````
---
title: "Uber trips Data Analysis"
output: github_document
author: 'Bharath Kumar Shivakumar'
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(ggthemes)
library(lubridate)
library(dplyr)
library(tidyr)
library(scales)
```

## Project goals
In this project, I have tried to analyze Uber Pickups in New York City
dataset.



## Including Code

```{r colors}
colors = c("#CC1011", "#665555", "#05a399", "#cfcaca", "#f5e840",
"#0683c9", "#e075b0")
```

```{r Import_datasets}
april_data <- read.csv("uber-raw-data-apr14.csv")
may_data <- read.csv("uber-raw-data-may14.csv")
jun_data <- read.csv("uber-raw-data-jun14.csv")
jul_data <- read.csv("uber-raw-data-jul14.csv")
aug_data <- read.csv("uber-raw-data-aug14.csv")
sep_data <- read.csv("uber-raw-data-sep14.csv")
```

```{r Format_data}
all_months <- rbind(april_data, may_data, jun_data, jul_data,
aug_data, sep_data)
all_months$Date.Time <- as.POSIXct(all_months$Date.Time, format =
"%m/%d/%Y %H:%M:%S")
all_months$Time <- format(as.POSIXct(all_months$Date.Time, format =
"%m/%d/%Y %H:%M:%S"), format="%H:%M:%S")
all_months$Date.Time <- ymd_hms(all_months$Date.Time)
all_months$day <- factor(day(all_months$Date.Time))
all_months$month <- factor(month(all_months$Date.Time, label = TRUE))
all_months$year <- factor(year(all_months$Date.Time))
all_months$dayofweek <- factor(wday(all_months$Date.Time, label =
TRUE))
all_months$hour <- factor(hour(hms(all_months$Time)))
all_months$minute <- factor(minute(hms(all_months$Time)))
all_months$second <- factor(second(hms(all_months$Time)))
````
````

```

## Including Plots

```{r Trip frequency, echo=FALSE}
hour_data <- all_months %>%
           group_by(hour) %>%
               dplyr::summarize(Total = n())
ggplot(hour_data, aes(hour, Total)) +
        geom_bar( stat = "identity", fill = "steelblue", color =
"red") +
            ggtitle("Trips Every Hour") +
            theme(legend.position = "none") +
            scale_y_continuous(labels = comma)

month_hour <- all_months %>%
          group_by(month, hour) %>%
              dplyr::summarize(Total = n())

ggplot(month_hour, aes(hour, Total, fill = month)) +
        geom_bar( stat = "identity") +
            ggtitle("Trips by Hour and Month") +
              scale_y_continuous(labels = comma)
```
```{r Everyday trips, echo=FALSE}
    day_group <- all_months %>%
              group_by(day) %>%
                 dplyr::summarize(Total = n())
    ggplot(day_group, aes(day, Total)) +
            geom_bar( stat = "identity", fill = "steelblue") +
               ggtitle("Trips Every Day") +
                theme(legend.position = "none") +
                scale_y_continuous(labels = comma)
```

```{r Trips by day of the week, echo=FALSE}
    day_month_group <- all_months %>%
               group_by(month, day) %>%
                   dplyr::summarize(Total = n())
    ggplot(day_month_group, aes(day, Total, fill = month)) +
            geom_bar( stat = "identity") +
               ggtitle("Trips by Day and Month") +
                scale_y_continuous(labels = comma) +
                scale_fill_manual(values = colors)
```
```{r Trips by month, echo=FALSE}
month_group <- all_months %>%
          group_by(month) %>%
             dplyr::summarize(Total = n())
ggplot(month_group , aes(month, Total, fill = month)) +
        geom_bar( stat = "identity") +
            ggtitle("Trips by Month") +
            theme(legend.position = "none") +
            scale_y_continuous(labels = comma) +
            scale_fill_manual(values = colors)
```
```

```

```{r Trips by day of the month, echo=FALSE}
    month_weekday <- all_months %>%
              group_by(month, dayofweek) %>%
                  dplyr::summarize(Total = n())
    ggplot(month_weekday, aes(month, Total, fill = dayofweek)) +
           geom_bar( stat = "identity", position = "dodge") +
              ggtitle("Trips by Day and Month") +
               scale_y_continuous(labels = comma) +
               scale_fill_manual(values = colors)
```
```{r Trip by base, echo=FALSE}
    ggplot(all_months, aes(Base)) +
     geom_bar(fill = "darkred") +
     scale_y_continuous(labels = comma) +
     ggtitle("Trips by Bases")
```

```{r Trips by Bases and month, echo=FALSE}
    ggplot(all_months, aes(Base, fill = month)) +
     geom_bar(position = "dodge") +
     scale_y_continuous(labels = comma) +
     ggtitle("Trips by Bases and Month") +
     scale_fill_manual(values = colors)
```

```{r Trips by Bases and Day of week, echo=FALSE}
    ggplot(all_months, aes(Base, fill = dayofweek)) +
     geom_bar(position = "dodge") +
     scale_y_continuous(labels = comma) +
     ggtitle("Trips by Bases and DayofWeek") +
     scale_fill_manual(values = colors)
```

```{r Heat map by hour and day, echo=FALSE}
day_and_hour <- all_months %>%
         group_by(day, hour) %>%
            dplyr::summarize(Total = n())
ggplot(day_and_hour, aes(day, hour, fill = Total)) +
            geom_tile(color = "white") +
              ggtitle("Heat Map by Hour and Day")
```

```{r Map visualization of rides in NYC}
min_lat <- 40.5774
max_lat <- 40.9176
min_long <- -74.15
max_long <- -73.7004
ggplot(all_months, aes(x=Lon, y=Lat, color = Base)) +
  geom_point(size=1) +
     scale_x_continuous(limits=c(min_long, max_long)) +
      scale_y_continuous(limits=c(min_lat, max_lat)) +
       theme_map() +
          ggtitle("NYC MAP BASED ON UBER RIDES DURING 2014 (APR-SEP)
```
```

```
by BASE")
```
Linear Regression

```{r Linear regression, echo=FALSE}
bookings = lm(Total ~ month + hour, data = month_hour)
summary(bookings)
```
Ploting linear regression
```{r Plot for ,inear regression}
plot(bookings$residuals, pch = 16, col = "blue")
```
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.