, 14, 15 and 16. Along with plotting bar charts, we found that the company with highest trips in March, June, September and December were Taxi Affiliation Service. However, we combined four brushable scatter plots with four bar charts as two linked views for March, June, September and December.

Figure 13. Bar chart represents the total number of trips of each taxi company in March

Figure 14. Bar chart represents the total number of trips of each taxi company in June

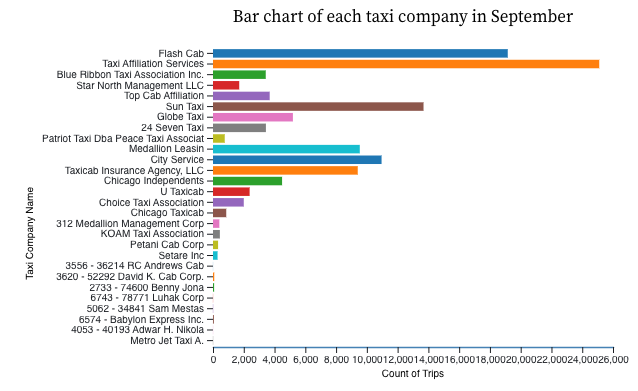
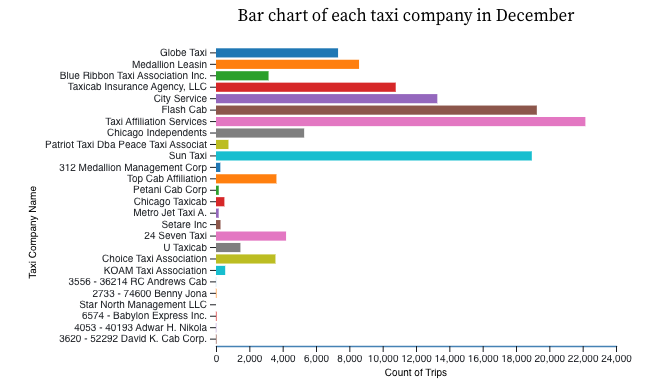


Figure 15. Bar chart represents the total number of trips of each taxi company in September

Figure 16. Bar chart represents the total number of trips of each taxi company in December

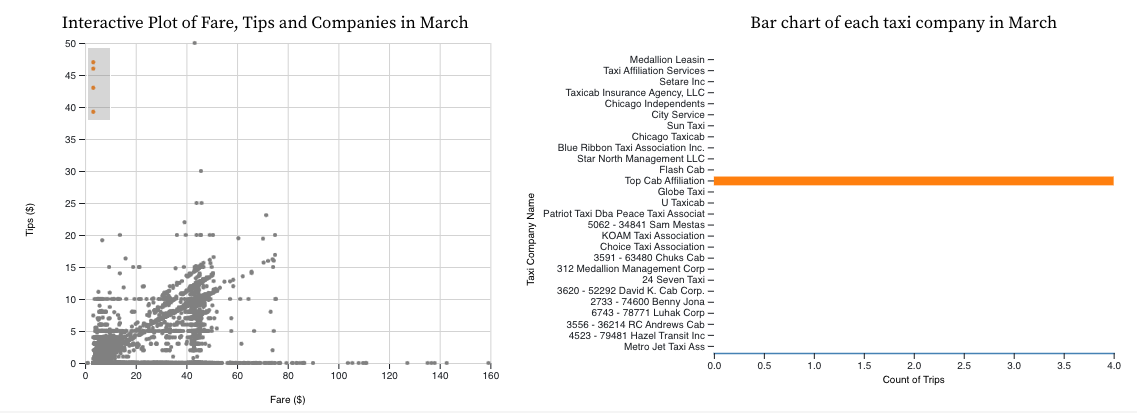
Since we knew the tipping distribution was based on the trips distance and fare from the first question in Figure 2. If the distance and fare of taxi trips increase, the tips also increase. We can estimate riders would pay more tips with lower fare because of the better service of taxi company. Therefore, we tried to hover the dots dots in higher tips and lower fare in scatter plots and obtained the company with highest number of trips from bar charts in Figure 17, Figure 18, Figure 19, and Figure 20. By comparing the two-linked-views of trips in March, June, September and Decemberin, we knew that the companies with better service than other competitors were Top Cab Affiliation, Taxi Affiliation Service and Medallion Leasin, Taxi Affiliation Service and Taxi Affiliation Service in March, June, September and December in 2021. 

Figure 17. Two Linked Views of taxi company with highest number of trips in March

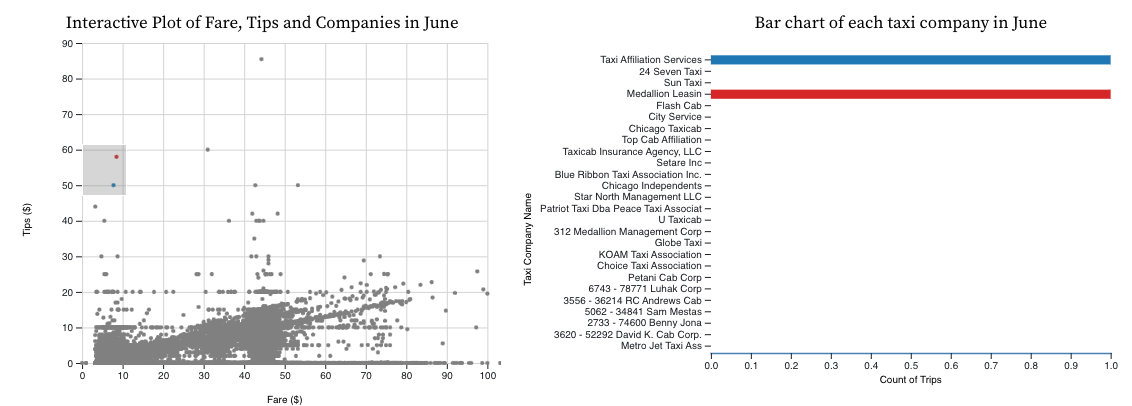


Figure 18. Two Linked Views of taxi company with highest number of trips in June

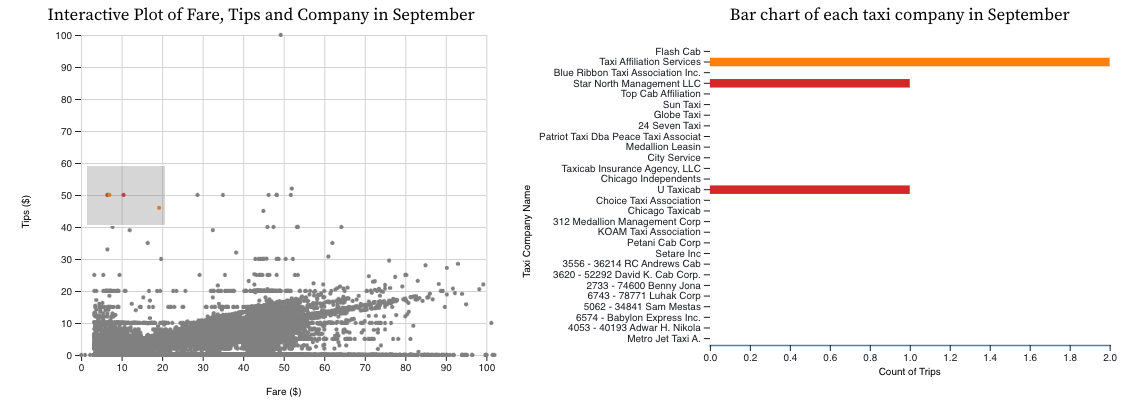


Figure 19. Two Linked Vies of taxi company with highest number of trips in September

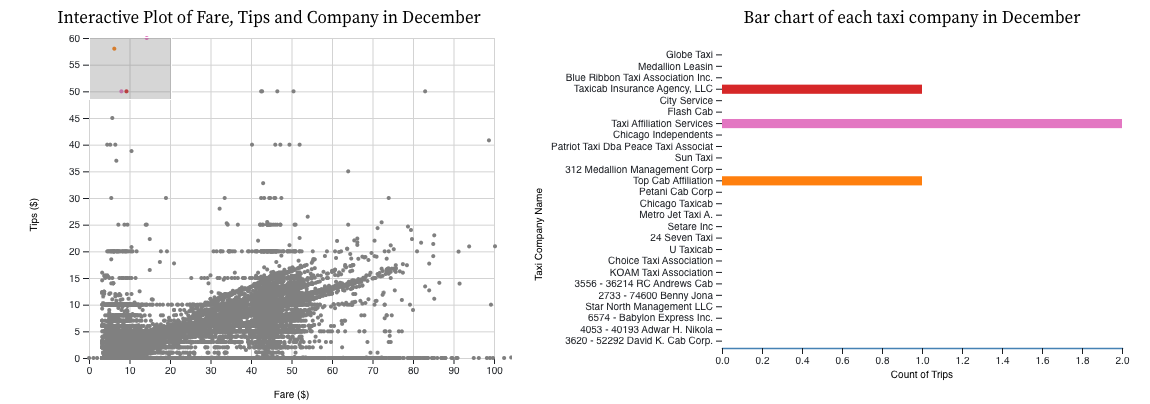


Figure 20. Two Linked Vies of taxi company with highest number of trips in December

Within these findings, we can answered the three questions we created. October and December have higher number of trips in 2021 and the tipping increase if the trips had long distance and higher fare. Along with comparison between taxi trips in March, June, September and December, the taxi company Taxi Affiliation Service has highest number of trips because of favorable fare and better service than other competitors in 2021.