title: "R Markdown, Knitr Assignment" author: Pamela date: Dec 2017

output: html_document: keep_md: yes

R Markdown, Knitr Assignment

1.Loading and preprocessing the data

```
MyData <- read.csv(file="activity.csv", header=TRUE, sep=",")
MyData_New <- subset(MyData, is.na(MyData$steps) == FALSE)
```

2.Mean total number of steps taken per day

```
require(stats)
MyData_Daysum <-tapply(MyData_New$steps, MyData_New$date, sum)
hist(MyData_Daysum, col="lightgreen", xlab ="Sum of Steps", main = "Histogram of Steps per Day")
```


Histogram of Steps per Day
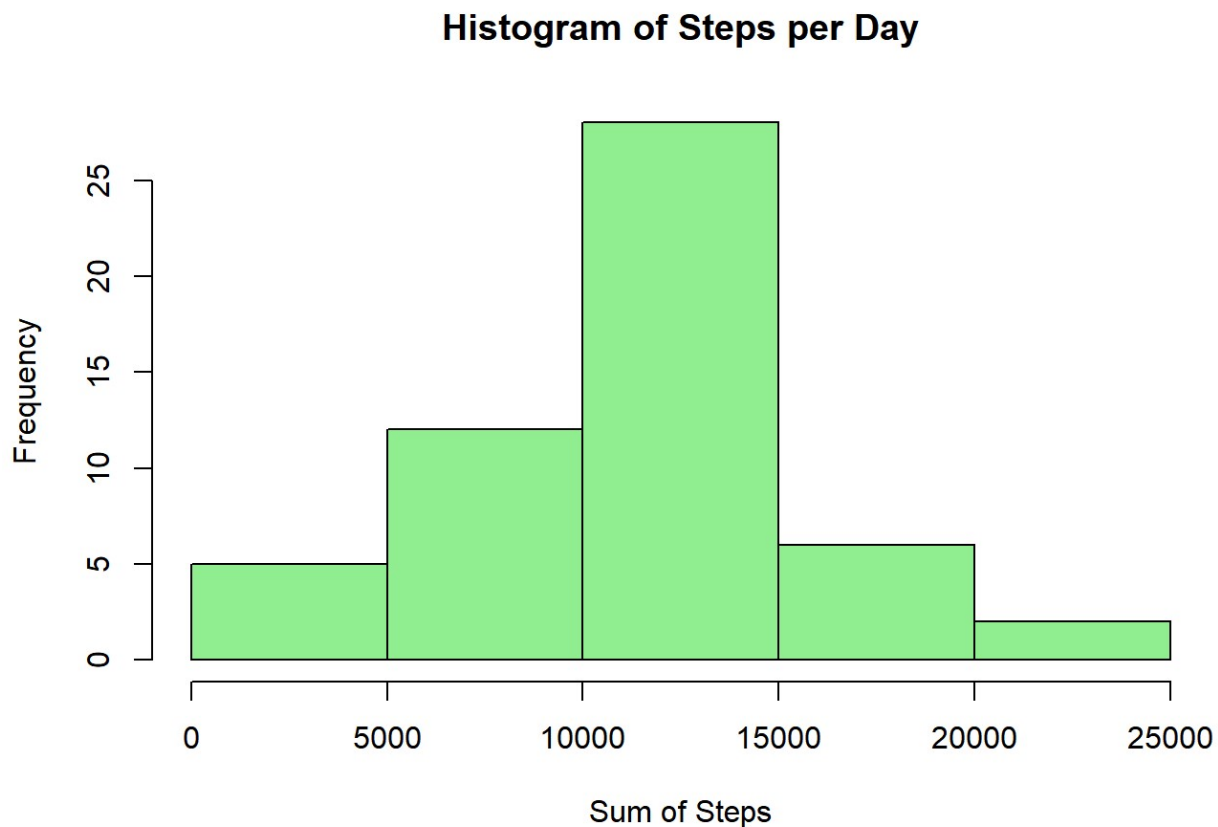
```
MyData_Daymean <-tapply(MyData_New$steps, MyData_New$date, mean, na.rm = TRUE)
MyData_Daymed <-tapply(MyData_New$steps, MyData_New$date, median , na.rm = TRUE)
```

The mean steps per day

: NA, 0.4375, 39.4166667, 42.0694444, 46.1597222, 53.5416667, 38.2465278, NA, 44.4826389, 34.375, 35.7777778, 60.3541667, 43.1458333, 52.4236111, 35.2048611, 52.375, 46.7083333, 34.9166667, 41.0729167, 36.09375, 30.6284722, 46.7361111, 30.9652778, 29.0104167, 8.6527778, 23.5347222, 35.1354167, 39.7847222, 17.4236111, 34.09375, 53.5208333, NA, 36.8055556, 36.7048611, NA, 36.2465278, 28.9375, 44.7326389, 11.1770833, NA, NA, 43.7777778, 37.3784722, 25.4722222, NA, 0.1423611, 18.8923611, 49.7881944, 52.4652778, 30.6979167, 15.5277778, 44.3993056, 70.9270833, 73.5902778, 50.2708333, 41.0902778, 38.7569444, 47.3819444, 35.3576389, 24.46875, NA

The median steps per day

: NA, 0, 0, 0, 0, 0, 0, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA, 0, 0, NA, 0, 0, 0, 0, NA, NA, 0, 0, 0, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA

3. Average daily activity pattern

```
MyData_Intrvmean <-tapply(MyData_New$steps, MyData_New$interval, mean)
plot(MyData_Intrvmean, type = "l",xlab ="Interval",ylab ="Mean Steps" , main = "Time s
eries of Average steps (across all days) vs 5-minute intervals", xaxt = "n", col="blu
e" )
axis(1, at=1:length(names(MyData_Intrvmean)), labels =names(MyData_Intrvmean) )
```

Time series of Average steps (across all days) vs 5-minute intervals
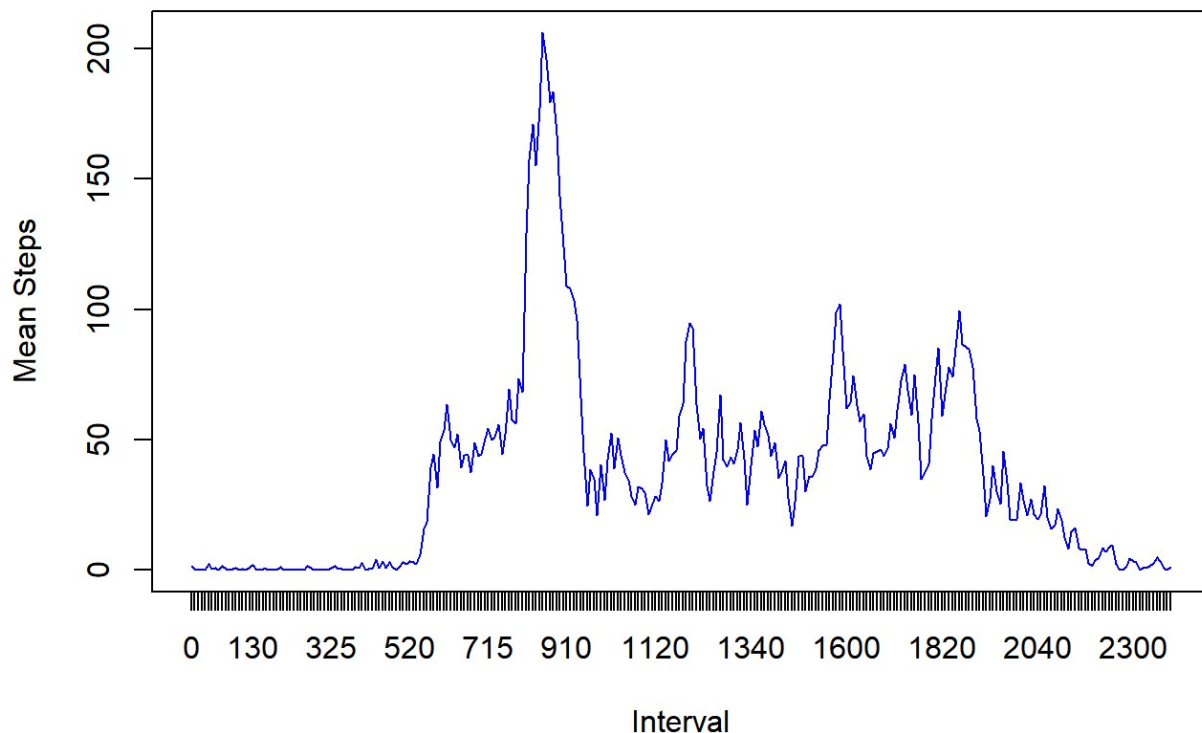
```
#Maximum average interval mean
MyData_Intrvmean_DF<- data.frame(interval=names(MyData_Intrvmean), value=MyData_Intrvm
ean)
x <-MyData_Intrvmean_DF[which.max(MyData_Intrvmean_DF$value),]
```

The maximum of average steps per day

206.1698113 falls at 835th interval

4. Imputing missing values

Count of missing values 2304 using nrow(MyData[!complete.cases(MyData),])

```
MyData2 <- MyData
cleanBase2 <- MyData2[!is.na(MyData2$steps),]
missingData <- is.na(MyData2$steps)
meanVals <- tapply(cleanBase2$steps, cleanBase2$interval, mean, na.rm=TRUE, simplify=T
RUE)
MyData2$steps[missingData] <- meanVals[as.character(MyData2$interval[missingData])]
#count of missing values in new data
nrow(MyData2[!complete.cases(MyData2),])
```

```
## [1] 0
```

```
#imputed histogram
MyData2_Daysum <-tapply(MyData2$steps, MyData2$date, sum)
hist(MyData2_Daysum, col="skyblue", xlab ="Sum of Steps", main = "Histogram of Steps p
er Day-Missing values imputed")
```

Histogram of Steps per Day-Missing values imputed
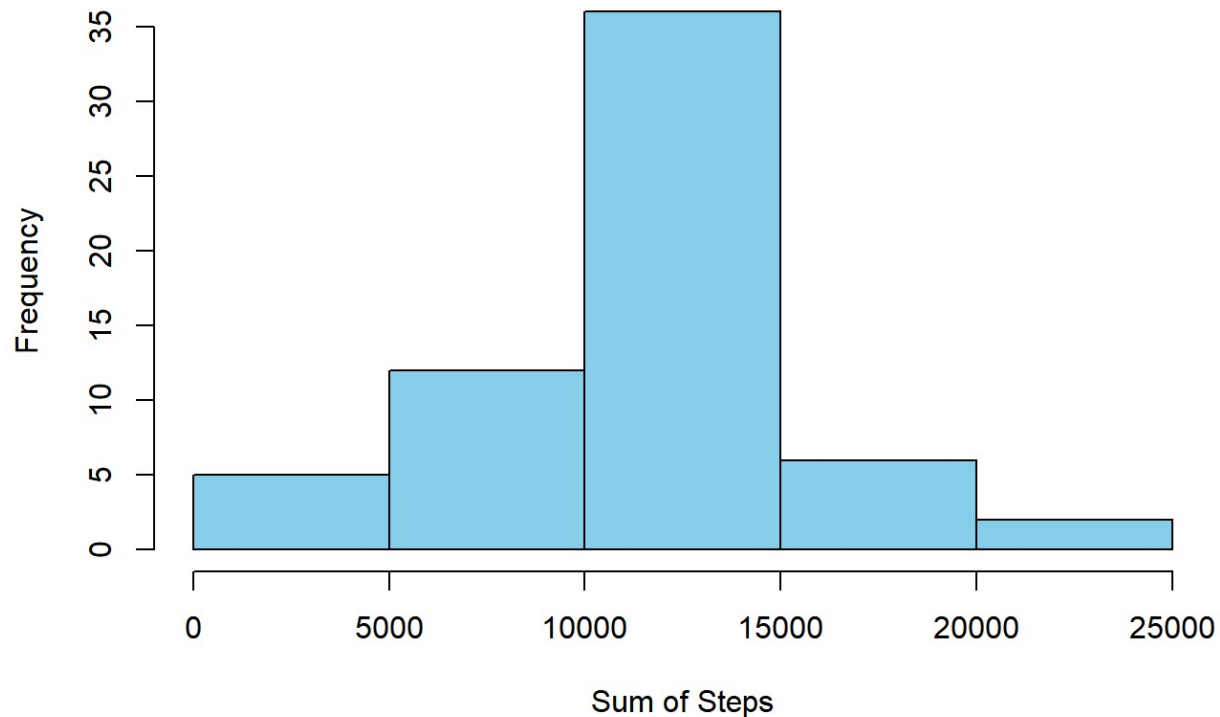

```
MyData2_Daymean <-tapply(MyData2$steps, MyData2$date, mean)
MyData2_Daymed <-tapply(MyData2$steps, MyData2$date, median)

#Calculate and report the mean and median total number of steps taken per day.
MyData_Daysum_DF<- data.frame(day=names(MyData_Daysum), value=MyData_Daysum)
MyData2_Daysum_DF<- data.frame(day=names(MyData2_Daysum), value=MyData2_Daysum)

D <-median(MyData_Daysum_DF$value, na.rm =  TRUE)
D2 <-median(MyData2_Daysum_DF$value, na.rm =  TRUE)
S <- round((sum(MyData_Daysum_DF$value, na.rm =  TRUE))/61,2)
S2 <-round((sum(MyData2_Daysum_DF$value, na.rm =  TRUE))/61,2)
```

**Mean before imputing is 9354.23

**After imputing is 1.07661910^{4}.

**Median before was 10765

**After is 1.076618910^{4}.

Since median is equal to mean after inputing we can say this data has stronger central tendency than the one before imputing

5. Are there differences in activity patterns between weekdays and weekends?

```
MyData2_Wk <- MyData2
MyData2_Wk$WkDay_WkEnd <-ifelse( weekdays(as.Date(MyData2_Wk$date)) == "Saturday" | we
ekdays(as.Date(MyData2_Wk$date)) == "Sunday", "WkEnd", "WkDay")
MyData2_WkDay <- MyData2_Wk[MyData2_Wk$WkDay_WkEnd == "WkDay", ]
MyData2_WkEnd <- MyData2_Wk[MyData2_Wk$WkDay_WkEnd == "WkEnd", ]
MyData2_WkDay <- tapply(MyData2_WkDay$steps, MyData2_WkDay$interval, mean, na.rm = TRU
E)
MyData2_WkEnd <- tapply(MyData2_WkEnd$steps, MyData2_WkEnd$interval, mean, na.rm = TRU
E)
##Plots of Weekday, Weekend
par(mfrow = c(1, 2), mar=c(5,4,4,0.5), cex= 1, cex.axis = 0.8 )
plot(MyData2_WkDay, type = "l",main ="Week Day",xlab="5 min Interval",ylab ="Daily Mea
n Steps" ,xaxt = "n", col="brown" )
axis(1, at=1:length(names(MyData2_WkDay)), labels =names(MyData2_WkDay) )
plot(MyData2_WkEnd, type = "l",main ="Week End",xlab="5 min Interval",ylab ="Daily Mea
n Steps", xaxt = "n", col="darkgreen" )
axis(1, at=1:length(names(MyData2_WkEnd)) )
```