assignment1

Jack Carbaugh

9/22/2021

R Markdown

This loads in the library Tidyverse

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Create the mtcars dataframe

```
# call built-in data mtcars.
data(mtcars)
```

filter out the row entries where mpg < 20

```
# Select only car models where mpg<20
mtcars_mpg2 <- mtcars[mtcars$mpg < 20,]
```

Limit columns to just mpg, cyl, disp, hp, gears

```
# Reduce the variables to mpg, cyl, disp, hp, gears
mtcars_mpg2 <- mtcars_mpg2[, c(1,2,3,4,10)]
```

Source the hand_functions.R file so that we can utilize its functions

```
# read the R file hand_functions.R so that it can be used
# notice that with echo = TRUE
source(file = "hand_functions.R", echo = TRUE)
```

```
##
## > sum_special <- function(df_x) {
## +     try(if (!is.data.frame(df_x))
## +         stop("Input data must be a data frame."))
## +     sp_means <- apply(df_ .... [TRUNCATED]
```

Uses the sum_special function that is present in hand_functions.R. This function takes in a dataframe is input. If the input is not a dataframe, the function is halted and an error meassge is given. Otherwise, the function gets the mean, variance, covariance, and correlation of the dataframe, adds each statistic into a list, and returns said list. We can also tun the debugger line by line (while using line breaks) to test our code.

```r
# Now use the function from hand_functions.R
sp_out <- sum_special(mtcars_mpg2)
sp_out
```

```
## $sp_means
##        mpg        cyl       disp         hp       gear
##  15.900000   7.555556 313.811111 191.944444   3.444444
##
## $sp_var
##          mpg          cyl         disp           hp         gear
##    7.5258824    0.7320261 9438.7645752 3253.5849673    0.6143791
##
## $sp_cov
##                mpg        cyl       disp          hp        gear
## mpg     7.5258824 -1.3176471 -188.79529  -75.81176   0.6352941
## cyl    -1.3176471  0.7320261   64.71111   28.44444  -0.2614379
## disp -188.7952941 64.7111111 9438.76458 2679.60065 -34.1934641
## hp    -75.8117647 28.4444444 2679.60065 3253.58497  15.2026144
## gear    0.6352941 -0.2614379  -34.19346   15.20261   0.6143791
##
## $sp_cor
##             mpg        cyl       disp          hp       gear
## mpg   1.0000000 -0.5613802 -0.7083614 -0.4844811  0.2954459
## cyl  -0.5613802  1.0000000  0.7784989  0.5828450 -0.3898406
## disp -0.7083614  0.7784989  1.0000000  0.4835389 -0.4490217
## hp   -0.4844811  0.5828450  0.4835389  1.0000000  0.3400314
## gear  0.2954459 -0.3898406 -0.4490217  0.3400314  1.0000000
```

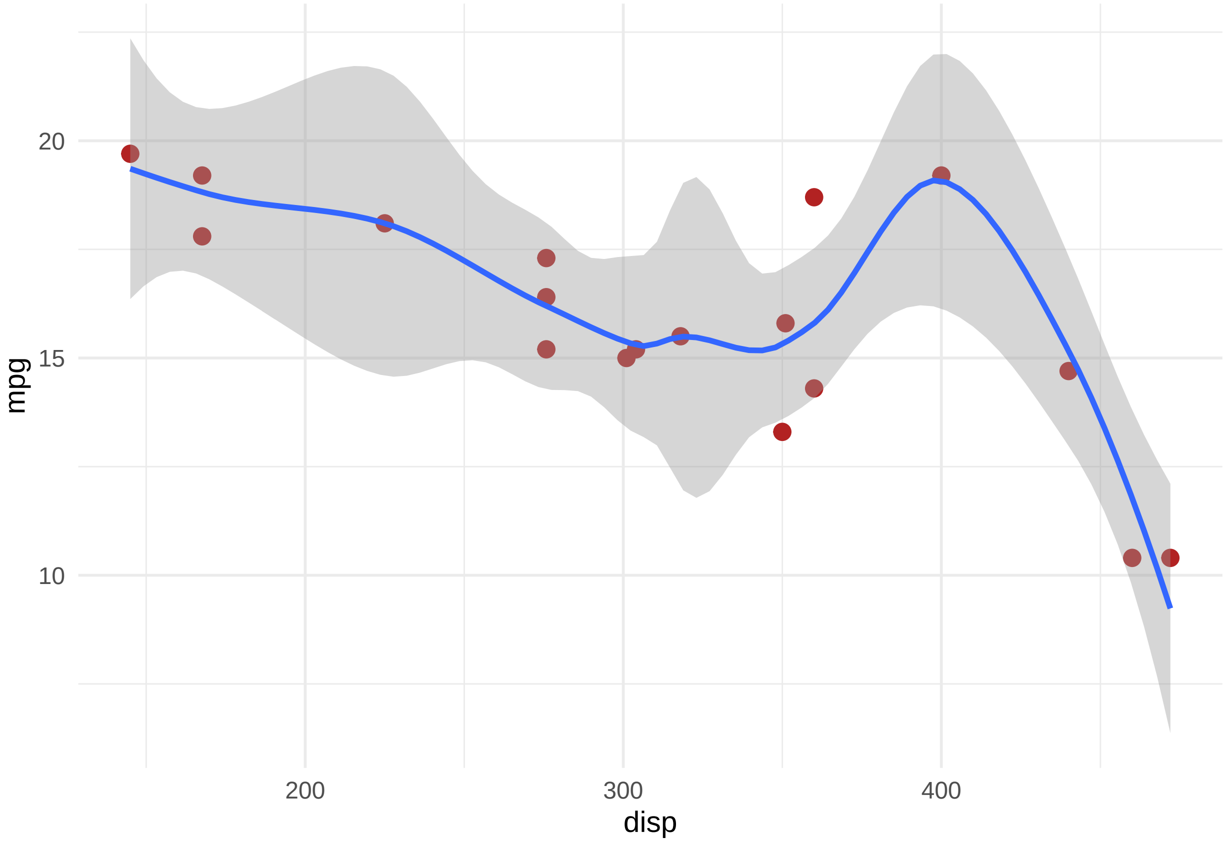
Load esquisse, allowing us to build a ggplot for mtcars with drag n drop

```r
# library(esquisse)
#
# esquisser(data = mtcars_mpg2, viewer = "browser")
```

Biulds a smooth ggplot showing the relationship between displacement and mpg.

```r
ggplot(mtcars_mpg2) +
  aes(x = disp, y = mpg) +
  geom_point(shape = "bullet", size = 4L, colour = "#B22222") +
  geom_smooth(span = 0.5) +
  theme_minimal()
```

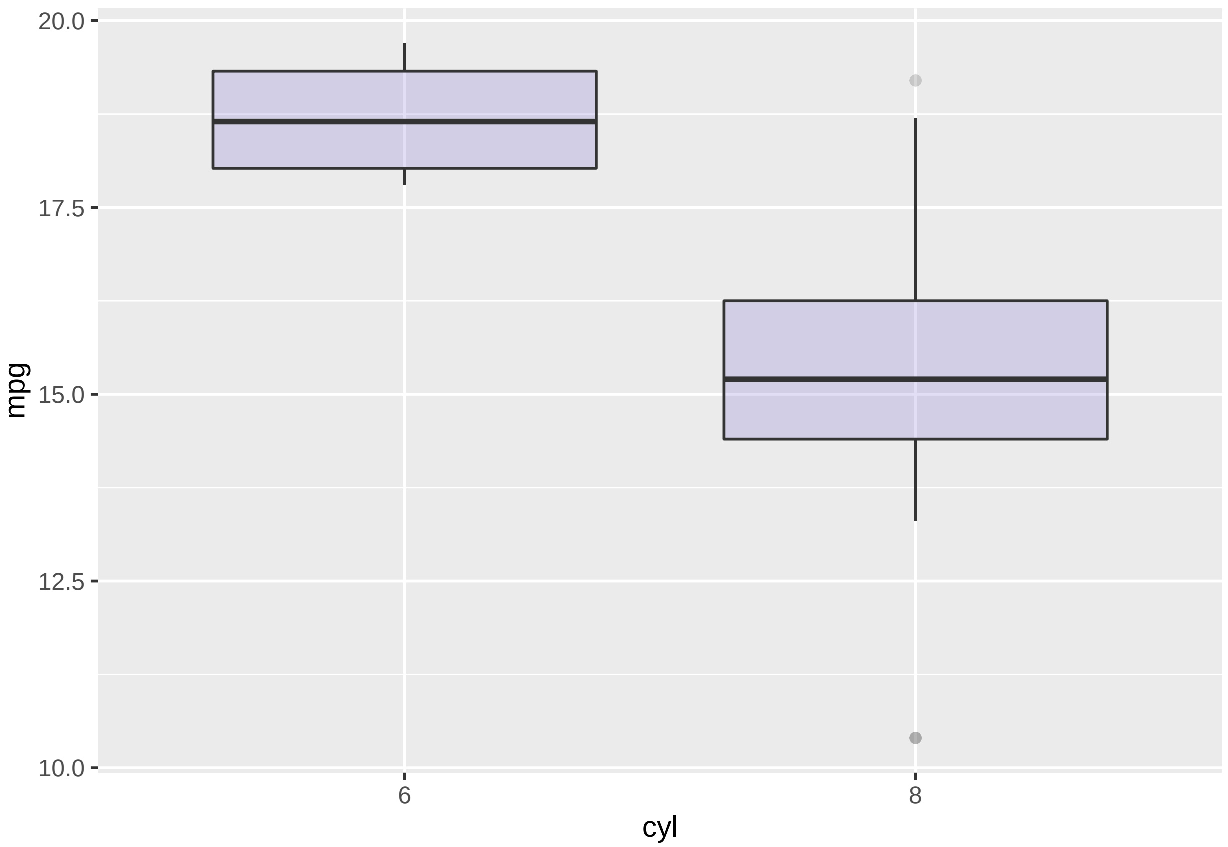
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Boxplot showing mpg, factored out by the cylinder type

```
# note that this boxplot cannot be made with esquisse() unless
# the data is adjusted.  What adjustment is needed?


ggplot(mtcars_mpg2, aes(x=as.factor(cyl), y=mpg)) +
  geom_boxplot(fill="slateblue", alpha=0.2) +
  xlab("cyl")
```


20.0
17.5
15.0
12.5
10.0
mpg
6
8
cyl