BA Pundits

Ushnik

December 12, 2016

NYC Yellow Taxi Traffic Data Analysis

Using NYC Taxi data available on the NYC Taxi and Limousine Commission website, we intend to analyze the geography of pickup and drop-offs made by the cabs during peak hours of the day. The data set includes 11.2 million trips made during the month of June 2016 and drills down each trip to particulars like pickup and drop-off dates, time, longitude, latitude, trip distance, fare amount, etc. We aim to identify the area that is the busiest during the peak hours and target the average commute time during weekdays and weekends. This would give us a sense of how much time it takes for a person to get from Point A to Point B with and without traffic (assuming weekends have little or no traffic as compared to weekdays).

```
####Install packages: "ggplot2" and "lubridate"
options(warn=-1)
library(ggplot2)
library(hexbin)
library(readr)
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(extrafont)
```

```
## Registering fonts with R
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following objects are masked from 'package:readr':
##
##     col_factor, col_numeric
```

```
library(grid)
library(RColorBrewer)
library(digest)
library(stringr)
library(methods)
library(lubridate)
```

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

```
## The following object is masked from 'package:base':
##
##     date
```

```
options(repr.plot.mimetypes = 'image/png', repr.plot.width=4, repr.plot.height=3,
repr.plot.res=300)

df <- read.csv("C:/Users/Ushnik/Downloads/yellow_tripdata_2016-06.csv", header=T)
str(df)
```

```
## 'data.frame':    11135470 obs. of  19 variables:
##  $ VendorID             : int  2 2 2 2 2 2 2 1 1 1 ...
##  $ tpep_pickup_datetime : Factor w/ 2395442 levels "2016-06-01 00:00:00",..: 706487 706487 70
6487 706487 706487 706487 706487 706488 706488 706488 ...
##  $ tpep_dropoff_datetime: Factor w/ 2399652 levels "1996-06-20 16:23:24",..: 707233 708555 70
7235 708614 707847 707606 708257 707452 707389 707881 ...
##  $ passenger_count      : int  2 1 1 1 1 1 5 1 1 2 ...
##  $ trip_distance        : num  0.79 5.22 1.26 7.39 3.1 2.17 6.02 1.4 1.2 1.9 ...
##  $ pickup_longitude     : num  -74 -74 -74 -74 -74 ...
##  $ pickup_latitude      : num  40.8 40.7 40.8 40.8 40.7 ...
##  $ RatecodeID           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ store_and_fwd_flag   : Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 1 ...
##  $ dropoff_longitude    : num  -74 -74 -74 -73.9 -74 ...
##  $ dropoff_latitude     : num  40.8 40.7 40.7 40.9 40.8 ...
##  $ payment_type         : int  2 1 1 1 1 1 2 1 2 1 ...
##  $ fare_amount          : num  6 22 6.5 26 13.5 10.5 21.5 8.5 8 12 ...
##  $ extra                : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
##  $ mta_tax              : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
##  $ tip_amount           : num  0 4 1.56 1 2.96 2.36 0 1.95 0 3.33 ...
##  $ tolls_amount         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ improvement_surcharge: num  0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 ...
##  $ total_amount         : num  7.3 27.3 9.36 28.3 17.76 ...
```

```
head(df)
```

```
##   VendorID tpep_pickup_datetime tpep_dropoff_datetime passenger_count
## 1        2  2016-06-09 21:06:36   2016-06-09 21:13:08               2
## 2        2  2016-06-09 21:06:36   2016-06-09 21:35:11               1
## 3        2  2016-06-09 21:06:36   2016-06-09 21:13:10               1
## 4        2  2016-06-09 21:06:36   2016-06-09 21:36:10               1
## 5        2  2016-06-09 21:06:36   2016-06-09 21:23:23               1
## 6        2  2016-06-09 21:06:36   2016-06-09 21:19:21               1
##   trip_distance pickup_longitude pickup_latitude RatecodeID
## 1          0.79        -73.98336        40.76094          1
## 2          5.22        -73.98172        40.73667          1
## 3          1.26        -73.99432        40.75107          1
## 4          7.39        -73.98236        40.77389          1
## 5          3.10        -73.98711        40.73317          1
## 6          2.17        -73.99520        40.73949          1
##   store_and_fwd_flag dropoff_longitude dropoff_latitude payment_type
## 1                  N         -73.97746         40.75398            2
## 2                  N         -73.98164         40.67024            1
## 3                  N         -74.00423         40.74217            1
## 4                  N         -73.92947         40.85154            1
## 5                  N         -73.98591         40.76645            1
## 6                  N         -73.99320         40.76264            1
##   fare_amount extra mta_tax tip_amount tolls_amount improvement_surcharge
## 1         6.0   0.5     0.5       0.00            0                   0.3
## 2        22.0   0.5     0.5       4.00            0                   0.3
## 3         6.5   0.5     0.5       1.56            0                   0.3
## 4        26.0   0.5     0.5       1.00            0                   0.3
## 5        13.5   0.5     0.5       2.96            0                   0.3
## 6        10.5   0.5     0.5       2.36            0                   0.3
##   total_amount
## 1         7.30
## 2        27.30
## 3         9.36
## 4        28.30
## 5        17.76
## 6        14.16
```

The data set includes 11135470 obs. of 19 variables

Data Filtering

We filter the data set to analyze peak hours of the morning. Assuming 8 AM to 9 AM to be the peak travelling hour for people going to work, we rewrite the data set with the filter. We then convert the pick up and drop-off dates to the day of the week. Finally we include the relevant data and create the data set we work with. The resulting data has a few latitude and longitude outliers, and we therefore establish the limits of our plot within NYC coordinates: Minimum Latitude = 40.5774, Maximum Latitude = 40.9176, Minimum Longitude = -74.15, Maximum Longitude = -73.7004.

```
df<- with(df,df[ hour (df$tpep_pickup_datetime)>=8 & hour(df$tpep_pickup_datetime)< 9 & hour
(df$tpep_dropoff_datetime)>=8 & hour(df$tpep_dropoff_datetime)< 9 , ] )
head(df,10)
```

```
##        VendorID tpep_pickup_datetime tpep_dropoff_datetime passenger_count
## 146436        1  2016-06-01 08:00:00   2016-06-01 08:00:37               1
## 146437        2  2016-06-01 08:00:00   2016-06-01 08:08:08               2
## 146438        2  2016-06-01 08:00:00   2016-06-01 08:11:48               6
## 146439        2  2016-06-01 08:00:00   2016-06-01 08:14:35               1
## 146440        2  2016-06-01 08:00:00   2016-06-01 08:05:22               1
## 146441        2  2016-06-01 08:00:00   2016-06-01 08:06:23               1
## 146442        1  2016-06-01 08:00:01   2016-06-01 08:14:07               1
## 146443        1  2016-06-01 08:00:01   2016-06-01 08:08:48               1
## 146444        2  2016-06-01 08:00:01   2016-06-01 08:15:15               1
## 146445        2  2016-06-01 08:00:01   2016-06-01 08:12:37               2
##        trip_distance pickup_longitude pickup_latitude RatecodeID
## 146436          1.30        -73.98035        40.74580          1
## 146437          1.39        -73.97506        40.79034          1
## 146438          0.72        -73.97356        40.76101          1
## 146439          2.31        -74.00020        40.74219          1
## 146440          0.66        -73.98213        40.77029          1
## 146441          0.72        -73.97213        40.75356          1
## 146442          2.00        -73.95631        40.76780          1
## 146443          1.20        -73.98000        40.74558          1
## 146444          2.15        -73.96409        40.77639          1
## 146445          0.84        -73.98239        40.76849          1
##        store_and_fwd_flag dropoff_longitude dropoff_latitude payment_type
## 146436                  N         -73.98153         40.74644            1
## 146437                  N         -73.97030         40.78390            1
## 146438                  N         -73.98186         40.76588            2
## 146439                  N         -73.97338         40.75508            1
## 146440                  N         -73.98729         40.77891            2
## 146441                  N         -73.98418         40.75870            1
## 146442                  N         -73.96792         40.78740            1
## 146443                  N         -73.99100         40.73148            1
## 146444                  N         -73.96641         40.79074            2
## 146445                  N         -73.99028         40.77182            1
##        fare_amount extra mta_tax tip_amount tolls_amount
## 146436         2.5     0     0.5       0.65            0
## 146437         7.5     0     0.5       1.66            0
## 146438         8.5     0     0.5       0.00            0
## 146439        11.0     0     0.5       1.00            0
## 146440         5.5     0     0.5       0.00            0
## 146441         6.0     0     0.5       0.50            0
## 146442        11.0     0     0.5       2.35            0
## 146443         7.5     0     0.5       2.05            0
## 146444        11.5     0     0.5       0.00            0
## 146445         9.0     0     0.5       1.96            0
##        improvement_surcharge total_amount
## 146436                   0.3         3.95
## 146437                   0.3         9.96
## 146438                   0.3         9.30
## 146439                   0.3        12.80
## 146440                   0.3         6.30
## 146441                   0.3         7.30
## 146442                   0.3        14.15
## 146443                   0.3        10.35
```

```
## 146444                  0.3        12.30
## 146445                  0.3        11.76
```

```
summary(df)
```

```
##     VendorID              tpep_pickup_datetime
##  Min.   :1.000   2016-06-21 08:13:12:    20
##  1st Qu.:1.000   2016-06-01 08:36:11:    17
##  Median :2.000   2016-06-03 08:10:57:    17
##  Mean   :1.531   2016-06-03 08:13:35:    17
##  3rd Qu.:2.000   2016-06-03 08:17:01:    17
##  Max.   :2.000   2016-06-03 08:17:28:    17
##                  (Other)            :390581
##         tpep_dropoff_datetime passenger_count trip_distance
##  2016-06-01 08:55:44:    18    Min.   :0.000   Min.   : 0.000
##  2016-06-01 08:50:04:    17    1st Qu.:1.000   1st Qu.: 0.900
##  2016-06-03 08:43:45:    17    Median :1.000   Median : 1.400
##  2016-06-03 08:44:06:    17    Mean   :1.597   Mean   : 2.078
##  2016-06-20 08:44:08:    17    3rd Qu.:1.000   3rd Qu.: 2.300
##  2016-06-02 08:44:20:    16    Max.   :9.000   Max.   :32.000
##  (Other)            :390584
##  pickup_longitude pickup_latitude   RatecodeID    store_and_fwd_flag
##  Min.   :-79.10   Min.   : 0.00   Min.   : 1.00   N:388728
##  1st Qu.:-73.99   1st Qu.:40.74   1st Qu.: 1.00   Y:  1958
##  Median :-73.98   Median :40.76   Median : 1.00
##  Mean   :-73.08   Mean   :40.26   Mean   : 1.02
##  3rd Qu.:-73.96   3rd Qu.:40.77   3rd Qu.: 1.00
##  Max.   :  0.00   Max.   :43.43   Max.   :99.00
##
##  dropoff_longitude dropoff_latitude  payment_type    fare_amount
##  Min.   :-118.19   Min.   : 0.00    Min.   :1.000   Min.   :-60.00
##  1st Qu.: -73.99   1st Qu.:40.74    1st Qu.:1.000   1st Qu.:  6.00
##  Median : -73.98   Median :40.76    Median :1.000   Median :  8.50
##  Mean   : -73.16   Mean   :40.31    Mean   :1.295   Mean   : 10.44
##  3rd Qu.: -73.97   3rd Qu.:40.77    3rd Qu.:2.000   3rd Qu.: 12.00
##  Max.   :   0.00   Max.   :45.63    Max.   :4.000   Max.   :450.00
##
##      extra              mta_tax          tip_amount       tolls_amount
##  Min.   :-0.500000   Min.   :-0.5000   Min.   : -8.000   Min.   :-5.5400
##  1st Qu.: 0.000000   1st Qu.: 0.5000   1st Qu.:  0.000   1st Qu.: 0.0000
##  Median : 0.000000   Median : 0.5000   Median :  1.280   Median : 0.0000
##  Mean   : 0.000304   Mean   : 0.4983   Mean   :  1.521   Mean   : 0.1489
##  3rd Qu.: 0.000000   3rd Qu.: 0.5000   3rd Qu.:  2.060   3rd Qu.: 0.0000
##  Max.   : 4.500000   Max.   : 0.5000   Max.   :175.000   Max.   :55.5500
##
##  improvement_surcharge  total_amount
##  Min.   :-0.3000       Min.   :-62.67
##  1st Qu.: 0.3000       1st Qu.:  7.80
##  Median : 0.3000       Median : 10.55
##  Mean   : 0.2997       Mean   : 12.91
##  3rd Qu.: 0.3000       3rd Qu.: 14.75
##  Max.   : 0.3000       Max.   :490.80
##
```

```
View(df)

df$day<- strptime(df[,2],"%Y-%m-%d")
df$day<-weekdays(df$day)
df1 <- subset(df,df$weekday == "Saturday" | df$weekday == "Sunday")
df2 <- subset(df,df$weekday != "Saturday" & df$weekday != "Sunday")
View(df)

df<-data.frame(df[c(2,3,5:7,10,11,19,20)])
head(df,10)
```

```
##        tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 146436  2016-06-01 08:00:00   2016-06-01 08:00:37          1.30
## 146437  2016-06-01 08:00:00   2016-06-01 08:08:08          1.39
## 146438  2016-06-01 08:00:00   2016-06-01 08:11:48          0.72
## 146439  2016-06-01 08:00:00   2016-06-01 08:14:35          2.31
## 146440  2016-06-01 08:00:00   2016-06-01 08:05:22          0.66
## 146441  2016-06-01 08:00:00   2016-06-01 08:06:23          0.72
## 146442  2016-06-01 08:00:01   2016-06-01 08:14:07          2.00
## 146443  2016-06-01 08:00:01   2016-06-01 08:08:48          1.20
## 146444  2016-06-01 08:00:01   2016-06-01 08:15:15          2.15
## 146445  2016-06-01 08:00:01   2016-06-01 08:12:37          0.84
##        pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 146436        -73.98035        40.74580         -73.98153         40.74644
## 146437        -73.97506        40.79034         -73.97030         40.78390
## 146438        -73.97356        40.76101         -73.98186         40.76588
## 146439        -74.00020        40.74219         -73.97338         40.75508
## 146440        -73.98213        40.77029         -73.98729         40.77891
## 146441        -73.97213        40.75356         -73.98418         40.75870
## 146442        -73.95631        40.76780         -73.96792         40.78740
## 146443        -73.98000        40.74558         -73.99100         40.73148
## 146444        -73.96409        40.77639         -73.96641         40.79074
## 146445        -73.98239        40.76849         -73.99028         40.77182
##        total_amount       day
## 146436         3.95 Wednesday
## 146437         9.96 Wednesday
## 146438         9.30 Wednesday
## 146439        12.80 Wednesday
## 146440         6.30 Wednesday
## 146441         7.30 Wednesday
## 146442        14.15 Wednesday
## 146443        10.35 Wednesday
## 146444        12.30 Wednesday
## 146445        11.76 Wednesday
```

```
df<- df %>% filter(df$trip_distance < 5)
View(df)

sprintf("# of Rows in Dataframe: %s", nrow(df))
```

```
## [1] "# of Rows in Dataframe: 362161"
```

```
sprintf("Dataframe Size: %s", format(object.size(df), units = "MB"))
```

```
## [1] "Dataframe Size: 387.9 Mb"
```

```
min_lat <- 40.5774
max_lat <- 40.9176
min_long <- -74.15
max_long <- -73.7004
```

The filtered data set now consists of 9.4 million entries.

Establishing Plot themes and parameters

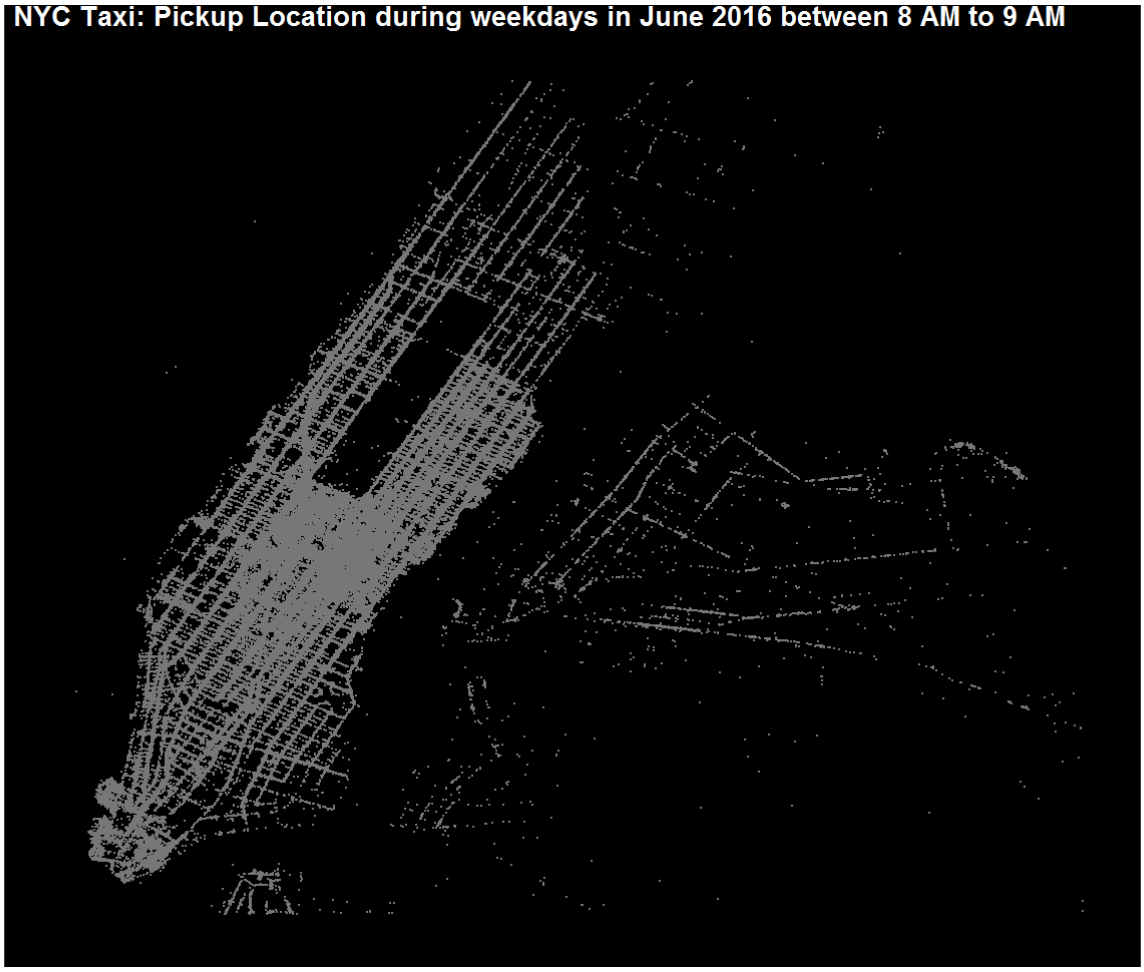
We create a theme for our plots with a black background and establish plot margins, title sizes and font sizes, and legend particulars

```
theme_map_dark <- function(palate_color = "Greys") {

  palate <- brewer.pal(palate_color, n=12)
  color.background = "black"
  color.grid.minor = "black"
  color.grid.major = "black"
  color.axis.text = palate[1]
  color.axis.title = palate[1]
  color.title = palate[1]

  font.title <- "Source Sans Pro"
  font.axis <- "Open Sans Condensed Bold"

  theme_bw(base_size=5) +
    theme(panel.background=element_rect(fill=color.background, color=color.background)) +
    theme(plot.background=element_rect(fill=color.background, color=color.background)) +
    theme(panel.border=element_rect(color=color.background)) +
    theme(panel.grid.major=element_blank()) +
    theme(panel.grid.minor=element_blank()) +
    theme(axis.ticks=element_blank()) +
    theme(plot.title=element_text(colour=color.title,family=font.title, size=10, face = 'bold'))
 +
    theme(axis.text.x=element_blank()) +
    theme(axis.text.y=element_blank()) +
    theme(axis.title.y=element_blank()) +
    theme(axis.title.x=element_blank()) +
    theme(plot.margin = unit(c(0.0, 0.5, 1, 0.75), "mm")) +
    theme(strip.background = element_rect(fill=color.background, color=color.background),strip.t
ext=element_text(size=7,colour=color.axis.title,family=font.title))

}
```

Data Analysis and Plot

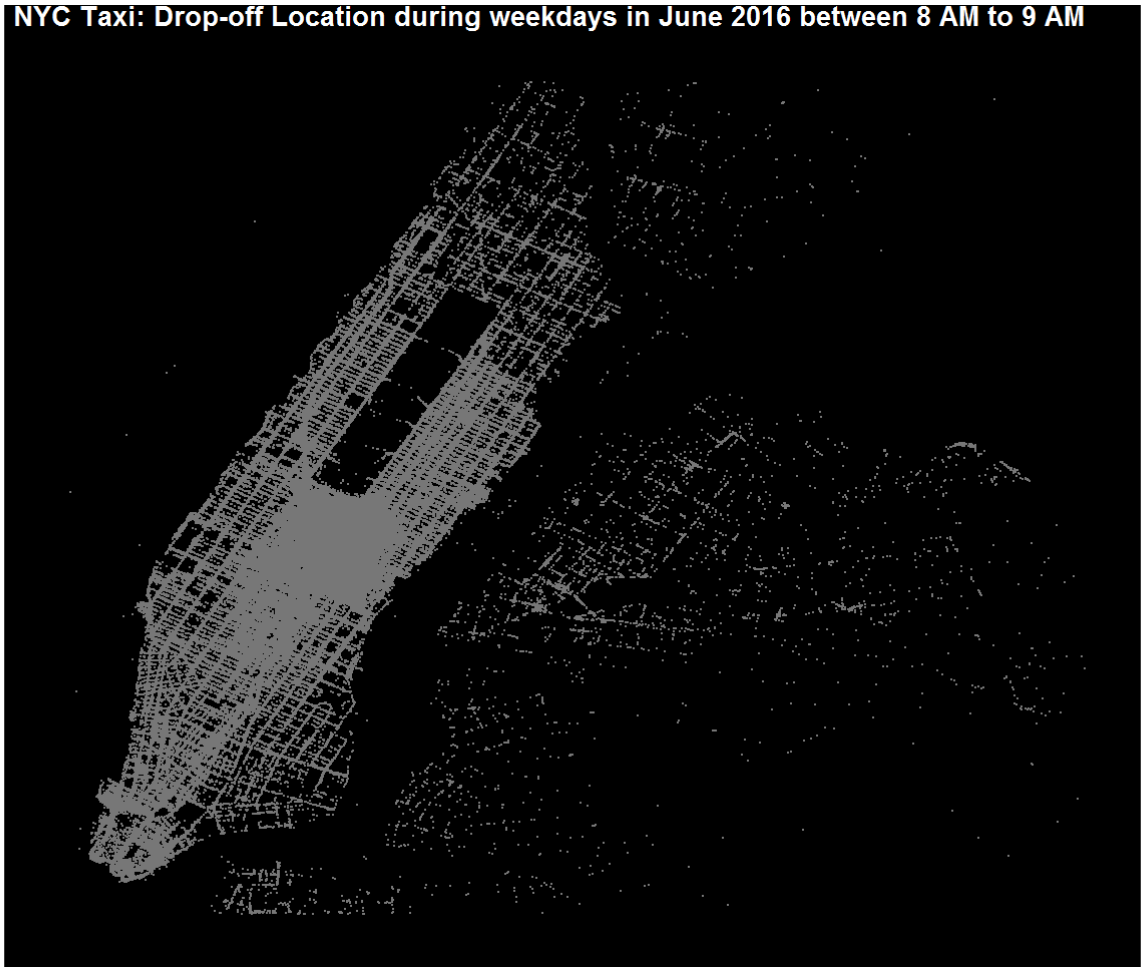
We analyze the data on the basis of pickup, drop-offs and whether it is a weekday or a weekend and plot 4 graphs which identify places in Manhattan that are the busiest.

```
######WEEKDAY#######

dfweekday<- df %>% filter(df$day!="Saturday" & df$day!= "Sunday")

#Pickup

plot1 <- ggplot(data=dfweekday , aes(x=dfweekday$pickup_longitude, y=dfweekday$pickup_latitude,
z=dfweekday$tpep_pickup_datetime)) +
  geom_point(size=0.06, color="#777777") +
  scale_x_continuous(limits=c(-74.0224, -73.8521)) +
  scale_y_continuous(limits=c(40.6959, 40.8348)) +
  theme_map_dark() +
  labs(title = "NYC Taxi: Pickup Location during weekdays in June 2016 between 8 AM to 9 AM") +
  coord_equal()
print(plot1)
```



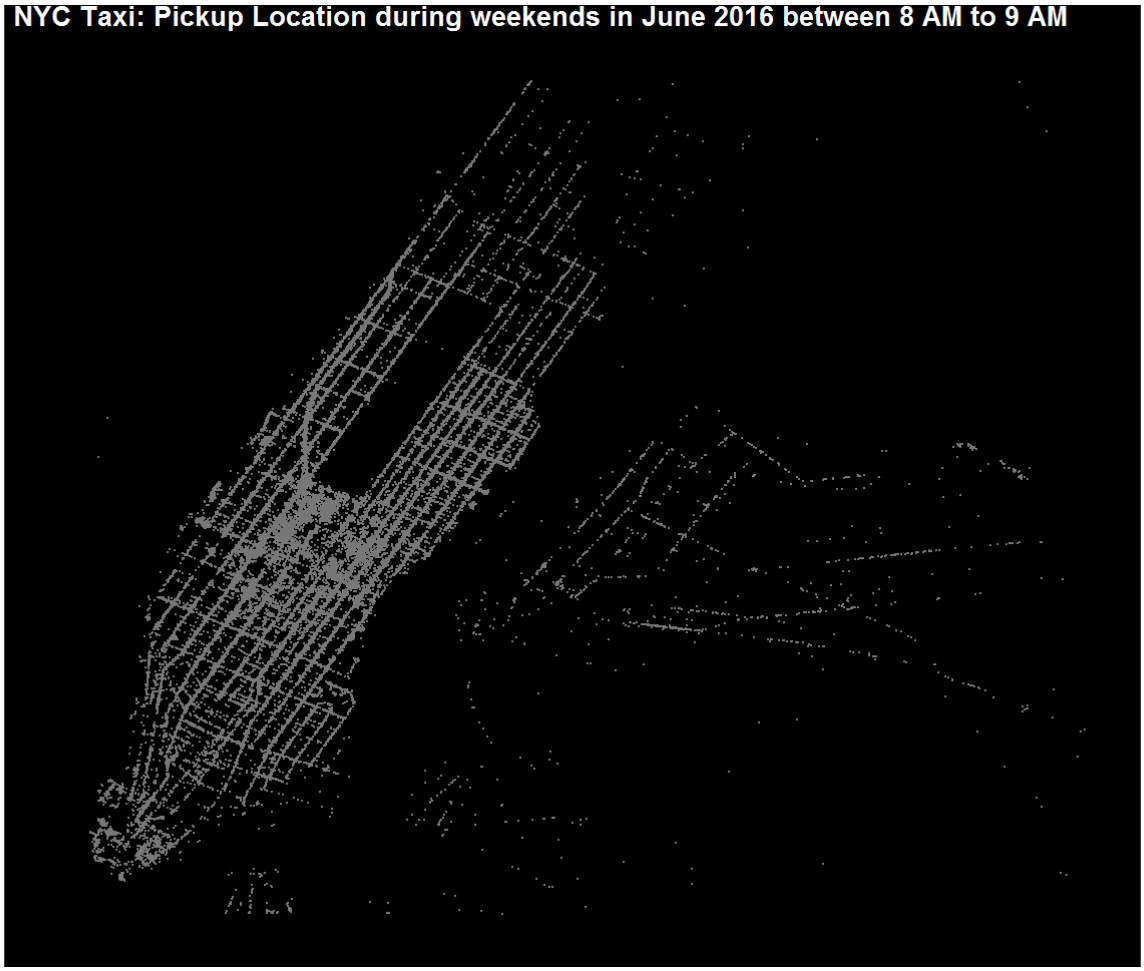
The plot above shows a high concentration of Taxi Pickups in Midtown East, Sutton Place, and areas between 6th and 8th Avenues and W42nd and W58th Streets.

```
#Drop-Off

plot2 <- ggplot(data=dfweekday , aes(x=dfweekday$dropoff_longitude,
y=dfweekday$dropoff_latitude, z=dfweekday$tpep_dropoff_datetime )) +
  geom_point(size=0.06, color="#777777") +
  stat_summary_hex(fun = dfweekday$tpep_dropoff_datetime, bins= 100, alpha=0.5) +
  scale_x_continuous(limits=c(-74.0224, -73.8521)) +
  scale_y_continuous(limits=c(40.6959, 40.8348)) +
  theme_map_dark() +
  scale_fill_gradient(low="#CCCCCC", high="#8E44AD", trans="log", breaks=c("00:30")) +
  labs(title = "NYC Taxi: Drop-off Location during weekdays in June 2016 between 8 AM to 9 AM")
+
  coord_equal()
print(plot2)
```



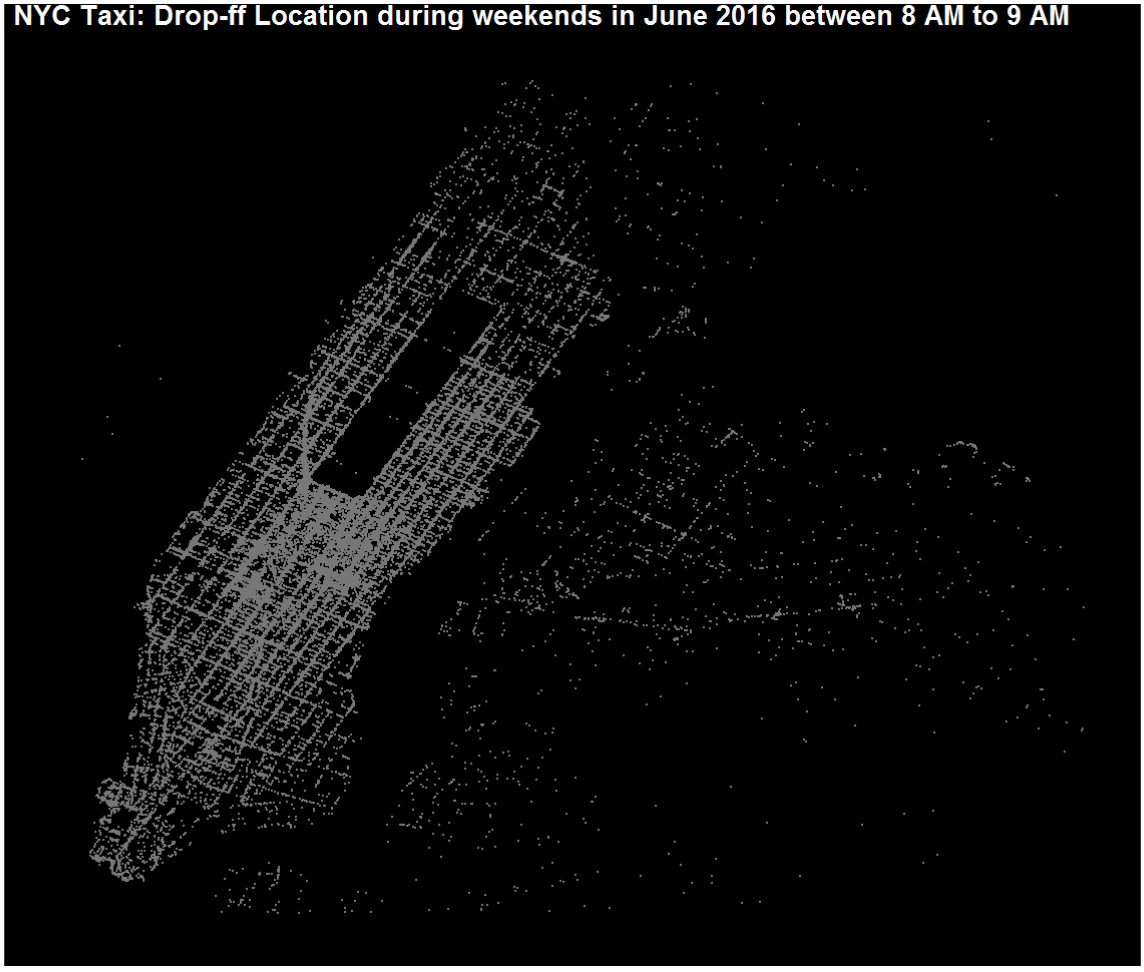
The plot above shows a high concentration of Taxi Drop-offs in Central and Eastern Midtown, some of which include the Garment District, Murray Hill, the Times Square Area, Columbus Circle, etc.

```
#####WEEKEND######

dfweekend<- df %>% filter(df$day=="Saturday" | df$day== "Sunday")

#Pickup

plot3 <- ggplot(data=dfweekend , aes(x=dfweekend$pickup_longitude, y=dfweekend$pickup_latitude,
z=dfweekend$tpep_pickup_datetime)) +
  geom_point(size=0.06, color="#777777") +
  stat_summary_hex(fun = dfweekend$tpep_pickup_datetime, bins= 100, alpha=0.5) +
  scale_x_continuous(limits=c(-74.0224, -73.8521)) +
  scale_y_continuous(limits=c(40.6959, 40.8348)) +
  theme_map_dark() +
  scale_fill_gradient(low="#FFFFFF", high="#E74C3C", trans="log", breaks=c("00:30")) +
  labs(title = "NYC Taxi: Pickup Location during weekends in June 2016 between 8 AM to 9 AM") +
  coord_equal()
print(plot3)
```



The weekend pickups show a relatively sparse distribution. Some of the areas with comparatively high pickups include the Times Square Area (a lot of hotels) and Upper Midtown on 5th Avenue (shopping).

```
#Drop-Off

plot4 <- ggplot(data=dfweekend , aes(x=dfweekend$dropoff_longitude,
y=dfweekend$dropoff_latitude, z=dfweekend$tpep_dropoff_datetime )) +
  geom_point(size=0.06, color="#777777") +
  stat_summary_hex(fun = dfweekend$tpep_dropoff_datetime, bins= 100, alpha=0.5) +
  scale_x_continuous(limits=c(-74.0224, -73.8521)) +
  scale_y_continuous(limits=c(40.6959, 40.8348)) +
  theme_map_dark() +
  scale_fill_gradient(low="#FFFFFF", high="#E74C3C", trans="log", breaks=c("00:30")) +
  labs(title = "NYC Taxi: Drop-ff Location during weekends in June 2016 between 8 AM to 9 AM") +
  coord_equal()
print(plot4)
```



Drop-offs during the weekends are concentrated mostly around the tourist attractions, some of which include Columbus Circle, Times Square and the Rockerfeller Center. People also travel to the St. Patricks Cathedral area on 50th street and Madison Avenue.

Calculation of time in spent in traffic

We consider 3 neighborhoods for this analysis - people who travel from Midtown East to the Garment District and from Midtown East to the Times Square area on weekdays and weekends respectively.

We define these areas according to the following coordinates:

Midtown East = Min Long: -73.9808, Max Long: -73.9591 & Min Lat: 40.7480, Max Lat: 40.7643 #### Garment District = Min Long: -73.9963, Max Long: -73.9841 & Min Lat: 40.7478, Max Lat: 40.7583 #### Times Square = Min Long: -73.9916, Max Long:-73.9808 & Min Lat: 40.7542, Max Lat: 40.7613

We calculate the average time differences between these coordinates on weekdays and weekends. Assuming that there is little traffic during the weekends, the average difference would give us the extra time a person spends sitting in traffic during the weekdays.

```
##Creating a new column for pickup and dropp-off times during weekdays and weekends

Hours1 <- format(as.POSIXct(strptime(dfweekday$tpep_pickup_datetime,"%Y-%m-%d  %H:%M:%S",tz="")) 
,format = "%H:%M")
head(Hours1)
```

```
## [1] "08:00" "08:00" "08:00" "08:00" "08:00" "08:00"
```

```
dfweekday$pickuptime<-Hours1

Hours2 <- format(as.POSIXct(strptime(dfweekday$tpep_dropoff_datetime,"%Y-%m-%d 
%H:%M:%S",tz="")) ,format = "%H:%M")
head(Hours2)
```

```
## [1] "08:00" "08:08" "08:11" "08:14" "08:05" "08:06"
```

```
dfweekday$dropofftime<-Hours2

Hours3 <- format(as.POSIXct(strptime(dfweekend$tpep_pickup_datetime,"%Y-%m-%d  %H:%M:%S",tz="")) 
,format = "%H:%M")
head(Hours3)
```

```
## [1] "08:07" "08:07" "08:07" "08:07" "08:07" "08:07"
```

```
dfweekend$pickuptime<-Hours3

Hours4 <- format(as.POSIXct(strptime(dfweekend$tpep_dropoff_datetime,"%Y-%m-%d 
%H:%M:%S",tz="")) ,format = "%H:%M")
head(Hours4)
```

```
## [1] "08:24" "08:16" "08:26" "08:20" "08:09" "08:11"
```

```
dfweekend$dropofftime<-Hours4


## Converting time to Hour and Minute format to be read by R and creating a column for time diff
erences

###Weekdays

pickupWeekday<-as.POSIXct(dfweekday$pickuptime,format="%H:%M")
dropoffWeekday<-as.POSIXct(dfweekday$dropofftime,format="%H:%M")

dfweekday$diffinmin <- difftime(dropoffWeekday,pickupWeekday,tz,units = "mins")

###Weekends

pickupWeekend<-as.POSIXct(dfweekend$pickuptime,format="%H:%M")
dropoffWeekend<-as.POSIXct(dfweekend$dropofftime,format="%H:%M")

dfweekend$diffinmin <- difftime(dropoffWeekend,pickupWeekend,tz,units = "mins")
```

Midtown East to Times Square

```
#For Weekdays (Monday to Friday)

M_T_Weekday<- subset(dfweekday,subset= dfweekday$pickup_longitude >=-73.9808 & dfweekday$pickup_
longitude <=-73.9591
             & dfweekday$pickup_latitude >= 40.7480 & dfweekday$pickup_latitude <= 40.7643
             & dfweekday$dropoff_longitude >= -73.9916  & dfweekday$dropoff_longitude <= -73.980
8
             & dfweekday$dropoff_latitude >= 40.7542 & dfweekday$dropoff_latitude <= 40.7613
)

head(M_T_Weekday)
```

```
##     tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 6    2016-06-01 08:00:00   2016-06-01 08:06:23          0.72
## 23   2016-06-01 08:00:05   2016-06-01 08:12:49          1.15
## 40   2016-06-01 08:00:09   2016-06-01 08:09:36          1.20
## 169  2016-06-01 08:03:42   2016-06-01 08:08:12          0.00
## 208  2016-06-01 08:00:32   2016-06-01 08:07:23          0.63
## 222  2016-06-01 08:00:34   2016-06-01 08:15:01          1.66
##     pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 6          -73.97213        40.75356         -73.98418         40.75870
## 23         -73.97250        40.76207         -73.98604         40.75847
## 40         -73.96573        40.75453         -73.98104         40.75611
## 169        -73.97315        40.75353         -73.98140         40.75775
## 208        -73.97465        40.75735         -73.98412         40.75770
## 222        -73.96870        40.76147         -73.99004         40.75647
##     total_amount       day pickuptime dropofftime diffinmin
## 6           7.30 Wednesday      08:00       08:06    6 mins
## 23         10.30 Wednesday      08:00       08:12   12 mins
## 40          9.95 Wednesday      08:00       08:09    9 mins
## 169         6.89 Wednesday      08:03       08:08    5 mins
## 208         7.82 Wednesday      08:00       08:07    7 mins
## 222        11.30 Wednesday      08:00       08:15   15 mins
```

```
mean(M_T_Weekday$trip_distance)
```

```
## [1] 0.9566856
```

```
#For Weekends (Saturday and Sunday)

M_T_Weekend<- subset(dfweekend,subset= dfweekend$pickup_longitude >=-73.9808 & dfweekend$pickup_
longitude <=-73.9591
             & dfweekend$pickup_latitude >=40.7480 & dfweekend$pickup_latitude <=40.7643
             & dfweekend$dropoff_longitude >=-73.9916  & dfweekend$dropoff_longitude <=-73.9808
             & dfweekend$dropoff_latitude >=40.7542 & dfweekend$dropoff_latitude <=40.7613
)
head(M_T_Weekend)
```

```
##     tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 213  2016-06-04 08:17:08   2016-06-04 08:27:16          1.47
## 240  2016-06-04 08:17:20   2016-06-04 08:22:07          0.90
## 363  2016-06-04 08:24:32   2016-06-04 08:27:16          0.53
## 476  2016-06-04 08:19:02   2016-06-04 08:25:18          1.10
## 501  2016-06-04 08:19:11   2016-06-04 08:25:56          1.61
## 707  2016-06-04 08:00:26   2016-06-04 08:04:28          0.90
##     pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 213        -73.97655        40.74805         -73.98595         40.75878
## 240        -73.97714        40.75862         -73.98866         40.75683
## 363        -73.97950        40.76386         -73.98590         40.75756
## 476        -73.97482        40.76241         -73.98942         40.75680
## 501        -73.97022        40.75955         -73.99081         40.75541
## 707        -73.96986        40.76403         -73.98162         40.75768
##     total_amount      day pickuptime dropofftime diffinmin
## 213         8.80 Saturday      08:17       08:27   10 mins
## 240         6.30 Saturday      08:17       08:22    5 mins
## 363         5.76 Saturday      08:24       08:27    3 mins
## 476         7.30 Saturday      08:19       08:25    6 mins
## 501         9.00 Saturday      08:19       08:25    6 mins
## 707         7.55 Saturday      08:00       08:04    4 mins
```

```
mean(M_T_Weekend$trip_distance)
```

```
## [1] 1.225527
```

```
##Calculating extra time taken during weekdays

mean(M_T_Weekend$diffinmin)
```

```
## Time difference of 7.85755 mins
```

```
mean(M_T_Weekday$diffinmin)
```

```
## Time difference of 11.07932 mins
```

```
Time_Difference1 <- mean(M_T_Weekday$diffinmin) - mean(M_T_Weekend$diffinmin)
Time_Difference1
```

```
## Time difference of 3.22177 mins
```

It takes an average of 3.22 mins more to get to Times Square from Midtown East on a weekday between 8AM and 9AM.

Midtown to Garmet District

```
#For Weekdays (Monday to Friday)

M_G_Weekday<- subset(dfweekday,subset= dfweekday$pickup_longitude >=-73.9808 & dfweekday$pickup_
longitude <=-73.9591
            & dfweekday$pickup_latitude >=40.7480 & dfweekday$pickup_latitude <=40.7643
            & dfweekday$dropoff_longitude >=-73.9963  & dfweekday$dropoff_longitude <=-73.9841
            & dfweekday$dropoff_latitude >=40.7478 & dfweekday$dropoff_latitude <=40.7583
)
head(M_G_Weekday)
```

```
##     tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 208  2016-06-01 08:00:32   2016-06-01 08:07:23          0.63
## 222  2016-06-01 08:00:34   2016-06-01 08:15:01          1.66
## 361  2016-06-01 08:03:54   2016-06-01 08:20:11          0.80
## 533  2016-06-01 08:04:01   2016-06-01 08:08:39          0.82
## 543  2016-06-01 08:04:03   2016-06-01 08:09:48          0.91
## 615  2016-06-01 08:01:31   2016-06-01 08:07:41          0.80
##     pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 208        -73.97465        40.75735         -73.98412         40.75770
## 222        -73.96870        40.76147         -73.99004         40.75647
## 361        -73.97387        40.74913         -73.98521         40.75288
## 533        -73.97863        40.75222         -73.98804         40.75214
## 543        -73.97594        40.76010         -73.98429         40.74857
## 615        -73.97821        40.75216         -73.98733         40.75114
##     total_amount       day pickuptime dropofftime diffinmin
## 208         7.82 Wednesday      08:00       08:07    7 mins
## 222        11.30 Wednesday      08:00       08:15   15 mins
## 361        14.12 Wednesday      08:03       08:20   17 mins
## 533         6.96 Wednesday      08:04       08:08    4 mins
## 543         8.16 Wednesday      08:04       08:09    5 mins
## 615         6.80 Wednesday      08:01       08:07    6 mins
```

```
mean(M_G_Weekday$trip_distance)
```

```
## [1] 1.207938
```

```
#For Weekends (Saturday and Sunday)

M_G_Weekend <- subset(dfweekend,subset= dfweekend$pickup_longitude >= -73.9808 & dfweekend$picku
p_longitude <= -73.9591
            & dfweekend$pickup_latitude >= 40.7480 & dfweekend$pickup_latitude <= 40.7643
            & dfweekend$dropoff_longitude >= -73.9963  & dfweekend$dropoff_longitude <= -73.984
1
            & dfweekend$dropoff_latitude >= 40.7478 & dfweekend$dropoff_latitude <= 40.7583
)

head(M_G_Weekend)
```

```
##     tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 70   2016-06-04 08:07:48   2016-06-04 08:12:49          1.08
## 149  2016-06-04 08:16:38   2016-06-04 08:24:53          1.43
## 240  2016-06-04 08:17:20   2016-06-04 08:22:07          0.90
## 280  2016-06-04 08:17:43   2016-06-04 08:22:50          1.01
## 363  2016-06-04 08:24:32   2016-06-04 08:27:16          0.53
## 381  2016-06-04 08:18:24   2016-06-04 08:26:45          1.70
##     pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 70         -73.97872        40.74991         -73.99385         40.74988
## 149        -73.97296        40.74910         -73.99301         40.74968
## 240        -73.97714        40.75862         -73.98866         40.75683
## 280        -73.97754        40.75455         -73.99055         40.75113
## 363        -73.97950        40.76386         -73.98590         40.75756
## 381        -73.96922        40.76064         -73.99085         40.75092
##     total_amount      day pickuptime dropofftime diffinmin
## 70          6.80 Saturday      08:07       08:12    5 mins
## 149         8.30 Saturday      08:16       08:24    8 mins
## 240         6.30 Saturday      08:17       08:22    5 mins
## 280         7.56 Saturday      08:17       08:22    5 mins
## 363         5.76 Saturday      08:24       08:27    3 mins
## 381        10.56 Saturday      08:18       08:26    8 mins
```

```
View(M_G_Weekend)
mean(M_G_Weekend$trip_distance)
```

```
## [1] 1.518019
```

```
mean(M_G_Weekend$diffinmin)
```

```
## Time difference of 9.068339 mins
```

```
mean(M_G_Weekday$diffinmin)
```

```
## Time difference of 12.88285 mins
```

```
Time_Difference2 <- mean(M_G_Weekday$diffinmin)-mean(M_G_Weekend$diffinmin)
Time_Difference2
```

```
## Time difference of 3.814514 mins
```

It takes an average of 3.82 mins longer on a weekday between 8AM and 9AM, to reach the Garment District from Midtown East.