Course Project 1 - Quantified self

Ana Clara

10/28/2020

Loading libraries

```
library(dplyr)
library(magrittr)
library(knitr)
library(lattice)
library(formatR)
```

Setting global options

```
options(scipen=1, digits=2)
knitr::opts_chunk$set(echo = TRUE)
options(dplyr.summarise.inform = FALSE)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

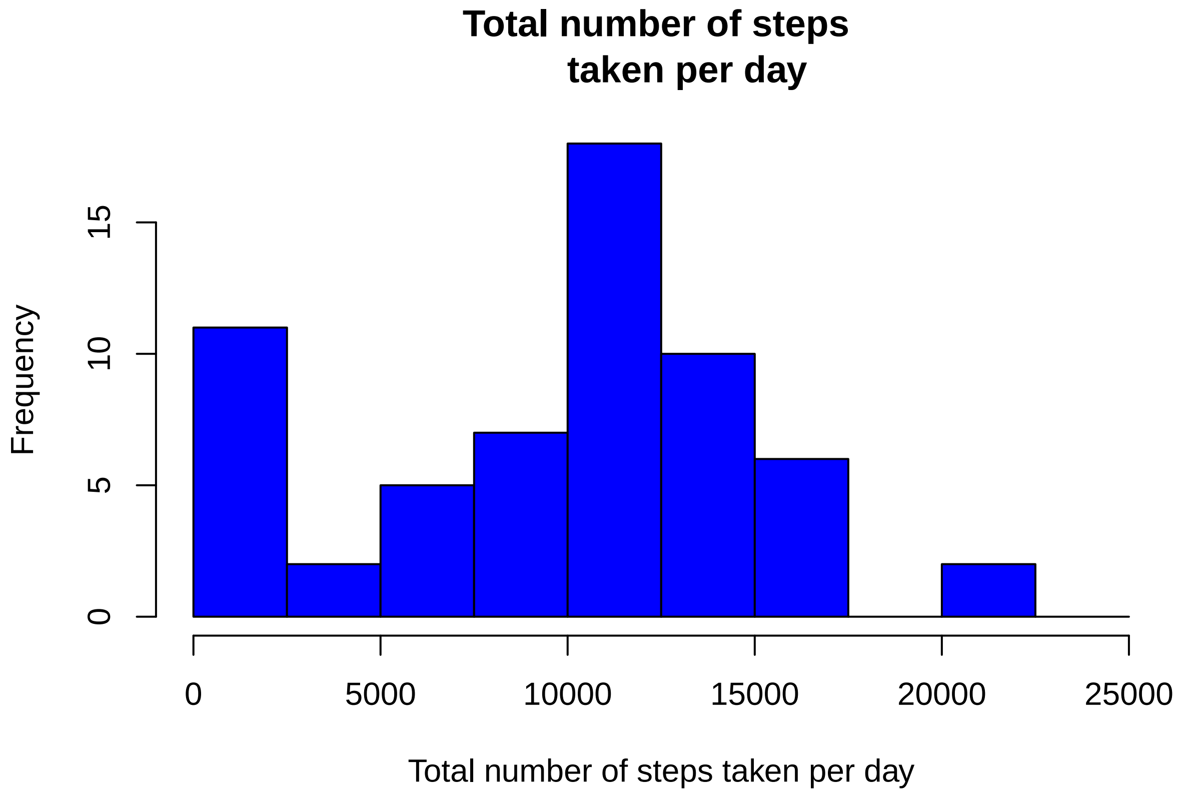
Unziping, reading and presenting pieces of raw data

```
data <- unzip("activity.zip")
data <- read.csv2("./activity.csv", sep = ",")

data$date <- as.POSIXct(data$date, "%Y-%m-%d", tz = "")
head(data)
```

```
##   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
## 6    NA 2012-10-01       25
```

What is the total number of steps taken per day?

```
## Calculating and presenting total of steps per day
sum_data <- data %>% group_by(date) %>% summarise(sum = sum(steps,
    na.rm = TRUE))
hist(sum_data$sum, breaks = seq(0, 25000, by = 2500), main = "Total number of steps
     taken per day",
    xlab = "Total number of steps taken per day", col = "blue")
```



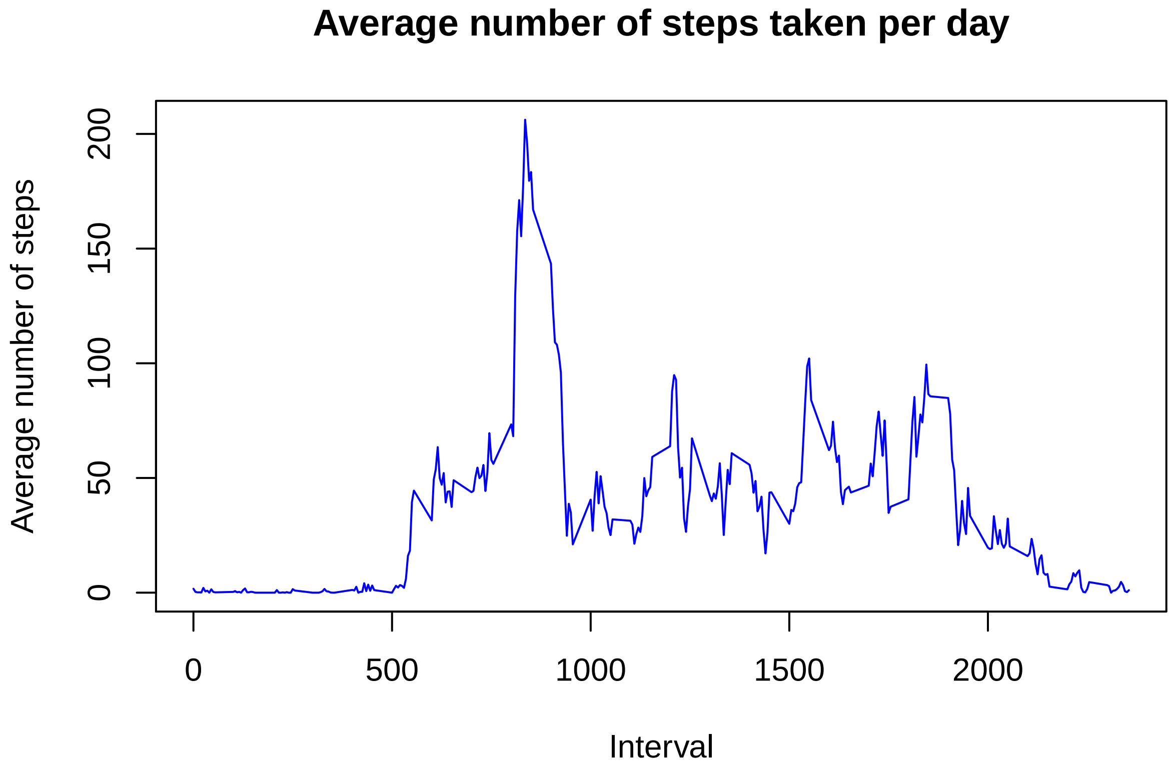
What is the mean and median number of the total steps taken per day?

```
## Calculating and presenting mean of total of steps per day
mean_data <- round(mean(sum_data$sum, na.rm = TRUE), 2)

median_data <- round(median(sum_data$sum, na.rm = TRUE), 2)
```

Mean of total steps is 9354.23. Median of total steps is 10395.

What is the average daily activity pattern?

```
## Calculating and presenting average of steps per interval
mean_int_data <- data %>% group_by(interval) %>% summarise(mean = (mean(steps,
    na.rm = TRUE)))
plot(mean_int_data$interval, mean_int_data$mean, type = "l",
    main = "Average number of steps taken per day", xlab = "Interval",
    ylab = "Average number of steps", col = "blue")
```



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
## Calculating and presenting max of steps per interval
max_interval <- mean_int_data[which.max(mean_int_data$mean),
    ]$interval
```

The 5-minute interval that contains the maximum number of steps on average across all days is: 835 , with a number of 104 steps

Imputing missing values

1. Calculate and report the total number of missing values in the dataset (i.e. the total number of rows with *NA*)

```r
## Calculating and presenting number of missing values
miss <- sum(is.na(data$steps))
```

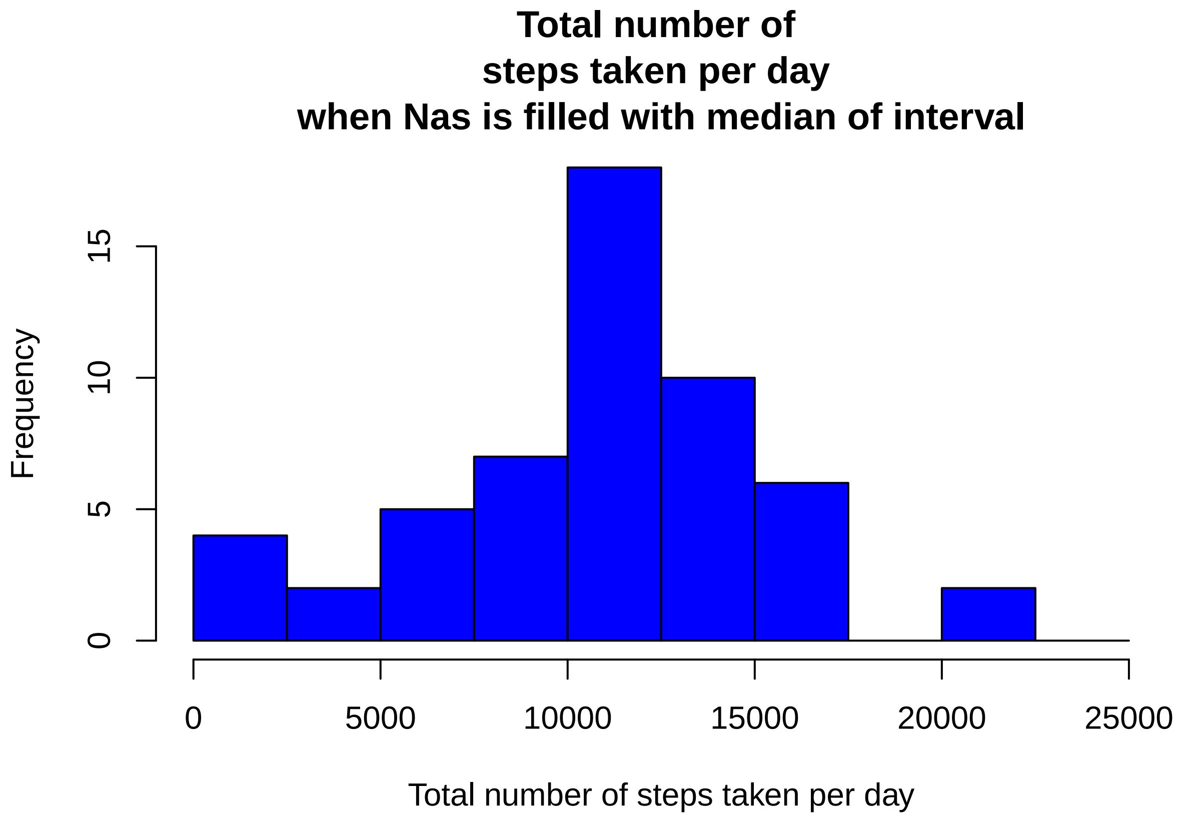
The number of total missing values id: 2304

2. Devise a strategy for filling in all of the missing values in the dataset. The strategy does not need to be sophisticated. For example, you could use the mean/median for that day, or the mean for that 5-minute interval, etc.
3. Identifying and filling NAs with median of interval

```r
## Identifying and filling NAs with median of interval
fill_miss_data <- data

median_int_data <- data %>% group_by(interval) %>% summarise(median = (median(steps,
    na.rm = TRUE)))

fill_miss_data$steps <- ifelse(is.na(data$steps), median_int_data[which(median_int_data$interval ==
    data$interval), ]$median, data$steps)
```

4. Make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day. Do these values differ from the estimates from the first part of the assignment? What is the impact of imputing missing data on the estimates of the total daily number of steps?

```r
## Calculating and presenting total of steps per day
sum_fill_data <- fill_miss_data %>% group_by(date) %>% summarise(sum = sum(steps))
hist(sum_fill_data$sum, breaks = seq(0, 25000, by = 2500), main = "Total number of
steps taken per day
when Nas is filled with median of interval",
    xlab = "Total number of steps taken per day", col = "blue")
```



```r
## Calculating and presenting mean of total of steps per day
mean_fill_data <- round(mean(sum_fill_data$sum, na.rm = TRUE),
    2)

median_fill_data <- round(median(sum_fill_data$sum, na.rm = TRUE),
    2)
```

Mean of total steps is: 10587.94.
Median of total steps is: 10682.5.

The impact is:

```r
## Calculating and presenting mean of total of steps per day
metrics <- cbind.data.frame(c(mean_data, median_data), c(mean_fill_data,
    median_fill_data))
names(metrics) <- c("Original data set", "Filling missing values")
row.names(metrics) <- c("Mean", "Median")
metrics
```

```
##        Original data set Filling missing values
## Mean                9354                  10588
## Median             10395                  10682
```

Are there differences in activity patterns between weekdays and weekends?

For this part the weekdays() function may be of some help here. Use the dataset with the filled-in missing values for this part.

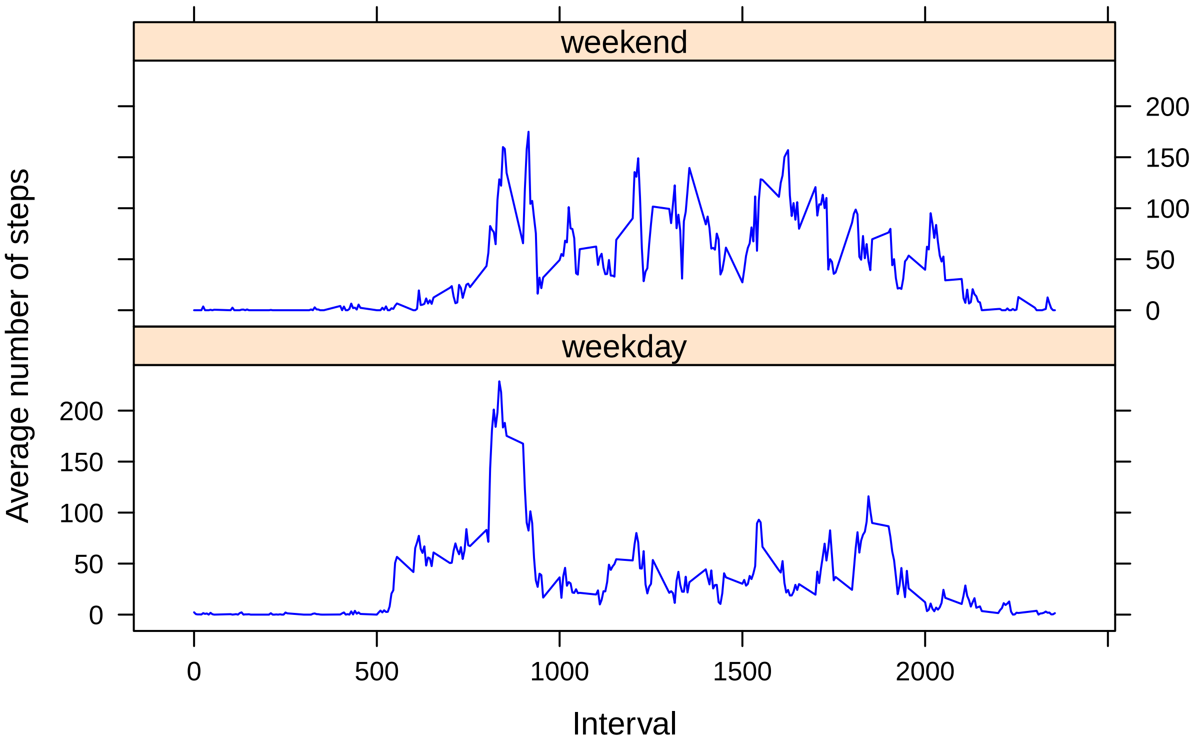
1. Create a new factor variable in the dataset with two levels – "weekday" and "weekend" indicating whether a given date is a weekday or weekend day.

```
## Adding 2 level factor weekday/weekend
fill_miss_data$days <- weekdays(fill_miss_data$date, FALSE)

fill_miss_data$wdays <- ifelse(fill_miss_data$days == "Saturday" |
    fill_miss_data$days == "Sunday", "weekend", "weekday")
fill_miss_data$wdays <- as.factor(fill_miss_data$wdays)

fill_miss_data <- fill_miss_data[, c(1, 2, 3, 5)]
```

2. Make a panel plot containing a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis). See the README file in the GitHub repository to see an example of what this plot should look like using simulated data.

```
## Calculating and presenting average of steps per interval
mean_int_wdays <- fill_miss_data %>% group_by(interval, wdays) %>%
    summarise(mean = (mean(steps, na.rm = TRUE)))

xyplot(mean_int_wdays$mean ~ mean_int_wdays$interval | mean_int_wdays$wdays,
    type = "l", main = "Average number of steps taken per day",
    xlab = "Interval", ylab = "Average number of steps", col = "blue",
    layout = c(1, 2))
```

Average number of steps taken per day


weekend
weekday
Average number of steps
Interval