Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

What is mean total number of steps taken per day?

For this part of the assignment, you can ignore the missing values in the dataset.

1. Make a histogram of the total number of steps taken each day

2. Calculate and report the mean and median total number of steps taken per day

```r
library(knitr)
opts_chunk$set(echo = TRUE)

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
```

```r
library(lubridate)
library(ggplot2)

data <- read.csv("activity.csv", header = TRUE, sep = ',', colClasses = c("numeric",
"character","integer"))

data$date <- ymd(data$date)

steps <- data %>%
  filter(!is.na(steps)) %>%
  group_by(date) %>%
  summarize(steps = sum(steps)) %>%
  print
```

```
## Source: local data frame [53 x 2]
##
##          date steps
##        (time) (dbl)
## 1  2012-10-02   126
## 2  2012-10-03 11352
## 3  2012-10-04 12116
## 4  2012-10-05 13294
## 5  2012-10-06 15420
## 6  2012-10-07 11015
## 7  2012-10-09 12811
## 8  2012-10-10  9900
## 9  2012-10-11 10304
## 10 2012-10-12 17382
## ..         ...   ...
```
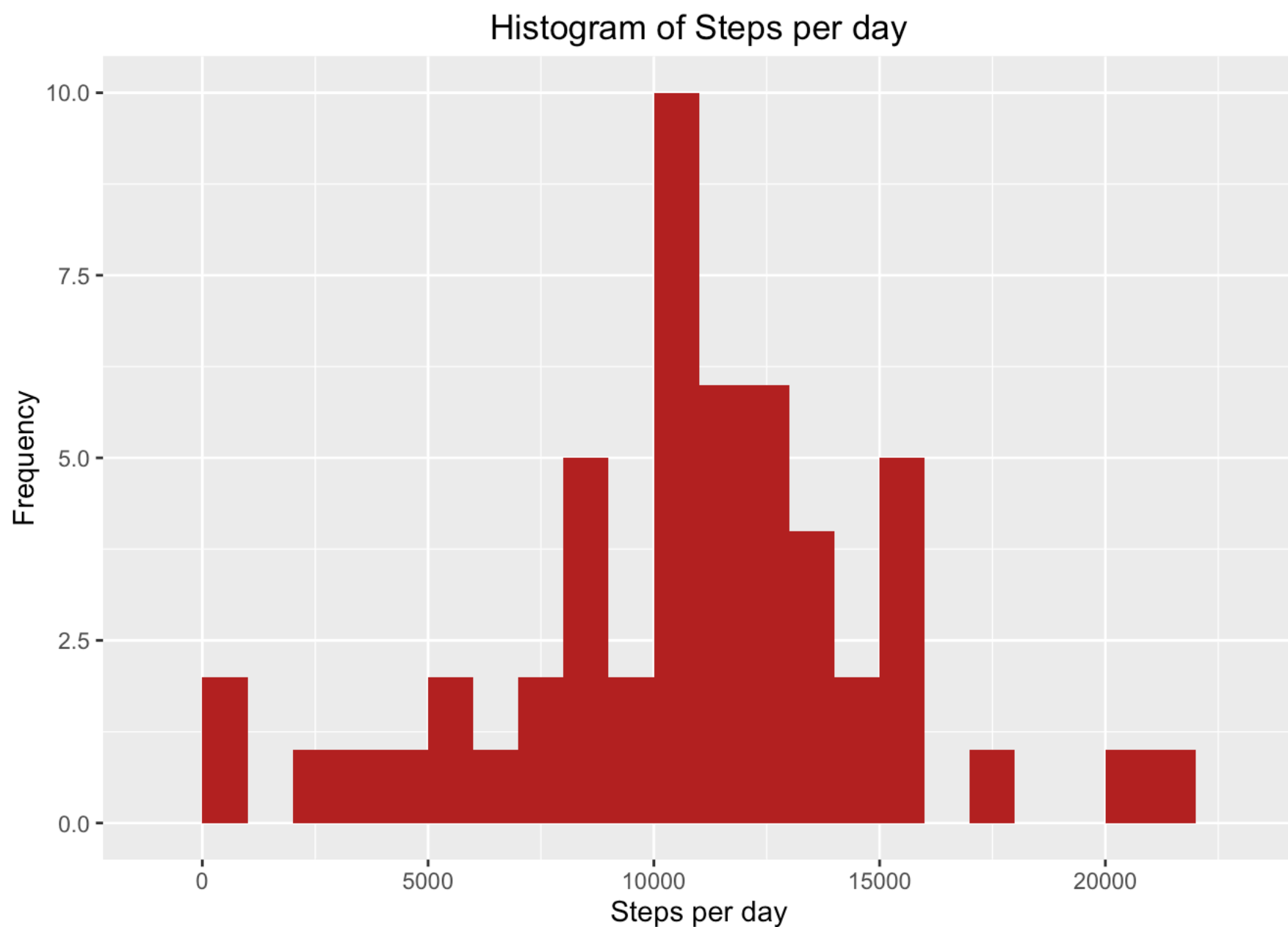
A histogram of the total number of steps taken each day:

Histogram of Steps per day

```
mean_steps <- mean(steps$steps, na.rm = TRUE)
median_steps <- median(steps$steps, na.rm = TRUE)
mean_steps
```
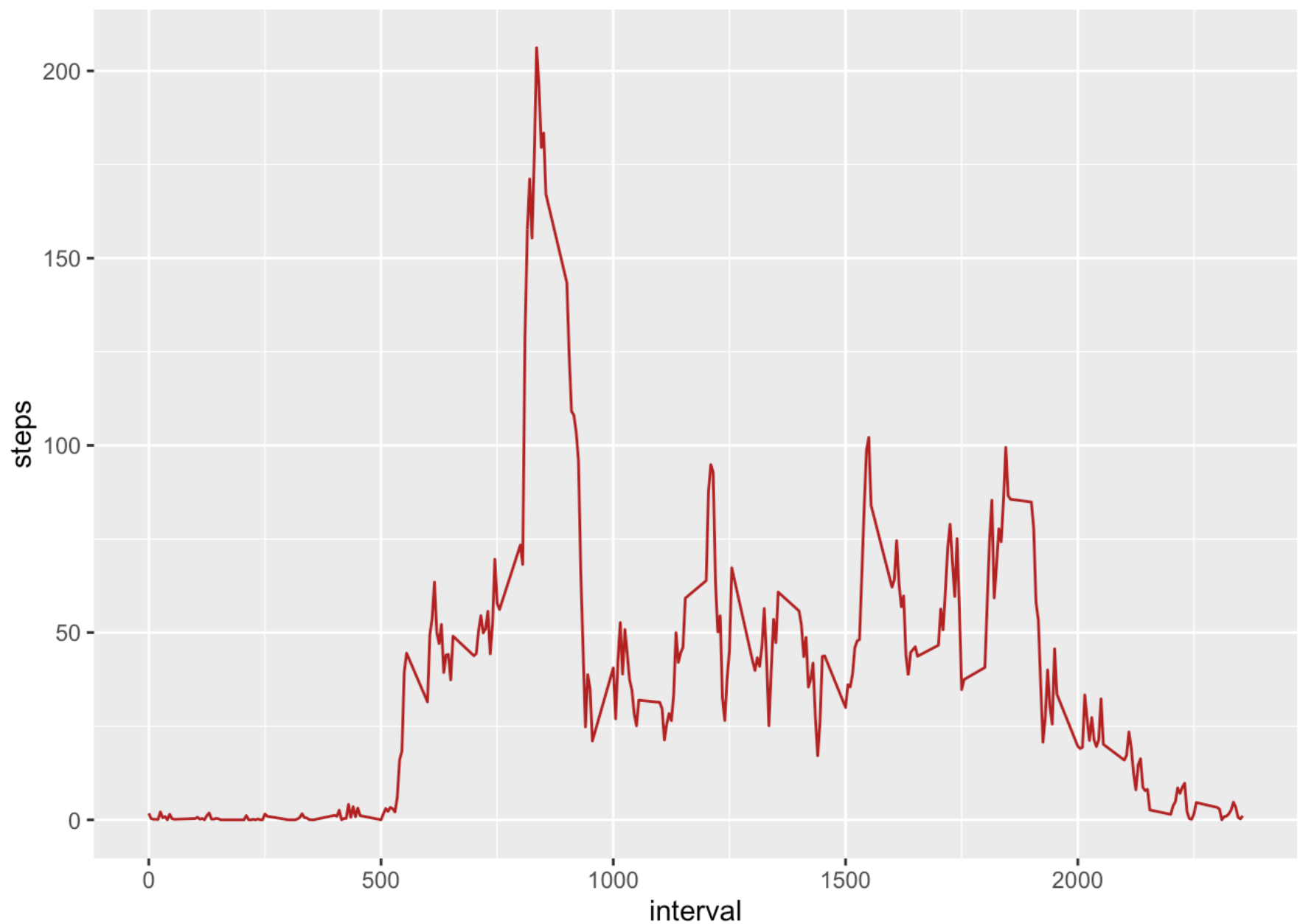
```
## [1] 10766.19
```

```
median_steps
```

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

What is the average daily activity pattern?

```
interval <- data %>%
    filter(!is.na(steps)) %>%
    group_by(interval) %>%
    summarize(steps = mean(steps))
```

The time series of the 5-minute interval and average steps taken:

Use which.max() to find out the maximum steps, on average, across all the days

```
interval[which.max(interval$steps),]
```

```
## Source: local data frame [1 x 2]
##
##   interval     steps
##      (int)     (dbl)
## 1      835  206.1698
```

Imputing missing values

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

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

```
data_full <- data
nas <- is.na(data_full$steps)
avg_interval <- tapply(data_full$steps, data_full$interval, mean, na.rm=TRUE, simplif
y=TRUE)
data_full$steps[nas] <- avg_interval[as.character(data_full$interval[nas])]
```

Calculate the number of steps taken in each 5-minute interval per day using dplyr and group by interval. Use ggplot for making the histogram:

```
steps_full <- data_full %>%
   filter(!is.na(steps)) %>%
   group_by(date) %>%
   summarize(steps = sum(steps)) %>%
   print
```

```
## Source: local data frame [61 x 2]
##
##             date     steps
##           (time)     (dbl)
## 1   2012-10-01 10766.19
## 2   2012-10-02   126.00
## 3   2012-10-03 11352.00
## 4   2012-10-04 12116.00
## 5   2012-10-05 13294.00
## 6   2012-10-06 15420.00
## 7   2012-10-07 11015.00
## 8   2012-10-08 10766.19
## 9   2012-10-09 12811.00
## 10 2012-10-10  9900.00
## ..         ...       ...
```

Create graph:

Histogram of Steps per day, including missing values
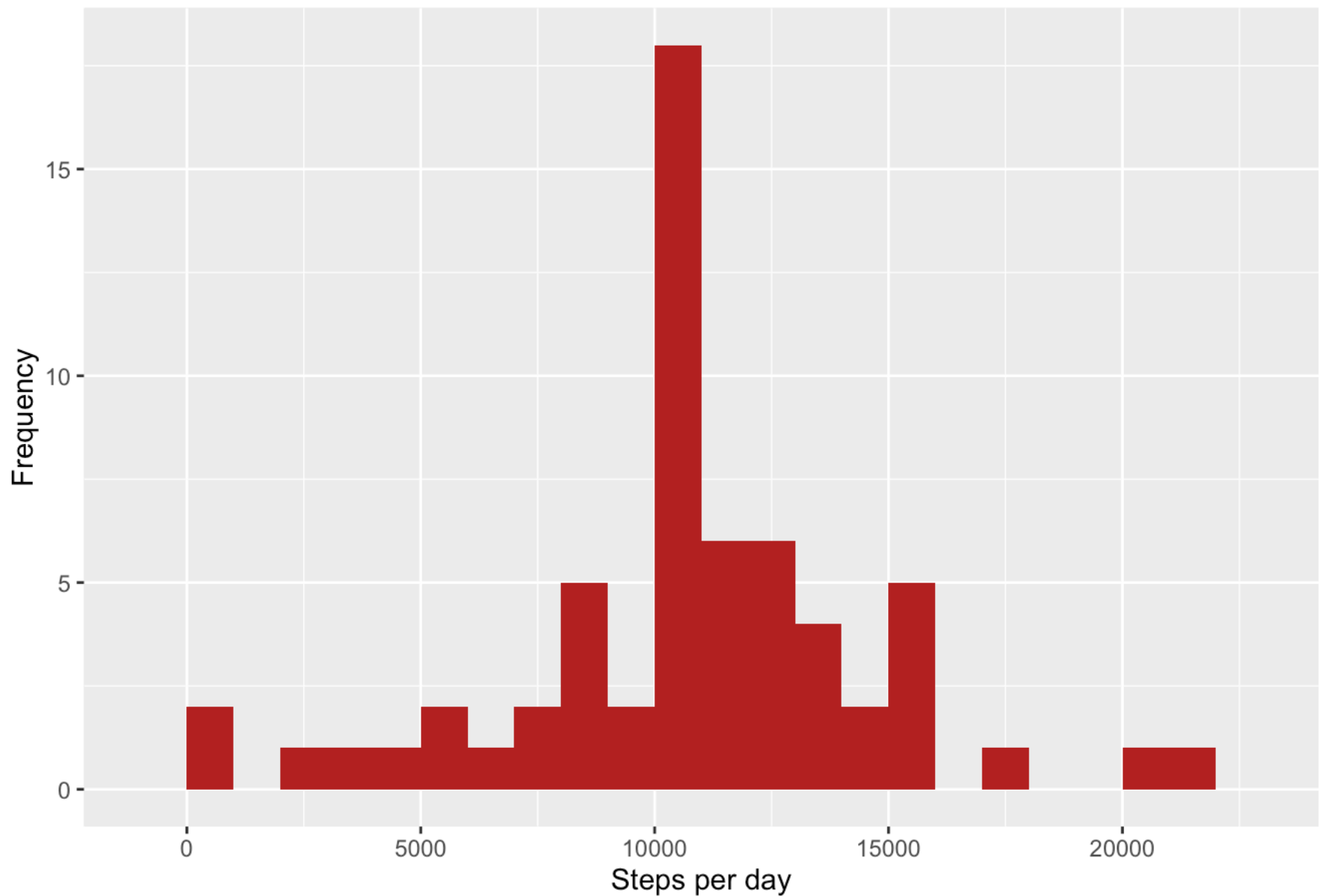
Calculate the mean and median steps with the filled in values:

```
mean_steps_full <- mean(steps_full$steps, na.rm = TRUE)
median_steps_full <- median(steps_full$steps, na.rm = TRUE)
mean_steps_full
```

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

```
median_steps_full
```

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

Are there differences in activity patterns between weekdays and weekends? Use dplyr and mutate to create a new column, weektype, and apply whether the day is weekend or weekday:

```
data_full <- mutate(data_full, weektype = ifelse(weekdays(data_full$date) == "Saturday" | weekdays(data_full$date) == "Sunday", "weekend", "weekday"))
data_full$weektype <- as.factor(data_full$weektype)
head(data_full)
```

```
##         steps        date interval weektype
## 1 1.7169811 2012-10-01        0  weekday
## 2 0.3396226 2012-10-01        5  weekday
## 3 0.1320755 2012-10-01       10  weekday
## 4 0.1509434 2012-10-01       15  weekday
## 5 0.0754717 2012-10-01       20  weekday
## 6 2.0943396 2012-10-01       25  weekday
```

Calculate the average steps in the 5-minute interval and use ggplot for making the time series of the 5-minute interval for weekday and weekend, and compare the average steps:

```
interval_full <- data_full %>%
   group_by(interval, weektype) %>%
   summarise(steps = mean(steps))
s <- ggplot(interval_full, aes(x=interval, y=steps, color = weektype)) + geom_line()
+ facet_wrap(~weektype, ncol = 1, nrow=2)
print(s)
```