

# Reproducible Research: Peer Assessment 1

*Alonggot Chawalitanont*

*June 26, 2016*

This report is a the first part of 2 peer assessment projects in Coursera's Reproducible Research course by Johns Hopkins University. The assignment makes use of data from a personal activity monitoring device. This device collects data at 5 minute intervals through out the day. The data consists of two months of data from an anonymous individual collected during the months of October and November, 2012 and include the number of steps taken in 5 minute intervals each day.

The aim of this report is to do the analysis on the dataset in order to answer the following questions:

- 1) What is mean total number of steps taken per day?
- 2) What is the average daily activity pattern?
- 3) Are there differences in activity patterns between weekdays and weekends?

to illustrate the visualize information in a literated format using R markdown, and also to make a reproducible research by providing a step-by-step analysis and codes involved in doing the analysis.

## Loading and Prepossessing the Data

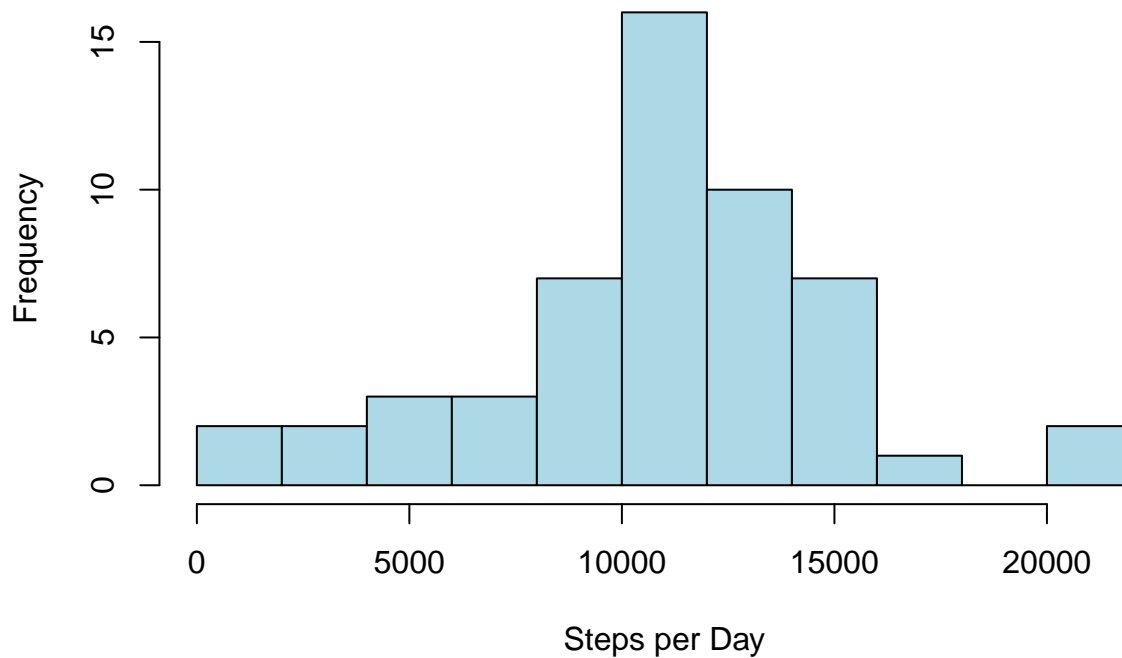
```
# Set working directory and read the data
setwd("C:/Users/HP/Desktop/RepData_PeerAssessment1")
options(scipen = 1, digits = 2)

DT <- read.csv("activity.csv", header = TRUE, sep = ",")
# Format the date and check the number observations
DT$date <- as.Date(DT$date,format="%Y-%m-%d")
obser <- nrow(DT)
```

- the total number of observations is 17568

Make histogram and find the mean and median of the total number of steps taken per day.

```
stepsD <- aggregate(DT$steps ~ DT$date, FUN = sum)
colnames(stepsD) <- c("date", "steps")
par(mfrow=c(1,1))
hist(stepsD$steps, main="", xlab="Steps per Day", col="light blue", breaks=15)
```

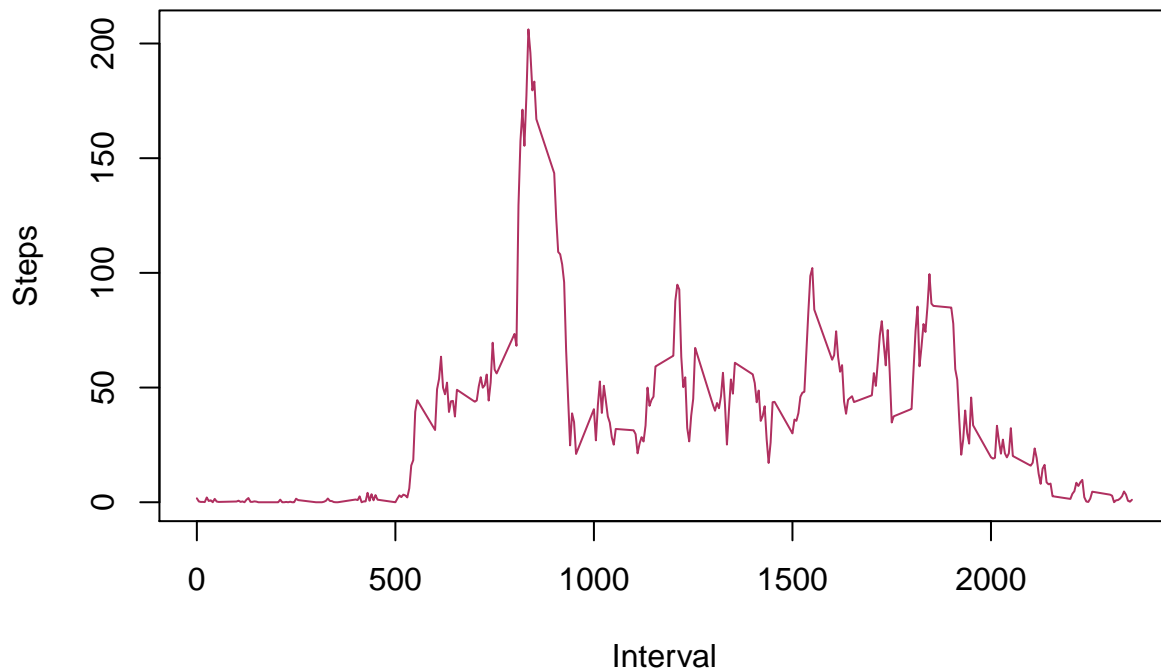


```
meanS <- mean(stepsD$steps)
medianS <- median(stepsD$steps)
```

- The *mean* total number of steps taken per day is 10766.19.
- The *median* total number of steps taken per day is = 10765

Illustrate the average daily activity patterns

```
itvSteps <- aggregate(steps~ interval, data=DT, FUN = mean)
plot(itvSteps, type="l", xlab="Interval", ylab="Steps", col="maroon")
```



```
maxItv <- itvSteps$interval[which.max(itvSteps$steps)]
maxSteps <- itvSteps$step[itvSteps$interval == maxItv]
```

- The 835th 5-minute interval contains the maximum number of steps, in which it has on average 206.17 steps.

## Imputing missing values

```
totNA <- sum(is.na(DT$steps))
```

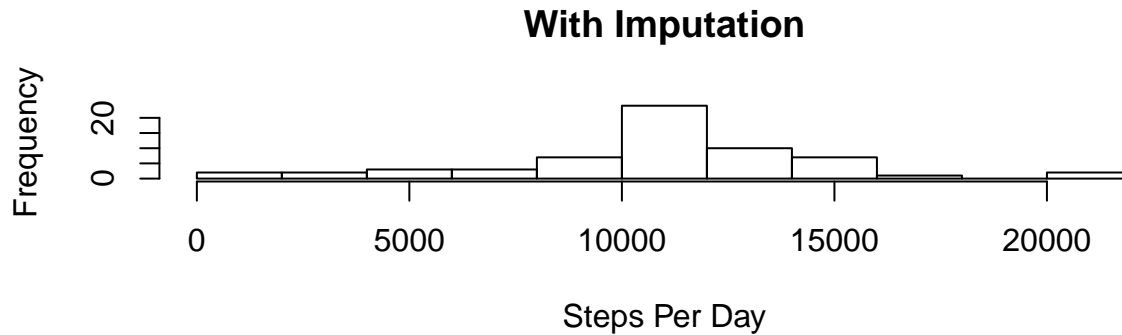
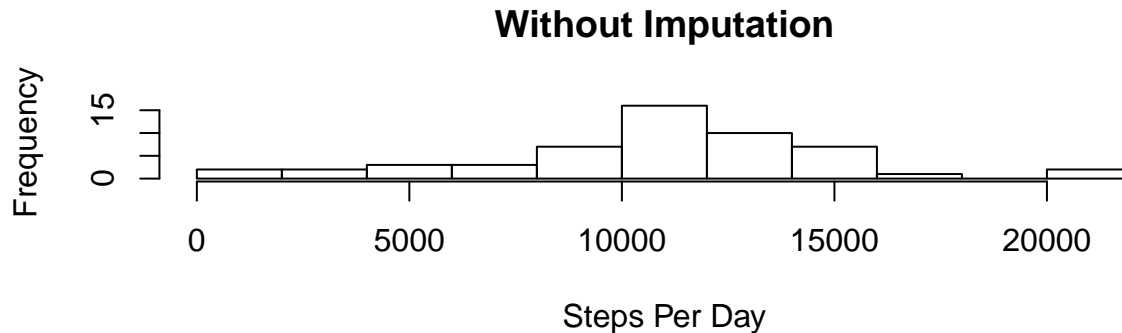
- There are 2304 missing values in the dataset.

```
fillDT <- DT

for (i in 1:nrow(fillDT)){
  if (is.na(fillDT[i, ]$steps)){
    fillDT[i, ]$steps <- itvSteps[itvSteps$interval == fillDT[i, ]$interval, ]$steps
  }
}

compDT <- aggregate(steps ~ date, data=fillDT, FUN=sum)
par(mfrow=c(2,1))
```

```
hist(stepsD$steps,breaks=10,main="Without Imputation", xlab="Steps Per Day")
hist(compDT$steps, breaks=10, main="With Imputation", xlab="Steps Per Day")
```



```
meanF <- mean(compDT$steps)
medianF <- median(compDT$steps)
```

- The *mean* total number of steps taken per day (with imputation) is 10766.19.
- The *median* total number of steps taken per day is (with imputation) = 10765.

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

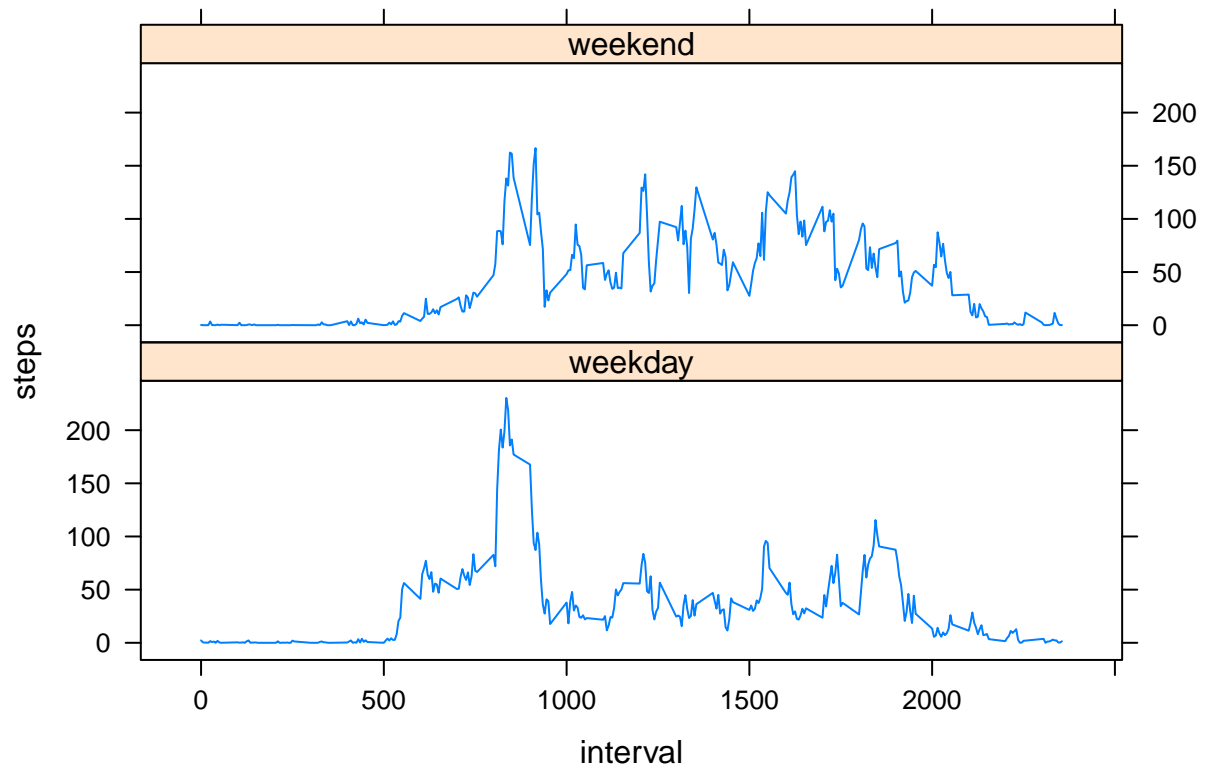
```
fillDT$day <- weekdays(as.Date(fillDT$date))

for (i in 1:nrow(fillDT)){
  if (fillDT[i,]$day %in% c("Saturday", "Sunday")){
    fillDT[i,]$day <- "weekend"
  }
  else{
    fillDT[i,]$day <- "weekday"
  }
}

fillDT$day <- as.factor(fillDT$day)
```

```
itvSteps2 <- aggregate(steps ~ interval + day, data = fillDT, FUN = mean)

library(lattice)
xyplot(steps ~ interval | day, data = itvSteps2, type = 'l', layout = c(1, 2))
```



- It can be seen that the activities are generally higher on Weekend than on Weekday. Although, the weekday activities are generally higher before afternoon while weekend activities are higher later afternoon.