Reproducible Research: Peer Assessment 1

Bui Dinh Ngoc

8 Dec 2015

## Loading and preprocessing the data

# 1.Load the data (i.e. read.csv()) some lib require here

library(readr)  
library(data.table)  
library(plyr)  
library(lattice)  
  
dat = read\_csv('activity.zip',na='NA',col\_types='dcd')  
  
names(dat)

## [1] "steps" "date" "interval"

head(dat)

## 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

str(dat)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 17568 obs. of 3 variables:  
## $ steps : num NA NA NA NA NA NA NA NA NA NA ...  
## $ date : chr "2012-10-01" "2012-10-01" "2012-10-01" "2012-10-01" ...  
## $ interval: num 0 5 10 15 20 25 30 35 40 45 ...

# Process/transform the data (if necessary) into a format suitable for your analysis

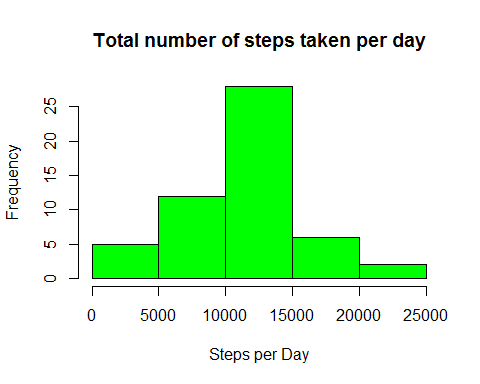
# Convert the 2nd column to Date type:

dat$date<-as.Date(dat$date, "%Y-%m-%d")  
head(dat)

## 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?

tidydata<-na.omit(dat)  
steps<-aggregate(steps ~ date, tidydata, sum)  
hist(steps$steps, main="Total number of steps taken per day", xlab="Steps per Day", col="green")



mean<-mean(steps$steps)  
median<-median(steps$steps)  
mean

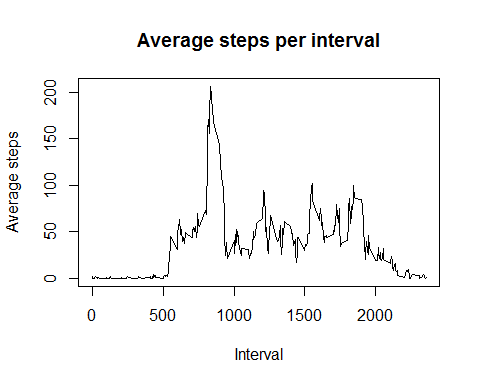
## [1] 10766.19

median

## [1] 10765

## What is the average daily activity pattern?

Interval<-unique(tidydata$interval)  
avgStep <- tapply(tidydata$steps, tidydata$interval, mean)  
##and plot  
plot(Interval,avgStep,type="l",ylab="Average steps", main="Average steps per interval")



maximum <- names(which.max(avgStep))  
maximum

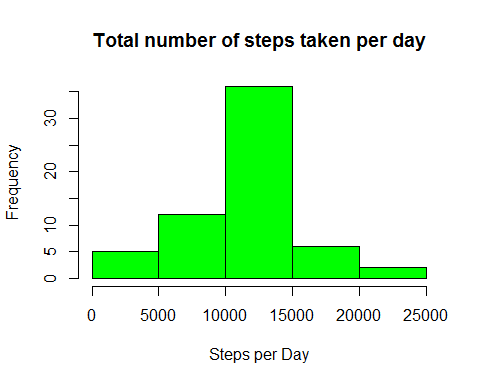
## [1] "835"

## Imputing missing values

missing <- nrow(dat)-nrow(tidydata)  
missing

## [1] 2304

isna <- is.na(dat)  
dat2 <- dat  
dat2$steps[which(isna)] <- avgStep[match(dat$interval[which(isna)], dat$interval)]  
  
steps2 <- aggregate(steps ~ date, dat2, sum)  
hist(steps2$steps, main="Total number of steps taken per day", xlab="Steps per Day", col="green")



mean <- mean(steps2$steps)  
median<- median(steps2$steps)  
mean

## [1] 10766.19

median

## [1] 10766.19

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

dat2$weekdays <- factor(format(dat2$date, "%A"))  
levels(dat2$weekdays)

## [1] "Friday" "Monday" "Saturday" "Sunday" "Thursday" "Tuesday"   
## [7] "Wednesday"

levels(dat2$weekdays) <- list(weekday = c("Monday", "Tuesday",  
 "Wednesday",   
 "Thursday", "Friday"),  
 weekend = c("Saturday", "Sunday"))  
levels(dat2$weekdays)

## [1] "weekday" "weekend"

table(dat2$weekdays)

##   
## weekday weekend   
## 12960 4608

avgSteps <- aggregate(dat2$steps,   
 list(interval = as.numeric(as.character(dat2$interval)),   
 weekdays = dat2$weekdays),  
 FUN = "mean")  
names(avgSteps)[3] <- "meanOfSteps"  
library(lattice)  
xyplot(avgSteps$meanOfSteps ~ avgSteps$interval | avgSteps$weekdays,   
 layout = c(1, 2), type = "l",   
 xlab = "Interval", ylab = "Number of steps")

