Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
unzip("repdata_data_activity.zip")
data <- read.csv("activity.csv")
```

What is mean total number of steps taken per day?

```
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.3
```

```
##
## 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
```

```
total_steps <- data %>%
  group_by(date) %>%
  summarise(daily_steps = sum(steps, na.rm = TRUE))
```

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

```
ggplot(total_steps, aes(daily_steps)) + geom_histogram(binwidth = 2000) +
  xlab("Total number of steps taken each day") +
  ylab("Frequency")
```
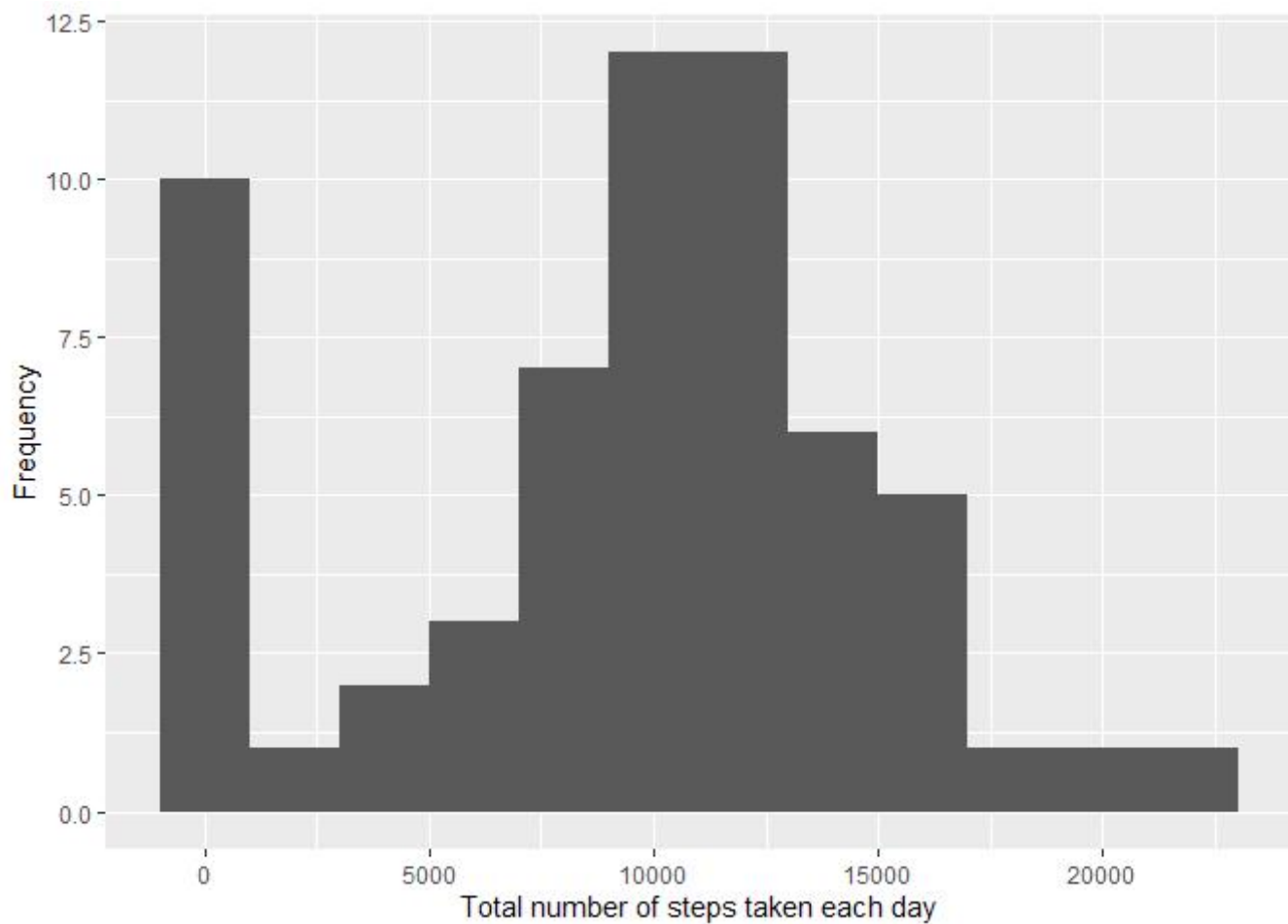
```
mean = mean(total_steps$daily_steps, na.rm=TRUE)
mean
```

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

```
median = median(total_steps$daily_steps, na.rm=TRUE)
median
```
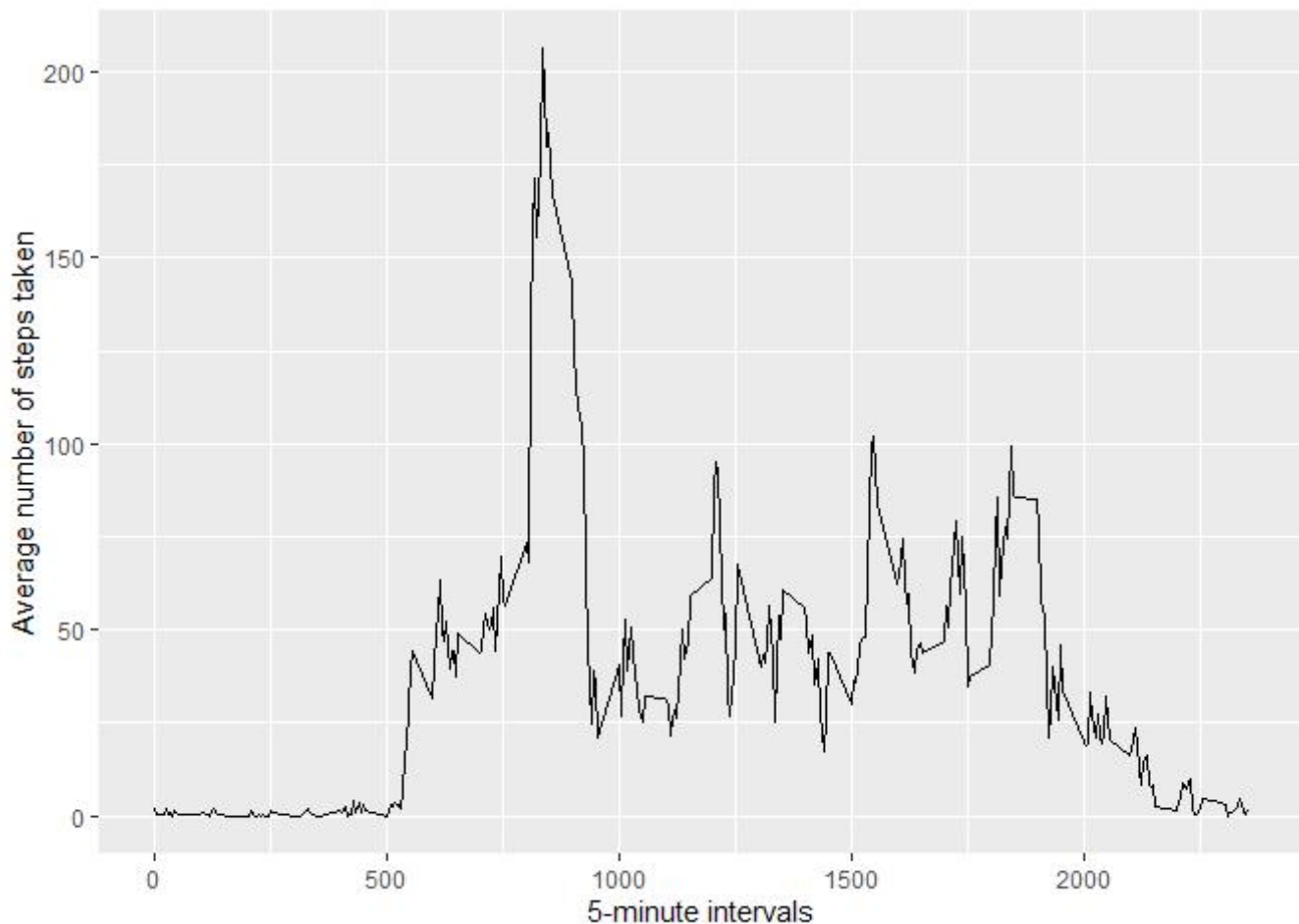
```
## [1] 10395
```

What is the average daily activity pattern?

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

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

```
ggplot(data=interval_steps, aes(x=interval, y=steps)) +
    geom_line() +
    xlab("5-minute intervals") +
    ylab("Average number of steps taken")
```


Imputing missing values

```
missing <- !complete.cases(data)
```

```
imputed_data <- data %>%
  mutate(
    steps = case_when(
      is.na(steps) ~ interval_steps$steps[match(data$interval, interval_steps$interval)],
      TRUE ~ as.numeric(steps)
    ))
```

```
imputed_total_steps <- imputed_data %>% group_by(date) %>% summarise(daily_steps = sum(steps))
```

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

```
ggplot(imputed_total_steps, aes(daily_steps)) +
  geom_histogram(binwidth = 2000) +
  xlab("Total number of steps taken each day") +
  ylab("Frequency")
```
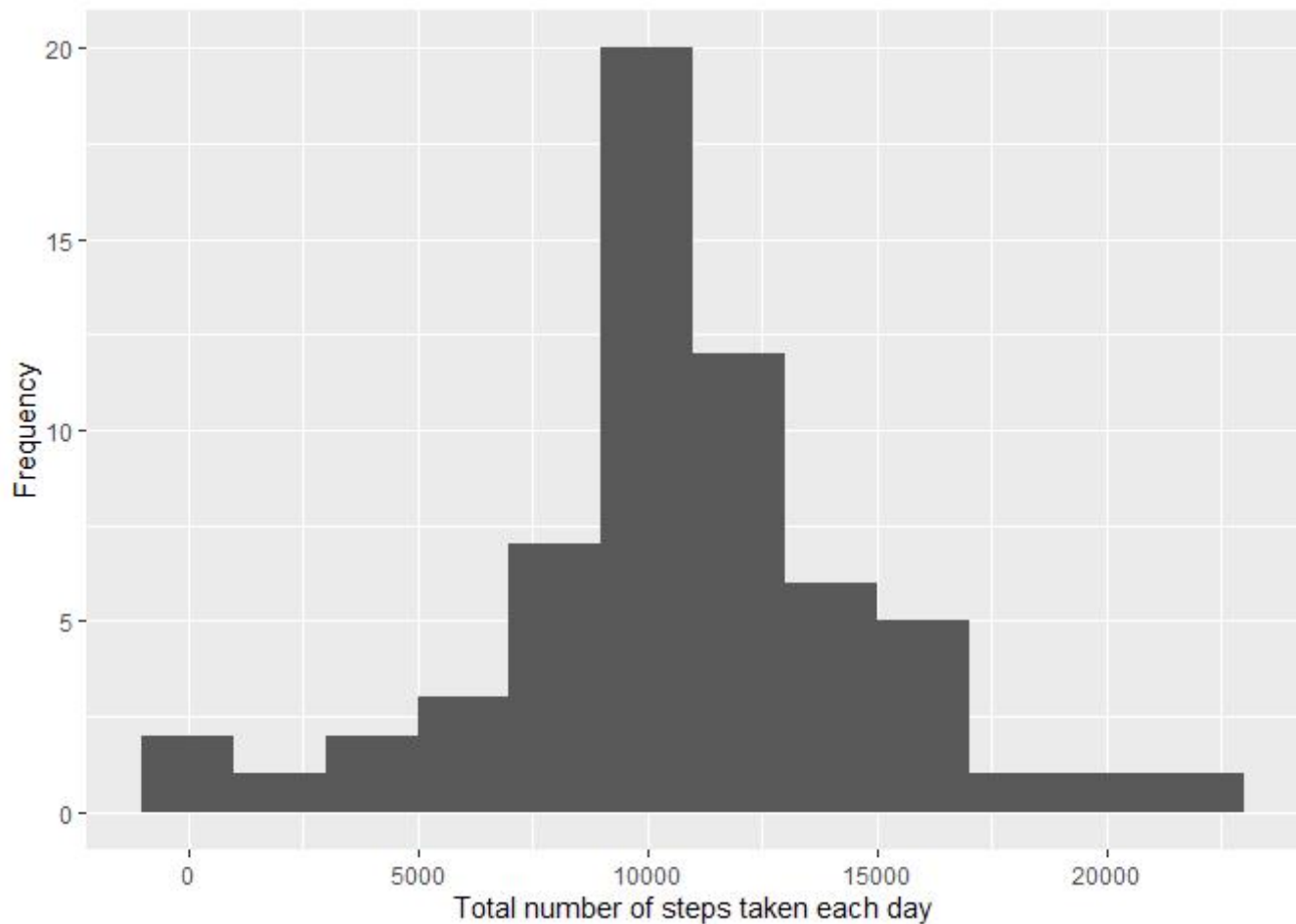


```
imputed_mean = mean(imputed_total_steps$daily_steps, na.rm=TRUE)
imputed_mean
```

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

```
imputed_median = median(imputed_total_steps$daily_steps, na.rm=TRUE)
imputed_median
```

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

```
mean_diff <- imputed_mean - mean
mean_diff
```

```
## [1] 1411.959
```

```
median_diff <- imputed_median - median
median_diff
```

```
## [1] 371.1887
```

Are there differences in activity patterns between weekdays and weekends?

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.0.3
```

```
##
## Attaching package: 'lubridate'
```

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

```
day_of_week <- imputed_data %>%
  mutate(
    date = ymd(date),
    weekday_or_weekend = case_when(wday(date) %in% 2:6 ~ "Weekday",
                                   wday(date) %in% c(1,7) ~ "Weekend")
  ) %>% select(-date) %>%
  group_by(interval, weekday_or_weekend) %>%
  summarise(
    steps = mean(steps)
  )
```

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

```
ggplot(day_of_week, aes(interval, steps)) +
  geom_line() +
  facet_wrap(~weekday_or_weekend, nrow = 2) +
  xlab("5-Minute intervals") +
  ylab("Average number of steps")
```