pA1__Template

Nimisha

21/05/2020

Reproducible research - course proj 1

```
setwd("C:/Users/NIMISHA/Documents/Reproducible research")
data1 <- read.csv("activity.csv",sep = ",")
summary(data1)
```

```
##      steps             date          interval
## Min.   :  0.00   2012-10-01:  288   Min.   :   0.0
## 1st Qu.:  0.00   2012-10-02:  288   1st Qu.: 588.8
## Median :  0.00   2012-10-03:  288   Median :1177.5
## Mean   : 37.38   2012-10-04:  288   Mean   :1177.5
## 3rd Qu.: 12.00   2012-10-05:  288   3rd Qu.:1766.2
## Max.   :806.00   2012-10-06:  288   Max.   :2355.0
## NA's   :2304     (Other)   :15840
```

```
dim(data1)
```

```
## [1] 17568      3
```

```
str(data1)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : Factor w/ 61 levels "2012-10-01","2012-10-02",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

```
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
##
## 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
```
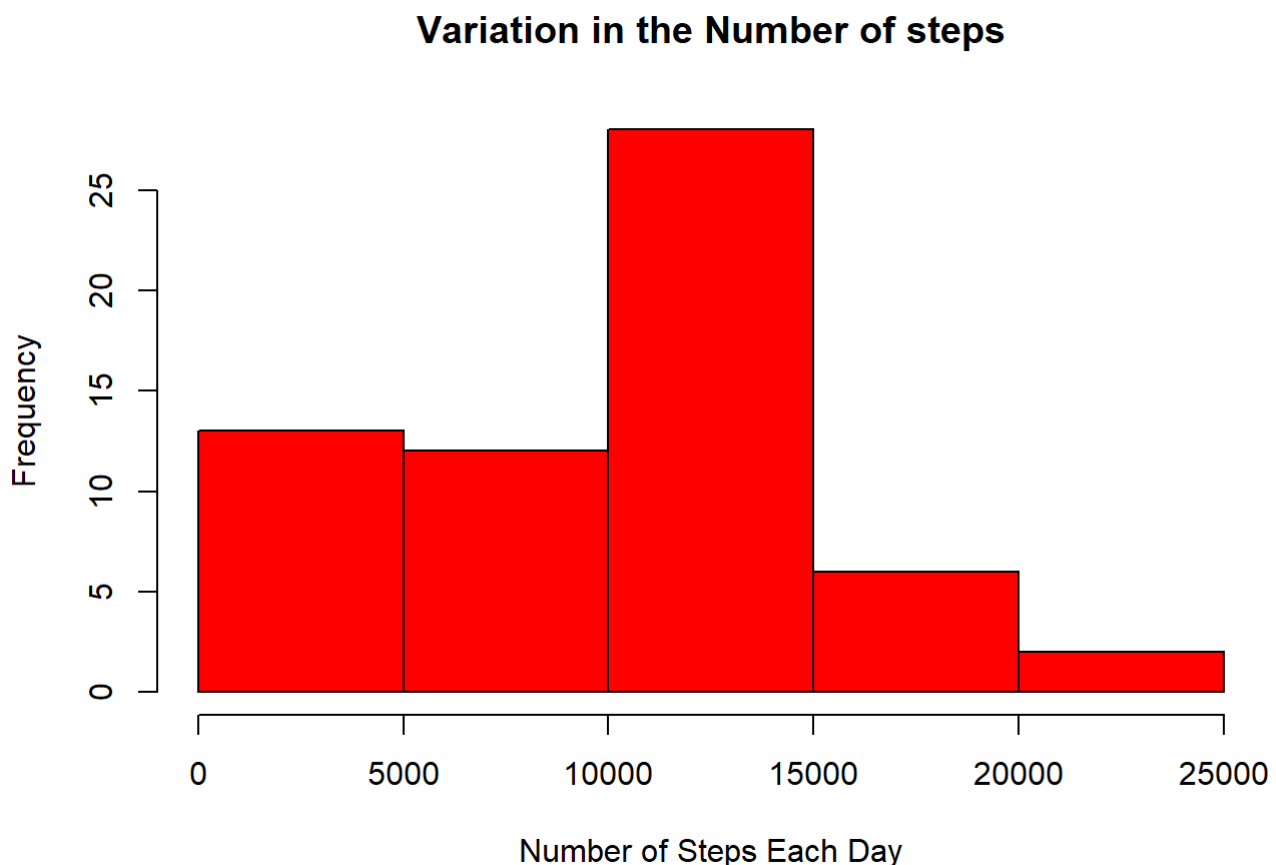
Calculating number of steps each day.

```
steps_each_day<-aggregate(x=list(steps=data1$steps),by=list(date=data1$date),sum,na.rm=TRUE)
head(steps_each_day)
```

```
##         date steps
## 1 2012-10-01     0
## 2 2012-10-02   126
## 3 2012-10-03 11352
## 4 2012-10-04 12116
## 5 2012-10-05 13294
## 6 2012-10-06 15420
```

3. Explaining the variation of the number of steps taken by the person each day

```
hist(steps_each_day$steps,col="red",main = "Variation in the Number of steps",xlab ="Number o
f Steps Each Day" )
```


4. Mean and median number of steps taken each day

```
mean_steps<-mean(steps_each_day$steps,na.rm=TRUE)
median_steps<-median(steps_each_day$steps,na.rm=TRUE)
```

5. Calculating average in each time interval across all days and making a line plot.

```
avg_steps_each_interval<- aggregate(x=list(steps=data1$steps),by=list(interval=data1$interva
l),mean,na.rm=TRUE)
max_<-which.max(avg_steps_each_interval$steps)
max_steps<-avg_steps_each_interval[max_,'steps']
ggplot(data = avg_steps_each_interval,aes(x=interval,y=steps))+geom_line(col="blue")+labs(tit
le = "Average Number of Steps in Each Interval")
```

Average Number of Steps in Each Interval
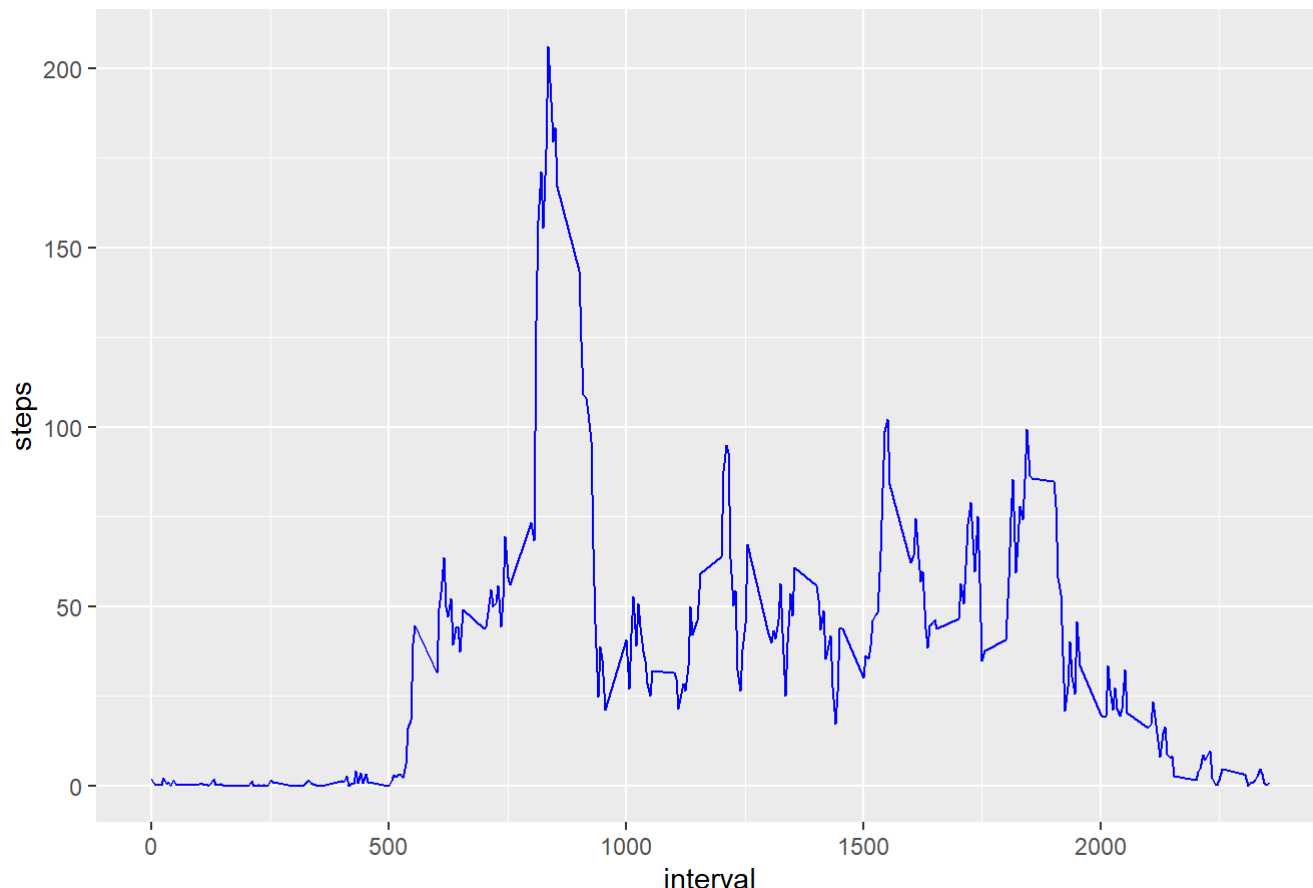
Finding index of all those entries which are filled as NA in the dataset

```
filled_data<-data1
na_index<-which(is.na(data1))
```

7. Imputing all those values in the dataset which are not available.

```
for(i in na_index){
        interval_of_data<-filled_data[i,'interval']
        filled_data[i,'steps']<-avg_steps_each_interval[avg_steps_each_interval$interval==int
erval_of_data,'steps']

}
```

Calculating number of steps taken each day and other basic statistics for complete data.

```
num_steps<-aggregate(x=list(steps=filled_data$steps),by=list(date=filled_data$date),sum)
mean_steps<-mean(num_steps$steps)
median_steps<-median(num_steps$steps)
ggplot(data=num_steps,aes(steps))+geom_histogram(col="red",bins = 8)+labs(title = "Number of
 Steps Taken Each Day")+xlab(label = c("Steps"))
```

Number of Steps Taken Each Day
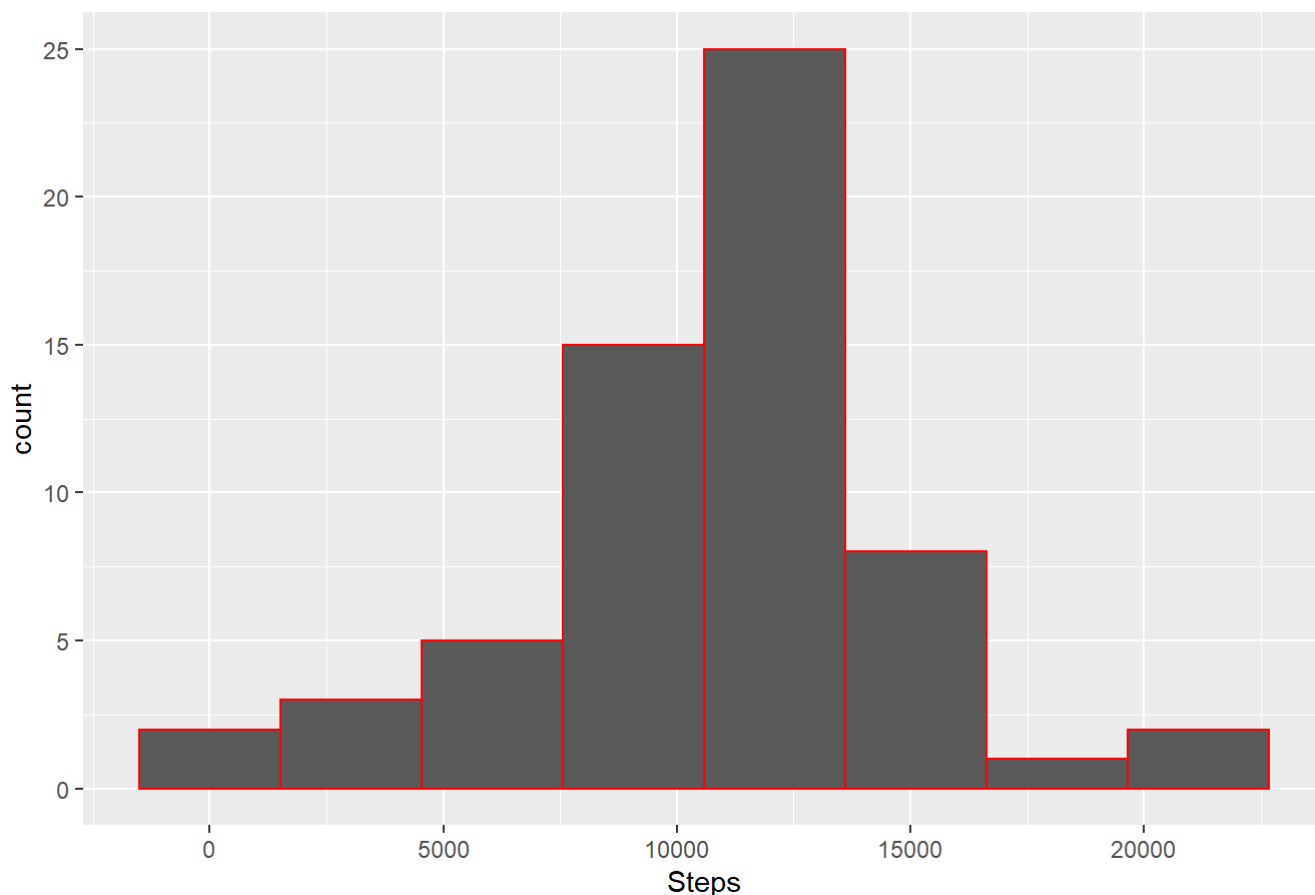
Making new column and filling the data that whether the day was a weekday or a weekend and plotting the acivity on weekend and weekday.

```
filled_data$weekday<-weekdays(as.Date(filled_data$date))

filled_data$weekday<-ifelse(filled_data$weekday=="Saturday"|filled_data$weekday=="Sunday","We
ekend","Weekday")
levels(filled_data$weekday)<-c("Weekend","Weekday")

data_for_plot<-aggregate(x=list(steps=filled_data$steps),by=list(interval=filled_data$interva
l,weekday=filled_data$weekday),mean)

g<-ggplot(data = data_for_plot,aes(x=interval,y=steps))+geom_line(col="blue")
g+facet_grid(weekday~.)+labs(title = "Activity on different type of days")+xlab(label = "5 mi
nute interval")
```

Activity on different type of days