
Analyzing Uber Rides in New York City

Harry Chapman
Ahmed Gurbuz
Akuete Kpodar
Nahielys La Fontaine
Rebekah Vinson

Contents

- Summary
- Data Preparation & Exploration
- Hypotheses
- Data Analysis & Data Visualizations



Summary

- Uber has become hugely popular in New York and its trips outpaced yellow taxis for the first-time last year.
- There are about 65,000 vehicles affiliated with Uber in the city, which provide more than 400,000 trips per day, according to the Taxi and Limousine Commission.
- The objective of this study is to provide an analysis of Uber rides in New York City.

Contents

- Data is retrieved from Amazon Web Services
- Data set includes the trips occurred in New York City between September 2014 and September 2015
- Data set consists of 31 million rows (1.5 GB)
- Random sample of 100 thousand rows of data is used for analysis
- 4078 rows of data from random sample is ignored because of missing values in the dataset
- No duplicate values are found in the dataset
- In total, 4% of the records are ignored and 96% of original data is retained for data analysis



Questions

Hypotheses

Tip

A hypothesis is a proposed explanation made on the basis of limited evidence



→ **What date of the year in 2015 are Uber trips requested more?**

Hypothesis 1: There will be more trips per day on Christmas and New Year's Eve in NYC

→ **Do Uber trips occur more often in the evening hours?**

Hypothesis 2: There will be a higher prevalence of Uber trips after 5:00 pm

→ **Do Uber trips occur more in the summer months?**

Hypothesis 3: There will be a higher prevalence of Uber trips in the month of July

→ **What is the relationship between trip distance and trip duration?**

Hypothesis 4: There is a positive correlation between trip distance and trip duration in NYC

Hypothesis 5: 95% of the variance in trip duration is explained by trip distance in NYC



→ What are popular pick-up and drop-off locations?

Hypothesis 6: Some pick-up and drop-off locations will be more popular to others

→ What is the relationship between staying in the same neighborhood and distance?

Hypothesis 7: There is a strong association between distance travelled and whether you stay or leave the neighborhood

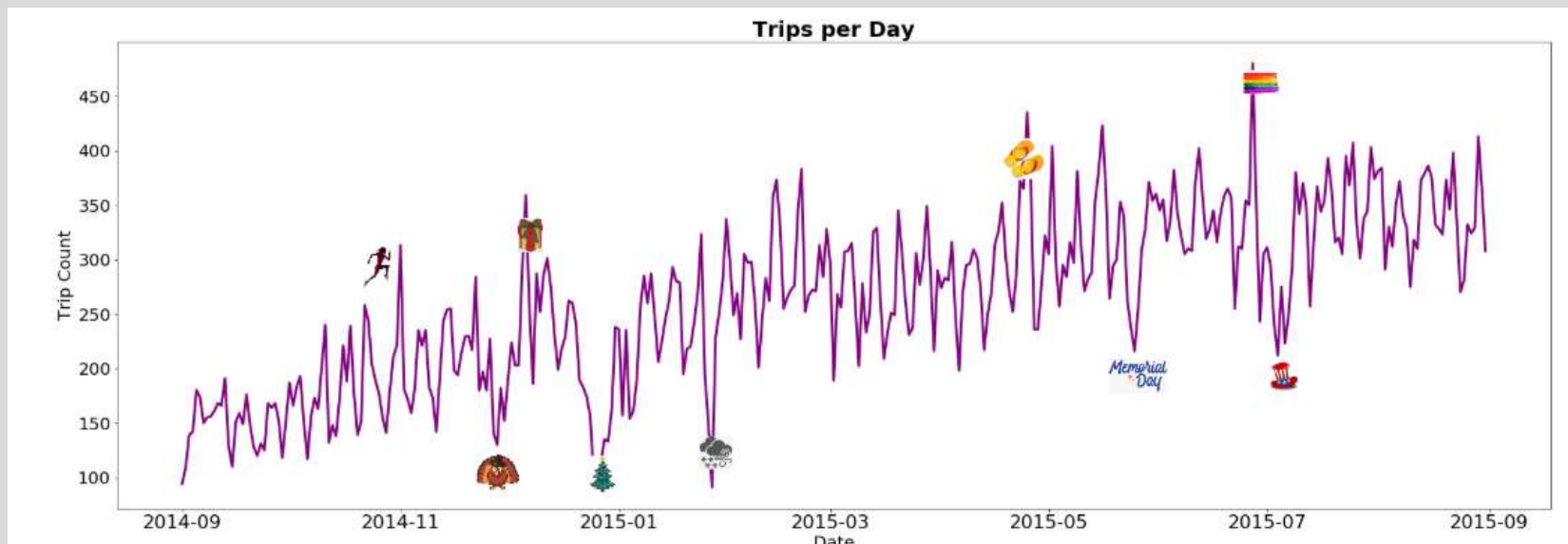
→ What is the relationship between leaving the neighborhood and distance?

Hypothesis 8: There is a negative correlation between staying in the neighborhood and the distance traveled

→ What is the difference between the average trip distance on New Year's Eve and rest of the year?

Hypothesis 9: Trip distance on New Year's is longer than average trip distance in NYC

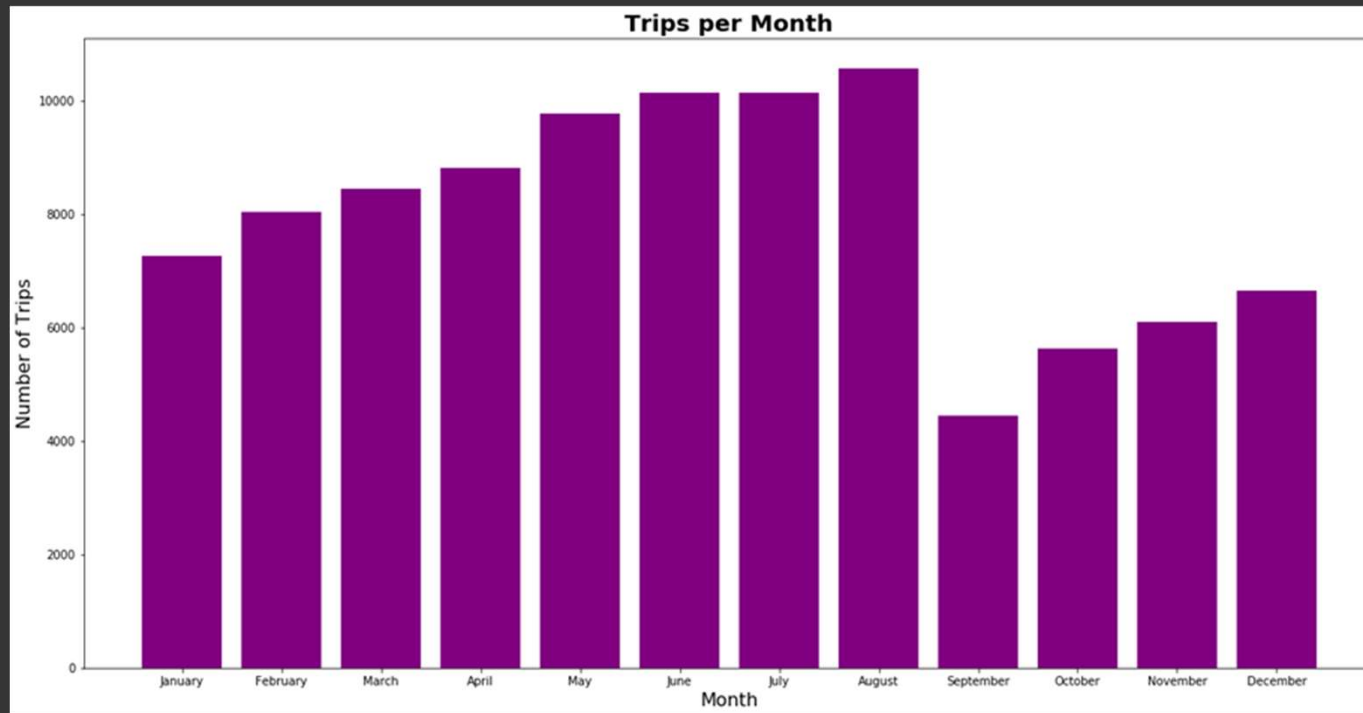
Most Uber trips per day in 2015



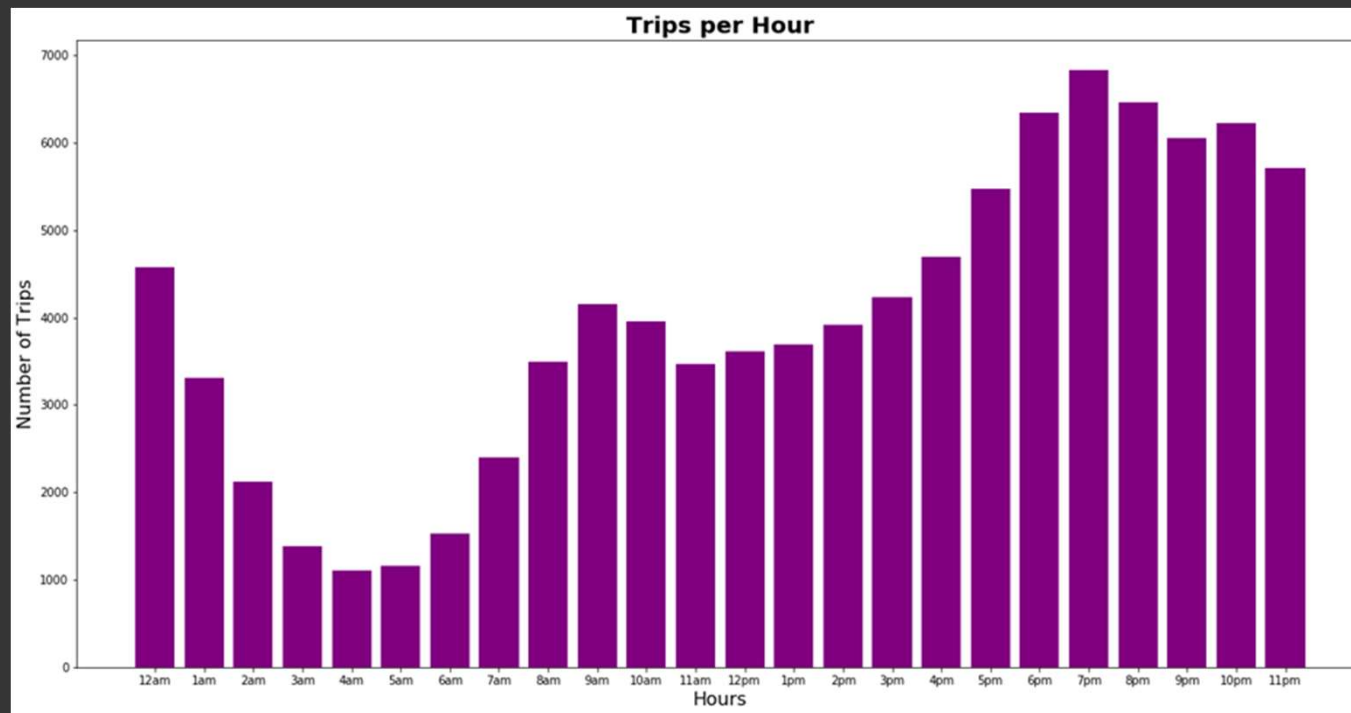
`matplotlib.offsetbox import (OffsetImage, AnnotationBbox)` to create the image box and `matplotlib._png`
`import read_png` to load icons (.png file) to (x,y) position

Most trips per day in 2015 was June 27th. On June 26th a historic US Supreme court decision legalized gay marriage. Uber NYC marched along with family and friends and offered free rides to the Pride Day parade.

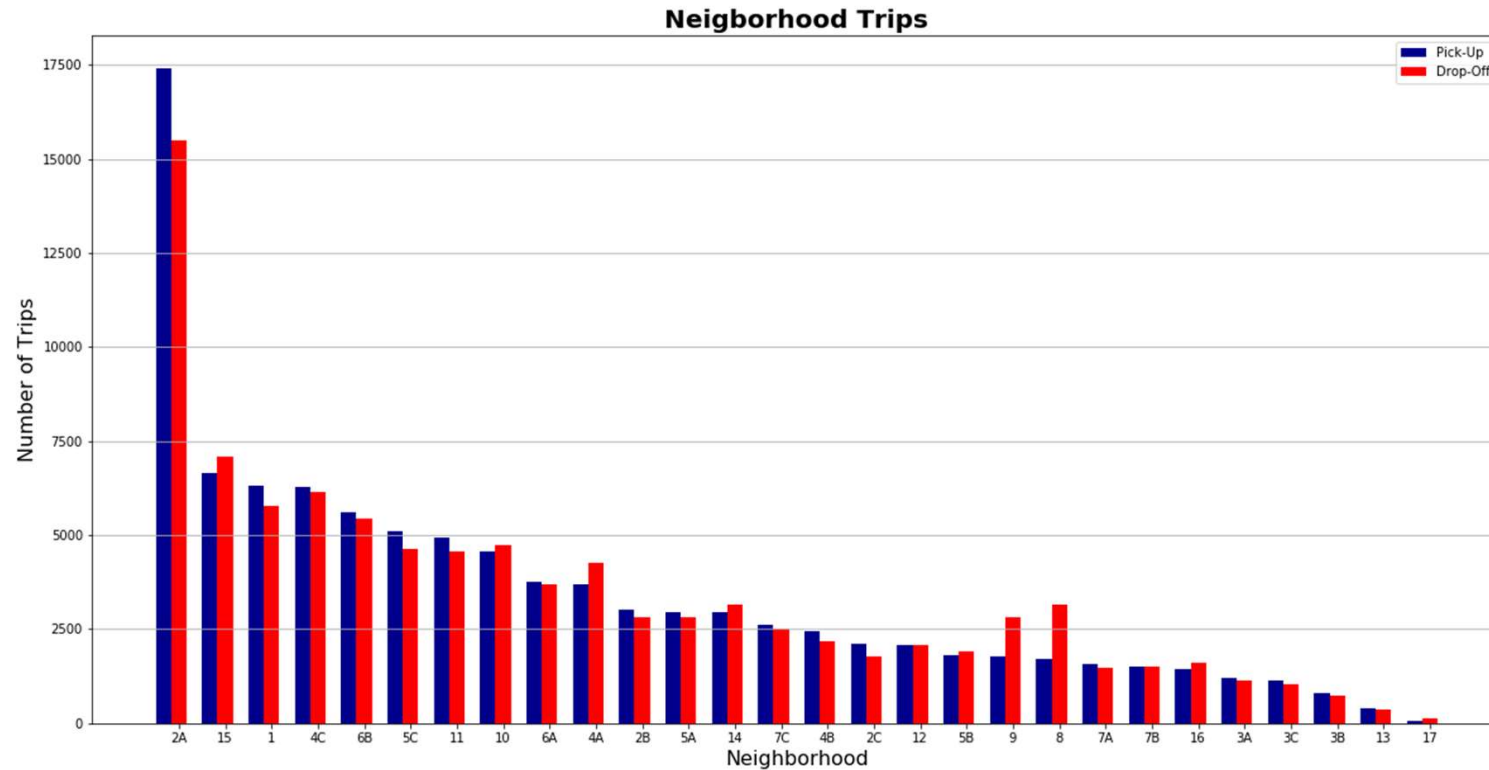
The highest number of trips occurred in the month of August



The highest number of trips occurred between 7pm and 8pm



— Popular Locations

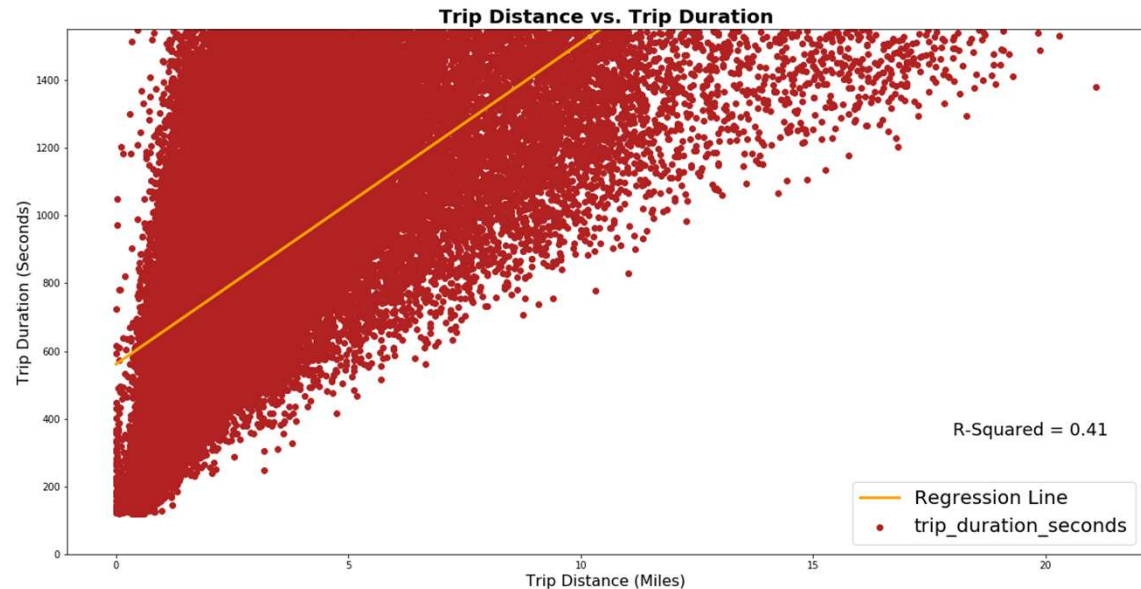


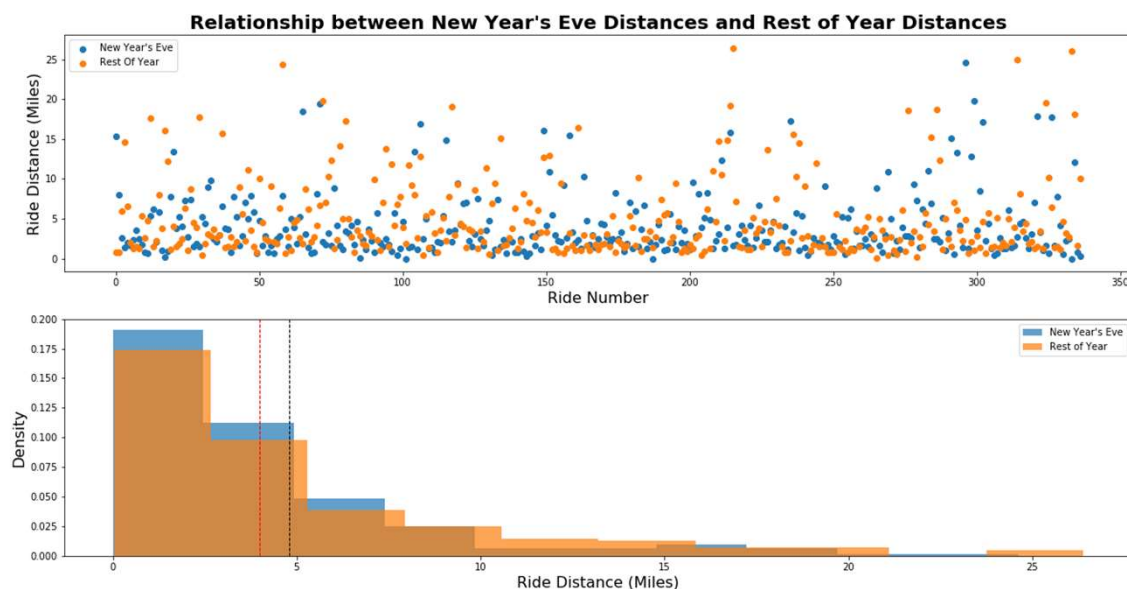
- The number of unique locations in the data set is 28.
- "2A" is the most popular destination with the highest number of trips.



The correlation between Distance and Duration is Significant but has a low R^2

The correlation between trip distance and duration was surprisingly low – at an R^2 of 0.41, the regression line bisects the data set and indicates that only 41% of trip duration is explained by trip distance





Number of trips occurred on New Year's Eve
337

Average daily number of trips rest of the year
263

$$\mu_1 = 3.99$$

$$\mu_2 = 4.80$$

Significance level

$$\alpha = 0.05$$

p-value = 0.02

$$(p < \alpha)$$

we reject H_0

$H_0: \mu_1 = \mu_2$ ("Average Uber trip distances occurred on New Year's Eve and rest of the year are equal")

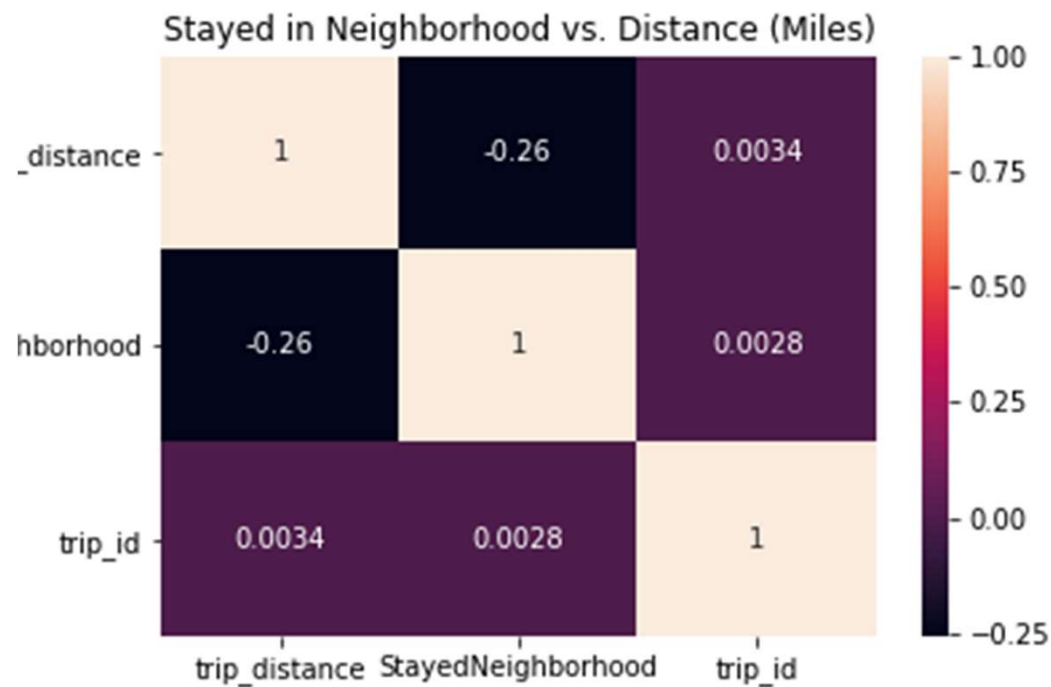
$H_1: \mu_1 \neq \mu_2$ ("Average Uber trip distances occurred on New Year's Eve and rest of the year are not equal")

We have sufficient evidence to conclude that average Uber trip distances occurred on New Year's Eve and rest of the year are not equal at 0.05 significance level.



There seem to be a slight positive correlation between “leaving the neighborhood” and “the distance traveled” which proves that there is some measure of dependence between the two variables. The value is not too close to 1 which signify that there is no strong positive correlation.





Conversely the relationship between “Staying in the Neighborhood” and the “Distance” has a negative correlation which signify the distance traveled is related to Staying in the neighborhood.



– Conclusions

- We need data sets from 2016, 2017 and 2018 and compare the same holidays, weather and cultural events. If each year has the same min days holidays or weather events, my hypothesis will continue to be invalid. Further research is needed to determine why these days are lowest days for Uber trips per day.
- There is a higher prevalence of Uber trips between 7:00pm – 8:00pm
- There is a higher prevalence of Uber trips in the month of August
- There is a positive correlation between trip distance and trip duration in NYC
- 95% of the variance in trip duration is explained by trip distance in NYC
- “2A” is a the most popular destination with the highest number of trips as pick-up and drop-off location
- Average trip distance on New Year’s Eve is shorter than daily average trip distance rest of the year

Appendix

Ride with Pride

https://www.youtube.com/watch?v=_57BDy5C9LA

Uber

Blizzard of 2015: Governors declare states of emergencies

<http://www.msnbc.com/msnbc/historic-snowstorm-expected-throughout-northeast>

01/26/15 08:23 AM—UPDATED 01/26/15 04:13 PM

NYC Pride March 2015: Photos of the historic celebration

AMNewYork

June 29, 2015 12:20 PM

B.Chappell NPR

Supreme Court Declares Same-Sex Marriage Legal In All 50 States

June 26, 2015 10:05 AM ET

Uber Logo Vector

Copyright 2008-2019 © seeklogo.com