Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
library(knitr)
library(swirl)
```

```
## Warning: le package 'swirl' a été compilé avec la version R 4.2.2
```

```
##
## | Hi! Type swirl() when you are ready to begin.
```

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

```
## Warning: le package 'lubridate' a été compilé avec la version R 4.2.2
```

```
##
## Attachement du package : 'lubridate'
```

```
## Les objets suivants sont masqués depuis 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'
```

```
## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag
```

```
## Les objets suivants sont masqués depuis 'package:base':
##
##     intersect, setdiff, setequal, union
```

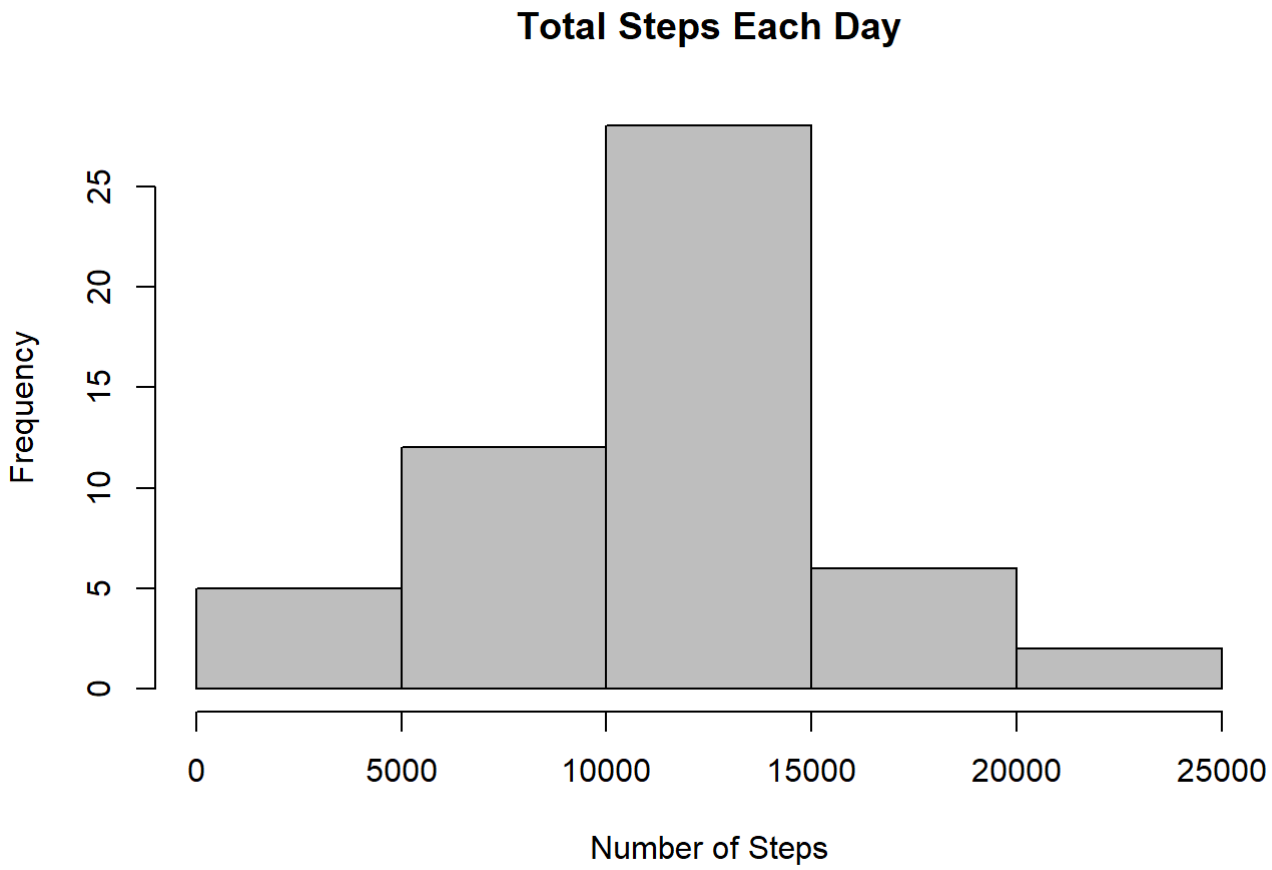
```
getwd()
```

```
## [1] "C:/Users/caroline/Documents/R/data/Rep_data"
```

```
setwd("C:/Users/caroline/Documents/R/data/Rep_data")
activity<-read.csv("activity.csv", header=T, quote="\"", sep=",")
activity<-na.omit(activity)
```

What is mean total number of steps taken per day?

```
totalSteps <- aggregate(steps ~ date, activity, sum)
hist(totalSteps$steps, main = paste("Total Steps Each Day"), col="grey", xlab="Number of Step
s")
```




```
#mean of the steps by day
mean_steps<-mean(totalSteps$steps)
mean_steps
```

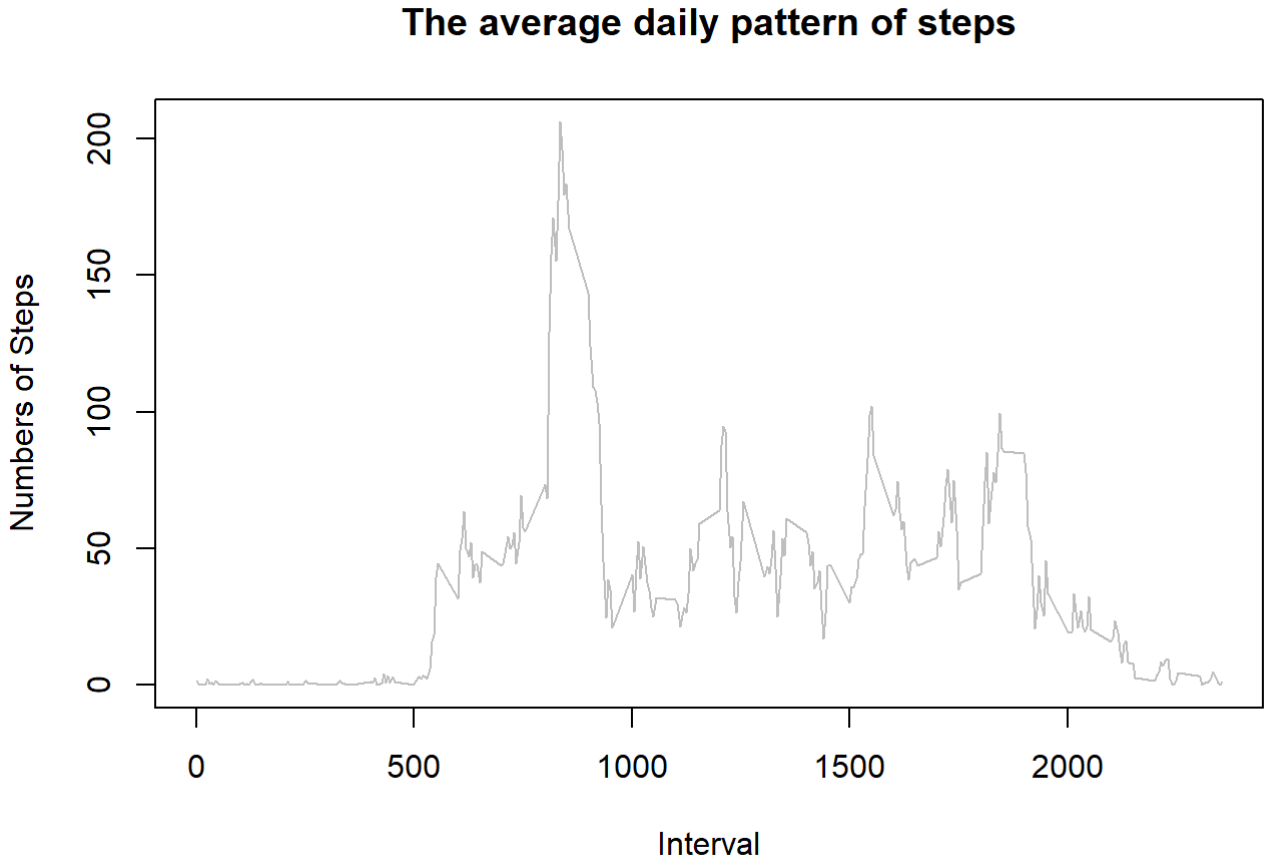
```
## [1] 10766.19
```

```
#median of the steps by day
median_steps<-median(totalSteps$steps)
median_steps
```

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

What is mean total number of steps taken per day?

```
#pattern of steps per interval
steps_per_interval<-aggregate(steps~interval,activity,mean)
plot(steps_per_interval$interval,steps_per_interval$steps,type="l",col="grey",xlab="Interva
l",ylab="Numbers of Steps",main="The average daily pattern of steps")
```



#maximum of steps in 5 min interval

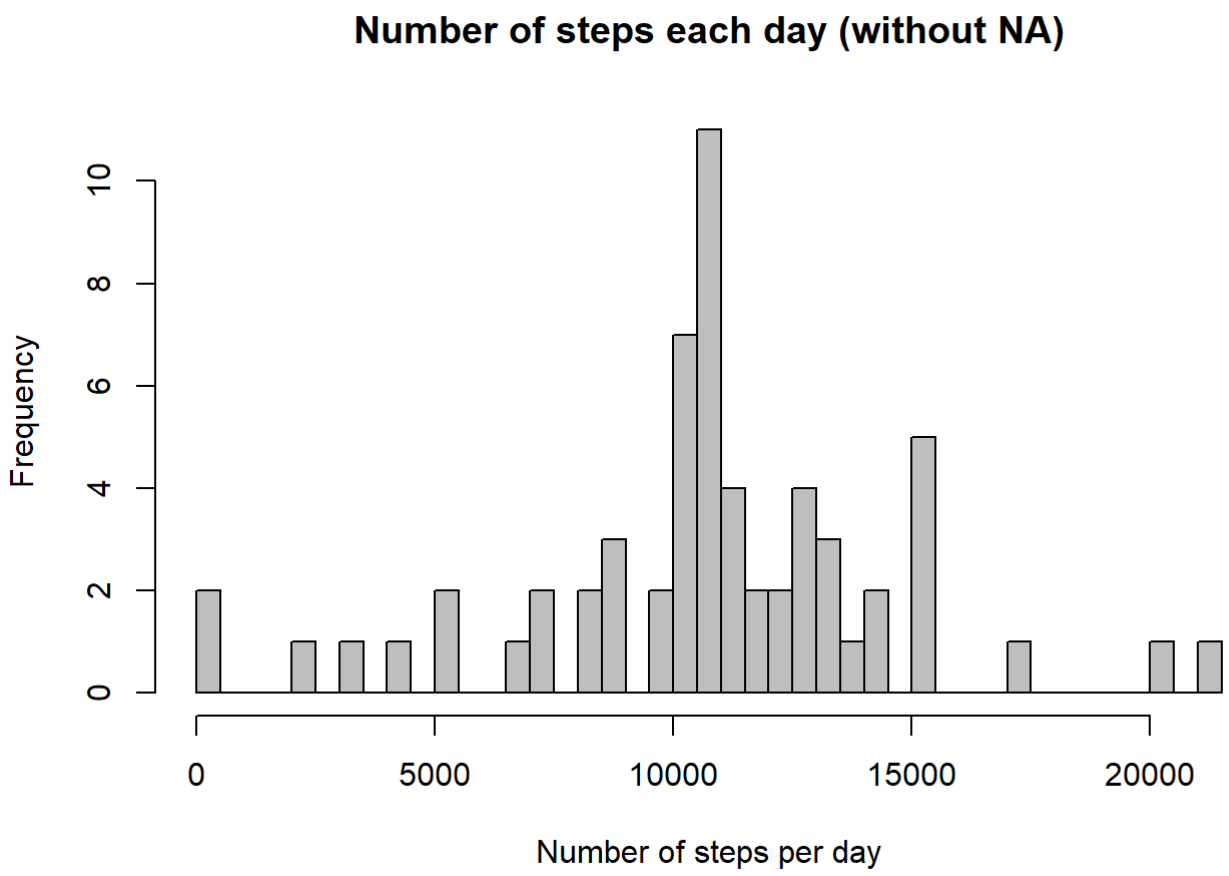
```
max_interval<-steps_per_interval$interval[which.max(steps_per_interval$steps)]
```

Imputing missing values

#calculating NA's

```
activity1<-read.csv("activity.csv")
steps_na<-sum(is.na(activity1$steps))
```

```
average_interva5 <- aggregate(steps ~ interval, activity1, FUN = mean, na.rm = TRUE)
fillNA<- activity1
na_steps <- is.na(activity1$steps)
aux <- na.omit(subset(average_interva5, interval == activity1$interval[na_steps]))
fillNA$steps[na_steps] <- aux[, 2]
na_steps_fillNA <- sum(is.na(fillNA))
steps_day_noNA<-aggregate(steps ~ date, data = fillNA, FUN = sum, na.rm = TRUE)
hist(steps_day_noNA$steps,main = "Number of steps each day (without NA)", xlab = "Number of s
teps per day",col = "grey", breaks = 40)
```



##Comparing the two datas with and withou NA's

```
mean_stepsNA<-mean(steps_day_noNA$steps)

median_stepsNA<-median(steps_day_noNA$steps)

summary(totalSteps$steps)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      41    8841   10765   10766   13294   21194
```

```
summary(steps_day_noNA$steps)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      41    9819   10766   10766   12811   21194
```

Are there differences in activity patterns between weekdays and weekends?

```
activity$date<-as.Date(x=activity$date,format="%Y%m%d")
weekdays_values = c("Monday","Tuesday", "Wednesday", "Thursday", "Friday")
date_type <- ifelse( weekdays(activity$date) %in% weekdays_values,"weekdays", "weekend")

activity$day<-factor(x=date_type)
steps_weekday <- aggregate(steps ~ interval + day, data = activity, FUN = mean, na.rm = TRUE)

ggplot(steps_weekday, aes(interval, steps, fill = day)) +
geom_line() +
facet_grid(day ~ .) +
xlab("5-minute interval") +
ylab("Average number of steps") +
ggtitle("Activity by the weekend or weekday")
```