

Reproducible Research - Peer Assessment 1

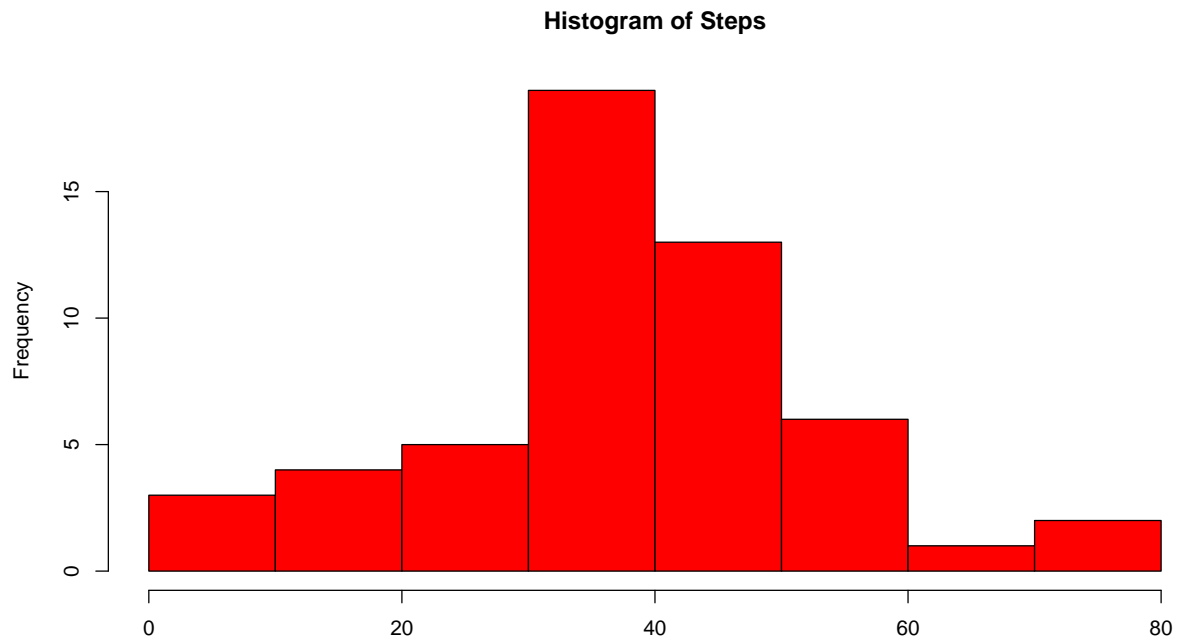
1. Loading and Processing the Data

```
data<-read.csv("activity.csv")  
#Calculating total no. of steps day wise.  
library(data.table)  
data1<-data.table(data)  
data_total<-data1[,mean(steps,na.rm=T),by=date]
```

2. What is Mean Total Number of Steps taken per day?

2.1 Plotting the Histogram

```
hist(data_total$V1,col="red",xlab="",main="Histogram of Steps")
```



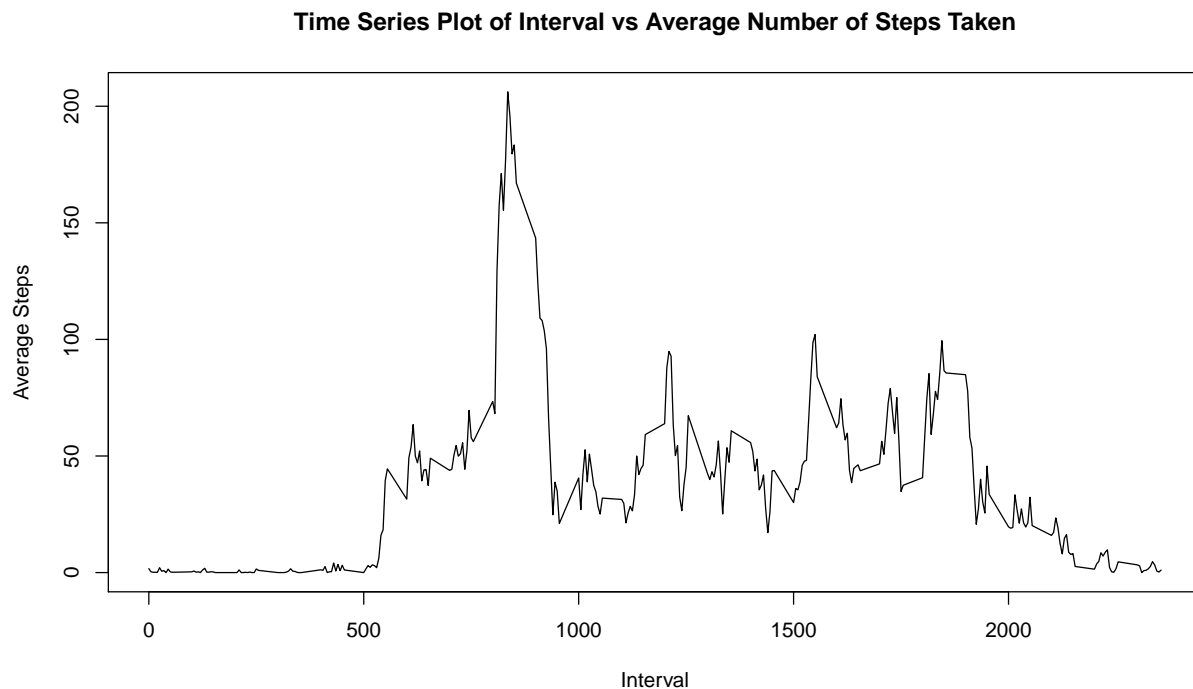
2.2 Calculating Mean and Median of the Number of Steps

```
mean<-mean(data_total$V1,na.rm=T)
median<-median(data_total$V1,na.rm=T)
```

The mean and median of total number of steps taken per day are 37.3826 and 37.3785 respectively.

3. What is the Average Daily Activity Pattern?

```
library(data.table)
data_table<-data.table(data)
avg_data<-data_table[,mean(steps,na.rm=T),by=interval]
plot(avg_data$interval,avg_data$V1,xlab="Interval",type="l", ylab="Average Steps",main="Time Series Plot")
```



```
maxima<-avg_data$interval[avg_data$V1==max(avg_data$V1)]
```

The interval which has the maximum no. of average steps is 835.

4. Imputing Missing Values

4.1 Calculating and Reporting the Number of Missing Values in the Dataset

```
nmis<-sum(is.na(data[,1]))
```

The number of missing values in the dataset is 2304.

4.2 Strategy for Imputing the Missing Values

The strategy used for imputing missing values is to replace the missing value by the mean of all other values in the same interval.

```
merged_data<-merge(data,avg_data,by="interval")
for(i in 1:nrow(data))
  {if(is.na(merged_data$steps[i])){merged_data$steps[i]=merged_data$V1[i]}}
```

4.3 Creating New Data after Imputing Missing Values

```
data_pi<-merged_data[,!(names(merged_data) %in% "V1")]
#Calculating total no. of steps day wise.
library(data.table)
data_pi1<-data.table(data_pi)
data_pi_total<-data_pi1[,mean(steps,na.rm=T),by=date]
```

4.4 Comparison of results of Before and After Imputing the Missing Values

4.4.1 Plotting the Histogram (After Imputing the Missing Values)

```
hist(data_pi_total$V1,col="red",xlab="",main="Histogram of Steps (After Imputing Missing Values)")
```



4.4.2 Calculating Mean and Median of the Number of Steps (After Imputing the Missing Values)

```
mean1<-mean(data_pi_total$V1,na.rm=T)
median1<-median(data_pi_total$V1,na.rm=T)
```

The mean and median of total number of steps taken per day in the new data (after imputing missing values) are 37.3826 and 37.3826 respectively.

Note: The technique which we used to impute the missing values replaces missing values by the average no. of steps in that interval across all the dates which remains close to zero, i.e. negligible and hence it doesn't alter the total no. of steps taken per day and gives approximately same figures of mean and median. Moreover the increased sum figure doesn't change the mean as the no. of observations without missing values is also increased.

5. Are there differences in Activity Patterns between Weekdays and Weekends?

5.1. Creating a New Factor Variable

```
data_pi<-merged_data[,!(names(merged_data) %in% "V1")]
data_pi$date <- strptime(as.character(data_pi$date), "%Y-%m-%d")
data_pi["flag"]<-NA
data_pi["flag"]<-weekdays(data_pi$date)

#Creating a factor variable to distinguish between weekday and weekend.
data_pi$flag[data_pi$flag == "Monday"]="Weekday"
data_pi$flag[data_pi$flag == "Tuesday"]="Weekday"
data_pi$flag[data_pi$flag == "Wednesday"]="Weekday"
data_pi$flag[data_pi$flag == "Thursday"]="Weekday"
data_pi$flag[data_pi$flag == "Friday"]="Weekday"
data_pi$flag[data_pi$flag == "Saturday"]="Weekend"
data_pi$flag[data_pi$flag == "Sunday"]="Weekend"
data_pi$flag<-as.factor(data_pi$flag)
```

5.2. Creating a two Panel Plot

```
#Plotting four graphs simultaneously.
par(mfrow=c(1,2))

weekday<-subset(data_pi,data_pi$flag=="Weekday")
weekend<-subset(data_pi,data_pi$flag=="Weekend")

library(data.table)
weekday_table<-data.table(weekday)
avg_weekday<-weekday_table[,mean(steps,na.rm=T),by=interval]
avg_weekday<-avg_weekday[order(avg_weekday$interval)]
```

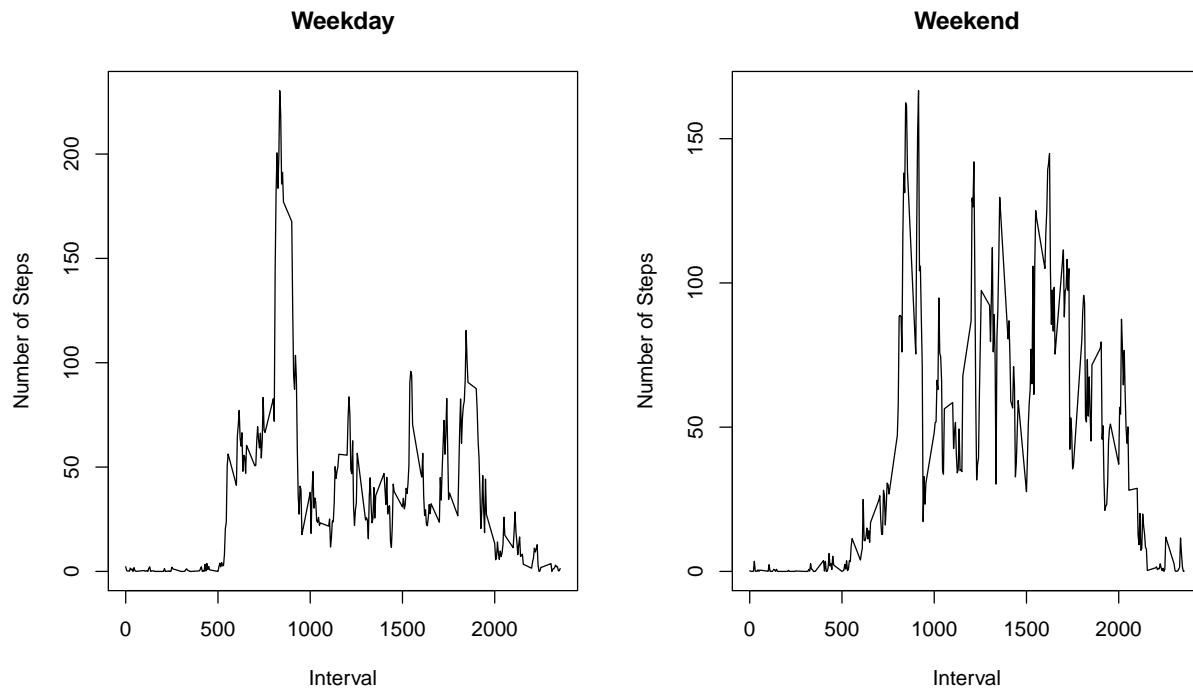
```

plot(as.numeric(avg_weekday$interval),avg_weekday$V1,xlab="Interval",type="l",
     ylab="Number of Steps",main="Weekday")

library(data.table)
weekend_table<-data.table(weekend)
avg_weekend<-weekend_table[,mean(steps,na.rm=T),by=interval]
avg_weekend<-avg_weekend[order(avg_weekend$interval)]

plot(as.numeric(avg_weekend$interval),avg_weekend$V1,xlab="Interval",type="l",
     ylab="Number of Steps",main="Weekend")

```



Conclusion: The activity patterns are observed to be different between Weekdays and Weekends. Weekend activity patterns are relatively more volatile (random) as compared to Weekday activity patterns.