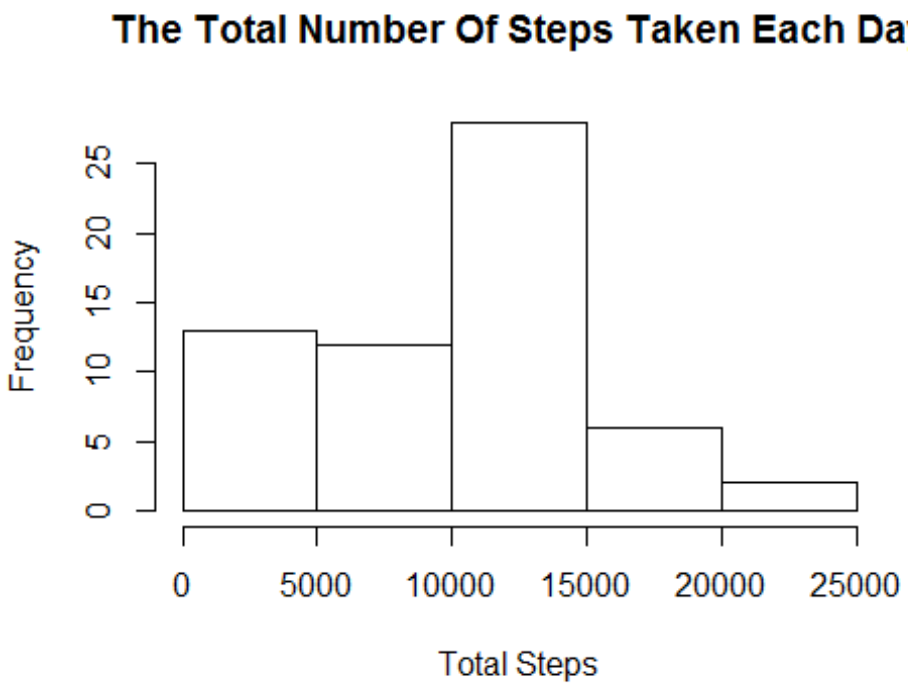
Reproducible Research: Peer Assessment 1

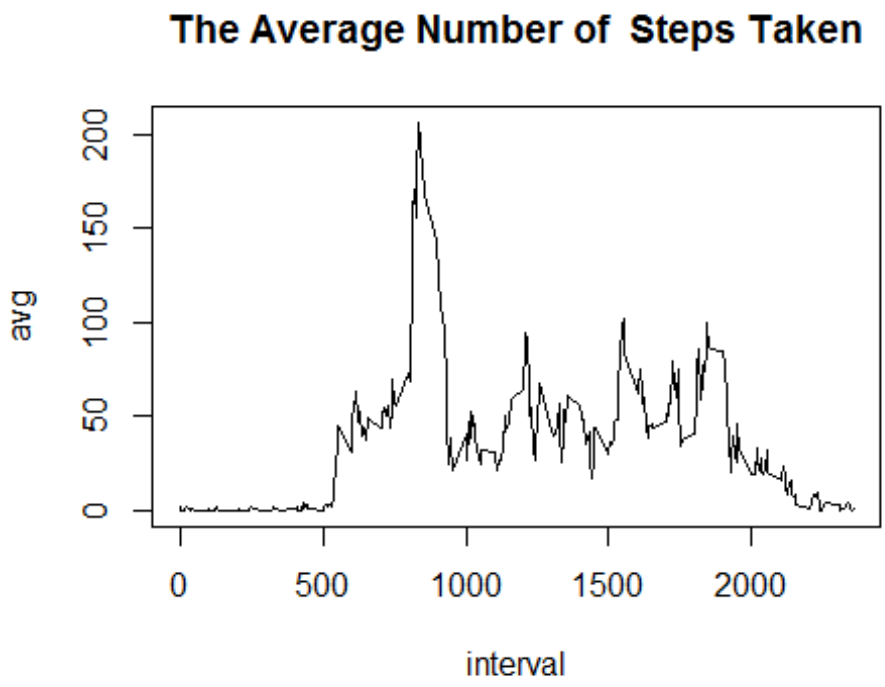
```
setwd("F:/-RdC-/Coursera/Coursera-DataScience/5、Reproducible Research/
Assignment/A1")

library(knitr)
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(lubridate)
library(ggplot2)

## Loading and preprocessing the data
Activity<-read.csv("./data/activity.csv",header=T,na.strings="NA")

Activity$date_time<-
  ymd_hms(paste(Activity$date," "
                ,substr(sprintf("%04.0f", Activity$interval),1,2)
                 ,":",substr(sprintf("%04.0f", Activity$interval),3,4),
":00",sep=""))

## What is mean total number of steps taken per day?
PerDateSum<-Activity %>%
  group_by(date) %>%
  summarise(sum=sum(steps,na.rm=T))

hist(PerDateSum$sum,main="The Total Number Of Steps Taken Each Day"
     ,xlab="Total Steps")
```



The mean total number of steps taken per day equal 9354.2295082

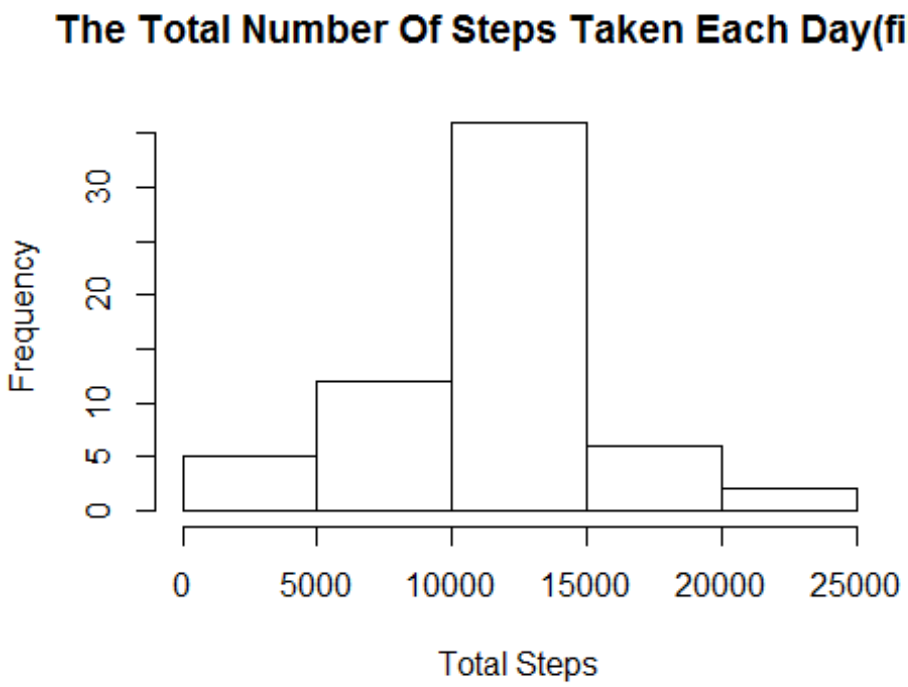
The median total number of steps taken per day equal 10395

```
## What is the average daily activity pattern?
IntervalAvg<-Activity %>%
  group_by(interval) %>%
  summarise(avg=mean(steps,na.rm=T))

plot(IntervalAvg,type="l",main="The Average Number of  Steps Taken")
```



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

```
## Imputing missing values
FilledData<-Activity

for (i in unique(FilledData$interval))
{
  FilledData[(is.na(FilledData$steps)
              & FilledData$interval==i),1]<-IntervalAvg[ which(Interval
Avg$interval==i),2]
}

FilledPerDateSum<-FilledData %>%
  group_by(date) %>%
  summarise(sum=sum(steps,na.rm=T))

hist(FilledPerDateSum$sum,main="The Total Number Of Steps Taken Each Da
y(filled)"
     ,xlab="Total Steps")
```



The mean total number of steps taken per day from the dataset which filled in all of the missing values equal 1.076618910^{4}

The median total number of steps taken per day from the dataset which filled in all of the missing values equal 1.076618910^{4}

These values differ from the estimates from the first part of the assignment .

```
## Are there differences in activity patterns between weekdays and week
ends?

i<-1
weekends<-""
while (i<nrow(FilledData)+1){
if(weekdays(FilledData$date_time[i]) %in% c("星期六","星期日"))
  {weekends[i]<-"weekends"}
else { weekends[i]<-"weekdays"}
i<-i+1
}

FilledData$WK<-weekends

FilledWkAvg<-  FilledData %>%
  group_by(WK,interval) %>%
  summarise(avg=mean(steps,na.rm=T)) %>%
  as.data.frame()
```

```
ggplot(FilledWkAvg,aes(x=interval,y=avg,group=WK))+geom_line()+facet_wrap(~WK,nrow=2)
```