Reproductible Reaseatch_Project1

Serge NYOKA

23 octobre 2016

Reproductible Research : Course Project 1

1. Loading and preprocessing the data

```
#Reading data file in the csv format.
dt_activity <- read.csv("activity.csv")
str(dt_activity)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : Factor w/ 61 levels "2012-10-01","2012-10-02",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

2. What is mean total number of steps taken per day ?

```
#Summing the steps for each day. 60 observations for 60 days
steps_daily <- aggregate(steps ~ date, dt_activity, sum)
head(steps_daily)
```

```
##         date steps
## 1 2012-10-02   126
## 2 2012-10-03 11352
## 3 2012-10-04 12116
## 4 2012-10-05 13294
## 5 2012-10-06 15420
## 6 2012-10-07 11015
```

Plotting Histogram

Total number of steps per day
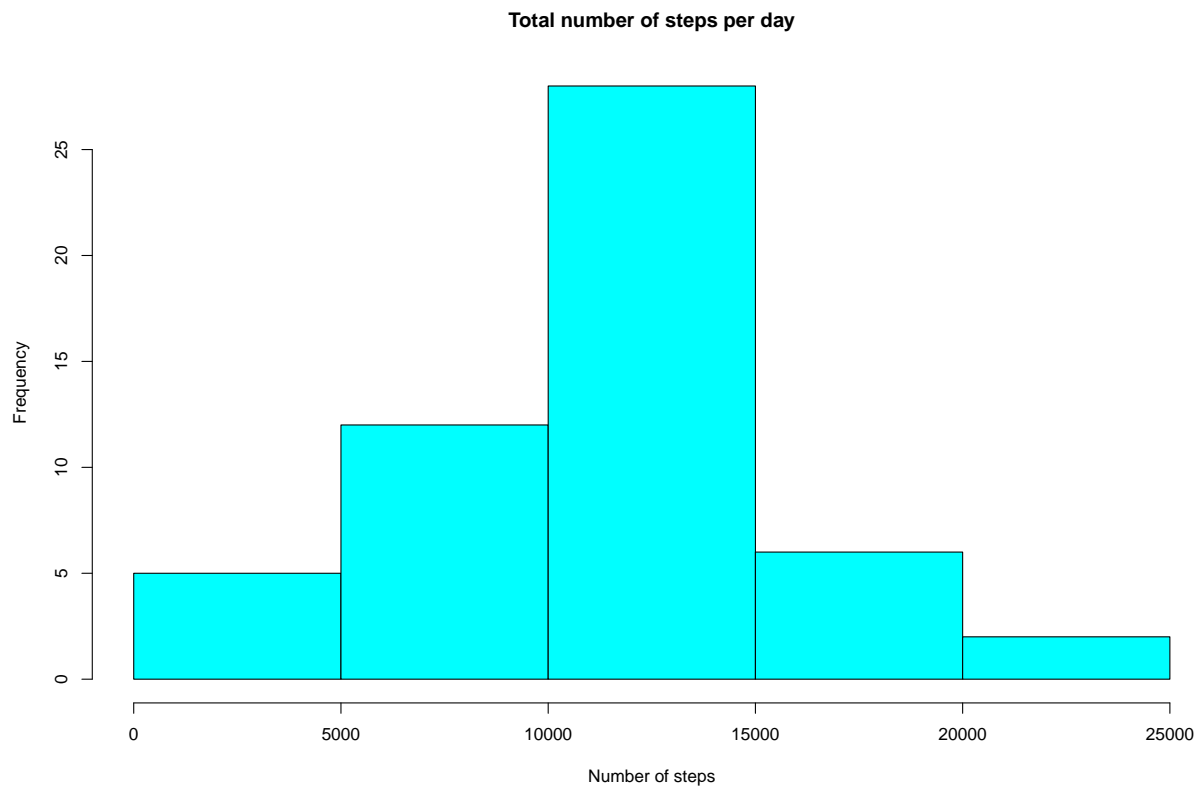

The mean is equal to 10766 and the median to 10765

summary(steps_daily)

3. What is the average daily activity pattern ?

```r
#Calculating the mean and the median of the total numberof steps per day
steps_mean <- mean(steps_daily$steps)
steps_median <- median(steps_daily$steps)

#For all days and for each interval we calculate the average steps
steps_interval <- aggregate(steps ~ interval, dt_activity, mean)
```

Plotting the average number of steps per day for each interval

Average of each day steps's number by interval
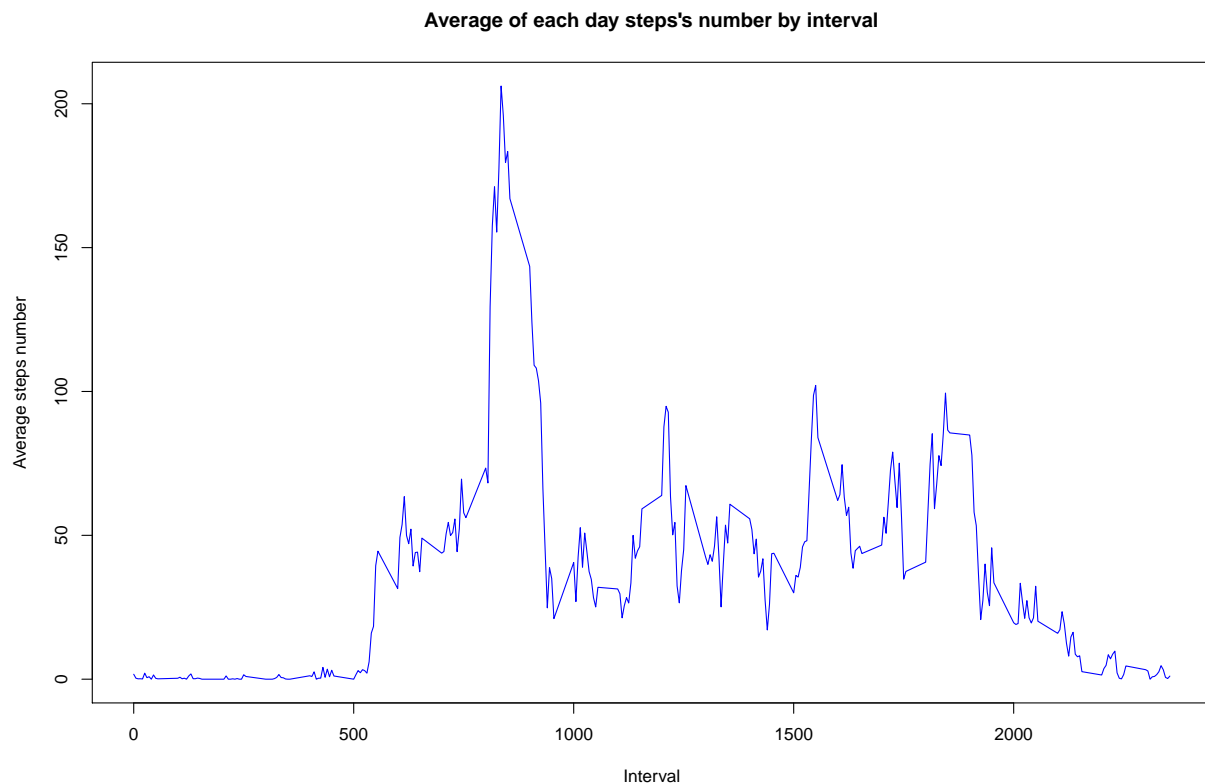

```
#Compute the maximum average steps of intervals
interval_maximum <- steps_interval[which.max(steps_interval$steps), 1]
interval_maximum
```

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

4. Imputing missing values

```
#Calculating and reporting the total number of missing value
missing_data_number <- sum(!complete.cases(dt_activity))
missing_data_number
```

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

```
dt_activity_imput <- transform(dt_activity, steps = ifelse(is.na(dt_activity$steps), steps_interval$step
dt_activity_imput[as.character(dt_activity_imput$date) == "2012-10-01", 1] <- 0

#Compute total steps per day imputed
steps_daily_imput <- aggregate(steps ~ date, dt_activity_imput, sum)
```
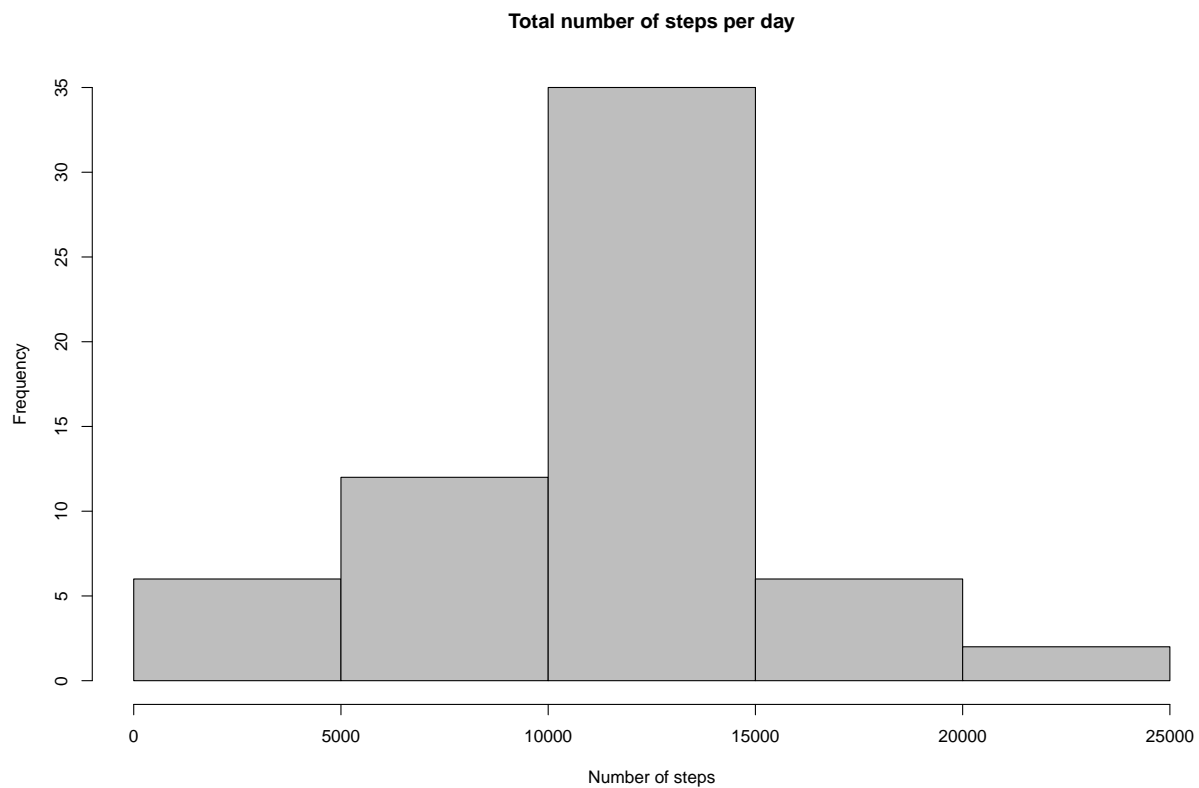
```
head(dt_activity_imput)
```

```
##   steps       date interval
```

```
## 1      0 2012-10-01          0
## 2      0 2012-10-01          5
## 3      0 2012-10-01         10
## 4      0 2012-10-01         15
## 5      0 2012-10-01         20
## 6      0 2012-10-01         25
```

Visualizing the imputed data without all NAs

Histogram of total number steps for each day

Total number of steps per day


```
#Calculate and report the mean and median total number of steps taken per day
steps_mean_imput <- mean(steps_daily_imput$steps)
steps_mean_imput
```

```
## [1] 10589.69
```

```
steps_median_imput <- median(steps_daily_imput$steps)
steps_median_imput
```

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

```
#Difference between mean, median and total number steps of the 2 dataset (imputed and non-imputed)
delta_mean <- steps_mean_imput - steps_mean
delta_mean
```

```
## [1] -176.4949
```

```
delta_median <- steps_median_imput - steps_median
delta_median
```

```
## [1] 1.188679
```

```
delta_steps <- sum(steps_daily_imput$steps) - sum(steps_daily$steps)
delta_steps
```

```
## [1] 75363.32
```

5. Are there differences in actyivity patterns between weekdays and weekends ?

Differences %in% activity patterns between weekdays and weekends

```
dayPos <- weekdays(as.Date(dt_activity_imput$date))
for(i in 1:length(dayPos)) {
  if(dayPos[i] %in% c("samedi", "dimanche")) dayPos[i] = "weekend"
  else dayPos[i] = "weekday"
}

dt_activity_imput$dayPosition <- as.factor(dayPos)

steps_interval_imput <- aggregate(steps ~ interval + dayPosition, dt_activity_imput, mean)

library(lattice)
```

Plotting . . .

Average steps for each day by interval
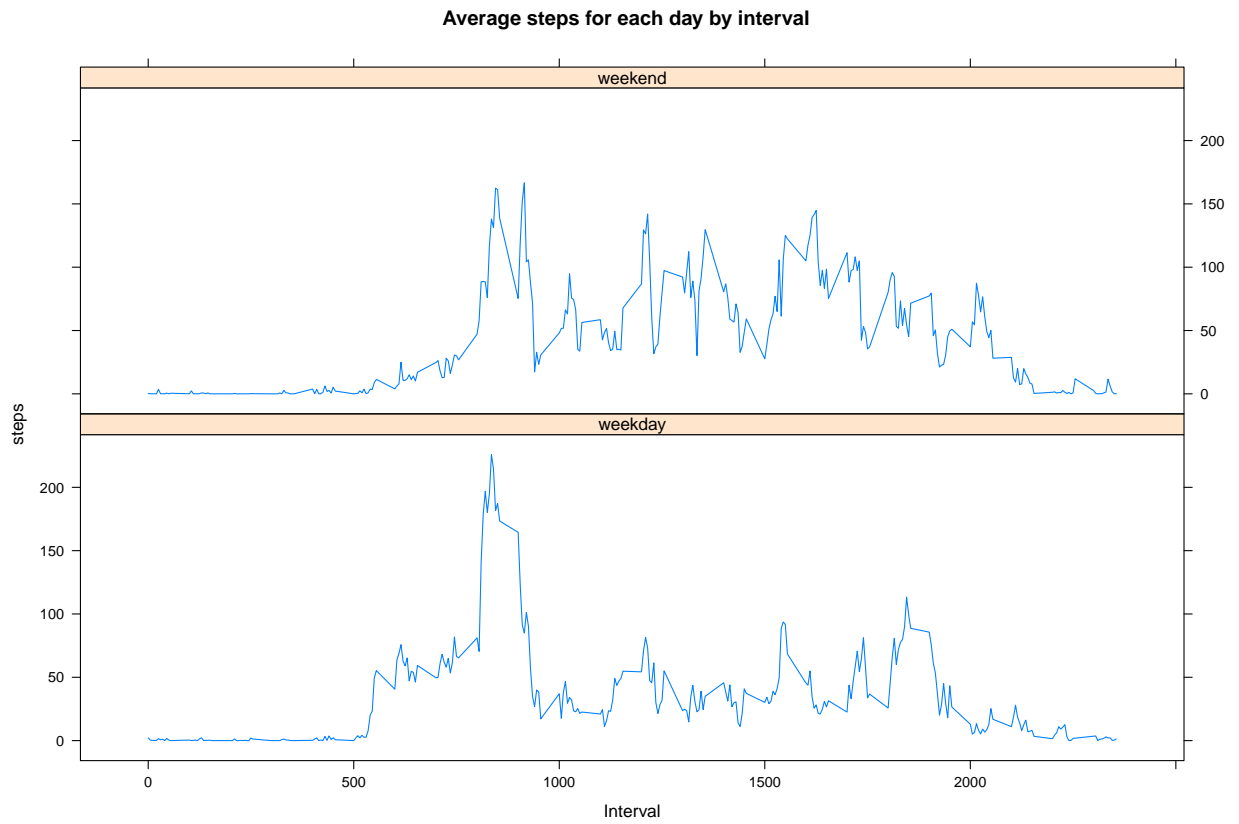
The difference is visible on these 2 charts