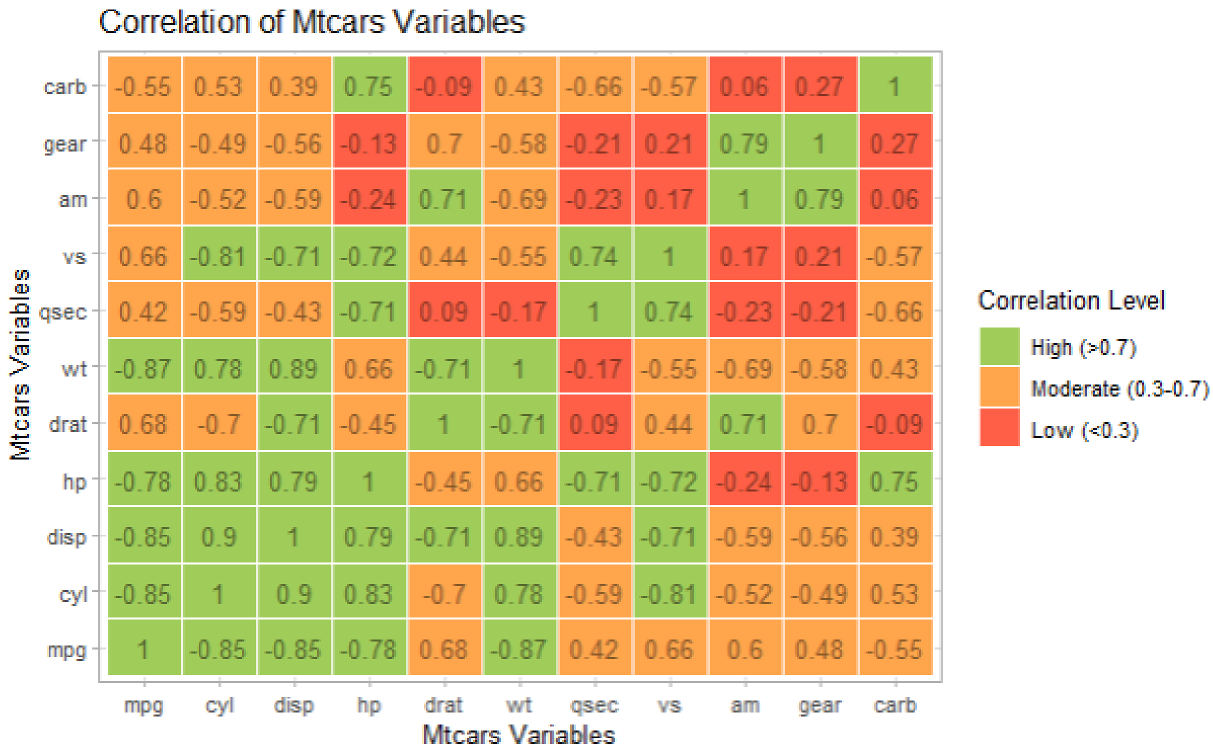
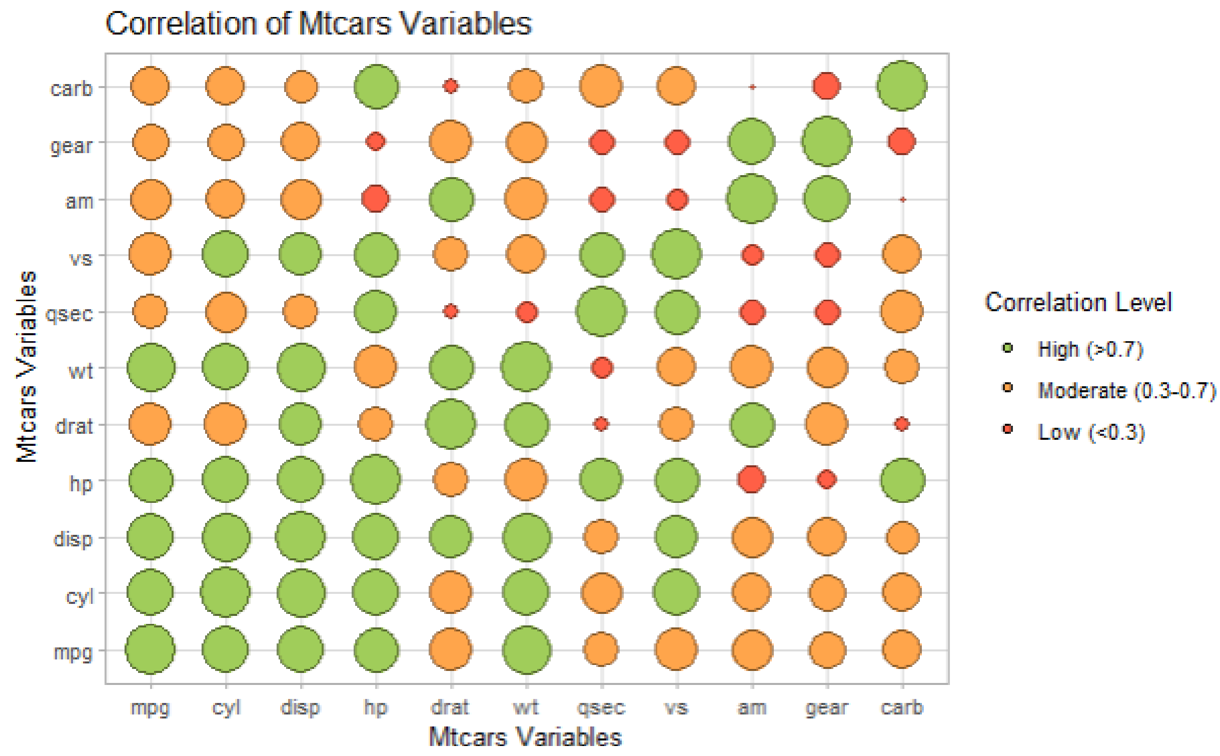
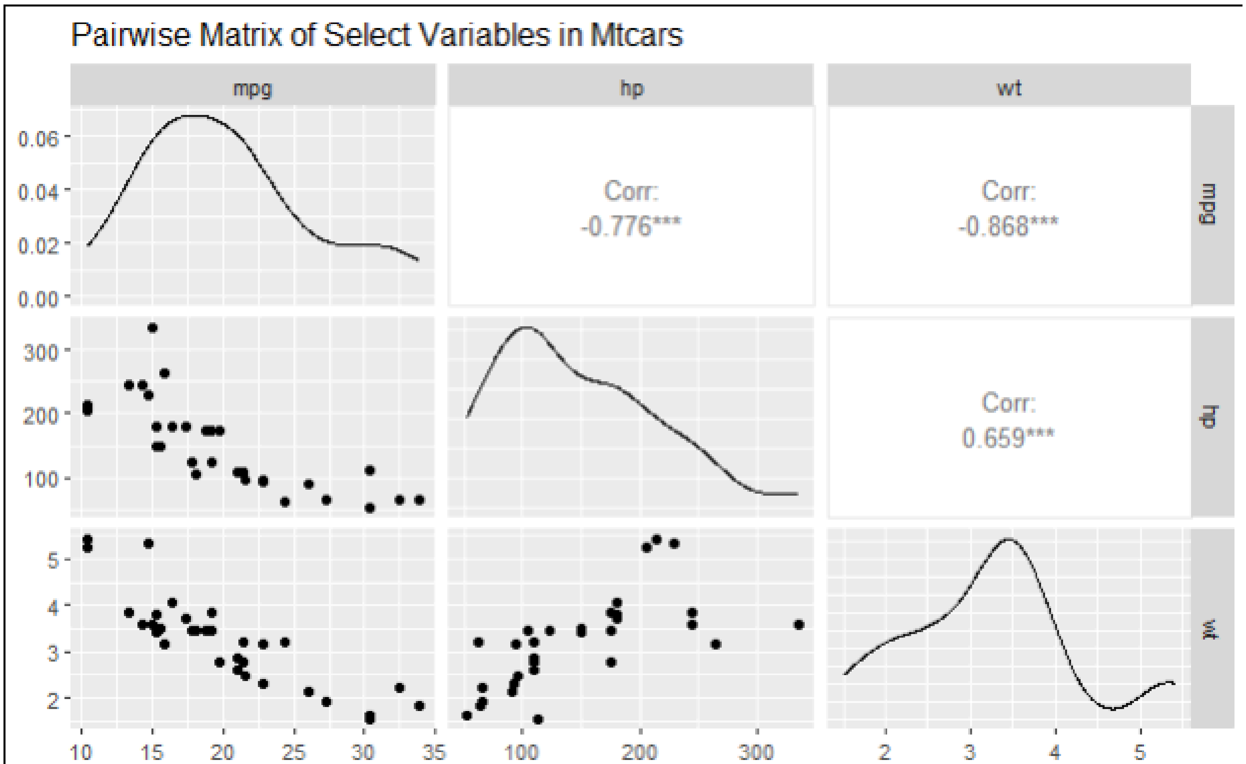
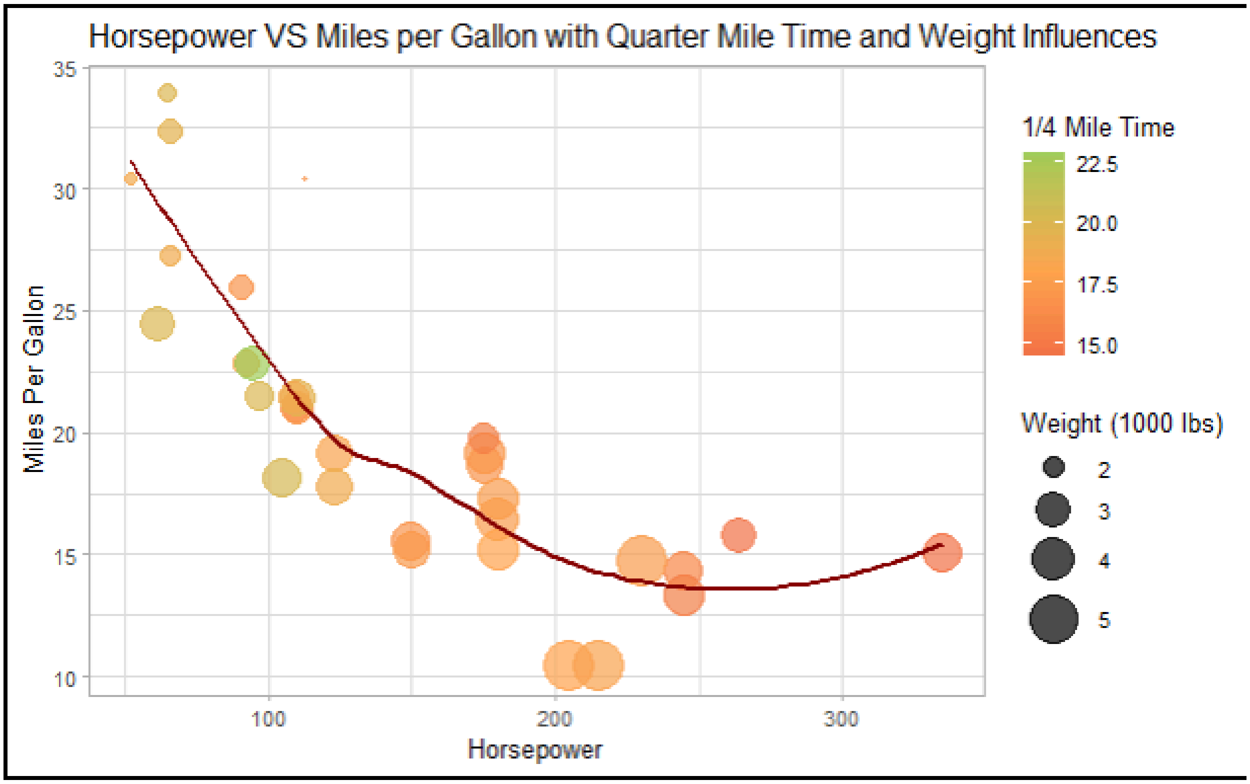
Module #8 Assignment

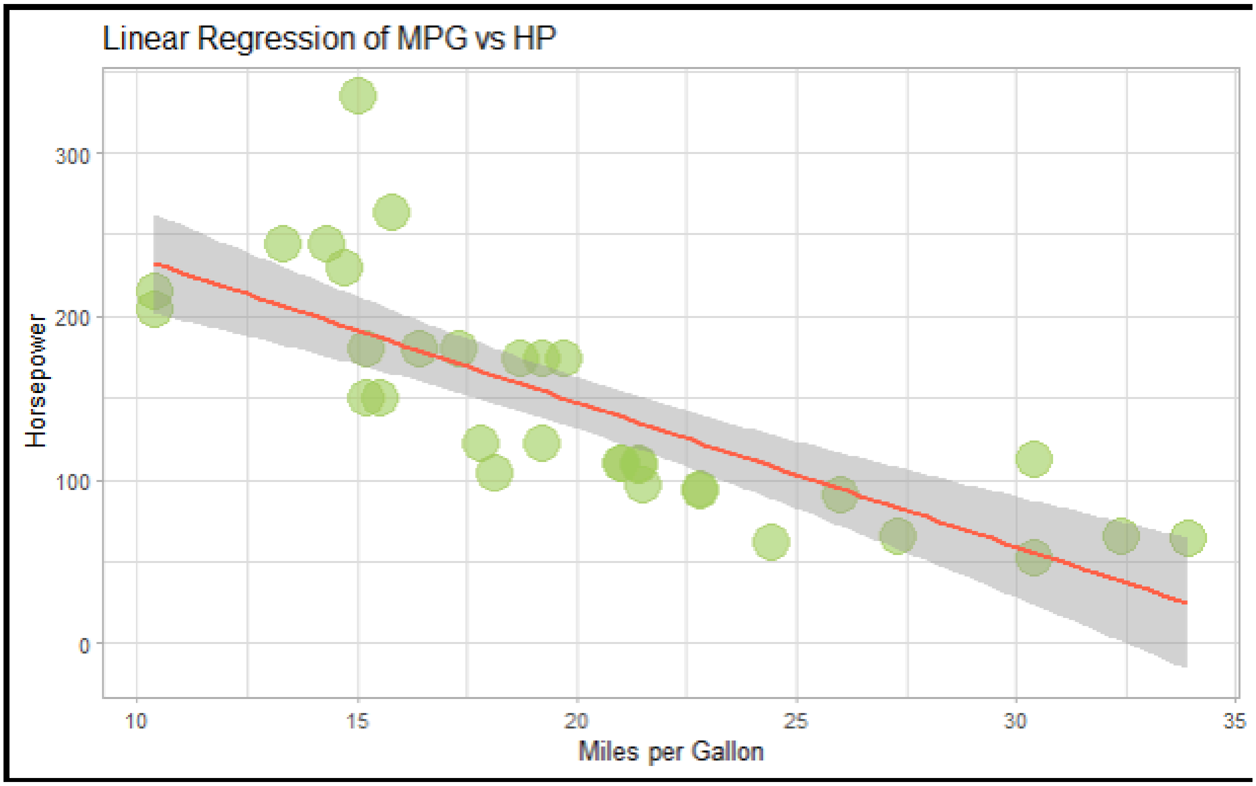
For this assignment, I made a series of correlation and regression visualizations using ggplot2 and mtcars.

```
```{r, "Mtcars Correlation Visualizations"}
library(ggplot2)
library(reshape2)
library(RColorBrewer)
library(GGally)
mtcars <- mtcars

mtcars_cor <- cor(mtcars)
long.mtcars_cor <- melt(mtcars_cor)

correlation <- function(x) {
  if (abs(x) > 0.7) {
    return("High Correlation")
  } else if (abs(x) >= 0.3 & abs(x) <= 0.7) {
    return("Moderate Correlation")
  } else {
    return("Low Correlation")
  }
}

long.mtcars_cor$cor_level <- sapply(long.mtcars_cor$value, correlation)
long.mtcars_cor$cor_level <- factor(long.mtcars_cor$cor_level,
                                    levels = c("High Correlation", "Moderate Correlation", "Low
Correlation"), labels = c("High (>0.7)", "Moderate (0.3-0.7)", "Low (<0.3)"))
```

Before creating the visualizations, I calculated the correlation coefficients between the variables in the mtcars dataset and converted the resulting correlation matrix into long format using the melt() function from the reshape2 package. This transformation was necessary to prepare the data for correlation visualization in ggplot2.

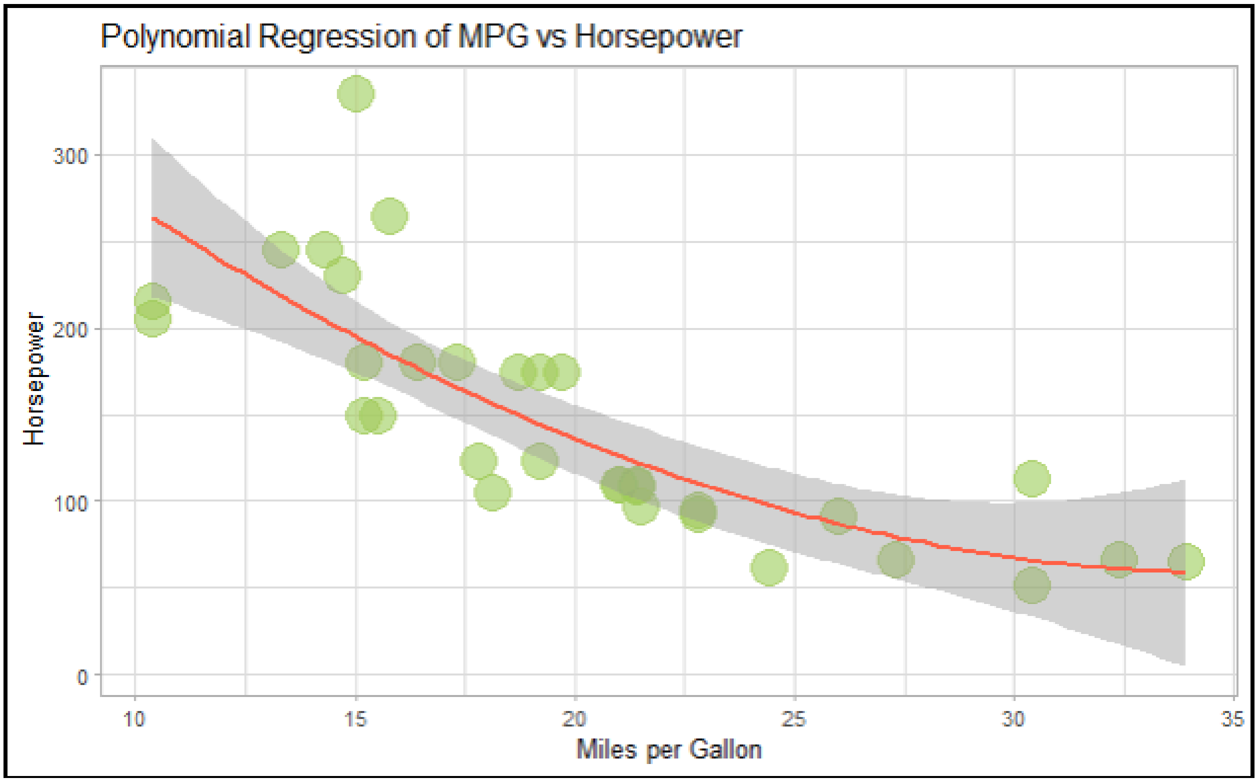
Correlation Visualizations:

```
ggplot(long.mtcars_cor, aes(x = Var1, y = Var2, fill = cor_level)) +
  geom_tile(color = "white") +
  geom_text(aes(label = round(value, 2), color = cor_level)) +
  scale_fill_manual(values = c("darkolivegreen3", "tan1", "tomato"),
                    name = "Correlation Level") +
  scale_color_manual(values = c("darkolivegreen", "tan4", "tomato4"), guide = "none")+
  theme_light() +
  theme(plot.caption = element_text(hjust = 0, size = 8)) +
  labs(x = "Mtcars Variables", y = "Mtcars Variables", title = "Correlation of Mtcars
Variables")
```
```



This heatmap shows levels of correlation between mtcars variables. Variables with strong correlations (above the absolute value of 0.7) are colored in green, moderate levels of correlation are colored orange, and low levels of correlations (below the absolute value of 0.3) are colored red.

```
ggplot(long.mtcars_cor, aes(x = Var1, y = Var2, fill = cor_level)) +
  geom_point(aes(size = abs(value), color = cor_level), shape = 21) +
  scale_fill_manual(values = c("darkolivegreen3", "tan1", "tomato"),
                    name = "Correlation Level") +
  scale_color_manual(values = c("darkolivegreen", "tan4", "tomato4"), guide = "none")+
  scale_size(range = c(1,10), guide = "none") +
  theme_light() +
  theme(plot.caption = element_text(hjust = 0, size = 8)) +
  labs(x = "Mtcars Variables", y = "Mtcars Variables", title = "Correlation of Mtcars
Variables")
```



This map is very similar to the previous heatmap, it is just a different way of presenting the same information. The primary difference between this heatmap and the previous one is that this heatmap uses size to denote the value of the correlation coefficient whereas the previous chart simply wrote it out. Both have merits, but also different uses.

```
pairwise_mtcars <- mtcars[, c("mpg", "hp", "wt")]
ggpairs(pairwise_mtcars, title = "Pairwise Matrix of Select Variables in Mtcars")
```



In a pairwise scatterplot matrix, each scatterplot shows the relationship between two variables, with points indicating whether the relationship is positive, negative, or neutral. The diagonal panels show the distribution of individual variables, while the upper panels display the correlation values for each pair.

```
ggplot(mtcars, aes(x = hp, y = mpg, size = wt, color = qsec)) +
  geom_point(alpha = 0.7) +
  geom_smooth(color = "darkred", se = FALSE, show.legend = FALSE) +
  scale_size_continuous(name = "Weight (1000 lbs)", range = c(1,10)) +
  scale_color_gradient2(low = "tomato2", mid = "tan1", high = "darkolivegreen3", midpoint = 18,
name = "1/4 Mile Time") +
  theme_light() +
  labs(title = "Horsepower VS Miles per Gallon with Quarter Mile Time and Weight Influences",
       x = "Horsepower",
       y= "Miles Per Gallon")
```



This chart visualizes the correlation between horsepower (x-axis) and miles per gallon (mpg) (y-axis) in the mtcars dataset. The size represents vehicle weight (heavier cars have larger circles), while the color gradient reflects quarter-mile time (green indicates faster times while red indicates slower times). This multi-variable approach helps illustrate how horsepower, weight, and acceleration interact and influence fuel mpg together.

Regression Visualizations:

```
```{r, "MTcars Regression Visualizations"}
ggplot(mtcars, aes(x=mpg, y =hp)) +
  geom_point(alpha = 0.6, color = "darkolivegreen3", size = 7) +
  geom_smooth(method = "lm", se = TRUE, color = "tomato") +
  theme_light() +
  labs (x = "Miles per Gallon", y = "Horsepower", title = "Linear Regression of MPG vs
Horsepower")
```
```



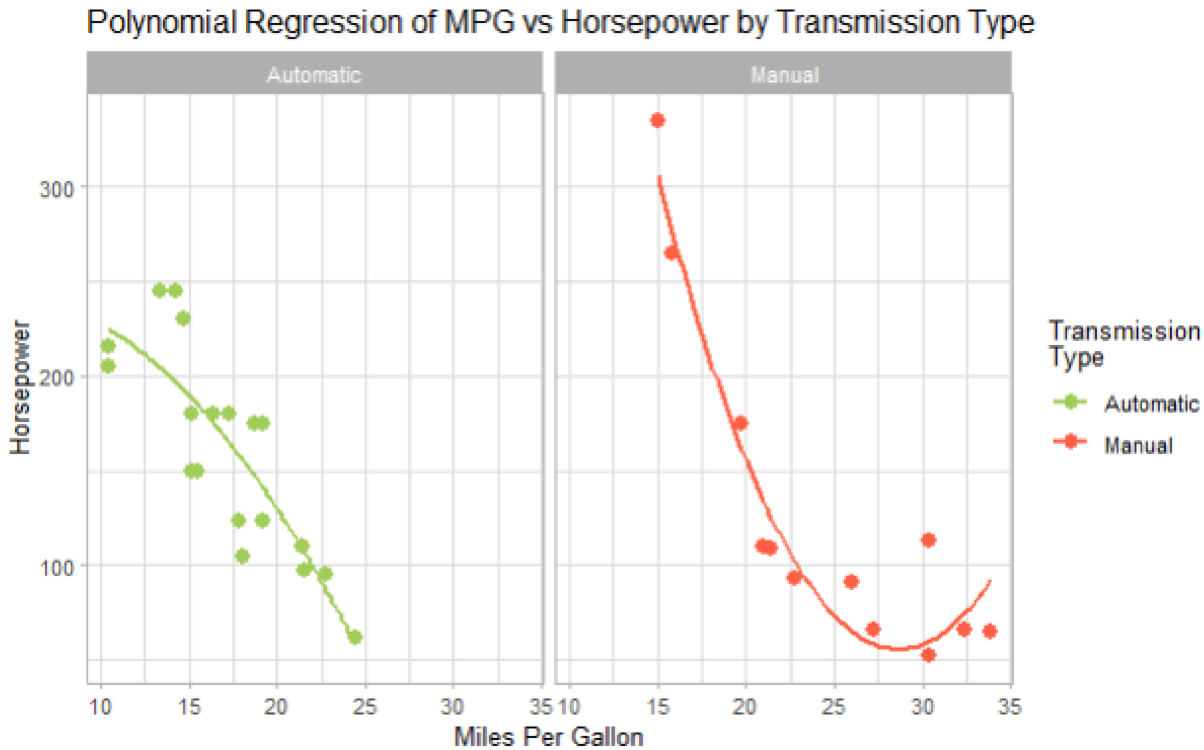
Linear regression is a statistical method that models the relationship between two variables by fitting a straight line through the data points. The chart indicates a negative correlation, meaning higher horsepower generally results in lower fuel efficiency. By using the relationship established between the **independent** (x) and **dependent** (y) variables, the regression line provides estimates for the hp value based on new inputs of the mpg variable.

```
ggplot(mtcars, aes(x = mpg, y = hp)) +
  geom_point(alpha = 0.6, color = "darkolivegreen3", size = 7) +
  geom_smooth(method = "lm", formula = y ~ poly(x,2), color = "tomato") +
    theme_light() +
  labs (x = "Miles per Gallon", y = "Horsepower", title = "Polynomial Regression of MPG vs
Horsepower")
```



Polynomial regression is a statistical method that models a non-linear relationship between variables by fitting a curved line to the data points. The curve in the chart can better capture the variation in horsepower as miles per gallon changes, offering a more flexible fit than a straight line.

```r
ggplot(mtcars, aes(x = mpg, y = hp, color = factor(am))) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", formula = y ~ poly(x,2),, se = FALSE) +
  facet_wrap(~ am, labeller = labeller(am = c("0" = "Automatic", "1" = "Manual"))) +
  scale_color_manual(values = c("0" = "darkolivegreen3", "1" = "tomato"),
                     labels = c("Automatic", "Manual"),
                     name = "Transmission \nType") +
  theme_light() +
  labs(title = "Polynomial Regression of MPG vs Horsepower by Transmission Type",
       x = "Miles Per Gallon",
       y = "Horsepower",
       color = "Transmission")
```



This chart uses **polynomial regression** to model the non-linear relationship between mpg and horsepower, split by transmission type using facet wrapping. The curved lines show that both automatic and manual transmissions exhibit a downward trend in horsepower as miles per gallon increases, but the polynomial fit captures the curvature, indicating more complexity in the relationship compared to a linear fit.

I believe I followed Few's guidelines a lot better in this set of visualizations. While I didn't use a monochrome color palette as Few typically recommends, I opted for a red/orange/green palette for all visualizations to both clearly represent the subject of correlation strength and maintain cohesion across all charts. This color scheme is commonly associated with denoting levels of intensity—green for strong, orange for moderate, and red for low—making it an intuitive choice for this analysis. The colors not only make the visualizations more engaging but also help to emphasize the differences in correlation strength in a clear and familiar manner.

October 20th, 2024 8:48pm