

ANA515_Final Project

Thi Huyen Nguyen

12/12/2021

Goal of the project: Use visualizations to analyze how time affected customer trips and how various users made trips from different bases.

Installing Packages

```
install.packages("tidyverse") install.packages("ggplot2") install.packages("ggthemes") install.packages("lubridate")  
install.packages("dplyr") install.packages("tidyr") install.packages("DT") install.packages("scales")
```

Calling installed packages

```
##  
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':  
##  
##    date, intersect, setdiff, union
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##    filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##    intersect, setdiff, setequal, union
```

Creating vector of colors to be implemented in our plots

```
#Using read.csv to import data from "C:/Users/16087/OneDrive/Documents",
```

Using summary to describe datasets

```
##   Date.Time      Lat      Lon      Base
## Length:564516  Min.   :40.07  Min.   :-74.77  Length:564516
## Class :character 1st Qu.:40.72  1st Qu.: -74.00  Class :character
## Mode  :character Median :40.74  Median : -73.98  Mode  :character
##                Mean  :40.74  Mean   : -73.98
##                3rd Qu.:40.76  3rd Qu.: -73.97
##                Max.   :42.12  Max.   : -72.07
```

```
##   Date.Time      Lat      Lon      Base
## Length:652435  Min.   :40.11  Min.   :-74.93  Length:652435
## Class :character 1st Qu.:40.72  1st Qu.: -74.00  Class :character
## Mode  :character Median :40.74  Median : -73.98  Mode  :character
##                Mean  :40.74  Mean   : -73.97
##                3rd Qu.:40.76  3rd Qu.: -73.97
##                Max.   :41.32  Max.   : -72.18
```

```
##   Date.Time      Lat      Lon      Base
## Length:663844  Min.   :39.96  Min.   :-74.86  Length:663844
## Class :character 1st Qu.:40.72  1st Qu.: -74.00  Class :character
## Mode  :character Median :40.74  Median : -73.98  Mode  :character
##                Mean  :40.74  Mean   : -73.97
##                3rd Qu.:40.76  3rd Qu.: -73.97
##                Max.   :41.32  Max.   : -72.70
```

```
##   Date.Time      Lat      Lon      Base
## Length:796121  Min.   :39.72  Min.   :-74.83  Length:796121
## Class :character 1st Qu.:40.72  1st Qu.: -74.00  Class :character
## Mode  :character Median :40.74  Median : -73.98  Mode  :character
##                Mean  :40.74  Mean   : -73.97
##                3rd Qu.:40.76  3rd Qu.: -73.97
##                Max.   :41.34  Max.   : -72.31
```

```
##   Date.Time      Lat      Lon      Base
## Length:829275  Min.   :39.66  Min.   :-74.77  Length:829275
## Class :character 1st Qu.:40.72  1st Qu.: -74.00  Class :character
## Mode  :character Median :40.74  Median : -73.98  Mode  :character
##                Mean  :40.74  Mean   : -73.97
##                3rd Qu.:40.76  3rd Qu.: -73.96
##                Max.   :41.32  Max.   : -72.34
```

```
## Date.Time          Lat          Lon          Base
## Length:1028136    Min.   :39.99    Min.   : -74.77    Length:1028136
## Class :character  1st Qu.:40.72    1st Qu.: -74.00    Class :character
## Mode  :character  Median :40.74    Median : -73.98    Mode  :character
##                  Mean   :40.74    Mean   : -73.97
##                  3rd Qu.:40.76    3rd Qu.: -73.96
##                  Max.   :41.35    Max.   : -72.72
```

Using rbind to bind dataframes

Preparing data by Converting and Creating factors of time objects

#Plotting the trips by the hours in a day: Tnumber of trips are higher in the evening around 5:00 and 6:00 PM.

Show entries

Search:

| | hour | Total |
|----|------|--------|
| 1 | 0 | 103836 |
| 2 | 1 | 67227 |
| 3 | 2 | 45865 |
| 4 | 3 | 48287 |
| 5 | 4 | 55230 |
| 6 | 5 | 83939 |
| 7 | 6 | 143213 |
| 8 | 7 | 193094 |
| 9 | 8 | 190504 |
| 10 | 9 | 159967 |

Showing 1 to 10 of 24 entries

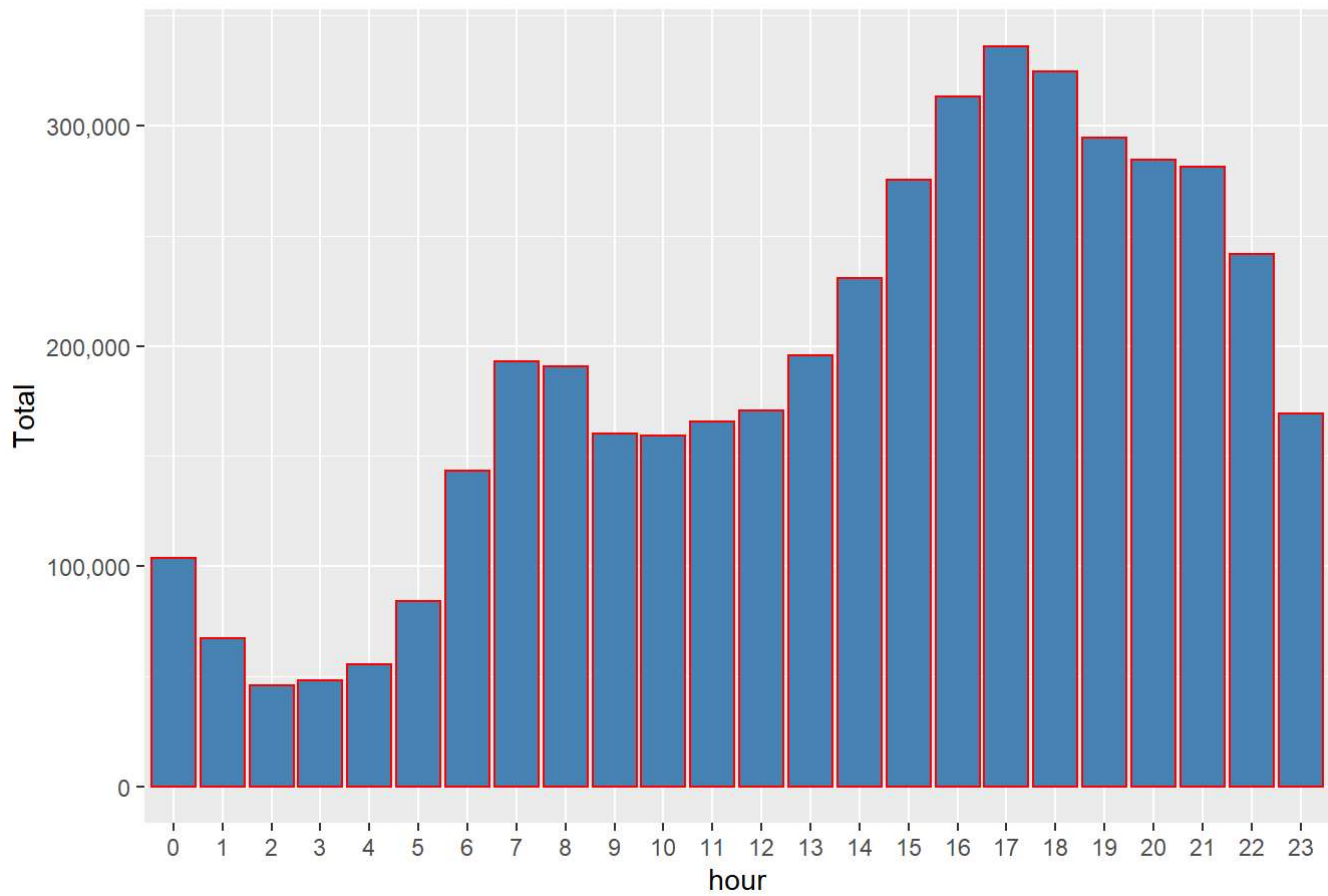
Previous

2

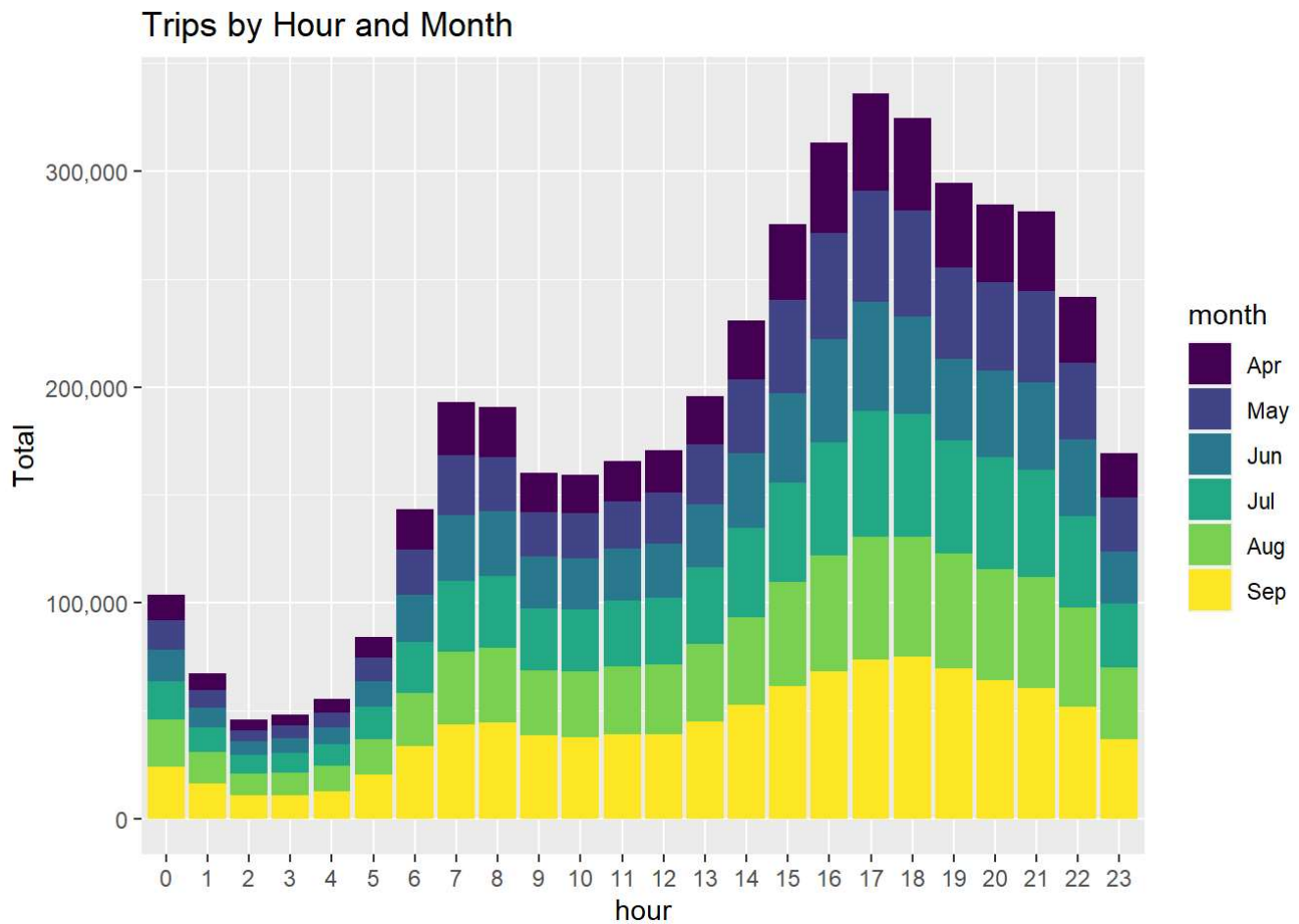
3

Next

Trips Every Hour



``summarise()`` has grouped output by 'month'. You can override using the ``.groups`` argument.



Plotting data by trips during every day of the month: number of trips are higher in the day 30th of the month.

Show entries

Search:

| day | | Total |
|-----|---|--------|
| 1 | 1 | 127430 |
| 2 | 2 | 143201 |
| 3 | 3 | 142983 |
| 4 | 4 | 140923 |
| 5 | 5 | 147054 |
| 6 | 6 | 139886 |
| 7 | 7 | 143503 |

| day | | Total |
|-----|----|--------|
| 8 | 8 | 145984 |
| 9 | 9 | 155135 |
| 10 | 10 | 152500 |

Showing 1 to 10 of 31 entries

Previous

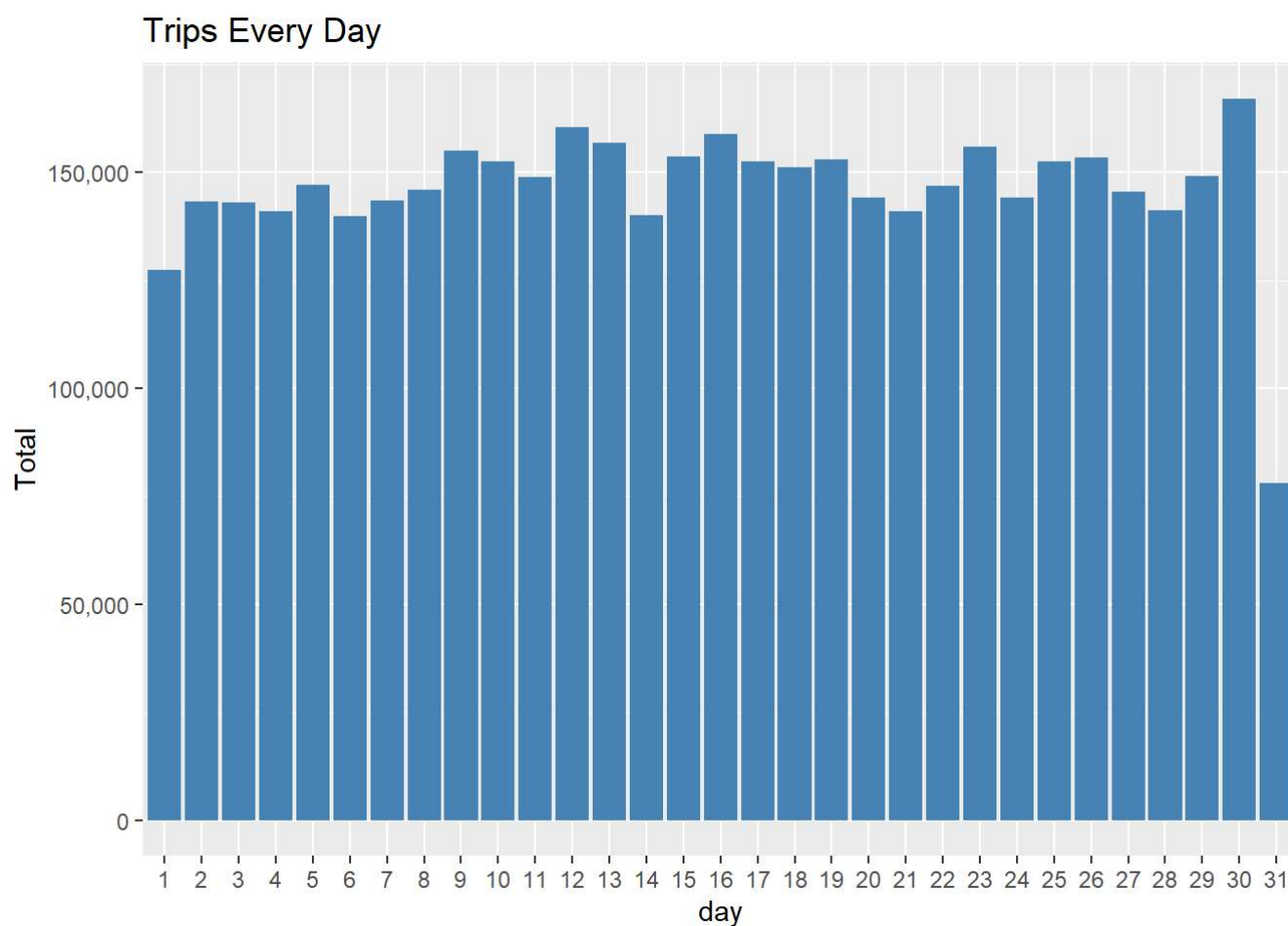
1

2

3

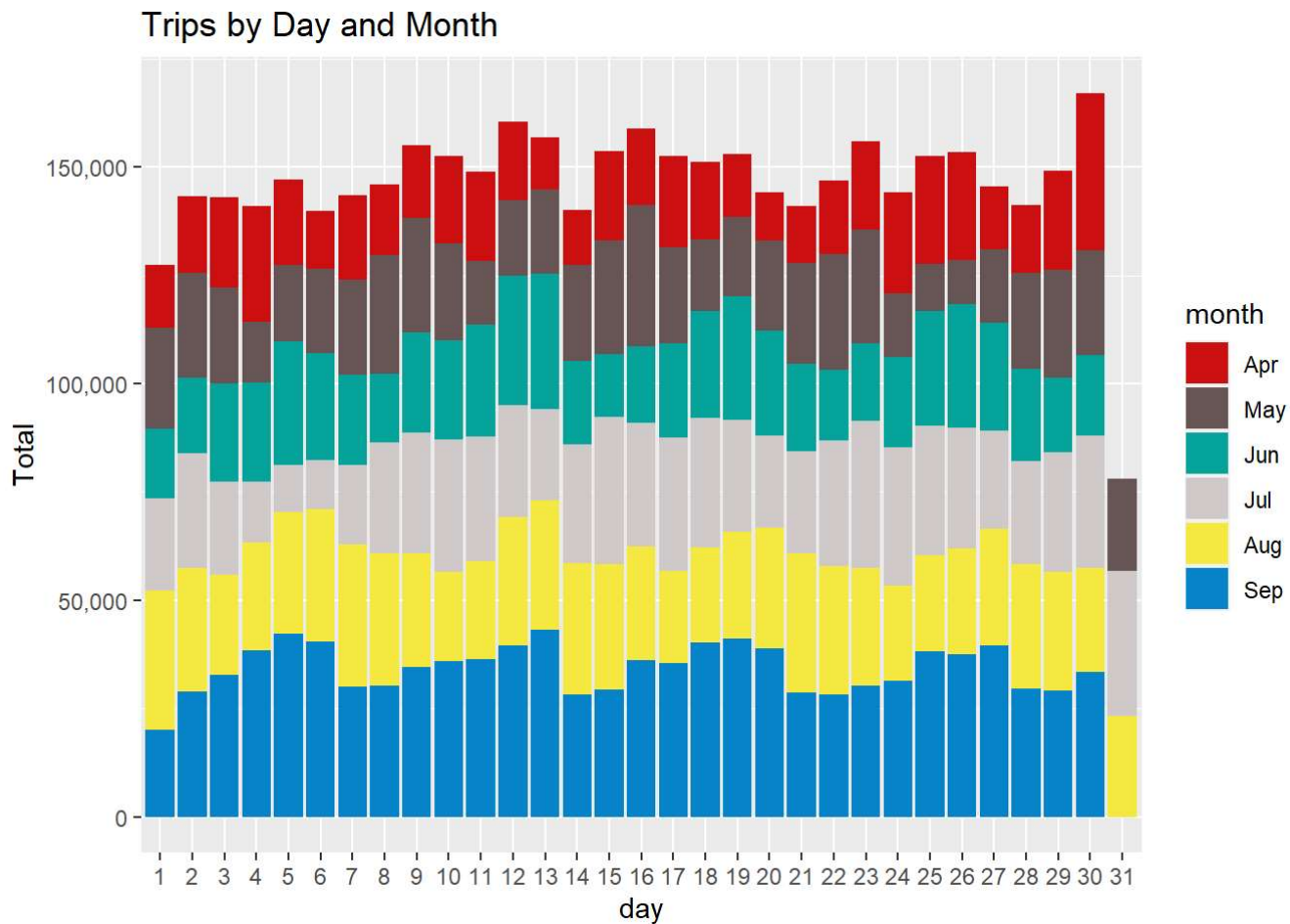
4

Next



Plotting data by trips by day and month:
number of trips are higher in day 30 and
September

```
## `summarise()` has grouped output by 'month'. You can override using the `.groups` argument.
```



Plotting Number of Trips taking place during months in a year: September is the most busy month of the year in the data

Show

10

 entries

Search:

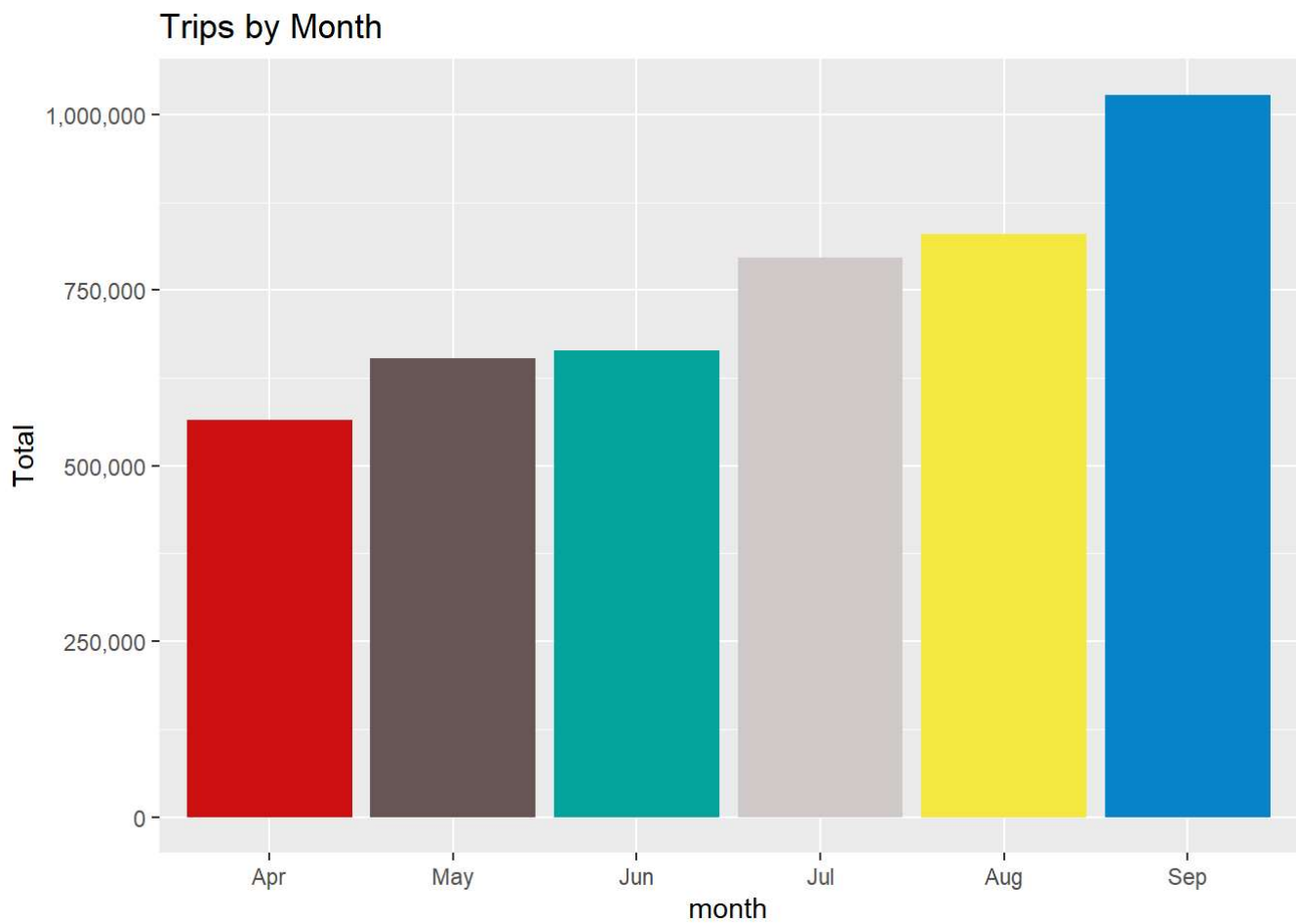
| month | | Total |
|-------|-----|---------|
| 1 | Apr | 564516 |
| 2 | May | 652435 |
| 3 | Jun | 663844 |
| 4 | Jul | 796121 |
| 5 | Aug | 829275 |
| 6 | Sep | 1028136 |

Showing 1 to 6 of 6 entries

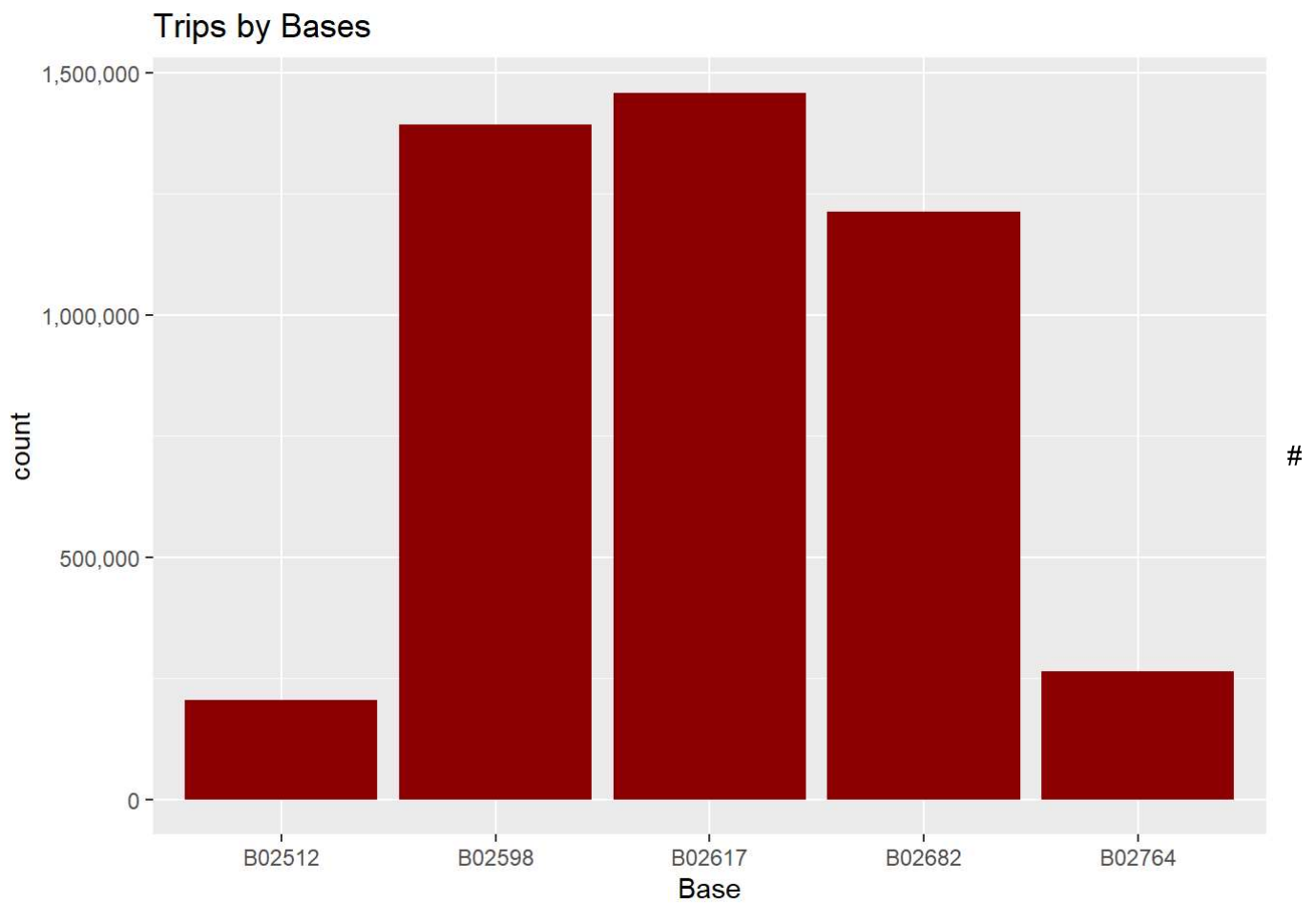
Previous

1

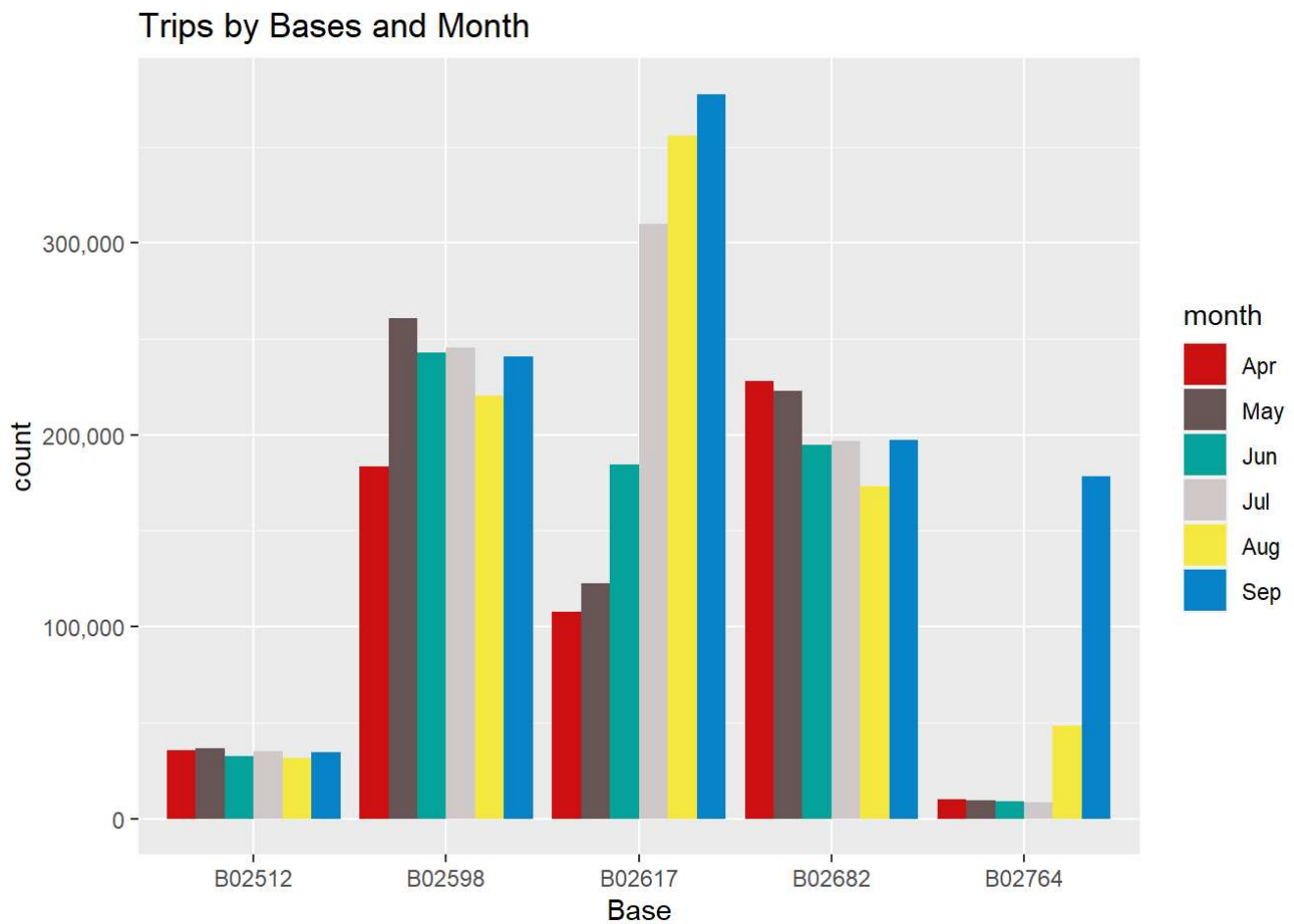
Next



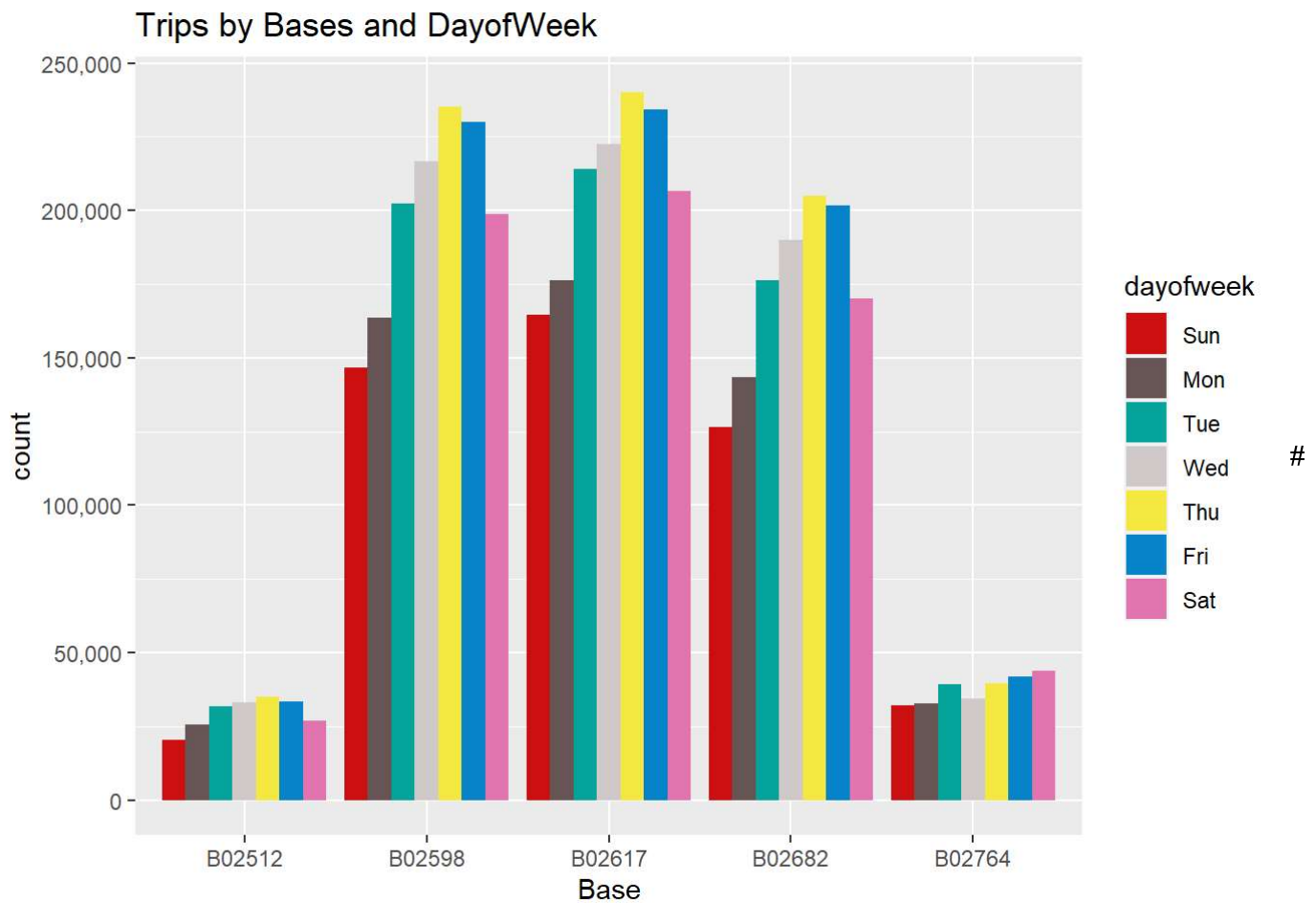
Plotting trip volume by Base: number of trips are higher in Base 02617



Plotting trips by Bases and Month: number of trips are higher Base 02617 and September



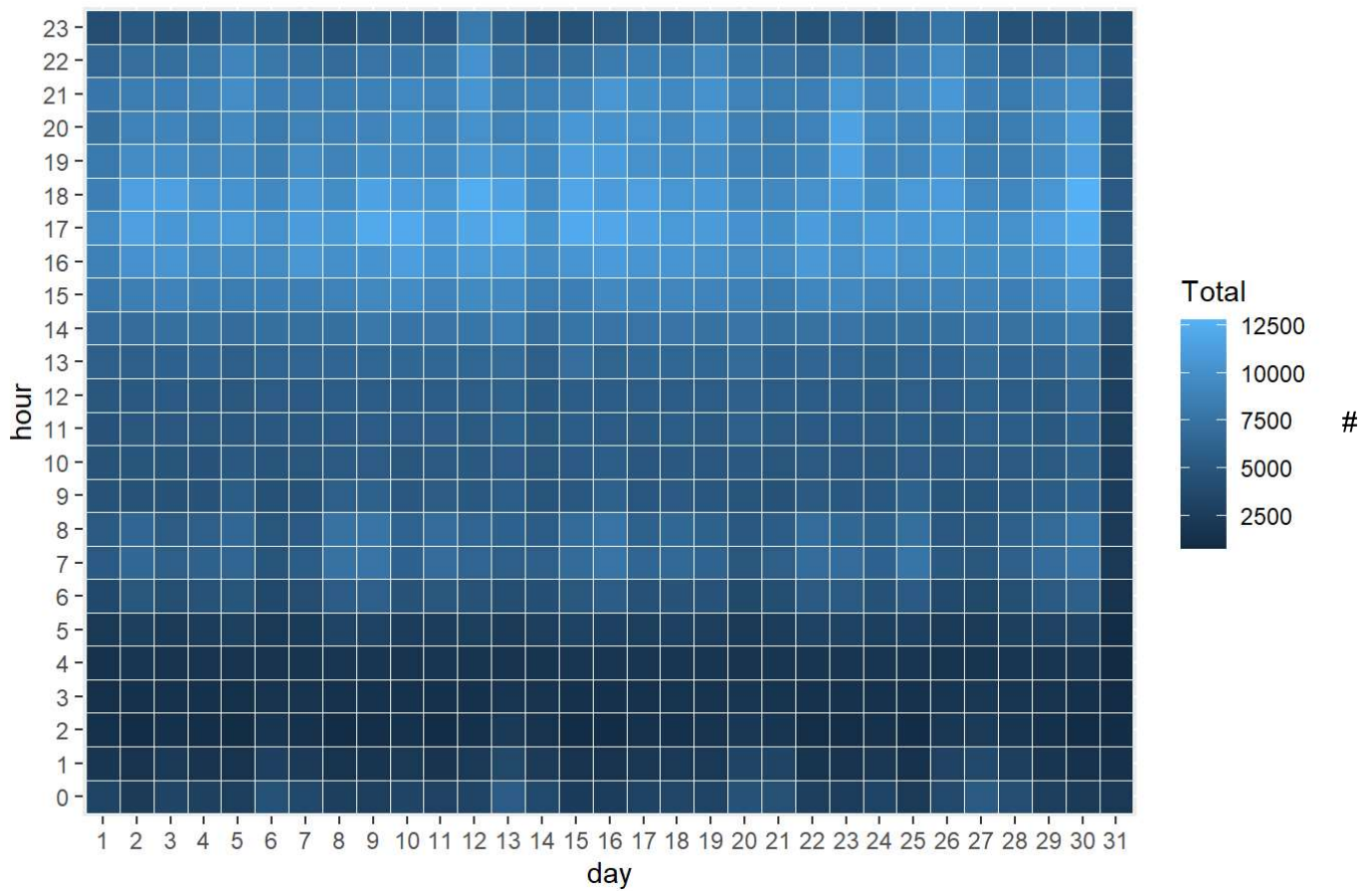
Plotting trip volume by Bases and Day of Week: number of trips are higher in Base 02617 and on Thursday.



Creating a Heatmap visualization of day, hour: number of trips are higher on the day 16th of the month and from 5PM to 6PM.

```
## `summarise()` has grouped output by 'day'. You can override using the `.groups` argument.
```

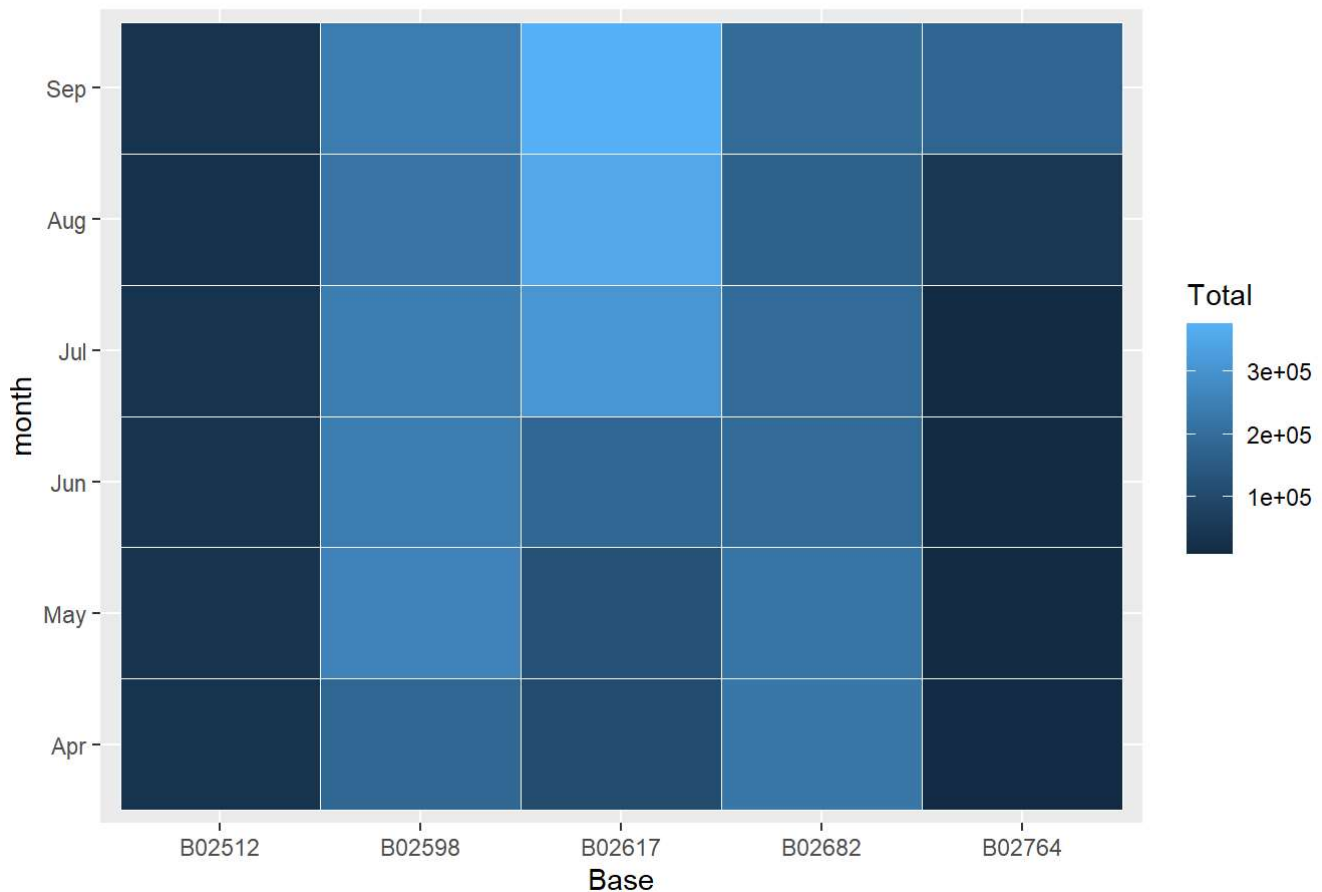
Heat Map by Hour and Day



Creating a Heatmap by month and bases

```
## `summarise()` has grouped output by 'Base'. You can override using the `.groups` argument.  
## `summarise()` has grouped output by 'Base'. You can override using the `.groups` argument.
```

Heat Map by Month and Bases

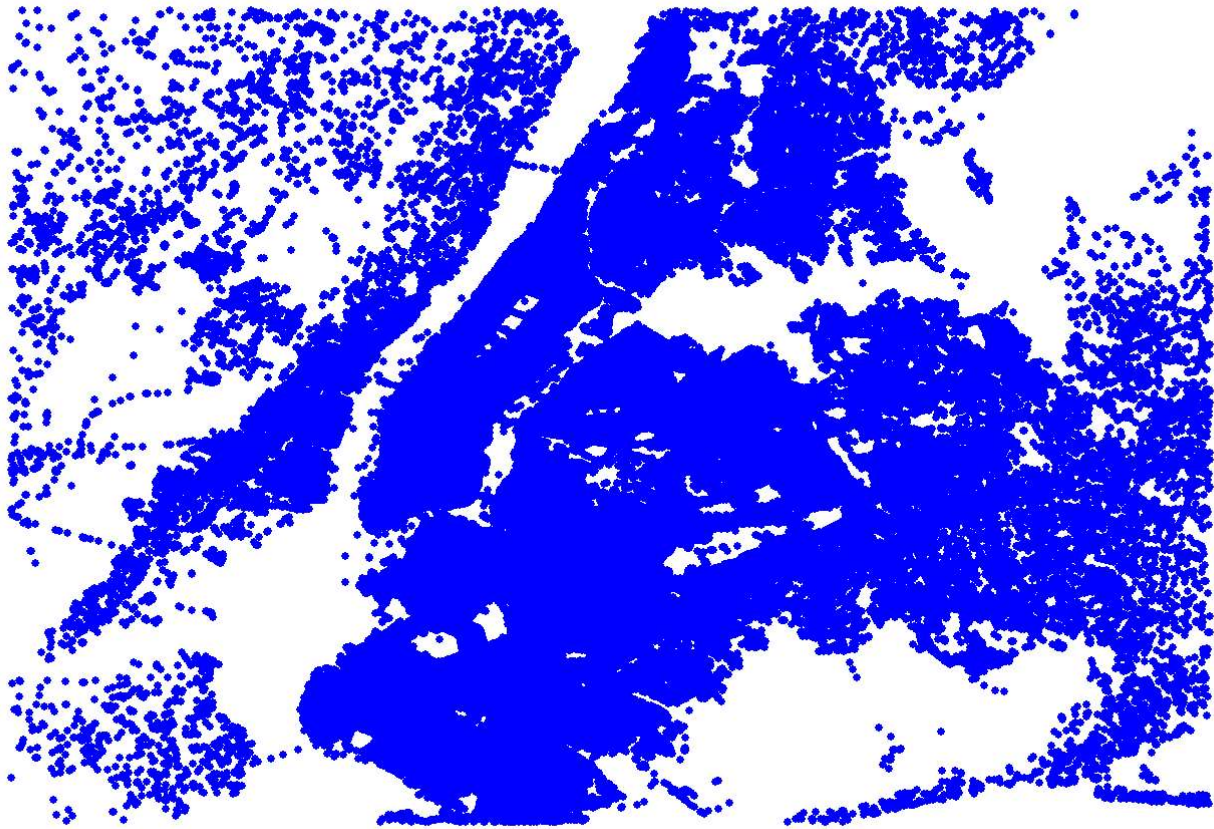


Creating heatmap by Bases and Day of the week: number of trips are higher in Base 02617 and on Thursday.

Creating a map visualization of rides in New York

```
## Warning: Removed 71701 rows containing missing values (geom_point).
```

NYC MAP BASED ON UBER RIDES DURING 2014 (APR-SEP)



```
## Warning: Removed 71701 rows containing missing values (geom_point).
```


NYC MAP BASED ON UBER RIDES DURING 2014 (APR-SEP) by BASE

