Reproducible Research Assignment 1

TamÃ¡s Jakab

November 20, 2016

Downloading, unzipping and loading the csv file into RStudio:

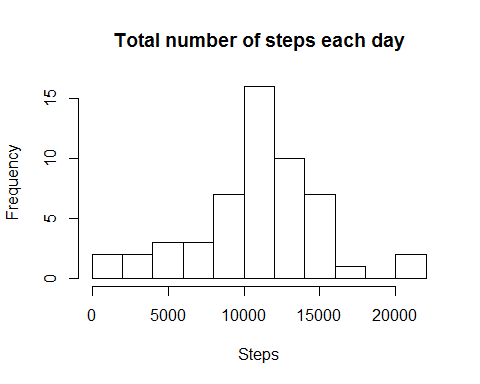
link <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"  
  
temp <- tempfile()  
download.file(link, temp)  
data <- read.csv(unz(temp, "activity.csv"),header = T, sep = ",")  
unlink(temp)

Calculating the total number of steps per day

perday <- aggregate(.~ date, data = data, sum)

Histogram of the total number of steps taken each day

hist(perday$steps, breaks = 10, main = "Total number of steps each day", xlab = "Steps")



Mean and median number of steps taken each day

mean(perday$steps)

## [1] 10766.19

median(perday$steps)

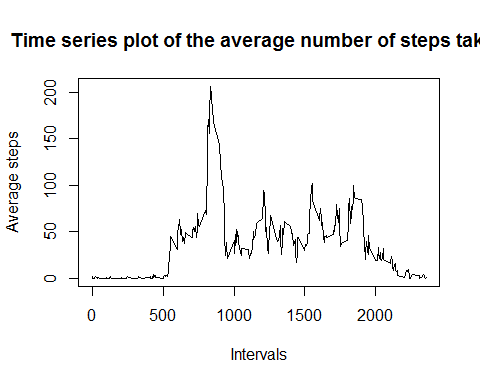
## [1] 10765

Calculating the average steps per interval

mean.inter <- aggregate(.~ interval, data = data, mean)

Time series plot of the average number of steps taken

plot(x=mean.inter$interval, y =mean.inter$steps, type = "l", main = "Time series plot of the average number of steps taken", xlab = "Intervals", ylab = "Average steps")



5-minute interval containing the maximum number of steps

mean.inter[mean.inter$steps == max(mean.inter$steps),1]

## [1] 835

Calculating the total number of rows with NAs

sum(is.na(data))

## [1] 2304

Imputing missing values

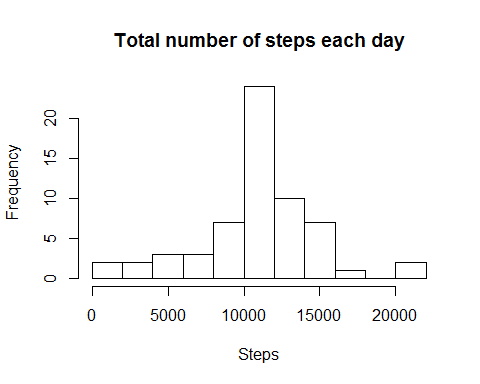
library(plyr)  
data2 <- data  
impute <- function(x, fun) {  
 missing <- is.na(x)  
 replace(x, missing, fun(x[!missing]))  
}  
data2 <-ddply(data2, ~ interval, transform, steps = impute(steps, mean))

Calculating the total number of steps per day(no NAs)

perday2 <- aggregate(.~ date, data = data2, sum)

Histogram of the total number of steps taken each day(no NAs)

hist(perday2$steps, breaks = 10, main = "Total number of steps each day", xlab = "Steps")



Mean and median number of steps taken each day(no NAs)

mean(perday2$steps)

## [1] 10766.19

median(perday2$steps)

## [1] 10766.19

Creating new factor variable

data2$w <- weekdays(as.Date(data2$date))  
data2$weekday <- ifelse(data2$w == "Saturday" | data2$w == "Sunday", data2$w <-"weekend",data2$w <-"weekday")  
data2$w <- NULL

Creating average steps based on weekdays/weekend and ploting them

par(mfrow = c(2,1))  
mean.interval.weekday <- aggregate(steps~ interval, data = subset(data2, weekday == "weekday"), mean)  
plot(x=mean.interval.weekday$interval, y =mean.interval.weekday$steps, type = "l", main = "Weekday", xlab = "Intervals", ylab = "Average steps")  
  
mean.interval.weekend <- aggregate(steps~ interval, data = subset(data2, weekday == "weekend"), mean)  
plot(x=mean.interval.weekend$interval, y =mean.interval.weekend$steps, type = "l", main = "Weekend", xlab = "Intervals", ylab = "Average steps")

