PA1_template.Rmd

Marcos Melendez

11/27/2020

Show any code that is needed to load the data and process and transform

```
activity <- read.csv("C:/Users/melen/Desktop/Exploratory Data Analysis//activity.csv", header = TRUE)

View(activity)

dim(activity)
```

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

```
str(activity)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : chr  "2012-10-01" "2012-10-01" "2012-10-01" "2012-10-01" ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

```
summary(activity)
```

```
##      steps            date              interval     
##  Min.   :  0.00   Length:17568       Min.   :   0.0  
##  1st Qu.:  0.00   Class :character   1st Qu.: 588.8  
##  Median :  0.00   Mode  :character   Median :1177.5  
##  Mean   : 37.38                      Mean   :1177.5  
##  3rd Qu.: 12.00                      3rd Qu.:1766.2  
##  Max.   :806.00                      Max.   :2355.0  
##  NA's   :2304
```

```
head(activity)
```

```
##   steps       date interval
## 1    NA 2012-10-01        0
## 2    NA 2012-10-01        5
## 3    NA 2012-10-01       10
## 4    NA 2012-10-01       15
## 5    NA 2012-10-01       20
## 6    NA 2012-10-01       25
```

```
tail(activity)
```

```
##       steps       date interval
## 17563    NA 2012-11-30     2330
## 17564    NA 2012-11-30     2335
## 17565    NA 2012-11-30     2340
## 17566    NA 2012-11-30     2345
## 17567    NA 2012-11-30     2350
## 17568    NA 2012-11-30     2355
```

What is mean total number of steps taken per day?

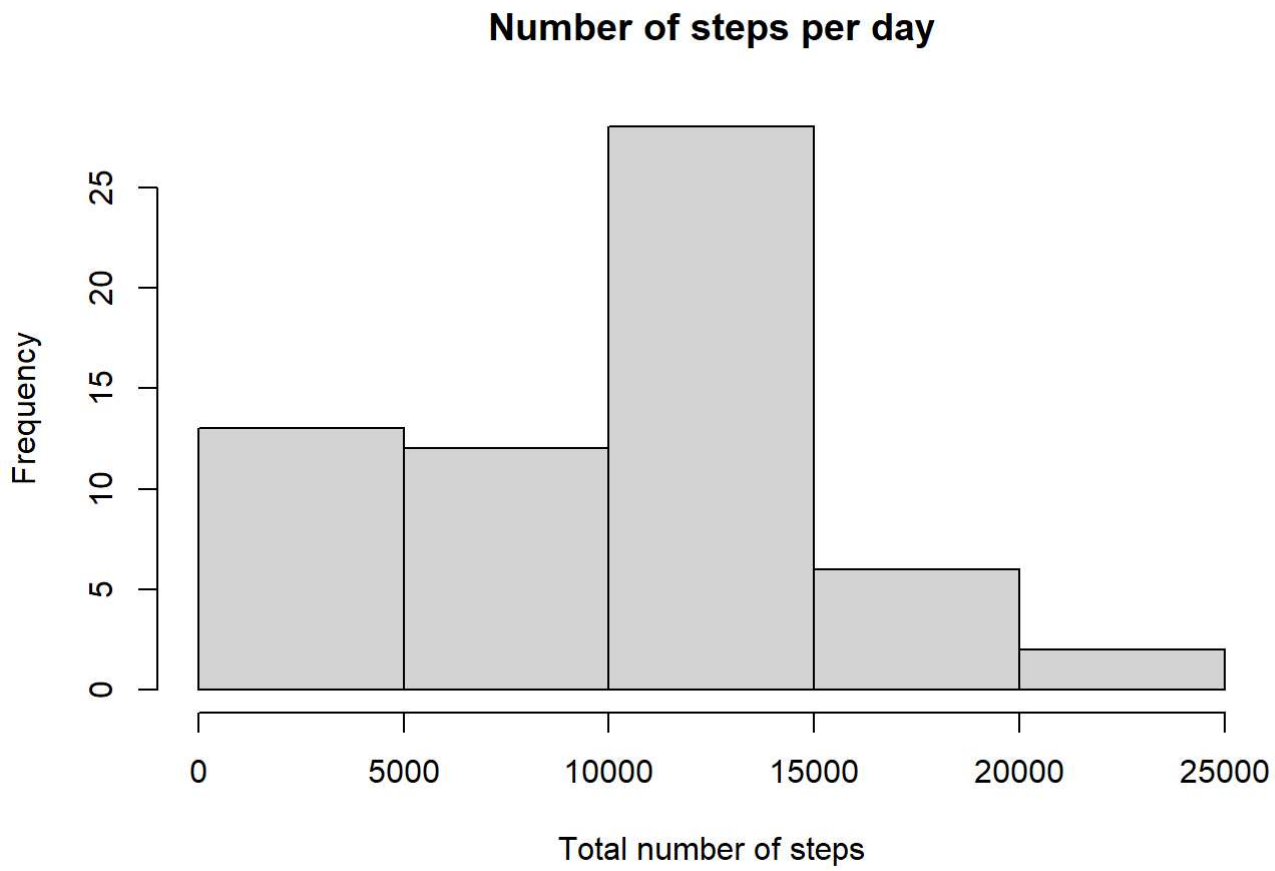
```
total_number_of_steps <- with(activity, tapply(steps, as.factor(activity$date), sum, na.rm = T))


summary(total_number_of_steps)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    6778   10395    9354   12811   21194
```

```
filtered_steps <- activity[!is.na(activity$steps),]
```

```
#Histogram
hist(total_number_of_steps, main = "Number of steps per day", xlab = "Total number of steps")
```




```
#Report the the mean and median of the total number of steps taken per day

summary(total_number_of_steps)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    6778   10395    9354   12811   21194
```

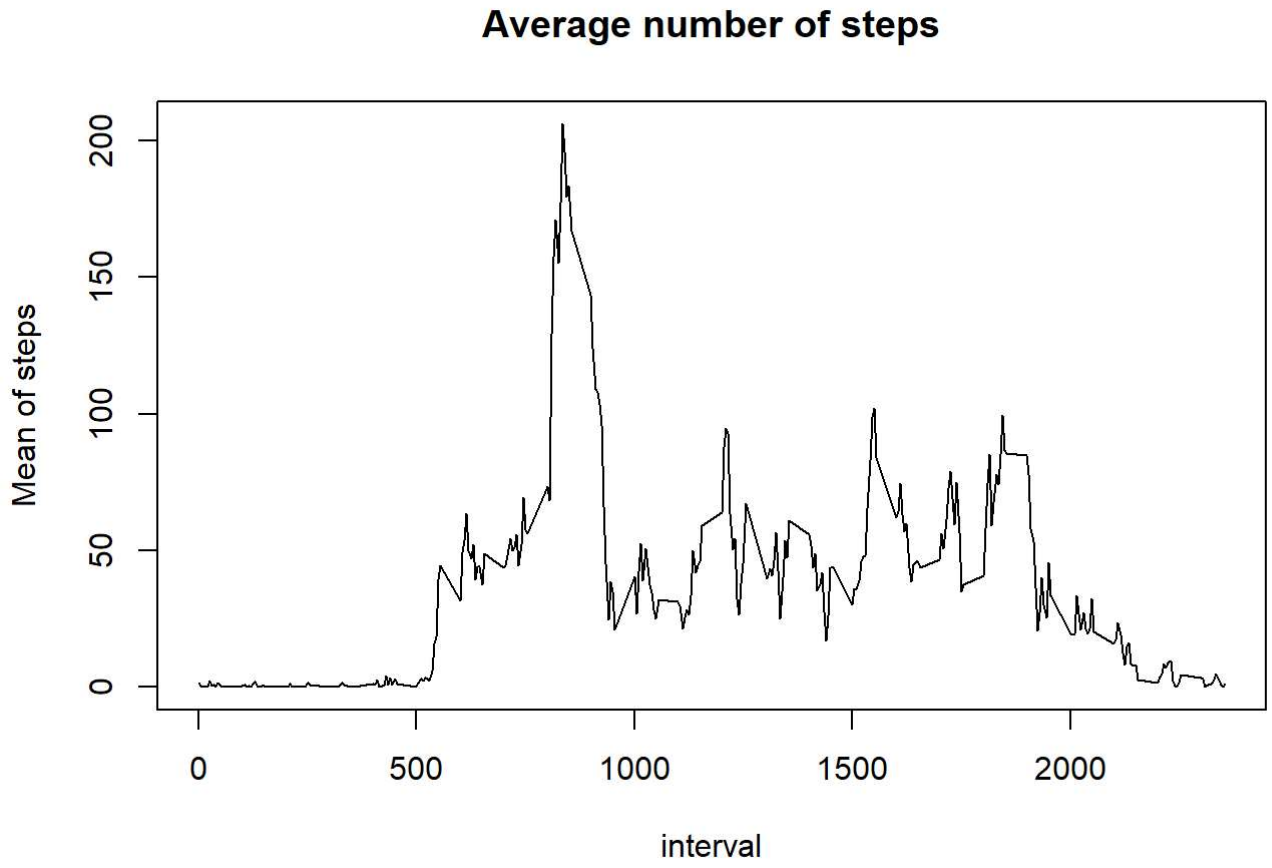
What is the average daily activity pattern?

```
mean_steps <- with(filtered_steps, tapply(steps, filtered_steps$interval, mean))

#mean_steps <- with(activity, tapply(steps, filtered_steps$interval, mean))

interval <- levels(as.factor(filtered_steps$interval))

plot(interval, mean_steps, type = "l", main = "Average number of steps", xlab = "interval", ylab
= "Mean of steps")
```



Imputing missing values

```
length(filtered_steps$steps)
```

```
## [1] 15264
```

Devise a strategy to fill all of the missing values in the data set

```
missing_data <- activity[is.na(activity$steps),]

newdataset <- rbind(filtered_steps, missing_data)
table <- data.frame(mean_steps, interval)
```

Are there differences in activity patterns between weekdays and weekends

library(ggplot2)

newdataset$days <- weekdays(as.Date(newdataset$date))