Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```r
#setwd("/Users/adakemia/Documents/Academic/Coursera/DataScienceSpecialization/05ReproducibleResearch/Pr
#test wd
setwd("/Users/adakemia/Documents/Academic/Coursera/DataScienceSpecialization/05ReproducibleResearch/Pro

url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"
zfile <- "activity.zip"
file <- "activity.csv"

if (!file.exists(file)) {
        if (!file.exists(zfile)) {
                method <- switch(Sys.info()[['sysname']],
                                 "Windows" = "internal",
                                 "Darwin" = "curl",
                                 "Linux" = "wget",
                                 "auto")
                download.file(url, destfile=zfile, method)
                }
        unzip(zfile, exdir="./")
}
library(data.table)
data <- fread(file, sep=",", stringsAsFactors=F, header=T,
              na.strings="NA", colClasses=c("numeric","character","numeric"))

str(data)
```

```
## Classes 'data.table' and 'data.frame':   17568 obs. of  3 variables:
##  $ steps   : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : chr  "2012-10-01" "2012-10-01" "2012-10-01" "2012-10-01" ...
##  $ interval: num  0 5 10 15 20 25 30 35 40 45 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:data.table':
##
##     hour, mday, month, quarter, wday, week, yday, year
```

```r
data[,date := ymd(date)]
```

```
##         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
##     ---
## 17564:      NA 2012-11-30      2335
## 17565:      NA 2012-11-30      2340
## 17566:      NA 2012-11-30      2345
## 17567:      NA 2012-11-30      2350
## 17568:      NA 2012-11-30      2355
```
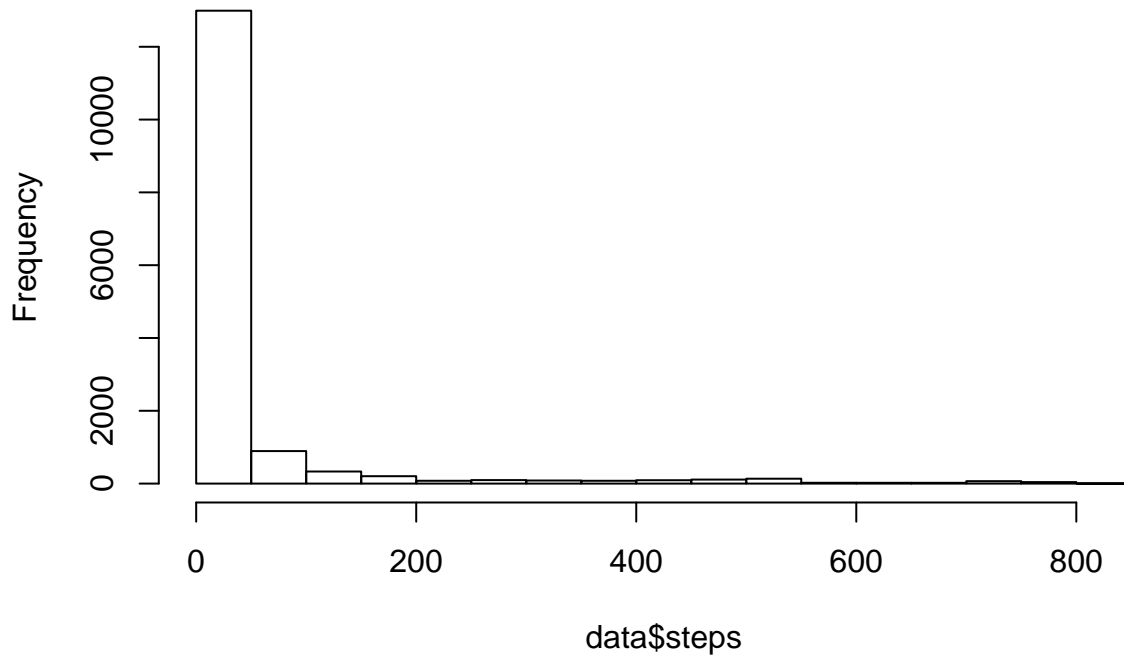
```
str(data)
```

```
## Classes 'data.table' and 'data.frame':   17568 obs. of  3 variables:
##  $ steps   : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : POSIXct, format: "2012-10-01" "2012-10-01" ...
##  $ interval: num  0 5 10 15 20 25 30 35 40 45 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
colSums(is.na(data))
```

```
##    steps      date interval
##     2304         0        0
```

```
hist(data$steps)
```

Histogram of data$steps

```
data[,unique(date)]
```

```
##  [1] "2012-10-01 UTC" "2012-10-02 UTC" "2012-10-03 UTC" "2012-10-04 UTC"
##  [5] "2012-10-05 UTC" "2012-10-06 UTC" "2012-10-07 UTC" "2012-10-08 UTC"
##  [9] "2012-10-09 UTC" "2012-10-10 UTC" "2012-10-11 UTC" "2012-10-12 UTC"
## [13] "2012-10-13 UTC" "2012-10-14 UTC" "2012-10-15 UTC" "2012-10-16 UTC"
## [17] "2012-10-17 UTC" "2012-10-18 UTC" "2012-10-19 UTC" "2012-10-20 UTC"
## [21] "2012-10-21 UTC" "2012-10-22 UTC" "2012-10-23 UTC" "2012-10-24 UTC"
## [25] "2012-10-25 UTC" "2012-10-26 UTC" "2012-10-27 UTC" "2012-10-28 UTC"
## [29] "2012-10-29 UTC" "2012-10-30 UTC" "2012-10-31 UTC" "2012-11-01 UTC"
## [33] "2012-11-02 UTC" "2012-11-03 UTC" "2012-11-04 UTC" "2012-11-05 UTC"
## [37] "2012-11-06 UTC" "2012-11-07 UTC" "2012-11-08 UTC" "2012-11-09 UTC"
## [41] "2012-11-10 UTC" "2012-11-11 UTC" "2012-11-12 UTC" "2012-11-13 UTC"
## [45] "2012-11-14 UTC" "2012-11-15 UTC" "2012-11-16 UTC" "2012-11-17 UTC"
## [49] "2012-11-18 UTC" "2012-11-19 UTC" "2012-11-20 UTC" "2012-11-21 UTC"
## [53] "2012-11-22 UTC" "2012-11-23 UTC" "2012-11-24 UTC" "2012-11-25 UTC"
## [57] "2012-11-26 UTC" "2012-11-27 UTC" "2012-11-28 UTC" "2012-11-29 UTC"
## [61] "2012-11-30 UTC"
```
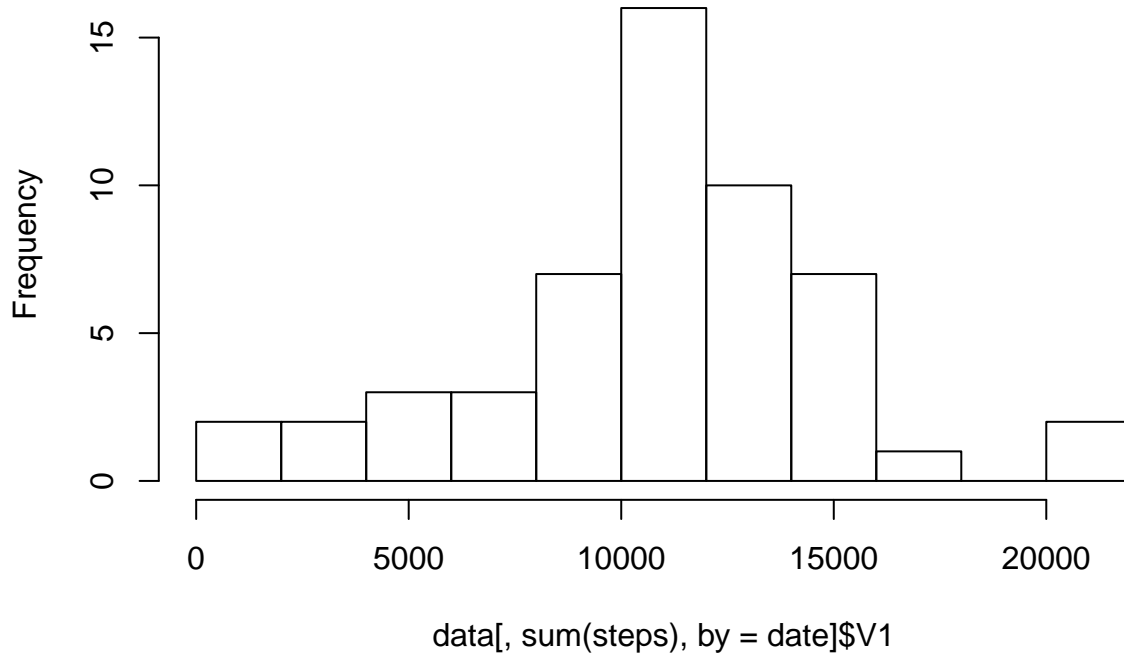
```
plot(data$interval[11:14], type="l")
```
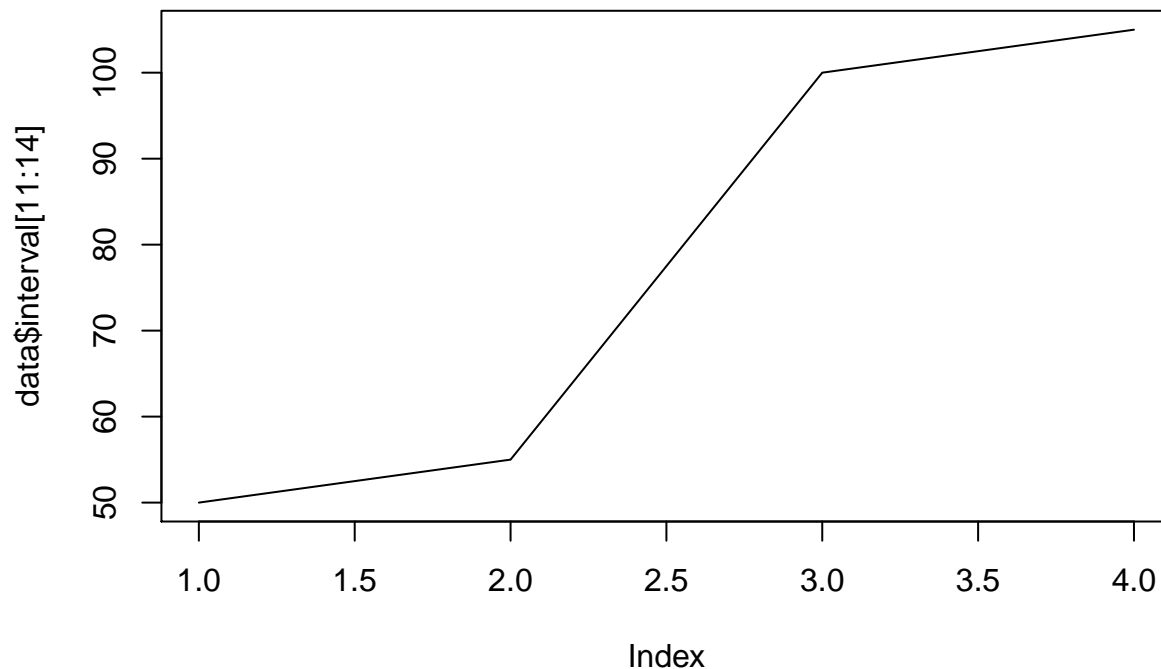


```
data[11:14]
```

```
##    steps       date interval
## 1:    NA 2012-10-01       50
## 2:    NA 2012-10-01       55
## 3:    NA 2012-10-01      100
## 4:    NA 2012-10-01      105
```

What is mean total number of steps taken per day?

```
hist(data[,sum(steps), by=date]$V1, breaks=8)
```

Histogram of data[, sum(steps), by = date]$V1
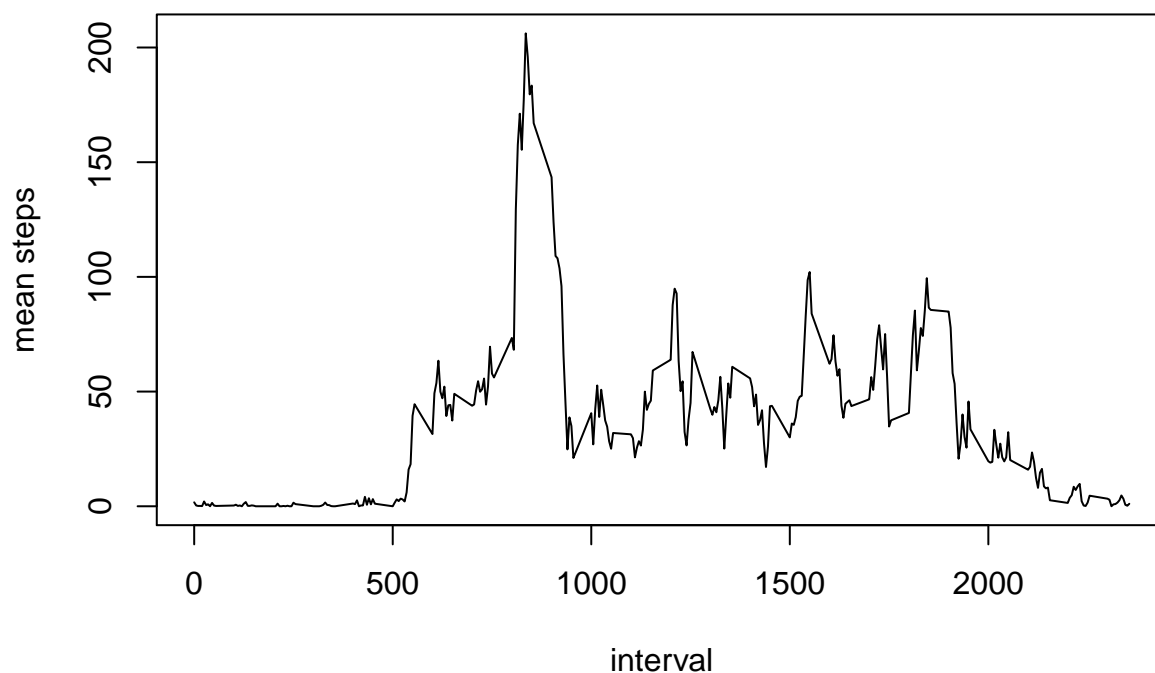


```
data[, .(sum = sum(steps)), by=date][,.(median = median(sum, na.rm=T),
                                         mean = mean(sum, na.rm=T))]
```

```
##     median     mean
## 1:  10765 10766.19
```

What is the average daily activity pattern?

```
ts <- data[, .(mean = round( mean(steps, na.rm=T), 2)), by=interval]

plot(ts$mean ~ ts$interval, type="l", main="Mean Steps Per Daily Interval",
     xlab="interval", ylab="mean steps")
```

Mean Steps Per Daily Interval


```
ts[which.max(mean)]
```

```
##    interval   mean
## 1:      835 206.17
```

Imputing missing values

```
dateNum <- unclass(data[,date])
```

Are there differences in activity patterns between weekdays and weekends?