HW3_KinjalRay

Kinjal

10/17/2018

```
#reading data
citibike <- read.csv("citibike.csv")
```

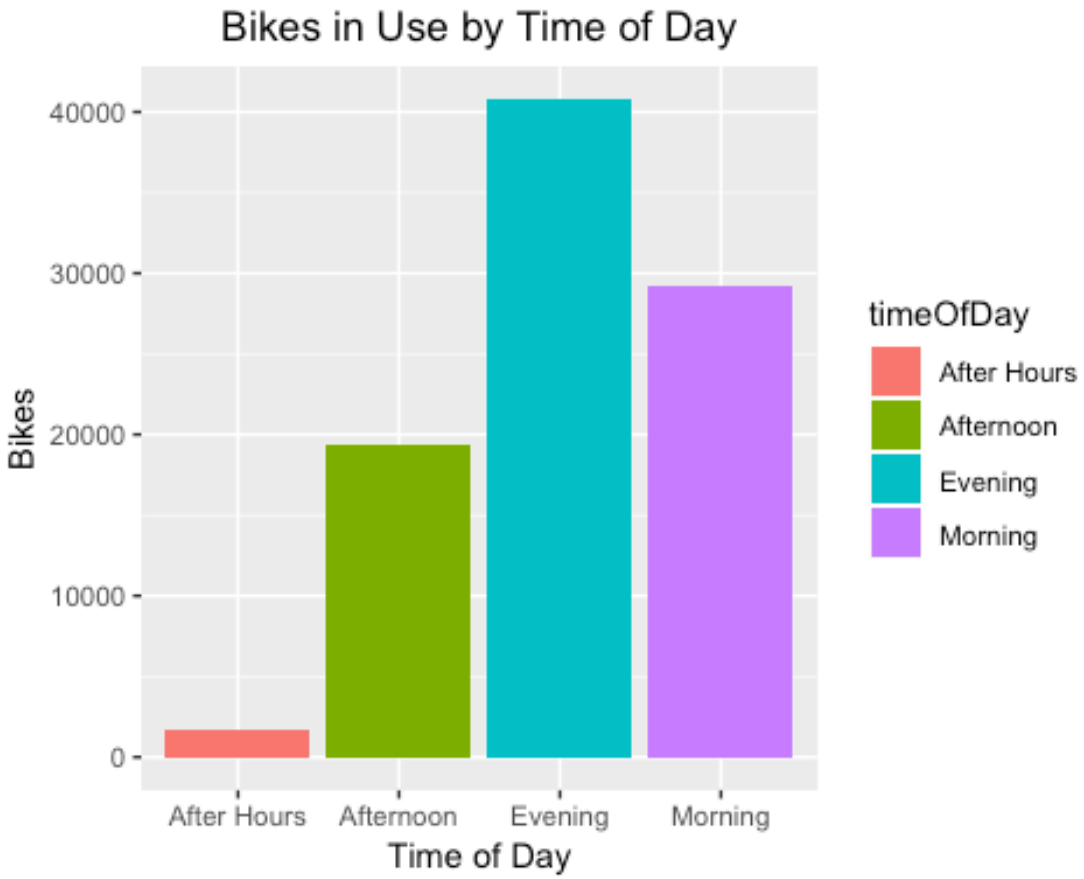
Males, on average, have shorter durations. Males, on average, travel longer distances.

```
tapply(citibike$tripduration, citibike$gender, mean, na.rm = TRUE)[-1]

##        1        2
## 624.6843 713.5721

library("geosphere")

## Warning: package 'geosphere' was built under R version 3.5.2

citibike$distance = distHaversine(citibike[ ,c("start.station.longitude",
"start.station.latitude")],citibike[ ,c("end.station.longitude",
"end.station.latitude")])

tapply(citibike$distance, citibike$gender, mean, na.rm = TRUE)[-1]

##        1        2
## 1427.471 1409.683

citibike$weekday <- factor(weekdays(as.Date(citibike$starttime, format =
"%m/%d/%Y")), levels =
c("Sunday","Monday","Tuesday","Wednesday","Thursday","Friday","Saturday"))

as.POSIXlt(citibike$starttime, format = "%m/%d/%Y %H:%M")
citibike$hourtime <- format(as.POSIXct(strptime(citibike$starttime,"%d/%m/%Y
%H:%M",tz="")) ,format = "%H")
citibike$hourtime <- as.integer(citibike$hourtime)
```

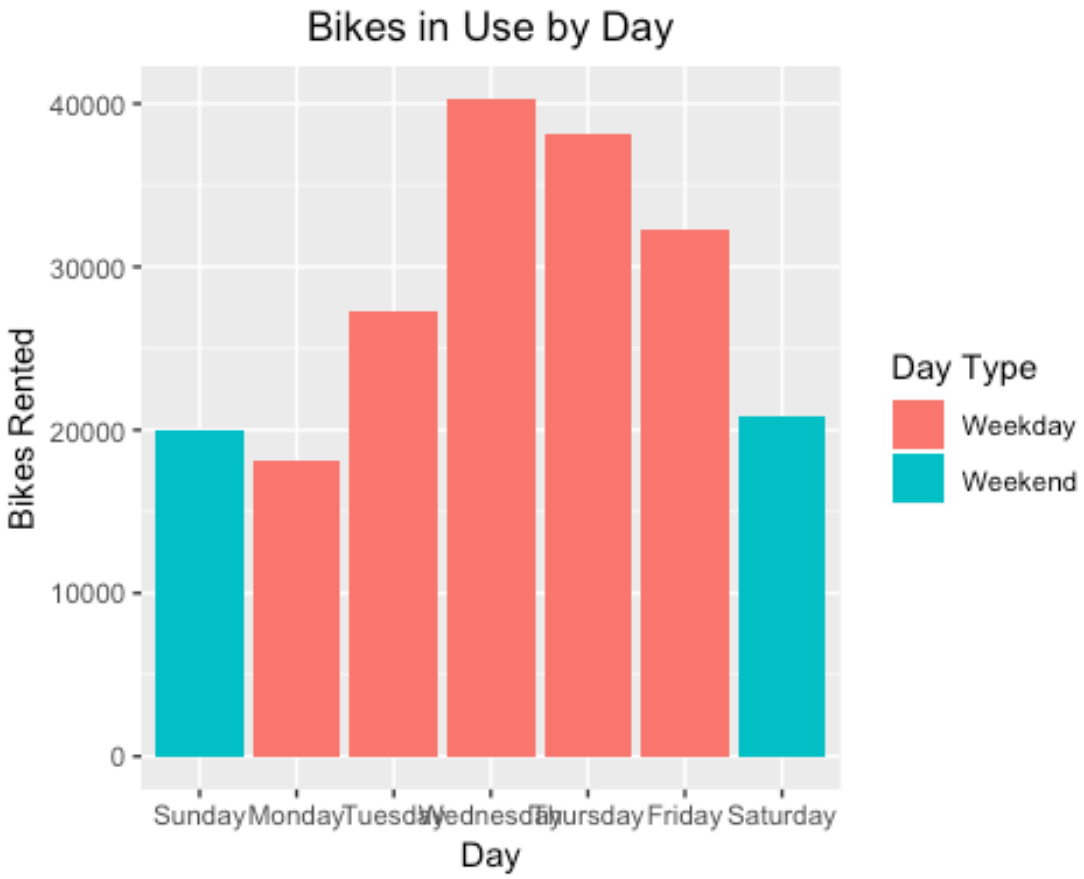
On weekdays, most traffic occurs, in the 7:30 AM - 10:00 AM and 3:00 PM - 6:00 PM ranges. On weekends, traffic is pretty consistent throughout the day, from 7:30 AM to 5:30 PM.

```
citibike$weekend <- ifelse(citibike$weekday == "Monday" | citibike$weekday ==
"Tuesday" |citibike$weekday == "Wednesday" |citibike$weekday == "Thursday"
|citibike$weekday == "Friday", "Weekday", "Weekend")

citibike1 <- subset(citibike, !is.na(hourtime))

citibike1$timeOfDay <- ifelse(citibike1$hourtime >= 04 & citibike1$hourtime <
12, "Morning",ifelse(citibike1$hourtime >= 12 & citibike1$hourtime < 16,
"Afternoon", ifelse(citibike1$hourtime >= 16 & citibike1$hourtime < 24,
```

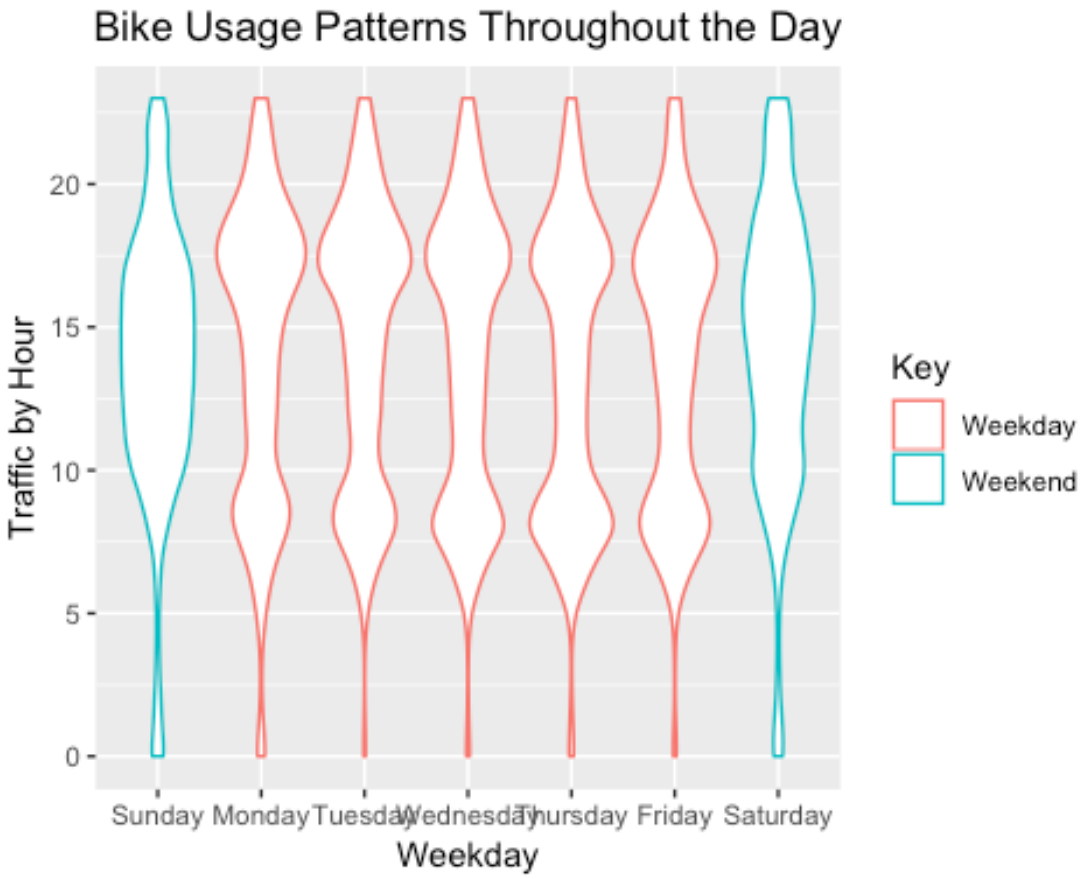
```
"Evening", "After Hours")))

library(ggplot2)

unique(citibike1$timeOfDay)

## [1] "After Hours" "Morning"     "Afternoon"   "Evening"

ggplot(citibike1, aes(timeOfDay, fill = timeOfDay)) + geom_bar() + labs(x =
"Time of Day", y = "Bikes")  + ggtitle("Bikes in Use by Time of Day") +
theme(plot.title = element_text(hjust = 0.5))
```




```
#barplot(table(citibike1$timeOfDay), main="Bikes in Use by Time of Day",
xlab="Time of Day", col= "green")

ggplot(citibike, aes(citibike$weekday, fill = citibike$weekend)) + geom_bar()
+ labs(x = "Day", y= "Bikes Rented", fill = "Day Type") + ggtitle("Bikes in
Use by Day") + theme(plot.title = element_text(hjust = 0.5))
```




```
ggplot(citibike1, aes(citibike1$weekday, citibike1$hourtime, colour =
citibike1$weekend)) + geom_violin() + ggtitle("Bike Usage Patterns Throughout
the Day") + theme(plot.title = element_text(hjust = 0.5)) + labs(x =
"Weekday", y = "Traffic by Hour", colour = "Key")
```



5 most popular stations by Time of Day are: In the morning: 8 Ave & W 31 St and W 41 St & 8 Ave are most popular In the afternoon: 1 Ave & E 15 St and 1 Ave & E 44 St are most popular In the evening: E 43 St & Vanderbilt Ave and Lafayette St & E 8 St are most popular In the after hours: Allen St & E Houston St and Lafayette St & E 8 St are most popular

```
library(plyr)

busystation <- tapply(citibike1$start.station.name, citibike1$timeOfDay,
count)

Morning <- as.data.frame(busystation$Morning)
Morning <- Morning[order(Morning$freq, decreasing = TRUE), ]
head(Morning, 5)

##                          x freq
## 15         8 Ave & W 31 St  767
## 288        W 41 St & 8 Ave  666
## 280        W 31 St & 7 Ave  446
## 16         8 Ave & W 33 St  361
## 232 Pershing Square South   338
```

```
Afternoon <- as.data.frame(busystation$Afternoon)
Afternoon <- Afternoon[order(Morning$freq, decreasing = TRUE), ]
head(Afternoon, 5)
```

```
##                   x freq
## 1  1 Ave & E 15 St  135
## 2  1 Ave & E 44 St   58
## 3 10 Ave & W 28 St   77
## 4 11 Ave & W 27 St   91
## 5 11 Ave & W 41 St   97
```

```
Evening <- as.data.frame(busystation$Evening)
Evening <- Evening[order(Evening$freq, decreasing = TRUE), ]
head(Evening, 5)
```

```
##                             x freq
## 130 E 43 St & Vanderbilt Ave  557
## 195    Lafayette St & E 8 St  542
## 105       E 17 St & Broadway  484
## 272          W 21 St & 6 Ave  441
## 203  Lexington Ave & E 24 St  378
```

```
afterHours <- as.data.frame(busystation$`After Hours`)
afterHours <- afterHours[order(afterHours$freq, decreasing = TRUE), ]
head(afterHours, 5)
```

```
##                            x freq
## 21  Allen St & E Houston St   29
## 157   Lafayette St & E 8 St   26
## 203     St Marks Pl & 2 Ave   25
## 37       Broadway & E 14 St   24
## 1           1 Ave & E 15 St   23
```

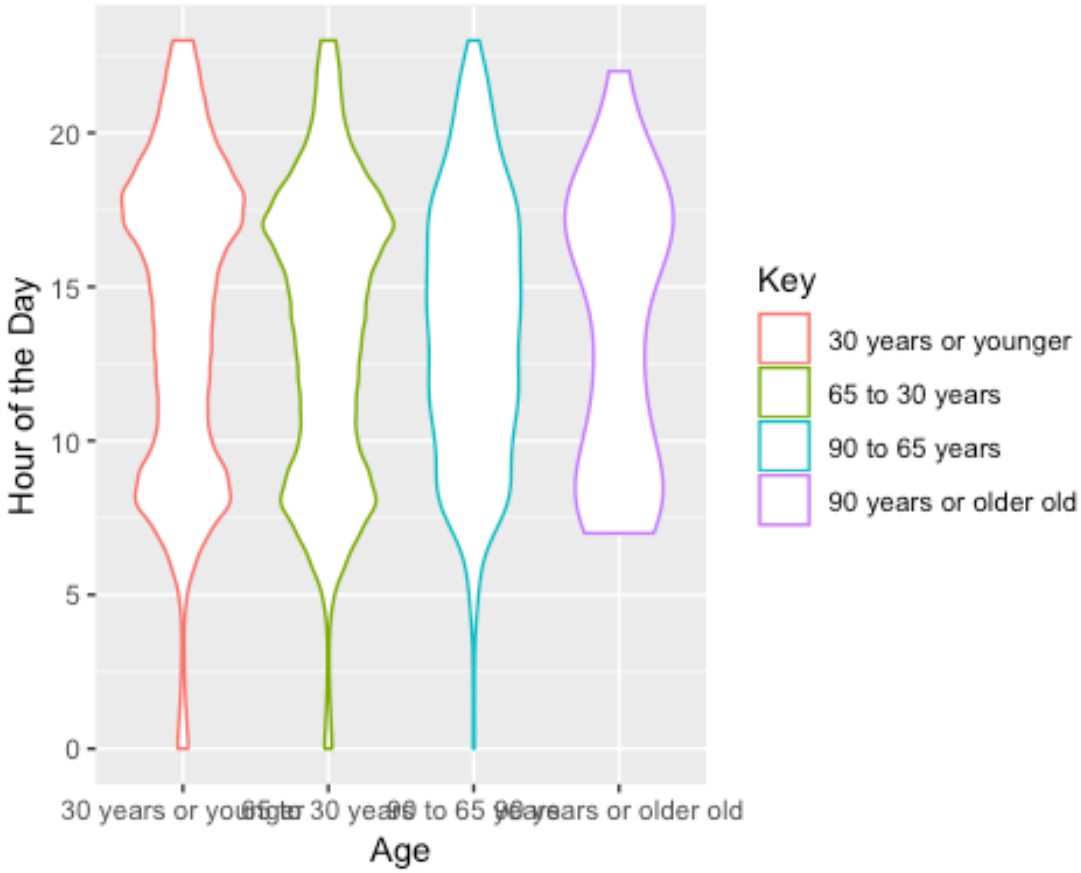
I wanted to explore how the age of the user affects the time of day they use the bikes.

I noticed younger demographics use bikes during morning and afternoon rush hour. I'm assuming these populations use the bikes to commute to work. The older group of 65-90 years old use the bikes consistently throughout the day. Though I find the data a bit questionable for those 90 years and older, it is interesting how none in this demographic use the bikes before 7 AM.

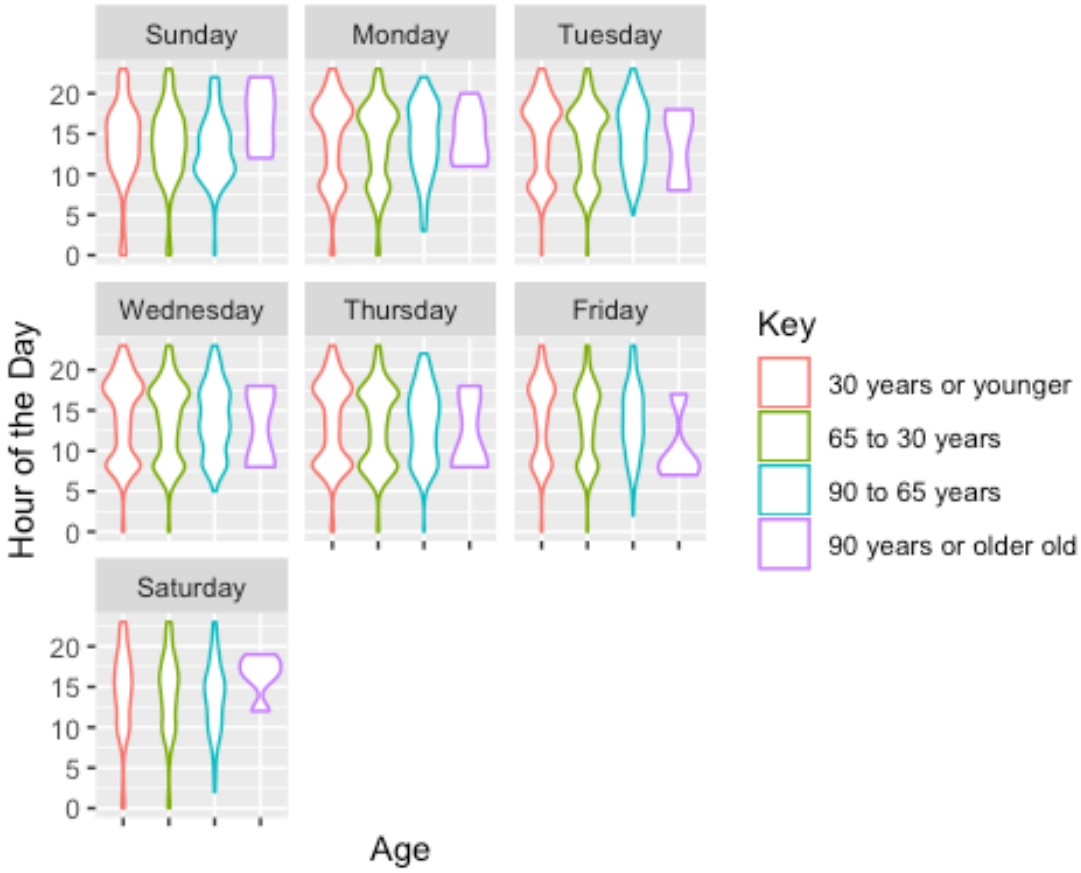
When splitting the data further by weekday, the oldest generation rental patterns are extremely varied, whereas the workforce generations remained relatively consistent.

```
citibike$demographic <- ifelse(citibike$birth.year >= 1900 &
citibike$birth.year < 1925, "90 years or older
old",ifelse(citibike$birth.year >= 1925 & citibike$birth.year < 1950, "90 to
65 years", ifelse(citibike$birth.year >= 1950 & citibike$birth.year < 1975,
"65 to 30 years", "30 years or younger")))

citibike2 <- subset(citibike, !is.na(demographic))
```

```
ggplot(citibike2, aes(citibike2$demographic, citibike2$hourtime, colour =
citibike2$demographic)) + geom_violin() + labs(x = "Age", y = "Hour of the
Day", colour = "Key")

## Warning: Removed 104701 rows containing non-finite values (stat_ydensity).
```




```
citibike2$weekday <- factor(weekdays(as.Date(citibike2$starttime, format =
"%m/%d/%Y")), levels =
c("Sunday","Monday","Tuesday","Wednesday","Thursday","Friday","Saturday"))

ggplot(citibike2, aes(citibike2$demographic, citibike2$hourtime, colour =
citibike2$demographic)) + geom_violin() + labs(x = "Age", y = "Hour of the
Day", colour = "Key") + facet_wrap(~ citibike2$weekday) +
theme(axis.text.x=element_blank())

## Warning: Removed 104701 rows containing non-finite values (stat_ydensity).
```


Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
0
5
10
15
20
Hour of the Day
Age
Key
30 years or younger
65 to 30 years
90 to 65 years
90 years or older old