My First R Marckdown Doc

My First markdwon file

I used I used the info here (https://rpubs.com/sureshbk/358994) and here (https://rstudio-pubs-static.s3.amazonaws.com/300758_367ce30144b44cd4901bedfa279bb64c.html) to complete the assignment.

I downloaded the data set, and I set my directory

```
setwd ("C:/Users/Charlotte/Documents/GitHub/RepData_PeerAssessment1")
```

Open/load the file

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

spent an hr with errors trying to figure out how install packages in Kniter until I figure out on the net U don't. Why wouldn't the lectures cover that … Wow? I used 1. install.packages("dplyr") 2. install.packages("ggplot2") 3. install.packages("magrittr") 4. library(tidyverse)

Calculate the total number of steps taken per day?

Get number of steps per date and sum results removing null

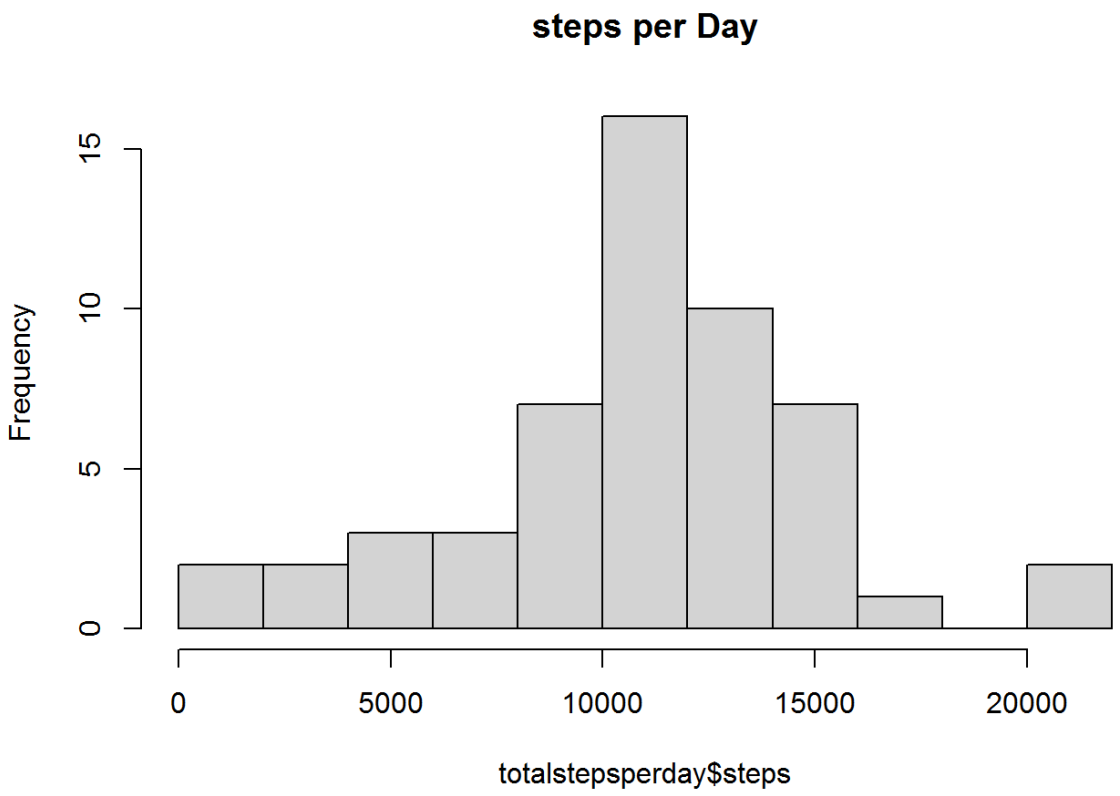
```
totalstepsperday <- aggregate(steps ~ date, data = DataFrame, FUN = sum, na.rm = TRUE)
```

put date in YYYYMMDD format

```
DataFrame$date <- as.Date(DataFrame$date, "%Y-%m-%d")
```

Make a histogram of the total number of steps taken each day. convert dates first A "Histograms are used to show distributions of variables while bar charts are used to compare variables. Via Forbes mag (https://www.forbes.com/sites/naomirobbins/2012/01/04/a-histogram-is-not-a-bar-chart/?sh=27f69e816d77)

```
hist(totalstepsperday$steps, main = "steps per Day", breaks = 10)
```



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

calculate the mean:

```
mean_steps <- mean(totalstepsperday$steps)
mean_steps
```

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

#Calculate median

```
median_steps <- median(totalstepsperday$steps)
median_steps
```

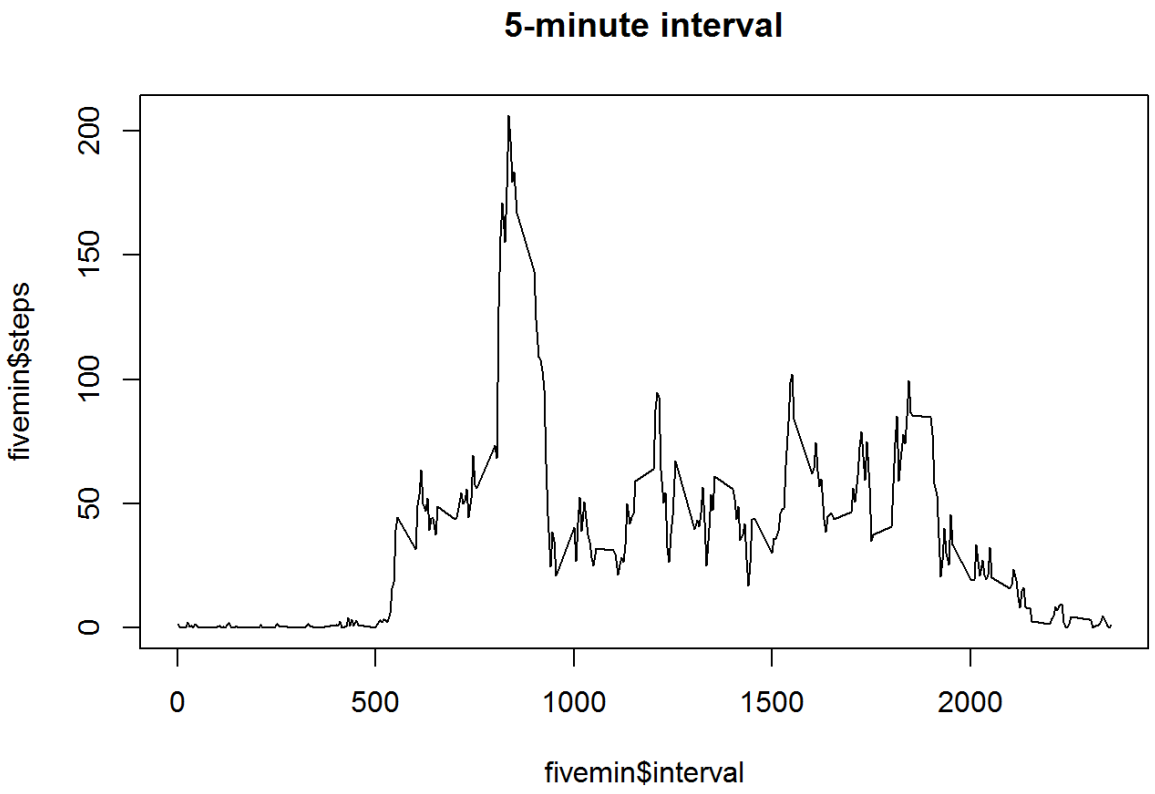
```
## [1] 10765
```

NEXT What is the average daily activity pattern?

Make a time series plot (of type 1) the 5-minute interval (x-axis)and the average number of steps taken, averaged across all days (y-axis).

#Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps? #From the DataFrame create dataset making Steps a dependent variable of interval

```
fivemin <- aggregate(steps ~ interval, data = DataFrame, FUN = mean, na.rm = TRUE)
plot (x = fivemin$interval, y = fivemin$steps, type = "l", main = "5-minute interval")
```



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

```
five_min_maxsteps <- fivemin$interval[which.max(fivemin$steps)]
five_min_maxsteps
```

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

Imputing missing values

Calculate and report the total number of missing values in the dataset Install tidyverse to replace missing values Duplicate the dataset

```
activity2 <- DataFrame
```

#Identify missing values of steps in the duplicated DataFrame

```
nas <- is.na(activity2$steps)
sum(is.na(activity2))
```

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

replace missing value with in activyt2

```
avg_interval <- tapply(activity2$steps, activity2$interval, mean, na.rm=TRUE, simplify = TRUE)
```

Make a panel plot containing a time series plot (-= L) of the 5-minute interval

(x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis). create number of steps per date from dataset with replaced NM values

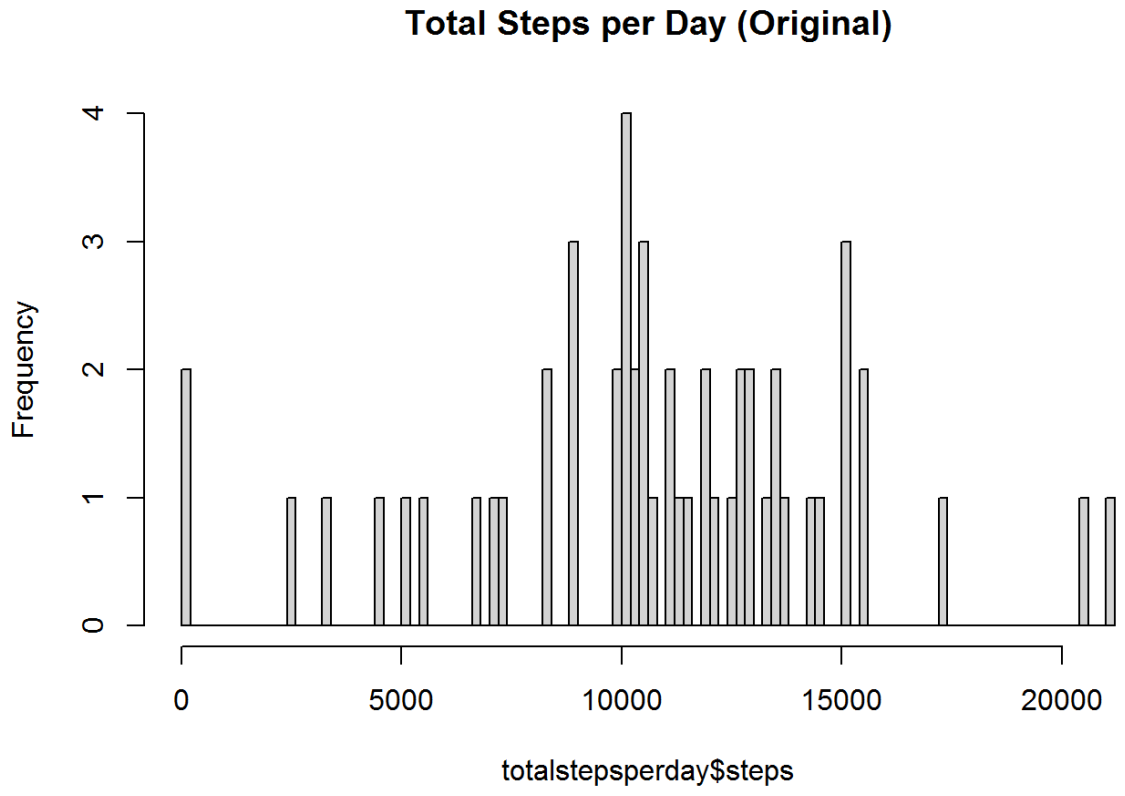
```
totalstepsperday2 <- aggregate(steps ~ date, data = DataFrame, FUN = sum, na.rm = TRUE)
```

#setup frame for two graphs

```
par(mfrow=c(1,2))
```

Histogram with the orginal dataset

```
hist(totalstepsperday$steps,
     main="Total Steps per Day (Original)",
     breaks=100)
```



```
hist(totalstepsperday2$steps,
     main = "Total Steps per Day (no-NA)",
     breaks=100)
```