Homework 1

Ethan Ashby

1/29/2020

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.2.0     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.1
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

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

1. Consider observing n = 10 pairs of points where no clear predictor/response relationship is obvious. Supposing that the relationship between the two variables appears to be linear, find the range of r (correlation coefficient) for which the null hypothesis H0 : corr = 0 will not be rejected at alpha = .05. (Refer to the t-test given at the end of lecture notes 2). (If you want to get quantiles from a t-distribution in R, you can use qt(quantile,df) where quantile is a number between 0 and 1 and df is the degrees of freedom.)

```
#find quantiles for df=8 where we reject the null hypothesis @ alpha-0.05. Remember two-tailed
qt(0.025, 8)
```

```
## [1] -2.306004
```

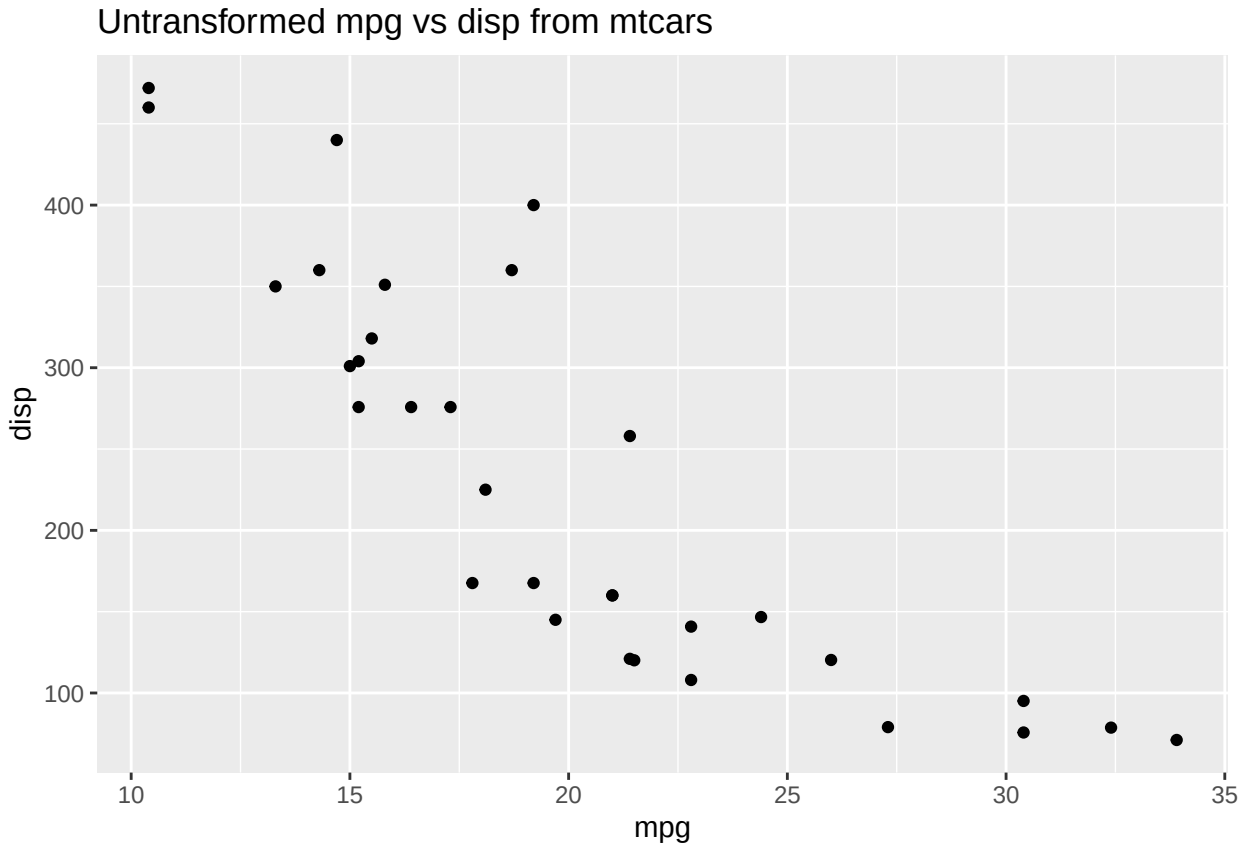
```
qt(0.975,8)
```

```
## [1] 2.306004
```

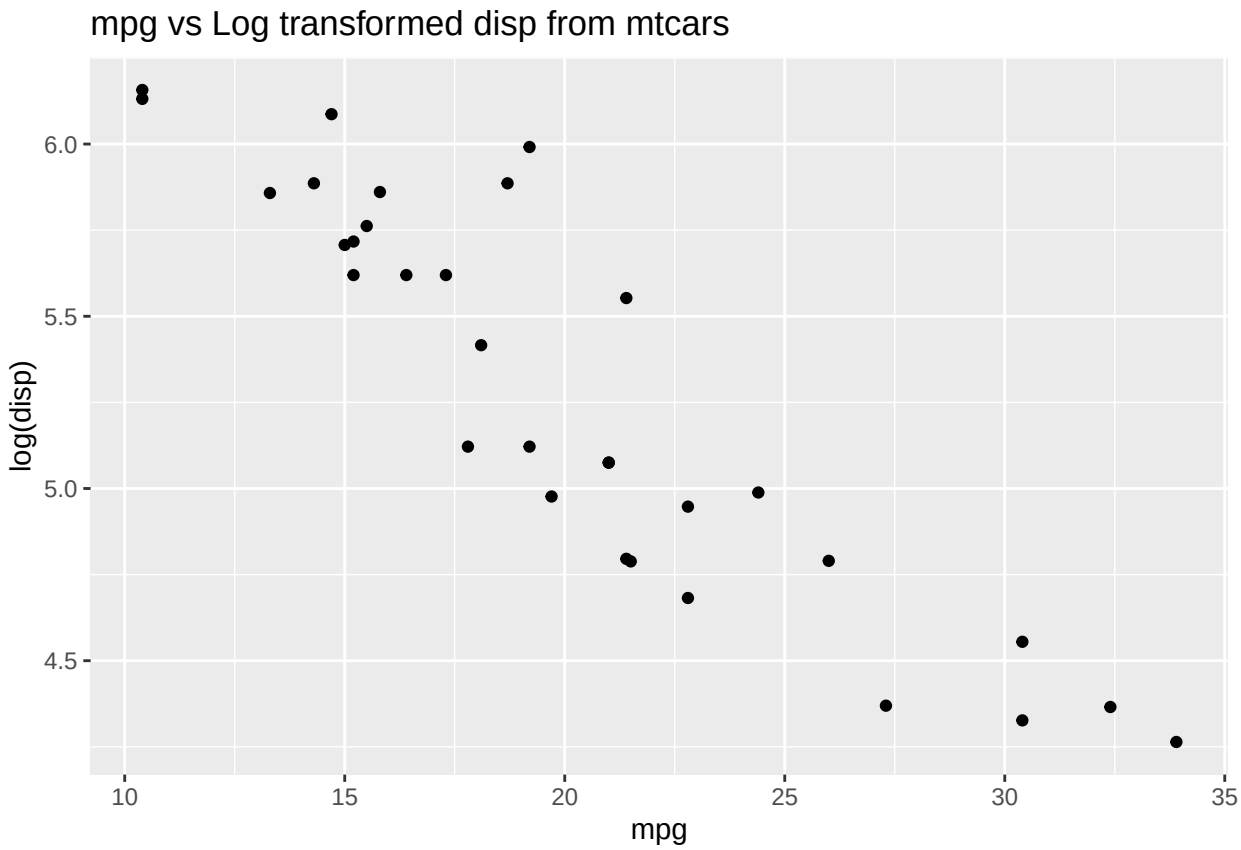
6. Using the data set mtcars (data(mtcars)) in R, find a transformation of the variables mpg (fuel efficiency) and disp (engine size) that allows for reasonable application of our linear model.

```
data(mtcars)

#diagnostic plot
mtcars %>% select(mpg, disp) %>% ggplot()+geom_point(aes(x=mpg, y=disp))+ggtitle("Untransformed mpg vs
```




```
#try a log transform on disp variable
mtcars %>% select(mpg, disp) %>% ggplot()+geom_point(aes(x=mpg, y=log(disp)))+ggtitle("mpg vs Log transf
```




```
#The data looks quite a bit more linear now... this is more reasonable for applying our linear model
```