knitr::opts\_chunk$set(echo =F)  
############Install the required packages  
library(ggplot2)  
library(ggthemes)  
library(lubridate)

##   
## Attaching package: 'lubridate'

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

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:lubridate':  
##   
## intersect, setdiff, union

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

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

library(tidyr)  
library(DT)  
library(scales)

####Description of the datast

## [1] 4534327 9

The dataset contains 9 columns and 4534327 rows.

## 'data.frame': 4534327 obs. of 9 variables:  
## $ Date.Time: POSIXct, format: "2014-04-01 00:11:00" "2014-04-01 00:17:00" ...  
## $ Lat : num 40.8 40.7 40.7 40.8 40.8 ...  
## $ Lon : num -74 -74 -74 -74 -74 ...  
## $ Base : Factor w/ 5 levels "B02512","B02598",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Time : chr "00:11:00" "00:17:00" "00:21:00" "00:28:00" ...  
## $ day : Factor w/ 31 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ month : Ord.factor w/ 6 levels "Apr"<"May"<"Jun"<..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ year : Factor w/ 1 level "2014": 1 1 1 1 1 1 1 1 1 1 ...  
## $ dayofweek: Ord.factor w/ 7 levels "Sun"<"Mon"<"Tue"<..: 3 3 3 3 3 3 3 3 3 3 ...

The data set contains two numeric variables, four categorical variables and one date variable

## Lat Lon   
## Min. :39.66 Min. :-74.93   
## 1st Qu.:40.72 1st Qu.:-74.00   
## Median :40.74 Median :-73.98   
## Mean :40.74 Mean :-73.97   
## 3rd Qu.:40.76 3rd Qu.:-73.97   
## Max. :42.12 Max. :-72.07

##   
## B02512 B02598 B02617 B02682 B02764   
## 0.04535910 0.30723699 0.32173529 0.26746836 0.05820026

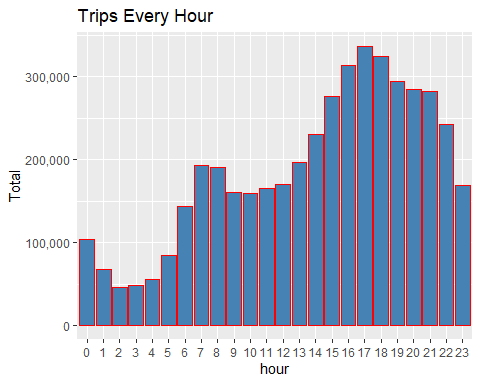
##   
## 1 2 3 4 5 6 7   
## 0.02810340 0.03158153 0.03153346 0.03107914 0.03243127 0.03085044 0.03164814   
## 8 9 10 11 12 13 14   
## 0.03219530 0.03421346 0.03363233 0.03282957 0.03542003 0.03460095 0.03090823   
## 15 16 17 18 19 20 21   
## 0.03390272 0.03504842 0.03363763 0.03337188 0.03376201 0.03179722 0.03112083   
## 22 23 24 25 26 27 28   
## 0.03240878 0.03441128 0.03179502 0.03366916 0.03383192 0.03212208 0.03113075   
## 29 30 31   
## 0.03287941 0.03686545 0.01721821

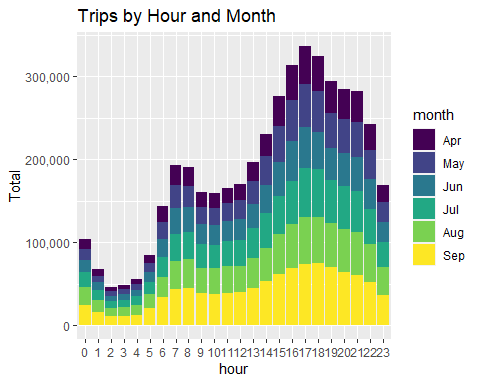
##   
## Apr May Jun Jul Aug Sep   
## 0.1244983 0.1438879 0.1464041 0.1755764 0.1828882 0.2267450

##   
## Sun Mon Tue Wed Thu Fri Sat   
## 0.1081042 0.1194162 0.1463920 0.1536034 0.1665396 0.1634507 0.1424939

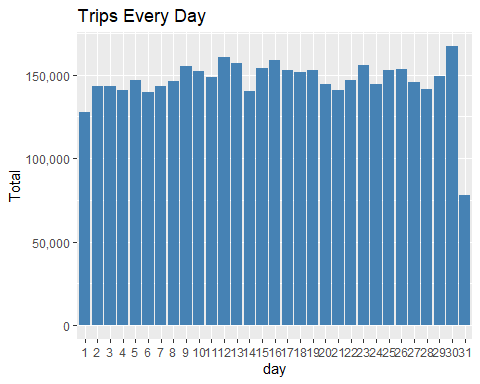
Base “B02617” has the highest proportion of pickups in 2014 with 32.1% of pick-ups while base B02512 has the lowest with 4.5% of picksup in 2014. The day with the highest pick-up is thursday while Sunday has the lowest pick-up. The month with the highest pickup is September with 22.7% of pickups while the day with the lowest amount of pickups is April with 12.4% of pickups.

##Plotting the trips by the hours in a day

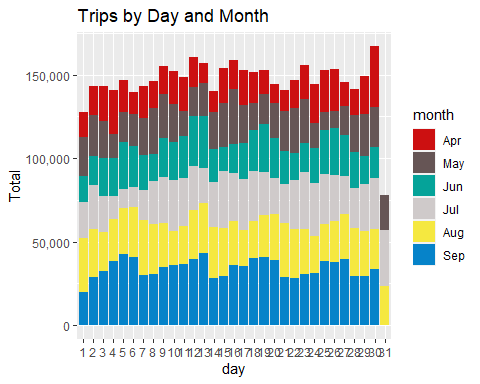
 The plot above presents the number of trips per hour of the day. We see that the number trips is the highest between the hours of 14:00 and 22:00 while trip is lowest between the hours of 00:00 and 5:00. The hours of 17:00 and 18:00 are the busiest.

 The plot shows that the distribution of trips by hours of the day is similar for all the months.

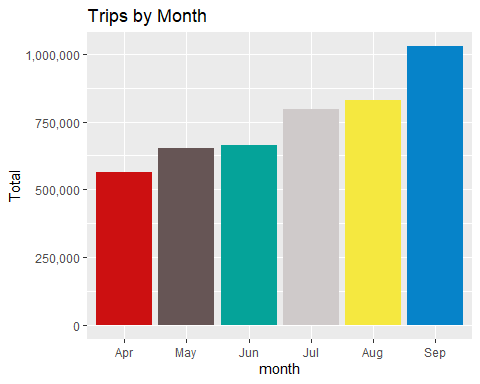
##Plotting data by trips during every day of the month



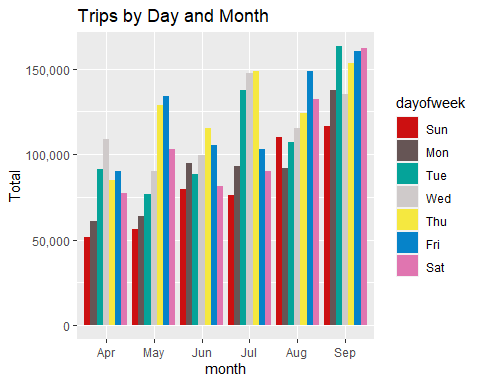
The plot above shows the number of trips per month. The result shows that the highest number of trips is on 30th day and the lowest os on 31st day.



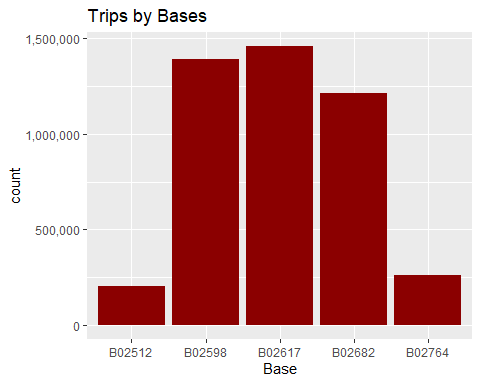
The result shows that most of the trip on the 30th day is contributed by the month April.

######Number of Trips taking place during months in a year 

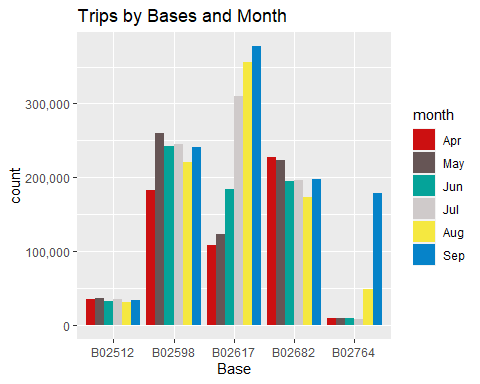
The plot shows that the number of trips becomes higher as time progress. September has the highest trip while April has the lowest.



The number of tips is not evenly distributed across months. For example, Wednesday has the highest number of trips in April. Friday has the highest number of trips in May. Thursday has the highest number of trips in June

###Finding out the number of Trips by bases 

The plot shows that base B02617 has the highest number of trips while B02512 has the lowest number of trips.

 The number of trips per month for base B02512 is close to one another. Looking at base B02764, we observed that most of the trips comes at September followed by August.For other bases large proprtion of the trips also came up in September.