Peer Assignment-1

Socrates

Saturday, February 14, 2015

### 1. Loading and preprocessing the data

#### Setting Working Directory

source\_url <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"  
target\_file <- "activity.zip"  
dest\_folder <- "C:\\courses\\5-Reproducible Research\\Project"  
data\_folder <- "C:\\courses\\5-Reproducible Research\\Project\\data"  
setwd(dest\_folder)

#### Downloading files

setInternet2(use = TRUE)  
download.file(source\_url, target\_file)

## Warning in download.file(source\_url, target\_file): downloaded length 53559  
## != reported length 53559

unzip(target\_file, exdir=data\_folder)

#### Loading Data into R working Enviornment

file <- list.files(data\_folder, full.names=TRUE)  
data <- read.csv(file, sep=",", header=TRUE, colClasses=c('numeric', 'character', 'numeric'))

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

#### Convert Date class from Character to Date

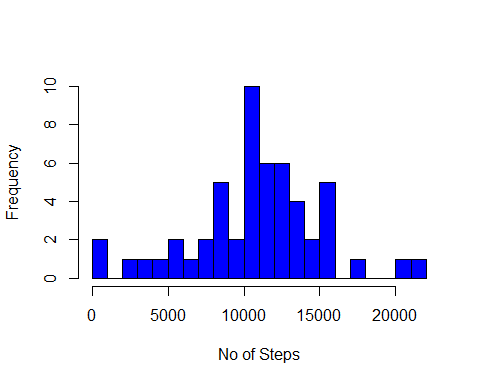
data$date <- as.Date(data$date, format="%Y-%m-%d")

#### Calculate the total number of steps taken per day

totalval <- aggregate(steps ~ date, data = data, sum)

#### Make a histogram of the total number of steps taken each day

hist(totalval[,2], xlab="No of Steps", main='', breaks = 30, col='blue')



#### Calculate Mean value

meanval <- round(mean(totalval[,2], na.rm=TRUE))  
meanval

## [1] 10766

#### Calculate Median value

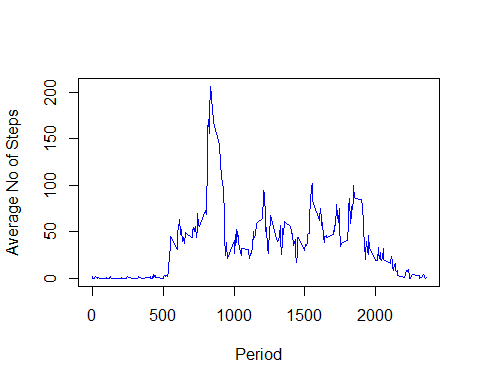
medianval <- round(median(totalval[,2], na.rm=TRUE))  
medianval

## [1] 10765

### 3. What is the average daily activity pattern?

#### Time series plot. x-axis: Average number of steps taken; y-axis: Averaged across all days

meaninterval <- aggregate(steps ~ interval, data=data, mean, na.rm=TRUE)  
plot(meaninterval$steps ~ meaninterval$interval, type = "l", xlab='Period', col='blue', ylab='Average No of Steps', main='')



#### Which 5-minute interval containing maximum number of steps?

meaninterval[(meaninterval$steps==max(meaninterval$steps)),1]

## [1] 835

### 4. Imputing missing values

#### Calculate and report the total number of missing values in the dataset

nrow(data[data$steps == 'NA',])

## [1] 2304

#### Strategy for filling in all of the missing values: The mean for that 5-minute interval.

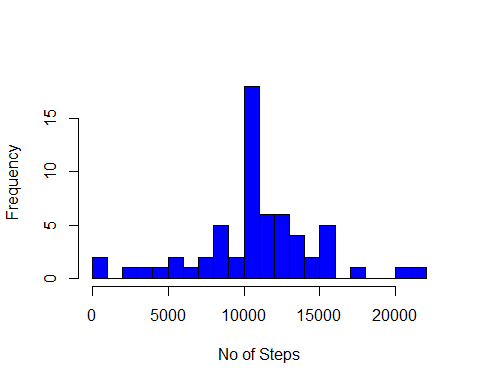
meaninterval <- aggregate(steps ~ interval, data=data, mean, na.rm=TRUE)  
meaninterval$steps <- round(meaninterval$steps)

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

for (i in 1:nrow(meaninterval)) {  
  
 # returns 0, 5, 10, 15, ...  
 intr.val <- meaninterval$interval[i]  
 upd.df <- length(data[which(data$interval == intr.val & is.na(data$steps)),]$steps)  
 if (upd.df > 0) {  
 data[which(data$interval == intr.val & is.na(data$steps)),]$steps <- meaninterval$steps[meaninterval$interval == intr.val]  
 }  
}

#### Make a histogram of the total number of steps taken each day on new dataset

totalval <- aggregate(steps ~ date, data = data, sum)  
hist(totalval[,2], xlab="No of Steps", main='', col='blue', breaks = 30)



#### Calculate and report Mean total number of steps taken per day.

meanval <- round(mean(totalval[,2], na.rm=TRUE))  
meanval

## [1] 10766

#### Calculate and report Median total number of steps taken per day.

medianval <- round(median(totalval[,2], na.rm=TRUE))  
medianval

## [1] 10762

#### Do these values differ from the estimates from the first part of the assignment?

YES, They differ.

#### What is the impact of imputing missing data on the estimates of the total daily number of steps?

After imputing missing data, the Total Sum got increased from 5.7060810^{5} to 6.5670410^{5}. However the mean and median values remain same. This is due to the fact that the no of days in original dataset increased from 53 to 61 after populating values for "NA"

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

#### Create a new factor variable in the dataset with two levels - "weekday" and "weekend"

data$dayname <- weekdays(data$date)  
data$daytype <- ifelse(data$dayname %in% c("Saturday", "Sunday"), "weekend", "weekday")

#### Panel plot containing a time series plot

library(lattice)   
meansteps <- aggregate(data$steps, by = list(data$interval, data$daytype), mean)  
names(meansteps) <- c("interval", "daytype", "steps")  
xyplot(steps ~ interval | daytype, meansteps, type = "l", layout = c(1, 2), xlab = "Interval", ylab = "Number of steps")

