

Reproducible Research Course Project 1

Liliana Grajales

April 3, 2016

Course Project 1

This document includes the code and graphs for Project 1 of the Reproducible Research Coursera Course. The data for this project is located at <https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip> (<https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip>).

The variables included in this dataset are:

- steps: Number of steps taking in a 5-minute interval (missing values are coded as)
- date: The date on which the measurement was taken in YYYY-MM-DD format
- interval: Identifier for the 5-minute interval in which measurement was taken

The dataset is stored in a comma-separated-value (CSV) and there are a total of 17,568 observations in this dataset.

```
#Libraries
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.2.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(base64)
```

```
## Warning: package 'base64' was built under R version 3.2.2
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
library(chron)
```

```
## Warning: package 'chron' was built under R version 3.2.4
```

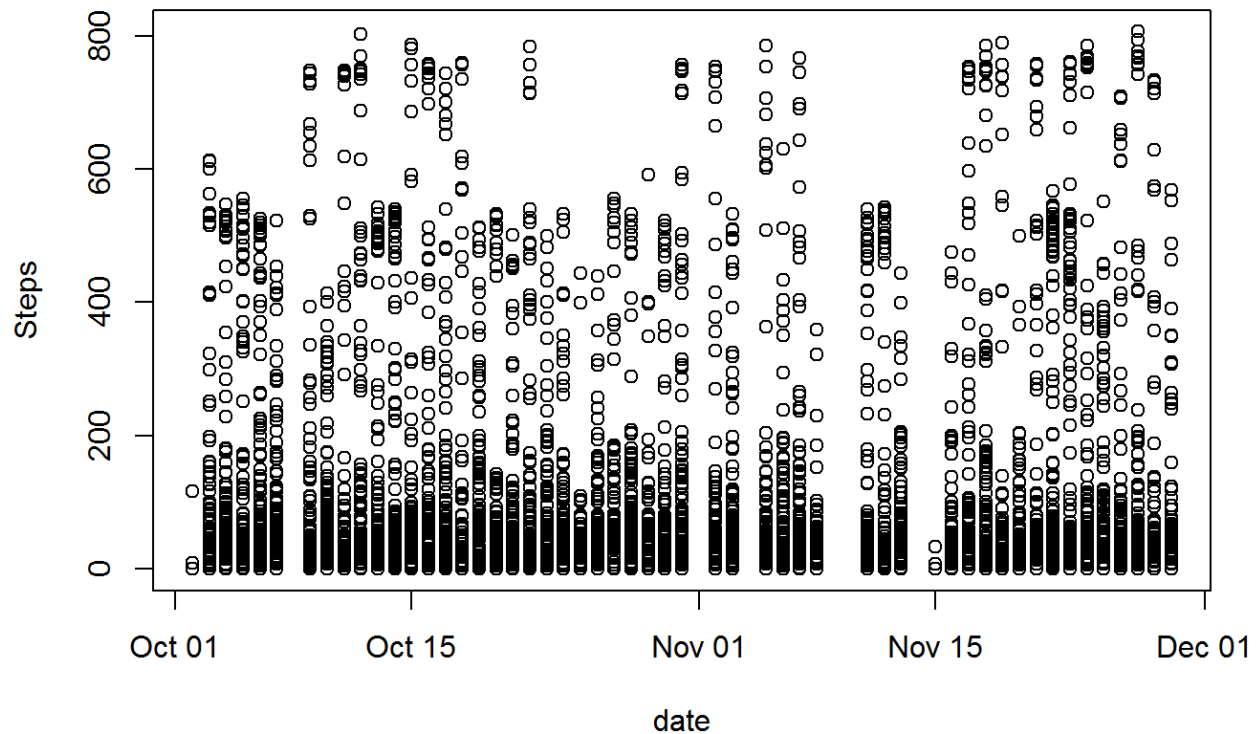
```
#Setting working directory path
setwd("D:/Liliana_2016/ReproducibleResearch/Script_Project/repdata-data-activit
y")

#*****
# Reading the table and separating the string of data into columns using sep
=";"
#*****
activity_all <- read.csv("activity.csv", sep=";", colClasses ="character")
#Ommiting NAs
activity <-na.omit(activity_all)
#*****
# Plotting the data
#*****
#First column date format description
day <- as.Date(activity[,2], "%Y-%m-%d")
steps <- as.numeric(activity[,1])
interval <-as.numeric(activity[,3])
steps_day <- data.frame(day,steps)
```

What is mean total number of steps taken per day?

In Plot #1 below each circle represents the number of steps taken in intervals of 5 minutes each day
For more detail time series plot at each 5-minute interval see Plot # 5

Number of Steps Taken each day (10/2-11/11), circle = steps in 5-min

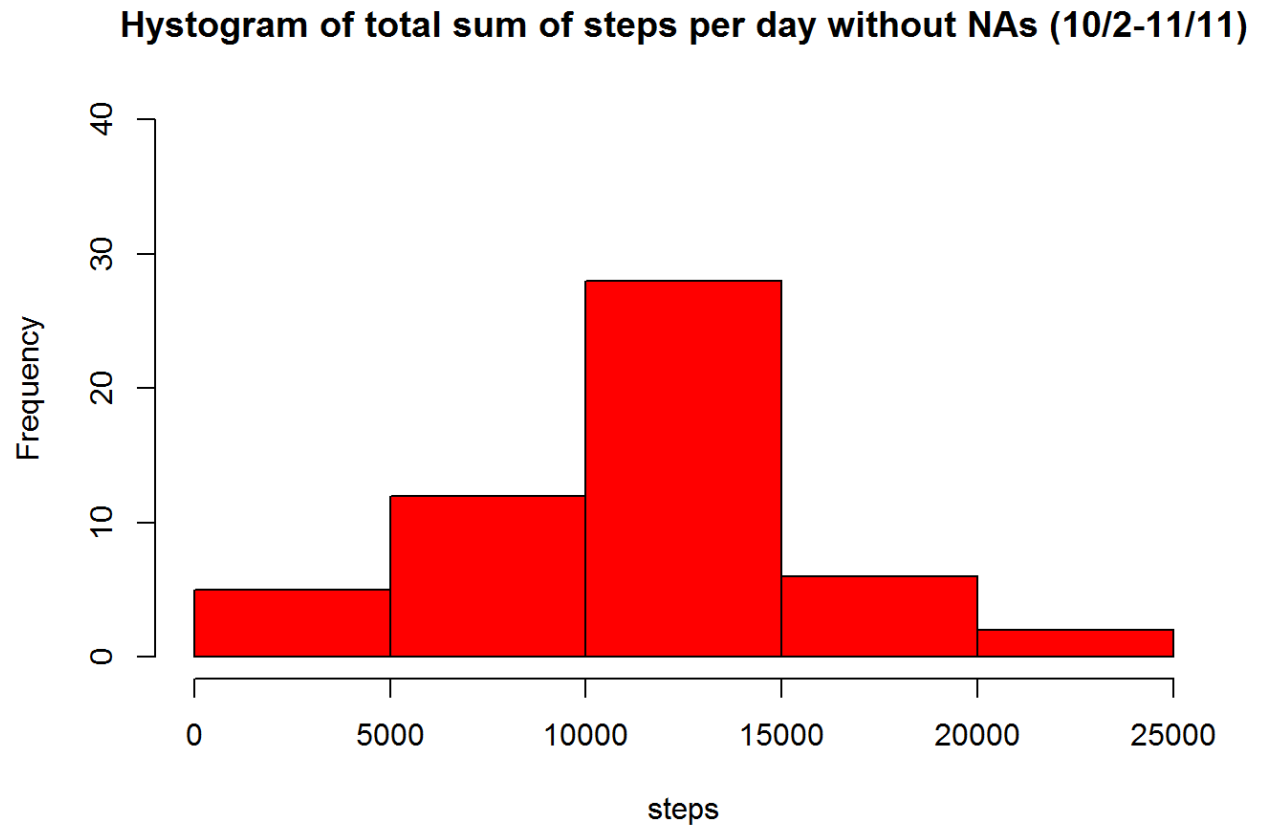


```

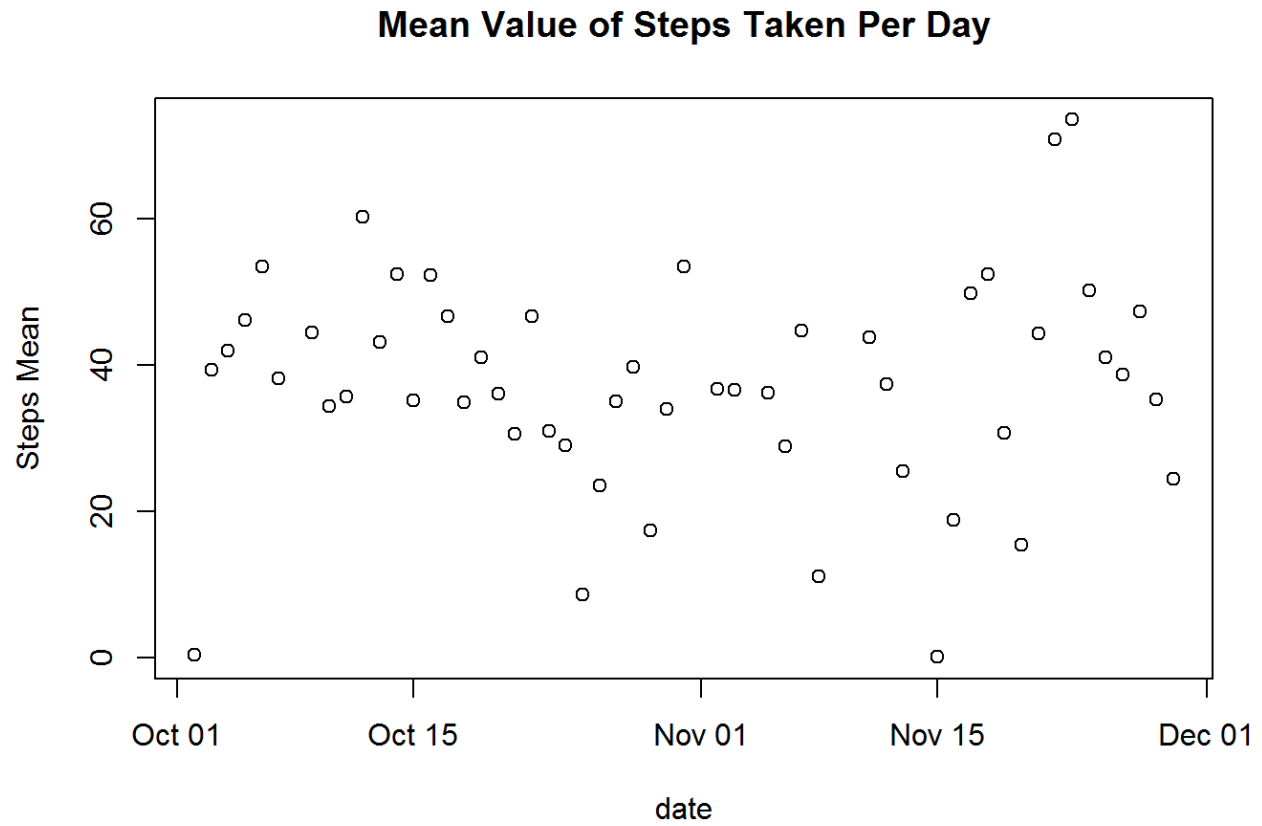
#*****
#Stats per day with NAs remmoved
#*****
sum_d <-with(steps_day, tapply(steps, day, sum))
mean_d <-with(steps_day, tapply(steps, day, mean))
std_d <-with(steps_day, tapply(steps, day, sd))
sqrt_dim_steps <-with(steps_day, tapply(steps, day, length))
sem_d<-std_d/sqrt(sqrt_dim_steps)
median_d <-with(steps_day, tapply(steps,day,median))

```

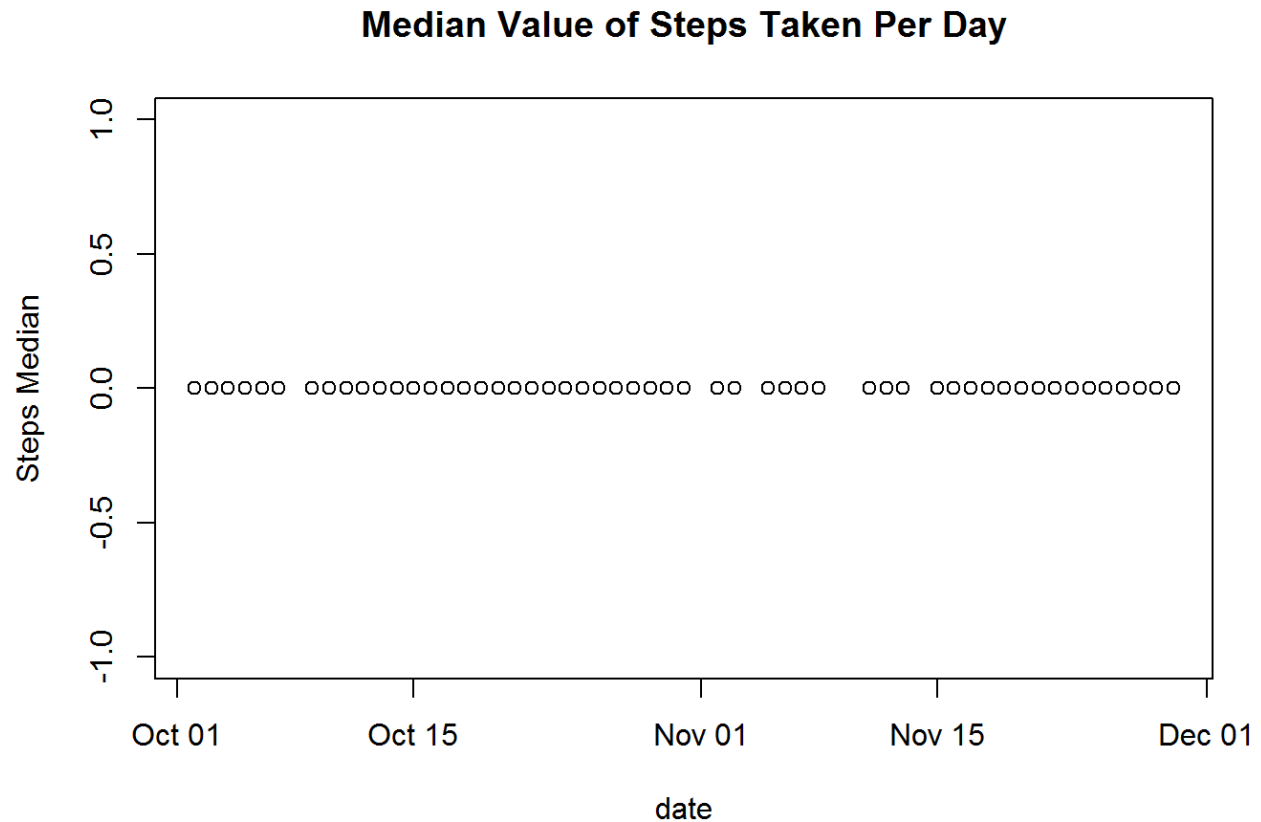
Plot #2 shows a Histogram of the total number of stepst taken each day.



Plot #3 shows the mean of the total number of steps taken per day



Plot #4 shows the median of the total number of steps taken per day

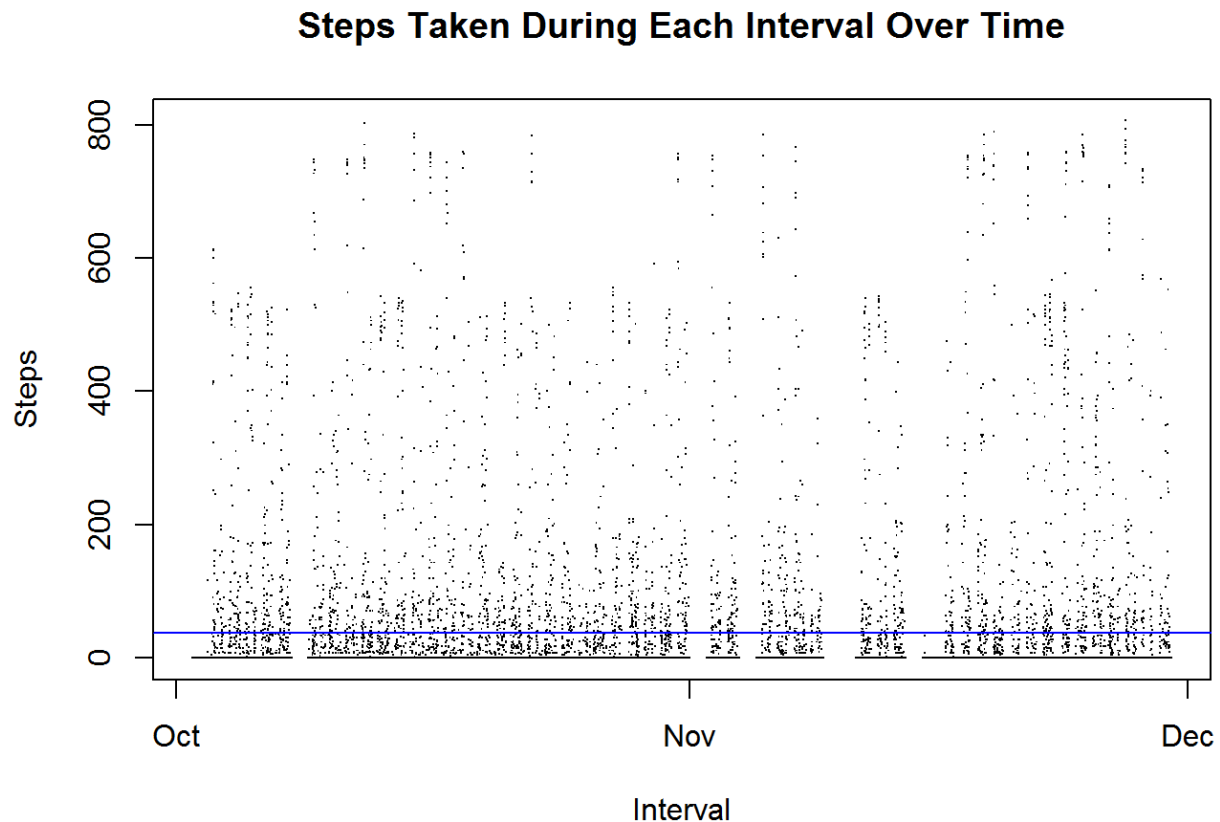


What is the average daily activity pattern?

Calculating the time series plot of the 5-minute interval and the average number of steps taken across all days

```
#####
#First two digit from left to right = Hour
#Last Two digits from left to right = Min
hour_d<-as.character(interval %/% 10^2) #First and second digit (2355= 23hour,
55 min)
minute_d<-as.character(interval %% 10^2) #last two digits
H_M_d <-paste(hour_d,minute_d, sep=":")
day_Hour_Min <-as.POSIXct(paste(day,H_M_d), format="%Y-%m-%d %H:%M")
all_steps_mean <-mean(steps)
```

Plot #5 shows a time series plot of the 5-minute interval and the average number of steps taken across all days (blue line)



Calculating Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
max_steps <- max(steps)
five_minute_interval <- data.frame(interval, steps)
five_minute_interval_max <- subset(five_minute_interval, five_minute_interval$steps == max_steps)
```

The maximum number of steps done in a 5-minute interval are:

```
max_steps
```

```
## [1] 806
```

The 5-minute interval with the maximum number of steps is:

```
five_minute_interval_max$interval
```

```
## [1] 615
```

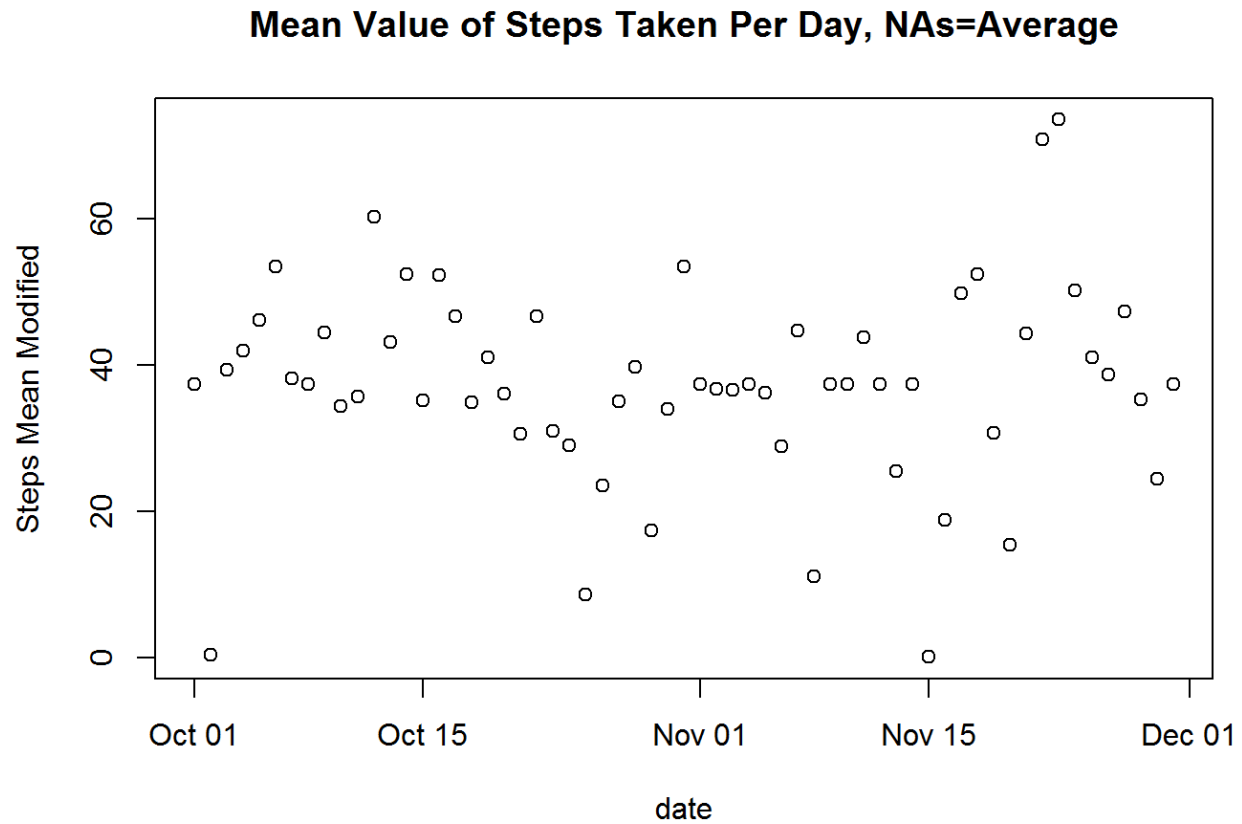
Imputing missing values

The following code describes how to modify the missing step values or NAs

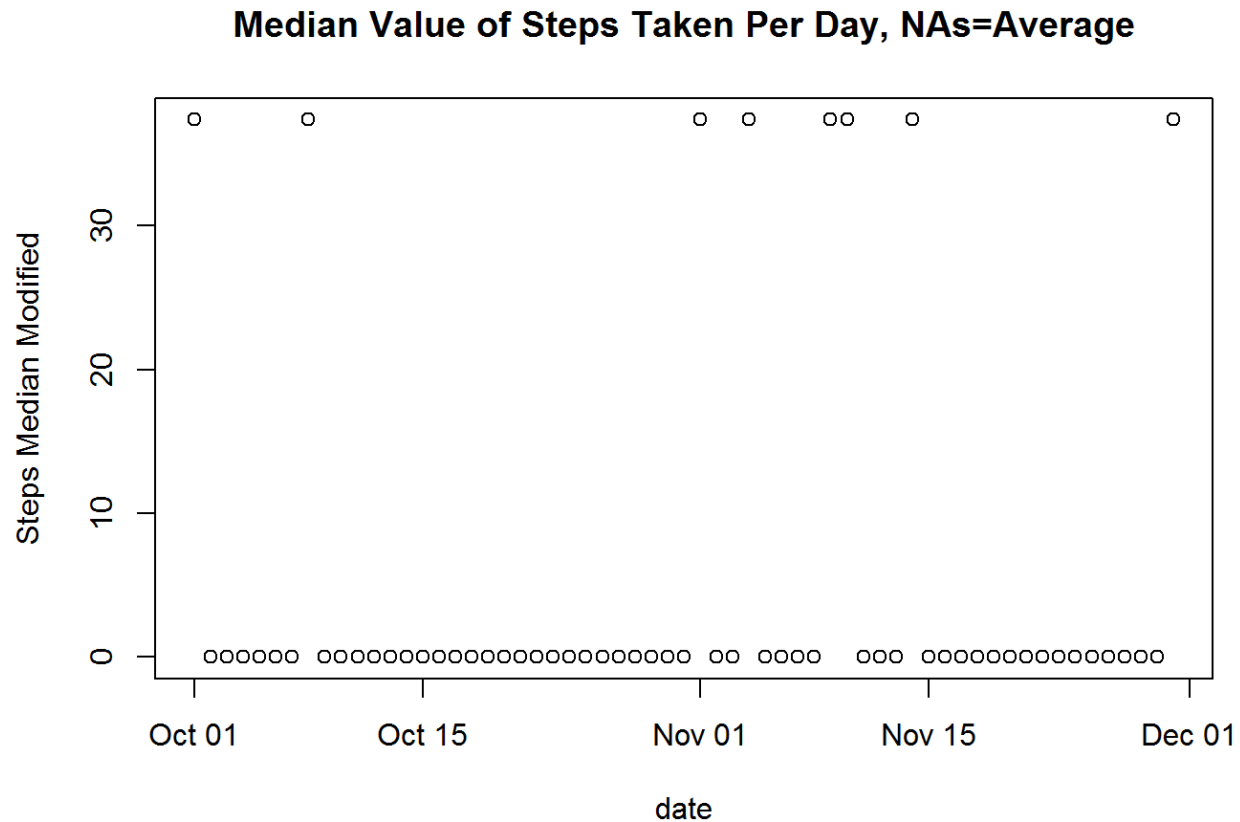
```
#####
#Imputting missing values
#####
number_NAs <- dim(activity_all)-dim(activity)
#Missing values number
number_NAs[1]
```

```
## [1] 2304
```

```
#####
date_m<- as.Date(activity_all[,2], "%Y-%m-%d")
steps_m <- as.numeric(activity_all[,1])
interval_m <-as.numeric(activity_all[,3])
activity_modified <-data.frame(date_m,steps_m,interval_m)
#Below is the code to modify all NAs with the mean of all the steps
activity_modified$steps_m[which(is.na(activity_modified$steps_m))]<- all_steps_
mean
#####
#Stats per day with NAs replaced by overall steps mean
#####
sum_m <-with(activity_modified, tapply(steps_m, date_m, sum))
mean_m <-with(activity_modified, tapply(steps_m, date_m, mean))
std_m <-with(activity_modified, tapply(steps_m, date_m, sd))
sqrt_dim_steps_m <-with(activity_modified, tapply(steps_m, date_m, length))
sem_m<-std_m/sqrt(sqrt_dim_steps_m)
median_m <-with(activity_modified, tapply(steps_m,date_m,median))
plot(unique(activity_modified$date_m), mean_m, type ="p", xlab="date", ylab= "S
taps Mean Modified", main="Mean Value of Steps Taken Per Day, NAs=Average")
```

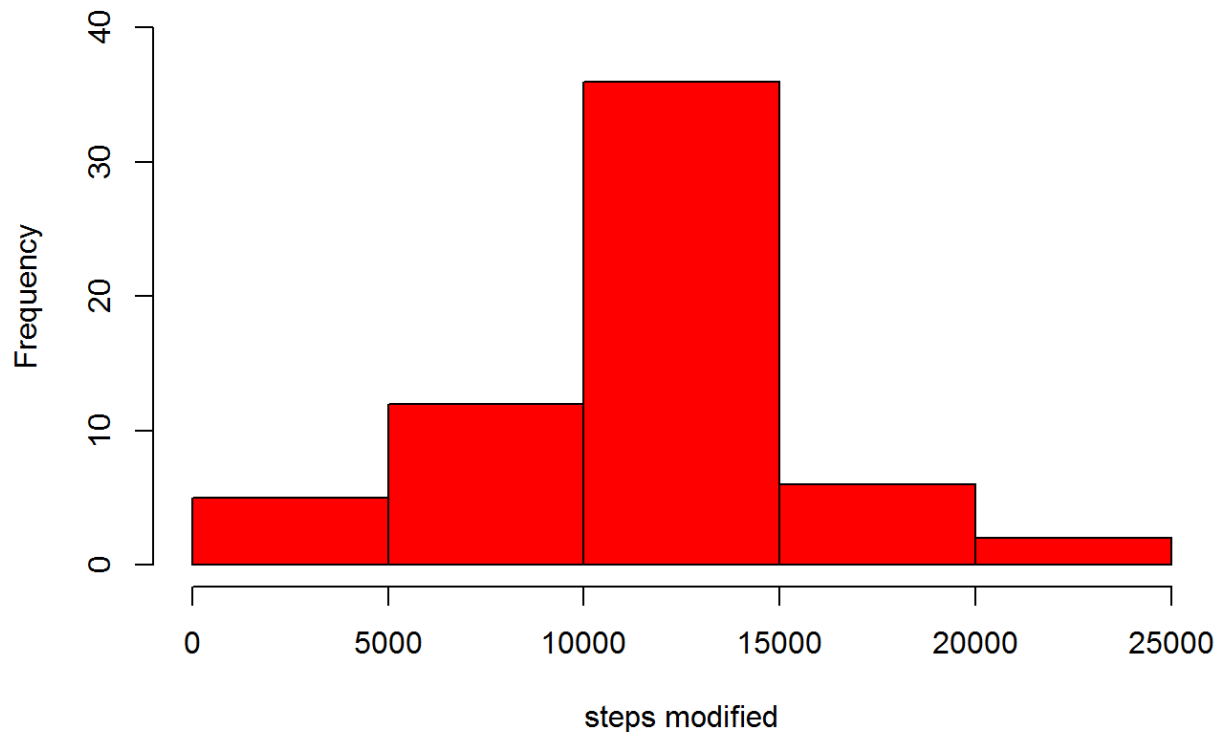
```
plot(unique(activity_modified$date_m), median_m, type = "p", xlab="date", ylab="Steps Median Modified", main="Median Value of Steps Taken Per Day, NAs=Averag e")
```



```
stats_day_m <- data.frame(date1_m=unique(activity_modified$date_m), sum_m, mean_
m, median_m, std_m, sem_m, sqrt_dim_steps_m)
```

Plotting the new histogram where the "NAs" were modified by the overall average number of steps

Histogram of total sum of steps per day with Modified NAs (10/2-11/11)



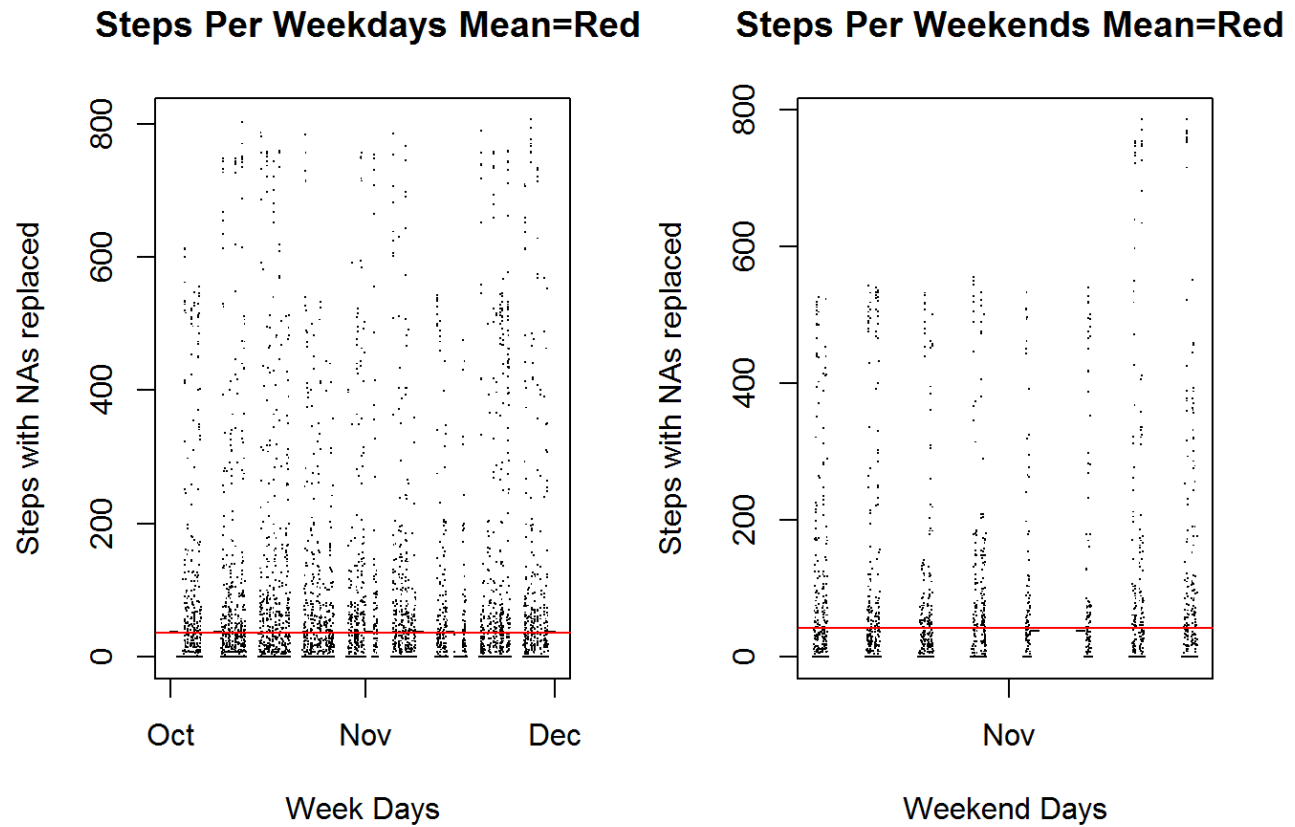
Are there differences in activity patterns between weekdays and weekends?

Calculating Steps done during the weekdays and the Weekends

```
#####
#Separating weekend data from week data
#####
hour_m<-as.character(interval_m %/% 10^2) #First and second digit (2355= 23hou
r, 55 min)
minute_m<-as.character(interval_m %% 10^2) #last two digits
H_M_m <-paste(hour_m,minute_m, sep=":")
day_Hour_Min_m <-as.POSIXct(paste(date_m,H_M_m), format="%Y-%m-%d %H:%M")
activity_modified$date_time <-day_Hour_Min_m

activity_modified_wkend <-subset(activity_modified,is.weekend(activity_modified
$date_m))
activity_modified_wk <-subset(activity_modified,!is.weekend(activity_modified$d
ate_m))
all_steps_mean_wkend <-mean(activity_modified_wkend$steps_m)
all_steps_mean_wk <-mean(activity_modified_wk$steps_m)
```

Panel plot comparing the average number of steps taken per 5-min interval across weekdays and weekend



```
## null device  
##           1
```