

# Steps

D. Mendez

14/09/2020

## Loading and preprocessing the data

Both dplyr and lubridate libraries are loaded. The data is read and stored in raw.data and the date data is turned into date format from character

```
library(dplyr)
```

```
##  
## 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
```

```
library(lubridate)
```

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

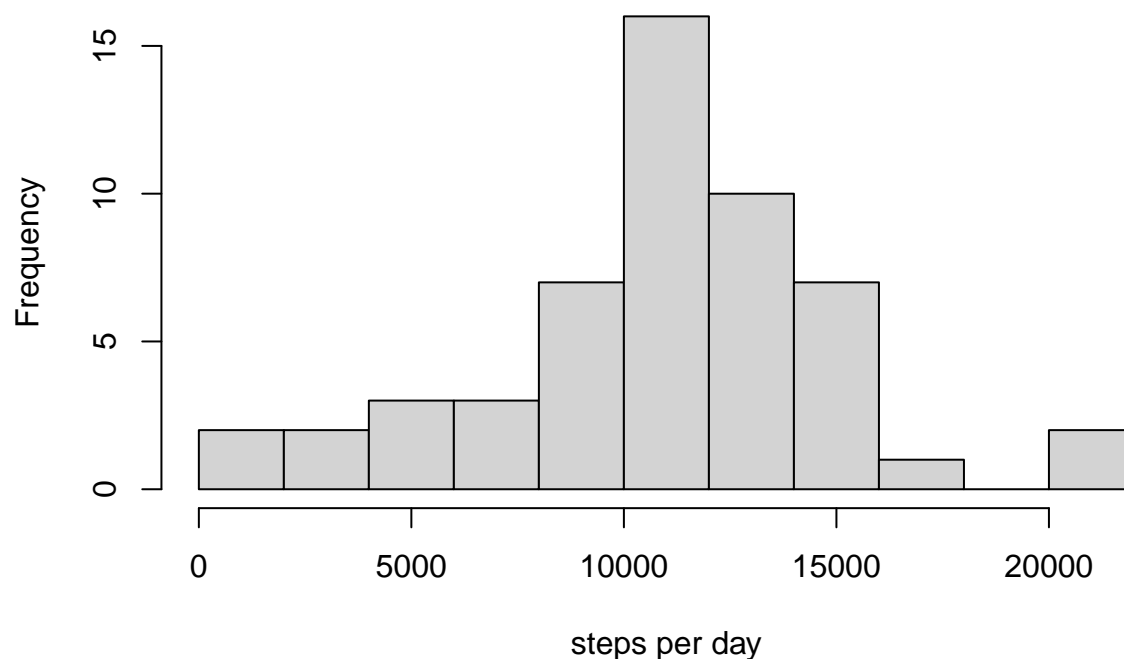
```
raw.data <- read.csv("activity.csv")  
raw.data$date <- ymd(raw.data$date)
```

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

An aggregate of steps per day is made to plot it in a histogram

```
steps.per.day <- aggregate(raw.data$steps,by=list(raw.data$date),FUN=sum,na.action=0)  
hist(steps.per.day$x,main="Histogram of steps per day",xlab="steps per day",breaks=10)
```

## Histogram of steps per day



The median steps per day is

```
median(steps.per.day$x, na.rm=TRUE)
```

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

And the mean steps per day is

```
mean(steps.per.day$x, na.rm=TRUE)
```

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

## Imputing missing values

The total of missing values is

```
sum(is.na(raw.data$steps))
```

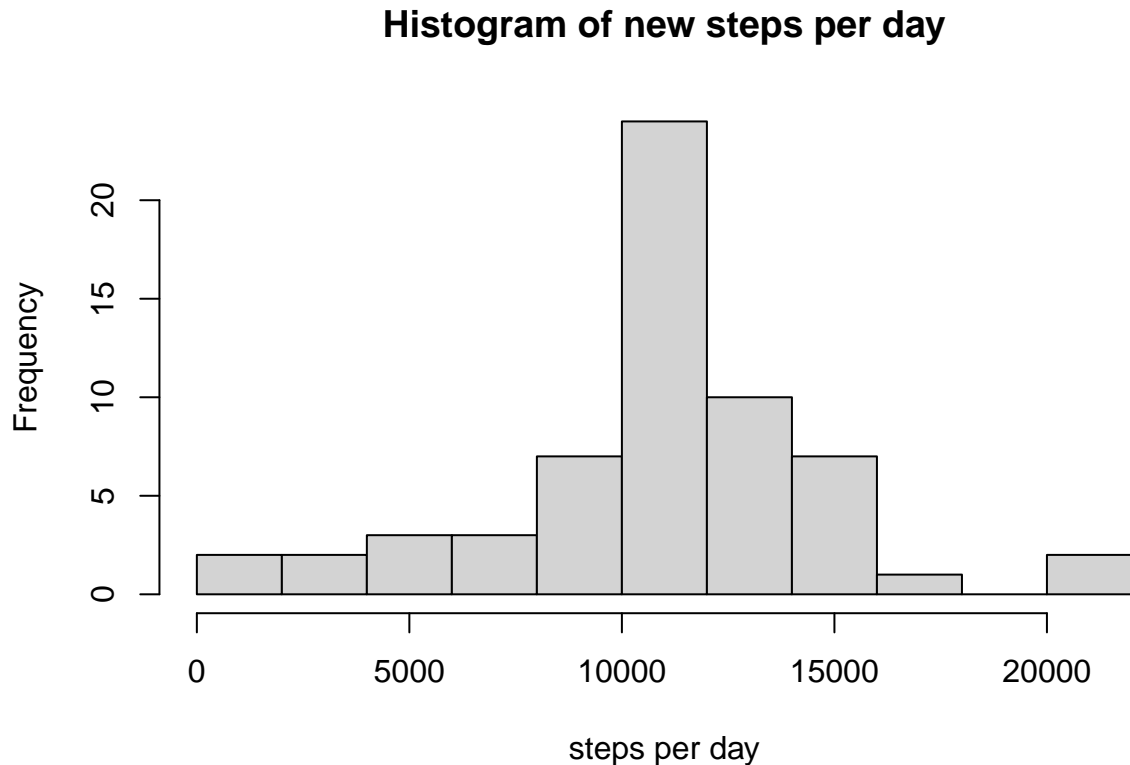
```
## [1] 2304
```

The strategy for filling that missing data will be to have it be equal to the average of steps per day /24/60\*5 for that interval That is saved in new.data

```
new.data <- raw.data
new.data$steps[is.na(raw.data$steps)] <- mean(steps.per.day$x,na.rm=TRUE)/24/60*5
```

Now we plot histogram of the new data to check if it's different

```
steps.per.day.new <- aggregate(new.data$steps,by=list(new.data$date),FUN=sum,na.action=0)
hist(steps.per.day.new$x,main="Histogram of new steps per day",xlab="steps per day",breaks=10)
```



The median steps per day is

```
median(steps.per.day.new$x,na.rm=TRUE)
```

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

And the mean steps per day is

```
mean(steps.per.day.new$x,na.rm=TRUE)
```

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

**Are there differences in activity patterns between weekdays and weekends?**

I'm creating a factor to indicate if the day is weekday or weekend

```

weekday <- c('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday')
new.data$day <- factor((weekdays(new.data$date) %in% weekday),
  levels=c(FALSE, TRUE), labels=c('weekend', 'weekday'))
par(mfrow=c(2,1))
new.data %>%
  filter(day=="weekend") %>%
  group_by(interval) %>%
  summarise(mean=mean(steps)) %>%
  plot(type="l",main="Weekend")

```

## 'summarise()' ungrouping output (override with '.groups' argument)

```

new.data %>%
  filter(day=="weekday") %>%
  group_by(interval) %>%
  summarise(mean=mean(steps)) %>%
  plot(type="l",main="weekday")

```

## 'summarise()' ungrouping output (override with '.groups' argument)

