Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
#Unzip the file
unzip(zipfile = "activity.zip")

#Read in the csv file
dat <- read.csv(file = "activity.csv",header = TRUE)
```

What is mean total number of steps taken per day?

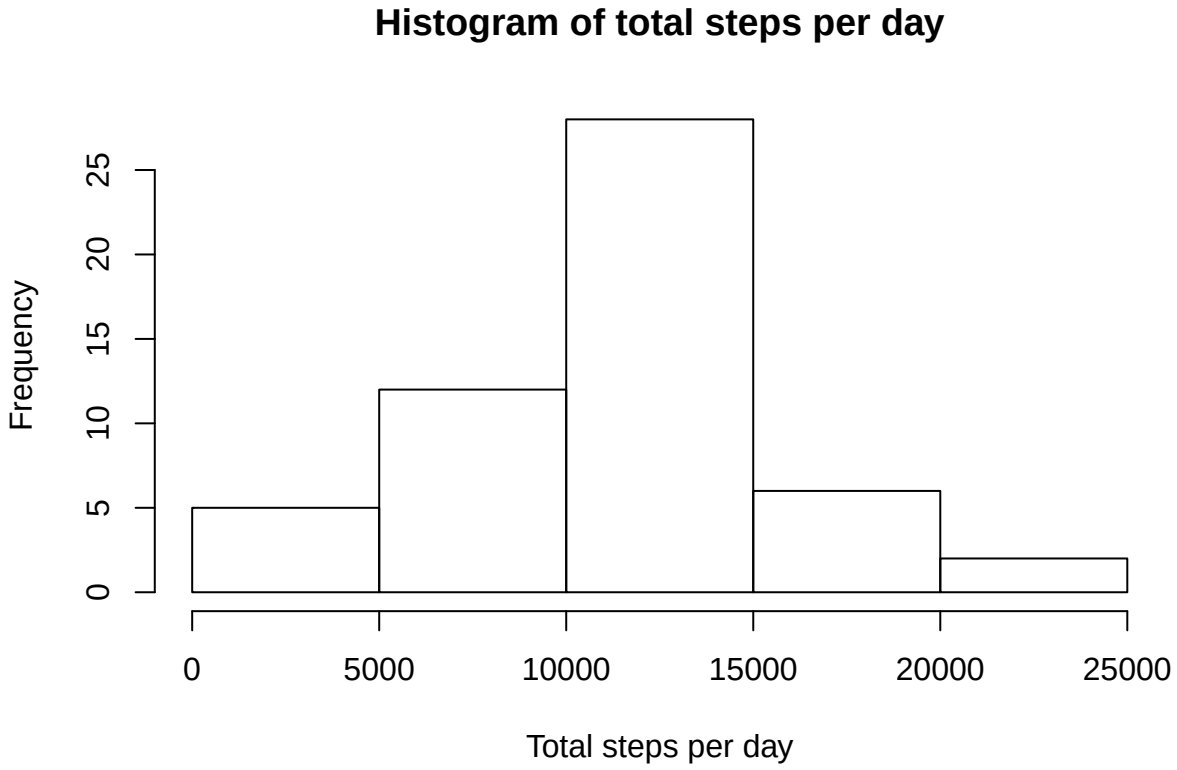
```
#Install the dplyr package
library(dplyr)
```

```
##
## 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
```

```
#Create a data frame with total steps per day
steps_per_day <- steps_per_day <- dat %>% group_by(date) %>% summarise(total_steps = sum(steps))

#Create a histogram of the total number of steps taken per dat

hist(steps_per_day$total_steps, main = "Histogram of total steps per day", xlab = "Total steps per day"
```



###Calculate the mean and median of the total steps per day

Mean total steps:

```
mean(steps_per_day$total_steps, na.rm = T)
```

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

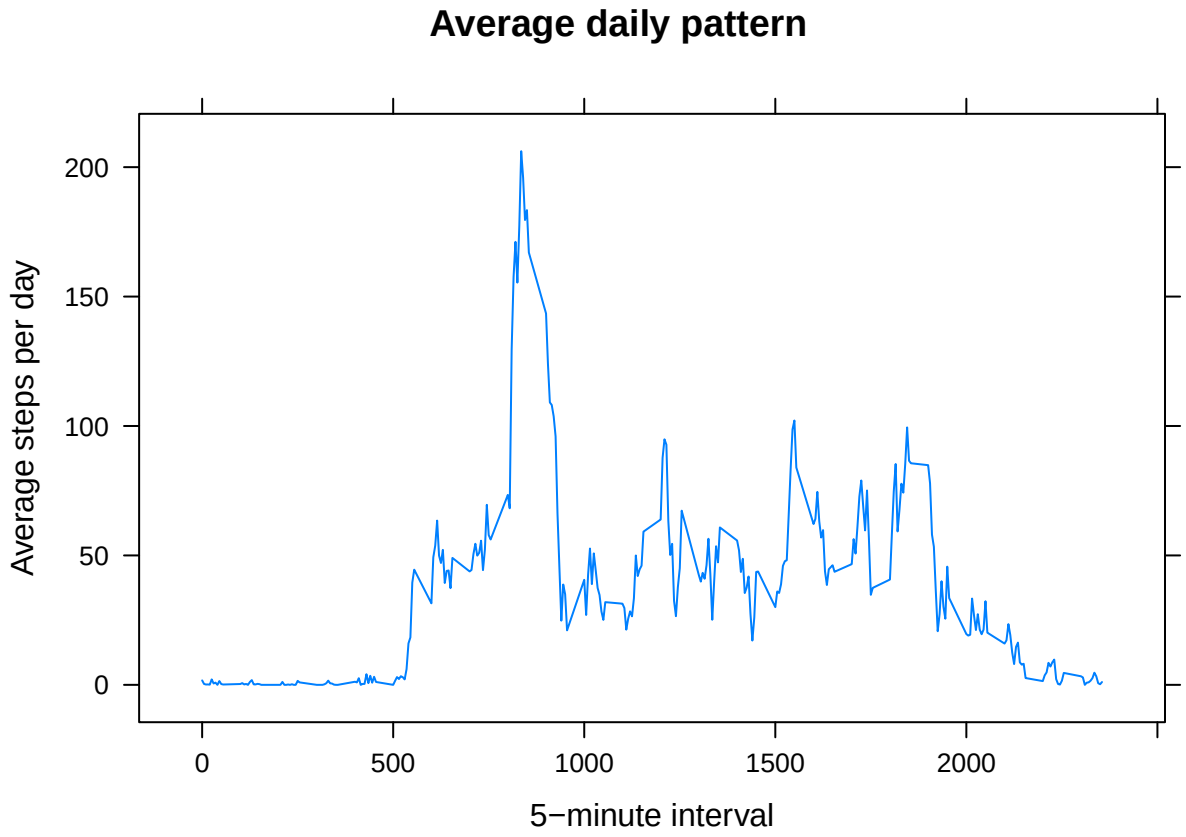
Meadian total steps:

```
median(steps_per_day$total_steps, na.rm = T)
```

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

What is the average daily activity pattern?

```
##Create a data frame with average steps per by time interval (after removing the NA values)

time_series <- dat %>% group_by(interval) %>% summarise(avg_steps_per_day = mean(steps, na.rm = T))

#Creating the time series plot
library(lattice)
xyplot(time_series$avg_steps_per_day ~ time_series$interval, type = "l", main = "Average daily pattern"
```




```
#Finding the interval that on average across all the days in the dataset, contains the maximum number o

max_interval <- time_series$interval[which.max(time_series$avg_steps_per_day)]
```

The 5-minute interval, which on average across all the days in the dataset, contains the maximum number of steps is 835

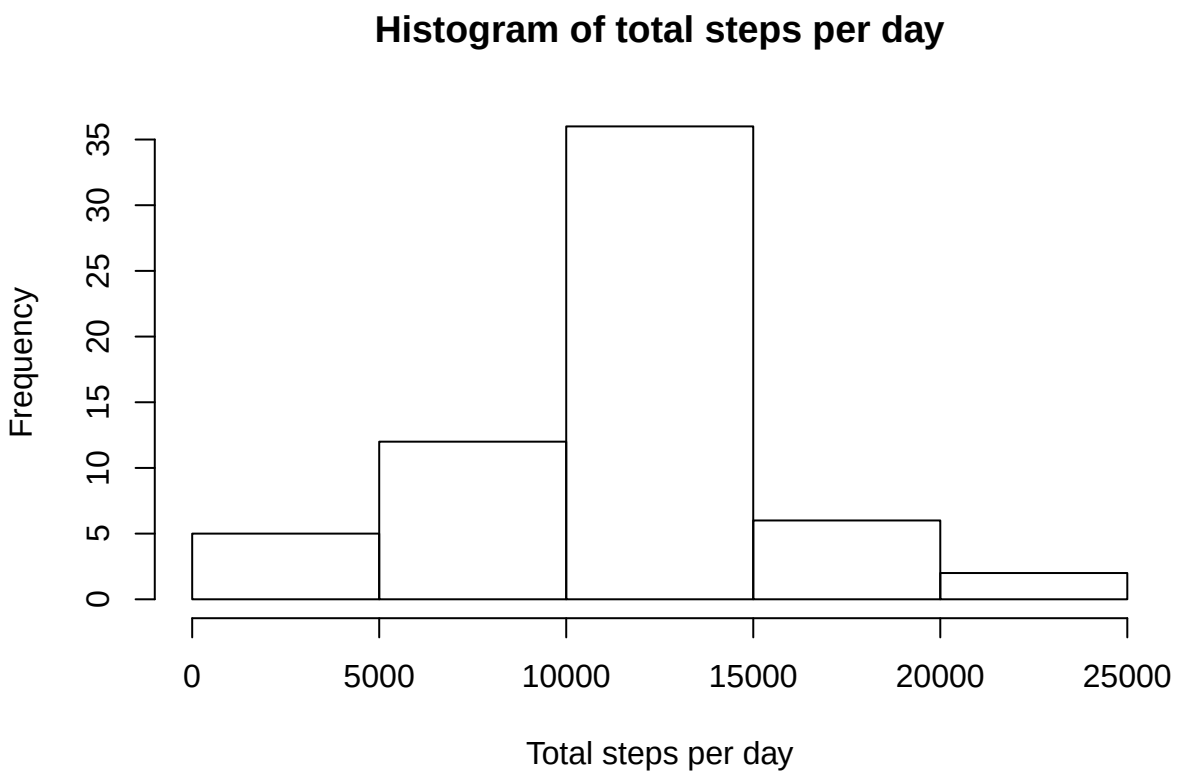
Inputing missing values

```
#Calculating total number of rows with missing values

missing_values <- sum(is.na(dat))
```

There are 2304 rows with missing values

```
#Creating a loop to replace NA's and creating a new dataset out of it

dat2 <- dat
i = nrow(dat)
val = 1
while(val <= i){
     if(is.na(dat2[val, "steps"]) == TRUE)
       dat2[val, "steps"] <- time_series$avg_steps_per_day[which(dat2[val, "interval"] == time_series$i
     val = val + 1
}

#Create a data frame from the new dataset with total steps per day
```

```r
steps_per_day_2 <- dat2 %>% group_by(date) %>% summarise(total_steps = sum(steps))

#Create a histogram of the total number of steps taken per dat

hist(steps_per_day_2$total_steps, main = "Histogram of total steps per day", xlab = "Total steps per da
```



###Calculate the mean and median of the total steps per day

Mean total steps:

```r
mean(steps_per_day_2$total_steps, na.rm = T)
```

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

Meadian total steps:

```r
median(steps_per_day_2$total_steps, na.rm = T)
```

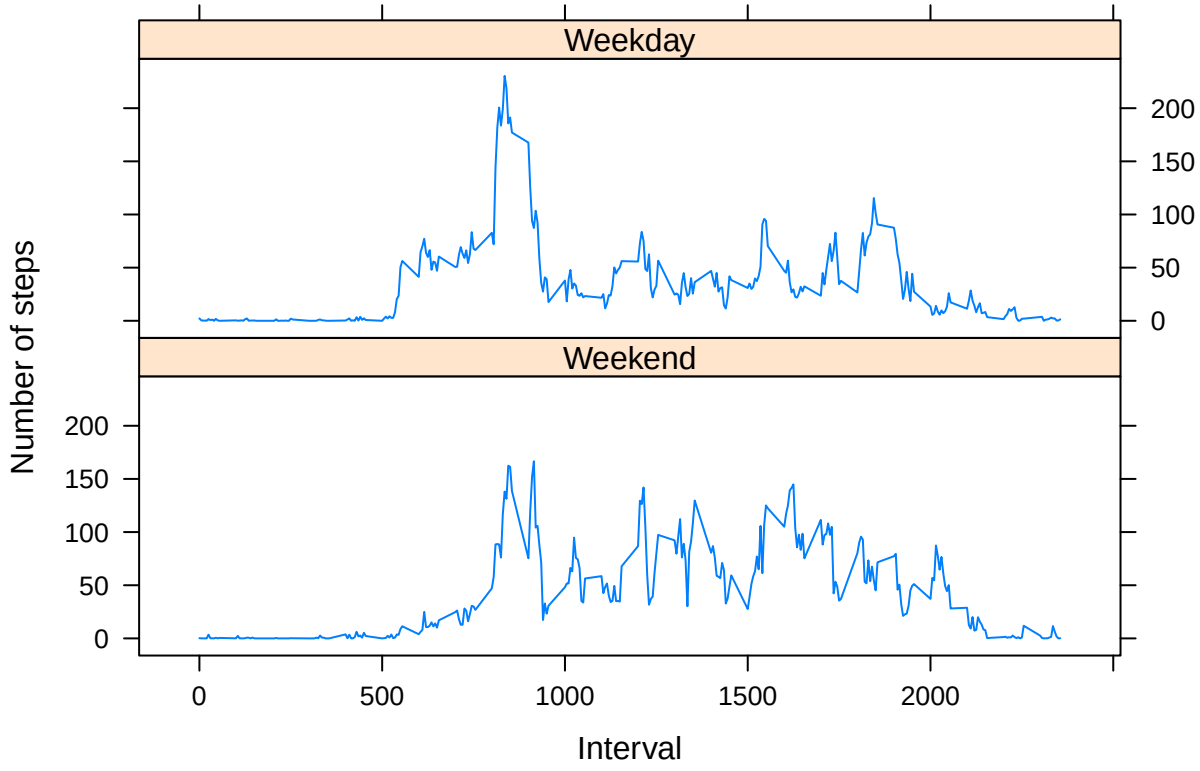
```
## [1] 10766.19
```

"Even though the frequency of the steps per day has increased after replacing the NA values with the mean values, there is no impact on the mean and the median of the total number of steps itself."

Are there differences in activity patterns between weekdays and weekends?

```r
#Adding a new factor variable to dat2 to differentiate between weekdays and weekends

dat2$date <- as.Date(dat$date)
```

```r
#Create a vector for weekdays
weekdays1 <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")

dat2$WeekDay <- factor((weekdays(dat2$date) %in% weekdays1), levels = c(FALSE, TRUE), labels = c("Weeke

#Creating a new data frame to have the average steps by weekday/weekend and 5 minute intervale

time_series_2 <- dat2 %>% group_by(interval, WeekDay) %>% summarise(average_steps = mean(steps))

#Creating a panel plot
library(lattice)
xyplot(average_steps ~ interval | WeekDay, data = time_series_2, layout = c(1, 2), type = "l", ylab = "
```



Based on the time series analysis, there does not seem to be a major difference of the average number of steps between weekdays and weekends