Reproducible Research: Peer Assessment 1

Jose Noriega

15/2/2020

This assigment is about to describe in multiple parts a monitoring example project using R Markdown. We will need to write a report that answers the questions detailed in “Instruction.pdf” file using the data sample download in the “activity.zip” file. completing the entire assignment in this single R markdown document which can be processed by knitr and be transformed into an HTML file. Following the results per questions:

## Loading and preprocessing the data

To load the data I used read.csv funtion, considering firt file as the headers and all the missing vaues, as follow:

data<-read.csv("activity.csv",header = TRUE,na.strings = "NA")  
head(data)

## steps date interval  
## 1 NA 2012-10-01 0  
## 2 NA 2012-10-01 5  
## 3 NA 2012-10-01 10  
## 4 NA 2012-10-01 15  
## 5 NA 2012-10-01 20  
## 6 NA 2012-10-01 25

## What is mean total number of steps taken per day?

I will calculate the mean total number steps taken by day, considery a tapply funtion which sum total steps by date, then calculate the mean of the StepsPerDay The result is the next:

StepsPerDay<-tapply(data$steps,data$date,sum,na.rm = TRUE)  
MeanStepPerDay<-mean(StepsPerDay)  
MeanStepPerDay

## [1] 9354.23

An histogram of the Total of number of steps by day is calculate and plot by the next code:

library(ggplot2)  
qplot(StepsPerDay,xlab = "Total Steps per day", ylab = "Frecuency",binwidth=500)



## What is the average daily activity pattern?

To calculate the Mean and the Median number step by day, I used the following code with their results:

MeanStepPerDay<-mean(StepsPerDay)  
MeanStepPerDay

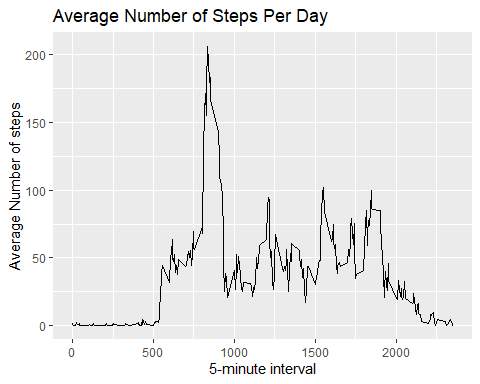
## [1] 9354.23

MedianStepPerDay<-median(StepsPerDay)  
MedianStepPerDay

## [1] 10395

For the Time series plot average number steps taken and the 5-minute interval that, on average, contains the maximun numer of step, resulting a graphics, I code the following:

AveDayActPatt<-aggregate(x=list(meanSteps=data$steps),by=list(interval=data$interval), FUN=mean,na.rm=TRUE)  
  
ggplot(data = AveDayActPatt,aes(x=interval,y=meanSteps))+geom\_line()+ggtitle("Average Number of Steps Per Day")+xlab("5-minute interval")+ylab("Average Number of steps")



## The 5-minutes interval on average per day in the data contains the maximun number of steps?

MaxSteps<-which.max(AveDayActPatt$meanSteps)  
MostOfSteps<-gsub("([0-9]{1,2})([0-9]{2})","\\1:\\2",AveDayActPatt[MaxSteps,"interval"])  
MostOfSteps

## [1] "8:35"

This “Interval number” indicates that 8.35 AM is the time when the average person is most active

## Code to describe and show a strategy for imputing missing data

The total number of missings values are calculate bu the next code

MissingValues<-length(which(is.na(data$steps)))  
MissingValues

## [1] 2304

## Make an Histogram of the number of total steps taken by day

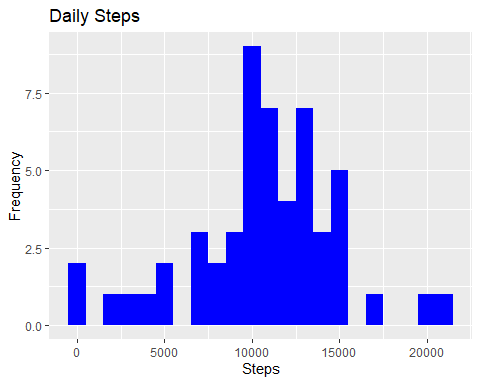
Following the histogram which show the total steps taken by day, in thi section I consider the advantage for the data.table function. Folowing the code and the histogram.

activity<-data.table::fread(input="activity.csv")  
TotalSteps<-activity[,c(lapply(.SD,sum)),.SDcols=c("steps"),by=.(date)]  
TotalSteps[,.(MeanSteps=mean(steps),MedianSteps=median(steps))]

## MeanSteps MedianSteps  
## 1: NA NA

ggplot(TotalSteps,aes(x=steps))+geom\_histogram(fill="blue",binwidth = 1000)+labs(title="Daily Steps",x="Steps",y="Frequency")

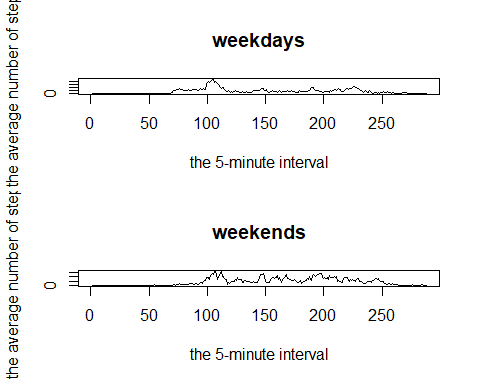
## Warning: Removed 8 rows containing non-finite values (stat\_bin).



## Are there differences in activity patterns between weekdays and weekends?

Building a factor variable considering weeks and weekends as follow:

data$date<-as.POSIXct(data$date)  
dataFix <-data  
for(i in unique(dataFix$interval)) {  
 dataFix$steps[is.na(dataFix$steps) & dataFix$interval == i] <- round(mean(dataFix$steps[data$interval == i],na.rm = TRUE))  
}  
dataFix$weekDay<-as.POSIXlt(data$date)$wday == 0 | as.POSIXlt(data$date)$wday ==6  
dataFix$weekDay<-factor(dataFix$weekDay,levels =c(F, T), labels=c("weekday","weekend"))  
  
stepsWeekDay<-tapply(dataFix$steps[dataFix$weekDay=="weekday"],dataFix$interval[dataFix$weekDay=="weekday"],mean)  
stepsWeekEnd<-tapply(dataFix$steps[dataFix$weekDay=="weekend"],dataFix$interval[dataFix$weekDay=="weekend"],mean)  
  
par(mfrow = c(2,1))  
plot(stepsWeekDay,type = "l",main = "weekdays"  
 ,xlab = "the 5-minute interval"  
 ,ylab = "the average number of steps")  
plot(stepsWeekEnd,type = "l",main = "weekends"  
 ,xlab = "the 5-minute interval"  
 ,ylab = "the average number of steps")

 ## Loading and processing the data

The data had been delivered by Coursera assignment in the link (<https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip>), indicating that "NA’ values are missing values