Reproducibe Research Proyect 1

Abner Aranda

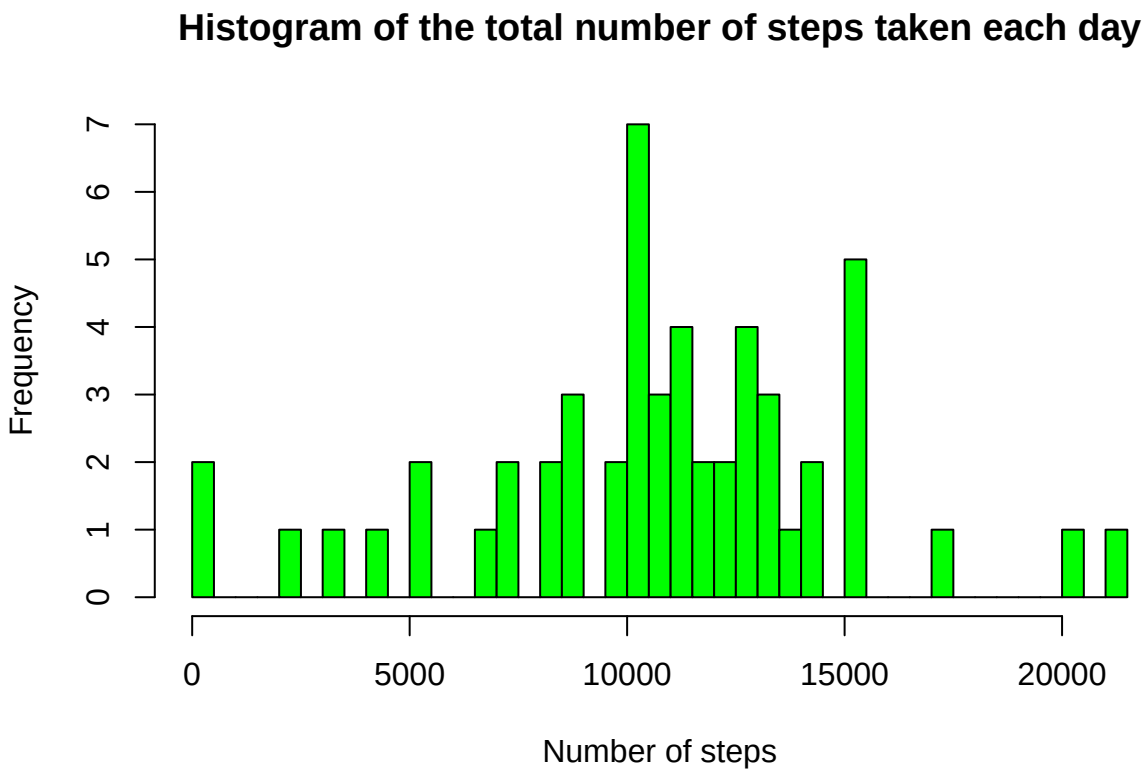
8/9/2021

Loading and preparing data

```
if (!file.exists("activity.csv") )
    {
     dlurl <- 'http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip'
     download.file(dlurl,destfile='repdata%2Fdata%2Factivity.zip',mode='wb')
     unzip('repdata%2Fdata%2Factivity.zip')
    }
activity <- read.csv("activity.csv")
```

What is the mean total number of steps taken per day? We start by creating a histogram of steps per day

```
Act_without_na <- subset(activity, !is.na(activity$steps))
step_per_day <- aggregate(steps ~ date, Act_without_na, sum)
hist(step_per_day $steps, breaks = 53, col = "green", xlab = "Number of steps", main = "Histogram of the
```



Calculate mean and median of the total number of steps taken per day Mean

```
act_mean <- mean(step_per_day$steps)
print(act_mean)
```

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

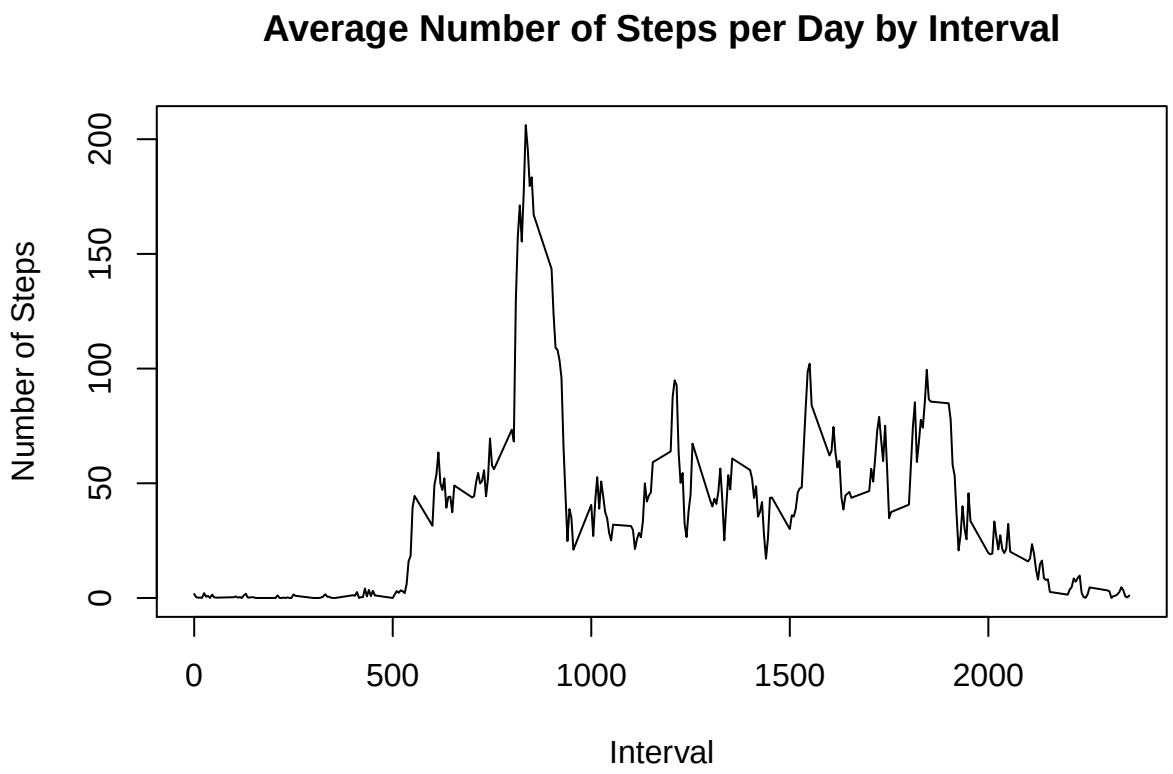
Median

```
act_med <- median(step_per_day$steps)
print(act_med)
```

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

What is the average daily activity pattern? Plot of steps per interval

```
steps_per_interval <- aggregate(steps ~ interval, Act_without_na, mean)
plot(steps_per_interval$interval, steps_per_interval$steps, type="l", xlab="Interval", ylab="Number of S
```



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
max_int <- steps_per_interval[which.max(steps_per_interval$steps),1]
print(max_int)
```

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

How many steps does that interval had?

```
max_int_numb <- steps_per_interval[steps_per_interval$interval == max_int,2]
print(max_int_numb)
```

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

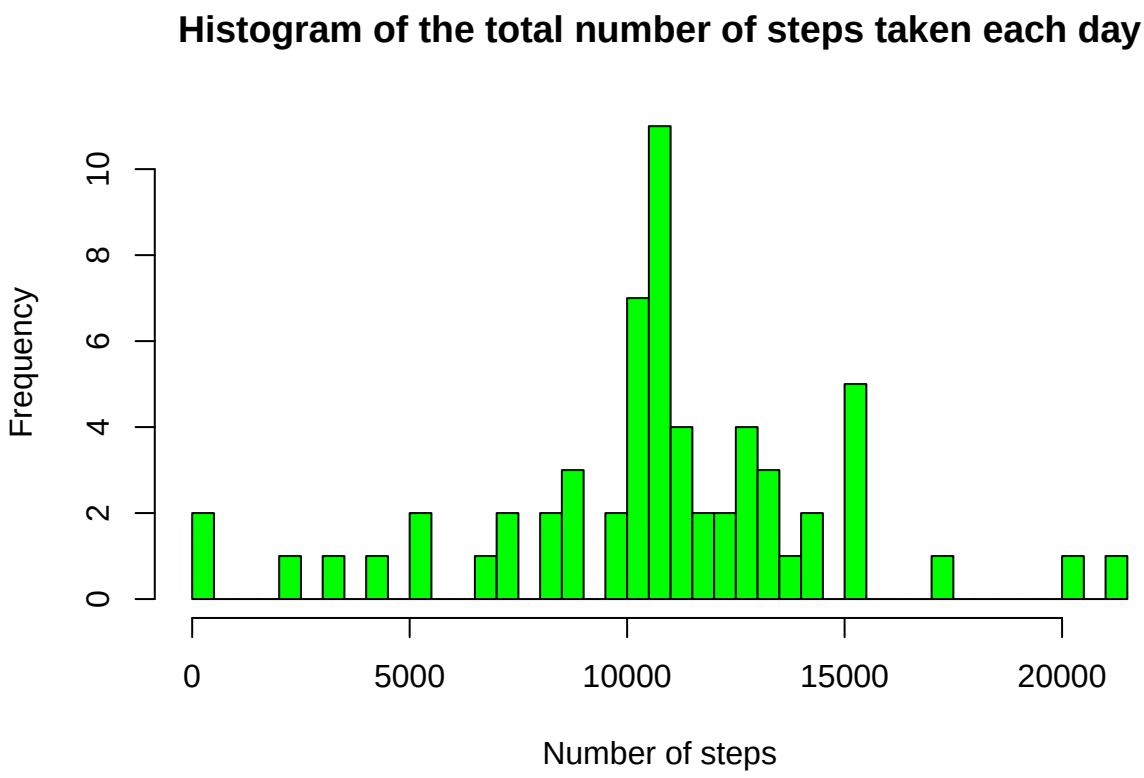
Imputing missing values Calculate and report the total number of missing values in the dataset (i.e. the total number of rows with NAs)

```
sum_NA <- sum(!complete.cases(activity))
sum_NA
```

```
## [1] 2304
```

Create a new dataset that is equal to the original dataset but with the missing data filled in. Create a histogram

```
na_index <- which(is.na(as.character(activity$steps)))
complete_act <- activity
complete_act[na_index, ]$steps <- unlist(lapply(na_index, FUN=function(na_index){steps_per_interval[acti
step_per_day_complete <- aggregate(steps ~ date, data = complete_act, sum)
hist(step_per_day_complete $steps, breaks = 53, col = "green", xlab = "Number of steps", main = "Histog
```



Mean

```
mean(step_per_day_complete$steps)
```

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

Median

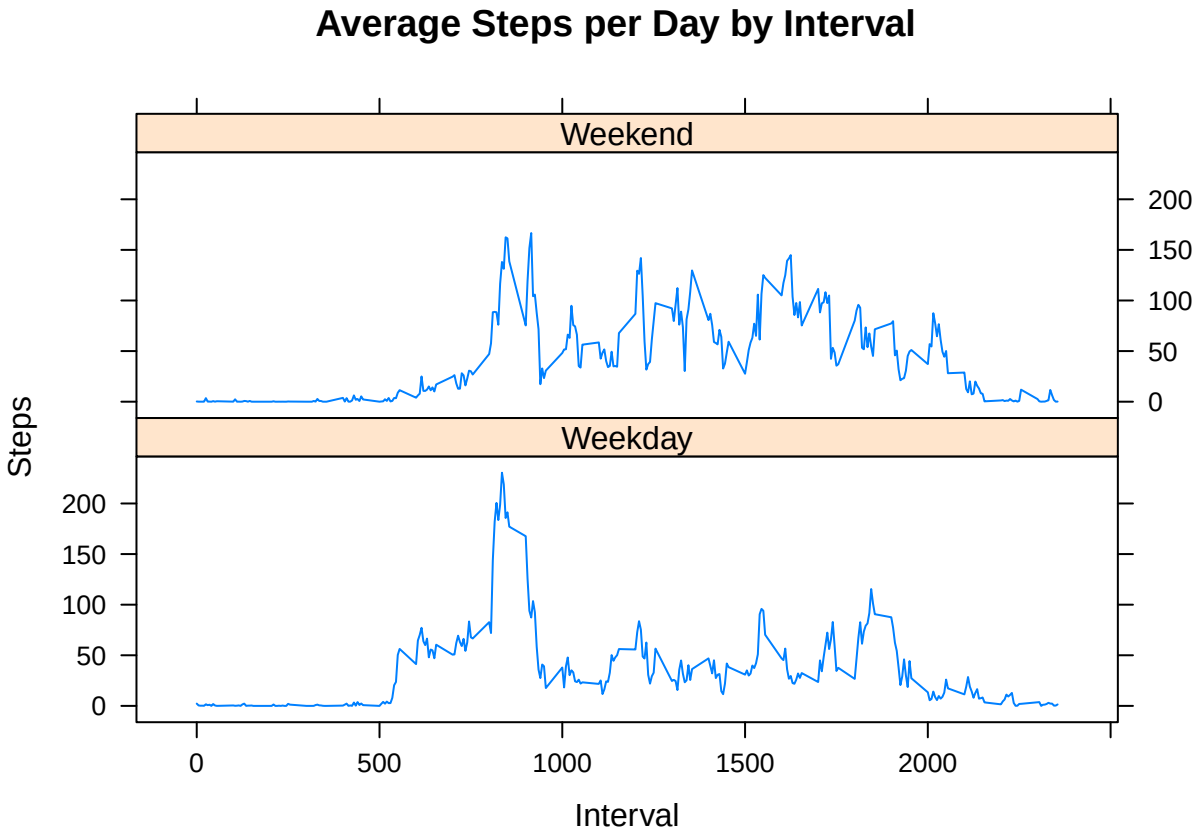
```
median(step_per_day_complete$steps)
```

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

Both mean and median has little to no changecompared with the incomplete data

Are there differences in activity patterns between weekdays and weekends? Lets separate the data between weekend and weekdays

```
complete_act$date <- as.Date(complete_act$date, format = "%Y-%m-%d")
weekdays <- c("Monday","Tuesday","Wednesday","Thursday","Friday")
complete_act $day_week = as.factor(ifelse(is.element(weekdays(as.Date(complete_act $date)),weekdays), "
steps_by_day_type <- aggregate(steps ~ interval + day_week, complete_act, mean)
library(lattice)
xyplot(steps_by_day_type$steps ~ steps_by_day_type$interval| steps_by_day_type$day_week, main="Average
```



The weekends do have more activity, although weekdays have the biggest peak of activity, pressumible during the morning