Course project 1

Mohanad

7/17/2020

Here we are gonna load the data and delete the NA Values

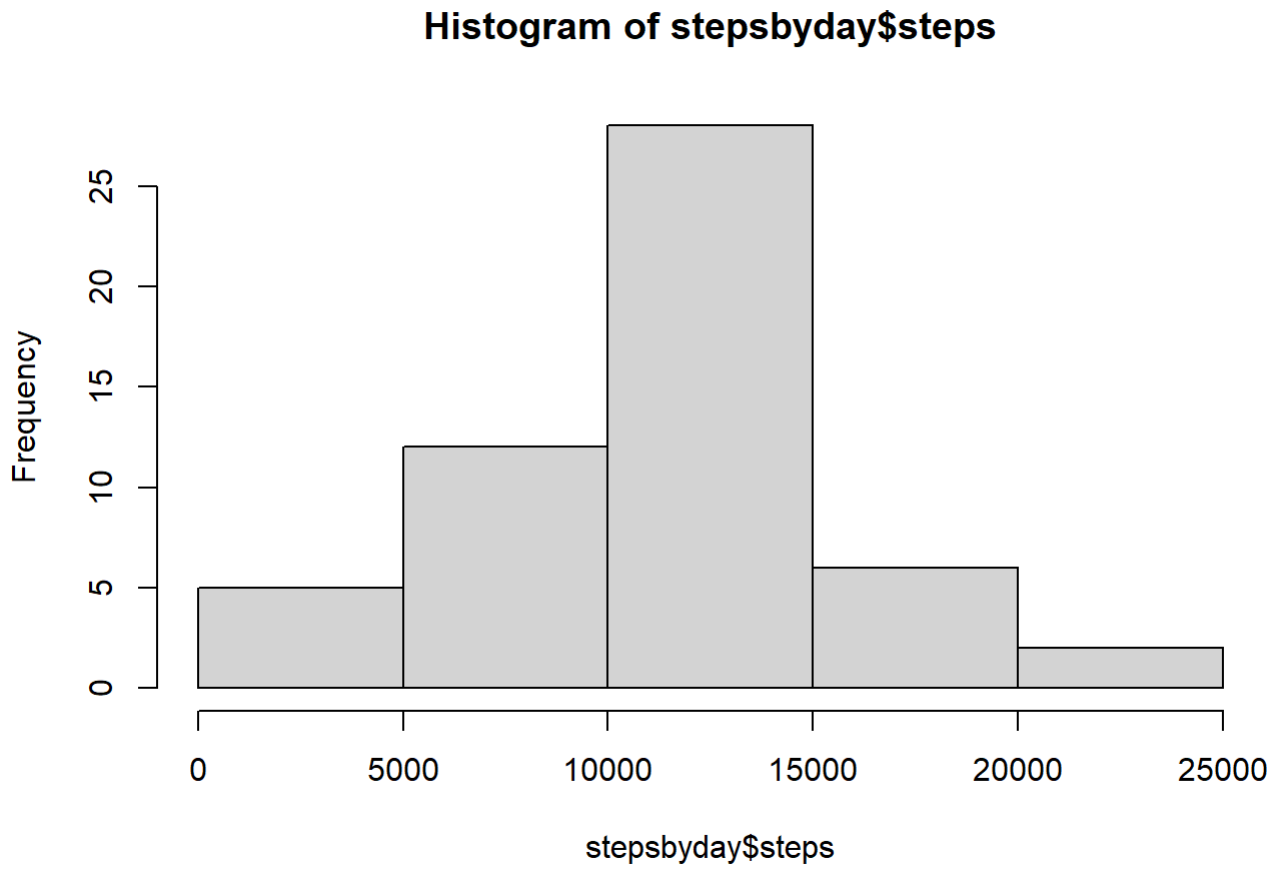
```
row_data=read.csv("./activity.csv")
data=na.omit(row_data)
data$date=as.Date(data$date)
```

let us make another data frame called “stepsbyday” and aggregate the steps by date

```
stepsbyday <- aggregate(steps~date,data=data,sum)
```

let us plot the histogram

```
hist(stepsbyday$steps)
```



we

will compute the mean and median of stepsbyday

```
mean(stepsbyday$steps)
```

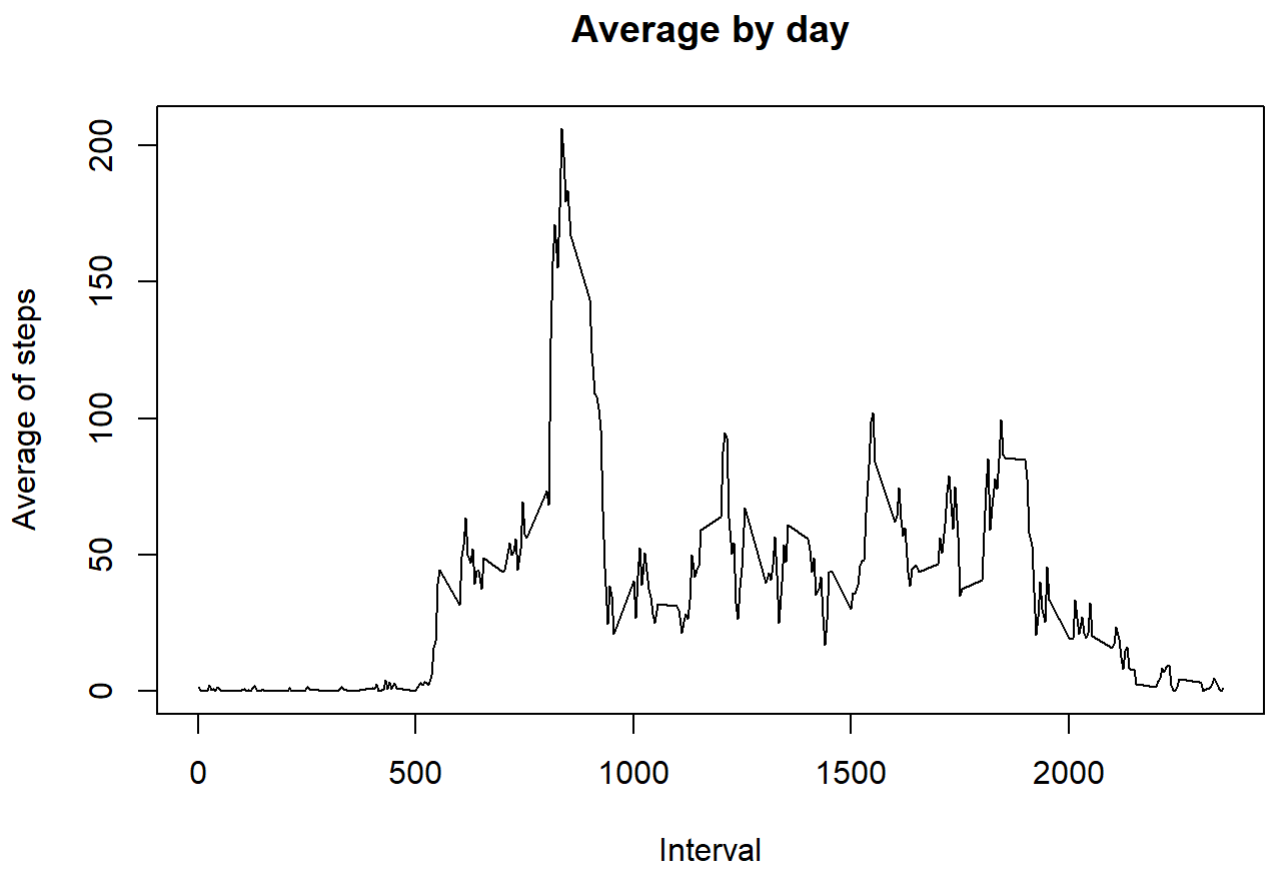
```
## [1] 10766.19
```

```
median(stepsbyday$steps)
```

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

we will now calculate the average by day and plot it

```
averagebyday <- aggregate(steps~interval,data=data,mean)
plot(averagebyday$interval,averagebyday$steps,type='l',main="Average by day",xlab="Interval",yla
b="Average of steps")
```




```
maximum=max(averagebyday$steps)
averagebyday[grepl(maximum,averagebyday$steps),]$interval
```

```
## [1] 835
```

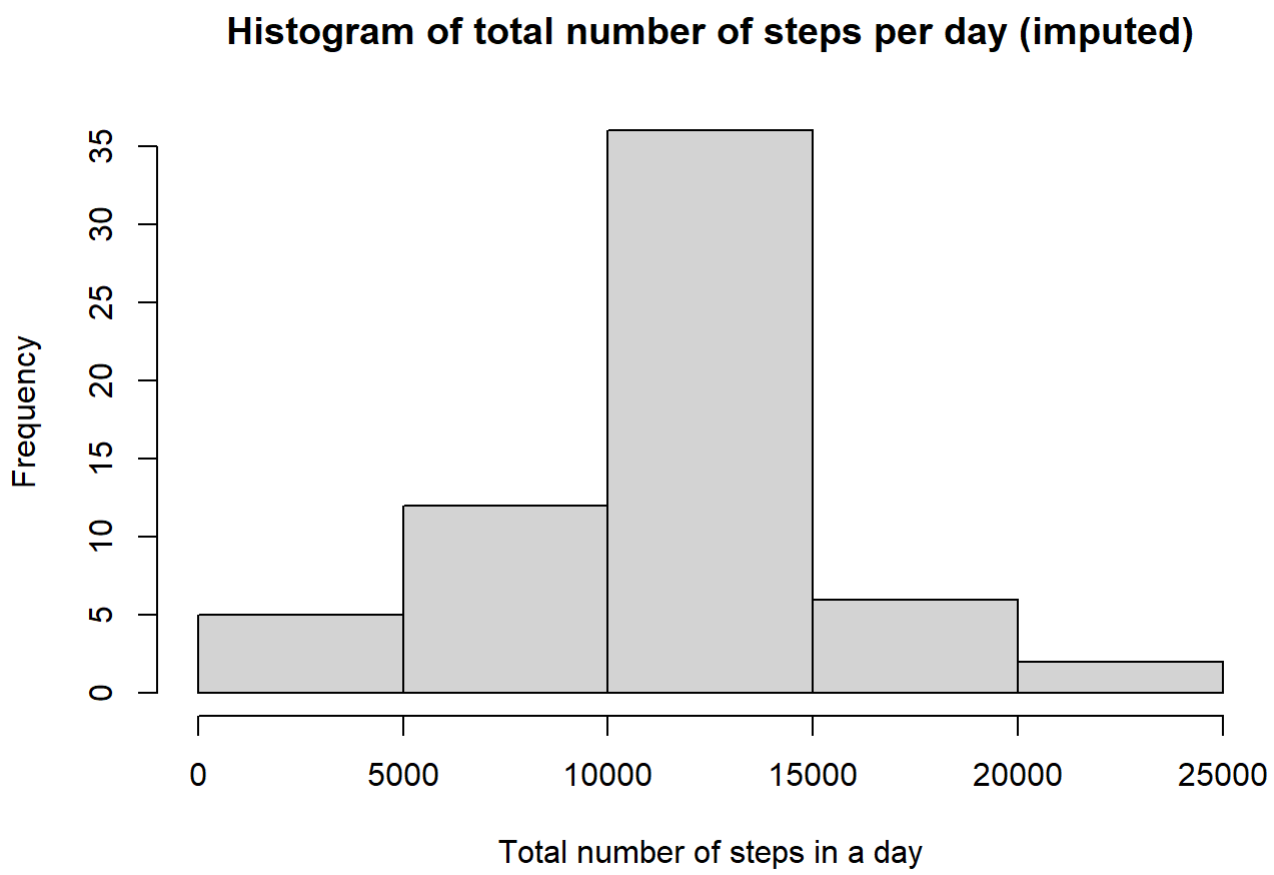
let us replace the number of rows with NA with the mean for 5-minute interval

```
data_imputed <- row_data
for (i in 1:nrow(data_imputed)) {
  if (is.na(data_imputed$steps[i])) {
     markup<-data_imputed$interval[i]
     data_imputed$steps[i]<-averagebyday[averagebyday$interval%in%markup,]$steps

  }
}
head(data_imputed)
```

```
##       steps       date interval
## 1 1.7169811 2012-10-01        0
## 2 0.3396226 2012-10-01        5
## 3 0.1320755 2012-10-01       10
## 4 0.1509434 2012-10-01       15
## 5 0.0754717 2012-10-01       20
## 6 2.0943396 2012-10-01       25
```

```
df_imputed_steps_by_day <- aggregate(steps ~ date, data_imputed, sum)
hist(df_imputed_steps_by_day$steps, main="Histogram of total number of steps per day (imputed)",
     xlab="Total number of steps in a day")
```



```
mean(df_imputed_steps_by_day$steps)
```

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

```
median(df_imputed_steps_by_day$steps)
```

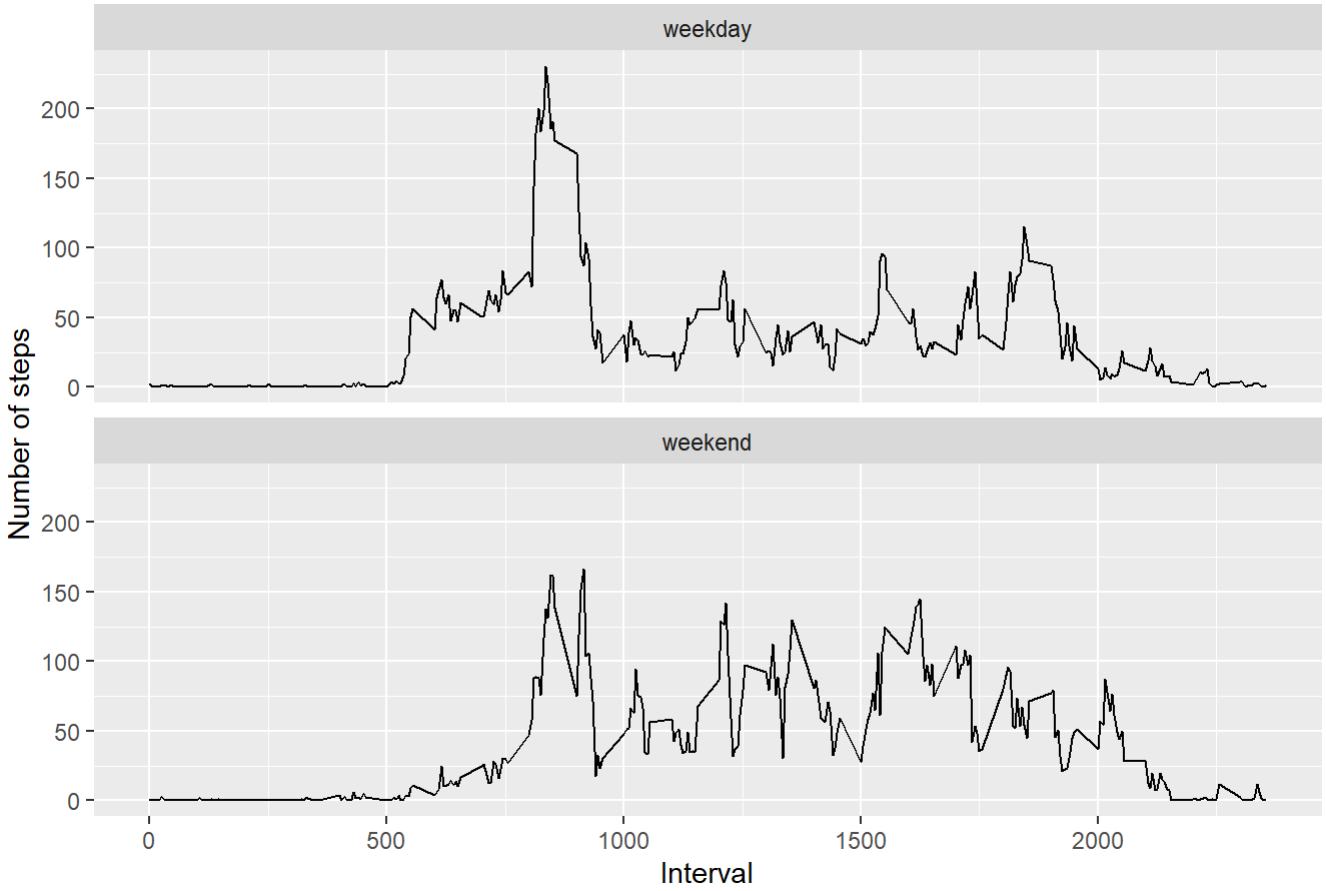
```
## [1] 10766.19
```

Are there differences in activity patterns between weekdays and weekends?

```
data_imputed['type_of_day'] <- weekdays(as.Date(data_imputed$date))
data_imputed$type_of_day[data_imputed$type_of_day  %in% c('Saturday','Sunday') ] <- "weekend"
data_imputed$type_of_day[data_imputed$type_of_day != "weekend"] <- "weekday"
data_imputed$type_of_day <- as.factor(data_imputed$type_of_day)

# calculate average steps by interval across all days
df_imputed_steps_by_interval <- aggregate(steps ~ interval + type_of_day, data_imputed, mean)
library(ggplot2)
# creat a plot
qplot(interval,
      steps,
      data = df_imputed_steps_by_interval,
      type = 'l',
      geom=c("line"),
      xlab = "Interval",
      ylab = "Number of steps",
      main = "") +
  facet_wrap(~ type_of_day, ncol = 1)
```

```
## Warning: Ignoring unknown parameters: type
```


weekday
weekend
Number of steps
Interval
0
50
100
150
200
0
500
1000
1500
2000