Reproducible Research

Loading necessary packages

```
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(readr)
library(ggplot2)
```

Loading and preprocessing the data

```
data_path <- list.files(pattern = "activity.csv",
                         recursive = TRUE,
                         ignore.case = TRUE)
data <- read_csv(data_path)
```

```
##
## -- Column specification ----------------------------------------------------
## cols(
##   steps = col_double(),
##   date = col_date(format = ""),
##   interval = col_double()
## )
```

What is mean total number of steps taken per day?

```
total_steps <- data %>%
  group_by(date) %>%
  summarise(Total_steps = sum(steps))
```

`summarise()` ungrouping output (override with `.groups` argument)

```
hist(total_steps$Total_steps, xlab = "Total steps", main = "Total number of steps taken per day")
```

Total number of steps taken per day
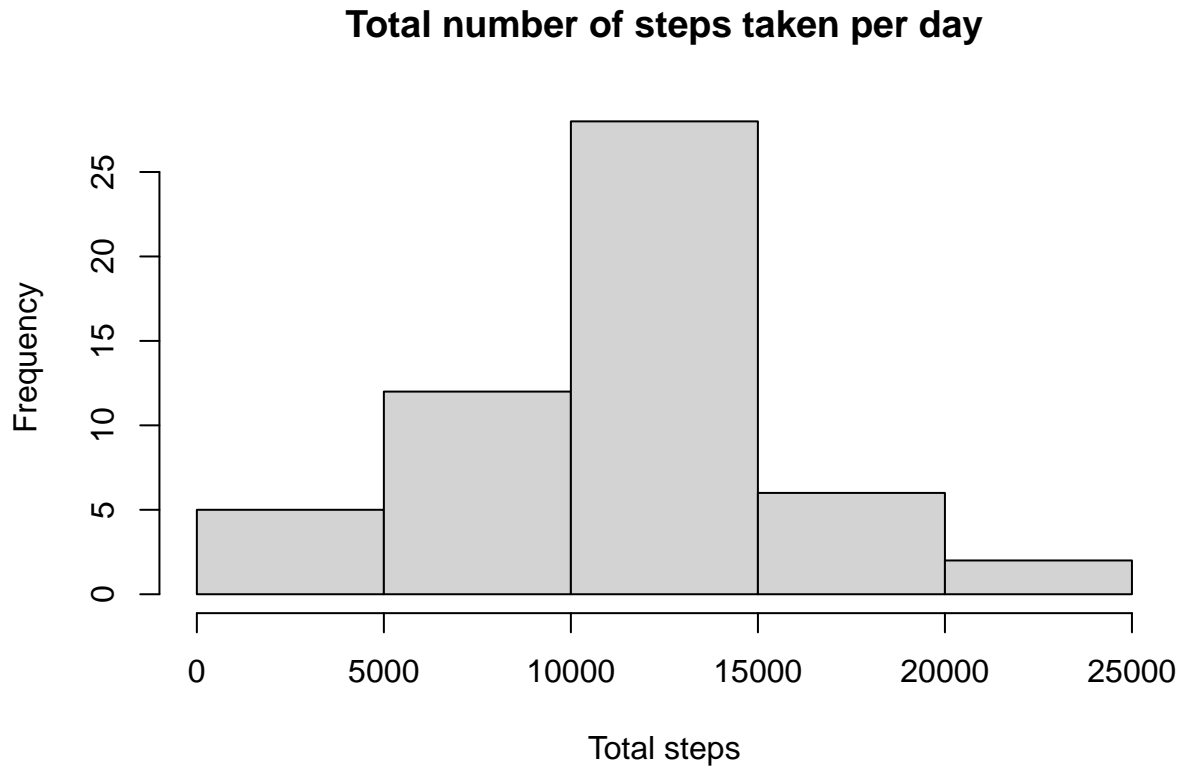

```
mean_total_steps <- mean(total_steps$Total_steps, na.rm = TRUE)
mean_total_steps
```

[1] 10766.19

```
median_total_steps <- median(total_steps$Total_steps, na.rm = TRUE)
median_total_steps
```
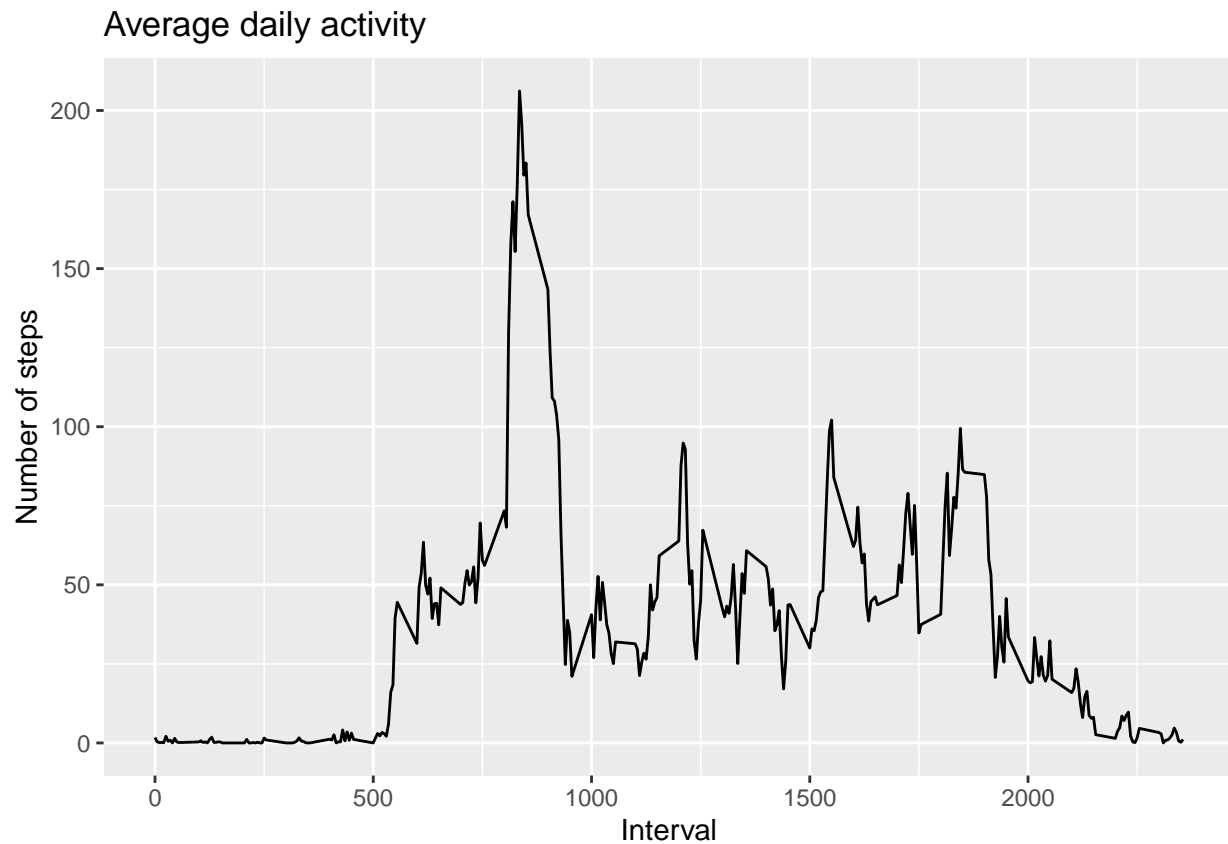
[1] 10765

What is the average daily activity pattern?

```
mean_steps <- data %>%
  group_by(interval) %>%
  summarise(Number_of_steps = mean(steps, na.rm = TRUE))
```

`summarise()` ungrouping output (override with `.groups` argument)

```
g <- ggplot(mean_steps, aes(x = interval, y = Number_of_steps))
g + geom_line() + xlab("Interval") + ylab("Number of steps") + ggtitle("Average daily activity")
```

Average daily activity


```
#interval with maximum mean steps
max_mean_steps <- mean_steps %>%
  filter(Number_of_steps == max(Number_of_steps))
max_mean_steps
```

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

Imputing missing values

```
#total number of missing values
NA_count <- sum(is.na(data))
NA_count
```

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

```r
#filling missing values by mean steps for the interval
data_wo_NA <- data %>%
  merge(mean_steps, by = "interval") %>%
  mutate(steps_wo_NA = if_else(is.na(steps) == TRUE, Number_of_steps, steps))
head(data_wo_NA, 10)
```

```
##    interval steps       date Number_of_steps steps_wo_NA
## 1         0    NA 2012-10-01        1.716981    1.716981
## 2         0     0 2012-11-23        1.716981    0.000000
## 3         0     0 2012-10-28        1.716981    0.000000
## 4         0     0 2012-11-06        1.716981    0.000000
## 5         0     0 2012-11-24        1.716981    0.000000
## 6         0     0 2012-11-15        1.716981    0.000000
## 7         0     0 2012-10-20        1.716981    0.000000
## 8         0     0 2012-11-16        1.716981    0.000000
## 9         0     0 2012-11-07        1.716981    0.000000
## 10        0     0 2012-11-25        1.716981    0.000000
```

```r
total_steps_wo_NA <- data_wo_NA %>%
  group_by(date) %>%
  summarise(Total_steps = sum(steps_wo_NA))
```

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

```r
hist(total_steps_wo_NA$Total_steps, xlab = "Total steps", main = "Total number of steps taken per day wo
```

Total number of steps taken per day wo NA
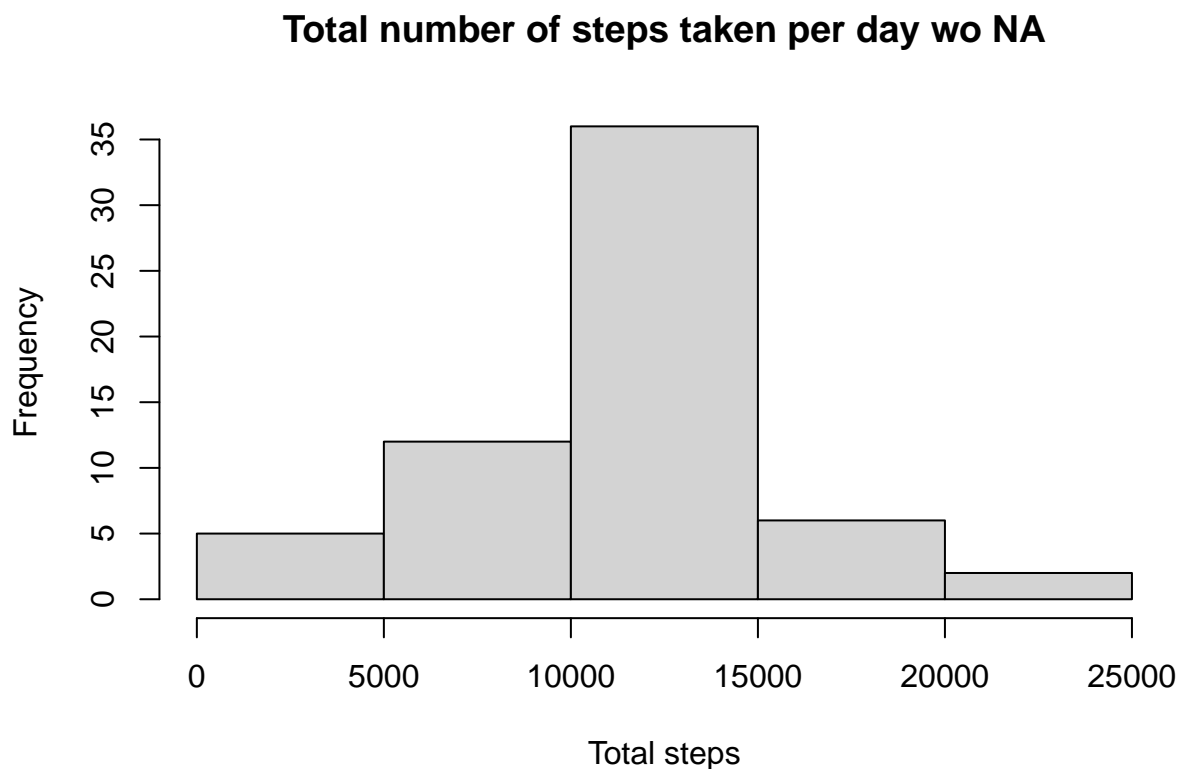
```r
#mean and median of the data without missing values
mean_total_steps_wo_NA <- mean(total_steps_wo_NA$Total_steps)
mean_total_steps_wo_NA
```

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

```r
median_total_steps_wo_NA <- median(total_steps_wo_NA$Total_steps)
median_total_steps_wo_NA
```

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

```r
#absolute differences
abs(mean_total_steps - mean_total_steps_wo_NA)
```

```
## [1] 0
```

```r
abs(median_total_steps - median_total_steps_wo_NA)
```

```
## [1] 1.188679
```

Are there differences in activity patterns between weekdays and weekends?

```r
mean_steps_wo_NA <- data_wo_NA %>%
  group_by(interval, weekend_t = chron::is.weekend(date)) %>%
  summarise(Number_of_steps = mean(steps_wo_NA)) %>%
  mutate(weekend_v = if_else(weekend_t == TRUE, "weekend", "weekday"))
```

```
## 'summarise()' regrouping output by 'interval' (override with '.groups' argument)
```

```r
g <- ggplot(mean_steps_wo_NA, aes(interval, Number_of_steps))
g + geom_line() + facet_grid(.~ weekend_v) + xlab("Interval") + ylab("Number of steps") +
  ggtitle("Total number of steps taken per day without missing values in the weekday and weekend day")
```

Total number of steps taken per day without missing values in the weekday