

# Peer Assessment 1

Boudewijn Aasman

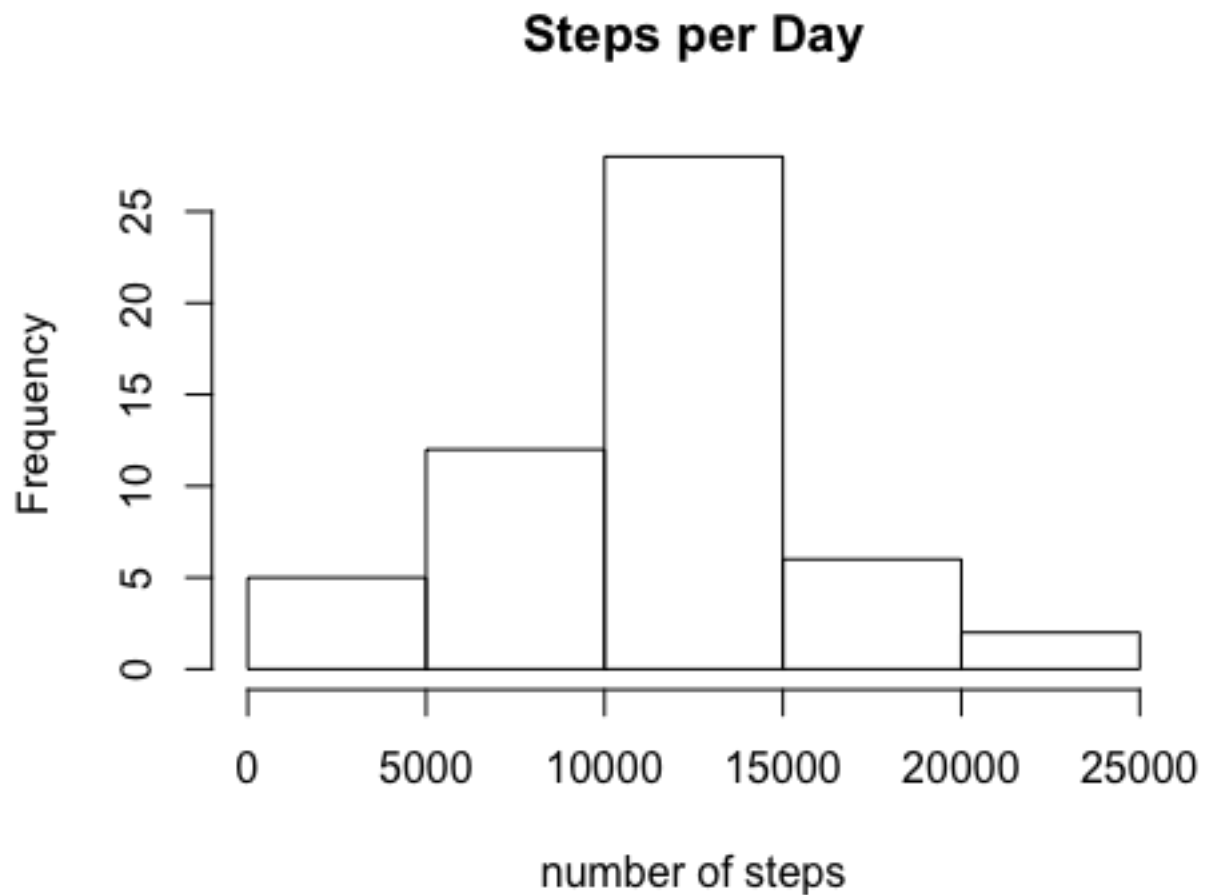
January 17, 2015

## Loading/Preprocessing data

```
rm(list = ls())  
library(ggplot2)  
setwd("~/Documents/repResearch")  
act = read.csv("activity.csv", header = TRUE)
```

## Mean number of steps

```
compAct = na.omit(act)  
t = ts(compAct$date)  
t = as.data.frame(t)  
compAct = cbind(compAct, t)  
sum = aggregate(steps ~ x, compAct, sum)  
hist(sum$steps, main = "Steps per Day", xlab = "number of steps")
```

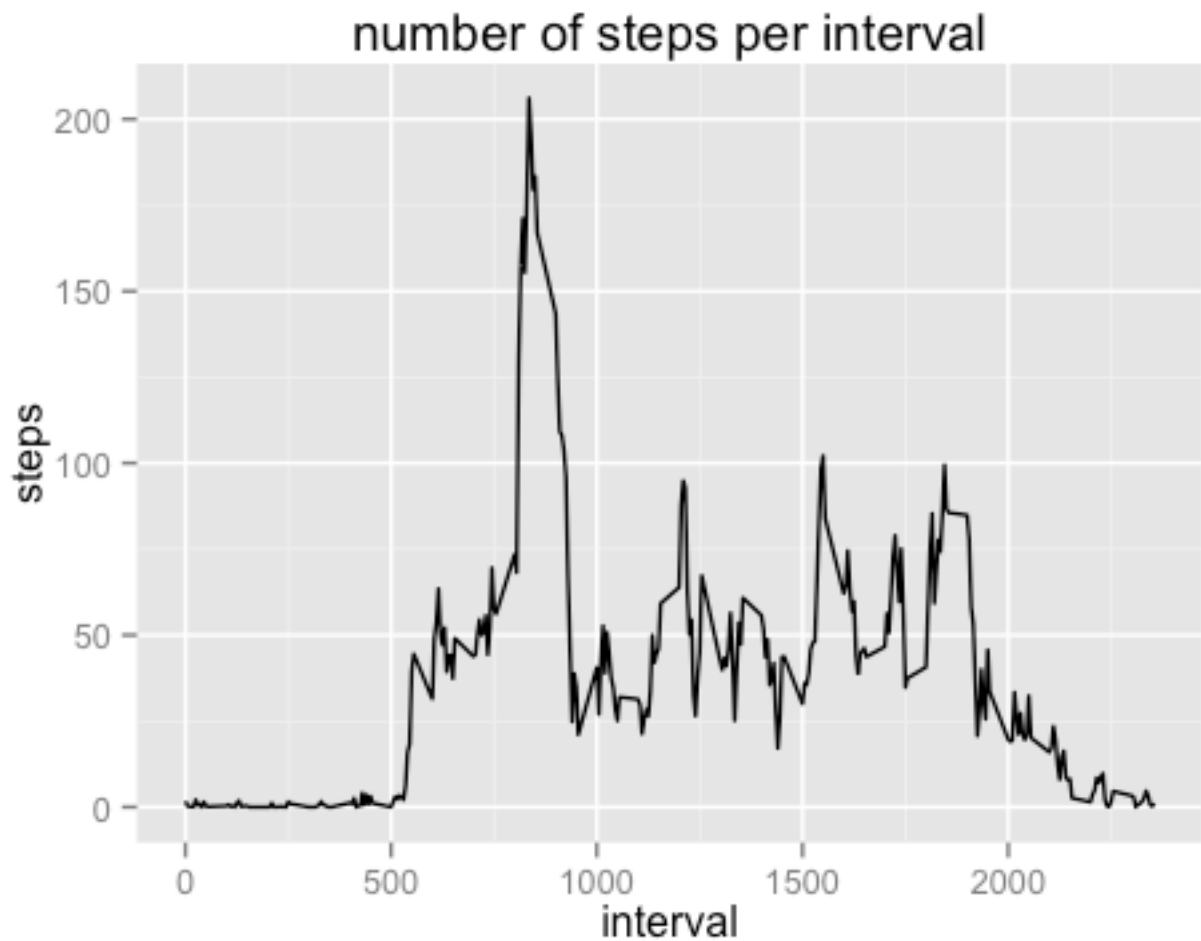


```
mean(sum$steps)
## [1] 10766.19
median(sum$steps)
## [1] 10765
```

- The mean number of steps per day comes out to 10766.19
- the median number of steps per day comes out to 10765

### Average Daily Activity Pattern

```
avg = aggregate(steps ~ interval, compAct, mean)
ggplot(avg, aes(x = interval, y = steps)) + geom_line() + ggtitle("number  
of steps per interval")
```



```
max = avg[order(-avg$steps),]  
head(max)
```

```
##      interval    steps  
## 104         835 206.1698  
## 105         840 195.9245  
## 107         850 183.3962  
## 106         845 179.5660  
## 103         830 177.3019  
## 101         820 171.1509
```

- As we can see, on average the most steps are taken in interval 835, with approximately 206 steps

### Missing Values

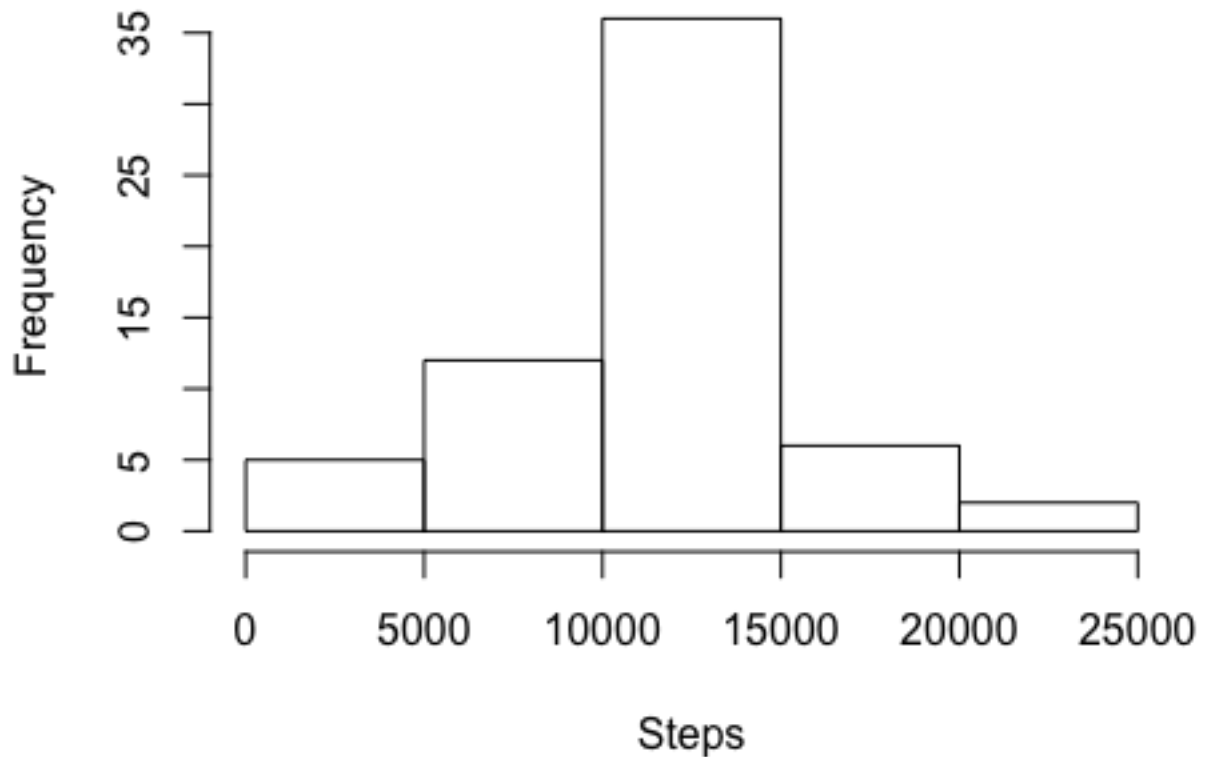
```
counter = 0  
for(i in 1:nrow(act)){  
  if(is.na(act$steps[i])){  
    counter = counter + 1  
  }  
}
```

```
}  
}  
counter  
## [1] 2304
```

- There are 2304 missing values in this data set

```
avgPerInt = aggregate(steps ~ interval, compAct, mean)  
t = ts(act$date)  
t = as.data.frame(t)  
act = cbind(act, t)  
  
merged = merge(act, avgPerInt, by = "interval")  
merged = merged[order(merged$x),]  
  
for(i in 1:nrow(merged)){  
  if(is.na(merged$steps.x[i])){  
    merged$steps.x[i] = merged$steps.y[i]  
  }  
}  
  
#produce identical data set to original  
newDataSet = merged[,c(2,3,1)]  
names(newDataSet)[1] = "steps"  
  
#histogram of total number of steps  
agg = aggregate(steps ~ date, newDataSet, FUN = "sum")  
hist(agg$steps, main = "Total Steps of each Day", xlab = "Steps")
```

## Total Steps of each Day



```
mean(agg$steps)
```

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

```
median(agg$steps)
```

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

- When including the imputed values, the mean is 10766.19, and the median becomes 10766.19 as well. The mean is identical to the one we found with the missing values. We would expect it to be very similar because we have a very large sample size and the imputed values are given using the averages per interval from the data set with no missing values. The median increased slightly, as we would expect with a larger number of values.

### Weekdays versus Weekends

```
newDataSet$date = as.Date(as.character(newDataSet$date))
```

```
day = weekdays(newDataSet$date)
```

```
newDataSet = cbind(newDataSet, day)
```

```

weekD = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")
which = NULL
for(i in 1:nrow(newDataSet)){
  if(newDataSet$day[i] %in% weekD){
    which[i] = "Weekday"
  } else{
    which[i] = "Weekend"
  }
}
newDataSet = cbind(newDataSet, which)
avgPerIntervalFinal = aggregate(steps ~ interval +
  which, newDataSet, mean)

ggplot(avgPerIntervalFinal, aes(x = interval, y = steps)) + geom_line() +
  facet_wrap(~ which)

```

