PA\_1 template Coursera Reproducible Research Project #1

Edward S.

December 17, 2017

# Set Up

library(Hmisc)

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, round.POSIXt, trunc.POSIXt, units

# 1A: Code for Reading in the Activity Dataset:

if(!file.exists("C:/Users/edwsp/Desktop/repdata%2Fdata%2Factivityactivity.csv")){  
 fileUrl<-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"  
 download.file(fileUrl,   
 destfile = paste0(getwd(),'/repdata%2Fdata%2Factivity.zip'))  
}  
  
if(!file.exists("activity.csv")){  
 unzip("C:/Users/edwsp/Desktop/repdata%2Fdata%2Factivity/activity.csv")}  
  
A <- read.csv("data/activity.csv", header=TRUE)

# 1A: ALternative Read In of Data:

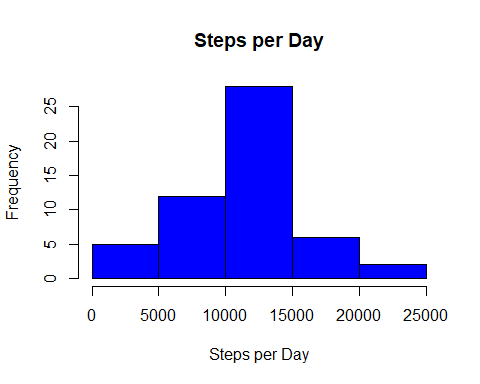
if(!file.exists("(paste0(getwd(),'/repdata%2Fdata%2Factivity.zip')")) {  
 temp <- tempfile()  
 download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip",  
 temp)  
 unzip(temp)  
 unlink(temp)  
}  
  
A<-read.csv("activity.csv")

# 1B: Code for Processing the Data:

str(A)  
attach(A)

# 2: Determine the number of Steps per day and Histogram of Total Number of Steps per Day:

SPD<-tapply(steps, date, sum)  
hist(SPD, main="Steps per Day", col="blue", xlab="Steps per Day")



# 3. Mean and Median Number of Steps per Day:

summary(SPD)

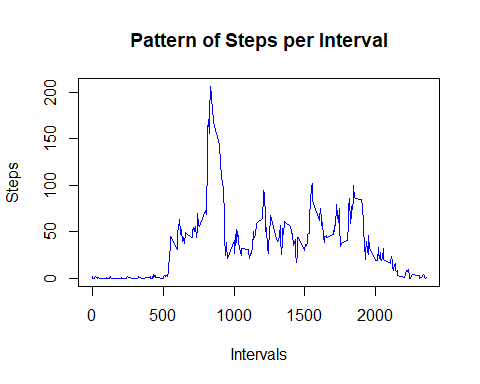
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 41 8841 10765 10766 13294 21194 8

# 4. Time Series Plot of Average Number of Steps Taken:

SPI.t<-tapply(steps, interval, mean, na.rm=TRUE)  
summary(SPI.t)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 2.486 34.113 37.383 52.835 206.170

plot(as.numeric(names(SPI.t)), SPI.t, type="l", main = "Pattern of Steps per Interval", xlab="Intervals", ylab="Steps", col="blue")



# 5. The 5 Minute Interval that on Average has the Maximum Number of Steps:

Max.steps<-names(sort(SPI.t, decreasing = TRUE)[1])  
Max.steps

## [1] "835"

# 6. Code to Describe and Show Strategy to Impute Missing Data:

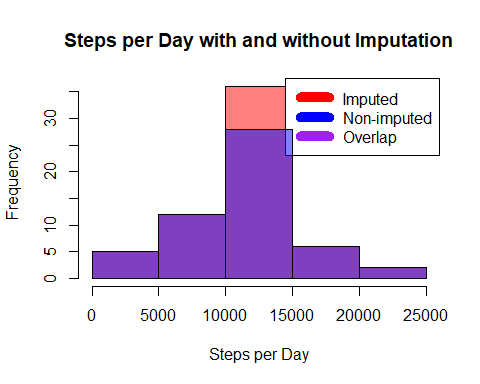
I'm simple- I'm just going to impute median number of steps for each interval when there's a missing value.

sum(is.na(steps))  
(sum(is.na(steps)))/(length(A$steps))  
  
SPI.a<-aggregate(steps, list(interval), mean, na.rm=TRUE)  
colnames(SPI.a)<-c("interval", "SPI")  
B<-merge(A, SPI.a, by="interval", all.x=TRUE)  
B$steps.i<-ifelse(is.na(B$steps), B$SPI, B$steps)  
attach(B)

## The following objects are masked from A:  
##   
## date, interval, steps

# 7. Histogram of the Total Number of Steps each Day after Imputation:

SPD.i<-tapply(steps.i, date, sum)  
  
hist(SPD.i, main="Steps per Day with and without Imputation", col=rgb(1,0,0,0.5), xlab="Steps per Day")   
  
legend("topright", c("Imputed", "Non-imputed", "Overlap"), col=c("red", "blue", "purple"), lwd=10)  
  
hist(SPD, main="Steps per Day", col=rgb(0,0,1,0.5), add=TRUE)



SPD.i.mean<-mean(SPD.i)  
SPD.i.median<-median(SPD.i)  
  
SPD.mean<-mean(SPD, na.rm=TRUE)  
SPD.median<-median(SPD, na.rm=TRUE)  
  
Mean.dif<-SPD.i.mean-SPD.mean  
Median.dif<-SPD.i.median-SPD.median  
  
SPD.i.mean

## [1] 10766.19

SPD.i.median

## [1] 10766.19

SPD.mean

## [1] 10766.19

SPD.median

## [1] 10765

Mean.dif

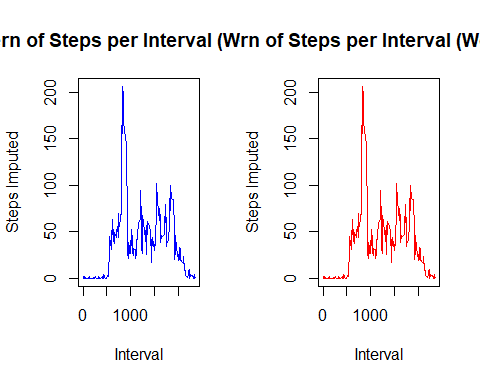
## [1] 0

Median.dif

## [1] 1.188679

# 8. Panel Plot comparing Average Number of Steps taken per 5 Min Interval across Weekends and Weekdays

library(lattice)  
  
B$Day<-weekdays(as.Date(B$date))  
B$Daytype<-ifelse(((B$Day=="Saturday")|(B$Day=="Sunday")), "Weekend", "Weekday")  
  
  
SPI.i<-aggregate(B$steps.i, list(B$interval), mean, na.rm=TRUE)  
colnames(SPI.i)<-c("interval", "SPI.i")  
C<-merge(B, SPI.i, by="interval", all.x=TRUE)  
  
SPI.i.Wkday<-C[which(C$Daytype=="Weekday"),]  
SPI.i.Wkend<-C[which(C$Daytype=="Weekend"),]  
  
par(mfrow=c(1,2))  
   
plot(SPI.i.Wkday$interval, SPI.i.Wkday$SPI.i,  
 xlab = "Interval",   
 ylab = "Steps Imputed",   
 main = "Pattern of Steps per Interval (Weekdays)",   
 type = "l",  
 col="blue")  
  
plot(SPI.i.Wkend$interval, SPI.i.Wkend$SPI.i,   
 xlab = "Interval",   
 ylab = "Steps Imputed",   
 main = "Pattern of Steps per Interval (Weekends)",   
 type = "l",  
 col="red")



# 9. All of the R Code needed to Reproduce the Plots can be found in the above report.