

Uber Insights: A Ride Through Data

Load necessary libraries

```
library(ggplot2)
library(dplyr)

##
## 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
library(lubridate)

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

Read the dataset

```
data <- read.csv("uber-raw-data-aug14.csv")
head(data)

##           Date.Time      Lat      Lon   Base
## 1 8/1/2014 0:03:00 40.7366 -73.9906 B02512
## 2 8/1/2014 0:09:00 40.7260 -73.9918 B02512
## 3 8/1/2014 0:12:00 40.7209 -74.0507 B02512
## 4 8/1/2014 0:12:00 40.7387 -73.9856 B02512
## 5 8/1/2014 0:12:00 40.7323 -74.0077 B02512
## 6 8/1/2014 0:13:00 40.7349 -74.0033 B02512
```

Examine the structure of the Date.Time column to understand its format and then convert Date.Time from character format to POSIXct for easier date-time manipulation.

The next step is to extract additional date-time features: Day, Weekday, and Hour

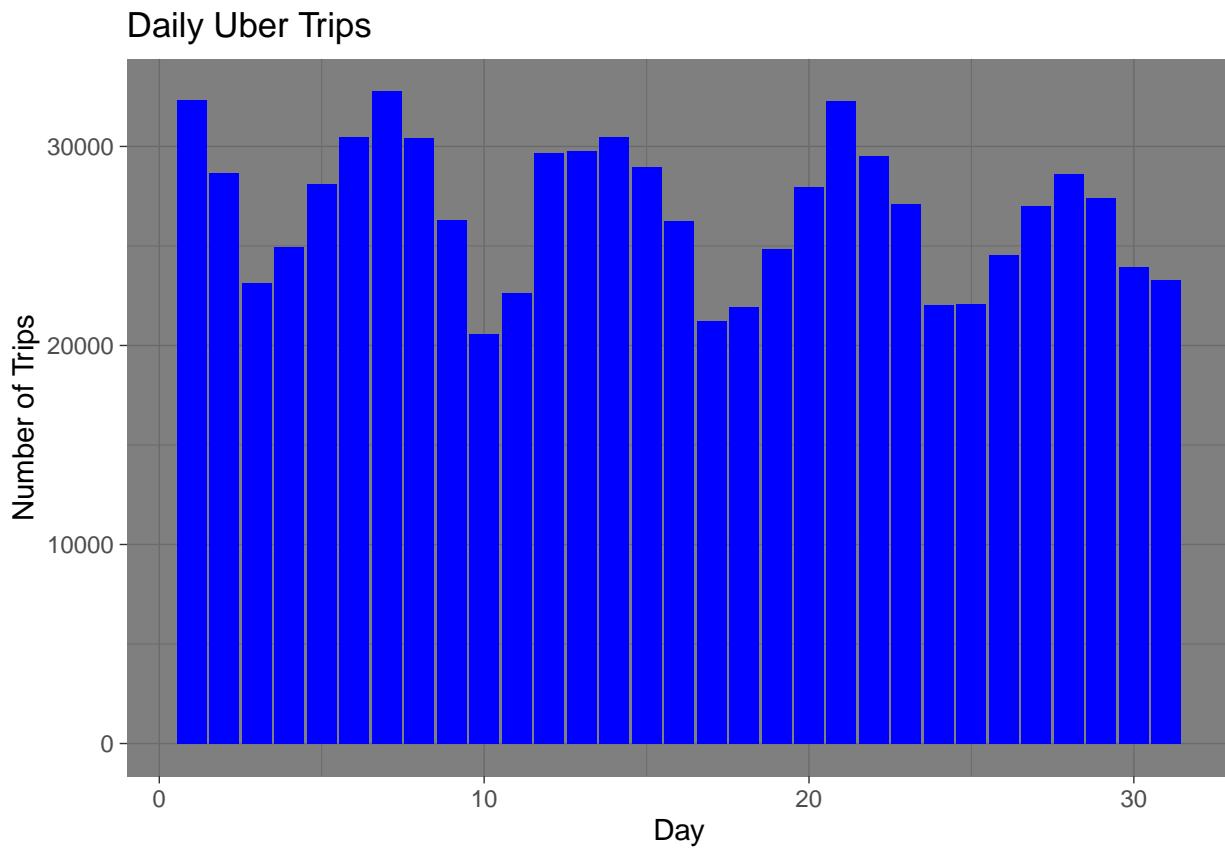
```
str(data$Date.Time)

##  chr [1:829275] "8/1/2014 0:03:00" "8/1/2014 0:09:00" "8/1/2014 0:12:00" ...
# Convert Date.Time from character to POSIXct
data$Date.Time <- as.POSIXct(data$Date.Time, format = "%m/%d/%Y %H:%M:%S")

data$Day <- day(data$Date.Time)
data$Weekday <- wday(data$Date.Time, label = TRUE)
data$Hour <- hour(data$Date.Time)
```

Create a bar plot to visualize the number of Uber trips per day

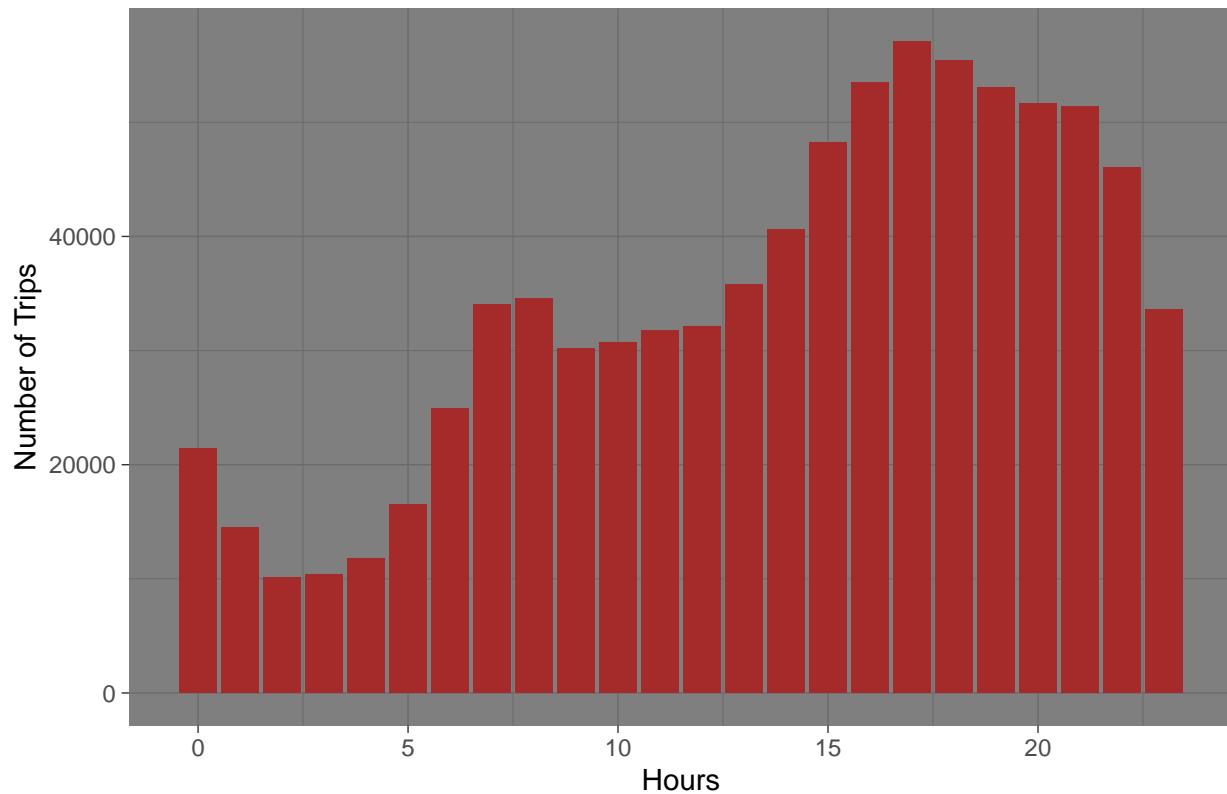
```
ggplot(data, aes(x = Day)) +  
  geom_bar(fill = "blue") +  
  labs(title = "Daily Uber Trips", x = "Day", y = "Number of Trips") +  
  theme_dark()
```



Create a bar plot to visualize the number of Uber trips per hour

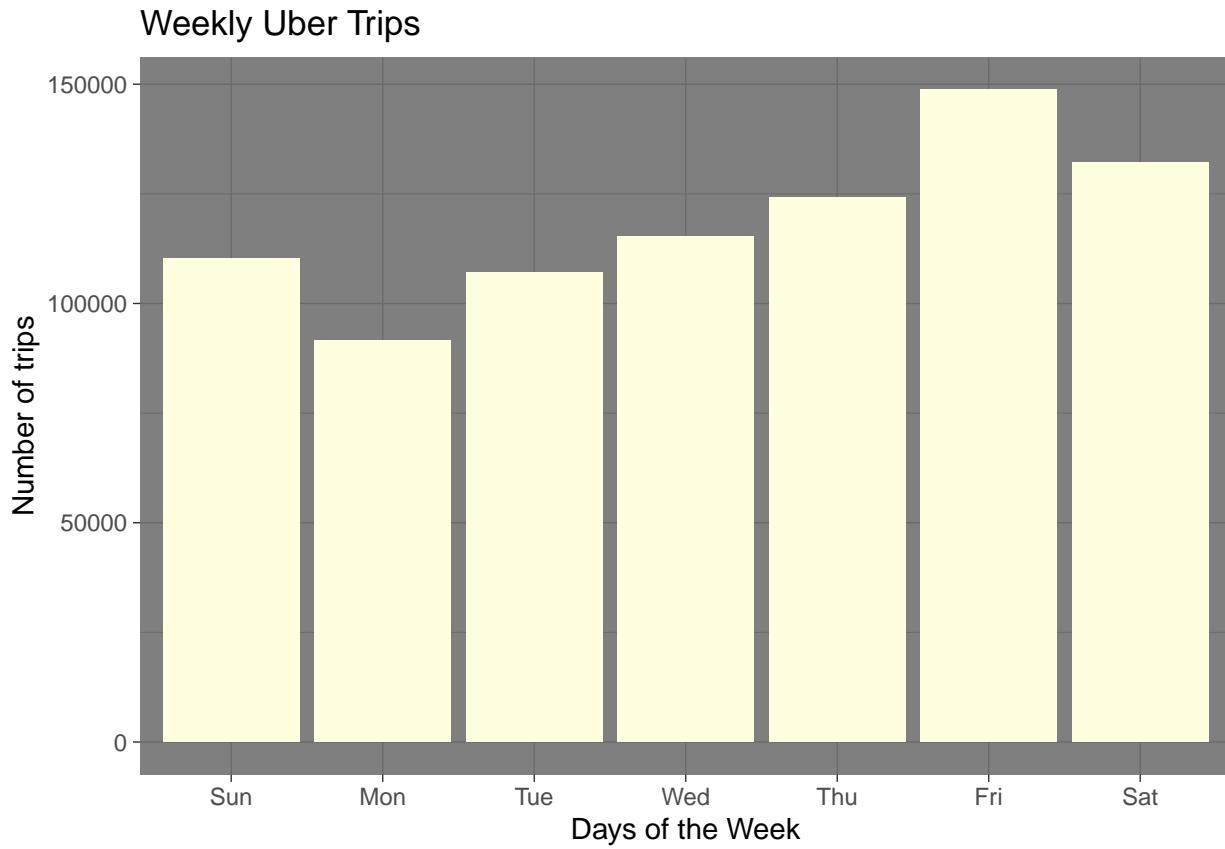
```
ggplot(data, aes(x=Hour)) +  
  geom_bar(fill = "brown") +  
  labs(title = "Hourly Uber Trips", x = "Hours", y = "Number of Trips") +  
  theme_dark()
```

Hourly Uber Trips



```
### Create a bar plot to visualize the number of Uber trips by weekday
```

```
ggplot(data, aes(x=Weekday)) +  
  geom_bar(fill = "lightyellow") +  
  labs(title = "Weekly Uber Trips", x = "Days of the Week", y = "Number of trips") +  
  theme_dark()
```



Display summary statistics of the dataset

```
summary(data)
```

```
##   Date.Time                  Lat          Lon
## Min.   :2014-08-01 00:00:00.00  Min.   :39.66  Min.   :-74.77
## 1st Qu.:2014-08-08 09:14:00.00  1st Qu.:40.72  1st Qu.:-74.00
## Median :2014-08-15 21:22:00.00  Median :40.74  Median :-73.98
## Mean   :2014-08-16 06:32:54.99  Mean   :40.74  Mean   :-73.97
## 3rd Qu.:2014-08-23 19:26:00.00  3rd Qu.:40.76  3rd Qu.:-73.96
## Max.   :2014-08-31 23:59:00.00  Max.   :41.32  Max.   :-72.34
##
##      Base           Day     Weekday       Hour
## Length:829275    Min.   : 1.00  Sun:110246  Min.   : 0.00
## Class :character  1st Qu.: 8.00  Mon: 91633   1st Qu.: 9.00
## Mode  :character  Median :15.00  Tue:107124   Median :15.00
##                   Mean   :15.67  Wed:115256   Mean   :14.06
##                   3rd Qu.:23.00  Thu:124117   3rd Qu.:19.00
##                   Max.   :31.00  Fri:148674   Max.   :23.00
##                   Sat:132225
```

Count the number of trips grouped by Weekday and Hour

```
# Count trips by weekday and hour
trip_counts <- data %>%
  group_by(Weekday, Hour) %>%
  summarise(Trip_Count = n())
```

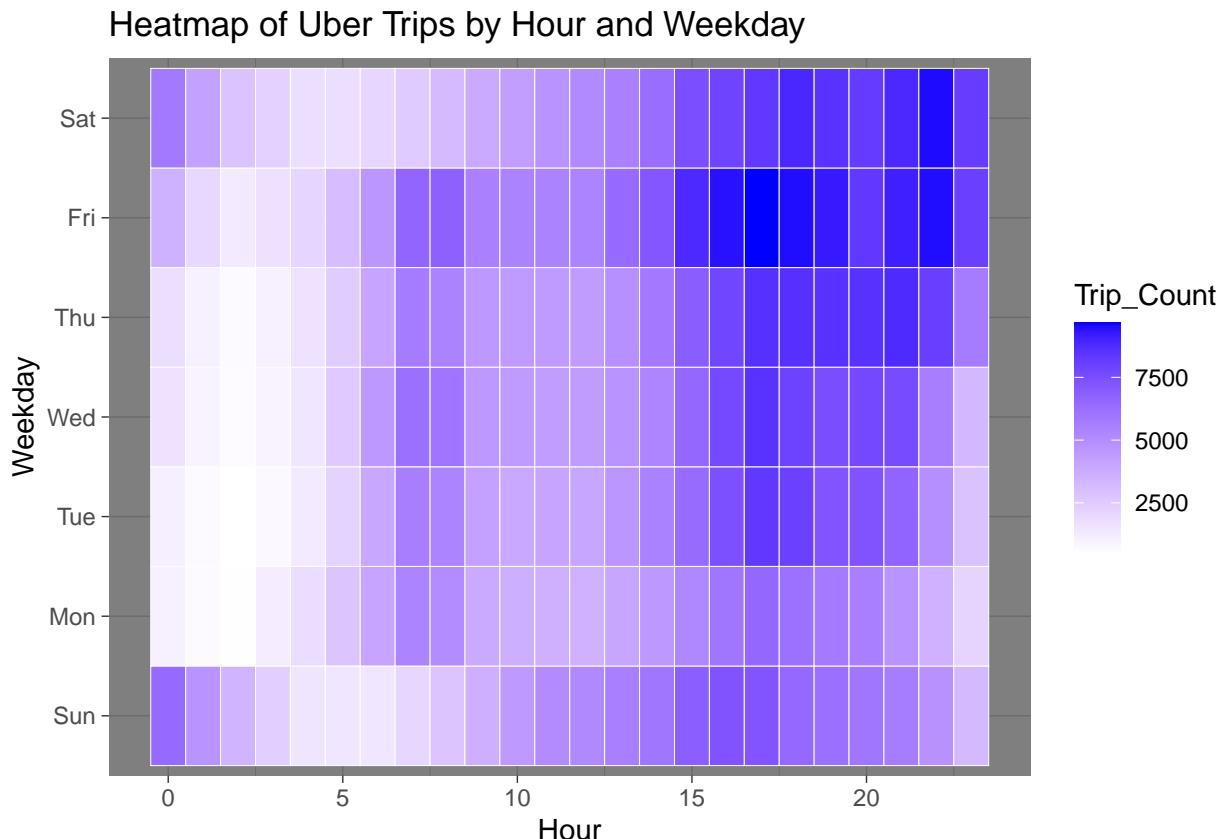
```
## `summarise()` has grouped output by 'Weekday'. You can override using the
## `.`groups` argument.
```

```
# View the first few rows of the summary
head(trip_counts)
```

```
## # A tibble: 6 x 3
## # Groups: Weekday [1]
##   Weekday Hour Trip_Count
##   <ord>    <int>     <int>
## 1 Sun        0       6429
## 2 Sun        1       4759
## 3 Sun        2       3439
## 4 Sun        3       2399
## 5 Sun        4       1530
## 6 Sun        5       1459
```

Create a heatmap to visualize the number of Uber trips by Hour and Weekday

```
# Create a heatmap
ggplot(trip_counts, aes(x = Hour, y = Weekday, fill = Trip_Count)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Heatmap of Uber Trips by Hour and Weekday", x = "Hour", y = "Weekday") +
  theme_dark()
```



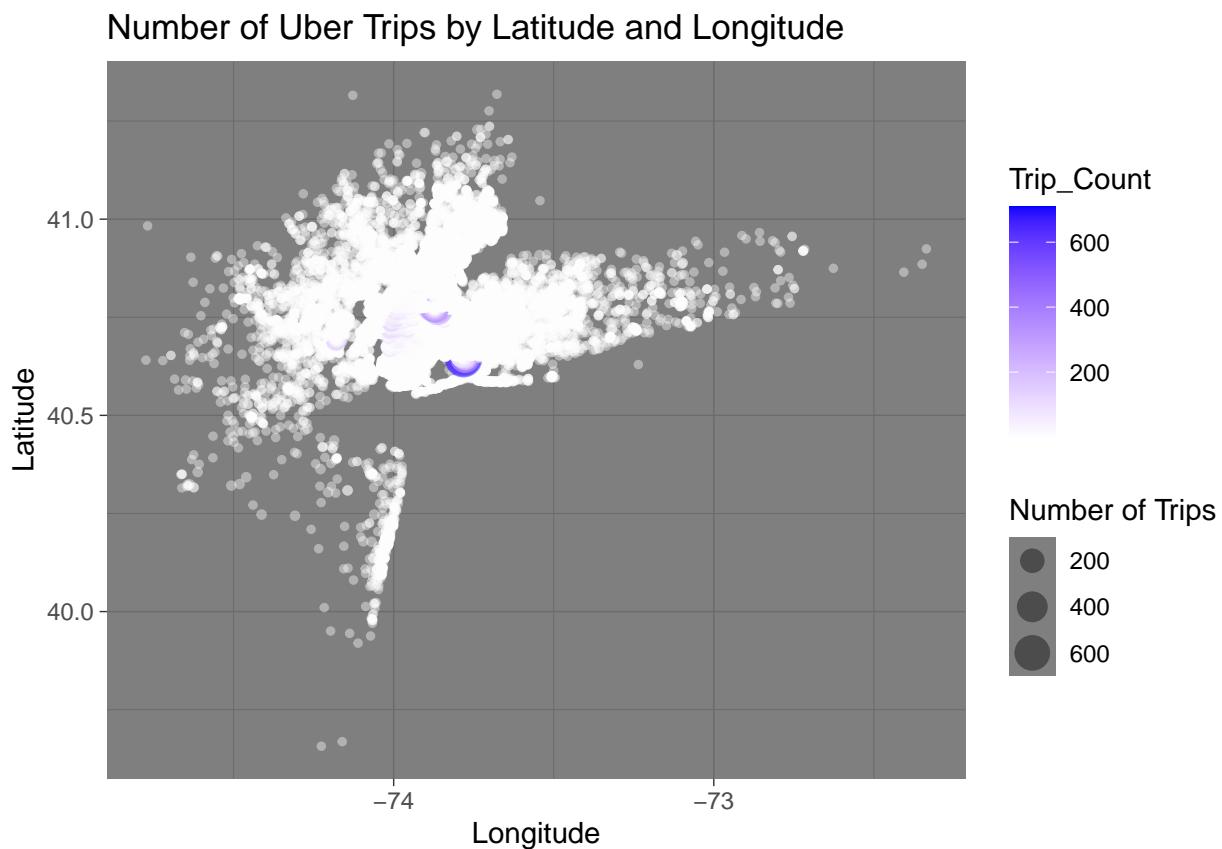
```
### Count the number of trips by Latitude and Longitude to visualize geographic distribution and create
a scatterplot to show the number of Uber trips based on Latitude and Longitude
```

```

trip_counts_geo <- data %>%
  group_by(Lat, Lon) %>%
  summarise(Trip_Count = n(), .groups = 'drop')

ggplot(trip_counts_geo, aes(x = Lon, y = Lat, size = Trip_Count, color = Trip_Count)) +
  geom_point(alpha = 0.4) +
  scale_color_gradient(low = "white", high = "blue") +
  labs(title = "Number of Uber Trips by Latitude and Longitude",
       x = "Longitude",
       y = "Latitude",
       size = "Number of Trips") +
  theme_dark()

```



Summary of Uber Trip Analysis

The analysis of Uber trip data for August 2014 reveals several key insights:

Peak Profitability: Friday stands out as the most profitable day for Uber, indicating a higher demand for rides as the weekend approaches.

Lower Usage on Mondays: Monday shows the least number of trips, suggesting that fewer people opt for Uber at the beginning of the week.

Busiest Hour: The period between 5 PM and 6 PM is identified as the busiest hour for Uber, likely due to the end of the workday when many commuters seek rides.

Early Morning Demand: There is a noticeable increase in Uber trips starting around 6 AM, indicating that early morning rides are common, possibly for commuters heading to work.

Geographic Concentration: Most Uber trips originate in the Manhattan region of New York, highlighting this area as a central hub for ride requests.

These insights can be valuable for understanding rider behavior and optimizing service strategies for Uber.