Reproducible Research: Peer Assessment 1

lwrctan

July 23, 2018

## Loading and preprocessing the data

activity <- read.csv("activity.csv", header = TRUE, colClasses = c("numeric", "character", "numeric"), na = "NA")

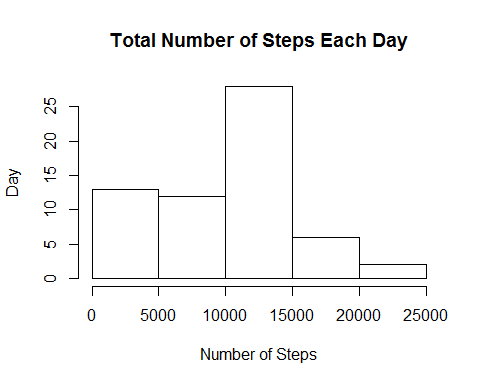
## What is mean total number of steps taken per day?

# sum up steps by day

StepDay <- tapply(activity$steps, activity$date, sum, na.rm=TRUE)

# plot histogram of total number of steps taken per day

hist(StepDay, main = "Total Number of Steps Each Day", xlab = "Number of Steps", ylab = "Day")



## What is the average daily activity pattern?

# Mean and median of number of steps taken each day

MeanDay <- mean(StepDay, na.rm = TRUE)  
MedianDay <- median(StepDay, na.rm = TRUE)  
MeanDay

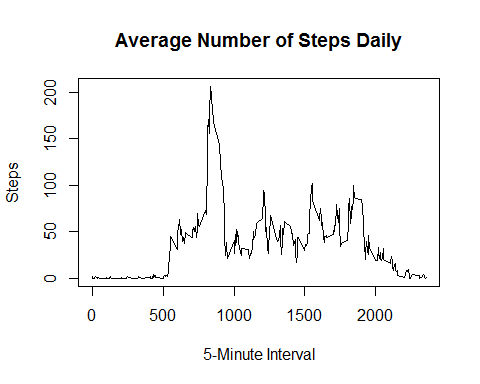
## [1] 9354.23

MedianDay

## [1] 10395

# Time series plot of the average number of steps taken

StepInterval <- tapply(activity$steps, activity$interval, mean, na.rm = TRUE)  
plot(as.numeric(names(StepInterval)),   
 StepInterval,   
 xlab = "5-Minute Interval",   
 ylab = "Steps",   
 main = "Average Number of Steps Daily",   
 type = "l")



# The 5-minute interval that, on average, contains the maximum number of steps

maxInterval <- names(sort(StepInterval, decreasing = TRUE)[1])  
maxSteps <- sort(StepInterval, decreasing = TRUE)[1]

## Imputing missing values

# Code to describe and show a strategy for imputing missing data

StepInterval <- tapply(activity$steps, activity$interval, mean, na.rm = TRUE)

# split activity data by interval

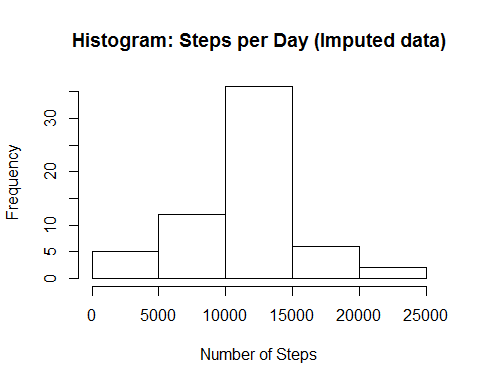
activity.split <- split(activity, activity$interval)

# fill in missing data for each interval

for(i in 1:length(activity.split)){  
 activity.split[[i]]$steps[is.na(activity.split[[i]]$steps)] <- StepInterval[i]  
}  
activity.imputed <- do.call("rbind", activity.split)  
activity.imputed <- activity.imputed[order(activity.imputed$date) ,]

# Histogram of the total number of steps taken each day after missing values are imputed

StepDay.imputed <- tapply(activity.imputed$steps, activity.imputed$date, sum)  
hist(StepDay.imputed, xlab = "Number of Steps", main = "Histogram: Steps per Day (Imputed data)")

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

# Panel plot comparing the average number of steps taken per 5-minute interval across weekday days and weekends

activity.imputed$day <- ifelse(weekdays(as.Date(activity.imputed$date)) == "Saturday" | weekdays(as.Date(activity.imputed$date)) == "Sunday", "weekend", "weekday")

# Calculate average steps per interval for weekends

StepInterval.weekend <- tapply(activity.imputed[activity.imputed$day == "weekend" ,]$steps, activity.imputed[activity.imputed$day == "weekend" ,]$interval, mean, na.rm = TRUE)

# Calculate average steps per interval for weekdays

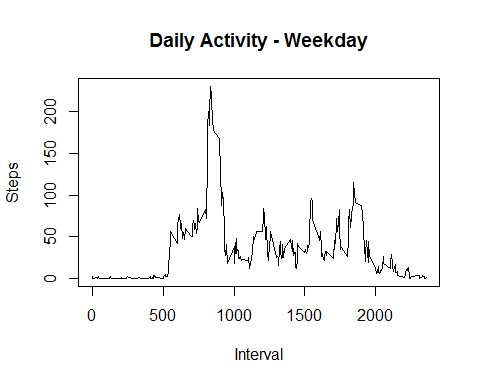
StepInterval.weekday <- tapply(activity.imputed[activity.imputed$day == "weekday" ,]$steps, activity.imputed[activity.imputed$day == "weekday" ,]$interval, mean, na.rm = TRUE)

# Set a 2 panel plot

par(mfrow=c(1,2))

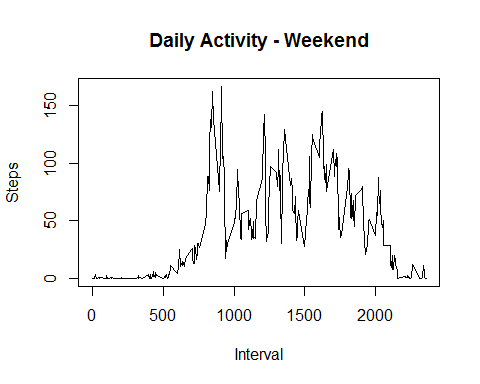
# Plot weekday activity

plot(as.numeric(names(StepInterval.weekday)),   
 StepInterval.weekday,   
 xlab = "Interval",   
 ylab = "Steps",   
 main = "Daily Activity - Weekday",   
 type = "l")



# Plot weekend activity

plot(as.numeric(names(StepInterval.weekend)),   
 StepInterval.weekend,   
 xlab = "Interval",   
 ylab = "Steps",   
 main = "Daily Activity - Weekend",   
 type = "l")



### end