

Reproducible Research

Ashitosh Nale

27/06/2020

Assignment week 2 Coursera

This is an R Markdown document for Assignment week 2 Coursera.

We are provided with the data from personal activity monitoring device. This device collects data at 5 interval through out the day. The data consists of two months of data from an anonymous individual collected during the month of October and November,2012 and include the number of steps taken in 5 intervals each day.

Loading the Data

```
data <- read.csv("activity.csv")
```

What is mean total number of steps taken per day ?

Firstly we will calculate the total number of steps taken per day

```
m <- tapply(data$steps,data$date, sum ,na.rm = FALSE)
```

Lets remove the missing values

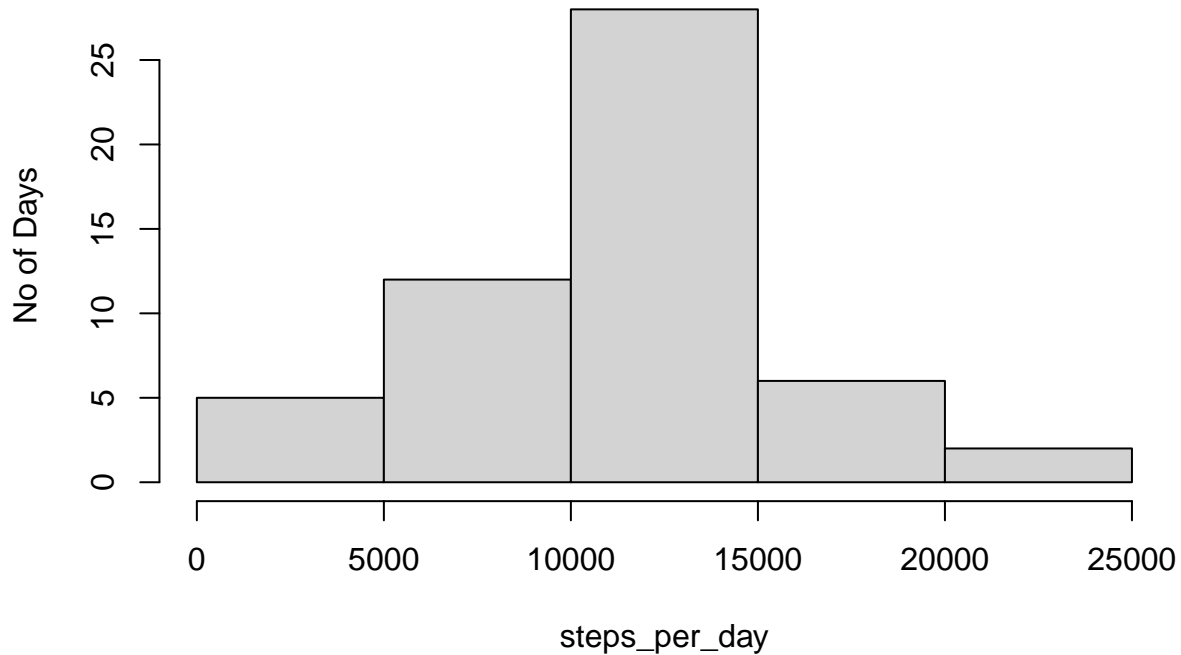
```
steps_per_day <- m[complete.cases(m)]  
head(steps_per_day)
```

```
## 2012-10-02 2012-10-03 2012-10-04 2012-10-05 2012-10-06 2012-10-07  
##          126          11352          12116          13294          15420          11015
```

Plot the histogram for Total number of steps taken each day

```
hist(steps_per_day, main = "Total Number of steps taken each day", ylab = "No of Days")
```

Total Number of steps taken each day



Calculate Mean and Median of total number of steps taken each day

```
Mean <- mean(steps_per_day)
Mean
```

```
## [1] 10766.19
```

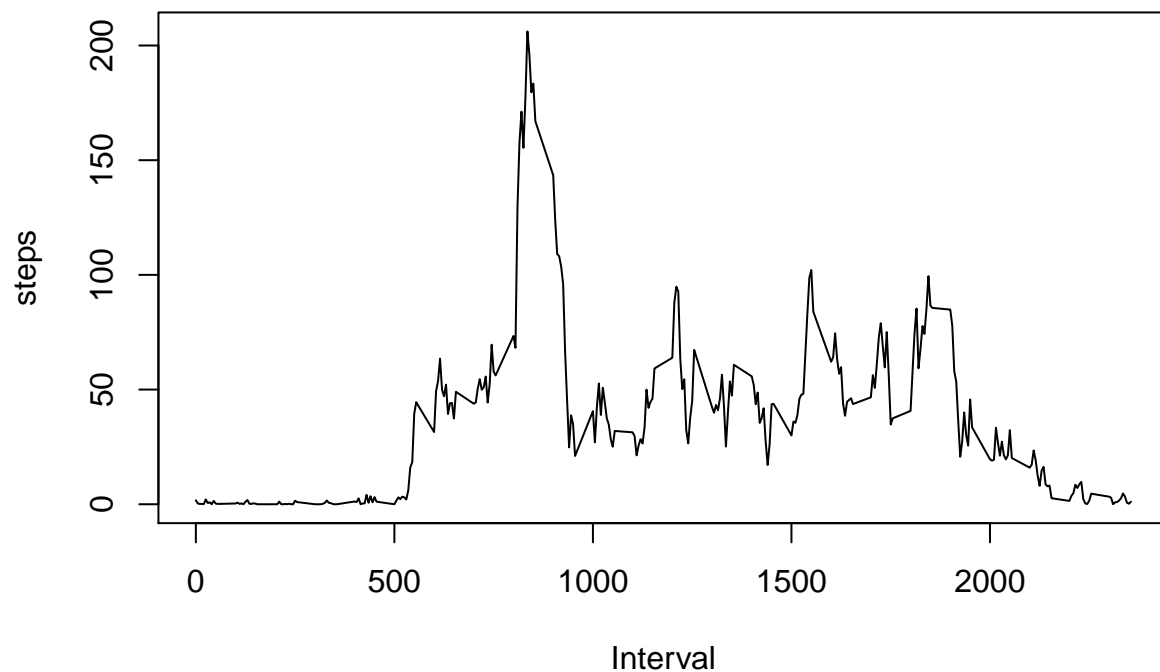
```
Median <- median(steps_per_day)
Median
```

```
## [1] 10765
```

So the calculated mean and median is 1.0766189×10^4 and 10765 respectively.

What is the average daily activity pattern ?

```
steps <- tapply(data$steps,data$interval,mean, na.rm = TRUE)
plot(names(steps),steps,type = "l",xlab = "Interval")
```



Which 5 min-Interval, on average across all the day in the dataset, contains the maximum number of steps?

```
names(which.max(steps))
```

```
## [1] "835"
```

So the 835th 5 min interval contains maximum number of steps.

```
n <- data.frame(tapply(data$steps,data$interval, mean ,na.rm = TRUE))
colnames(n)= "Mean"
n[is.na(n)] <- 0
head(n)
```

```
##           Mean
## 0  1.7169811
## 5  0.3396226
## 10 0.1320755
## 15 0.1509434
## 20 0.0754717
## 25 2.0943396
```

The NAs in steps column are replaced by taking average of steps of corresponding Intervals

```
data[is.na(data)] = n$Mean
```

Now we have the data with zero NA values.

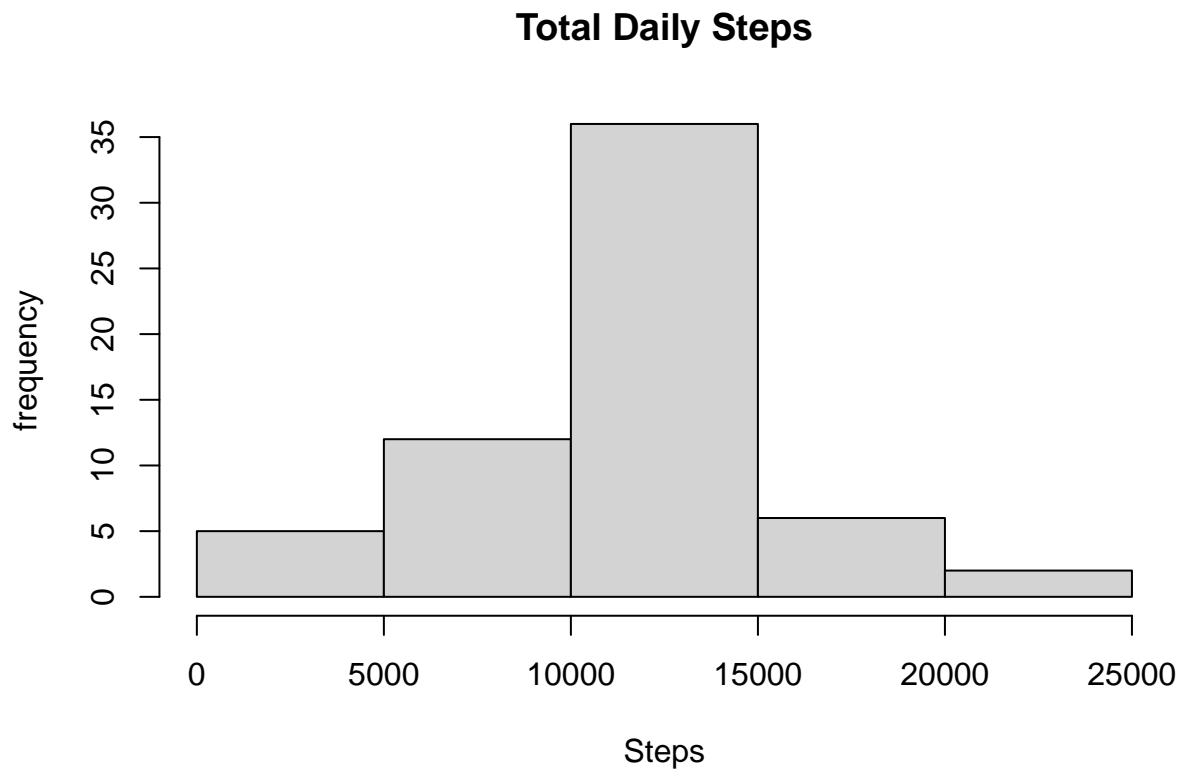
```
sum(is.na(data))
```

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

```
steps <- tapply(data$steps,data$date,sum)
```

Histogram of the total number of steps taken each day & calculate and report mean & median total number of steps taken per day.

```
hist(steps,xlab = "Steps",ylab = "frequency", main = "Total Daily Steps")
```



Calculate Mean and Median of total number of steps taken each day

```
Mean1 <- mean(steps)  
Mean1
```

```
## [1] 10766.19
```

```
Median1 <- median(steps)
Median1
```

```
## [1] 10766.19
```

Compare Old Mean Values and new one

```
Mean
```

```
## [1] 10766.19
```

```
Mean1
```

```
## [1] 10766.19
```

```
Median
```

```
## [1] 10765
```

```
Median1
```

```
## [1] 10766.19
```

Are there differences in activity patterns between weekdays and weekends?

Adding the column “Days” is data as Weekend or Weekday

```
data$date <- as.Date(data$date,"%Y-%m-%d")
data$day <- weekdays(data$date)
data$Days <- ifelse(data$day == "Saturday" | data$day == "Sunday", "Weekend", "Weekday" )
```

Averaging no. of steps along with Interval and Days

```
data1 <- aggregate(data$steps,list(data$Days,data$interval),mean)
names(data1) <- c("Days","Interval","Steps")
```

Comparison of Number of Steps in each Interval

```
library(ggplot2)
g <- ggplot(data1,aes(Interval,Steps,color = Days))
g + geom_line() + facet_grid(Days~.) + ggtitle("Comparison of Number of Steps in each Interval")
```

Comparison of Number of Steps in each Interval

