Reproducible Research: Peer Assessment 1

```
##
## 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
```

Loading and preprocessing the data

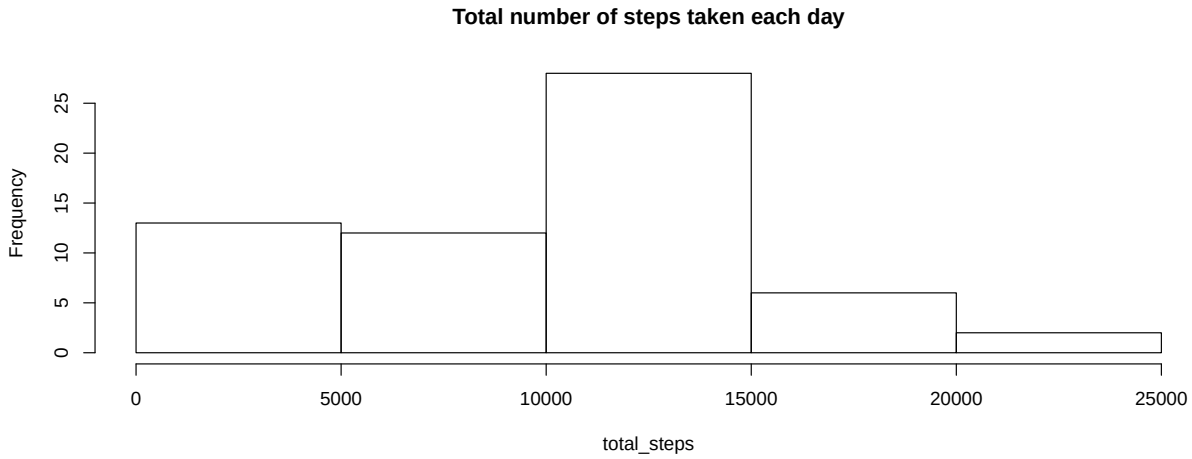
```
 download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip",destfile = "rawdata
unzip("rawdata.zip")
data<- read.csv("activity.csv")
```

What is mean total number of steps taken per day?

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

```
data<- read.csv("activity.csv")
dat <- select(data, steps, date) %>%
    group_by(date) %>%
    summarize(total_steps= sum(steps, na.rm = T))

    with(dat, hist(total_steps, main="Total number of steps taken each day"))
```



Report the mean and median total number of steps taken per day.

The mean and median total number of steps taken per day:

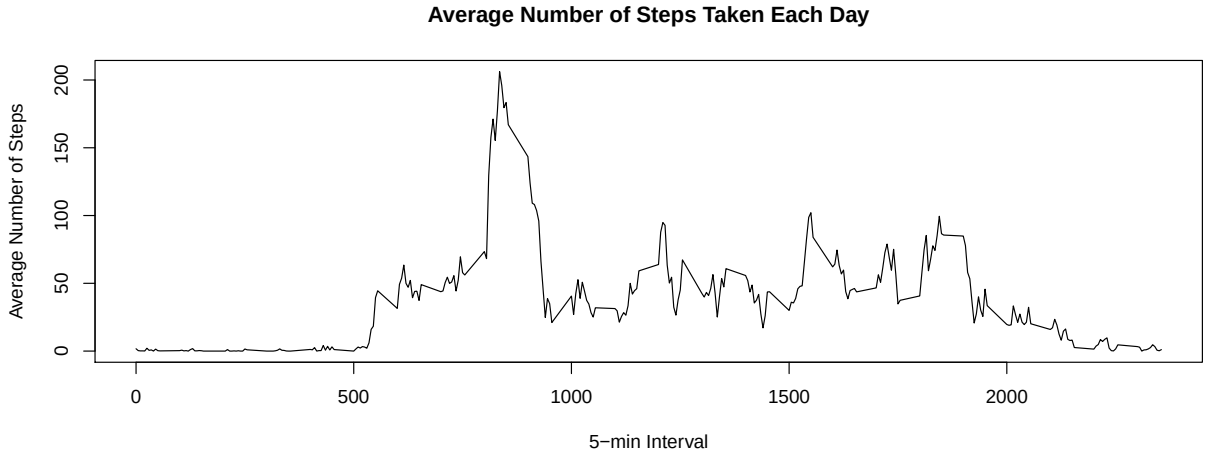
```
select(dat, total_steps) %>%
summarise(mean= mean(total_steps), median= median(total_steps)) %>% as.data.frame
```

```
##      mean median
## 1 9354.23  10395
```

What is the average daily activity pattern?

Make a time series plot of the 5-minute interval and the average number of steps taken, averaged across all days.

```
library(lubridate)
dat <- group_by(data, interval) %>%
    summarize(mean_steps= mean(steps, na.rm = T)) %>%
    with(plot(interval, mean_steps,main = "Average Number of Steps Taken Each Day", type = 'l', xlab =
```



The 5-minute interval contains the maximum number of steps

```
    group_by(data, interval) %>%
    summarize(mean_steps= mean(steps, na.rm = T)) %>%
    filter(mean_steps==max(mean_steps), na.rm = T) %>%
    select(the_interval_maximize_steps= interval) %>% as.data.frame
```

```
##   the_interval_maximize_steps
## 1                         835
```

Imputing missing values

The total number of rows with missing data is:

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

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

The strategy for filling in all of the missing values is to replace the missing values with the mean for that 5-minute interval.

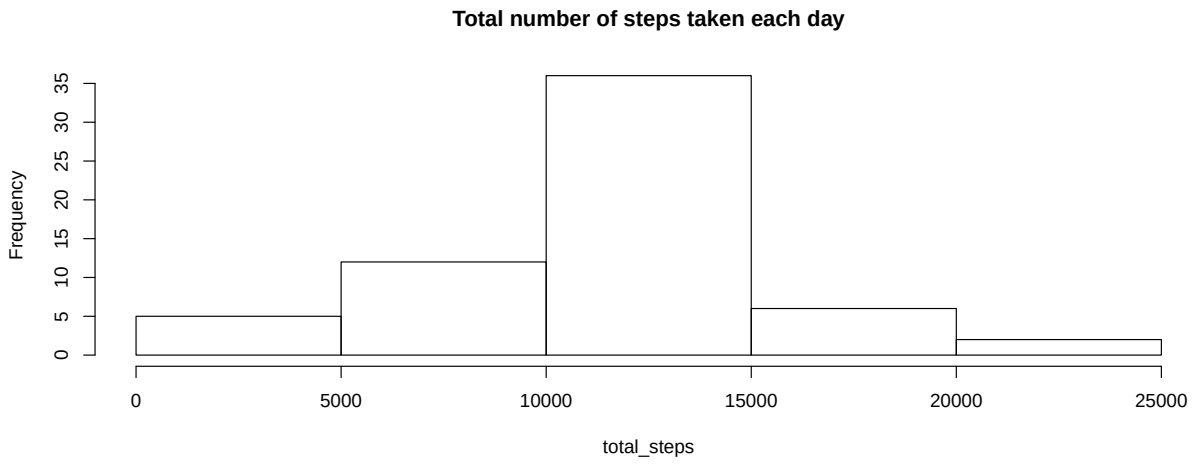
```
## mean steps table for intervals
dat <- group_by(data, interval) %>%
    summarize(mean_steps= mean(steps, na.rm = T)) %>% as.data.frame

## the replace function to replace the NAs with the means steps across days
replace_value <- function(steps,interval_value){
    if (!is.na(steps)) {steps}
    else {dat[dat$interval==interval_value,2]}
}

## replace the NAs in the data
n= length(data$steps)
for (i in 1:n){
    data$steps[i]= replace_value(data$steps[i],data$interval[i])
}
```

analysis with the new data

```
dat <- select(data, steps, date) %>%
    group_by(date) %>%
    summarize(total_steps= sum(steps))
    with(dat, hist(total_steps, main="Total number of steps taken each day"))
```



The mean and median total number of steps taken per day.

```
select(dat, total_steps) %>%
summarise(mean= mean(total_steps), median= median(total_steps)) %>% as.data.frame
```

```
##       mean   median
## 1 10766.19 10766.19
```

From the report we can see that both values are increased. and the histigram becomes more concentrated to the central.

Are there differences in activity patterns between weekdays and weekends?

create a new vairable indicating the type of days

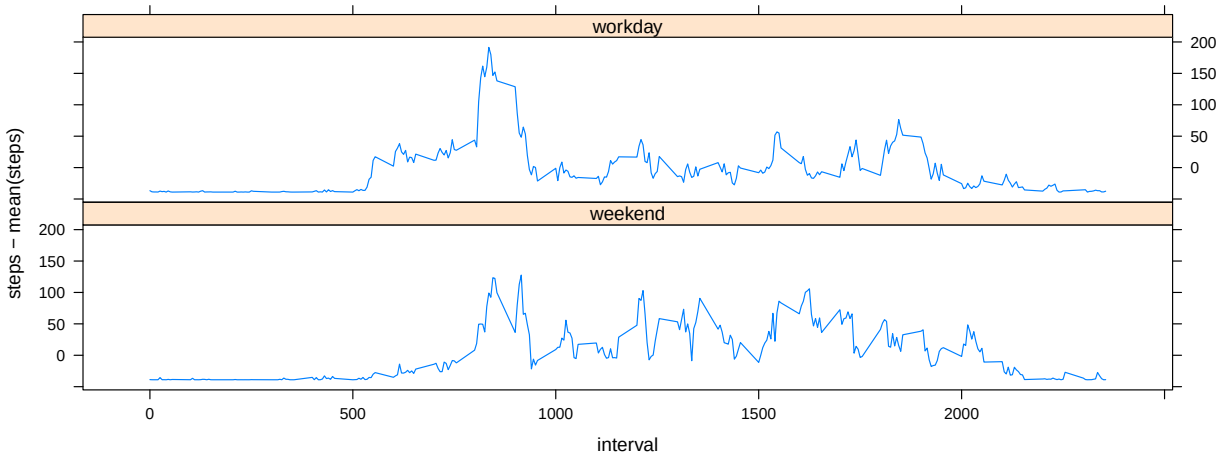
```
data<- mutate(data, date= as.Date(date))
day_type<- function(date){
        if (weekdays(date)=="Sunday"|weekdays(date)=="Saturday") {"weekend"}
        else {"workday"}
}

dat <- mutate(data, type= sapply(date,day_type)) %>%
    mutate(type = as.factor(type))
```

From the following pictures we can see that the partterns are different.

```
plotdat <- select(dat, steps, interval, type) %>%
    group_by(interval,type) %>%
    summarize(steps= mean(steps))

    xyplot(steps~mean(steps)~interval|type, layout= c(1,2), data = plotdat, type ="l")
```



```
qplot(interval, steps, data = plotdat, facets = .~type, geom = c("point","smooth"),method="loess")
```