Reproducibe Research Proyect 1

Abner Aranda

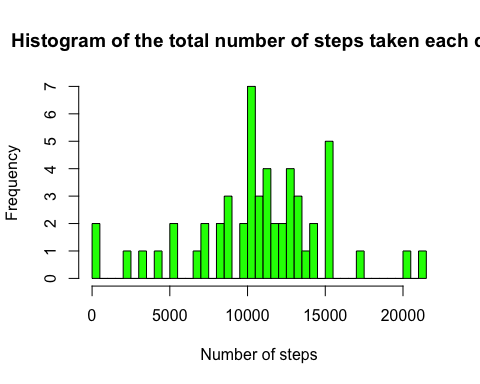
8/9/2021

Loading and preparing data

if (!file.exists("activity.csv") )  
 {  
 dlurl <- 'http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip'   
 download.file(dlurl,destfile='repdata%2Fdata%2Factivity.zip',mode='wb')   
 unzip('repdata%2Fdata%2Factivity.zip')  
 }  
activity <- read.csv("activity.csv")

What is the mean total number of steps taken per day? We start by creating a histogram of steps per day

Act\_without\_na <- subset(activity, !is.na(activity$steps))  
step\_per\_day <- aggregate(steps ~ date, Act\_without\_na, sum)  
hist(step\_per\_day $steps, breaks = 53, col = "green", xlab = "Number of steps", main = "Histogram of the total number of steps taken each day")

 Calculate mean and median of the total number of steps taken per day Mean

act\_mean <- mean(step\_per\_day$steps)  
print(act\_mean)

## [1] 10766.19

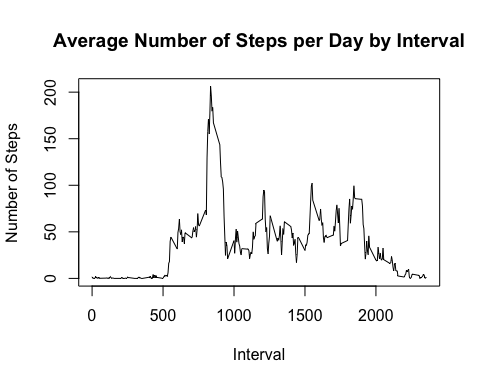
Median

act\_med <- median(step\_per\_day$steps)  
print(act\_med)

## [1] 10765

What is the average daily activity pattern? Plot of steps per interval

steps\_per\_interval <- aggregate(steps ~ interval, Act\_without\_na, mean)  
plot(steps\_per\_interval$interval, steps\_per\_interval$steps, type="l", xlab="Interval", ylab="Number of Steps", main="Average Number of Steps per Day by Interval")

 Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

max\_int <- steps\_per\_interval[which.max(steps\_per\_interval$steps),1]  
print(max\_int)

## [1] 835

How many steps does that interval had?

max\_int\_numb <- steps\_per\_interval[steps\_per\_interval$interval == max\_int,2]  
print(max\_int\_numb)

## [1] 206.1698

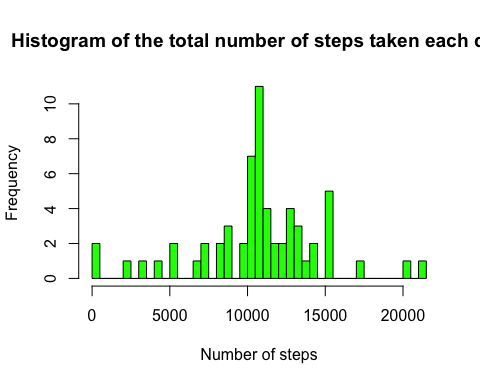
Imputing missing values Calculate and report the total number of missing values in the dataset (i.e. the total number of rows with NAs)

sum\_NA <- sum(!complete.cases(activity))  
sum\_NA

## [1] 2304

Create a new dataset that is equal to the original dataset but with the missing data filled in. Create a histogram

na\_index <- which(is.na(as.character(activity$steps)))  
complete\_act <- activity  
complete\_act[na\_index, ]$steps <- unlist(lapply(na\_index, FUN=function(na\_index){steps\_per\_interval[activity[na\_index, ]$interval== steps\_per\_interval$interval, ]$steps}))  
step\_per\_day\_complete <- aggregate(steps ~ date, data = complete\_act, sum)  
hist(step\_per\_day\_complete $steps, breaks = 53, col = "green", xlab = "Number of steps", main = "Histogram of the total number of steps taken each day")

 Mean

mean(step\_per\_day\_complete$steps)

## [1] 10766.19

Median

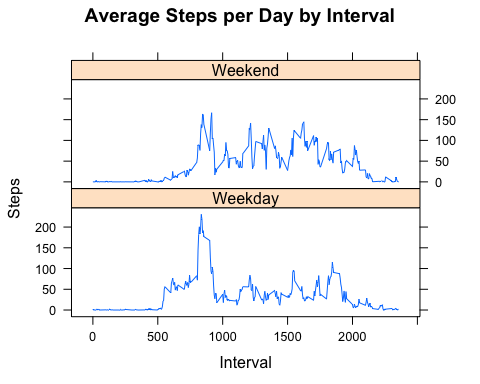
median(step\_per\_day\_complete$steps)

## [1] 10766.19

Both mean and median has little to no changecompared with the incomplete data

Are there differences in activity patterns between weekdays and weekends? Lets separate the data between weekend and weekdays

complete\_act$date <- as.Date(complete\_act$date, format = "%Y-%m-%d")  
weekdays <- c("Monday","Tuesday","Wednesday","Thursday","Friday")  
complete\_act $day\_week = as.factor(ifelse(is.element(weekdays(as.Date(complete\_act $date)),weekdays), "Weekday", "Weekend"))  
steps\_by\_day\_type <- aggregate(steps ~ interval + day\_week, complete\_act, mean)  
library(lattice)  
xyplot(steps\_by\_day\_type$steps ~ steps\_by\_day\_type$interval| steps\_by\_day\_type$day\_week, main="Average Steps per Day by Interval",xlab="Interval", ylab="Steps",layout=c(1,2), type="l")

 The weekends do have more activity, although weekdays have the biggest peak of activity, pressumible during the morning