Reproducible Research Project 1

Output: html\_document: keep\_md: true

## Loading and preprocessing the data

setwd("./R/Reproducible Research")  
  
activityData <- read.csv(file = "activity.csv", sep =",", colClasses=c("integer","Date","integer"))  
str(activityData)

## 'data.frame': 17568 obs. of 3 variables:  
## $ steps : int NA NA NA NA NA NA NA NA NA NA ...  
## $ date : Date, format: "2012-10-01" "2012-10-01" "2012-10-01" ...  
## $ interval: int 0 5 10 15 20 25 30 35 40 45 ...

summary(activityData)

## steps date interval   
## Min. : 0.00 Min. :2012-10-01 Min. : 0.0   
## 1st Qu.: 0.00 1st Qu.:2012-10-16 1st Qu.: 588.8   
## Median : 0.00 Median :2012-10-31 Median :1177.5   
## Mean : 37.38 Mean :2012-10-31 Mean :1177.5   
## 3rd Qu.: 12.00 3rd Qu.:2012-11-15 3rd Qu.:1766.2   
## Max. :806.00 Max. :2012-11-30 Max. :2355.0   
## NA's :2304

## What is Mean Total Number of Steps Taken Per Day

Mean Number of Steps

totalSteps <- tapply(activityData$steps, activityData$date, sum, na.rm=TRUE)  
stepMean <- mean(totalSteps)  
stepMean

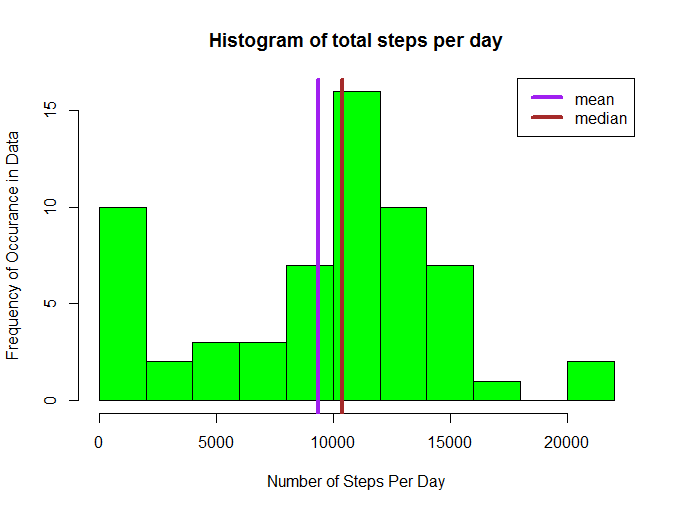
## [1] 9354.23

Median   
stepMedian <- median(totalSteps)  
stepMedian

## [1] 10395

## Histogram of Total Steps Taken Each Day

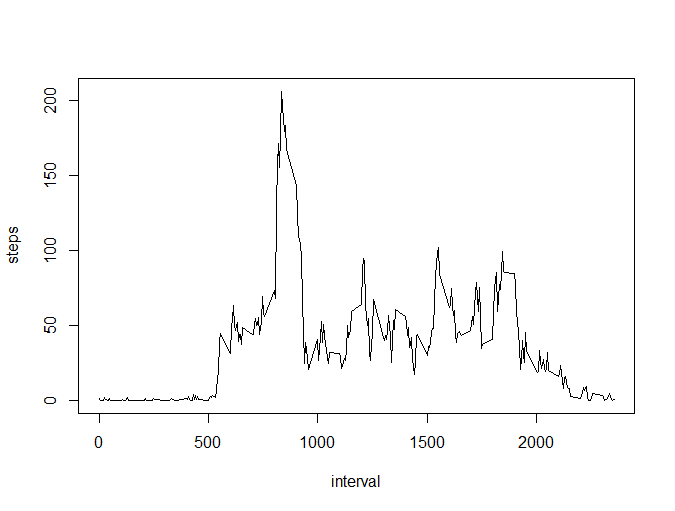
hist(totalSteps, breaks = 12,col =' green'  
 xlab = "Number of Steps Per Day",  
 ylab = "Frequency of Occurance in Data",  
 main = "Histogram of total steps per day")  
  
abline(v=stepMean, col="purple", lwd ="4")  
abline(v=stepMedian, col="brown", lwd ="4")  
legend(x = "topright", legend = c("mean", "median"), col=c("purple","brown"), lwd = 4)



plot of chunk unnamed-chunk-3

## Average Daily Activity Pattern

library(plyr)  
dailyActivity <- ddply(activityData, .(interval), summarize, steps = mean(steps, na.rm = TRUE))  
with(dailyActivity, plot(interval, steps, type = "l"))



plot of chunk unnamed-chunk-4

5 Minutes Containing the Maximum Number of Steps

maxSteps <- dailyActivity[which.max(dailyActivity$steps),]$interval

## Inputing Missing Values

Find Out How Many Missing Values

sum(is.na(activityData$steps))

## [1] 2304

# Filling in All the Missing Values in the Dataset

stepValues <- data.frame(activityData$steps)  
stepValues [is.na(stepValues),] <- ceiling(tapply(X=activityData$steps,INDEX=activityData$interval,FUN=mean,na.rm=TRUE))  
  
newData <- cbind(stepValues, activityData[,2:3])  
colnames(newData) <- c("Steps", "Date", "Interval")  
  
summary(newData)

## Steps Date Interval   
## Min. : 0.00 Min. :2012-10-01 Min. : 0.0   
## 1st Qu.: 0.00 1st Qu.:2012-10-16 1st Qu.: 588.8   
## Median : 0.00 Median :2012-10-31 Median :1177.5   
## Mean : 37.45 Mean :2012-10-31 Mean :1177.5   
## 3rd Qu.: 27.00 3rd Qu.:2012-11-15 3rd Qu.:1766.2   
## Max. :806.00 Max. :2012-11-30 Max. :2355.0

Newly Inputed Dataset from the Old Dataset

totalSteps2 <- tapply(newData$Steps, newData$Date, sum, na.rm=TRUE)  
stepMean2 <- mean(totalSteps2)  
stepMean2

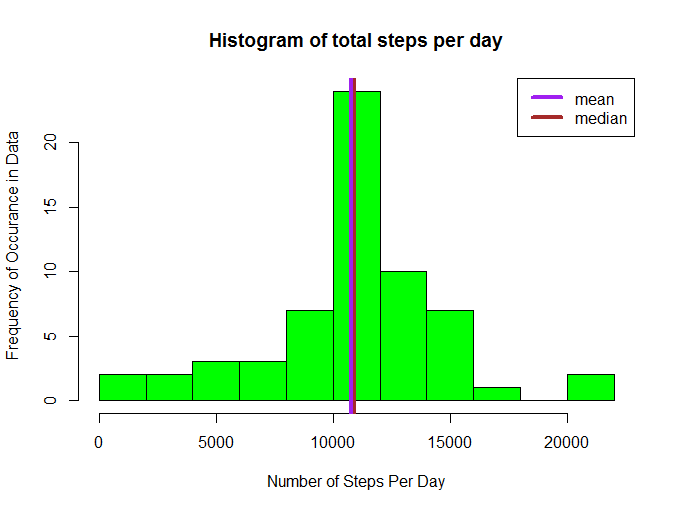
## [1] 10784.92

stepMedian2 <- median(totalSteps2)  
stepMedian2

## [1] 10909

## New Input Histogram Data

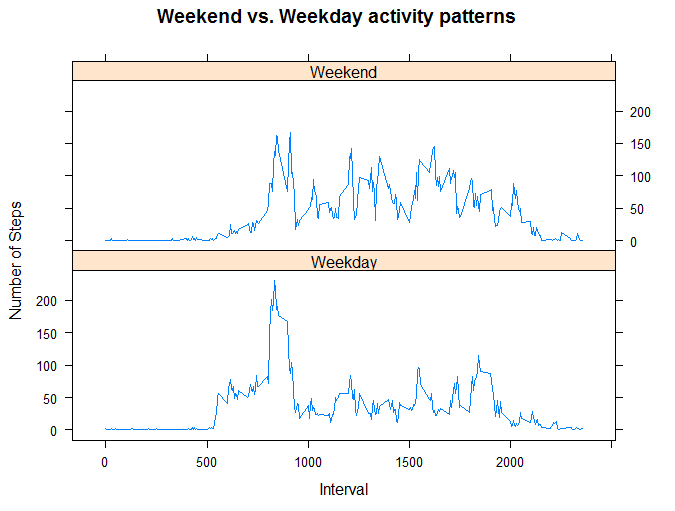
hist(totalSteps2, breaks = 12,col =' green'  
 xlab = "Number of Steps Per Day",  
 ylab = "Frequency of Occurance in Data",  
 main = "Histogram of total steps per day")  
  
abline(v=stepMean2, col="purple", lwd ="5")  
abline(v=stepMedian2, col="brown", lwd ="3")  
legend(x = "topright", legend = c("mean", "median"), col=c("purple","brown"), lwd = 4)



plot of chunk unnamed-chunk-9

## Differences in Activity Patterns Between Weekend and Weekdays

newData$Weekend <- weekdays(newData$Date) == "Saturday" | weekdays(newData$Date) == "Sunday"  
  
newData$Weekend <- factor(newData$Weekend, levels = c(F, T), labels = c("Weekday", "Weekend"))  
  
activity <- ddply(newData, .(Interval, Weekend), summarize, steps = mean(Steps, na.rm = TRUE))  
  
library(lattice)  
xyplot(steps ~ Interval | Weekend, activity, type = "l", layout = c(1, 2), ylab = "Number of Steps", xlab = "Interval", main = "Weekend vs. Weekday activity patterns")



plot of chunk unnamed-chunk-10