

Peer-graded-assgnm-1

pz266

17 05 2022

```
#Load packages
```

```
knitr::opts_chunk$set(echo = TRUE)
library(knitr)
library(ggplot2)
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.3

##
## Attaching package: 'dplyr'

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

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

Reading dataset & processing it

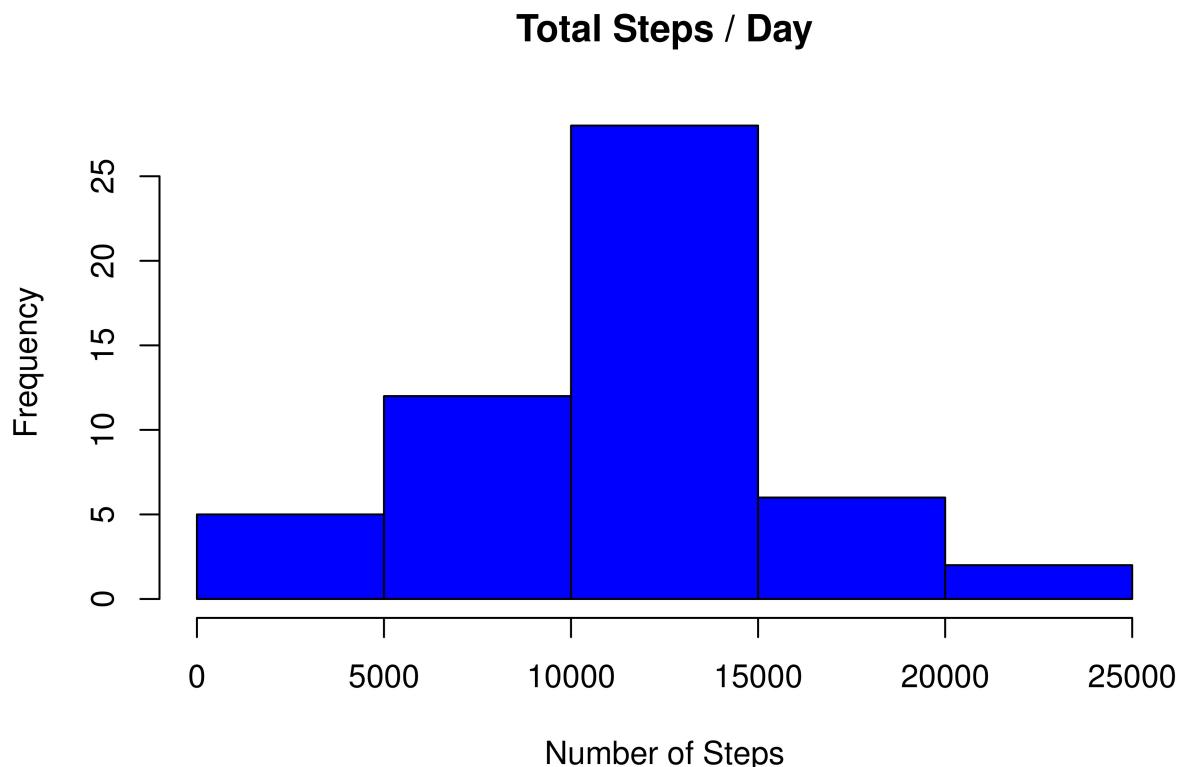
```
knitr::opts_chunk$set(echo = TRUE)

if(!file.exists("getdata-projectfiles-UCI HAR Dataset.zip")) {
  temp <- tempfile()
  download.file("http://d396qusza40orc.cloudfront.net/reddata%2Factivity.zip",temp)
  unzip(temp)
  unlink(temp)
}

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

Histogram of total N (steps) / day

```
knitr::opts_chunk$set(echo = TRUE)
steps_by_day <- aggregate(steps ~ date, data, sum)
hist(steps_by_day$steps, main = paste("Total Steps / Day"), col="blue", xlab="Number of Steps")
```



Mean N(steps) / day and median

```
knitr::opts_chunk$set(echo = TRUE)
rmean <- mean(steps_by_day$steps)
rmean
```

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

```
rmedian <- median(steps_by_day$steps)
rmedian
```

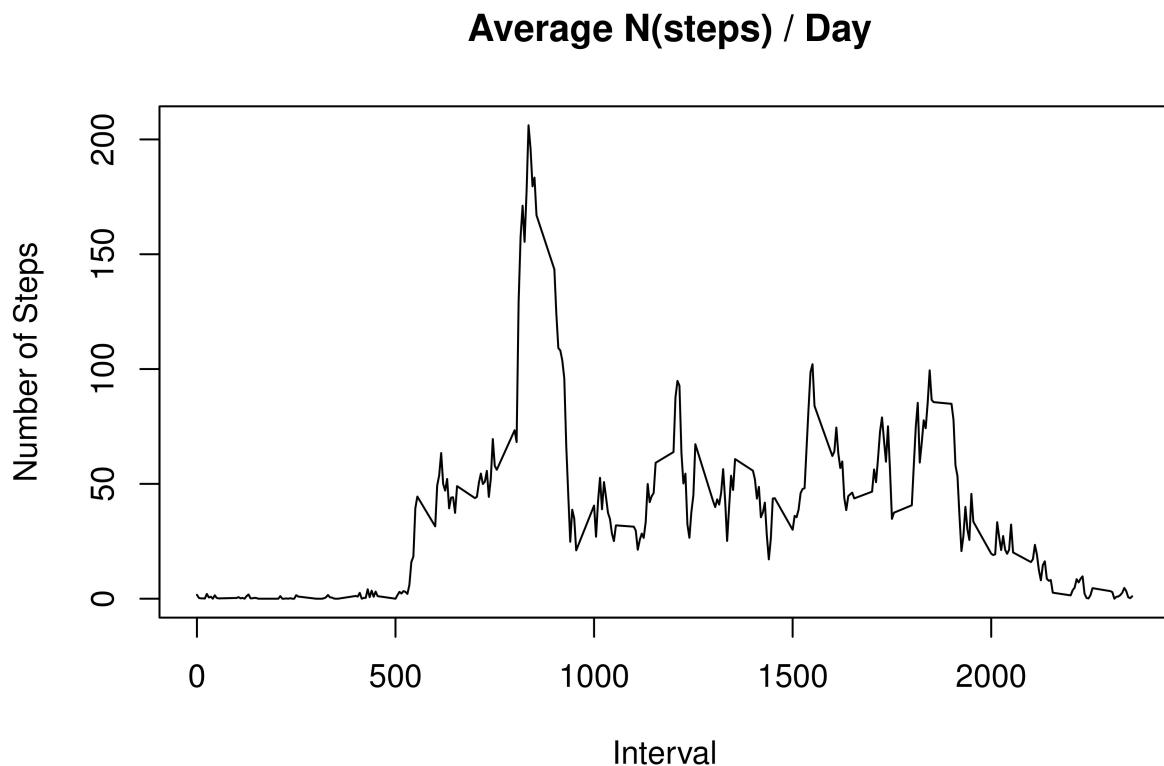
```
## [1] 10765
```

Plot of the average N (steps) / day

```

knitr::opts_chunk$set(echo = TRUE)
steps_by_interval <- aggregate(steps ~ interval, data, mean)
par(bg = 'white')
plot(steps_by_interval$interval, steps_by_interval$steps, type="l", xlab="Interval",
     ylab="Number of Steps", main=" Average N(steps) / Day")

```



5-minute interval average Maximum N(steps)

```

knitr::opts_chunk$set(echo = TRUE)
max_interval <- steps_by_interval[which.max(steps_by_interval$steps),1]
max_interval

## [1] 835

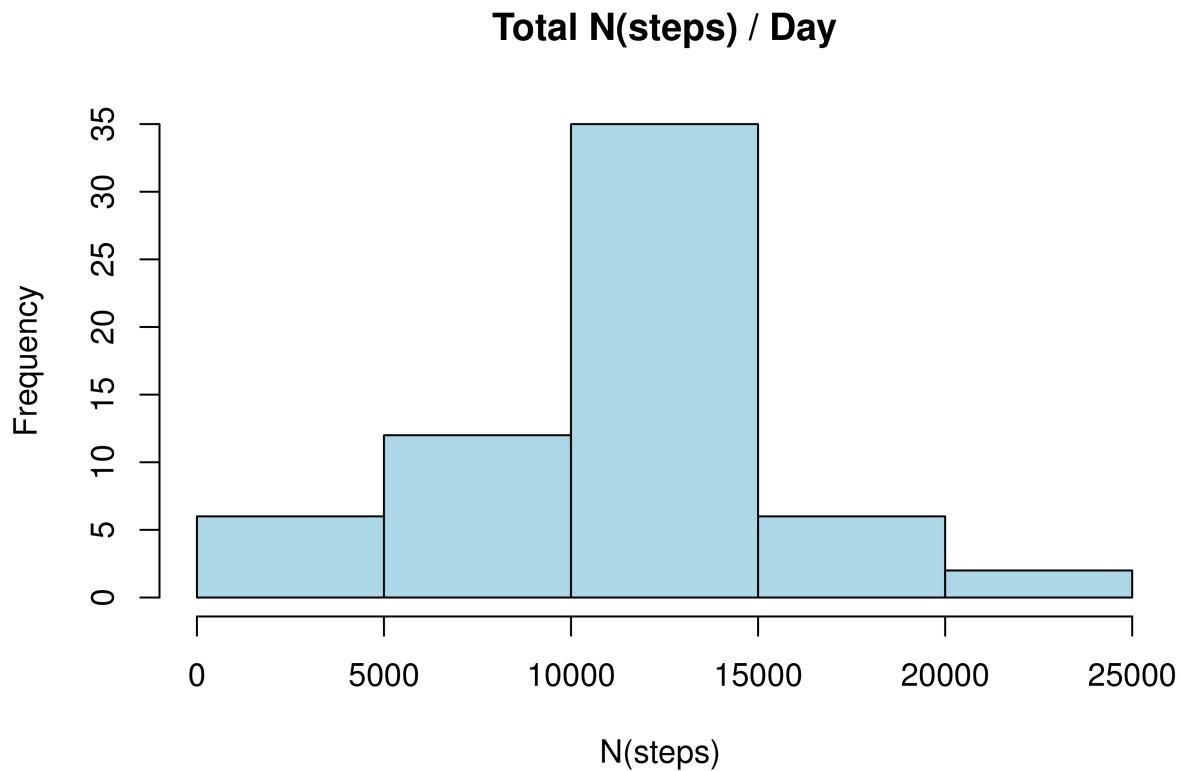
#imputing missing data

knitr::opts_chunk$set(echo = TRUE)
incomplete <- sum(!complete.cases(data))
imputed_data <- transform(data, steps = ifelse(is.na(data$steps), steps_by_interval$steps[match(data$date,
imputed_data[as.character(imputed_data$date) == "2012-10-01", 1] <- 0

```

Histogram of the Total N(steps) / day, imputed miss. values

```
knitr::opts_chunk$set(echo = TRUE)
steps_by_day_i <- aggregate(steps ~ date, imputed_data, sum)
hist(steps_by_day_i$steps, main = paste("Total N(steps) / Day"), col="Lightblue", xlab="N(steps) ")
```



Average N(steps) taken per 5-minute interval, weekdays & weekends

New mean and median for imputed data

```
knitr::opts_chunk$set(echo = TRUE)
rmean.i <- mean(steps_by_day_i$steps)
rmedian.i <- median(steps_by_day_i$steps)
```

Difference between imputed and non-imputed data

```
knitr::opts_chunk$set(echo = TRUE)
mean_diff <- rmean.i - rmean
med_diff <- rmedian.i - rmedian
```

Tot.difference

```
knitr::opts_chunk$set(echo = TRUE)
total_diff <- sum(steps_by_day_i$steps) - sum(steps_by_day$steps)

weekdays <- c("Monday", "Tuesday", "Wednesday", "Thursday",
             "Friday")
imputed_data$dow = as.factor(ifelse(is.element(as.Date(imputed_data$date)), weekdays), "Weekday")

steps_by_interval_i <- aggregate(steps ~ interval + dow, imputed_data, mean)

library(lattice)

xyplot(steps_by_interval_i$steps ~ steps_by_interval_i$interval | steps_by_interval_i$dow, main="Average N(steps) / Day")
```

Average N(steps) / Day

