Reproducible Research First Assignment

Khaled Alzafari

September 16, 2015

load the Data:

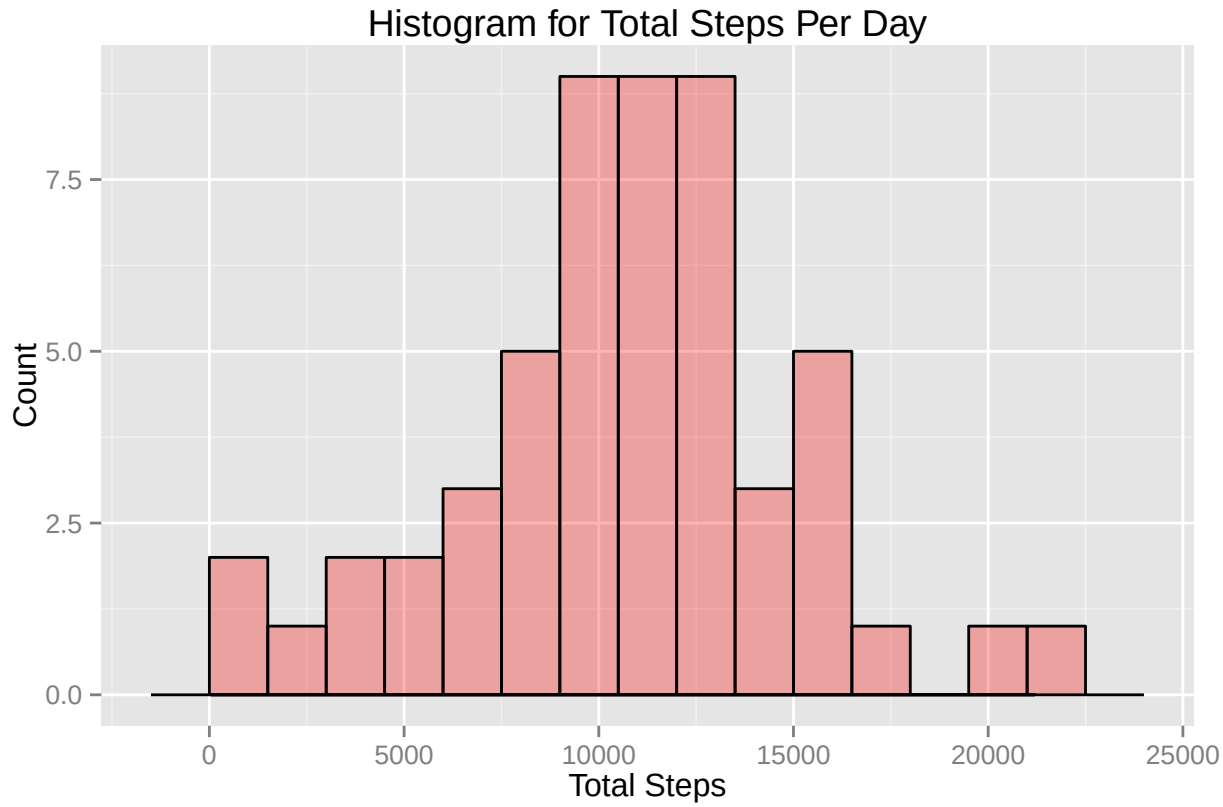
```
echo = TRUE
mydata= read.csv(file="activity.csv", header=TRUE, sep=",")
```

Histogram for the total number of steps per day:

```
echo = TRUE
totalSteps <- aggregate(steps ~ date, data = mydata, sum, na.rm = TRUE)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```

```
ggplot(data=totalSteps, aes(totalSteps$steps)) + geom_histogram( col="black", fill="red", alpha = 0.3, l
```



the mean and median total number of steps taken per day:

```
echo = TRUE
Mean_steps = mean(totalSteps$steps,  na.rm = TRUE)
Median_steps = median(totalSteps$steps,  na.rm = TRUE)
Mean_steps
```

```
## [1] 10766.19
```

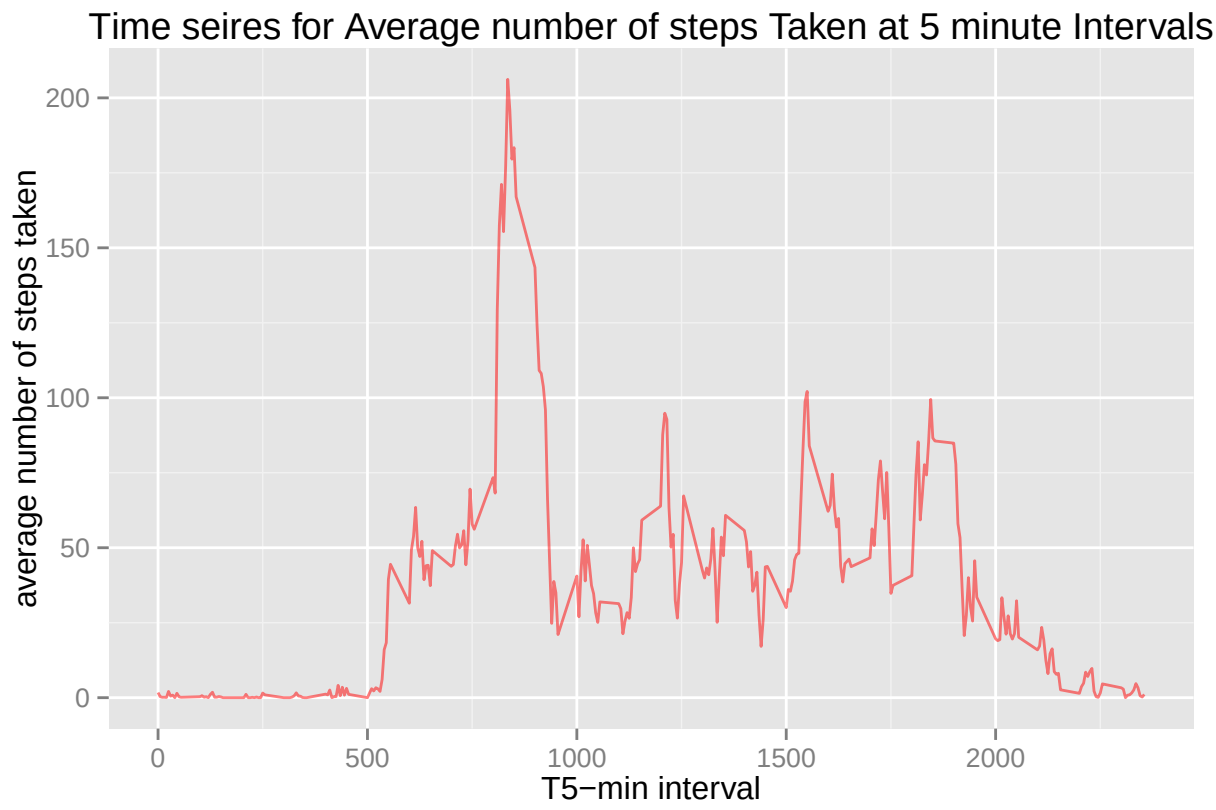
```
Median_steps
```

```
## [1] 10765
```

The average daily activity pattern:

```
echo = TRUE
meanSteps <- aggregate(steps ~ interval, data = mydata, mean, na.rm = TRUE)

dp= ggplot(meanSteps, aes(meanSteps$interval, meanSteps$steps))+ geom_line(col="red", fill="red", alpha
dp + labs(title="Time seires for Average number of steps Taken at 5 minute Intervals") + labs(x="T5-min
```



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

```
echo = TRUE
maxinterval = meanSteps[which.max(meanSteps$steps), ]
maxinterval
```

```
##     interval    steps
## 104      835 206.1698
```

```
### It is the interval number 835  that has the value of 206.2
```

Imputing missing values

Number of Missing Value

```
echo = TRUE
missing_rows <- sum(!complete.cases(mydata))
missing_rows
```

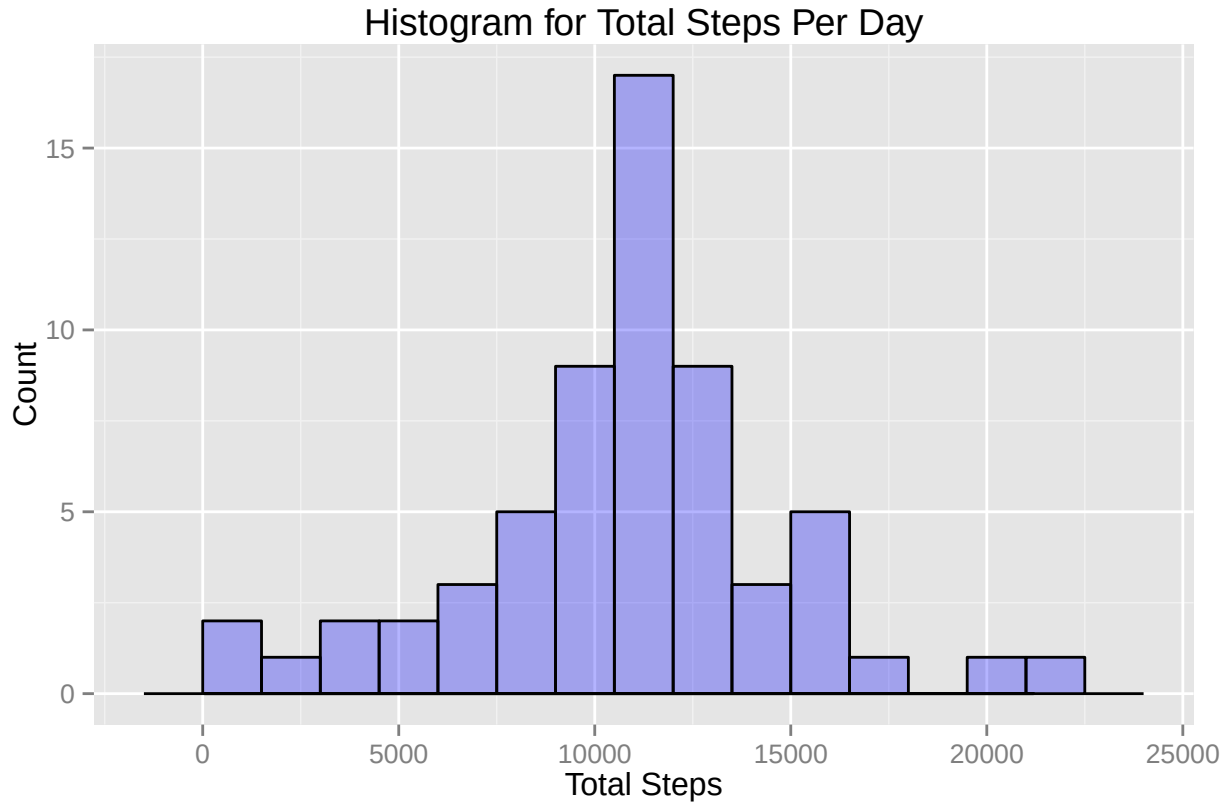
```
## [1] 2304
```

Devise a strategy for filling in all of the missing values in the dataset the new stratgy depend on replace NA with the mean of the "steps" variable

```
echo = TRUE
newdataset= mydata
newdataset$steps[which(is.na(newdataset$steps))] = mean(newdataset$steps, na.rm = T)
```

histogram for the new Data Set:

```
echo = TRUE
newtotalSteps <- aggregate(steps ~ date, data = newdataset, sum, na.rm = TRUE)
library(ggplot2)
ggplot(data=newtotalSteps, aes(newtotalSteps$steps)) + geom_histogram( col="black", fill="blue", alpha =
```



The mean and median total number of steps taken per day for the new data set:

```
echo = TRUE
newMean_steps = mean(newtotalSteps$steps)
newMedian_steps = median(newtotalSteps$steps)
newMean_steps
```

```
## [1] 10766.19
```

```
newMedian_steps
```

```
## [1] 10766.19
```

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

The mean for both data (before and after removing missing Data NA) while our strategy was based on replacing NA with the mean of the same variable.whereas, the there is a small difference for the median before removeing NA and after.

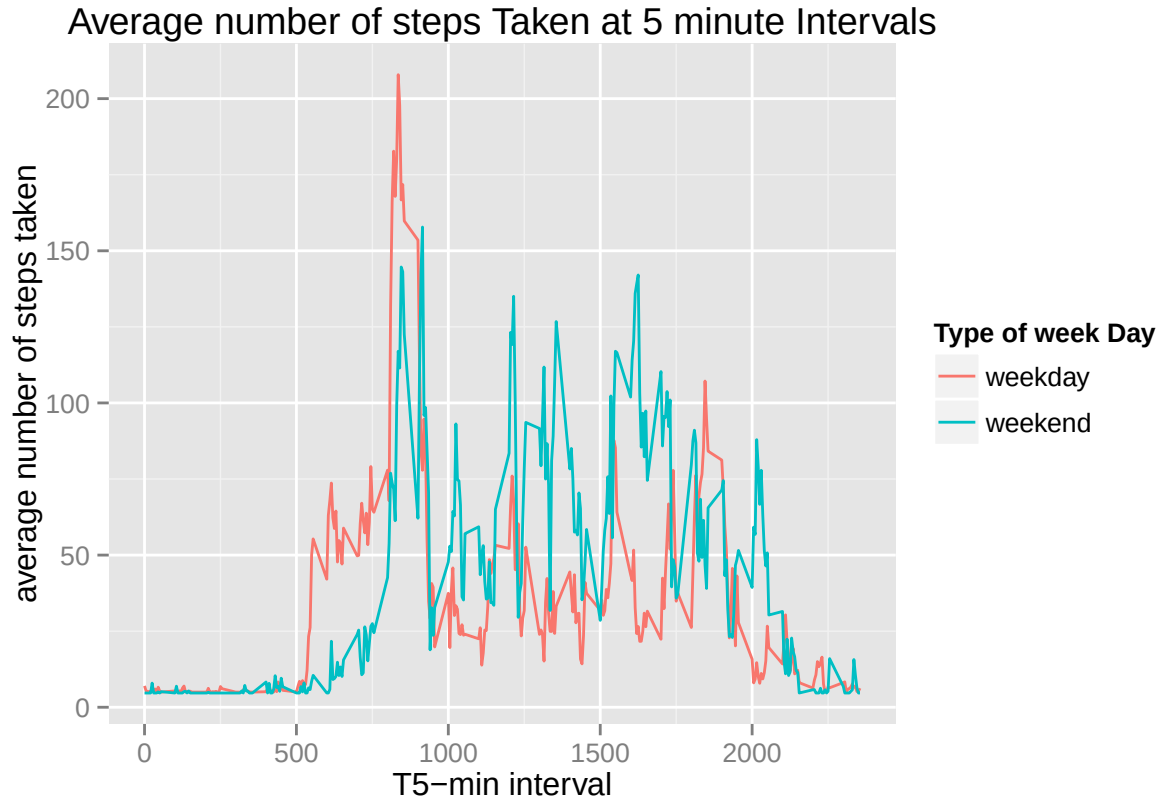
Are there differences in activity patterns between weekdays and weekends?

```
echo = TRUE
newdataset$days= weekdays(as.Date(newdataset$date))
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.2.1
```

```
newdataset$weeks <- revalue(newdataset$days, c("Saturday"="weekend", "Sunday"="weekend", "Monday"="weeke
newmeanSteps <- aggregate(steps ~ interval + weeks, data = newdataset, mean, na.rm = TRUE)

library(ggplot2)
X1 = ggplot (newmeanSteps, aes(newmeanSteps$interval, newmeanSteps$steps, colour = newmeanSteps$weeks)
X1 + labs(title="Average number of steps Taken at 5 minute Intervals") + labs(x="T5-min interval", y="a
```



The difference is clear between the normal working days and weekend days...