Course 5 week 2

Jaime Gómez

2024-08-27

```
library(magrittr)
```

```
## Warning: package 'magrittr' was built under R version 4.1.3
```

```
library(dplyr)
```

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

```
##
## Attaching package: 'dplyr'
```

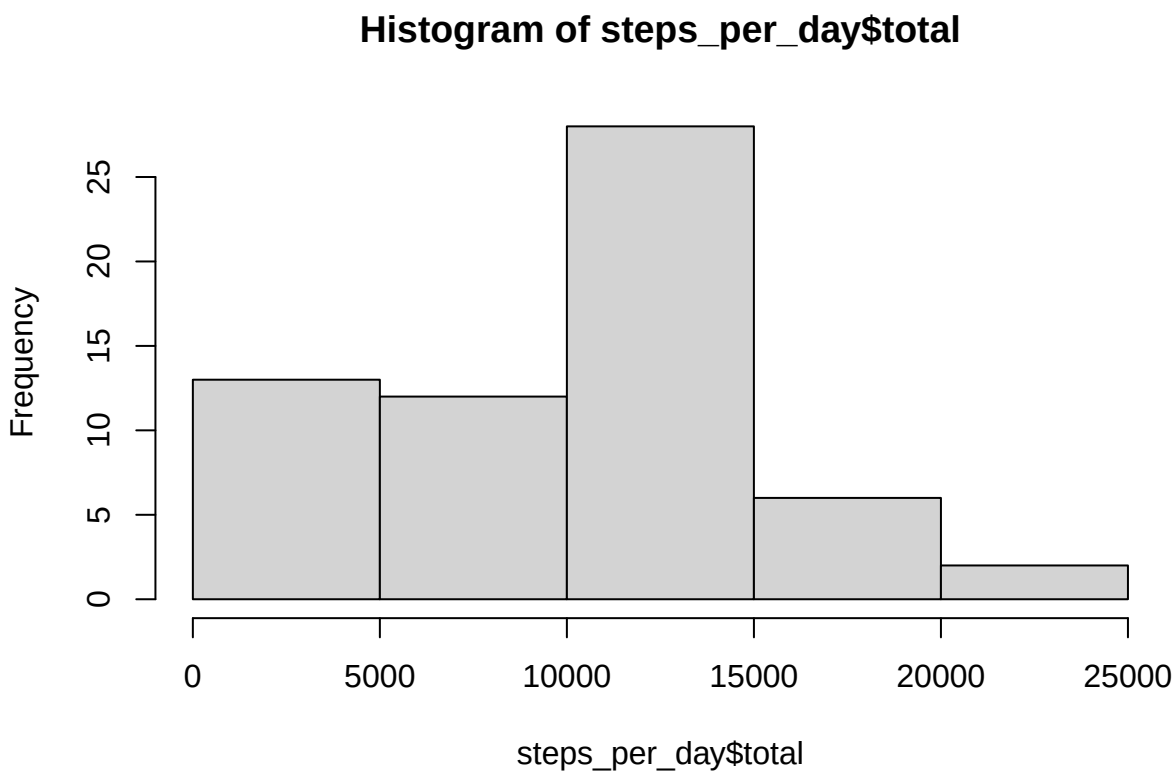
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
activity<-read.csv("C:/Users/jgomez/Desktop/Cursos 2024/activity.csv")
steps_per_day <- activity %>% group_by(date) %>% summarise(total=sum(steps, na.rm = T))
hist(steps_per_day$total)
```



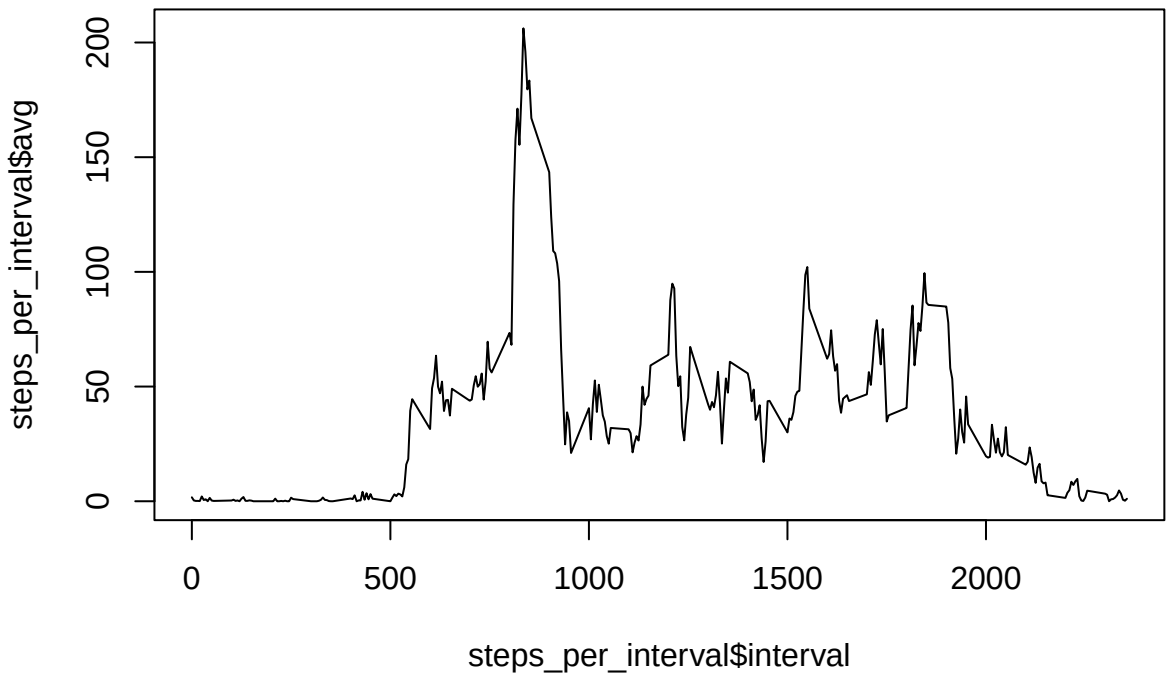

```
mean_steps <- mean(steps_per_day$total, na.rm=T)
print(mean_steps)
```

```
## [1] 9354.23
```

```
median_steps <- median(steps_per_day$total, na.rm = T)
print(median_steps)
```

```
## [1] 10395
```

```
steps_per_interval <- activity %>% group_by(interval) %>% summarise(avg=mean(steps, na.rm = T))
plot<-plot(steps_per_interval$interval, steps_per_interval$avg, type="l")
```




```
max <- steps_per_interval %>% subset(steps_per_interval$avg==max(steps_per_interval$avg))
print(max)
```

```
## # A tibble: 1 x 2
##   interval   avg
##      <int> <dbl>
## 1      835  206.
```

Imputing missing values

We‘re going to replace the Nas with the median of the corresponding interval

```
na<-summary(activity)
print(na)
```

```
##      steps            date              interval     
##  Min.   :  0.00   Length:17568       Min.   :   0.0  
##  1st Qu.:  0.00   Class :character   1st Qu.: 588.8  
##  Median :  0.00   Mode  :character   Median :1177.5  
##  Mean   : 37.38                      Mean   :1177.5  
##  3rd Qu.: 12.00                      3rd Qu.:1766.2  
##  Max.   :806.00                      Max.   :2355.0  
##  NA's   :2304
```

```
imputing <- activity %>% group_by(interval) %>% summarise(med=median(steps, na.rm = T))
activity_imputing <- merge(activity, imputing, by.x = "interval")

activity_imputing <- activity_imputing %>% mutate(steps= if_else(is.na(steps), med, steps))

steps_per_day_imputed <- activity_imputing %>% group_by(date) %>% summarise(total=sum(steps, na.rm = T))
hist(steps_per_day_imputed$total)
```


```
mean_steps_imputed <- mean(steps_per_day_imputed$total, na.rm=T)
print(mean_steps_imputed)
```

```
## [1] 9503.869
```

```
median_steps_imputed <- median(steps_per_day_imputed$total, na.rm = T)
print(median_steps_imputed )
```

```
## [1] 10395
```

The median is not affected, while the mean it´s a little bit higher

Weekdays

```
activity_imputing$date <- as.Date.character(activity_imputing$date)
activity_imputing <- activity_imputing %>% mutate(weekday_1= weekdays(date))
activity_imputing <- activity_imputing %>% mutate(weekday = if_else(weekday_1 %in% c("sábado", "domingo
steps_per_interval_weekday <- activity_imputing %>% group_by(interval, weekday) %>% summarise(avg=mean(
```

```
## `summarise()` has grouped output by 'interval'. You can override using the
## `.groups` argument.
```

```
ggplot(steps_per_interval_weekday,aes(x=interval, y = avg) )+geom_line()+facet_wrap(facets="weekday")
```