Assignment 1 - Week 2 - Reproducible Research

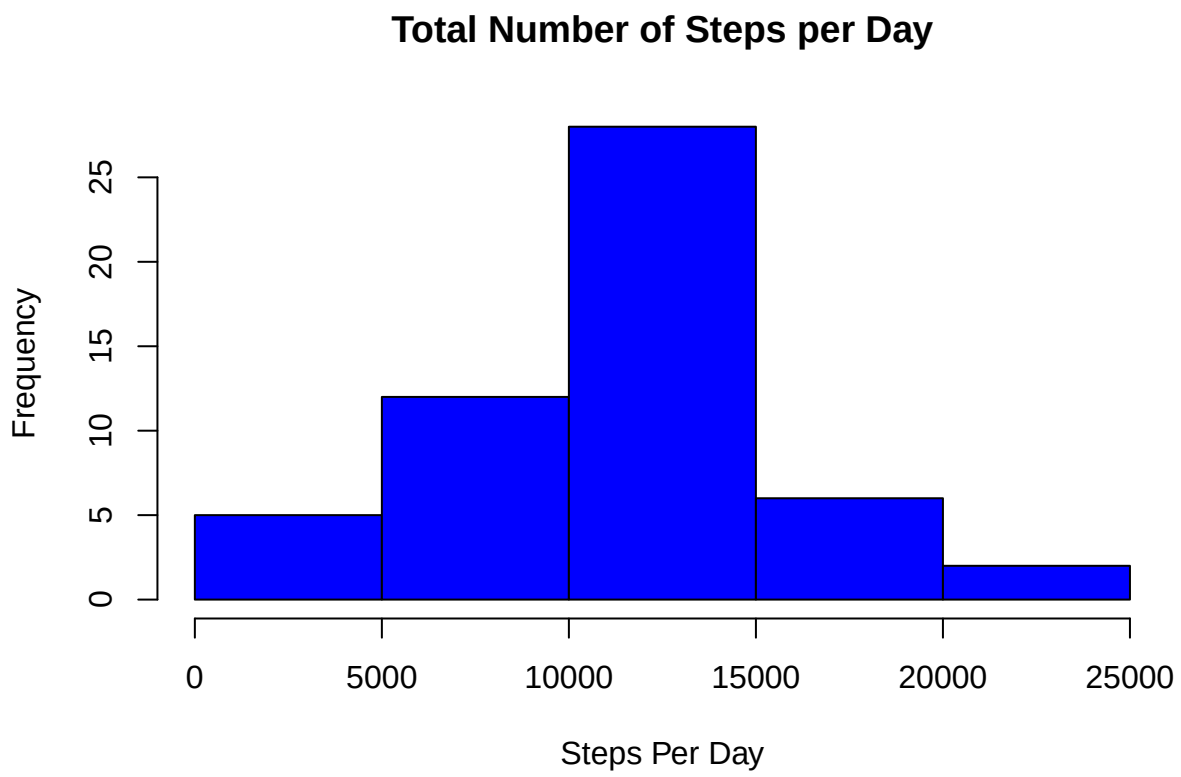
1. Code for reading in the dataset and/or processing the data

Load ggplot library and read the csv file

```
library(ggplot2)
dataset <- read.csv("./activity.csv")
```

2. Histogram of the total number of steps taken each day

```
dataset$Date <- as.Date(dataset$date)
stepsday <- aggregate(steps ~ Date, data=dataset, FUN=sum, na.rm=T)
hist(stepsday$steps, xlab="Steps Per Day", main="Total Number of Steps per Day", col="blue")
```



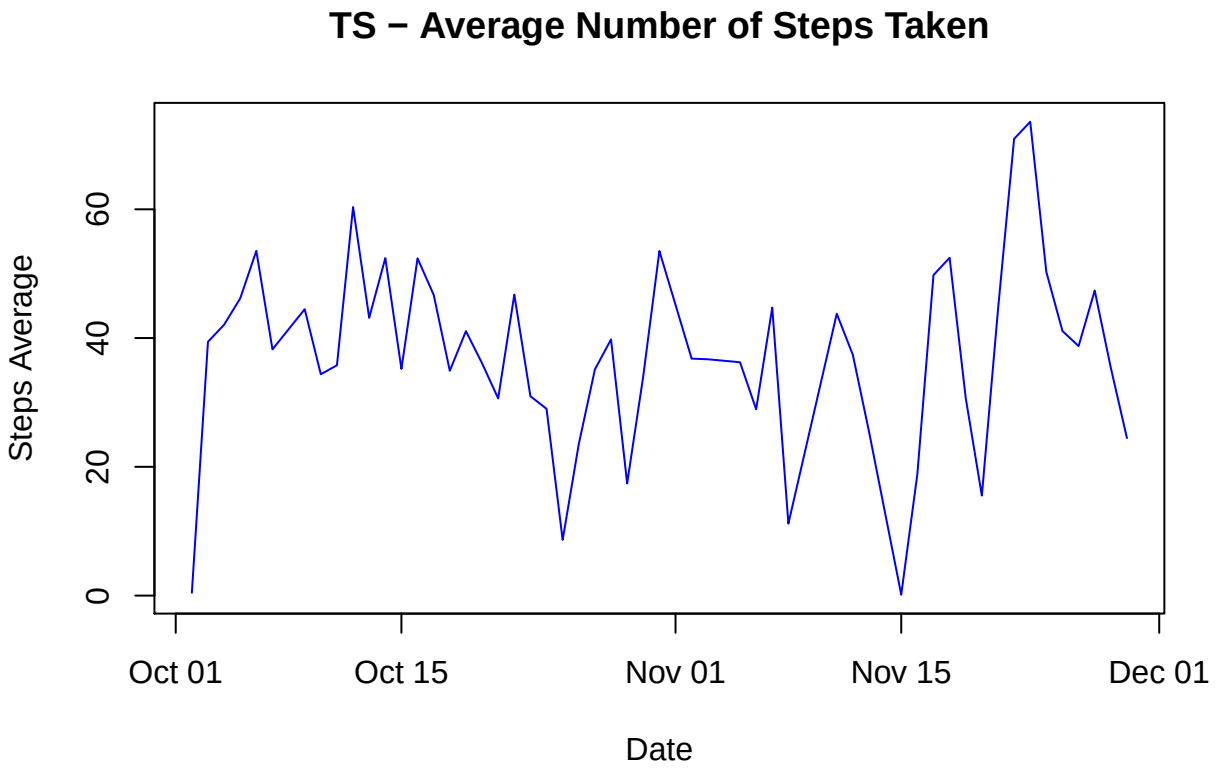
3. Mean and median number of steps taken each day

```
summary(stepsday)
```

```
##       Date               steps
##  Min.   :2012-10-02   Min.   :   41
##  1st Qu.:2012-10-16   1st Qu.: 8841
##  Median :2012-10-29   Median :10765
##  Mean   :2012-10-30   Mean   :10766
##  3rd Qu.:2012-11-16   3rd Qu.:13294
##  Max.   :2012-11-29   Max.   :21194
```

4. Time series plot of the average number of steps taken

```
stepsday_mean <- aggregate(steps ~ Date, data=dataset, FUN=mean, na.rm=T)
plot(stepsday_mean, type="l", ylab="Steps Average", main="TS - Average Number of Steps Taken", col="blu
```



5. The 5-minute interval that, on average, contains the maximum number of steps

```
stepsday_interval <- aggregate(steps ~ interval, data=dataset, FUN=mean, na.rm=T)
print(max_interval <- max(stepsday_interval$steps))
```

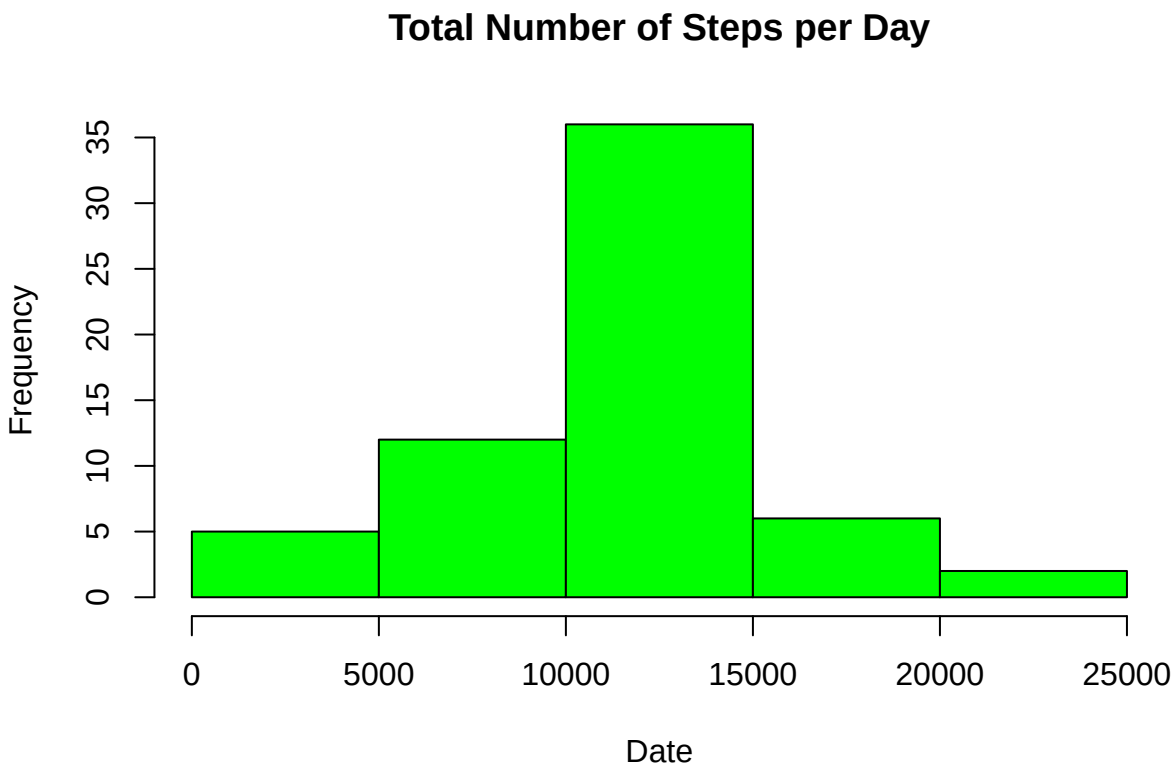
```
## [1] 206.1698
```

6. Code to describe and show a strategy for imputing missing data

```
dataset1 <- dataset
dataset1$steps[is.na(dataset1$steps)] <- mean(dataset1$steps[!is.na(dataset1$steps)])
```

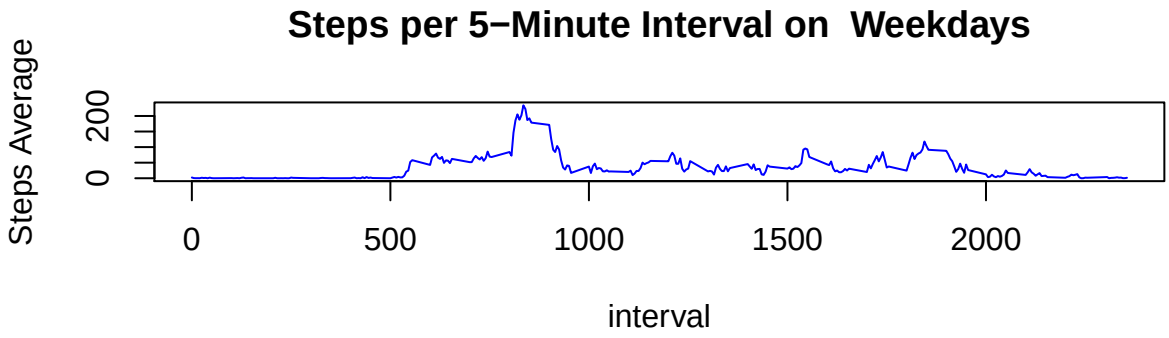
7. Histogram of the total number of steps taken each day after missing values are imputed

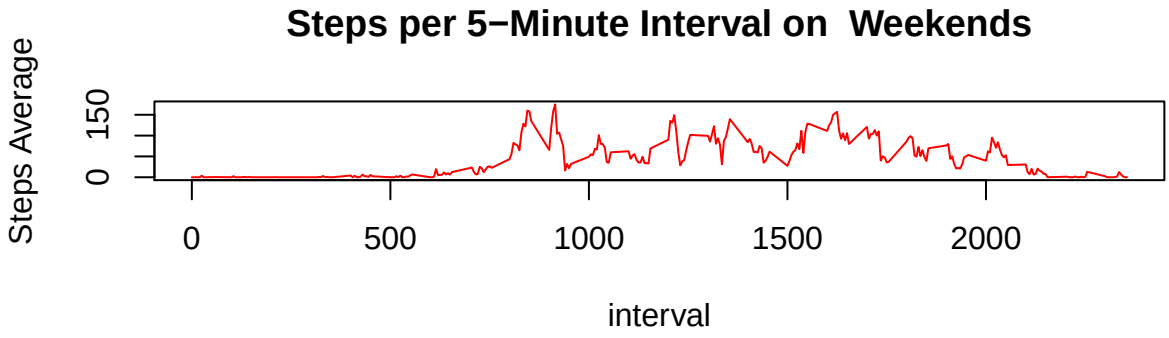
```
stepsday_sum <- aggregate(steps ~ Date, data=dataset1, FUN=sum)
hist(stepsday_sum$steps, xlab="Date", main="Total Number of Steps per Day", col="green")
```



8. Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and

```
dataset2 <- dataset
dataset2$weekd <- weekdays(dataset2$Date)
dataset2_wday <- dataset2[dataset2$weekd %in% c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"),]
dataset2_wend <- dataset2[dataset2$weekd %in% c("Saturday", "Sunday"),]
dataset2_wday_int <- aggregate(steps ~ interval, data=dataset2_wday, FUN=mean, na.rm=T)
dataset2_wend_int <- aggregate(steps ~ interval, data=dataset2_wend, FUN=mean, na.rm=T)
par(mfrow=c(2,1))
plot(dataset2_wday_int, type="l", ylab="Steps Average", main="Steps per 5-Minute Interval on  Weekdays")
plot(dataset2_wend_int, type="l", ylab="Steps Average", main="Steps per 5-Minute Interval on  Weekends")
```






```
dev.off()
```

```
## null device
##           1
```