

Reproducible Research: Peer Assessment 1

Juan Perez-Bermejo

August 16, 2014

Load and process data. Also loads dependencies

This code chunk will check if the file is already unzipped. If not, it will unzip it. During the file loading process, it will set the classes of the columns. Resulting data frame details can be seen in the summary table below.

```
library(xtable)
library(ggplot2)
library(reshape2)

if(!file.exists("activity.csv")){
  unzip("./activity.zip")
}

act<-read.csv("activity.csv", colClasses=c("integer", "Date", "integer"))
sum<-xtable(summary(act))
print(sum, type="html")
```

steps

date

interval

1

Min. : 0.0

Min. :2012-10-01

Min. : 0

2

1st Qu.: 0.0

1st Qu.:2012-10-16

1st Qu.: 589

3

Median : 0.0

Median :2012-10-31

Median :1178

4

Mean : 37.4

Mean :2012-10-31

Mean :1178

5

3rd Qu.: 12.0
3rd Qu.:2012-11-15
3rd Qu.:1766
6
Max. :806.0
Max. :2012-11-30
Max. :2355
7
NA's :2304

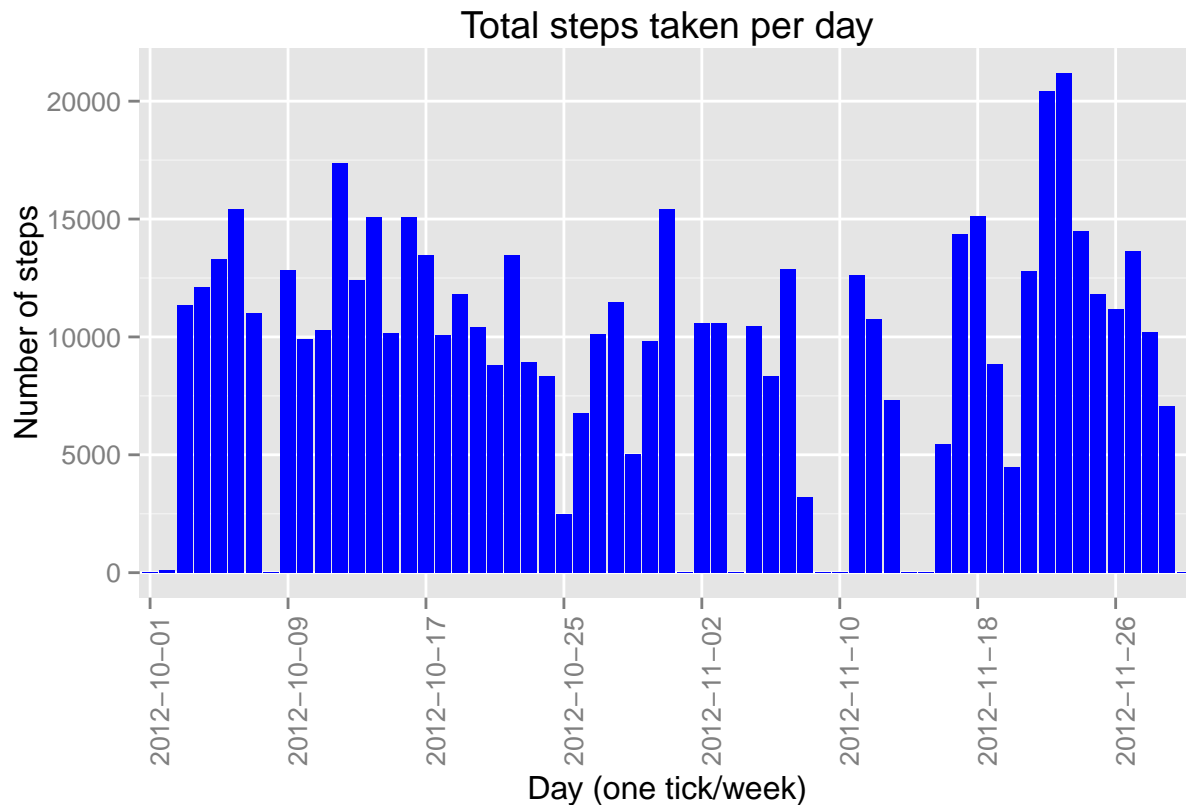
What is the mean total number of steps taken per day?

```
stepsDay<-tapply(act$steps, act$date, sum, na.rm=T)
stepsDay<-melt(stepsDay)
meanStepsDay<-mean(stepsDay$value, na.rm=T)
medianStepsDay<-median(stepsDay$value, na.rm=T)
```

On average, the subject took 9354.2295 steps per day. The median value is 10395.

A histogram depicting the total number of steps per day is shown below. Note that some days had no available records at all (maybe because the subject did not activate the device?). Those are represented as blanks. No single day for which the device was recording presented zero steps. The histogram reveals that some days the subject walked around twice the average steps per day, while others he just didn't walk much (this can also be due to device malfunction/not being active).

```
sd<-ggplot(data=stepsDay, aes(x=Var1, y=value))
sd<-sd+geom_bar(stat="identity", fill="blue")
sd<-sd+labs(title="Total steps taken per day", x="Day (one tick/week)", y="Number of steps")
sd<-sd+theme(axis.text.x=element_text(angle=90, v=1, h=1))
sd<-sd+scale_x_discrete(breaks=stepsDay$Var1[seq(1, length(stepsDay[,2]), 8)])
sd
```

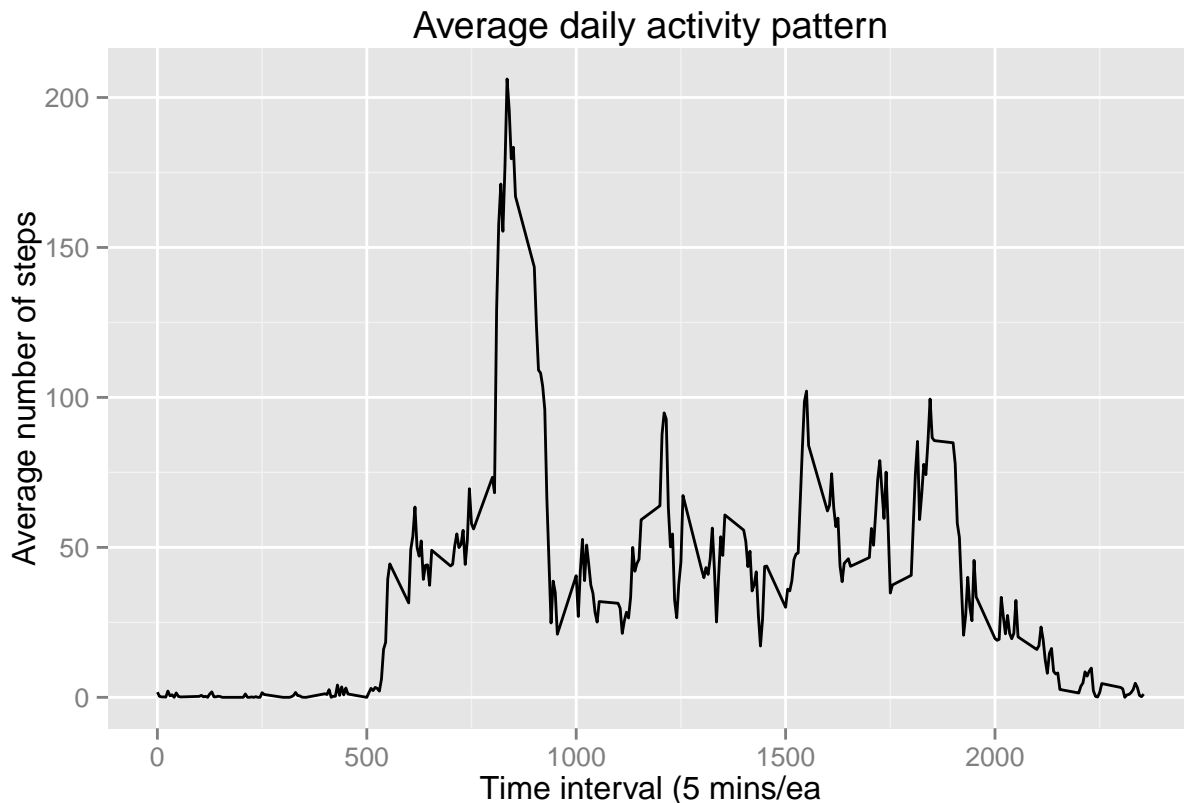


What is the average daily activity pattern?

The plot below shows the average number of steps per day, for each one of the time intervals (times) of the day.

```
actPattern<-tapply(act$steps, act$interval, mean, na.rm=T)
actPattern<-melt(actPattern)
maxActInterval<-actPattern$Var1[actPattern$value==max(actPattern$value)]

ap<-ggplot(data=actPattern, aes(x=Var1, y=value))
ap<-ap+geom_line(stat="identity")
ap<-ap+labs(title="Average daily activity pattern",
            x="Time interval (5 mins/ea", y="Average number of steps")
ap
```



The 5 minute interval of the day during which the subject took the higher number of steps was at 835 in the morning. Probably he commutes to work on foot, or maybe goes for a run in the morning?

Imputing missing values

```
nMissingVal<-table(complete.cases(act))[1]
```

There are in total 2304 observations missing in the dataset.

To fill those, I decided to take the average of steps in the same weekday, on the same time interval. I did this as the activity of the individual is most likely the same in the same time interval on the same day of the week. One caveat is that the number of datapoints is different for every weekday.

```
act$weekday<-weekdays(act$date)

#Generates a table with average steps per interval per day of the week
stepsWD<-tapply(act$steps, list(act$interval, act$weekday), mean, na.rm=T)
stepsWD<-melt(stepsWD)
act2<-act
for (n in 1:nrow(act2)){
  o<-act2[n,]
  if(!complete.cases(o)){
    act2[n,]$steps<-stepsWD$value[stepsWD$Var1==o$interval & stepsWD$Var2==o$weekday]
  }
}
```

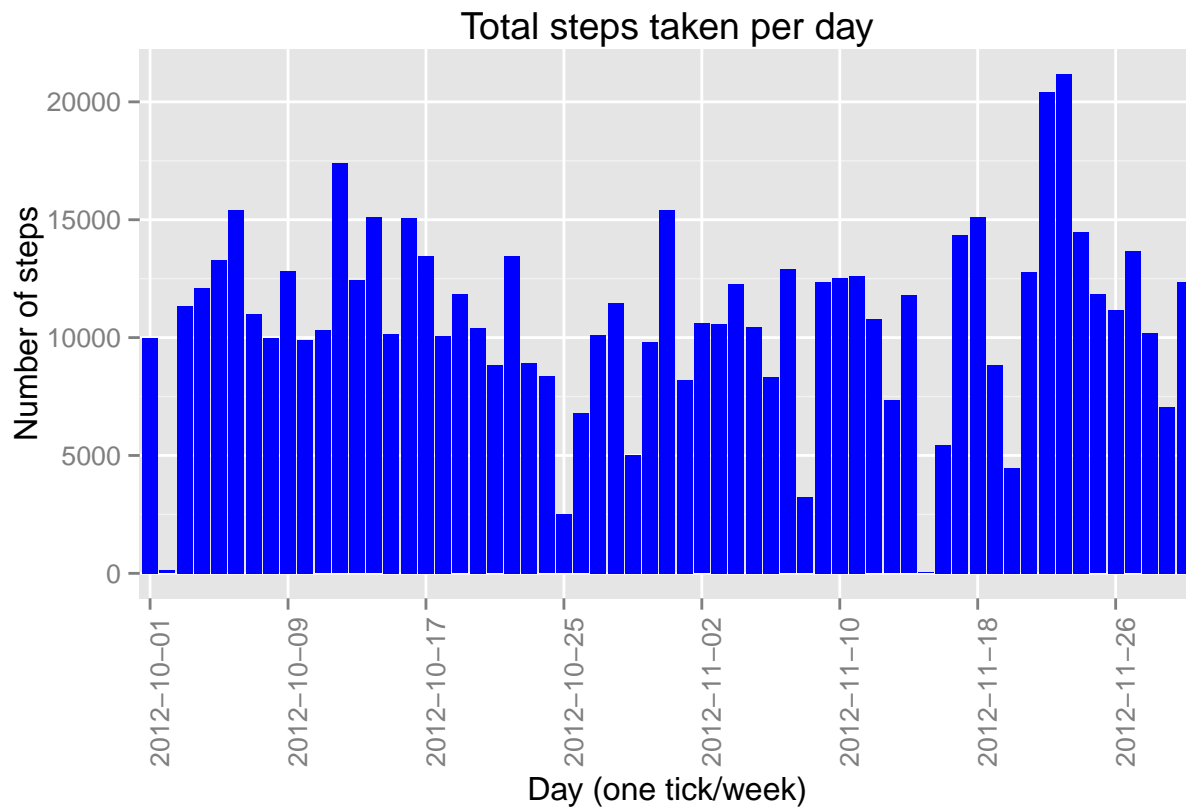
This imputation allows us to recalculate the total steps taken per day. *Note: this code block is identical to the one used in the first part of the assignment, but this time using the 'filled in' dataset.

```

stepsDay2<-tapply(act2$steps, act2$date, sum, na.rm=T)
stepsDay2<-melt(stepsDay2)
meanStepsDay2<-mean(stepsDay2$value)
medianStepsDay2<-median(stepsDay2$value)

sd2<-ggplot(data=stepsDay2, aes(x=Var1, y=value))
sd2<-sd2+geom_bar(stat="identity", fill="blue")
sd2<-sd2+labs(title="Total steps taken per day", x="Day (one tick/week)", y="Number of steps")
sd2<-sd2+theme(axis.text.x=element_text(angle=90, v=1, h=1))
sd2<-sd2+scale_x_discrete(breaks=stepsDay2$Var1[seq(1, length(stepsDay2[,2]), 8)])
sd2

```



As it can be observed in the plot, there are still some days on which the individual walks much more (or less) than in others. The average total steps per day is now , median being .

Are there differences in activity patterns between weekdays and weekends?