CM515Assignment5

Klaudia Poplawski

Instructions

The aim of this assignment is to provide you with an opportunity to sharpen your skills in using ggplot(). While some of the plots you create may resemble those we covered in class, it is essential that they are original.

Rules for the Assignment:

1.All data sets except iris are fair game. That includes past data sets used in the class, sets built into R, your own data, and even online data.

2. All graphs must include axis labels, plot title, a theme of your choice, and a brief description/interpretation of the plot (2-3 sentences).

It's worth noting that ggplot() is a widely-used tool, and there are numerous online resources available for you to explore. We strongly recommend that you take advantage of these resources to deepen your understanding of ggplot().

Knit the document into a PDF and submit it to Canvas by 02/28/2024 at 11:59 pm.

Load Packages and Data

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

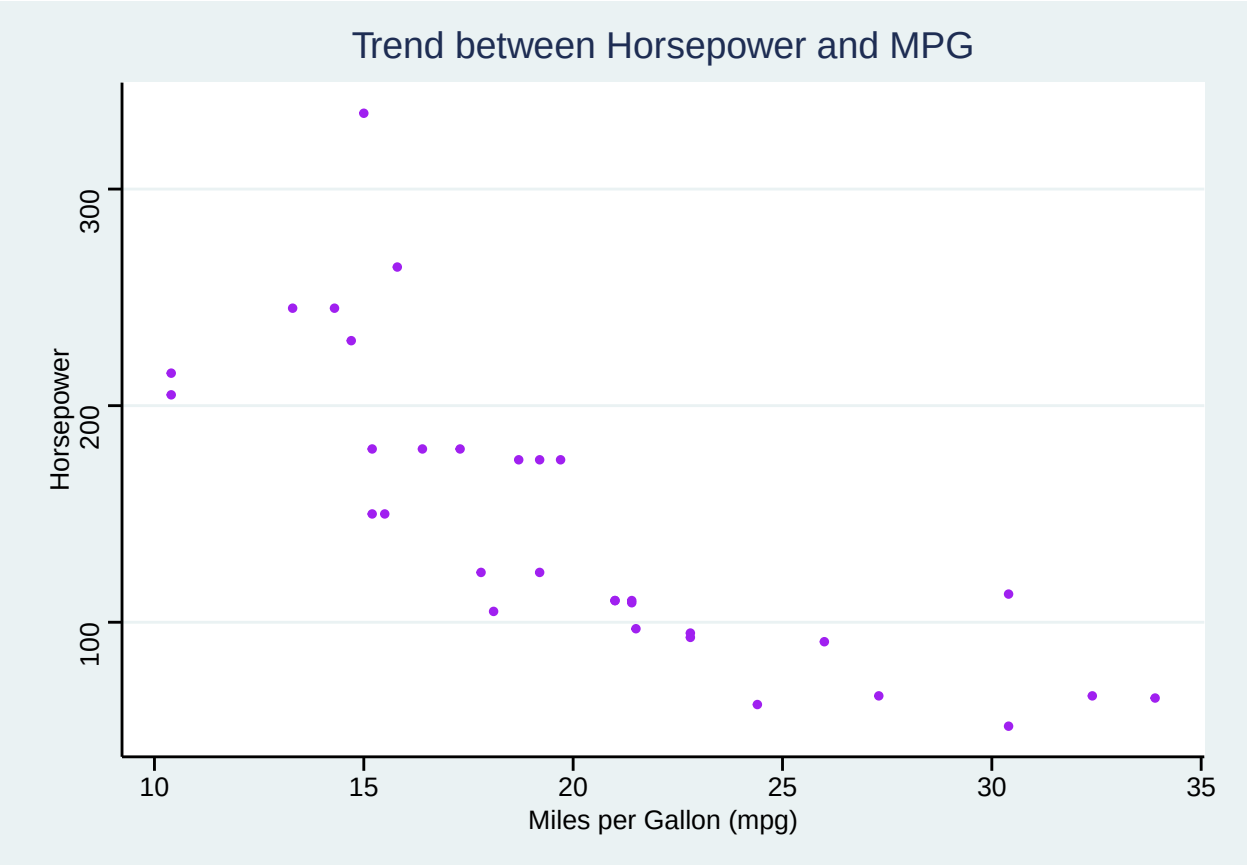
```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
```

```
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

---

1. Make a Scatter Plot With Customized Point Size and Transparency (3pts)

##Description and Interpretation: For this plot, I chose two variables that I thought would have a relationship with one another based on my limited knowledge of cars. This scatterplot just plotted these variables against each other, nothing special or fancy. Just basic plotting with a few customized geometric features.

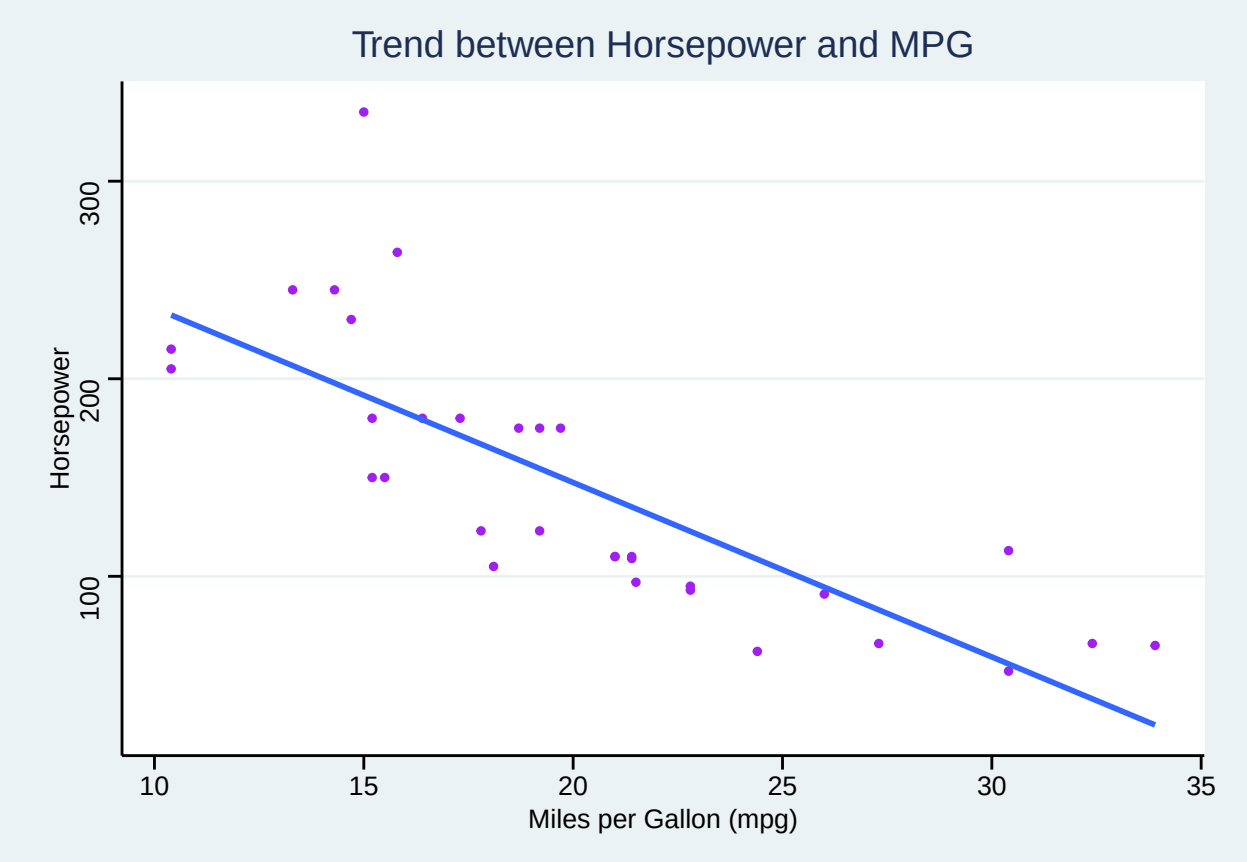


---

2. Fit a Line Through Your Scatter Plot From 1. (3pts)

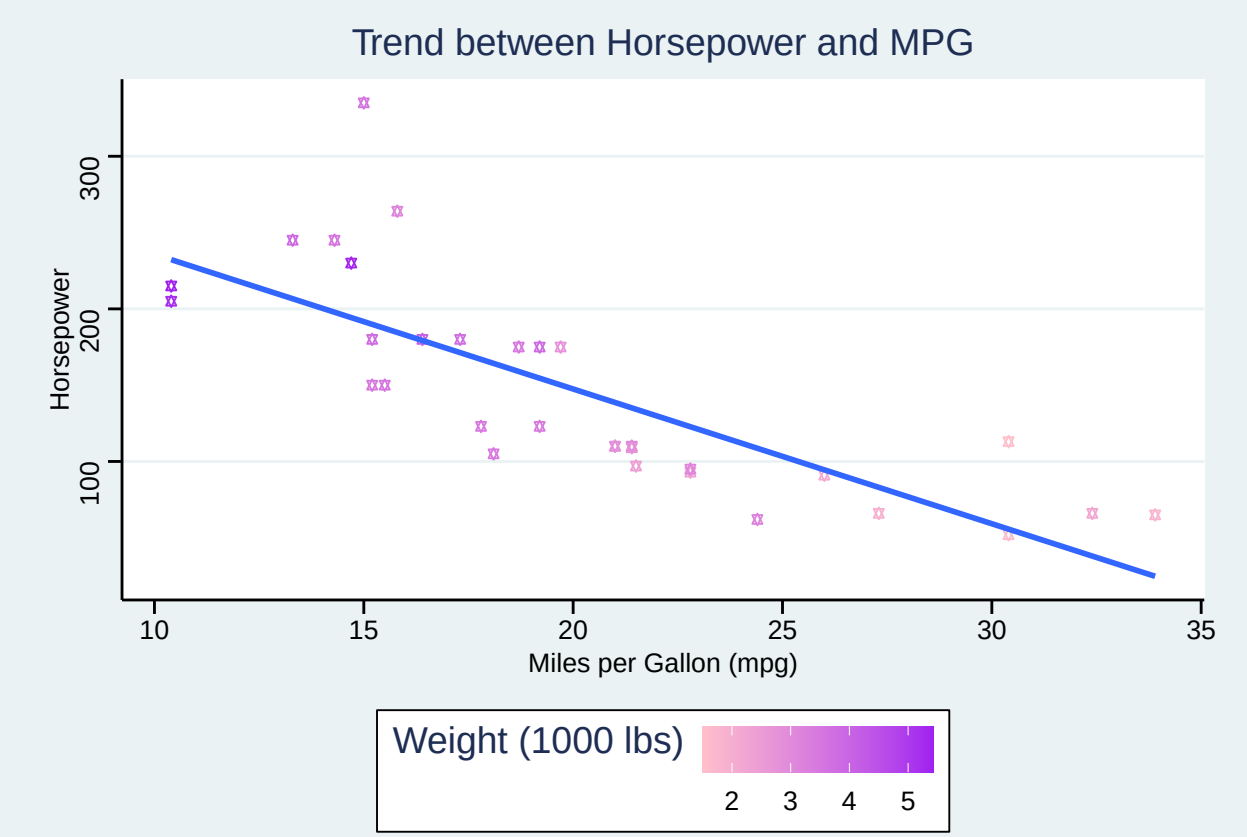
##Description and Interpretation: This plot is the same as above with an added trendline that allows us to see a direct trend in the relationship between the variables. I was getting a warning but with Dr. King we figured out that the warning is not anything crazy and vanishes when the trendline is gone. Just has to do

with the trendline not fitting all the points which is very reasonable with how the distribution looks in the scatter plot.



##3. Set the transparency, shape, and color of your graph from 1 based on expressions. Scale the color. (3pts)

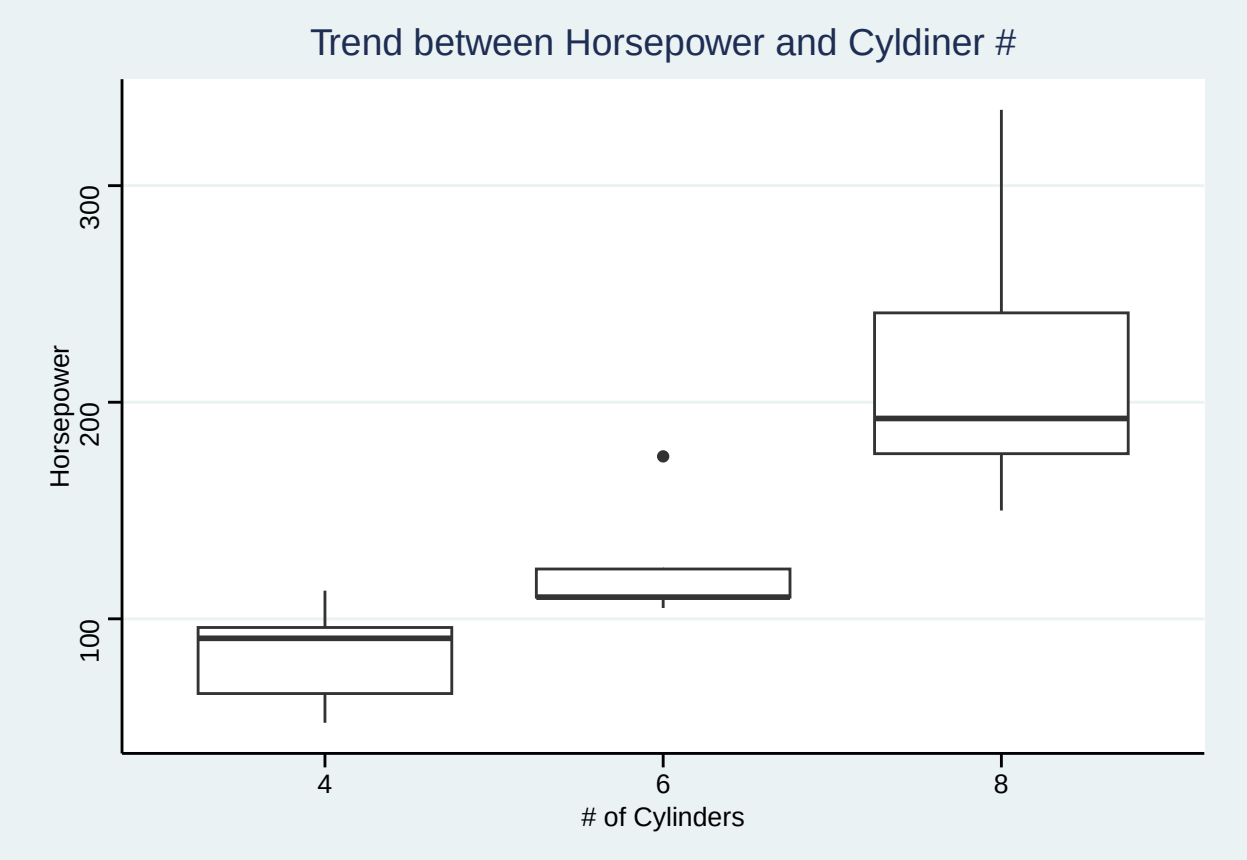
##Description and Interpretation: I liked playing around with the shape for this one, but couldn't figure out how to fill the shape. Regardless, I actually chose to do a color gradient based on a third variable (being weight) to introduce another variable for comparison in the relationship. For the most part, apart from a few, there seemed to be a relationship between increase in horsepower, weight with a decrease in mpg.I decided to set the transparency to 1 because there weren't really any overlapping points and this made the colors more visible. Again, I got another warning, this had to do with the trend line interacting with the color gradient but did not affect my plot or the way the data was shown.



---

4. Make a Boxplot With Customized Boxplot Width (3pts)

##Description and Interpretation: At this step, I actually went and spent about an hour trying to figure out which built in dataset was best because I don't have my own excel sheets to work with currently and I realized the data I originally chose was not really compatible with a box plot. For this, I chose mtcars for the homework because the cylinders variable (cyl) had different categories and was easy to separate. This plot is really useful for the categorization of data in groups because especially for this type of dataset, 4 cylinder cars run very differently compared to 8 and you can compare datapoints within a given category.

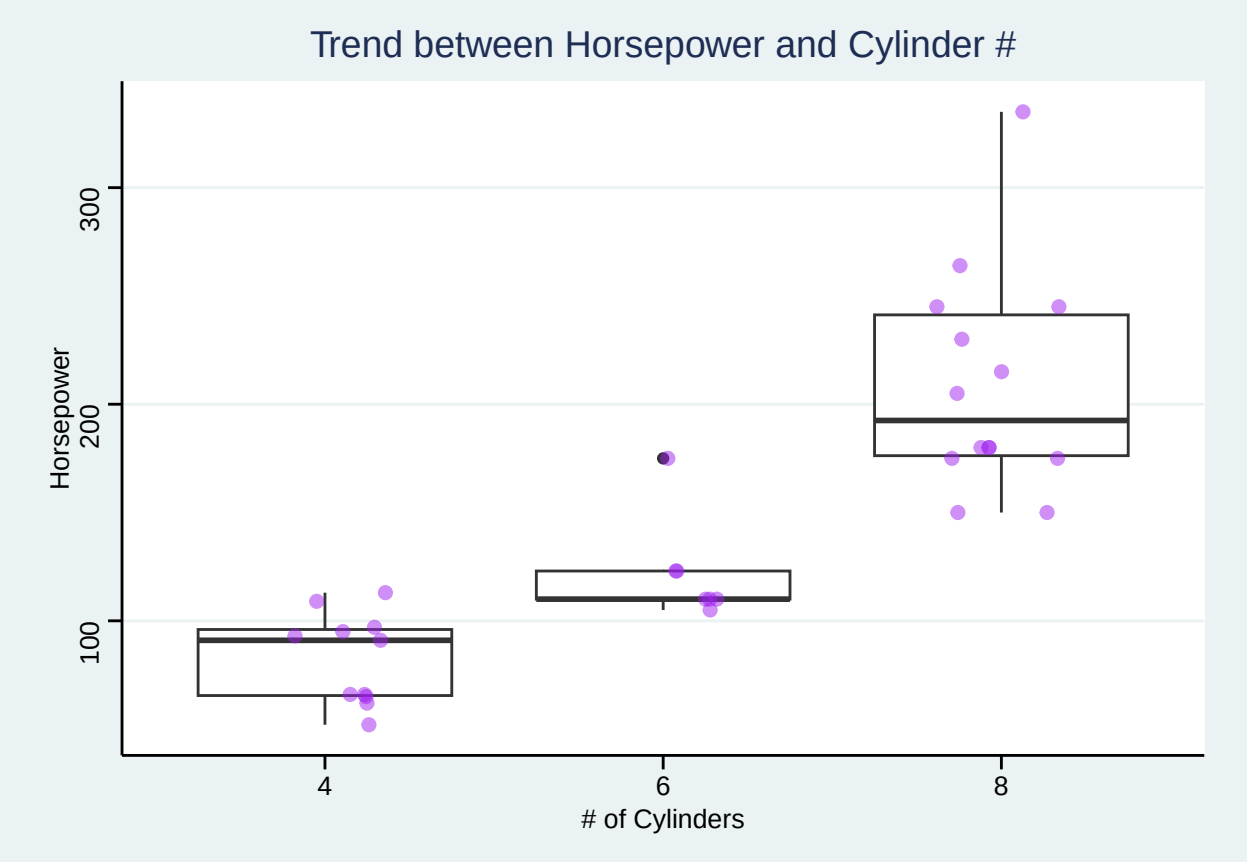


#*******

5. Overlay the Individual Points Over Your Box Plot From 3. and Adjust the Point Size and Transparency as Needed (3pts)

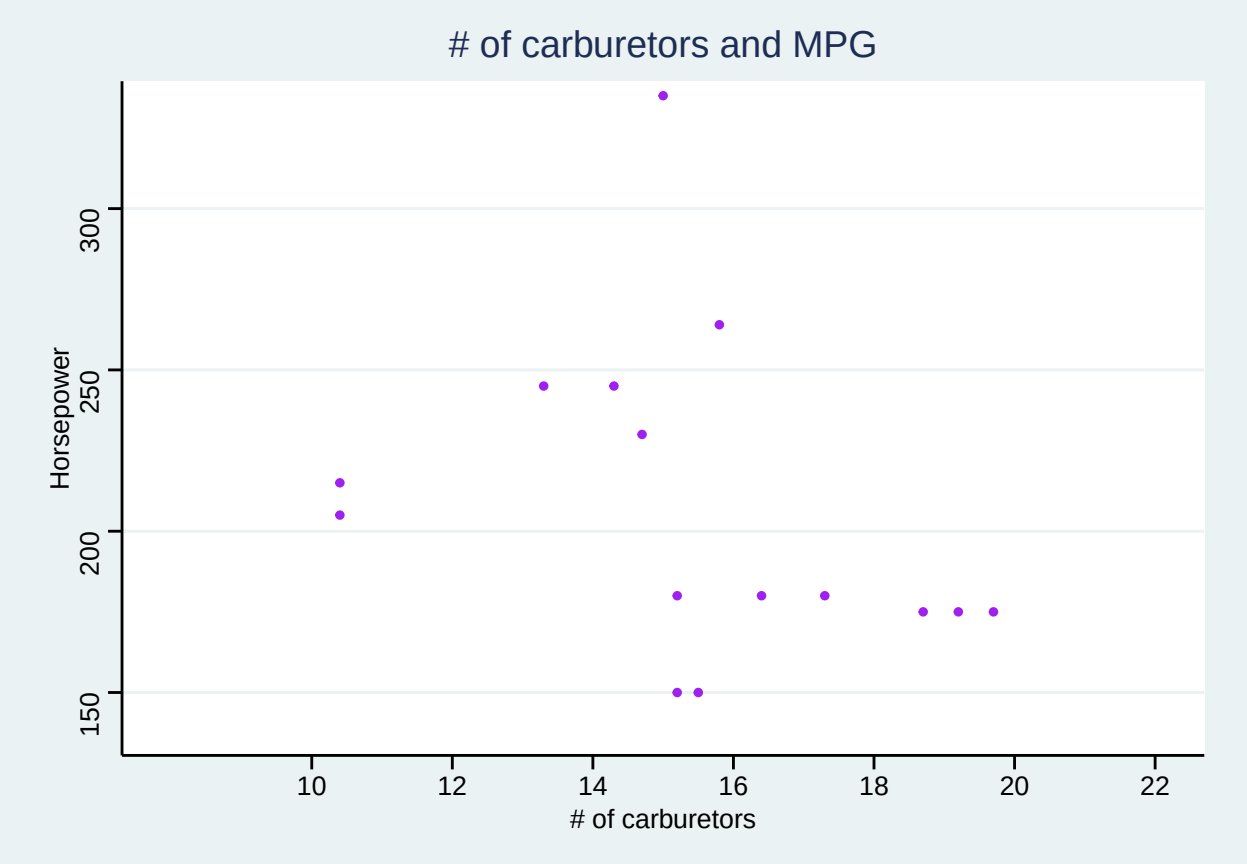
#******

##Description and Interpretation: This type of plot is not my favorite because it is very messy in my opinion but I understand that there are situations where using this combination of plots is best, however I usually find box plots are good and simple ways of representing this. In this case, I did have some data overlapping because of my chosen variables put against each other so I set the transparency lower so you can see the overlap and when the points were bigger they took too much room in my opinion.



6. Make a plot and modify the scale in a beneficial way using scale_x_continuous(), scale_y_continuous(), and coordinate_cartesian(). Explain how your modifications to the scales improved the plot. (4pts)

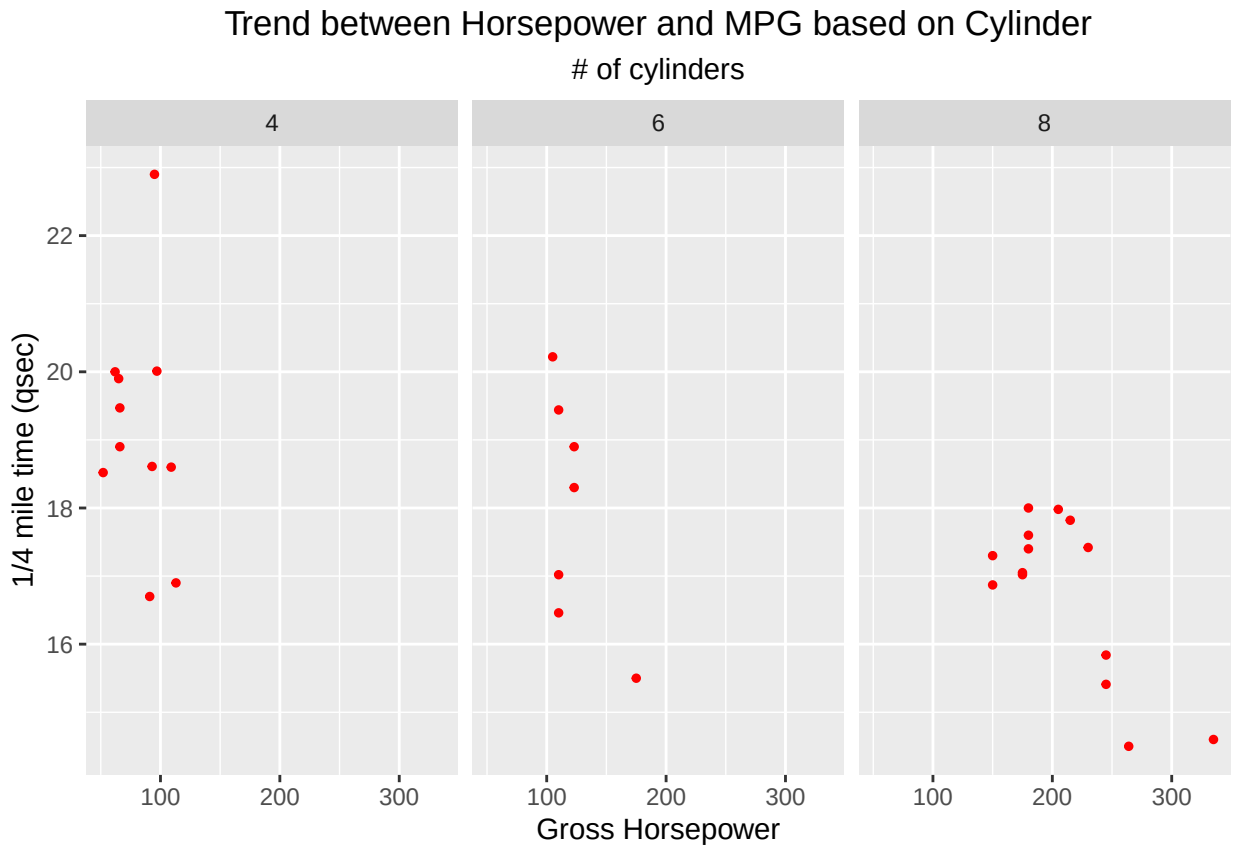
##Description and Interpretation: I tried to narrow down my data by cars that had a higher mpg as if I were looking to buy a car and of course in today's economy, I want a high mpg car. I think the limits I set really helped the scope of the data be more specific and especially on the x-axis, I had so much room so I was able to change the axis limits so that there wasn't so much blank space and it was easier to read where stuff falls along the x-axis.



---

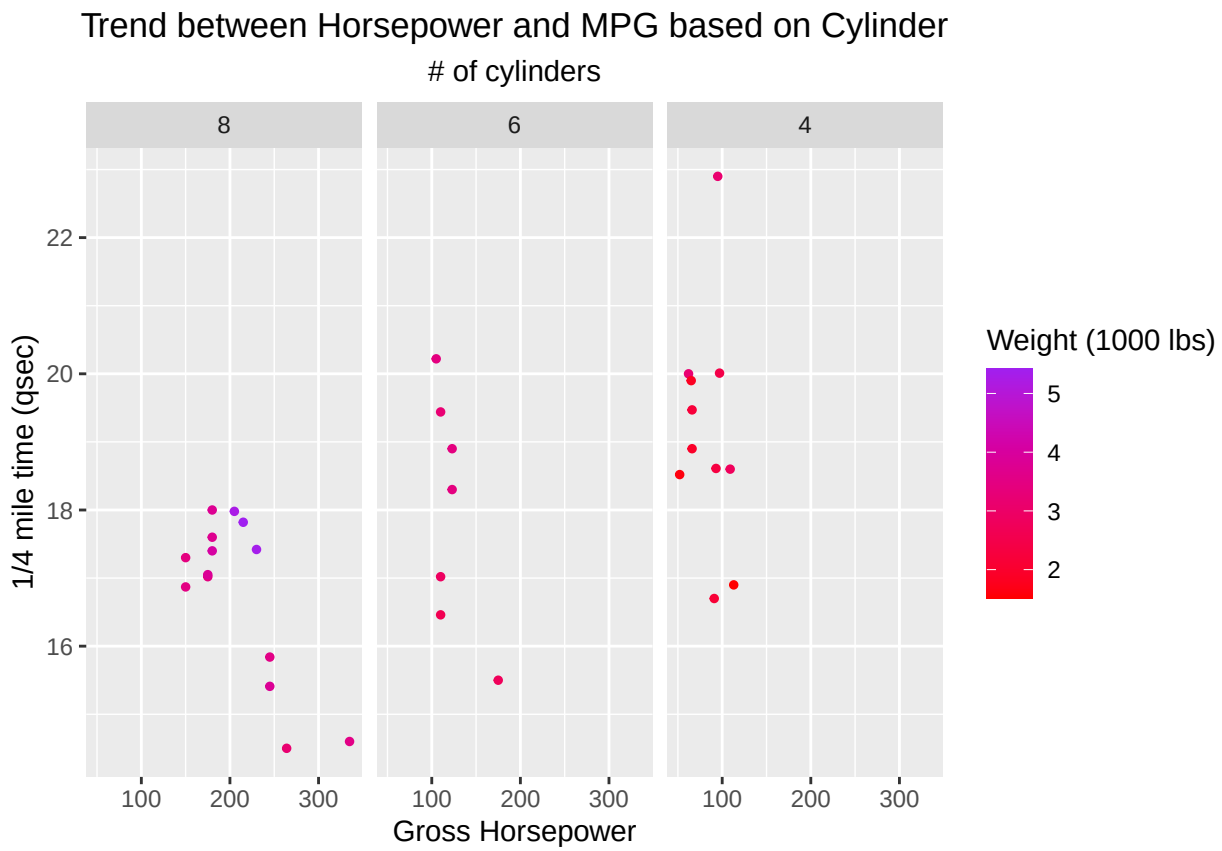
7. Make a New (don't use any plots from 1-4) Plot with facet_wrap() (4pts)

##Description and Interpretation: It took me a second to figure this out because I was not in class on Wednesday but this looks like the equivalent of a box plot but because it is separate graphs for each cylinder type, we were able to plot 2 new variables against each other for analysis based on cylinder number. I like this better than the boxplot overlayed with scatter plot because it has less "stuff" on it.



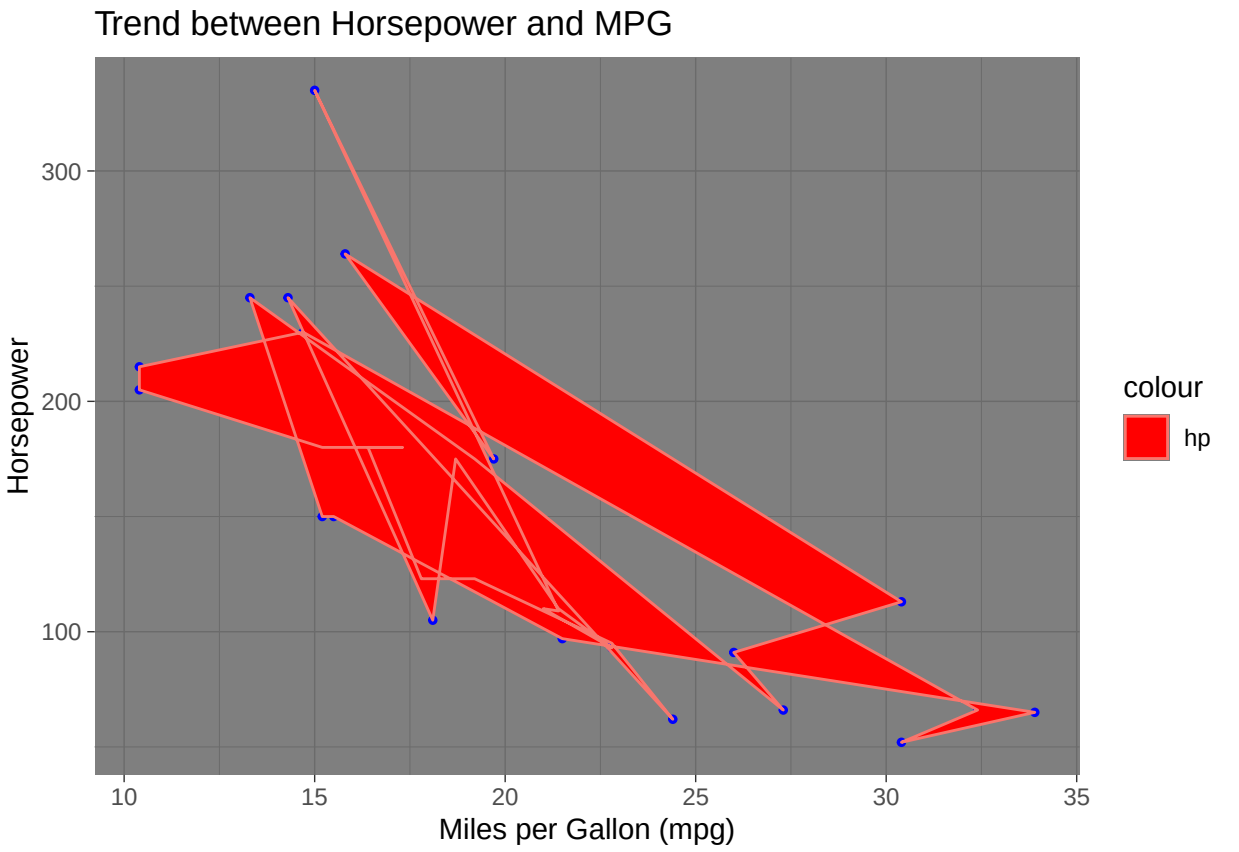
8. Using the plot from 5, scale the colors and reorder your facet to reverse order. (4pts)

##Description and Interpretation: This one, I used what I learned in Meg's tidyverse module in order to flip the order of the facets. For the scale, I did what I previously did in an earlier question where I used a third variable for the scale color gradient to introduce an extra variable for comparison. I will note that for this graph and the one in question 7, I was not able to implement a theme without the graph failing to publish so I did not incorporate themes for these two.



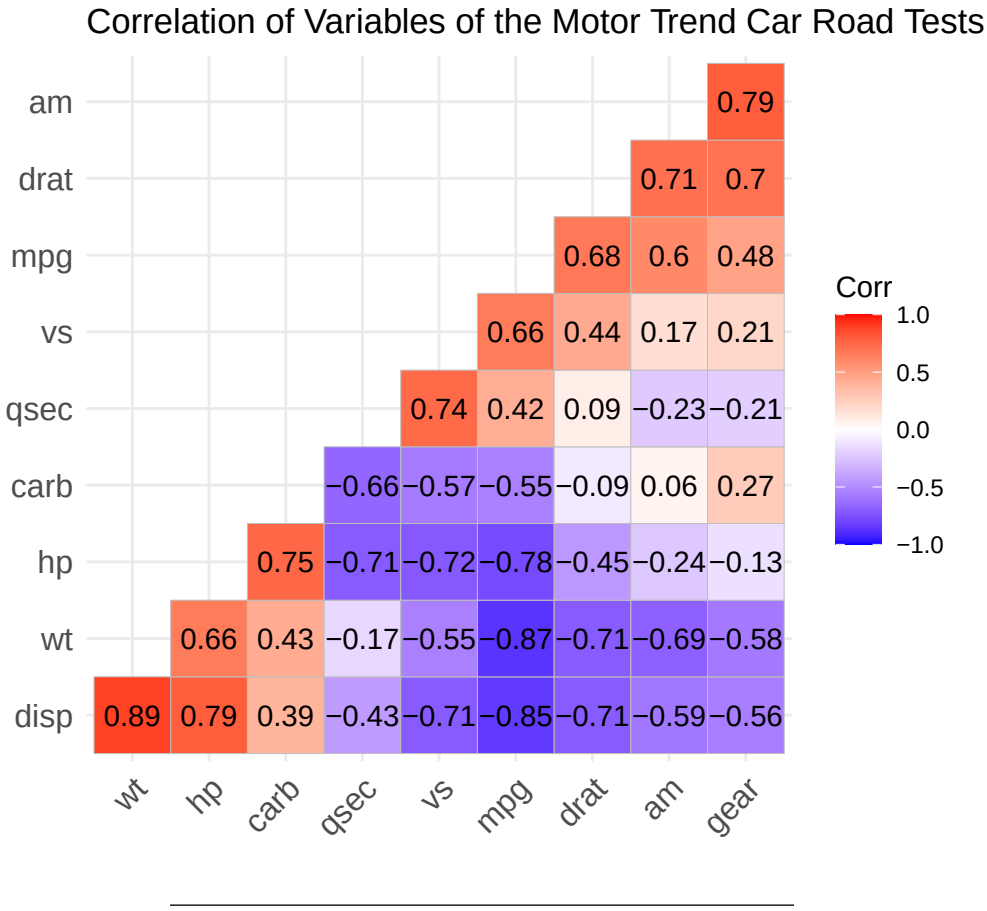
9. Make a Plot Using a Geometry That Was Not Covered in Class (4pts)

##Description and Interpretation: This one was so fun to make because it looks so awful and not representative of the data at all. I just figured its a good time to learn to make a graph, even if it might never apply to my research and I had to switch the colors around cause not every color was showing up and I had a hard time trying to make the lines a different color but I at least got the points to be blue. It was nice to experiment and I look forward to trying other ones in the r graph gallery :).



10. Make a Plot Using a ggplot Extension That Was Not Covered in Class. Explain what the extension does. (4pts)

##Description and Interpretation: So for this one, I originally wanted to experiment with gganimate but it wasn't very compatible with mtcars so I chose ggcorrplot because this one I knew would be due to so many variables having possible relations to one another. I actually was surprised with the results because I thought more variables would correlate with one another so that just goes to show how little I know about cars.This is definitely something I see myself using in a future lab when testing several variables.



Appendix

```
knitr::opts_chunk$set(echo = F, message = F, warning = F)
####
#00#
####
library(datasets)
library(tidyverse)
library(knitr)
library(ggplot2)
library(ggthemes)
library(dplyr)
library(ggcorrplot)


head(mtcars)
summary(mtcars)

####
#01#
####

mtcars %>%
```

```r
  ggplot(aes(x = mpg, y = hp, color=mpg)) +
  labs(x = "Miles per Gallon (mpg)", y = "Horsepower",
  title = "Trend between Horsepower and MPG") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_stata() +
  geom_point(alpha=1, color="purple", size=1)

####
#02#
####
#your code here
mtcars %>%
  ggplot(aes(x = mpg, y = hp, color=mpg)) +
  labs(x = "Miles per Gallon (mpg)", y = "Horsepower",
  title = "Trend between Horsepower and MPG") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_stata() +
  geom_point(alpha=1, color="purple", size=1) +
  geom_smooth(method = "lm", se = FALSE)

####
#03#
####
#your code here
#can I use the gradient for a third variable? aka volume in this case?
#the warning message coming up: I tried asking chatGPT what it meant but when I tried the grouping thin

mtcars %>%
  ggplot(aes(x = mpg, y = hp, color=wt)) +
  labs(x = "Miles per Gallon (mpg)", y = "Horsepower",
  title = "Trend between Horsepower and MPG") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_stata() +
  geom_point(alpha=1, size=1, shape="star") +
  scale_color_gradient(low = "pink", high = "purple", name = "Weight (1000 lbs)") +
  geom_smooth(method = "lm", se = FALSE)

####
#04#
####
#your code here
mtcars$cyl <- factor(mtcars$cyl)
#Why did I have to turn it into a factor but Alex didn't have to for the iris example

mtcars %>%
  ggplot(aes(x = cyl, y = hp)) +
  labs(x = "# of Cylinders", y = "Horsepower",
  title = "Trend between Horsepower and Cyldiner #") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_stata() +
  geom_boxplot()
```

```r
####
#05#
####
#your code here

mtcars %>%
  ggplot(aes(x = cyl, y = hp)) +
  labs(x = "# of Cylinders", y = "Horsepower",
  title = "Trend between Horsepower and Cylinder #") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_stata() +
  geom_boxplot() +
  geom_jitter(size = 2, alpha = .5, color = "purple", width = .2, height = 0)

####
#06#
####
#your code here

mtcars %>%
  ggplot(aes(x = mpg, y = hp, color=mpg)) +
  labs(x = "# of carburetors", y = "Horsepower",
  title = "# of carburetors and MPG") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_stata() +
  geom_point(alpha=1, color="purple", size=1) +
  scale_y_continuous(limits = c(150,350)) +
  scale_x_continuous(breaks = seq(10, 34, by = 2)) +
  coord_cartesian(ylim=c(140,330), xlim=c(8,22))


#07#
####
#your code here
mtcars %>%
  ggplot(aes(x= hp, y= qsec)) +
  labs(x = "Gross Horsepower", y = "1/4 mile time (qsec)",
  title = "Trend between Horsepower and MPG based on Cylinder", subtitle = "# of cylinders") +
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle = element_text(hjust = 0.5)) +
  facet_wrap(~cyl) +
  geom_point(alpha=1, color="red", size=1)

#08#
####
#your code here

mtcars %>%
  mutate(cyl = factor(cyl, levels = c(8, 6, 4))) %>%
  ggplot(aes(x= hp, y= qsec, color=wt)) +
 labs(x = "Gross Horsepower", y = "1/4 mile time (qsec)",
  title = "Trend between Horsepower and MPG based on Cylinder", subtitle = "# of cylinders") +
  theme(plot.title = element_text(hjust = 0.5), plot.subtitle = element_text(hjust = 0.5)) +
  facet_wrap(~cyl) +
```

```r
  geom_point(alpha=1, size=1) +
  scale_color_gradient(low="red", high="purple", name = "Weight (1000 lbs)")

####
#09#
####
#your code here

mtcars %>%
  ggplot(aes(x = mpg, y = hp, color="hp")) +
  labs(x = "Miles per Gallon (mpg)", y = "Horsepower",
  title = "Trend between Horsepower and MPG") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_dark() +
  geom_point(alpha=1, color="blue", size=1) +
  geom_polygon(fill = "red")
####
#10#
####
#your code here
library(dplyr)
library(ggcorrplot)

mtcars %>%
  select_if(is.numeric) %>%
  cor() %>%
  ggcorrplot(hc.order = TRUE, type = "lower", lab = TRUE, title = "Correlation of Variables of the Moto
```