Reproducible Research: Peer Assessment 1

```
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(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 3.4.3
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(lubridate)
```

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

```
## The following object is masked from 'package:base':
##
##     date
```

Loading and preprocessing the data

```
#list what is in the zip file
unzip("activity.zip", list=TRUE)
```

```
##           Name Length                Date
## 1 activity.csv 350829 2014-02-11 10:08:00
```

```
#read the activity.csv
input_data <- read.csv(unzip(zipfile="activity.zip"))
```

What is mean total number of steps taken per day?

1. Calculate the total number of steps taken per day
2. If you do not understand the difference between a histogram and a barplot, research the difference between them. Make a histogram of the total number of steps taken each day

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

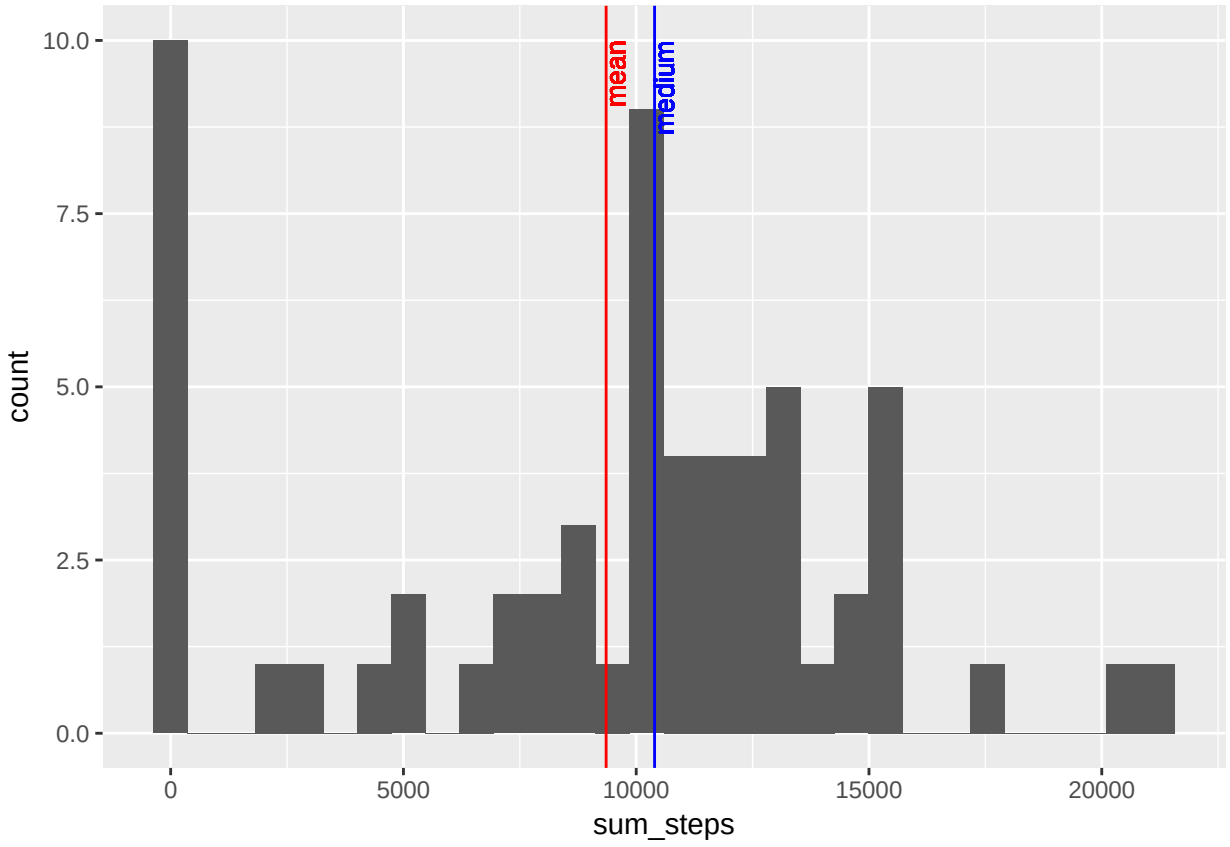
```
Total_Daily_Steps <- input_data %>%
                     group_by(date) %>%
                     summarize(sum_steps = sum(steps, na.rm=TRUE))

mean_daily = round(mean(Total_Daily_Steps$sum_steps),0)
median_daily = round(median(Total_Daily_Steps$sum_steps),0)

#qplot(data=Total_Daily_Steps, x=sum_steps)

ggplot(Total_Daily_Steps,aes(sum_steps)) +
    geom_histogram() +
    geom_vline(xintercept = mean_daily, color="red") +
    geom_text(aes(x=mean_daily, y=10), label='mean', color="red", size=3.5, angle=90, hjust=1, vjust=1)
    geom_vline(xintercept = median_daily, color="blue") +
    geom_text(aes(x=median_daily, y=10),label='medium', color="blue",size=3.5,angle=90, hjust=1, vjust=
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The mean number of steps taken per day is The median number of steps taken per day is

What is the average daily activity pattern?

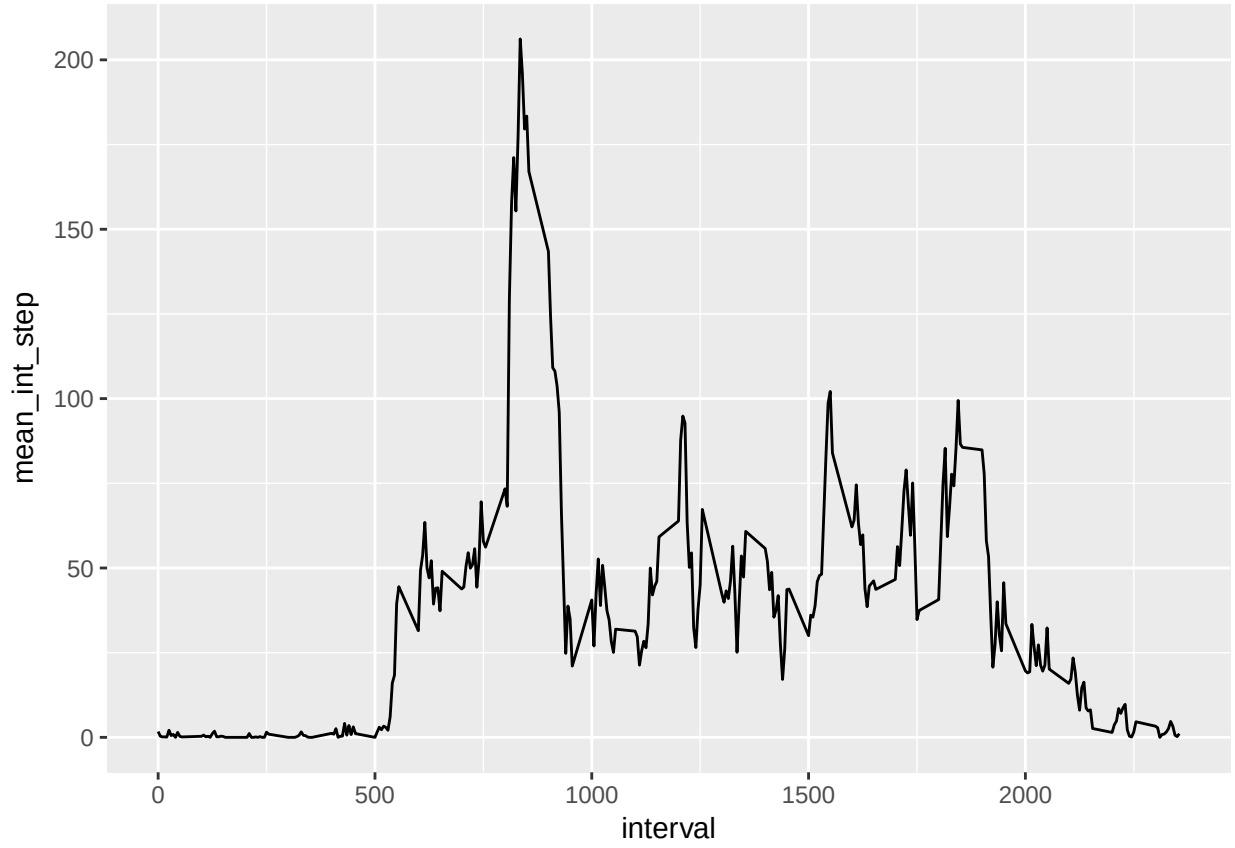
1. Make a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)

2. Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
Total_Daily_Steps$date <- ymd(Total_Daily_Steps$date)


interval_mean <- input_data %>%
                 group_by(interval)%>%
                 summarize(mean_int_step = mean(steps, na.rm=TRUE))

ggplot(interval_mean, aes(interval, mean_int_step)) + geom_line()
```




```
max_step = interval_mean[which.max(interval_mean$mean_int_step),'interval']
```

At interval “”r max_step“”, on average across all the days in the dataset, contains the maximum number of steps.

Imputing missing values

Are there differences in activity patterns between weekdays and weekends?