

Course Project 1

Marco Lopez

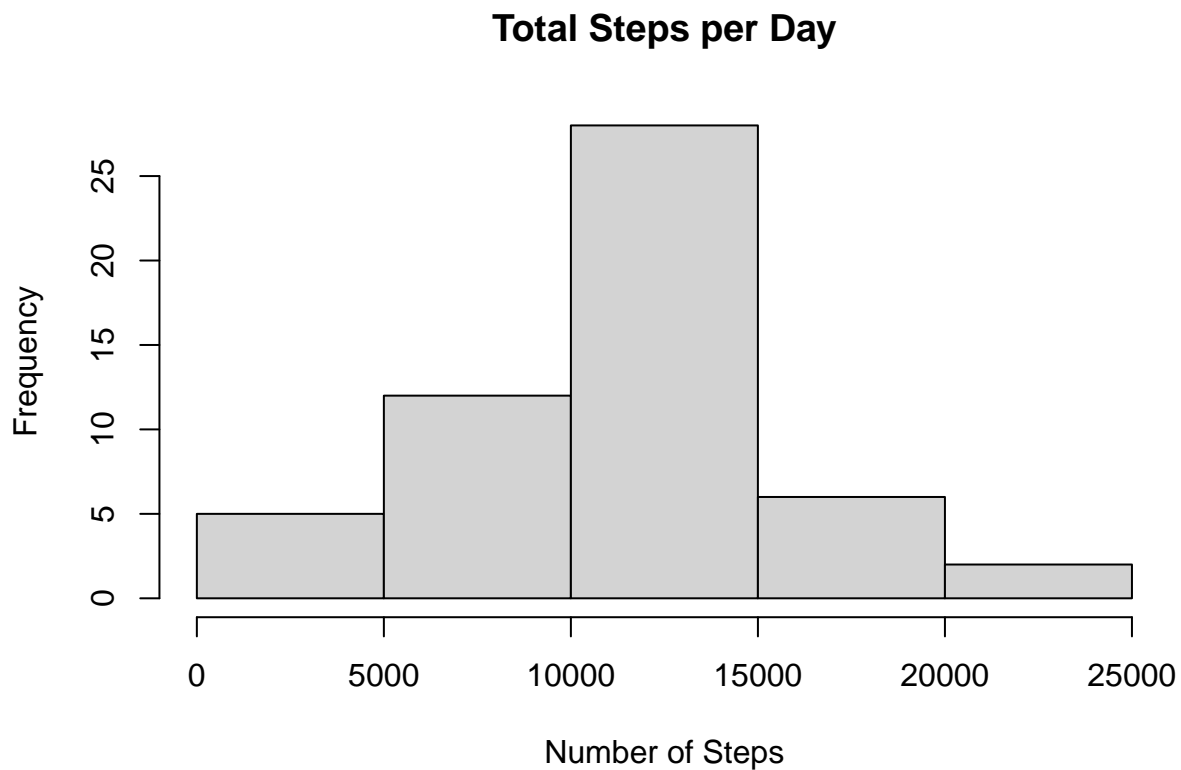
31/3/2022

Loading and preprocessing the data

```
Data <- read.csv(file="activity.csv", header=TRUE)
```

What is mean total number of steps taken per day?

Make a histogram of the total number of steps taken each day



Mean and median number of steps taken each day

```
meanS <- mean(TStepforday$steps, na.rm = TRUE)
medD <- median(TStepforday$steps, na.rm = TRUE)
summary(meanS)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10766   10766   10766   10766   10766   10766
```

```
summary(medD)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10765   10765   10765   10765   10765   10765
```

What is the average daily activity pattern?

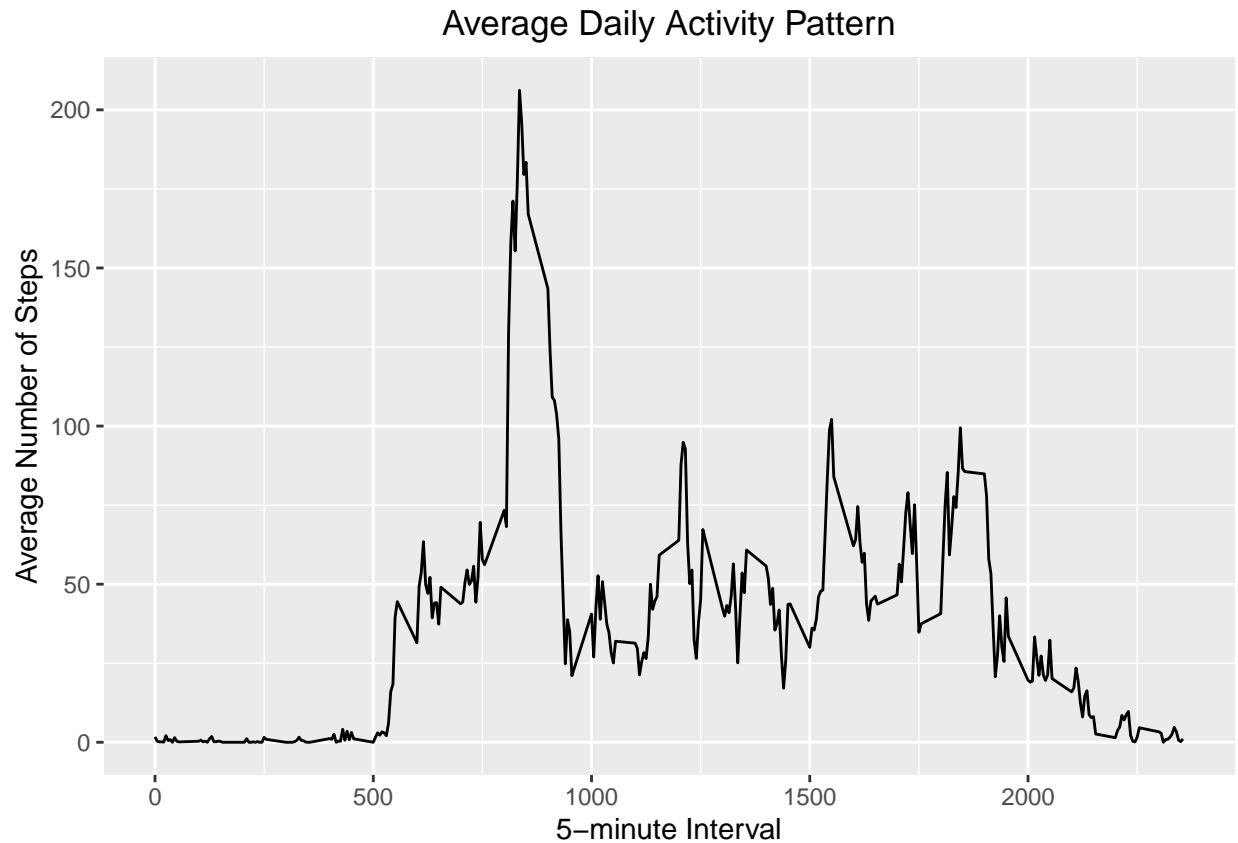
Make a time series plot

Time series plot of the average number of steps taken (averaged across all days) versus the 5-minute intervals

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
meanSteps <- aggregate(steps ~ interval, Data, mean)
ggplot(data = meanSteps, aes(x = interval, y = steps)) + geom_line() + ggtitle("Average Daily Activity Pattern")
```



Report give the 5-minute interval that, on average, contains the maximum number of steps

```
maxstepinterval <- meanSteps[which.max(meanSteps$steps),]
summary(maxstepinterval)
```

```
##      interval      steps
## Min.   :835  Min.   :206.2
## 1st Qu.:835  1st Qu.:206.2
## Median :835  Median :206.2
## Mean   :835  Mean   :206.2
## 3rd Qu.:835  3rd Qu.:206.2
## Max.   :835  Max.   :206.2
```

Calculate and report the total number of missing values

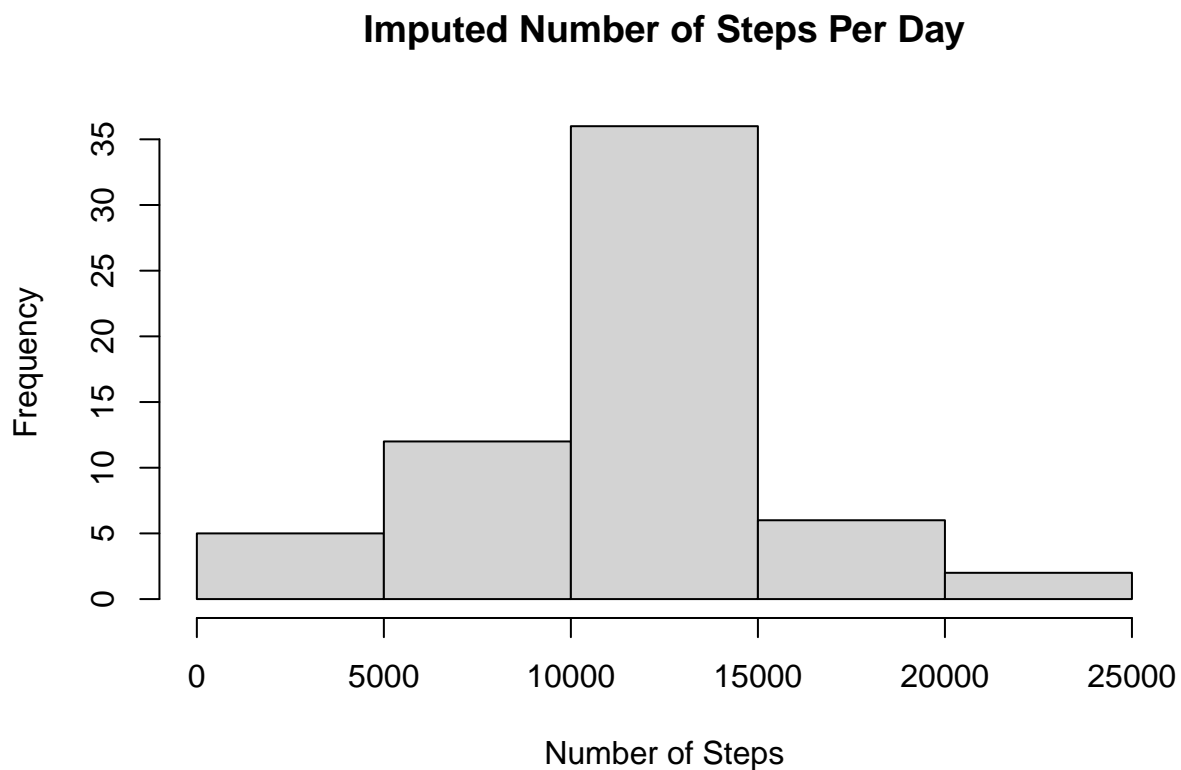
```
missingV <- is.na(Data$steps)
```

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

```
miss_Data <- transform(Data, steps = ifelse(is.na(Data$steps),  
                                           + meanSteps$steps[match(Data$interval,  
                                                                    + meanSteps$interval)],  
                                           + Data$steps))
```

Report contain a histogram of the total number of steps taken each day after missing values were imputed

```
iStepsbInt <- aggregate(steps ~ date, miss_Data, FUN=sum)  
hist(iStepsbInt$steps, main = "Imputed Number of Steps Per Day", xlab = "Number of Steps")
```



```
impMeanSteps <- mean(iStepsbInt$steps, na.rm = TRUE)  
impMedSteps <- median(iStepsbInt$steps, na.rm = TRUE)  
diffMean = impMeanSteps - meanSteps  
diffTotal = sum(iStepsbInt$steps) - sum(TStepforday$steps)
```

Are there differences in activity patterns between weekdays and weekends?

Create a new factor variable in the dataset with two levels – “weekday” and “weekend”

```
DayType <- function(date) {  
  day <- weekdays(date)  
  if (day %in% c('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday')) return ("weekday")  
  else if (day %in% c('Saturday', 'Sunday'))  
    return ("weekend")  
  else stop ("Format no valid") }  
}
```

#For this part the weekdays() function may be of some help here.

```
miss_Data$dateType <- ifelse(as.POSIXlt(miss_Data$date)$wday %in% c(0,6), 'weekend', 'weekday')
```

Make a panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends

```
AVG <- aggregate(steps ~ interval + dateType, data=miss_Data, mean)  
ggplot(AVG, aes(interval, steps)) + geom_line() + facet_grid(dateType ~ .) + xlab("5-minute interval")
```

