Reproducible Research Peer Grade Assessment Project 1

Sandip Mondal

Data, i.e. activity.zip is downloadef from [this GitHub repository](http://github.com/rdpeng/RepData_PeerAssessment1) on March 19, 2018 and kept in R working directory

## Loading and preprocessing the data

### Unzipping the data to get “activity.csv”

unzip("activity.zip")

### Reading “activity.csv” and displaying first few sample rows

Acts <- read.csv("activity.csv")  
head(Acts)

## 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?

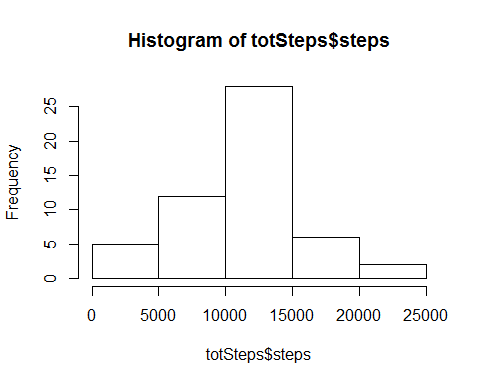
### Calculating total number of steps taken per day

totSteps <- aggregate(steps ~ date, data = Acts, sum, na.rm = TRUE)  
print(totSteps)

## date steps  
## 1 2012-10-02 126  
## 2 2012-10-03 11352  
## 3 2012-10-04 12116  
## 4 2012-10-05 13294  
## 5 2012-10-06 15420  
## 6 2012-10-07 11015  
## 7 2012-10-09 12811  
## 8 2012-10-10 9900  
## 9 2012-10-11 10304  
## 10 2012-10-12 17382  
## 11 2012-10-13 12426  
## 12 2012-10-14 15098  
## 13 2012-10-15 10139  
## 14 2012-10-16 15084  
## 15 2012-10-17 13452  
## 16 2012-10-18 10056  
## 17 2012-10-19 11829  
## 18 2012-10-20 10395  
## 19 2012-10-21 8821  
## 20 2012-10-22 13460  
## 21 2012-10-23 8918  
## 22 2012-10-24 8355  
## 23 2012-10-25 2492  
## 24 2012-10-26 6778  
## 25 2012-10-27 10119  
## 26 2012-10-28 11458  
## 27 2012-10-29 5018  
## 28 2012-10-30 9819  
## 29 2012-10-31 15414  
## 30 2012-11-02 10600  
## 31 2012-11-03 10571  
## 32 2012-11-05 10439  
## 33 2012-11-06 8334  
## 34 2012-11-07 12883  
## 35 2012-11-08 3219  
## 36 2012-11-11 12608  
## 37 2012-11-12 10765  
## 38 2012-11-13 7336  
## 39 2012-11-15 41  
## 40 2012-11-16 5441  
## 41 2012-11-17 14339  
## 42 2012-11-18 15110  
## 43 2012-11-19 8841  
## 44 2012-11-20 4472  
## 45 2012-11-21 12787  
## 46 2012-11-22 20427  
## 47 2012-11-23 21194  
## 48 2012-11-24 14478  
## 49 2012-11-25 11834  
## 50 2012-11-26 11162  
## 51 2012-11-27 13646  
## 52 2012-11-28 10183  
## 53 2012-11-29 7047

### Creating histogram of the total number of steps taken each day

hist(totSteps$steps)



### Calculating mean of the total number of steps taken per day

mean(totSteps$steps)

## [1] 10766.19

### Calculating median of the total number of steps taken per day

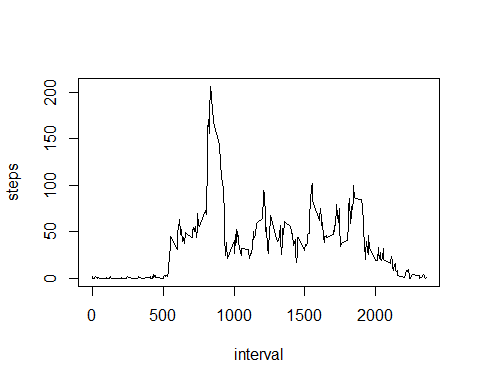
median(totSteps$steps)

## [1] 10765

## What is the average daily activity pattern?

### Makeing a time series plot of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)

intv\_stpAvg <- aggregate(steps ~ interval, data = Acts, mean, na.rm=TRUE)  
plot(steps ~ interval, data = intv\_stpAvg, type = "l")



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

intv\_stpAvg[which.max(intv\_stpAvg$steps), ]$interval

## [1] 835

As per dataset being considered here 835 interval contains the maximum number of steps average across all the days.

## Imputing missing values

### Calculating and reporting the total number of missing values in the dataset (i.e. the total number of rows with NAs)

sum(is.na(Acts))

## [1] 2304

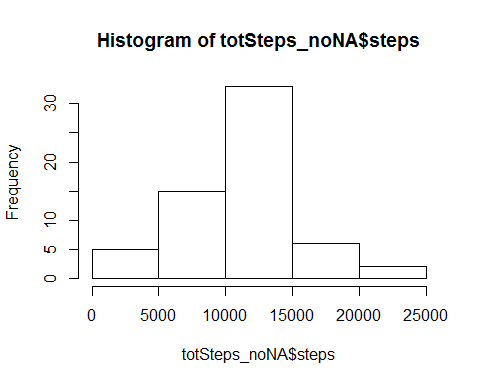
As per dataset being considered total number rows with missing value is 2304.

### Missing values are being replaced by mean values for that days

acts\_noNA <- Acts  
  
acts\_noNA$day\_name <- weekdays(as.POSIXlt(Acts$date), abbreviate = FALSE)  
  
for (i in 1:nrow(acts\_noNA)) {  
 if(is.na(acts\_noNA[i, ]$steps)) {  
 acts\_noNA[i, ]$steps <-mean(acts\_noNA[acts\_noNA$day\_name == acts\_noNA[i, ]$day\_name,]$steps, na.rm = TRUE)  
 }  
}

### Creating histogram of the total number of steps taken each day

totSteps\_noNA <- aggregate(steps ~ date, data = acts\_noNA, sum)  
hist(totSteps\_noNA$steps)



### Calculating mean of the total number of steps taken per day

mean(totSteps\_noNA$steps)

## [1] 10821.21

### Calculating median of the total number of steps taken per day

median(totSteps\_noNA$steps)

## [1] 11015

Imputing missing data on the estimates of the total daily number of steps is **making the result different**.

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

### Creating a new factor variable in the dataset with two levels - “weekday” and “weekend” indicating whether a given date is a weekday or weekend day

acts\_noNA$week\_or\_weekend <- ifelse(acts\_noNA$day\_name %in% c("Saturday", "Sunday"), "weekend", "weekday")

### Making a panel plot containing a time series plot (i.e. type = “l”) of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis).

library(lattice)  
acts\_noNA <- aggregate(steps ~ interval + week\_or\_weekend, acts\_noNA, mean)  
xyplot(steps ~ interval | factor(week\_or\_weekend), data = acts\_noNA, aspect = 1/2,   
 type = "l")

