

final ba1

Arnav

12/12/2016

```
dfmain <- read.csv("/Users/arnavsomani/Desktop/NYU COURSE/sem 3/ba/r programming csv files/06yellow.csv", header=T,stringsAsFactors=F)
```

```
head(dfmain,10)
```

```

## 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
## 7 2 2016-06-09 21:06:36 2016-06-09 21:30:13 5
## 8 1 2016-06-09 21:06:37 2016-06-09 21:16:47 1
## 9 1 2016-06-09 21:06:37 2016-06-09 21:15:44 1
## 10 1 2016-06-09 21:06:37 2016-06-09 21:23:57 2
## 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
## 7 6.02 -73.98048 40.74168 1
## 8 1.40 -73.98570 40.74674 1
## 9 1.20 -73.98428 40.77501 1
## 10 1.90 -73.97792 40.75009 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
## 7 N -73.90296 40.74874 2
## 8 N -73.98251 40.76277 1
## 9 N -73.96850 40.76643 2
## 10 N -73.98724 40.74936 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
## 7 21.5 0.5 0.5 0.00 0 0.3
## 8 8.5 0.5 0.5 1.95 0 0.3
## 9 8.0 0.5 0.5 0.00 0 0.3
## 10 12.0 0.5 0.5 3.33 0 0.3
## total_amount
## 1 7.30
## 2 27.30
## 3 9.36
## 4 28.30
## 5 17.76
## 6 14.16
## 7 22.80
## 8 11.75

```

```
## 9          9.30
## 10         16.63
```

```
summary(dfmain)
```

```
##      VendorID      tpep_pickup_datetime tpep_dropoff_datetime passenger_count
## Min.      :1.00      Length:11135470      Length:11135470      Min.      :0.000
## 1st Qu.:1.00      Class :character      Class :character      1st Qu.:1.000
## Median :2.00      Mode  :character      Mode  :character      Median :1.000
## Mean     :1.53
## 3rd Qu.:2.00
## Max.     :2.00
## trip_distance      pickup_longitude      pickup_latitude      RatecodeID
## Min.      : 0.00      Min.      :-118.19      Min.      : 0.00      Min.      : 1.000
## 1st Qu.: 1.00      1st Qu.: -73.99      1st Qu.:40.74      1st Qu.: 1.000
## Median : 1.71      Median : -73.98      Median :40.75      Median : 1.000
## Mean     : 3.04      Mean     : -73.05      Mean     :40.24      Mean     : 1.044
## 3rd Qu.: 3.23      3rd Qu.: -73.97      3rd Qu.:40.77      3rd Qu.: 1.000
## Max.     :71732.70      Max.      : 0.00      Max.     :64.10      Max.     :99.000
## store_and_fwd_flag dropoff_longitude dropoff_latitude      payment_type
## Length:11135470      Min.      :-118.19      Min.      : 0.00      Min.      :1.00
## Class :character      1st Qu.: -73.99      1st Qu.:40.73      1st Qu.:1.00
## Mode  :character      Median : -73.98      Median :40.75      Median :1.00
##                               Mean     : -73.12      Mean     :40.28      Mean     :1.35
##                               3rd Qu.: -73.96      3rd Qu.:40.77      3rd Qu.:2.00
##                               Max.      :106.25      Max.     :60.04      Max.     :5.00
## fare_amount          extra          mta_tax          tip_amount
## Min.      : -450.0      Min.      :-41.2300      Min.      :-2.7000      Min.      : -67.700
## 1st Qu.: 6.5      1st Qu.: 0.0000      1st Qu.: 0.5000      1st Qu.: 0.000
## Median : 10.0      Median : 0.0000      Median : 0.5000      Median : 1.350
## Mean     : 13.5      Mean     : 0.3407      Mean     : 0.4973      Mean     : 1.842
## 3rd Qu.: 15.5      3rd Qu.: 0.5000      3rd Qu.: 0.5000      3rd Qu.: 2.460
## Max.     :628544.7      Max.     :597.9200      Max.     :60.3500      Max.     :854.850
## tolls_amount          improvement_surcharge      total_amount
## Min.      : -12.5000      Min.      : -0.3000      Min.      : -450.8
## 1st Qu.: 0.0000      1st Qu.: 0.3000      1st Qu.: 8.8
## Median : 0.0000      Median : 0.3000      Median : 12.3
## Mean     : 0.3402      Mean     : 0.2997      Mean     : 16.8
## 3rd Qu.: 0.0000      3rd Qu.: 0.3000      3rd Qu.: 18.4
## Max.     :970.0000      Max.     :11.6400      Max.     :629033.8
```

```
str(dfmain)
```

```
## 'data.frame':    11135470 obs. of  19 variables:
## $ VendorID          : int  2 2 2 2 2 2 2 1 1 1 ...
## $ tpep_pickup_datetime : chr  "2016-06-09 21:06:36" "2016-06-09 21:06:36" "2016-06-09 21:06:36" "2016-06-09 21:06:36" ...
## $ tpep_dropoff_datetime: chr  "2016-06-09 21:13:08" "2016-06-09 21:35:11" "2016-06-09 21:13:10" "2016-06-09 21:36:10" ...
## $ 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   : chr  "N" "N" "N" "N" ...
## $ 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 ...
```

```
library(lubridate)
```

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

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

```
df<-data.frame(dfmain[c(2,3,5,6,7,10,11)])
```

```
head(df,5)
```

```
##   tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 1 2016-06-09 21:06:36 2016-06-09 21:13:08          0.79
## 2 2016-06-09 21:06:36 2016-06-09 21:35:11          5.22
## 3 2016-06-09 21:06:36 2016-06-09 21:13:10          1.26
## 4 2016-06-09 21:06:36 2016-06-09 21:36:10          7.39
## 5 2016-06-09 21:06:36 2016-06-09 21:23:23          3.10
##   pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 1      -73.98336      40.76094      -73.97746      40.75398
## 2      -73.98172      40.73667      -73.98164      40.67024
## 3      -73.99432      40.75107      -74.00423      40.74217
## 4      -73.98236      40.77389      -73.92947      40.85154
## 5      -73.98711      40.73317      -73.98591      40.76645
```

```
str(df)
```

```
## 'data.frame':    11135470 obs. of  7 variables:
## $ tpep_pickup_datetime : chr  "2016-06-09 21:06:36" "2016-06-09 21:06:36" "2016-06-09 21:06:36" "2016-06-09 21:06:36" ...
## $ tpep_dropoff_datetime: chr  "2016-06-09 21:13:08" "2016-06-09 21:35:11" "2016-06-09 21:13:10" "2016-06-09 21:36:10" ...
## $ 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 ...
## $ dropoff_longitude    : num  -74 -74 -74 -73.9 -74 ...
## $ dropoff_latitude     : num  40.8 40.7 40.7 40.9 40.8 ...
```

```
library(lubridate)
str(df)
```

```
## 'data.frame':    11135470 obs. of  7 variables:
## $ tpep_pickup_datetime : chr  "2016-06-09 21:06:36" "2016-06-09 21:06:36" "2016-06-09 21:06:36" "2016-06-09 21:06:36" ...
## $ tpep_dropoff_datetime: chr  "2016-06-09 21:13:08" "2016-06-09 21:35:11" "2016-06-09 21:13:10" "2016-06-09 21:36:10" ...
## $ 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 ...
## $ dropoff_longitude    : num  -74 -74 -74 -73.9 -74 ...
## $ dropoff_latitude     : num  40.8 40.7 40.7 40.9 40.8 ...
```

```
dfsub<- with(df,df[ hour (df$tpep_pickup_datetime)>=8 & hour(df$tpep_pickup_datetime)< 9
& df$trip_distance<5 , ] )
```

```
head(dfsub,5)
```

```
##          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
##          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
```

```
dfsub$day<- strptime(dfsub[,1],"%Y-%m-%d")
dfsub$weekday<-weekdays(dfsub$day)
df1 <- subset(dfsub,dfsub$weekday == "Saturday" | dfsub$weekday == "Sunday")
df2 <- subset(dfsub,dfsub$weekday != "Saturday" & dfsub$weekday != "Sunday")
```

```
head(df1,5)
```

```
##          tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 1420121 2016-06-04 08:07:17 2016-06-04 08:24:39          3.60
## 1420123 2016-06-04 08:07:18 2016-06-04 08:16:46          4.00
## 1420124 2016-06-04 08:07:18 2016-06-04 08:26:58          4.60
## 1420125 2016-06-04 08:07:18 2016-06-04 08:20:16          4.07
## 1420126 2016-06-04 08:07:18 2016-06-04 08:09:07          0.38
##          pickup_longitude pickup_latitude dropoff_longitude
## 1420121      -73.99523      40.73369      -73.95515
## 1420123      -73.96806      40.75130      -73.93251
## 1420124      -74.00023      40.74293      -73.96124
## 1420125      -74.01493      40.71642      -73.98834
## 1420126      -73.99815      40.74599      -73.99481
##          dropoff_latitude      day weekday
## 1420121      40.76536 2016-06-04 Saturday
## 1420123      40.79488 2016-06-04 Saturday
## 1420124      40.80116 2016-06-04 Saturday
## 1420125      40.76658 2016-06-04 Saturday
## 1420126      40.75029 2016-06-04 Saturday
```

```
head(df2,5)
```

```
##          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
##          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
##          day weekday
## 146436 2016-06-01 Wednesday
## 146437 2016-06-01 Wednesday
## 146438 2016-06-01 Wednesday
## 146439 2016-06-01 Wednesday
## 146440 2016-06-01 Wednesday
```

```
str(df1)
```

```
## 'data.frame':    60042 obs. of  9 variables:
## $ tpep_pickup_datetime : chr  "2016-06-04 08:07:17" "2016-06-04 08:07:18" "2016-06-04 08:07:18" "2016-06-04 08:07:18" ...
## $ tpep_dropoff_datetime: chr  "2016-06-04 08:24:39" "2016-06-04 08:16:46" "2016-06-04 08:26:58" "2016-06-04 08:20:16" ...
## $ trip_distance       : num  3.6 4 4.6 4.07 0.38 1.1 1.6 1.2 1.76 1.16 ...
## $ pickup_longitude    : num  -74 -74 -74 -74 -74 ...
## $ pickup_latitude     : num  40.7 40.8 40.7 40.7 40.7 ...
## $ dropoff_longitude   : num  -74 -73.9 -74 -74 -74 ...
## $ dropoff_latitude    : num  40.8 40.8 40.8 40.8 40.8 ...
## $ day                 : POSIXlt, format: "2016-06-04" "2016-06-04" ...
## $ weekday             : chr  "Saturday" "Saturday" "Saturday" "Saturday" ...
```

```
Hours1 <- format(as.POSIXct(strptime(df1$tpep_pickup_datetime,"%Y-%m-%d %H:%M:%S",tz=""))
,format = "%H:%M")
head(Hours1,5)
```

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

```
df1$pickuptime<-Hours1
```

```
Hours2 <- format(as.POSIXct(strptime(df1$tpep_dropoff_datetime,"%Y-%m-%d %H:%M:%S",tz=""))
) ,format = "%H:%M")
head(Hours2,5)
```

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

```
df1$dropofftime<-Hours2
```

```
Hours3 <- format(as.POSIXct(strptime(df2$tpep_pickup_datetime,"%Y-%m-%d %H:%M:%S",tz=""))
,format = "%H:%M")
head(Hours3,5)
```

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

```
df2$pickuptime<-Hours3
```

```
Hours4 <- format(as.POSIXct(strptime(df2$tpep_dropoff_datetime,"%Y-%m-%d %H:%M:%S",tz=""))
) ,format = "%H:%M")
head(Hours4,5)
```

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

```
df2$dropofftime<-Hours4
```

Midtown East = lon: >=-73.9808 lon: <=-73.9591 lat: >=40.7480 lat: <=40.7643

Garment District = lon: >=-73.9963 lon: <=-73.9841 lat: >=40.7478 lat: <=40.7583

Times Square = lon: >=-73.9896 lon: <=-73.9855 lat: >=40.7563 lat: <=40.7603

Midtown to Garment District

For Weekends (Saturday and Sunday)

```
td1<- subset(df1,subset= df1$pickup_longitude >=-73.9808 & df1$pickup_longitude <=-73.9591
& df1$pickup_latitude >=40.7480 & df1$pickup_latitude <=40.7643
& df1$dropoff_longitude >=-73.9963 & df1$dropoff_longitude <=-73.9841
& df1$dropoff_latitude >=40.7478 & df1$dropoff_latitude <=40.7583
)
head(td1,5)
```

```
##          tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 1420434 2016-06-04 08:07:48 2016-06-04 08:12:49          1.08
## 1420711 2016-06-04 08:16:38 2016-06-04 08:24:53          1.43
## 1420901 2016-06-04 08:17:20 2016-06-04 08:22:07          0.90
## 1420947 2016-06-04 08:17:43 2016-06-04 08:22:50          1.01
## 1421109 2016-06-04 08:24:32 2016-06-04 08:27:16          0.53
##          pickup_longitude pickup_latitude dropoff_longitude
## 1420434      -73.97872      40.74991      -73.99385
## 1420711      -73.97296      40.74910      -73.99301
## 1420901      -73.97714      40.75862      -73.98866
## 1420947      -73.97754      40.75455      -73.99055
## 1421109      -73.97950      40.76386      -73.98590
##          dropoff_latitude      day weekday pickuptime dropofftime
## 1420434      40.74988 2016-06-04 Saturday      08:07      08:12
## 1420711      40.74968 2016-06-04 Saturday      08:16      08:24
## 1420901      40.75683 2016-06-04 Saturday      08:17      08:22
## 1420947      40.75113 2016-06-04 Saturday      08:17      08:22
## 1421109      40.75756 2016-06-04 Saturday      08:24      08:27
```

```
mean(td1$trip_distance)
```

```
## [1] 1.534835
```

Midtown to Garment District

For Weekdays (Monday to Friday)

```
td2<- subset(df2,subset= df2$pickup_longitude >=-73.9808 & df2$pickup_longitude <=-73.9591
& df2$pickup_latitude >=40.7480 & df2$pickup_latitude <=40.7643
& df2$dropoff_longitude >=-73.9963 & df2$dropoff_longitude <=-73.9841
& df2$dropoff_latitude >=40.7478 & df2$dropoff_latitude <=40.7583
)
head(td2,5)
```



```
##          tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 146810    2016-06-01 08:00:32    2016-06-01 08:07:23          0.63
## 146825    2016-06-01 08:00:34    2016-06-01 08:15:01          1.66
## 147062    2016-06-01 08:03:54    2016-06-01 08:20:11          0.80
## 147261    2016-06-01 08:04:01    2016-06-01 08:08:39          0.82
## 147272    2016-06-01 08:04:03    2016-06-01 08:09:48          0.91
##          pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 146810          -73.97465         40.75735         -73.98412         40.75770
## 146825          -73.96870         40.76147         -73.99004         40.75647
## 147062          -73.97387         40.74913         -73.98521         40.75288
## 147261          -73.97863         40.75222         -73.98804         40.75214
## 147272          -73.97594         40.76010         -73.98429         40.74857
##          day    weekday pickuptime dropofftime
## 146810 2016-06-01 Wednesday      08:00      08:07
## 146825 2016-06-01 Wednesday      08:00      08:15
## 147062 2016-06-01 Wednesday      08:03      08:20
## 147261 2016-06-01 Wednesday      08:04      08:08
## 147272 2016-06-01 Wednesday      08:04      08:09
```

```
mean(td2$trip_distance)
```

```
## [1] 1.257072
```

Midtown East = lon: >=-73.9808 lon: <=-73.9591 lat: >=40.7480 lat: <=40.7643

Garment District = lon: >=-73.9963 lon: <=-73.9841 lat: >=40.7478 lat: <=40.7583

Times Square = lon: >=-73.9896 lon: <=-73.9855 lat: >=40.7563 lat: <=40.7603

Midtown to Times Square

For Weekends (Saturday and Sunday)

```
td3<- subset(df1,subset= df1$pickup_longitude >=-73.9808 & df1$pickup_longitude <=-73.95
91
          & df1$pickup_latitude >=40.7480 & df1$pickup_latitude <=40.7643
          & df1$dropoff_longitude >=-73.9896 & df1$dropoff_longitude <=-73.9855
          & df1$dropoff_latitude >=40.7563 & df1$dropoff_latitude <=40.7603
)
head(td3,5)
```

```
##          tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 1420830 2016-06-04 08:17:08 2016-06-04 08:27:16 1.47
## 1420901 2016-06-04 08:17:20 2016-06-04 08:22:07 0.90
## 1421109 2016-06-04 08:24:32 2016-06-04 08:27:16 0.53
## 1421283 2016-06-04 08:19:02 2016-06-04 08:25:18 1.10
## 1422258 2016-06-04 08:09:13 2016-06-04 08:17:38 1.49
##          pickup_longitude pickup_latitude dropoff_longitude
## 1420830 -73.97655 40.74805 -73.98595
## 1420901 -73.97714 40.75862 -73.98866
## 1421109 -73.97950 40.76386 -73.98590
## 1421283 -73.97482 40.76241 -73.98942
## 1422258 -73.96571 40.75876 -73.98660
##          dropoff_latitude day weekday pickuptime dropofftime
## 1420830 40.75878 2016-06-04 Saturday 08:17 08:27
## 1420901 40.75683 2016-06-04 Saturday 08:17 08:22
## 1421109 40.75756 2016-06-04 Saturday 08:24 08:27
## 1421283 40.75680 2016-06-04 Saturday 08:19 08:25
## 1422258 40.75835 2016-06-04 Saturday 08:09 08:17
```

```
mean(td3$trip_distance)
```

```
## [1] 1.187447
```

Midtown to Times Square

For Weekdays (Monday to Friday)

```
td4<- subset(df2,subset= df2$pickup_longitude >=-73.9808 & df2$pickup_longitude <=-73.95
91
          & df2$pickup_latitude >=40.7480 & df2$pickup_latitude <=40.7643
          & df2$dropoff_longitude >=-73.9896 & df2$dropoff_longitude <=-73.9855
          & df2$dropoff_latitude >=40.7563 & df2$dropoff_latitude <=40.7603
)
head(td4)
```

```
##          tpep_pickup_datetime tpep_dropoff_datetime trip_distance
## 146460 2016-06-01 08:00:05 2016-06-01 08:12:49 1.15
## 150558 2016-06-01 08:07:47 2016-06-01 08:22:29 1.79
## 151449 2016-06-01 08:16:01 2016-06-01 08:23:32 0.79
## 153060 2016-06-01 08:09:55 2016-06-01 08:24:55 1.22
## 154409 2016-06-01 08:20:14 2016-06-01 08:29:42 1.10
## 155913 2016-06-01 08:06:51 2016-06-01 08:16:52 1.00
##          pickup_longitude pickup_latitude dropoff_longitude dropoff_latitude
## 146460 -73.97250 40.76207 -73.98604 40.75847
## 150558 -73.97317 40.75545 -73.98865 40.75864
## 151449 -73.97696 40.75164 -73.98656 40.75816
## 153060 -73.97016 40.76022 -73.98639 40.75689
## 154409 -73.97349 40.76374 -73.98679 40.75682
## 155913 -73.97421 40.75053 -73.98819 40.75960
##          day weekday pickuptime dropofftime
## 146460 2016-06-01 Wednesday 08:00 08:12
## 150558 2016-06-01 Wednesday 08:07 08:22
## 151449 2016-06-01 Wednesday 08:16 08:23
## 153060 2016-06-01 Wednesday 08:09 08:24
## 154409 2016-06-01 Wednesday 08:20 08:29
## 155913 2016-06-01 Wednesday 08:06 08:16
```

```
mean(td4$trip_distance)
```

```
## [1] 1.04629
```

```
str(df1)
```

```
## 'data.frame': 60042 obs. of 11 variables:
## $ tpep_pickup_datetime : chr "2016-06-04 08:07:17" "2016-06-04 08:07:18" "2016-06-04 08:07:18" "2016-06-04 08:07:18" ...
## $ tpep_dropoff_datetime: chr "2016-06-04 08:24:39" "2016-06-04 08:16:46" "2016-06-04 08:26:58" "2016-06-04 08:20:16" ...
## $ trip_distance : num 3.6 4 4.6 4.07 0.38 1.1 1.6 1.2 1.76 1.16 ...
## $ pickup_longitude : num -74 -74 -74 -74 -74 ...
## $ pickup_latitude : num 40.7 40.8 40.7 40.7 40.7 ...
## $ dropoff_longitude : num -74 -73.9 -74 -74 -74 ...
## $ dropoff_latitude : num 40.8 40.8 40.8 40.8 40.8 ...
## $ day : POSIXlt, format: "2016-06-04" "2016-06-04" ...
## $ weekday : chr "Saturday" "Saturday" "Saturday" "Saturday" ...
## $ pickuptime : chr "08:07" "08:07" "08:07" "08:07" ...
## $ dropofftime : chr "08:24" "08:16" "08:26" "08:20" ...
```

```
str(df2)
```

```
## 'data.frame':   396815 obs. of  11 variables:
## $ tpep_pickup_datetime : chr  "2016-06-01 08:00:00" "2016-06-01 08:00:00" "2016-06-01 08:00:00" "2016-06-01 08:00:00" ...
## $ tpep_dropoff_datetime: chr  "2016-06-01 08:00:37" "2016-06-01 08:08:08" "2016-06-01 08:11:48" "2016-06-01 08:14:35" ...
## $ trip_distance       : num  1.3 1.39 0.72 2.31 0.66 0.72 2 1.2 2.15 0.84 ...
## $ pickup_longitude    : num  -74 -74 -74 -74 -74 ...
## $ pickup_latitude     : num  40.7 40.8 40.8 40.7 40.8 ...
## $ dropoff_longitude   : num  -74 -74 -74 -74 -74 ...
## $ dropoff_latitude    : num  40.7 40.8 40.8 40.8 40.8 ...
## $ day                 : POSIXlt, format: "2016-06-01" "2016-06-01" ...
## $ weekday             : chr  "Wednesday" "Wednesday" "Wednesday" "Wednesday" ...
## $ pickuptime          : chr  "08:00" "08:00" "08:00" "08:00" ...
## $ dropofftime         : chr  "08:00" "08:08" "08:11" "08:14" ...
```

```
pickup1<-as.POSIXct(df1$pickuptime,format="%H:%M")
head(pickup1,5)
```

```
## [1] "2016-12-12 08:07:00 EST" "2016-12-12 08:07:00 EST"
## [3] "2016-12-12 08:07:00 EST" "2016-12-12 08:07:00 EST"
## [5] "2016-12-12 08:07:00 EST"
```

```
dropoff1<-as.POSIXct(df1$dropofftime,format="%H:%M")
head(dropoff1)
```

```
## [1] "2016-12-12 08:24:00 EST" "2016-12-12 08:16:00 EST"
## [3] "2016-12-12 08:26:00 EST" "2016-12-12 08:20:00 EST"
## [5] "2016-12-12 08:09:00 EST" "2016-12-12 08:11:00 EST"
```

```
str(pickup1)
```

```
## POSIXct[1:60042], format: "2016-12-12 08:07:00" "2016-12-12 08:07:00" ...
```

```
str(dropoff1)
```

```
## POSIXct[1:60042], format: "2016-12-12 08:24:00" "2016-12-12 08:16:00" ...
```

```
df1$diffinmin <- difftime(dropoff1,pickup1,tz,units = "mins")
```

```
pickup2<-as.POSIXct(df2$pickuptime,format="%H:%M")
head(pickup2,5)
```

```
## [1] "2016-12-12 08:00:00 EST" "2016-12-12 08:00:00 EST"
## [3] "2016-12-12 08:00:00 EST" "2016-12-12 08:00:00 EST"
## [5] "2016-12-12 08:00:00 EST"
```

```
dropoff2<-as.POSIXct(df2$dropofftime,format="%H:%M")
head(dropoff2,5)
```

```
## [1] "2016-12-12 08:00:00 EST" "2016-12-12 08:08:00 EST"
## [3] "2016-12-12 08:11:00 EST" "2016-12-12 08:14:00 EST"
## [5] "2016-12-12 08:05:00 EST"
```

```
str(pickup2)
```

```
## POSIXct[1:396815], format: "2016-12-12 08:00:00" "2016-12-12 08:00:00" ...
```

```
str(dropoff2)
```

```
## POSIXct[1:396815], format: "2016-12-12 08:00:00" "2016-12-12 08:08:00" ...
```

```
df2$diffinmin <- difftime(dropoff2,pickup2,tz,units = "mins")
```

```
mean(tdl$diffinmin)
```

```
## Warning in mean.default(tdl$diffinmin): argument is not numeric or logical:
## returning NA
```

```
## [1] NA
```

```
mean(td2$diffinmin)
```

```
## Warning in mean.default(td2$diffinmin): argument is not numeric or logical:
## returning NA
```

```
## [1] NA
```

```
midtowneastgarment <- mean(td2$diffinmin) - mean(tdl$diffinmin)
```

```
## Warning in mean.default(td2$diffinmin): argument is not numeric or logical:
## returning NA
```

```
## Warning in mean.default(tdl$diffinmin): argument is not numeric or logical:
## returning NA
```

```
midtowneastgarment
```

```
## [1] NA
```

```
mean(td3$diffinmin)
```

```
## Warning in mean.default(td3$diffinmin): argument is not numeric or logical:  
## returning NA
```

```
## [1] NA
```

```
mean(td4$diffinmin)
```

```
## Warning in mean.default(td4$diffinmin): argument is not numeric or logical:  
## returning NA
```

```
## [1] NA
```

```
midtowntimes<-mean(td4$diffinmin)-mean(td3$diffinmin)
```

```
## Warning in mean.default(td4$diffinmin): argument is not numeric or logical:  
## returning NA
```

```
## Warning in mean.default(td3$diffinmin): argument is not numeric or logical:  
## returning NA
```

```
midtowntimes
```

```
## [1] NA
```

```
midtowneastgarment1<-as.numeric(midtowneastgarment)  
midtowneastgarment1
```

```
## [1] NA
```

```
midtowntimes1 <- as.numeric(midtowntimes)  
str(midtowntimes1)
```

```
## num NA
```

```
midtowntimes1
```

```
## [1] NA
```

