Fitbit

Marwa

12/16/2020

Activity Monitoring Data

Reading the data

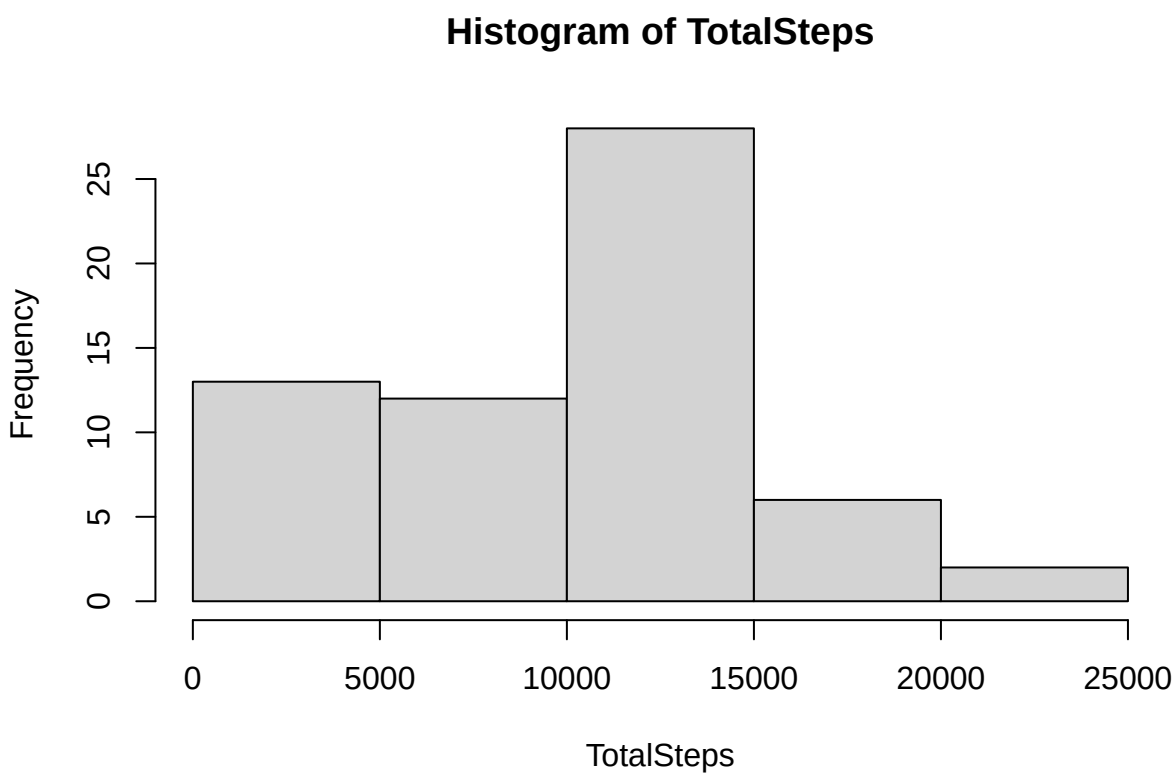
```
activity <- read.csv("activity.csv")
```

Transforming the data

```
TotalSteps <- tapply(activity$steps,activity$date,sum, na.rm=TRUE)
```

Plot for steps taken each day

```
hist(TotalSteps)
```



The mean and median total number of steps taken per day

```
Mean <- mean(TotalSteps)
Median <- median(TotalSteps)

Mean
```

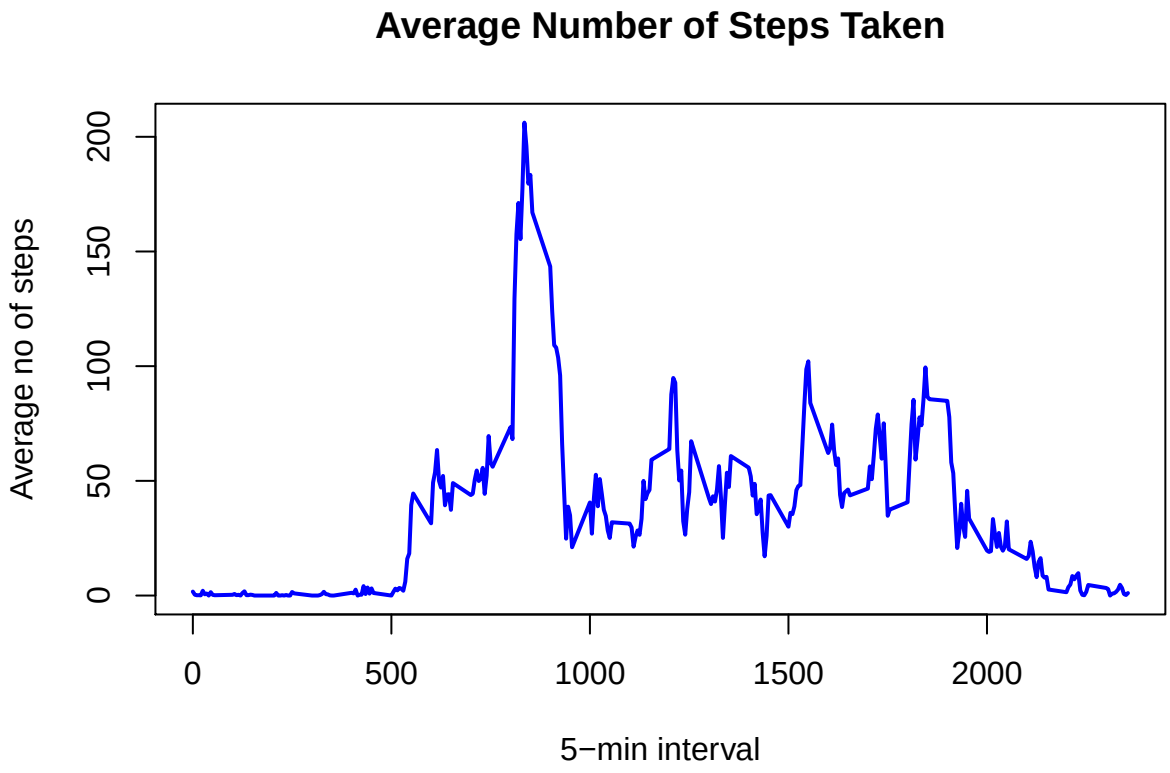
```
## [1] 9354.23
```

```
Median
```

```
## [1] 10395
```

Time series plot of the average number of steps taken

```
average_step <- aggregate(steps ~ interval, data = activity, mean, na.rm = TRUE)
plot(average_step$interval, average_step$steps, type = "l", lwd = 2, col = "blue",
     main = "Average Number of Steps Taken", axes =TRUE,
     xlab = "5-min interval", ylab = "Average no of steps")
```



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

```
av <- tapply(activity$steps, activity$interval, mean, na.rm = T)
which.max(av)
```

```
## 835
## 104
```

average_steps Code to describe and show a strategy for imputing missing data a. the total number of missing values in the dataset (i.e. the total number of rows with NAs)

```
length(which(is.na(activity$steps)))
```

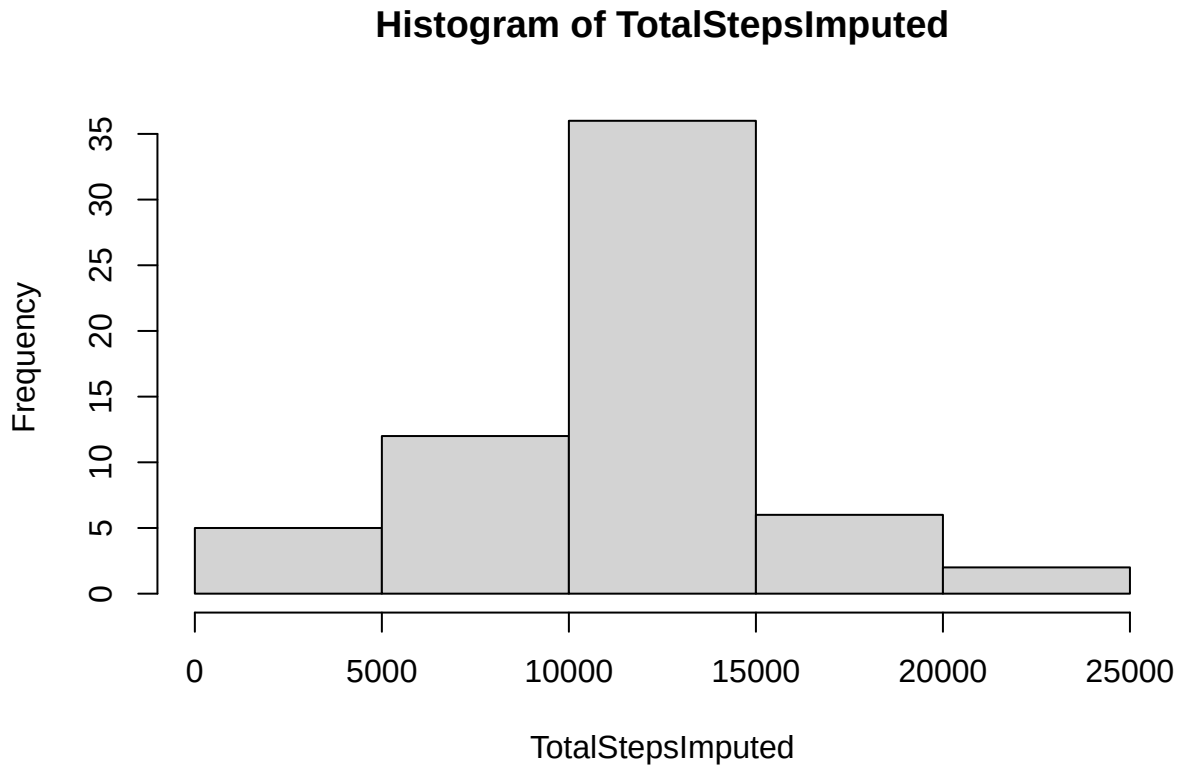
```
## [1] 2304
```

b. Imputed values

```
missingdata_Imputed <- activity
missingdata_Imputed$steps <- impute(activity$steps, fun=mean)
```

plotting a histogram total number of steps taken each day

```
TotalStepsImputed <- tapply(missingdata_Imputed$steps,missingdata_Imputed$date,sum, na.rm=TRUE)
hist(TotalStepsImputed)
```



The New median and mean

```r
Mean <- mean(TotalStepsImputed)
Median <- median(TotalStepsImputed)

Mean
```

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

```r
Median
```

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