Reproducible Research: Peer Assessment 1

Jennyfer Combariza

December 5, 2016

Loading and preprocessing the data

```
data <- read.csv("activity.csv")
```

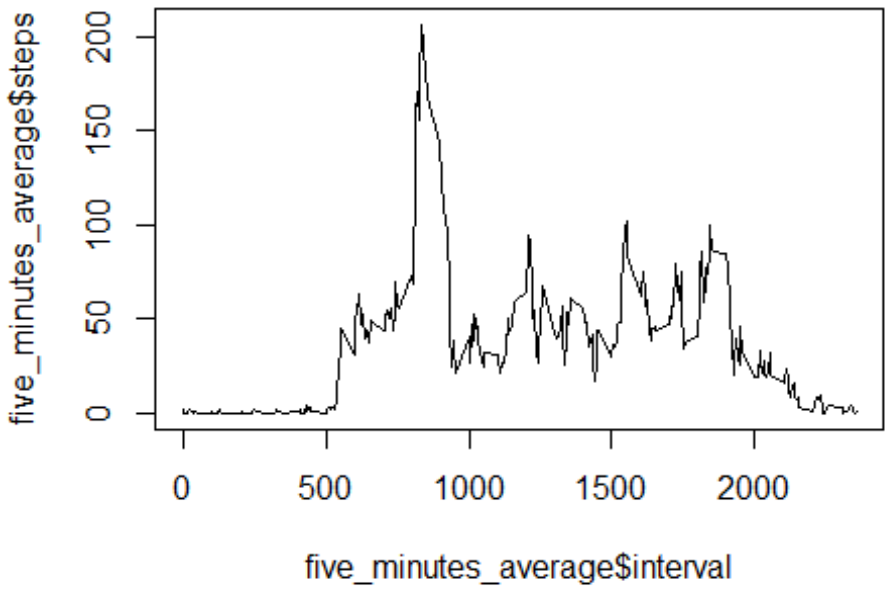
What is mean total number of steps taken per day?

```
library(plyr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)

total.steps <- tapply(data$steps, data$date, FUN = sum, na.rm = TRUE)

#Mean
mean(total.steps)

## [1] 9354.23

#Median
median(total.steps)

## [1] 10395
```

What is the average daily activity pattern?

We make a time series plot of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis) and we plot the result.

```
five_minutes_average <- aggregate(steps~interval, data=data, FUN=mean,
na.rm=TRUE)
plot(x = five_minutes_average$interval, y = five_minutes_average$steps, type
= "l")
```

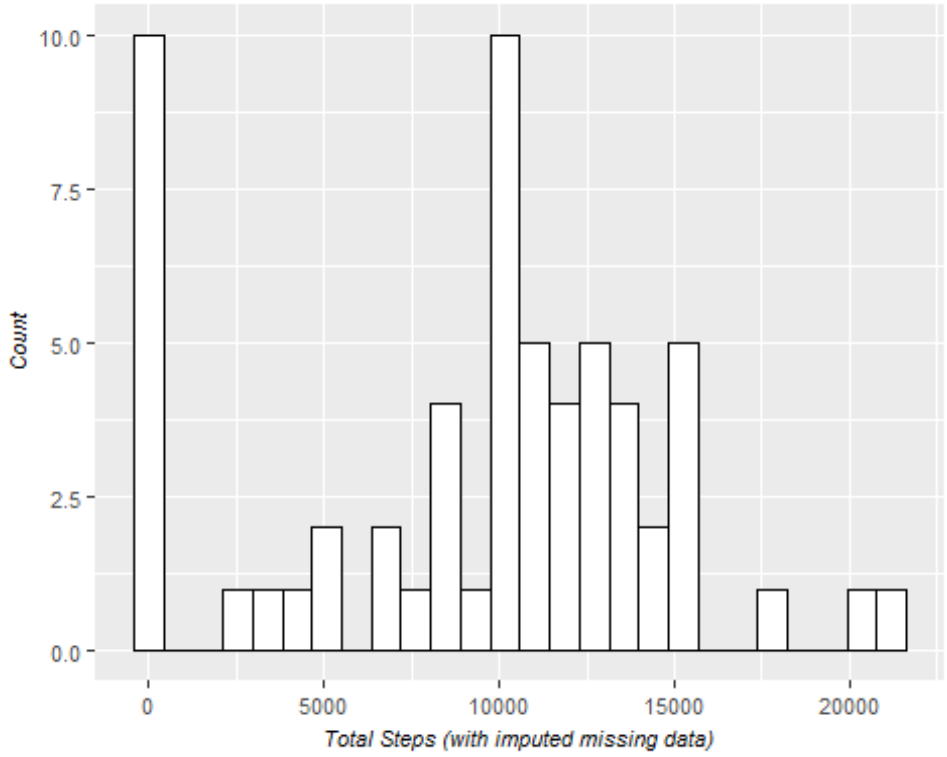



```
png("average.png", width=750)
plot(x = five_minutes_average$interval, y = five_minutes_average$steps, type
= "l")
dev.off()

## png
##   2

max_steps <- max(five_minutes_average$steps)
for (i in 1:288)
{
  if (five_minutes_average$steps[i] == max_steps)
    five_minute_interval_at_max_steps <- five_minutes_average$interval[i]
}
five_minute_interval_at_max_steps

## [1] 835
```

Imputing missing values

```
sum(!complete.cases(data))

## [1] 2304
```

```
fillfunc <- function(step, interval) {
  ifelse(is.na(step), averages[averages$interval == interval, ]$steps, step)
}

data_fill <- data
tot_steps_day <- aggregate(steps ~ date, data=data_fill, FUN=sum,
                           na.rm=TRUE, na.action=NULL)

granularity = diff(range(tot_steps_day$steps)) / 25
ggplot(data=tot_steps_day, aes(x=tot_steps_day$steps))              +
  geom_histogram(binwidth=granularity, col="black", fill="white") +
  labs(x="Total Steps (with imputed missing data)", y="Count")    +
  theme(axis.text=element_text(size=8),
        axis.title=element_text(size=8, face="italic"))
```



Are there differences in activity patterns between weekdays and weekends?

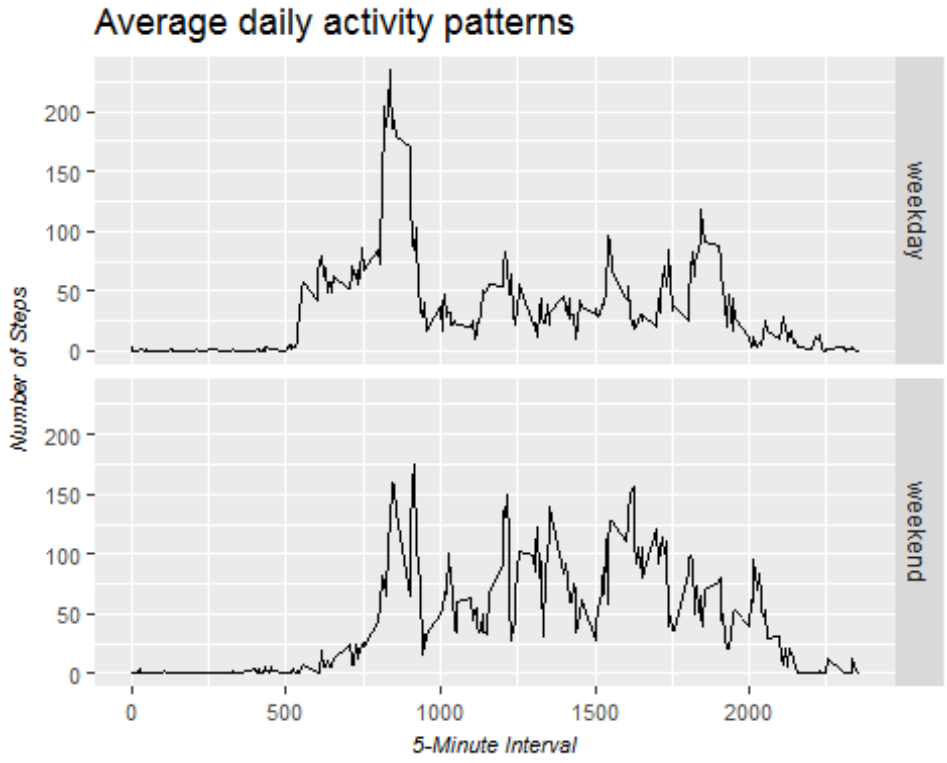
We start by creating a new factor variable in the dataset with two levels -- weekday and weekend indicating whether a given date is a weekday or weekend day.

```
switch(Sys.info()[[ 'sysname' ]],
    Windows = { lctime <- "English" }, { lctime <- "C" })
Sys.setlocale("LC_TIME", lctime)

## [1] "English_United States.1252"
```

```
data_fill$date <- as.Date(data_fill$date, "%Y-%m-%d")
daytype <- function(day) {
    ifelse(weekdays(day, abbreviate=FALSE) %in% c("Saturday", "Sunday"),
           "weekend", "weekday")
}
data_fill$daytype <- mapply(daytype, data_fill$date)
```

And we end this document by plotting the two resulting datasets:

```
averages <- aggregate(steps ~ interval + daytype, data=data_fill, FUN=mean)
ggplot(data=averages, aes(x=interval, y=steps))   +
       geom_line()                                +
       facet_grid(daytype ~ .)                    +
       ggtitle("Average daily activity patterns") +
       xlab("5-Minute Interval")                  +
       ylab("Number of Steps")                    +
       theme(axis.text=element_text(size=8),
             axis.title=element_text(size=8, face="italic"))
```



Answer

Yes, there are differences in activity patterns between weekdays and weekends.