Reproducible\_Research\_Project\_1.R

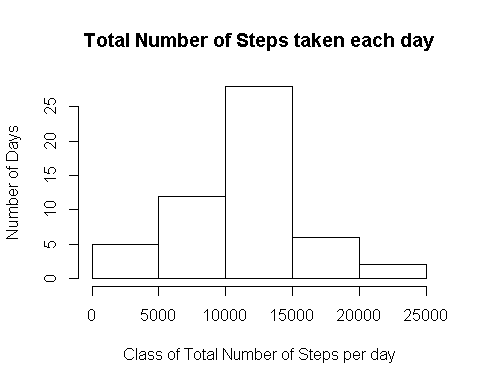
JOlynick

2020-10-14

#Step 1  
if(!file.exists("getdata-projectfiles-UCI HAR Dataset.zip")) {  
 temp <- tempfile()  
 download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip",temp)  
 unzip(temp)  
 unlink(temp)  
}  
  
activity<-read.csv("activity.csv")  
head(activity)

## 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

#Step 2  
totalStepsByDay<-aggregate(steps~date, activity, sum)  
  
hist(totalStepsByDay$steps, xlab="Class of Total Number of Steps per day", ylab="Number of Days", main="Total Number of Steps taken each day")



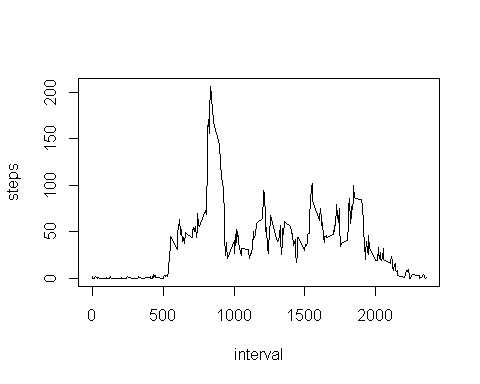
#Step 3  
mean\_raw<-mean(totalStepsByDay$steps)  
mean\_raw

## [1] 10766.19

median\_raw<-median(totalStepsByDay$steps)  
median\_raw

## [1] 10765

#Step 4  
averageStepsbyInterval<-aggregate(steps~interval, activity, mean)  
  
with(averageStepsbyInterval, plot(interval, steps, type = "l"))



#Step 5  
averageStepsbyInterval[which.max(averageStepsbyInterval[,2]),1]

## [1] 835

#Step 6  
missingIndex<-is.na(activity[,1])  
  
#Step 7  
m<-mean(averageStepsbyInterval$steps)  
  
activity1<-activity  
activity1[missingIndex,1]<-m  
head(activity1)

## steps date interval  
## 1 37.3826 2012-10-01 0  
## 2 37.3826 2012-10-01 5  
## 3 37.3826 2012-10-01 10  
## 4 37.3826 2012-10-01 15  
## 5 37.3826 2012-10-01 20  
## 6 37.3826 2012-10-01 25

#Step 8  
totalStepsByDay1<-aggregate(steps~date, activity1, sum)  
hist(totalStepsByDay1$steps, xlab="Class of Total Number of Steps per day", ylab="Number of Days", main="Number of Steps taken each day after missing values are imputed")  
  
totalStepsByDay1<-aggregate(steps~date, activity1, sum)  
mean\_afterImput<-mean(totalStepsByDay1$steps)  
mean\_afterImput

## [1] 10766.19

median\_afterImput<-median(totalStepsByDay1$steps)  
median\_afterImput

## [1] 10766.19

#Step 9  
activity1$date<-as.Date(activity1$date)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.3

##   
## 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

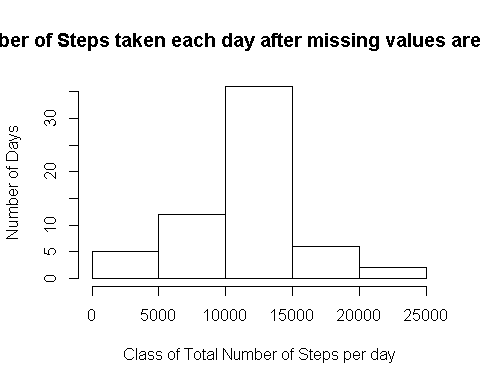
#Step 10  
activity2<-activity1%>%  
 mutate(dayType= ifelse(weekdays(activity1$date)=="Saturday" | weekdays(activity1$date)=="Sunday", "Weekend", "Weekday"))  
head(activity2)

## steps date interval dayType  
## 1 37.3826 2012-10-01 0 Weekday  
## 2 37.3826 2012-10-01 5 Weekday  
## 3 37.3826 2012-10-01 10 Weekday  
## 4 37.3826 2012-10-01 15 Weekday  
## 5 37.3826 2012-10-01 20 Weekday  
## 6 37.3826 2012-10-01 25 Weekday

#Step 11  
averageStepByDayTypeAndInterval<-activity2 %>%  
 group\_by(dayType, interval) %>%  
 summarize(averageStepByDay=sum(steps))  
  
head(averageStepByDayTypeAndInterval)

## # A tibble: 6 x 3  
## # Groups: dayType [1]  
## dayType interval averageStepByDay  
## <chr> <int> <dbl>  
## 1 Weekday 0 315.  
## 2 Weekday 5 242.  
## 3 Weekday 10 231.  
## 4 Weekday 15 232.  
## 5 Weekday 20 228.  
## 6 Weekday 25 283.

#Step 12  
library(lattice)



with(averageStepByDayTypeAndInterval,   
 xyplot(averageStepByDay ~ interval | dayType,   
 type = "l",   
 main = "Total Number of Steps within Intervals by dayType",  
 xlab = "Daily Intervals",  
 ylab = "Average Number of Steps"))

