Course Project 1

Oscar Chamberlain

February 27th, 2017

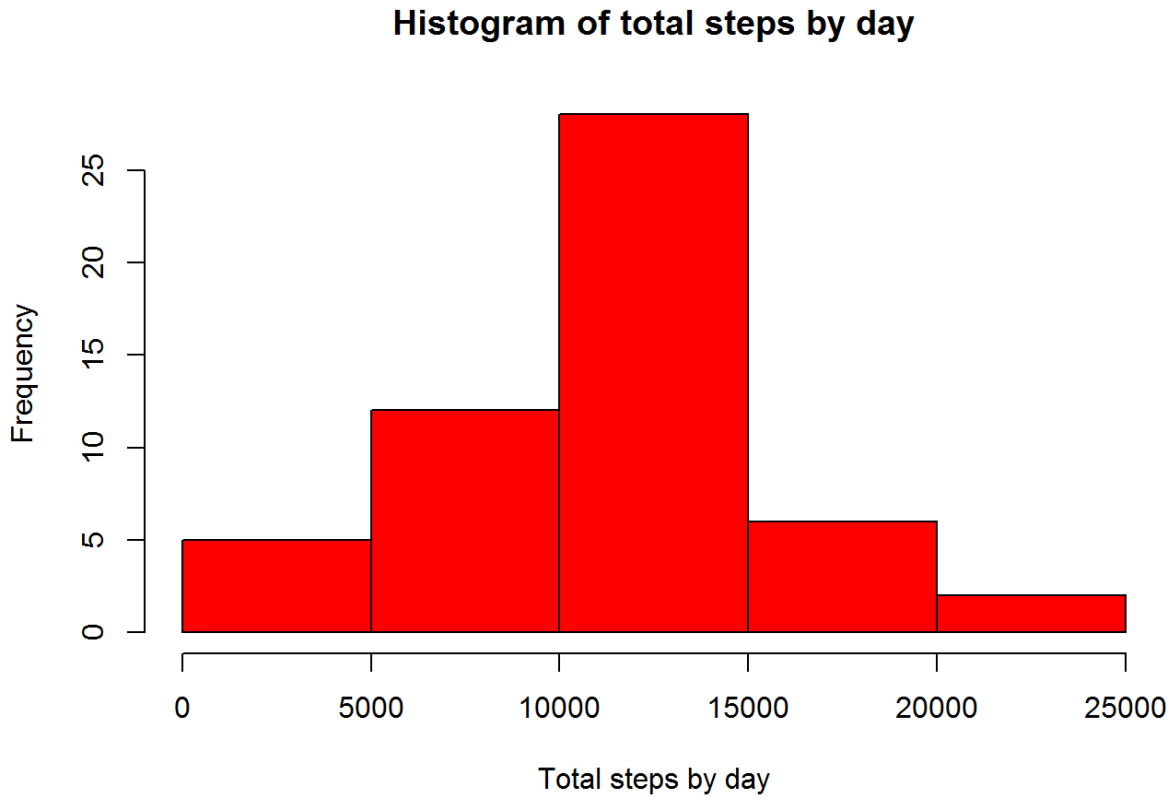
Introduction

It is now possible to collect a large amount of data about personal movement using activity monitoring devices such as a Fitbit, Nike Fuelband, or Jawbone Up. These type of devices are part of the “quantified self” movement - a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. But these data remain under-utilized both because the raw data are hard to obtain and there is a lack of statistical methods and software for processing and interpreting the data.

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.

Histogram with the total number of steps taken per day

```
actday<-group_by(dataset,date)
total_step_day<-summarise(actday, total_step=sum(steps))
hist(total_step_day$total_step, col="red", main="Histogram of total steps b
y day", xlab="Total steps by day")
```



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

```
mean(total_step_day$total_step, na.rm=TRUE)
```

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

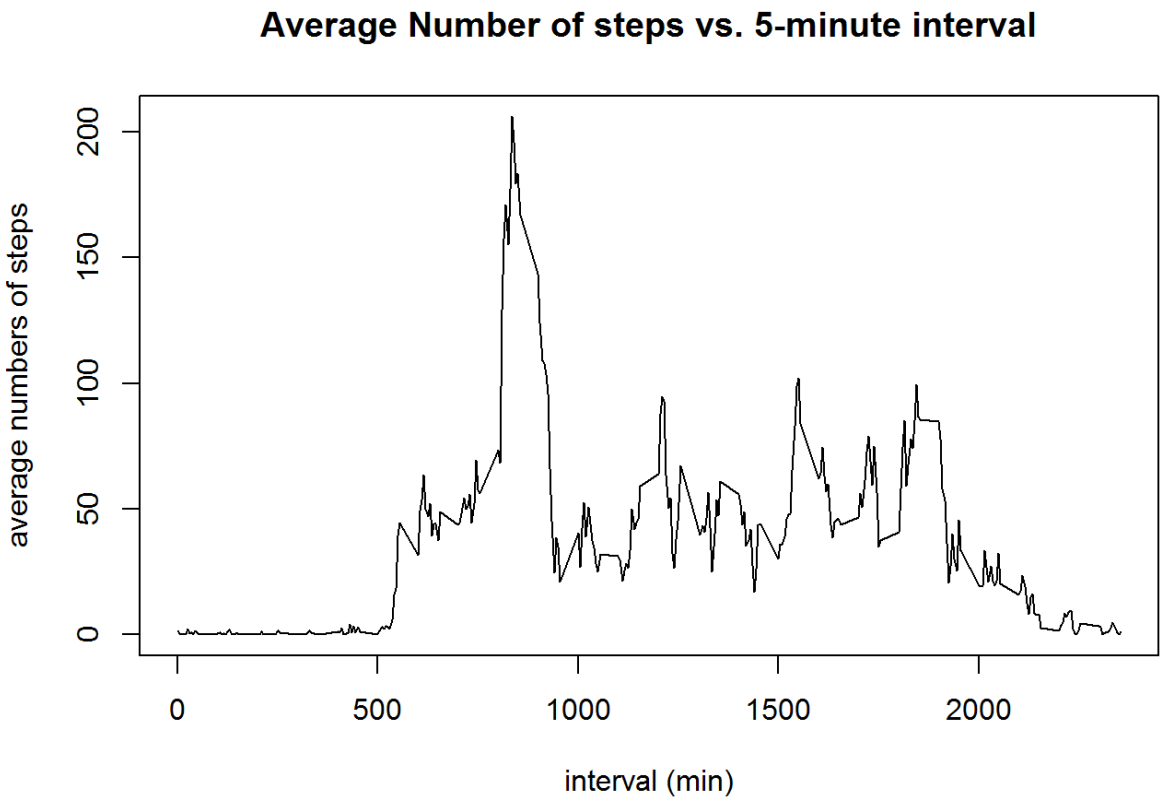
```
median(total_step_day$total_step, na.rm=TRUE)
```

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

Time series plot with the average number of

steps taken, averaged across all days (y-axis)

```
act_interval<-group_by(dataset, interval)
avg_step_interval<-summarise(act_interval, avg_step=mean(steps, na.rm = TRU
E))
plot(avg_step_interval$interval, avg_step_interval$avg_step, type= "l", main
="Average Number of steps vs. 5-minute interval", xlab = "interval (min)", y
lab="average numbers of steps")
```



The maximum number of steps within 5-minute interval, on average across all the days in the dataset

```
max_avg<-which.max(avg_step_interval$avg_step)
avg_step_interval$interval[max_avg]
```

```
## [1] 835
```

The total number of missing values in the dataset

```
N<-summary(dataset)
N[7]
```

```
## [1] "NA's   :2304  "
```

Filling in all of the missing values in the dataset.

```
install.packages("zoo", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/ocham/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)
```

```
## package 'zoo' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\ocham\AppData\Local\Temp\RtmpO4RTou\downloaded_packages
```

```
library(zoo)
```

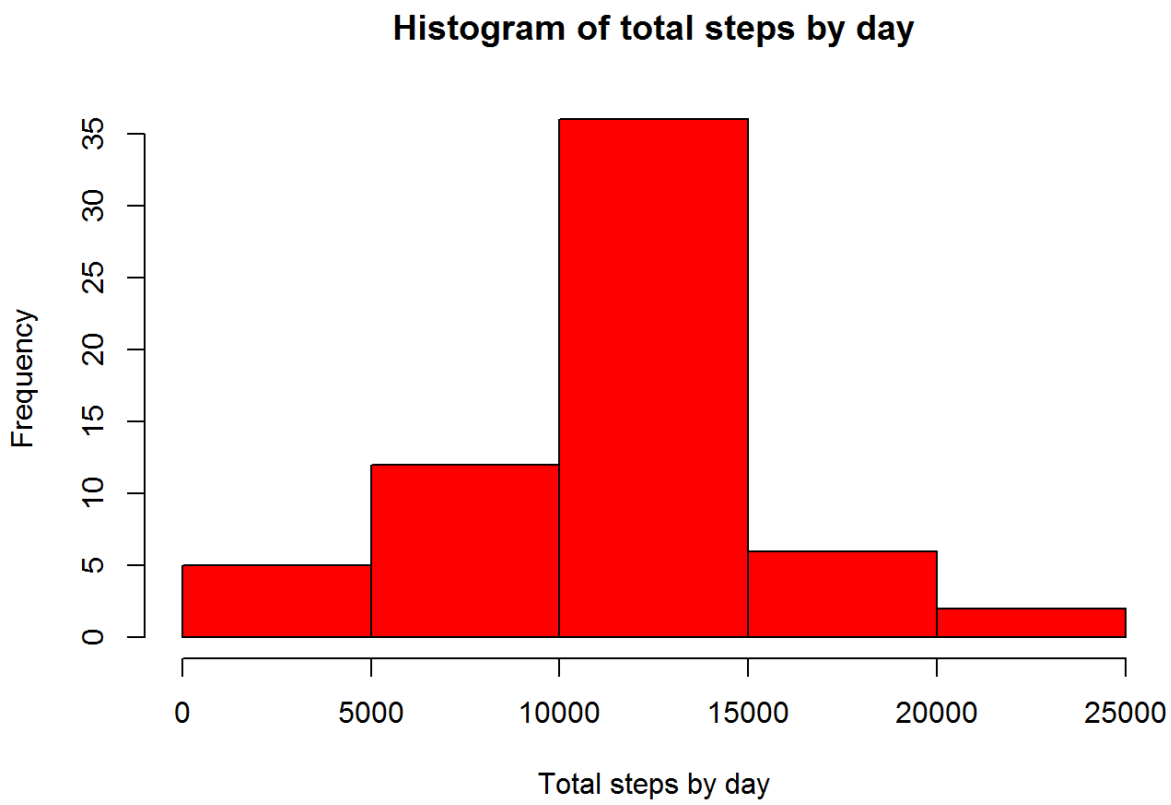
```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
new_dataset<-dataset
new_dataset$steps<-na.aggregate(new_dataset$steps)
```

Histogram of the total number of steps taken each day

```
new_actday<-group_by(new_dataset,date)
new_total_step_day<-summarise(new_actday, new_total_step=sum(steps))
hist(new_total_step_day$new_total_step, col="red", main="Histogram of total
steps by day", xlab="Total steps by day")
```



The mean and median total number of steps taken per day.

The median changed to the same value of the mean.

```
mean(new_total_step_day$new_total_step)
```

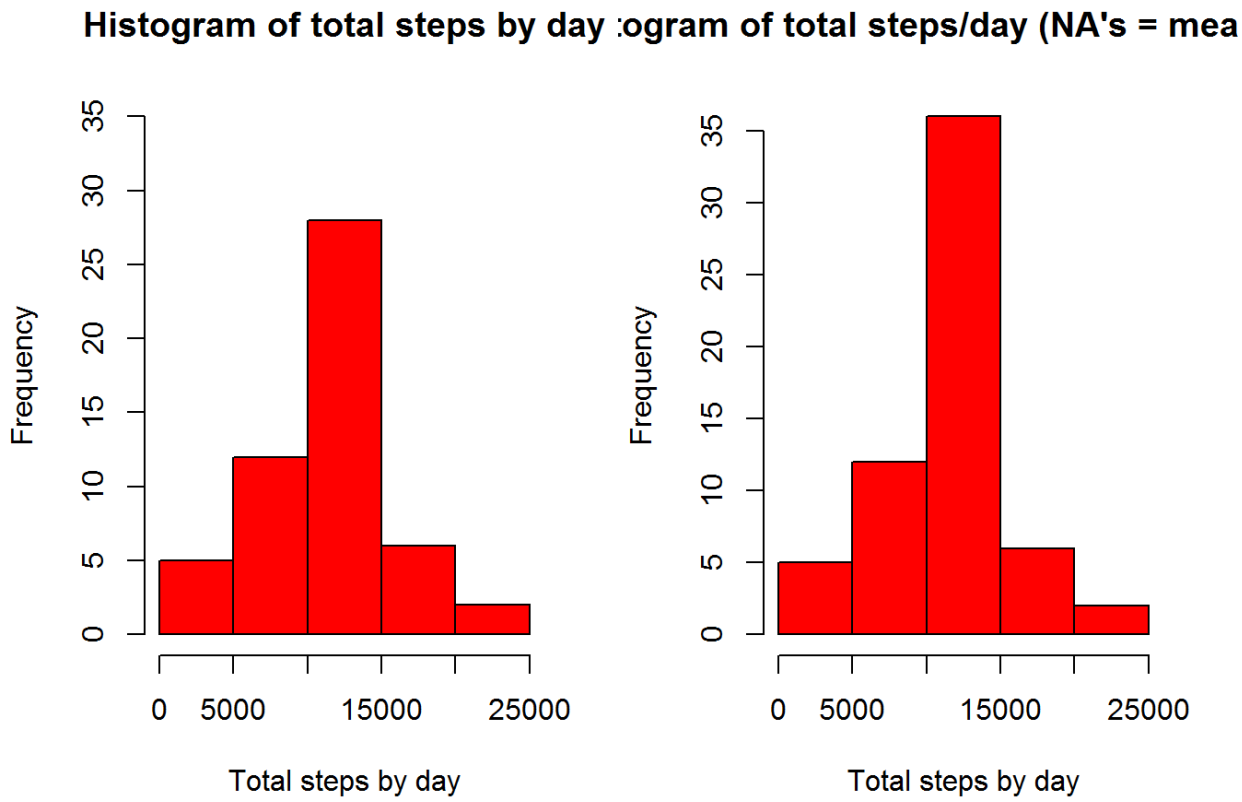
```
## [1] 10766.19
```

```
median(new_total_step_day$new_total_step)
```

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

Imputing missing data on the estimates of the total daily number of steps, will increase the frequency of media values

```
par(mfrow=c(1,2))
hist(total_step_day$total_step, col="red", main="Histogram of total steps b
y day", xlab="Total steps by day", ylim = c(0,35))
hist(new_total_step_day$new_total_step, col="red", main="Histogram of total
steps/day (NA's = mean value)", xlab="Total steps by day")
```



Are there differences in activity patterns between weekdays and weekends?

Yes the activity seems to increase later during the weekends. This can be observed in the plot below.

```
new_actday$weekday <-factor(ifelse(weekdays(new_actday$date) %in% c("sábad
o", "domingo"), "weekend", "weekday"))

install.packages("lattice", repos = "http://cran.us.r-project.org")
```

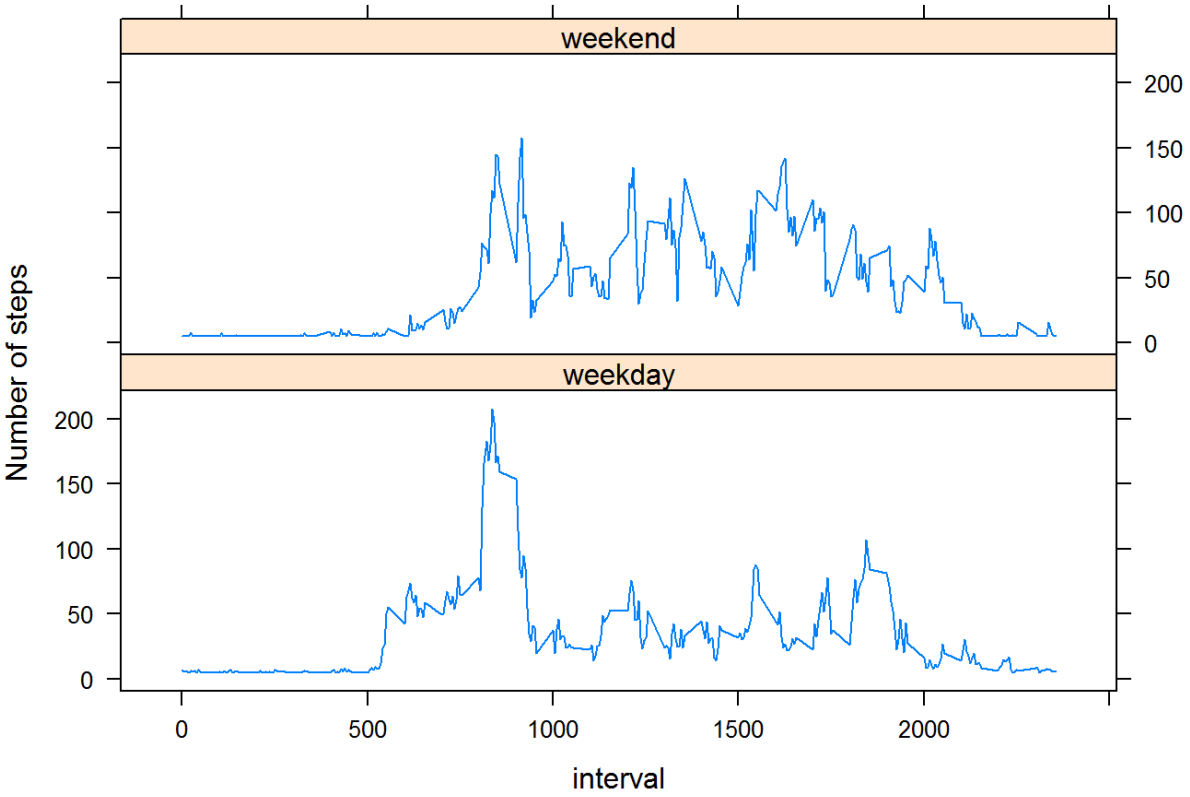
```
## Installing package into 'C:/Users/ocham/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)
```

```
## package 'lattice' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\ocham\AppData\Local\Temp\RtmpO4RTou\downloaded_packages
```

```
library(lattice)

w1<-subset(new_actday, weekday=="weekday")
w2<-subset(new_actday, weekday=="weekend")
w1_step_day<-group_by(w1, interval)
w2_step_day<-group_by(w2, interval)
w1_avg_step<-summarise(w1_step_day, avg_step=mean(steps))
w1_avg_step$weekday<-factor("weekday")
w2_avg_step<-summarise(w2_step_day, avg_step=mean(steps))
w2_avg_step$weekday<-factor("weekend")
ww<-rbind(w1_avg_step,w2_avg_step)

xyplot(avg_step ~interval | weekday, data = ww, layout = c(1,2), type="l" ,
ylab = ("Number of steps"))
```


weekend
weekday
Number of steps
interval
0
50
100
150
200
0
500
1000
1500
2000