Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(scales)
```

```
## Warning: package 'scales' was built under R version 4.0.5
```

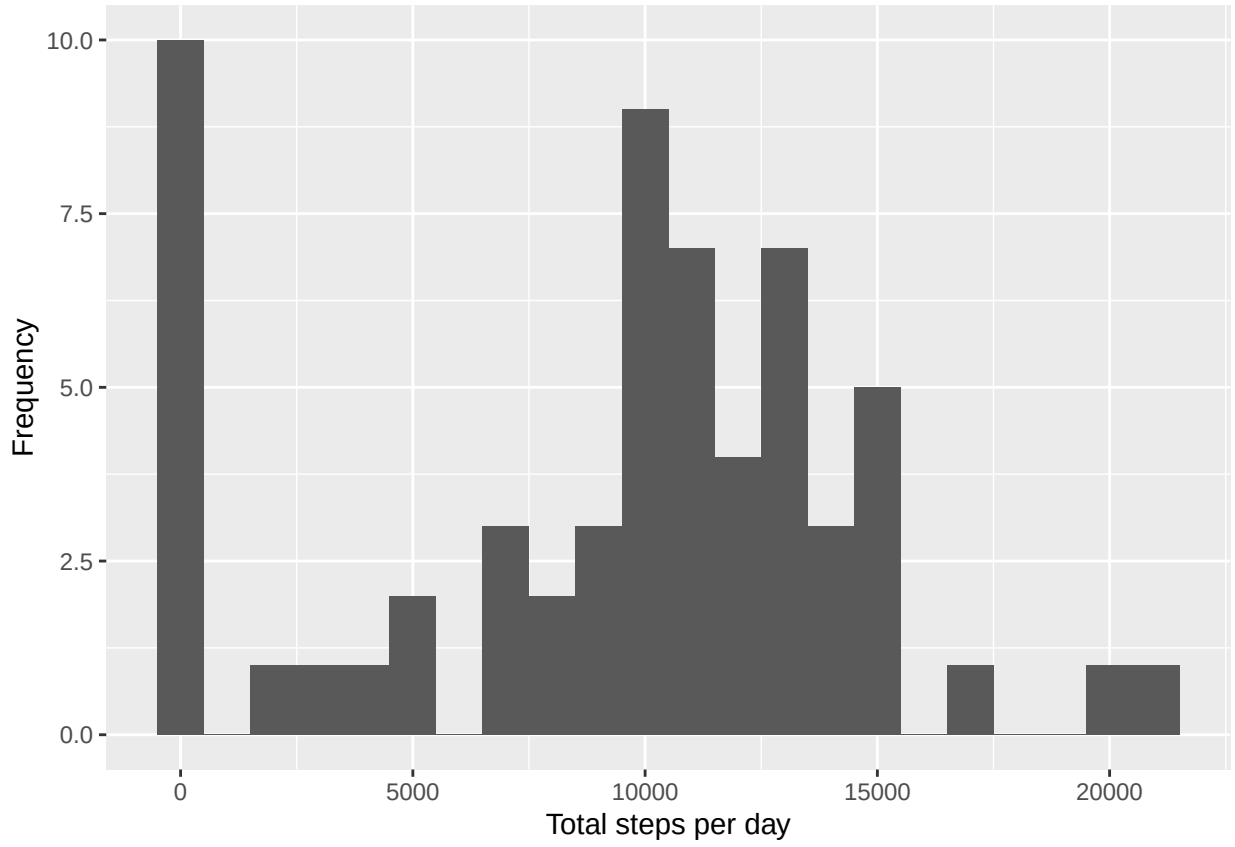
```
library(Hmisc)
```

```
## Warning: package 'Hmisc' was built under R version 4.0.5

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

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


#flag for NA vaues
NA_flag<-is.na(data$steps)
#data excluding NA values
clean_data<-data[!NA_flag,]
```

What is mean total number of steps taken per day?

```
Total_steps_per_day=tapply(data$steps,data$date,FUN=sum,na.rm=TRUE)
qplot(Total_steps_per_day, xlab='Total steps per day', ylab='Frequency', binwidth=1000)
```

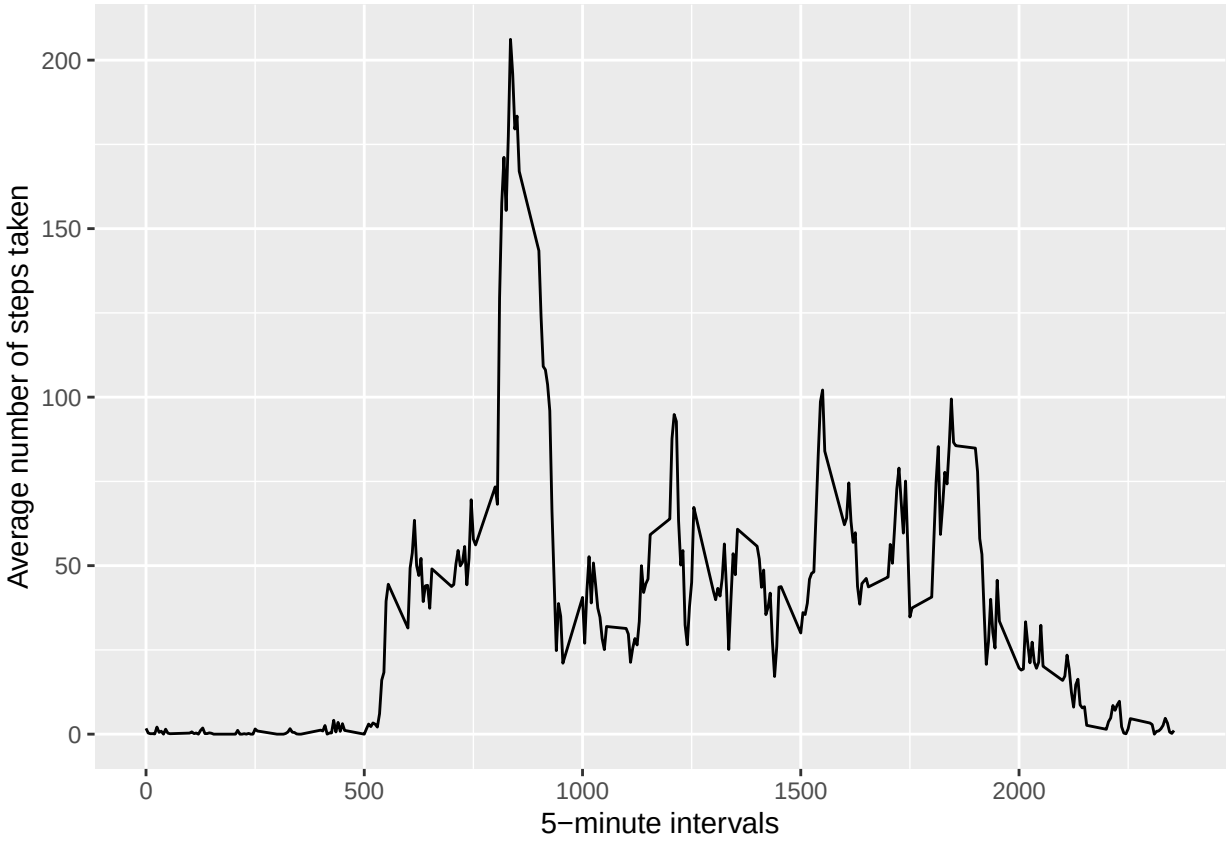



```
m<-mean(Total_steps_per_day)
med<-median(Total_steps_per_day)
```

- Mean is 9354.2295082
- Median is 10395

What is the average daily activity pattern?

```
average <- aggregate(x=list(mean_steps=data$steps), by=list(interval=data$interval), FUN=mean, na.rm=TRU
ggplot(data=average, aes(x=interval, y=mean_steps)) +
    geom_line() +
    xlab("5-minute intervals") +
    ylab("Average number of steps taken")
```




```
max_interval <- average[which.max(average$mean_steps),]
```

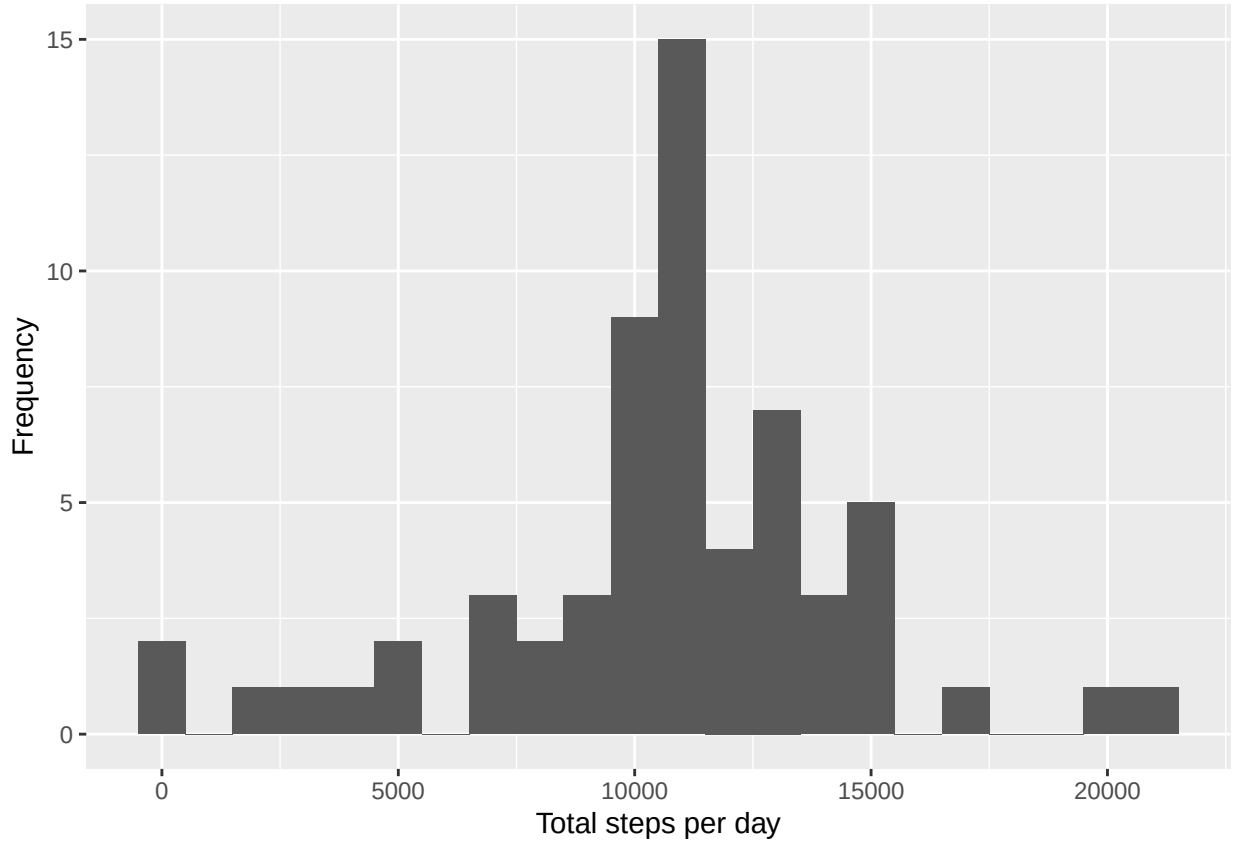
5-minute interval with maximum number of steps is: 835, 206.169811320755

Imputing missing values

```
missing_number<-sum(NA_flag)
```

The total number of missing values is 2304

```
imputed_data<-data
imputed_data$steps <- impute(data$steps, fun=mean)
Total_steps_imputed <- tapply(imputed_data$steps, imputed_data$date, sum)
qplot(Total_steps_imputed, xlab='Total steps per day', ylab='Frequency', binwidth=1000)
```

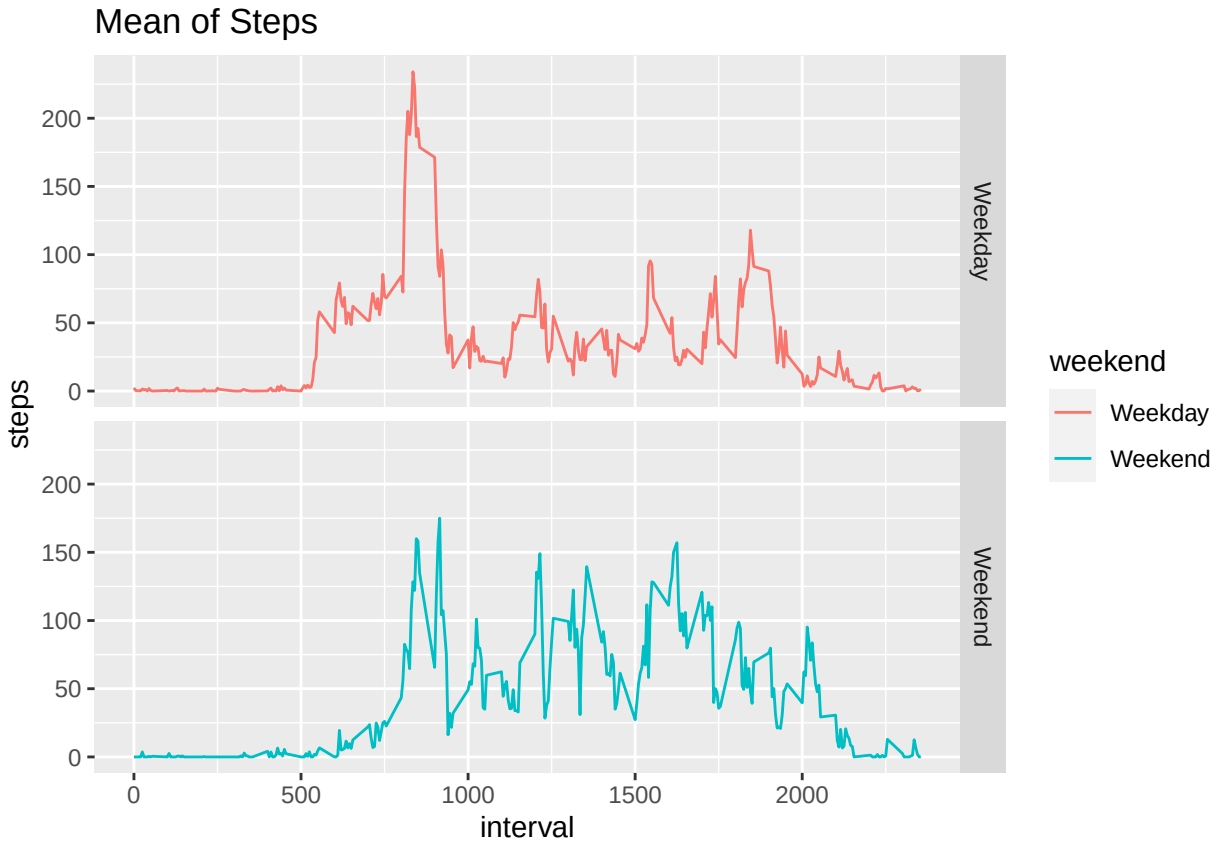



```
new_mean<-mean(Total_steps_imputed)
new_med<-median(Total_steps_imputed)
```

Original mean was 9354.2295082; new mean is 1.0766189×10^4 Original median was 10395; new median is `new_med` Effect: Median of data is pushed toward the mean

Are there differences in activity patterns between weekdays and weekends?

```
data$weekday <- weekdays(as.Date(as.character(data$date)))
data$weekend <- ifelse(data$weekday == "Saturday" | data$weekday == "Sunday", "Weekend", "Weekday")

new_average <- aggregate(list(new_steps=data$steps), by=list(weekend=data$weekend, new_interval=data$in

ggplot(new_average, aes(x =new_interval, y=new_steps, color=weekend)) +
  geom_line() +
  facet_grid(weekend ~ .) +
  labs(title = "Mean of Steps", x = "interval", y = "steps")
```



There is some variation on weekdays whereas not much variation on weekends.