UBER ANALYSIS:

Pickups in New York City

Introduction

This project aims to use data in order to gain insights about the uber pickups in NYC from April-September 2014.

We would want to use visualization techniques, like those employed by the library “ggplot2”, to understand how data could be used to create business decisions that is beneficial to the company.

Dataset

There are six file of raw data on Uber pickups in NYC from April to September 2014. The files are separated by month and each has the following columns:

- Date/Time: the date and time of the Uber pickup
- Lat: the latitude of the Uber pickup
- Lon: the longitude of the Uber pickup
- Base: the TLC base company code affiliated with the Uber pickup

The dataset is available on:

- Kaggle: https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city?select=uber-raw-data-apr14.csv

Libraries

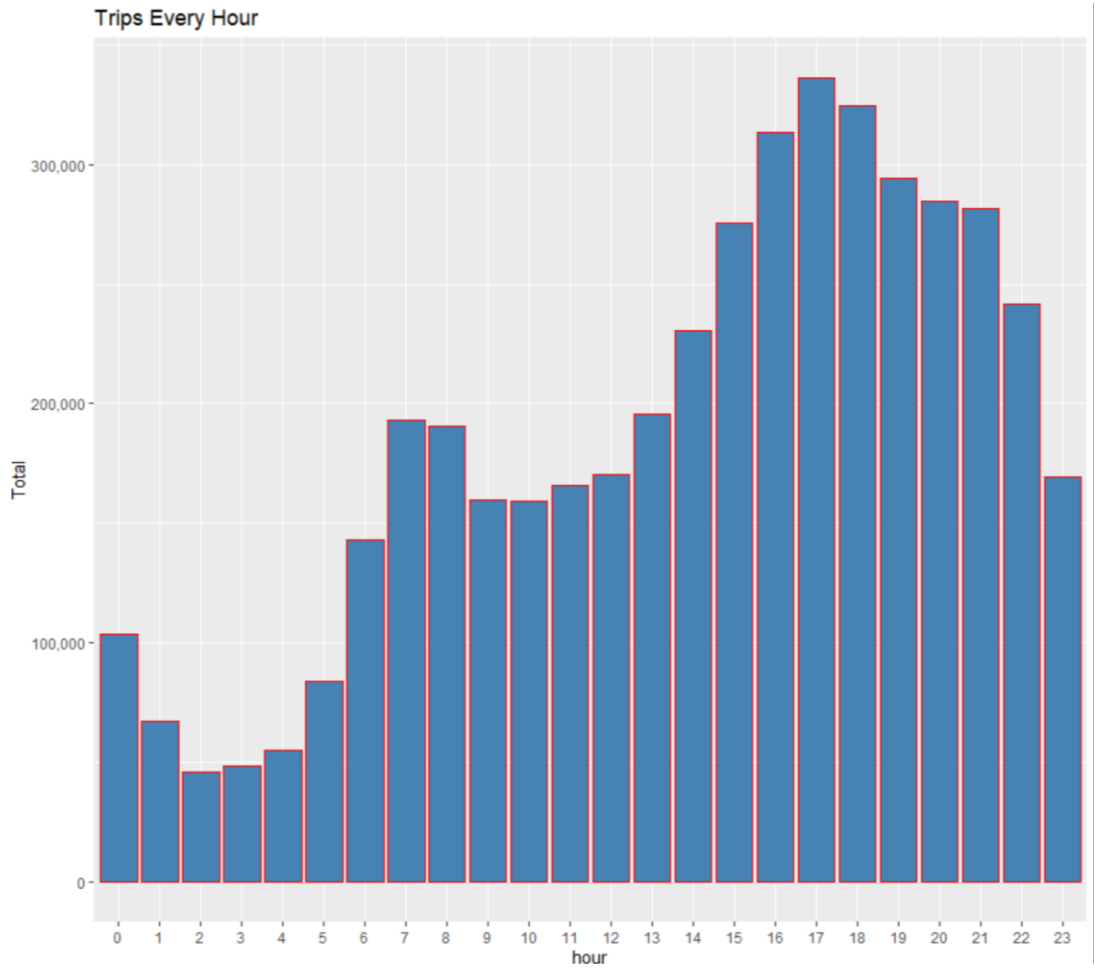
For this project, the following libraries are used:

- ggplot2: for data visualization
- ggthemes: an add-on for ggplot2
- lubridate: make dealing with Dates a little easier
- dplyr: for data manipulation
- tidyr: to tidy the data
- DT: provides an R interface to the JavaScript library
- scales: to map the data to the correct scales

Discerning Which Hour has the Highest # of Trips

Looking at the graph on the right, the 17th hour of the day, 5 pm, has the highest number of trips being made. This could be potentially explained by workers who worked 9-5 jobs. In fact, the range starting from the 14th to 21st hour have the highest numbers. We could use this findings to suggest the company to have more drivers out at these times to accommodate more customers.

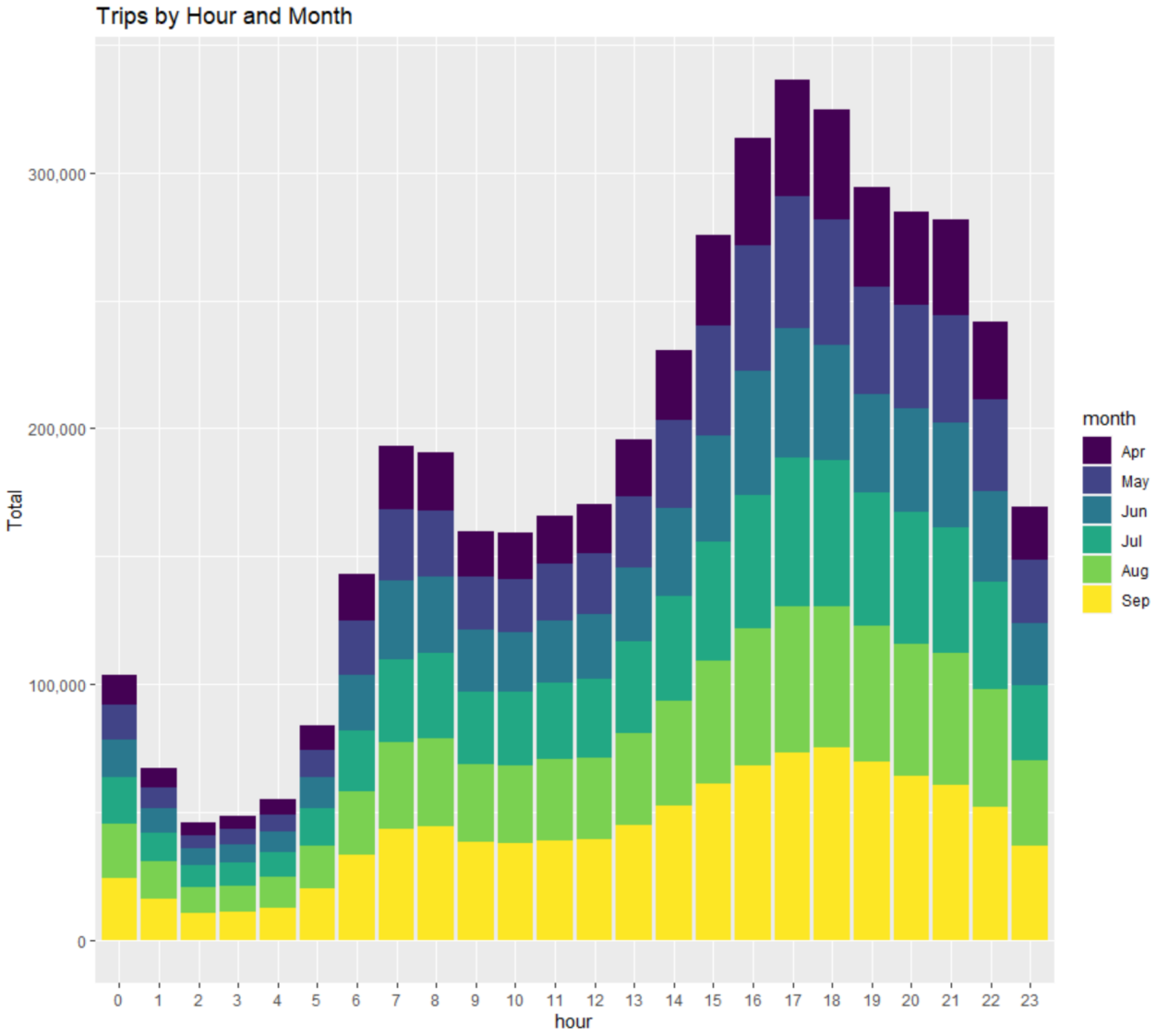
At the same time, we could suggest not a lot of drivers at the 0 – 5 hour range as it has the lowest number of trips.



Distinguish Which Month has the Highest # of Trips

The difference between this graph and the one before is that this one dictates which month specifically obtains the highest number of trips by hour. As seen on the image, it is the month of September.

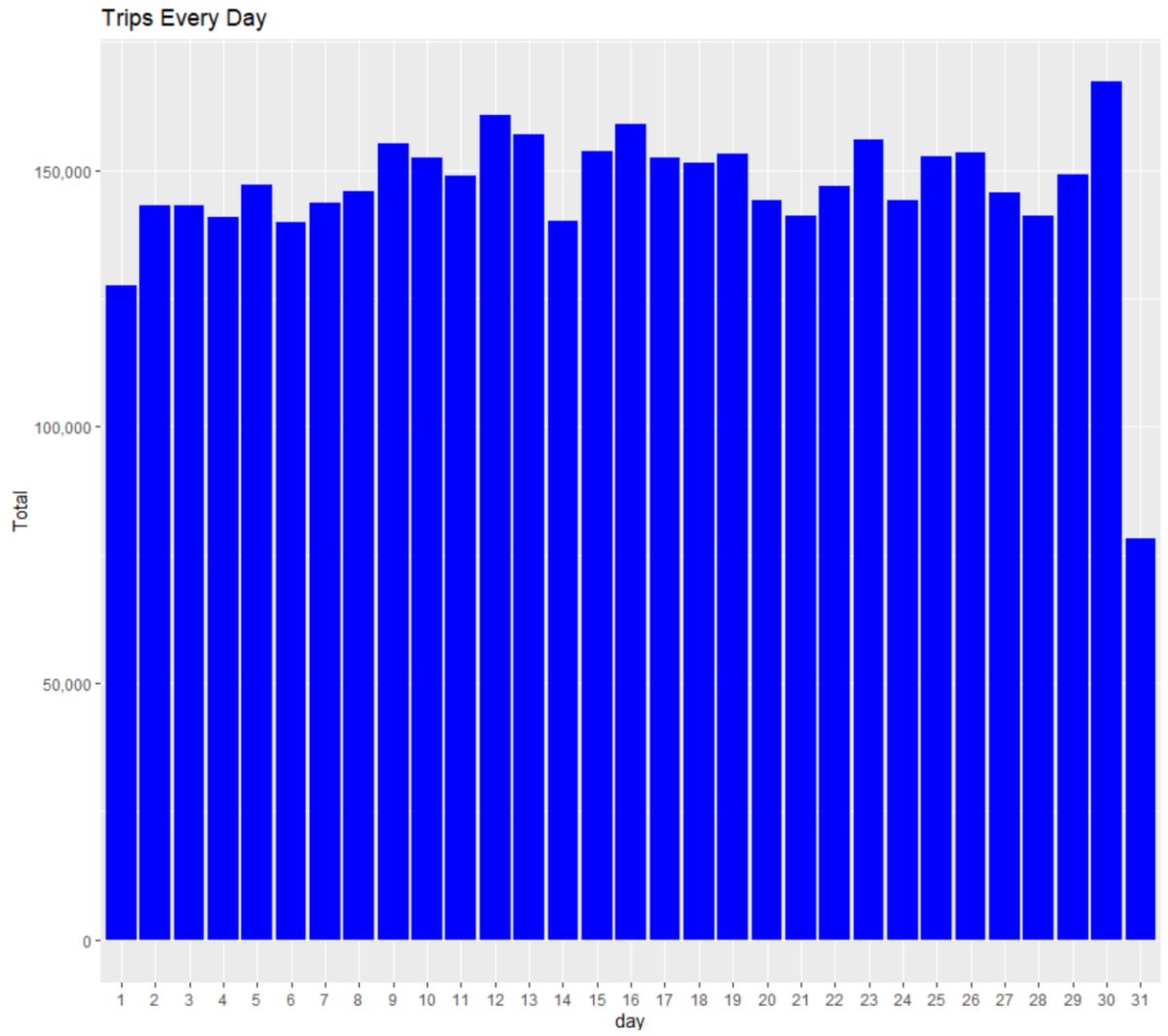
One of the reasons could be because September is when the visitor season starts. This is a time of transition. It's still technically summer until around the 21st, but the city takes on an autumn frame of mind after Labor Day.



Which Day of the Month has the Most Trips?

Looking at the graph, we can see that there is not much variation on the number of trips per day. Though the 31st having the lowest number makes sense as not all months have 31 days on them.

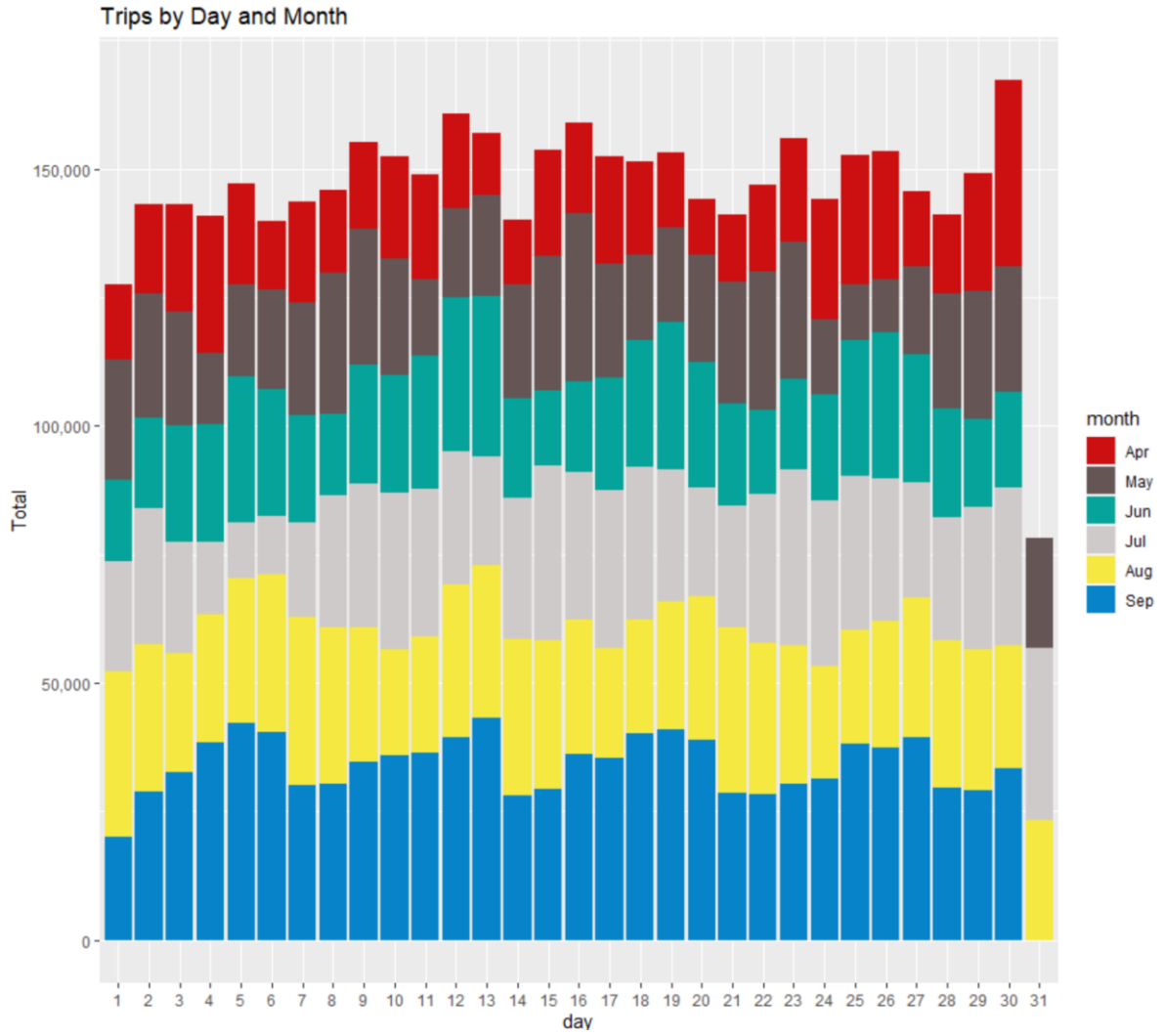
The 30th day having the highest number and the 1st day of the month having the lowest is very interesting.



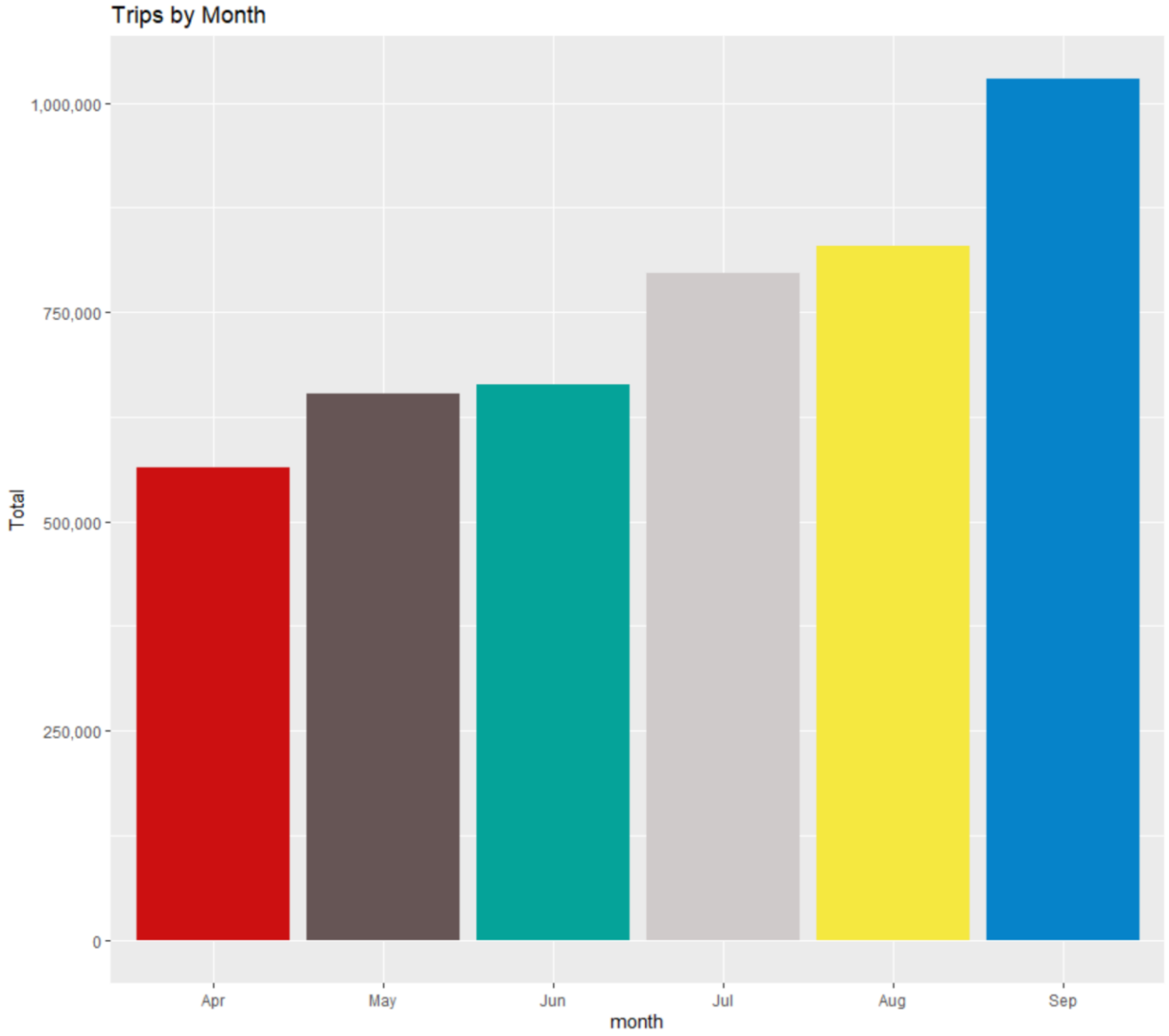
Which Month has the Highest # of Trips per day?

Like the graph on page 6, the month of September has the most number of trips per day with the exception of the following days:

- 1
- 7
- 14
- 15
- 21
- 23
- 24
- 30
- 31

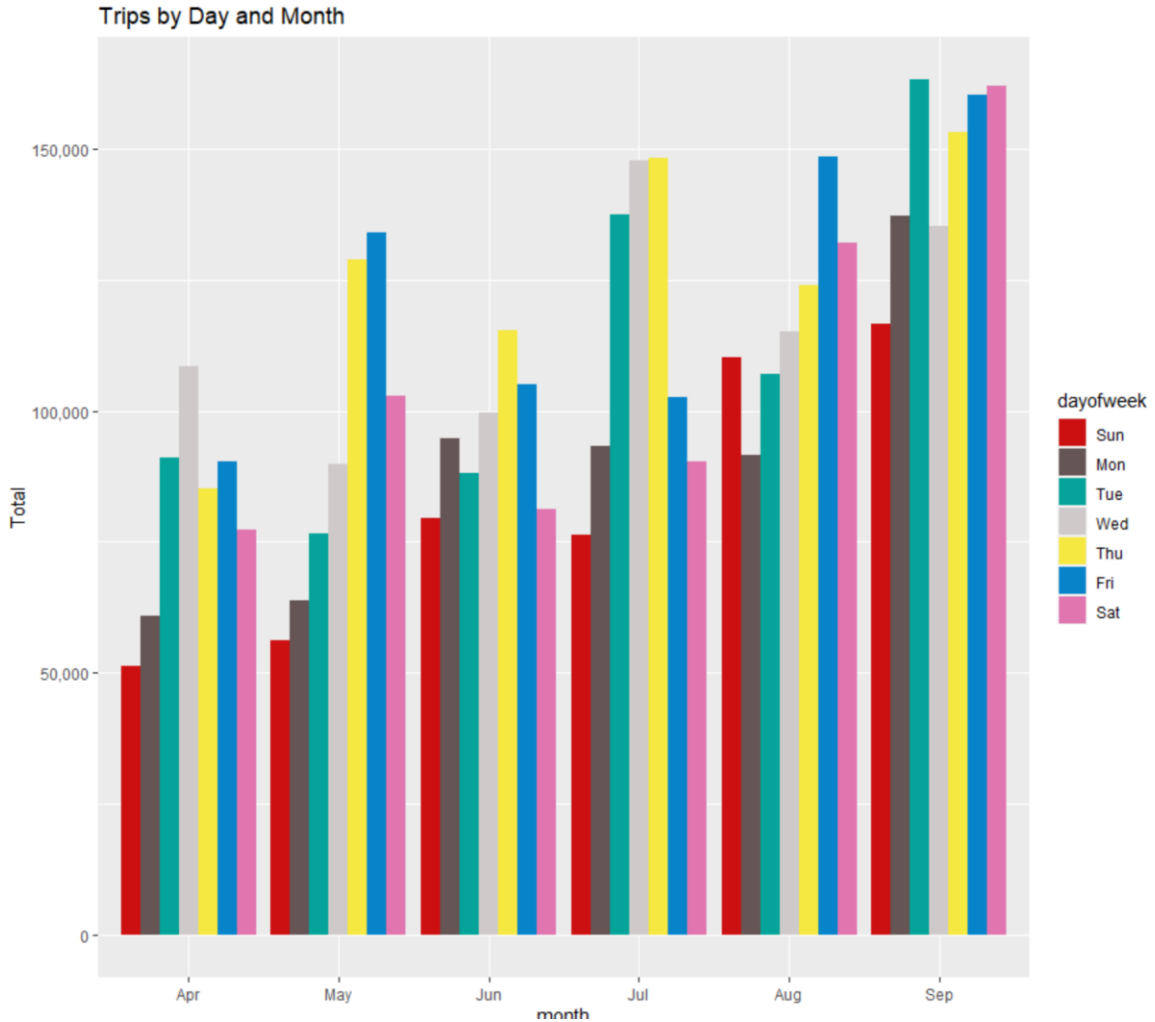


This graph further proves what is stated in pages 6 and 8. September is the month with the highest number of trips.



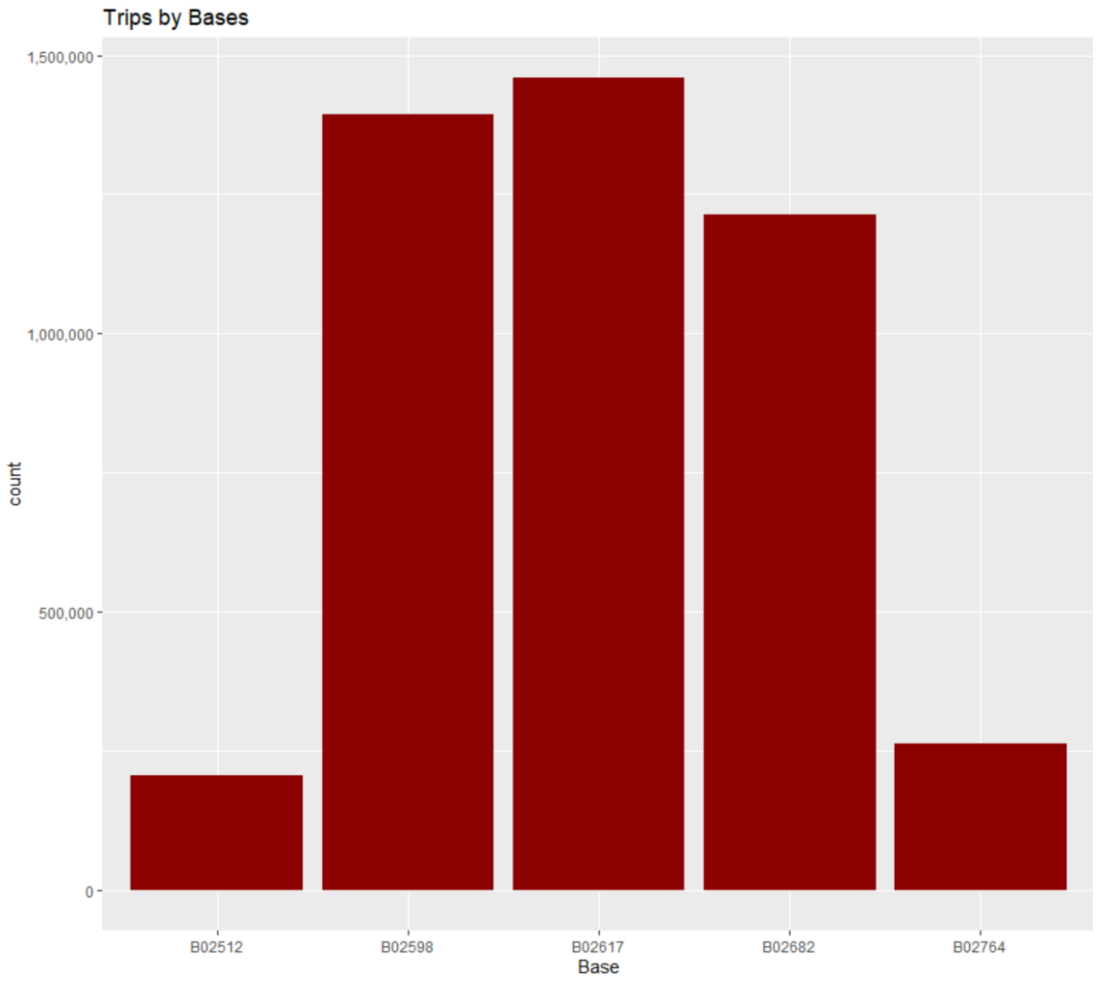
Distinguish which Day of the Week has the highest # of Trips

In the month of April, Wednesday has the most of trips in a week. For May, it is Friday, though Thursday is not far off. It is the opposite for June. For August, it is Tuesday. Lastly, September has the highest number on Tuesday, though Friday and Saturday have nearly the same numbers.



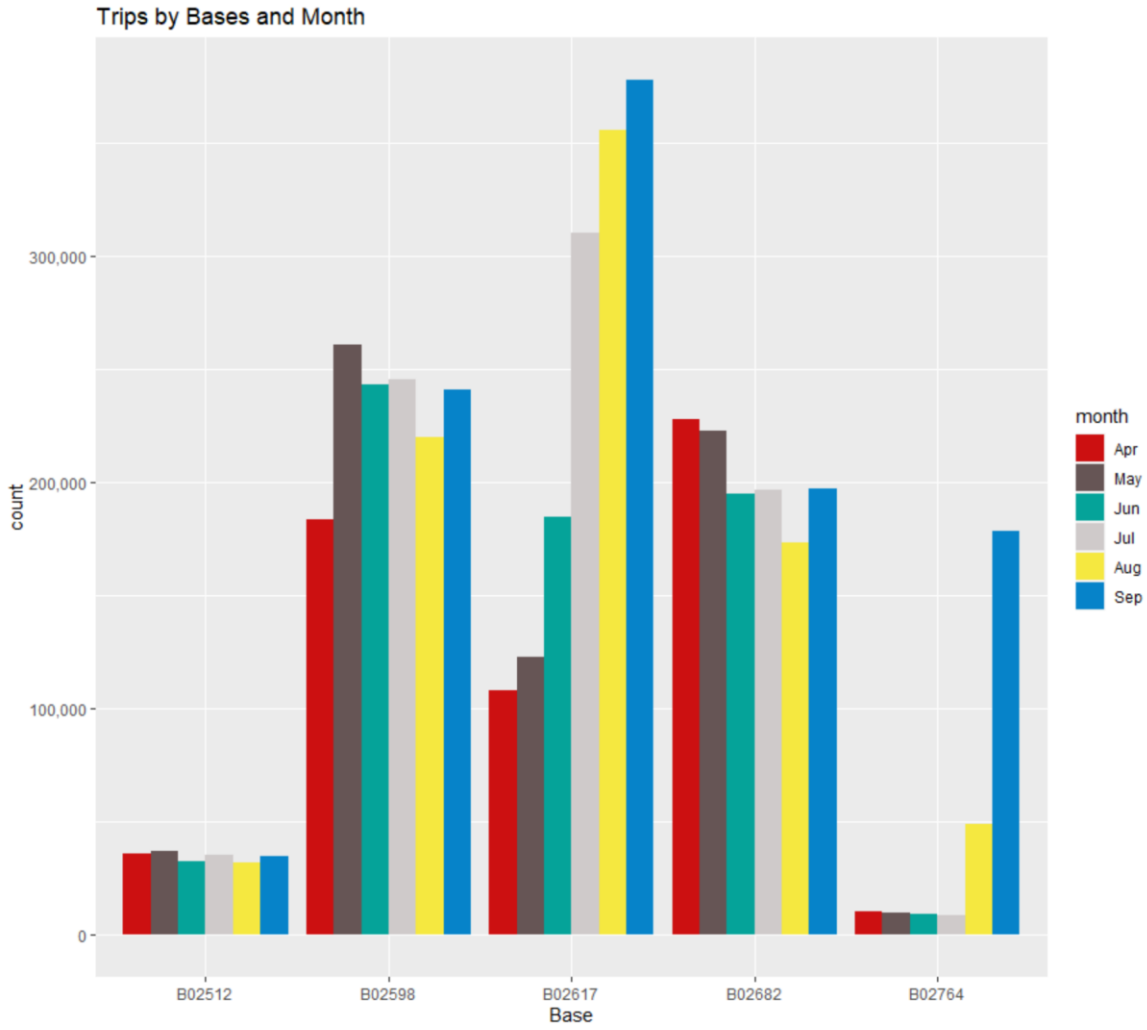
Distinguish Which Base has the highest # of Trips

The base company code B02617 has the highest number of trips followed by B02598 and then B02682. The base codes B02764 and B02512 has the lowest number.



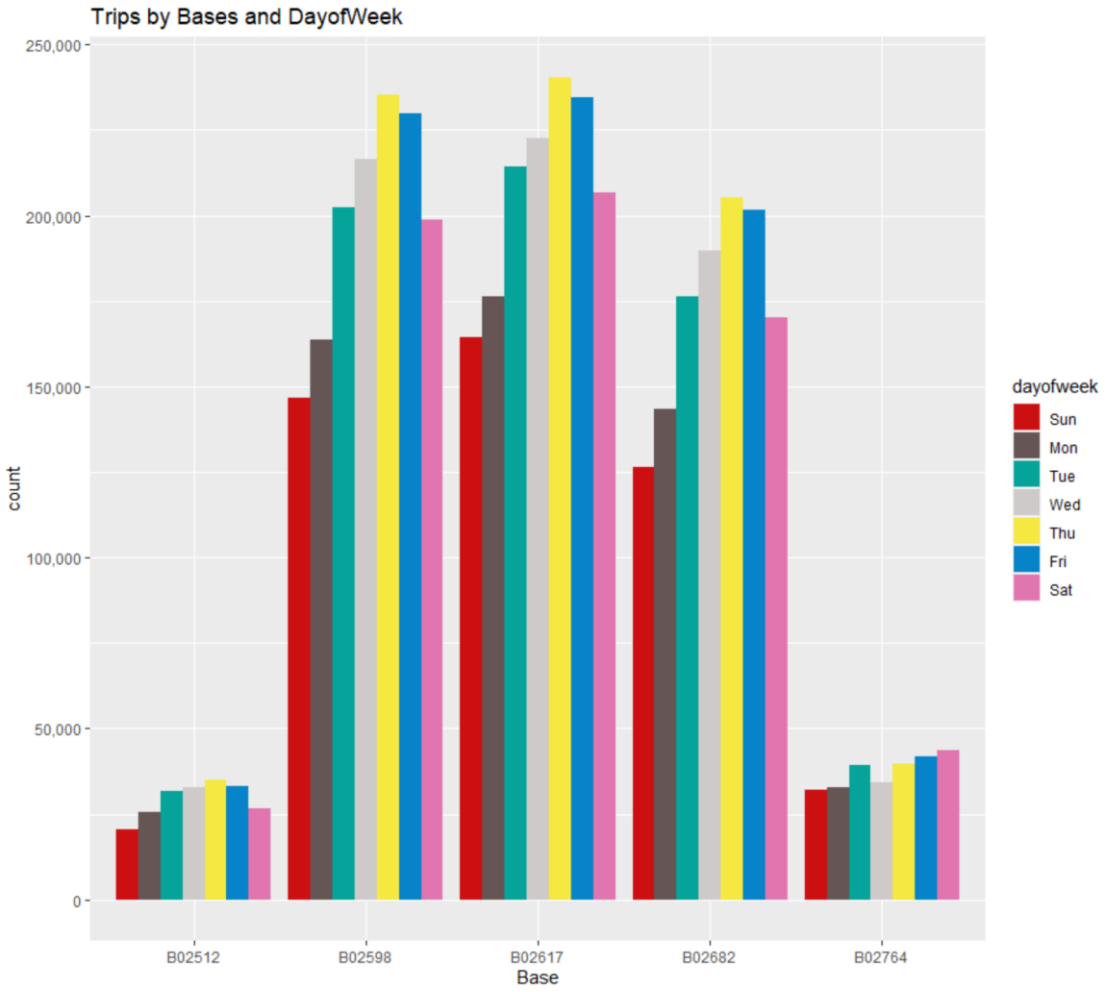
Distinguish Which Month has the Highest # of Pickup Based on Bases

- September has the highest number of pickup on the base company codes B02764 and B02617.
- B02598 has the highest number of trips on the moth of May.
- April is the month with the highest for the base code B02682.
- The base code B02512 has approximately the same number of trips for each month.



Distinguish Which Day of the Week has the highest # of Trips based on Base:

In the month of April, Wednesday has the most of trips in a week. For May, it is Friday, though Thursday is not far off. It is the opposite for June. For August, it is Tuesday. Lastly, September has the highest number on Tuesday, though Friday and Saturday have nearly the same numbers.

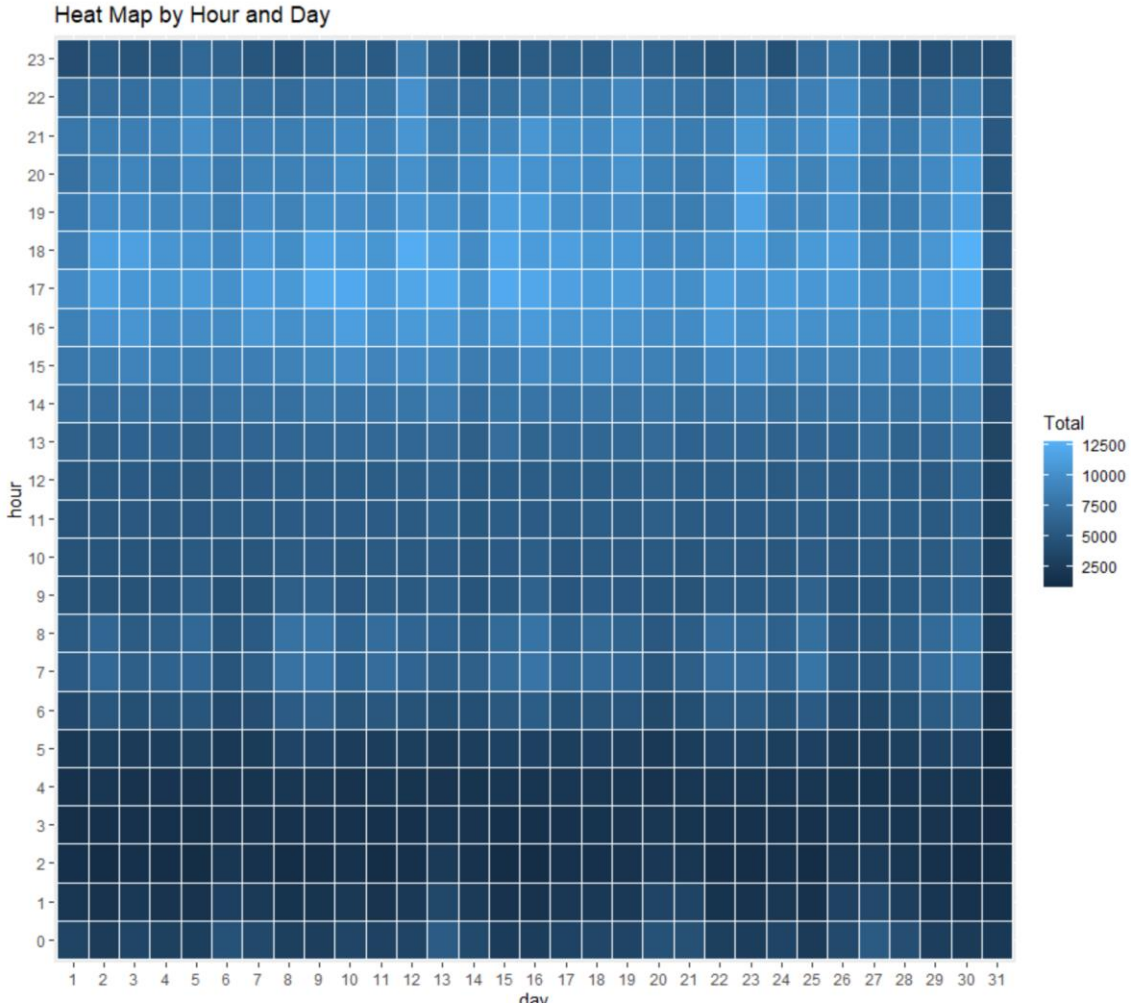


Heat Map: Hour & Day

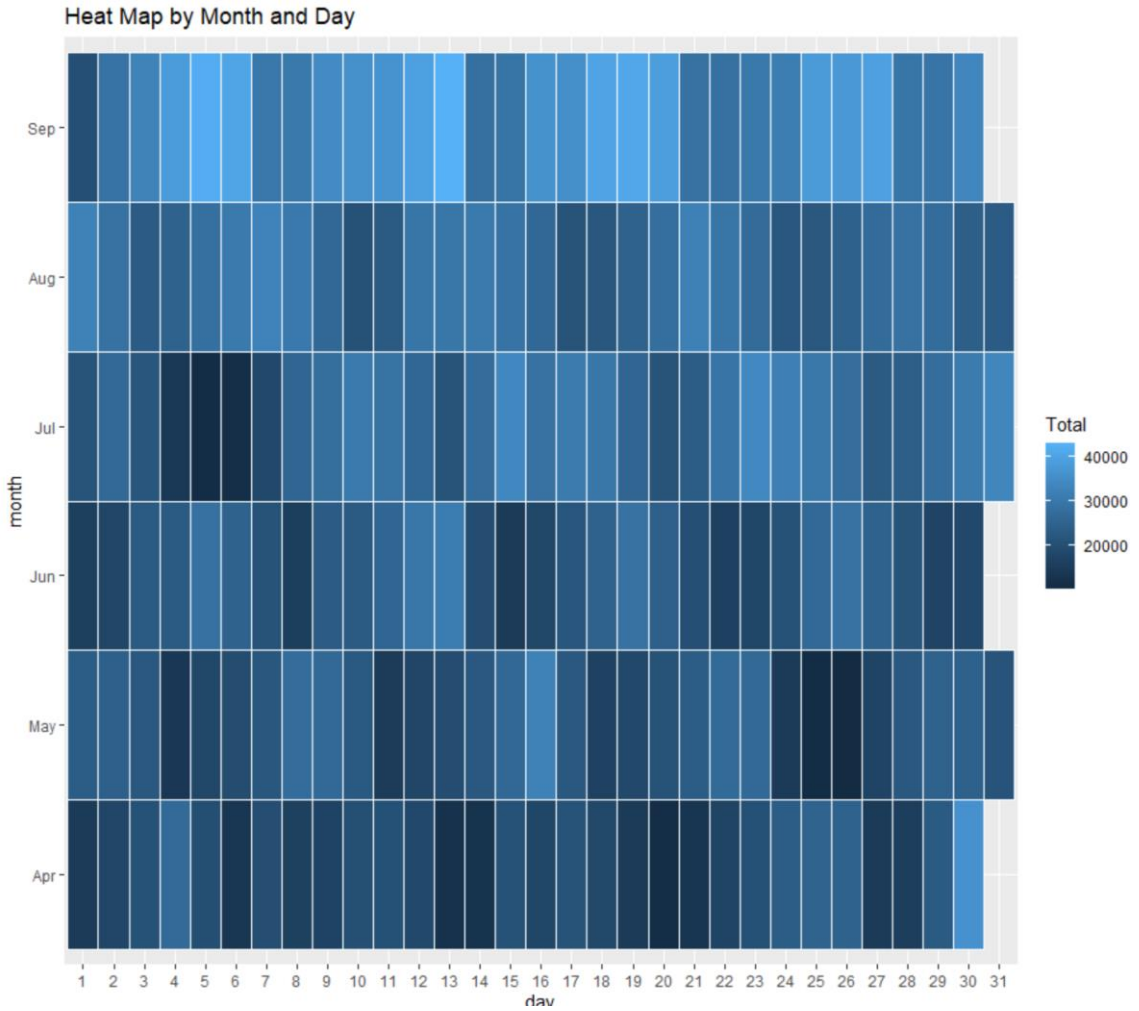
- A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors.

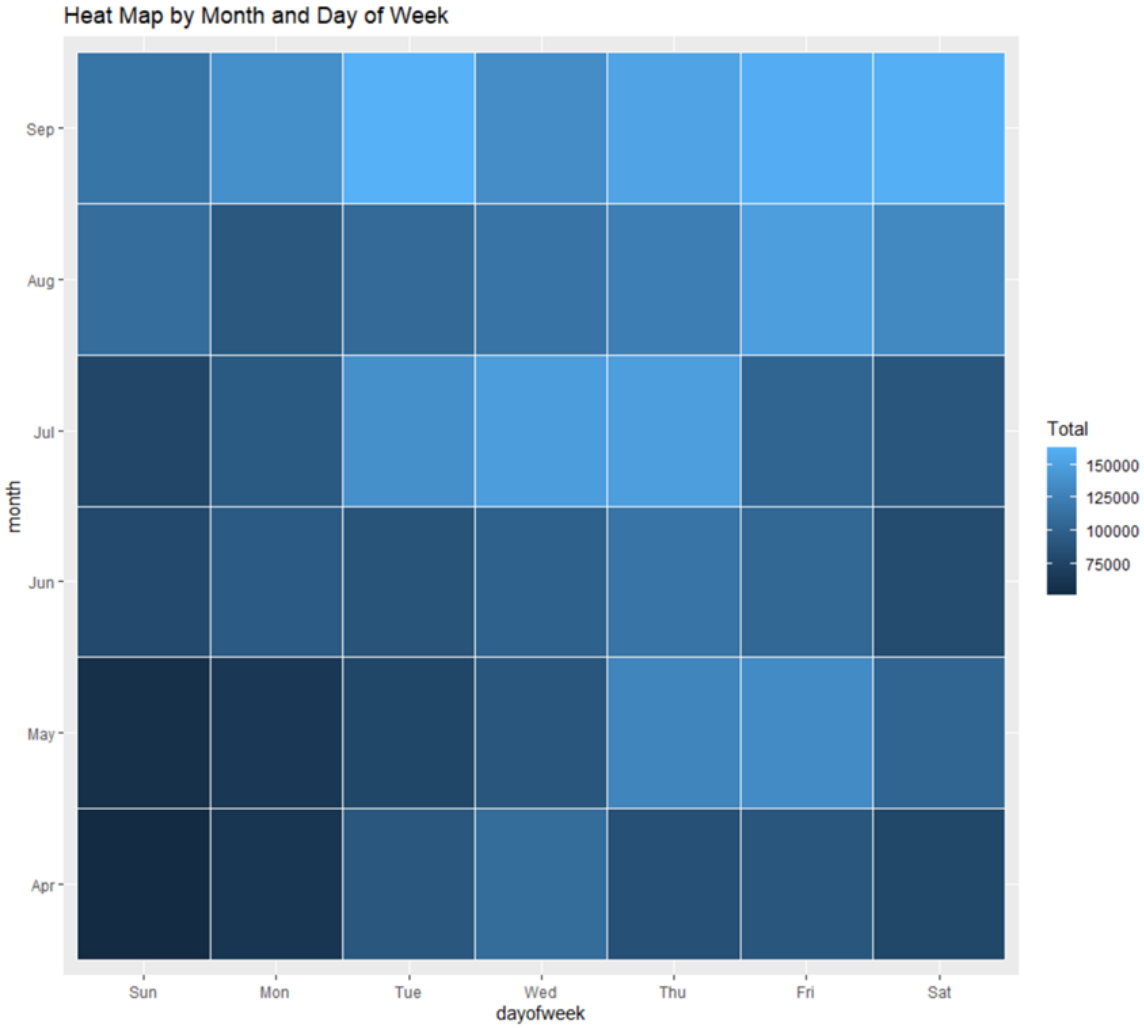
- The colors in the heatmap on the right represents the total number of trips grouped by hour and day.

<u>The lighter the color the higher the total.</u>

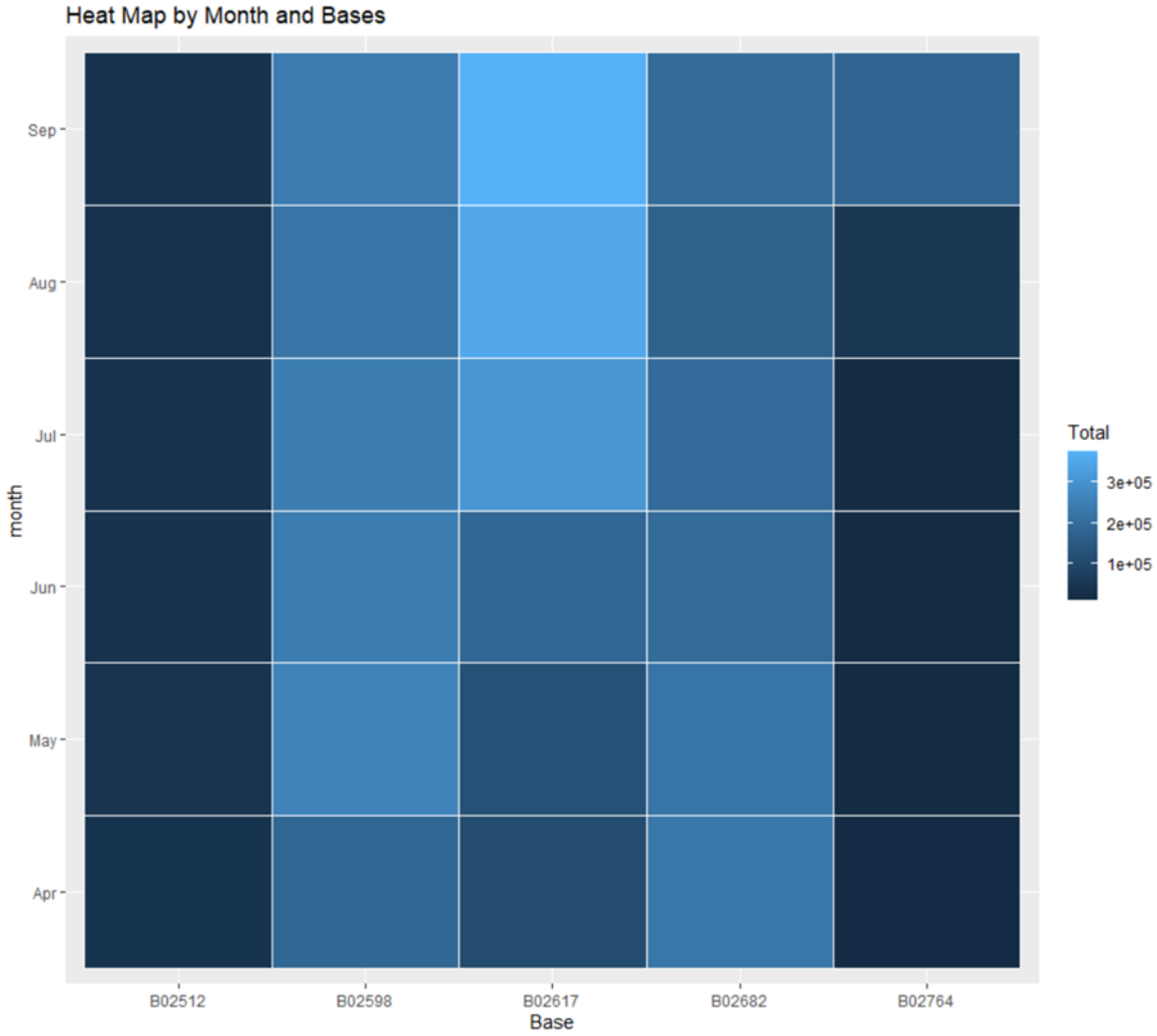


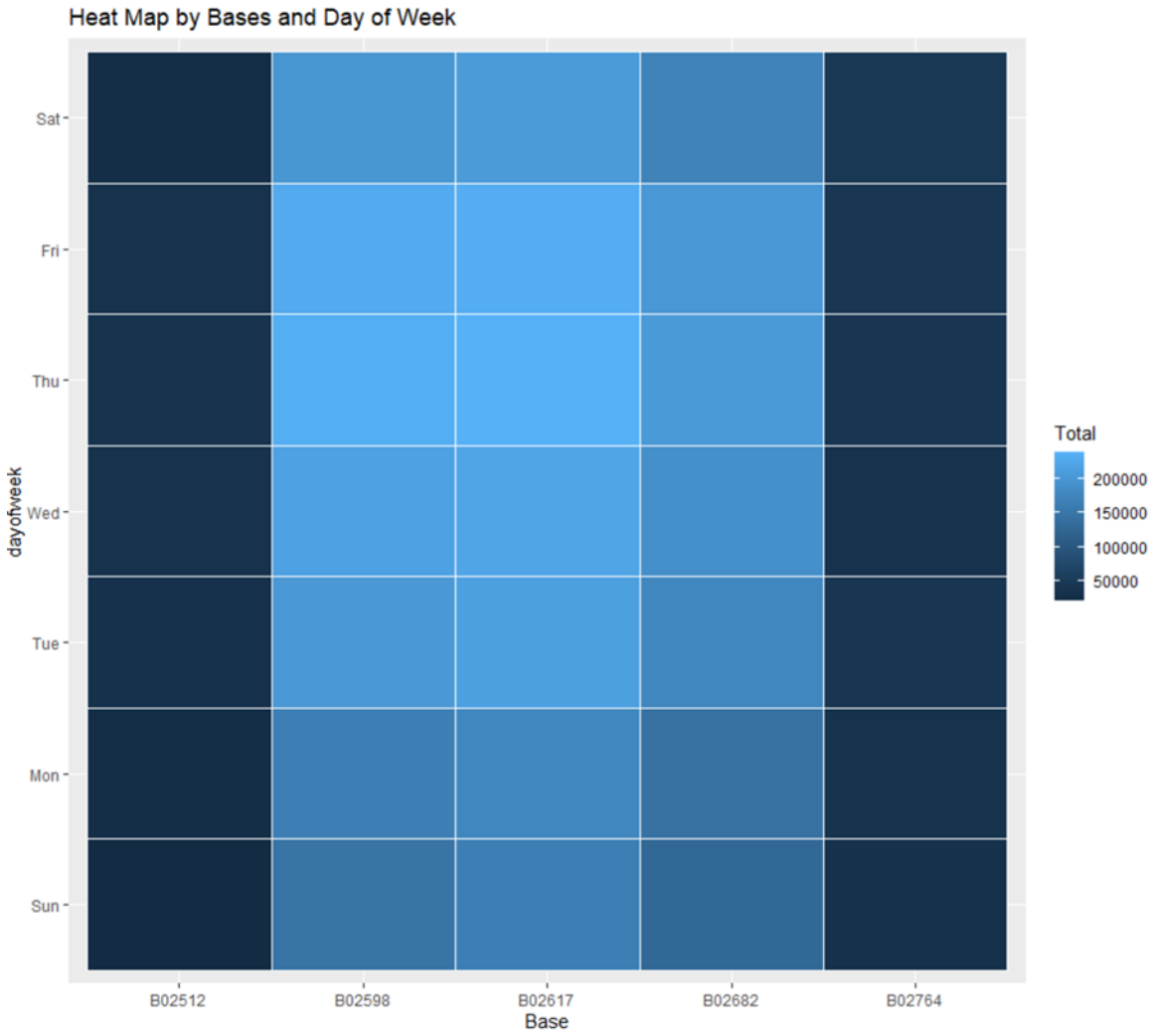
More Heat Maps





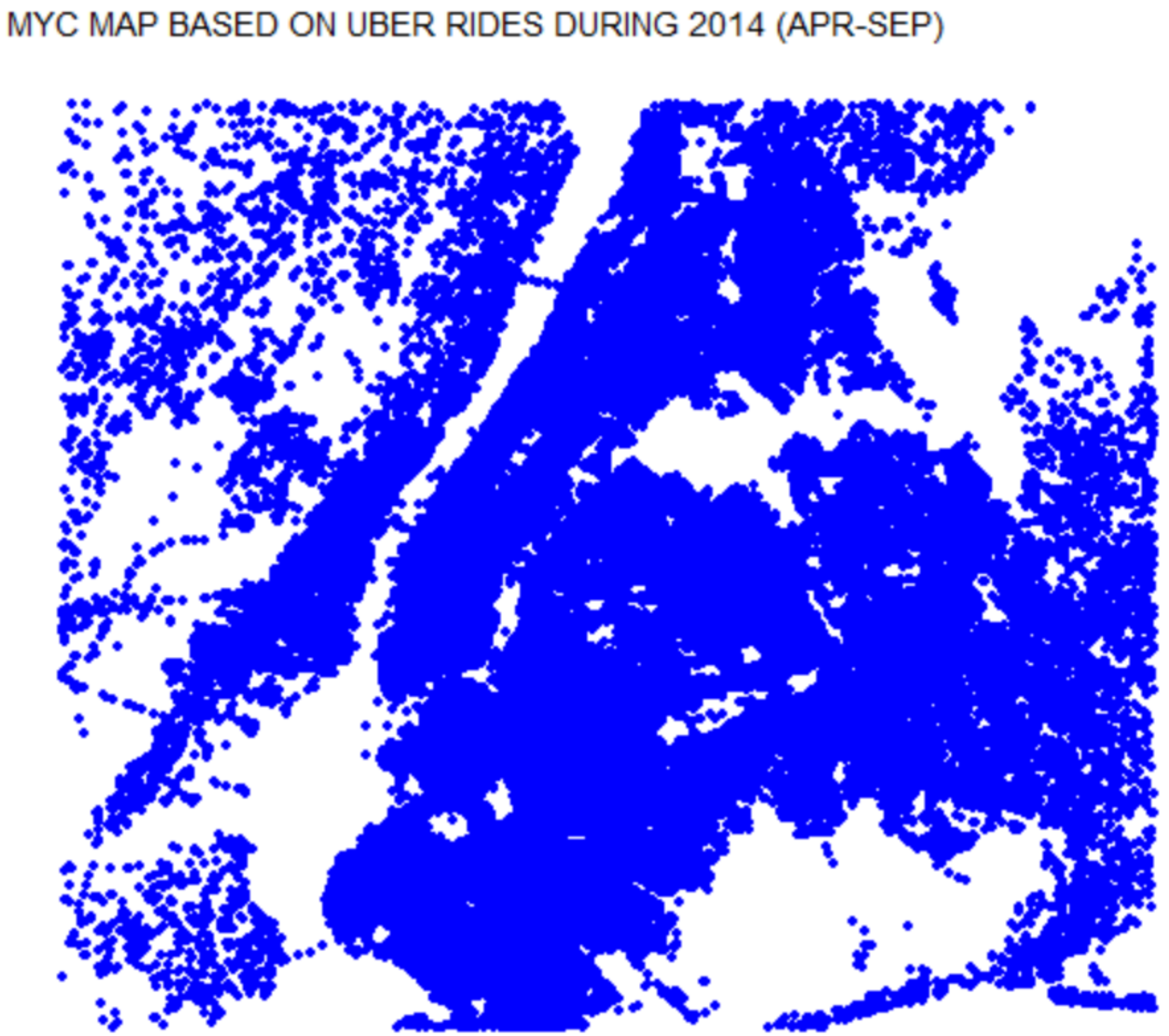
Heat Map


Heat Map by Month and Bases
Sep
Aug
Jul
Jun
May
Apr
month
B02512
B02598
B02617
B02682
B02764
Base
Total
3e+05
2e+05
1e+05


Heat Map by Bases and Day of Week
Sat
Fri
Thu
Wed
Tue
Mon
Sun
dayofweek
B02512
B02598
B02617
B02682
B02764
Base
Total
200000
150000
100000
50000

Map Visualization of Rides in NYC

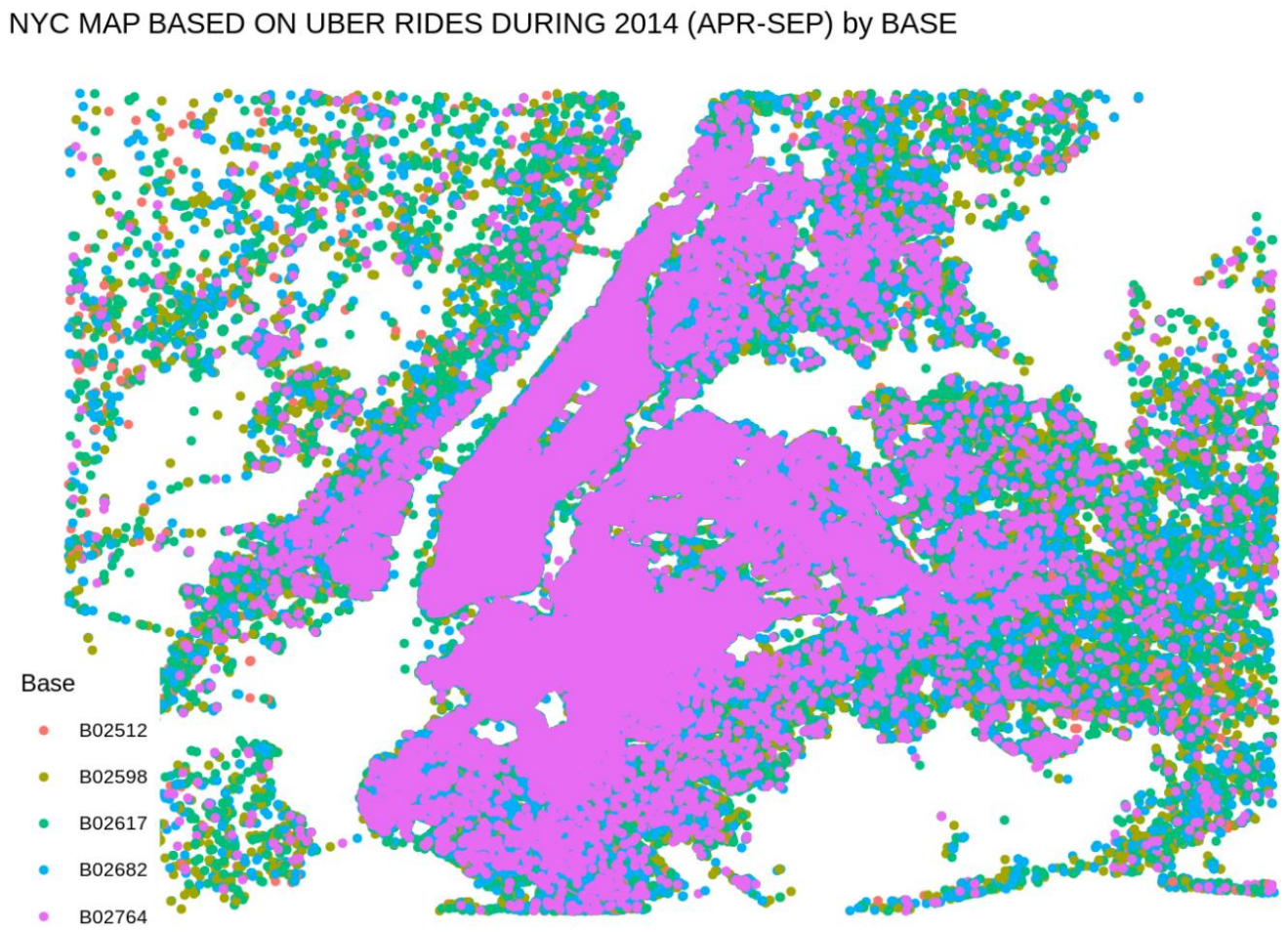
Each dot represents a pickup that happened in NYC. The more clustered an area is the more trips is made.



Map Visualization of Rides in NYC Differentiated by Base

This map is the same from the previous page, but is more transparent about which base code a pickup is affiliated with.

<u>As seen, the base code B02764 has the highest number of pickup in NYC.</u>




THANK YOU