Submission

2022-09-27

Code for reading in the dataset and/or processing the data

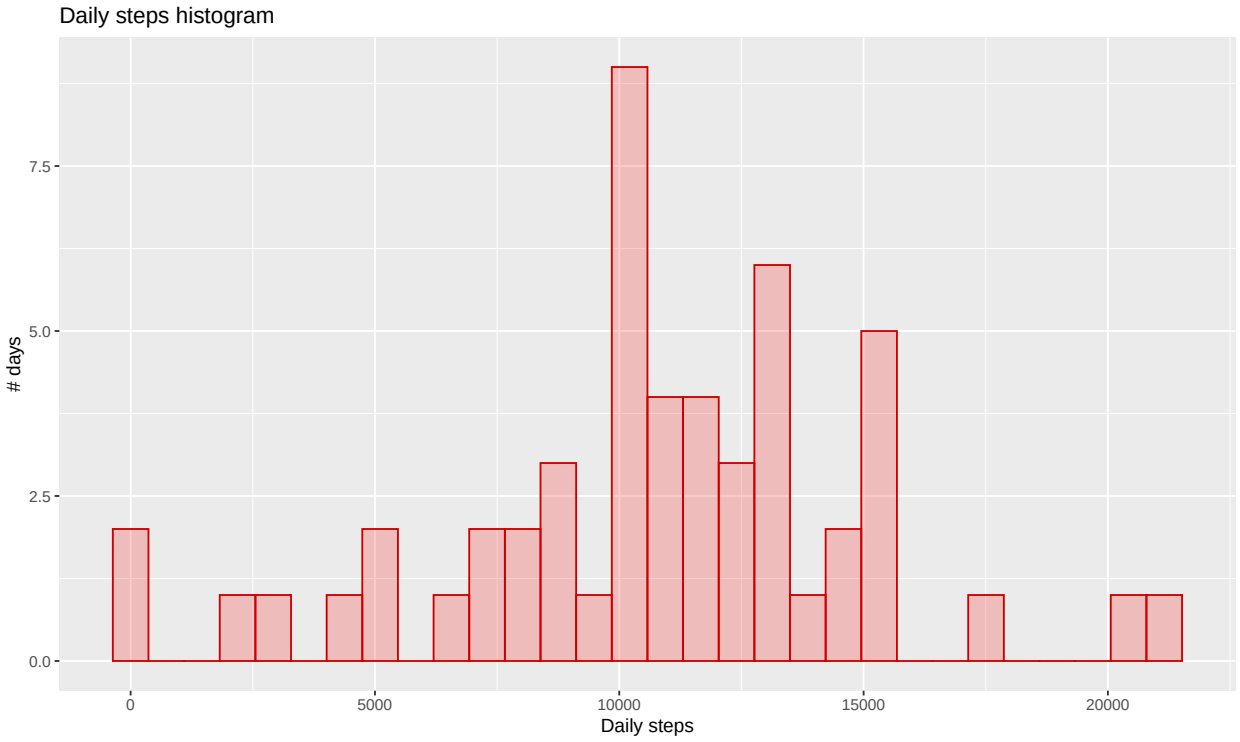
```
df <- read.csv("activity.csv")
```

Histogram of the total number of steps taken each day

```
library(ggplot2)
library(dplyr)

tdf <- df %>%
    filter(!is.na(steps)) %>%
    group_by(date) %>%
    summarise(t_steps = sum(steps))

his <- ggplot(tdf, aes(x = t_steps)) + geom_histogram(color = "red3", fill = "red",
    alpha = 0.2) + labs(title = "Daily steps histogram", y = "# days", x = "Daily steps")
his
```



Mean and median number of steps taken each day

```
tdf %>%
    select(t_steps) %>%
    summarise(mean = mean(t_steps), median = median(t_steps))
```

```
## # A tibble: 1 x 2
##     mean median
##    <dbl>  <int>
## 1 10766.  10765
```

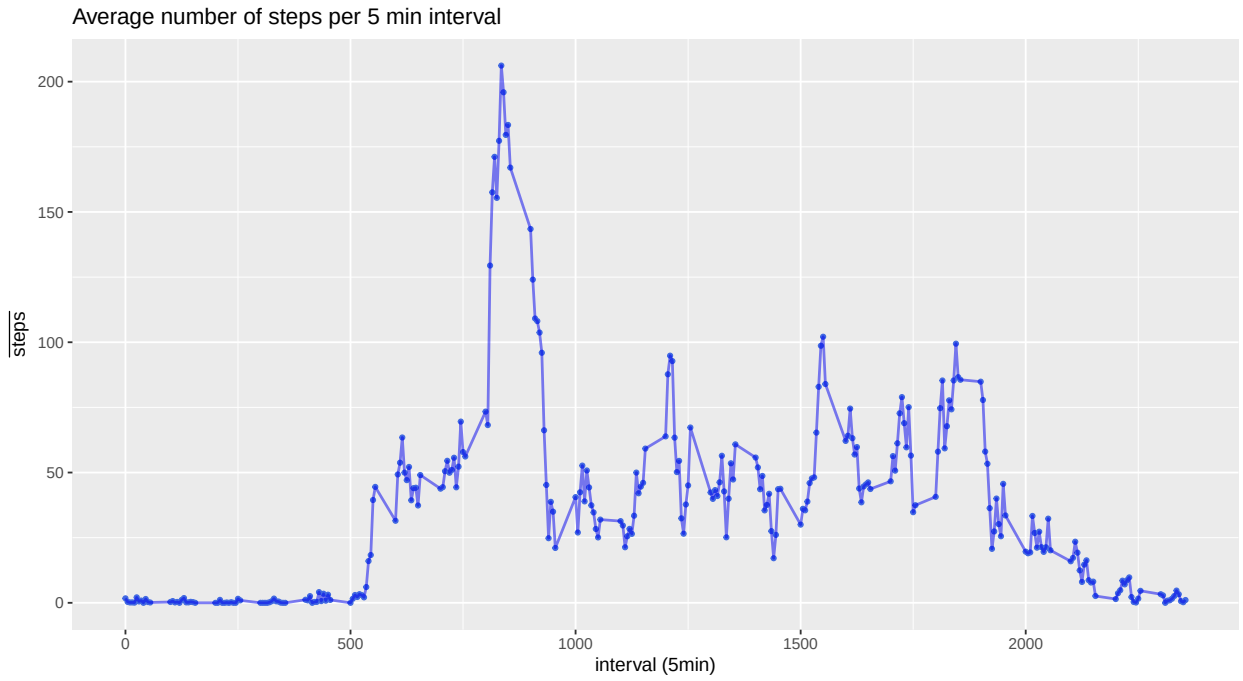
Time series plot of the average number of steps taken

```
library(latex2exp)

ts_df <- df %>%
    filter(!is.na(steps)) %>%
    group_by(interval) %>%
    summarise(avg_steps = mean(steps)) %>%
    mutate(pretty_interval = paste(as.character(interval%/%100),
        "h", as.character(interval%%100)))

ts <- ggplot(ts_df, aes(x = interval, y = avg_steps)) +
    geom_point(size = 1, color = "royalblue", fill = "blue") +
    geom_line(size = 0.75, alpha = 0.5, color = "blue2") +
    labs(title = "Average number of steps per 5 min interval",
        x = "interval (5min)") + ylab(TeX("$\\bar{steps}$"))

ts
```



The 5-minute interval that, on average, contains the maximum number of steps

```
ts_df %>%
    top_n(1, avg_steps) %>%
    select(pretty_interval)
```

```
## # A tibble: 1 x 1
##   pretty_interval
##   <chr>
## 1 8 h 35
```