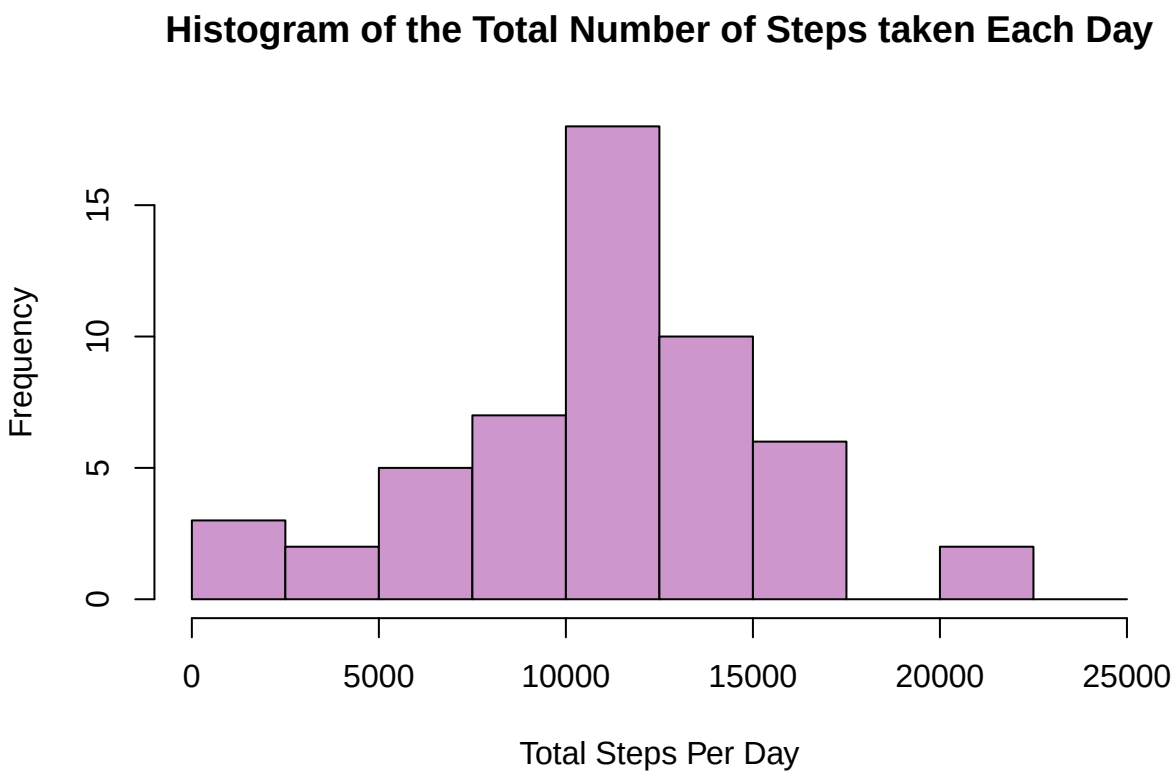
Reproducible Research

ludimila

15/10/2020

```
knitr::opts_chunk$set(warning=FALSE)

#require library

library(ggplot2)

#download file

fileurl = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"
if(!file.exists("./activity")){dir.create("./activity")}
download.file(fileurl, destfile = "./activity.zip")

#unzip and load data

unzip(zipfile = "./activity.zip", exdir = "./activity")

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


### 1 - What is mean number of steps taken per day?

TOTALstepday <- aggregate(steps~date, activity, sum)

hist(TOTALstepday$steps, main = "Histogram of the Total Number of Steps taken Each Day",
     xlab = "Total Steps Per Day", ylab = "Frequency", col = "plum3", breaks = seq(0,25000, by=2500))
```



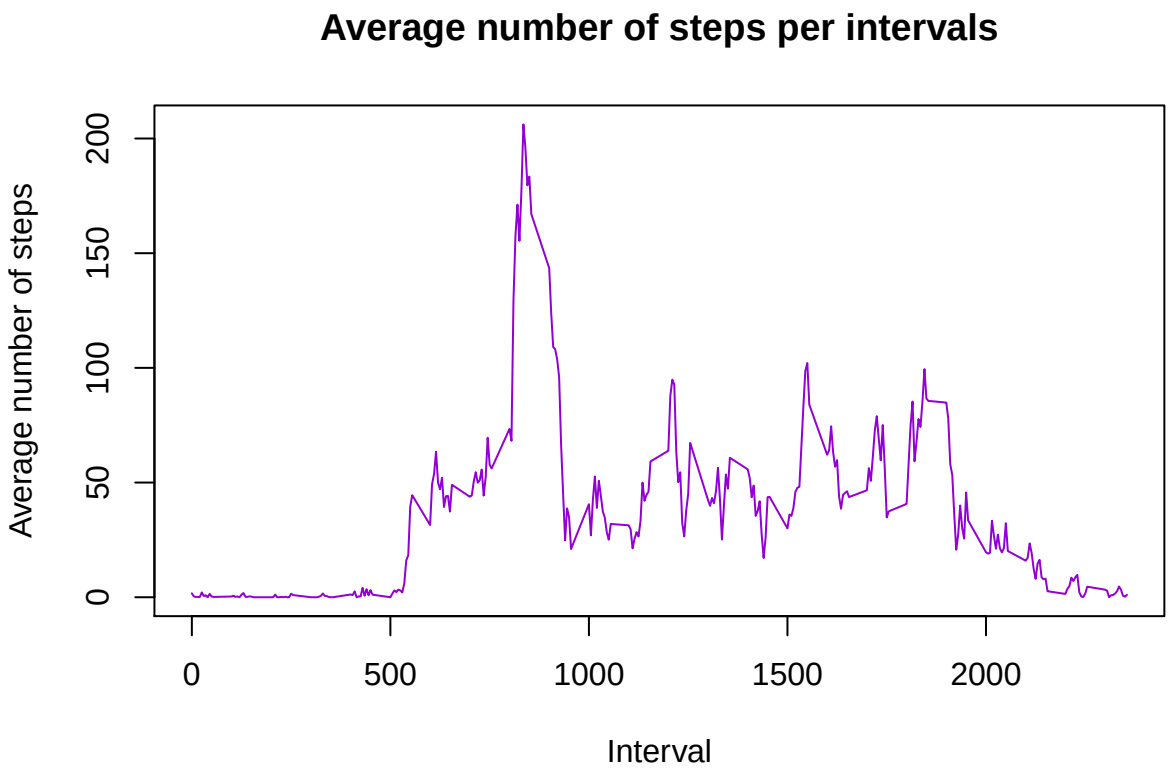

```
#mean of the total number of steps taken per day

mean(TOTALstepday$steps)
```

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

```
#median of the total number of steps taken per day

median(TOTALstepday$steps)
```

```
## [1] 10765
```

```
## 2 - What is the average daily activity pattern?

averagedailyact <- aggregate(steps~interval, activity, mean)

with(averagedailyact, plot(interval, steps, type = "l", main = "Average number of steps per intervals",
                           ylab = "Average number of steps", xlab = "Interval", col = "darkviolet"))
```




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

```
averagedailyact[which.max(averagedailyact[,2]),1]
```

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

```
## 3 - Imputing missing values
```

```
missingvalues <- is.na(activity[,1])
```

```
stepmeaninterval <- mean(averagedailyact$steps)
```

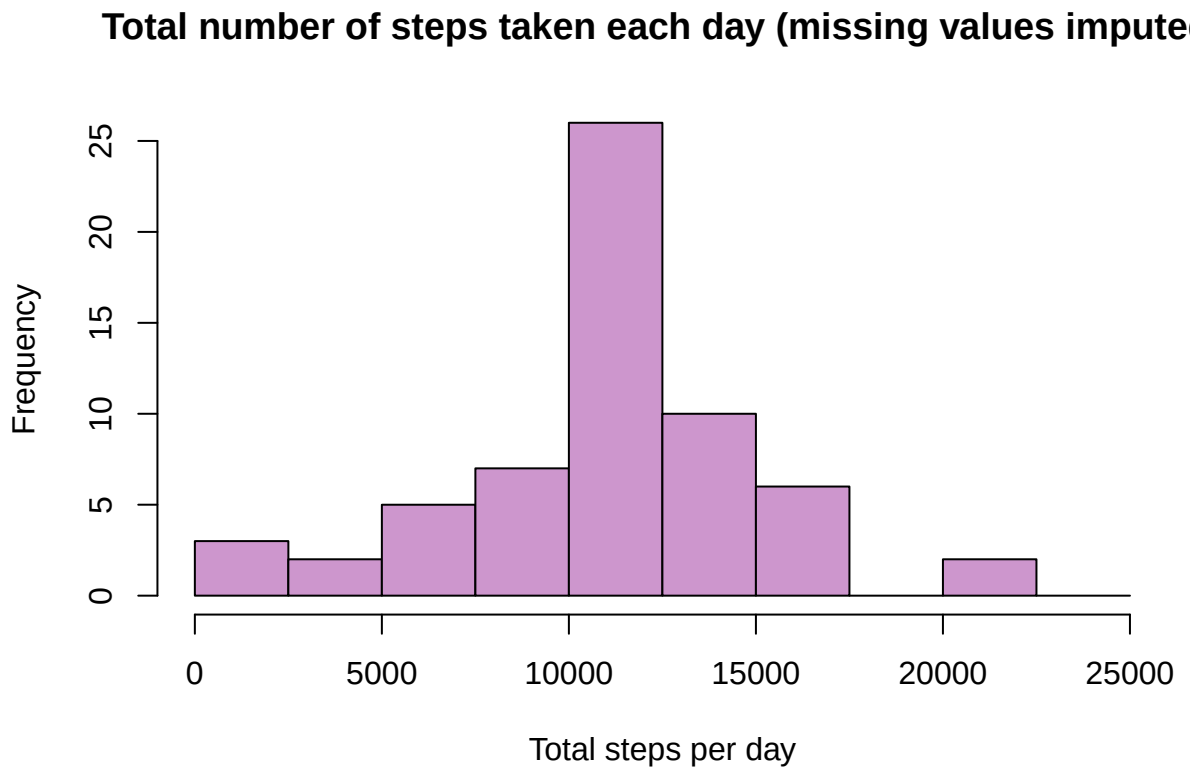
```
activityfilled <- activity
```

```
activityfilled[missingvalues,1] <- stepmeaninterval
```

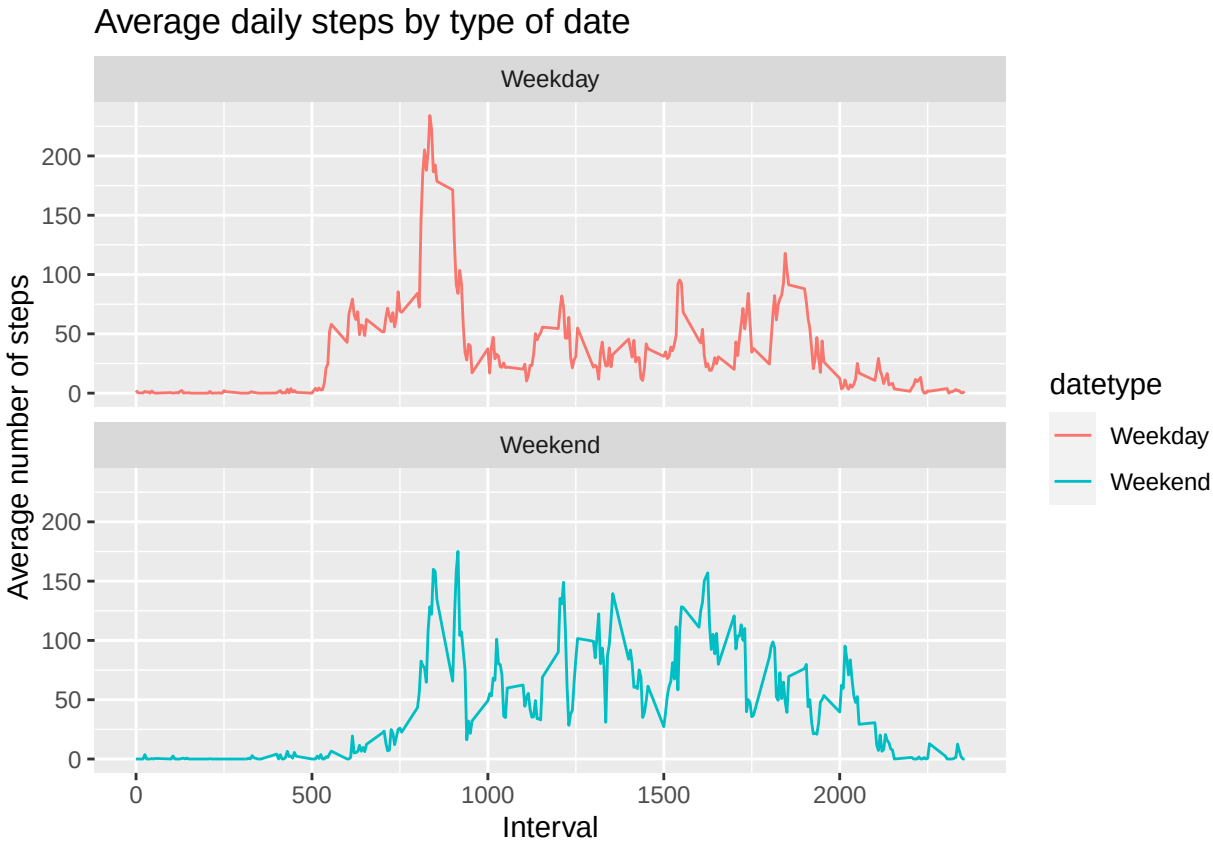
```
head(activityfilled)
```

```
##     steps       date interval
## 1 37.3826 2012-10-01        0
## 2 37.3826 2012-10-01        5
## 3 37.3826 2012-10-01       10
## 4 37.3826 2012-10-01       15
## 5 37.3826 2012-10-01       20
## 6 37.3826 2012-10-01       25
```

```r
TOTALstepday_filled <- aggregate(steps~date, activityfilled, sum)

hist(TOTALstepday_filled$steps, main = "Total number of steps taken each day (missing values imputed)",
     xlab = "Total steps per day", ylab = "Frequency", col = "plum3", breaks = seq(0,25000, by=2500))
```




```r
## 4 - Are there differences in activity patterns between weekdays and weekends?


activity$date <- as.Date(strptime(activity$date, format="%Y-%m-%d"))
activity$datetype <- sapply(activity$date, function(x) {
  if (weekdays(x) == "sábado" | weekdays(x) =="domingo")
  {y <- "Weekend"} else
  {y <- "Weekday"}
  y
})

activityDATE <- aggregate(steps~interval + datetype, activity, mean, na.rm = TRUE)

ggplot(activityDATE, aes(x = interval , y = steps, color = datetype)) +
  geom_line() +
  labs(title = "Average daily steps by type of date", x = "Interval", y = "Average number of steps") +
  facet_wrap(~datetype, ncol = 1, nrow = 2)
```


Average daily steps by type of date
Weekday
Weekend
Average number of steps
0
50
100
150
200
0
500
1000
1500
2000
Interval
datetype
Weekday
Weekend