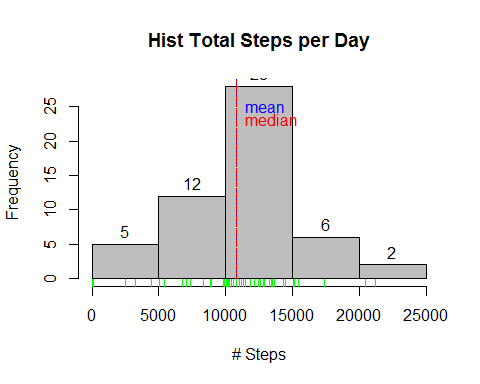
RRproject1

John Fields

July 29, 2017

Read Data and Develop Histogram of Daily steps

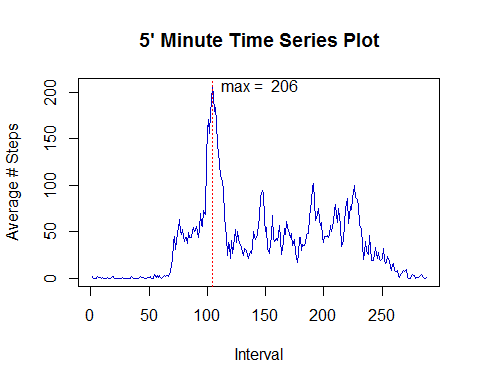
Ractivity <- read.csv("activity.csv")  
  
origData <- complete.cases(Ractivity)  
Rsteps <- subset(Ractivity, complete.cases(Ractivity))  
  
ByDay <- split(Rsteps,Rsteps$date, drop=TRUE) # split the complete cases by date   
dailySteps <- sapply(ByDay, function(x) sum(x$steps)) # build a numeric vector w/ daily sum of steps   
hist(dailySteps, main="Hist Total Steps per Day", xlab="# Steps",   
 col="gray", labels = TRUE) # plot a histogram and lable values  
abline(v=mean(dailySteps), lty=5, col="blue") # draw a blue line thru the mean   
abline(v=median(dailySteps), lty=6, col="red") # draw a red line thru the median   
text(mean(dailySteps),25,labels="mean", pos=4, col="blue") # label the mean   
text(mean(dailySteps),23,labels="median", pos=4, col="red") # label the median  
  
rug(dailySteps, col="green")



The mean is 'mean(dailysteps)' The median is 'median(dailysteps)' The sum is 'sum(dailysteps)'

Create a Time Series Chart

splitMI <- split(Rsteps,Rsteps$interval, drop=TRUE) # split the complete cases by date   
intervalAvg <- sapply(splitMI, function(x) mean(x$steps)) # vector of Avg. steps per interval   
plot(intervalAvg, type="l",   
 main="5' Minute Time Series Plot",   
 ylab="Average # Steps",   
 xlab="Interval", col="blue3") # plot the 5' time series  
abline(v=which.max(intervalAvg), lty=3, col="red") # draw a red line thru the median   
text(which.max(intervalAvg),max(intervalAvg),   
 labels=paste("max = ",as.character(round(max(intervalAvg)))),   
 pos=4, col="black")



splitmax <- names(which.max(intervalAvg))  
splitavg <- round(max(intervalAvg))  
splitname <- which.max(intervalAvg)

The name index where the max number of step taken is 104 The max number of steps is 206

dataNA <- sum(is.na(Ractivity$steps))

The number of missing values is 2304

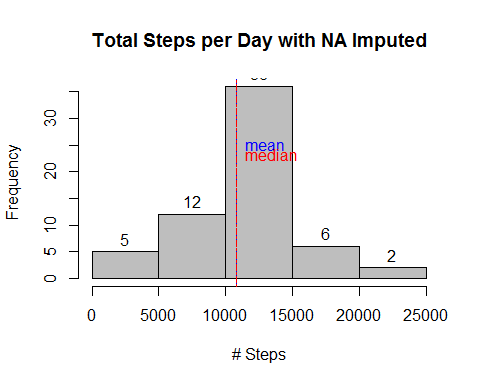
data1 <- sum(!is.na(Ractivity$steps))

The number of values is 15264

five\_avg <- rep(intervalAvg, 61)  
data2 <- Ractivity  
  
for(i in 1:length (data2[,1])){  
   
 if(is.na(data2 [i,1])==TRUE){  
 data2[i,1] = five\_avg [i]  
 }}  
  
daily1<-c()  
  
  
for (i in 1:61){ #the total number of days in October and November is 31+30=61  
 start<-(i-1)\*288+1 #there are 288 five-minute steps in a day; 24\*60/5=288  
 last<-(i-1)\*288+288  
 temp<-data2[start:last,1] #extracting all 5-minute steps for each day  
 daily1<-c(daily1,sum(temp)) #concatenating the daily totals   
}

Devolop the New Histogram with the NA values imputed

hist(daily1, main="Total Steps per Day with NA Imputed", xlab="# Steps",   
 col="gray", labels = TRUE) # plot a histogram and lable values  
abline(v=mean(daily1), lty=5, col="blue") # draw a blue line thru the mean   
abline(v=median(daily1), lty=6, col="red") # draw a red line thru the median   
text(mean(daily1),25,labels="mean", pos=4, col="blue") # label the mean   
text(mean(daily1),23,labels="median", pos=4, col="red")



Summarize the two datasets, 1 with NA values and the other with imputed values

summary(dailySteps)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 41 8841 10760 10770 13290 21190

summary(daily1)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 41 9819 10770 10770 12810 21190

Plot Weekday versus Weedend Steps Taken

data2$date<-as.Date(data2$date)  
data2$day<-weekdays(data2$date)  
  
data2\_weekdays<-data2[(!data2$day %in% c("Saturday","Sunday")),] # weekdays  
data2\_weekend<-data2[(data2$day %in% c("Saturday","Sunday")),] # weekend  
  
weekday\_steps<-data2\_weekdays[,1]  
temp<-matrix(weekday\_steps,nrow=288)  
weekday\_steps\_average<-apply(temp,1,mean)  
  
  
weekend\_steps<-data2\_weekend[,1]  
temp<-matrix(weekend\_steps,nrow=288)  
weekend\_steps\_average<-apply(temp,1,mean)  
  
par(mfrow=c(2,1))  
  
plot(Ractivity$interval[1:288],weekday\_steps\_average, type="l",xlab='Intervals',ylab="Number of steps",  
 col='black',lwd=2, main="Weekday")  
  
plot(Ractivity$interval[1:288],weekend\_steps\_average, type="l", xlab='Intervals',ylab="number of steps",  
 col='red',lwd=2,main="Weekend")

