Example SLIDES CAN BE FOUND HERE: bit.ly/CPSC298Example

DATASET/SUMMARY

I am using the mtcars dataset, which is found in base R. It provides 12 stats (such as mpg, engine shape, transmission type…) for 32 different cars.

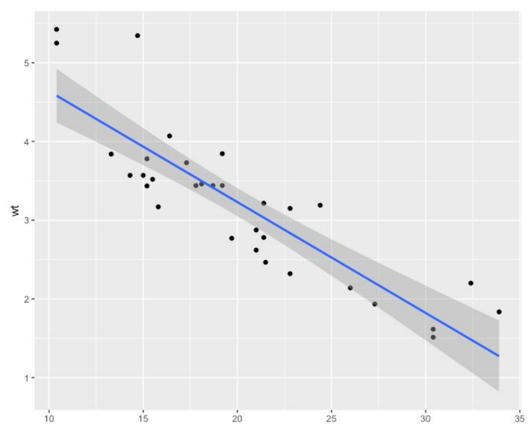
You will need to have a longer discussion of your data, telling me what each variable you use means

Summary statistics for this dataset can be found below:

```
      mpg         cyl         disp             hp             drat             wt
 Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760   Min.   :1.513
 1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080   1st Qu.:2.581
 Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695   Median :3.325
 Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597   Mean   :3.217
 3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920   3rd Qu.:3.610
 Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930   Max.   :5.424
      qsec        vs     am     gear        carb       cluster
 Min.   :14.50   0:18   0:19   3:15   Min.   :1.000   1:12
 1st Qu.:16.89   1:14   1:13   4:12   1st Qu.:2.000   2:20
 Median :17.71                 5: 5   Median :2.000
 Mean   :17.85                        Mean   :2.812
 3rd Qu.:18.90                        3rd Qu.:4.000
 Max.   :22.90                        Max.   :8.000
```

ANALYSES/GRAPHS

First, I ran **a regression model** in R to explore the relationship between weight of cars and MPG.

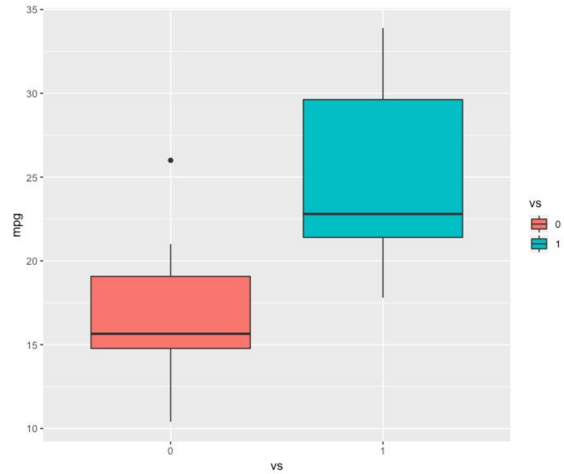



```
Call:
lm(formula = wt ~ mpg, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6516 -0.3490 -0.1381  0.3190  1.3684

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.04726    0.30869  19.590  < 2e-16 ***
mpg         -0.14086    0.01474  -9.559 1.29e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4945 on 30 degrees of freedom
Multiple R-squared:  0.7528,    Adjusted R-squared:  0.7446
F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

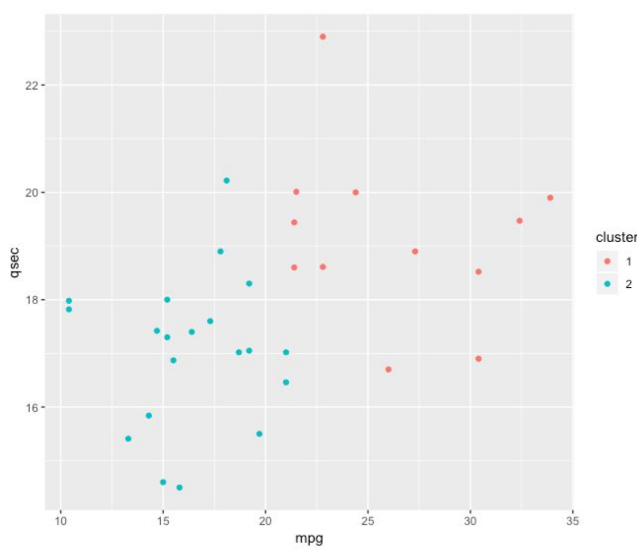
There was a significant ($p < 0.05$) relationship between weight and mpg. The relationship between weight and mpg is negative. As weight goes down, mpg goes up. So lighter cars tend to have better mpg which makes sense, since it takes more energy to move large objects.

Secondly, I was interested in whether engine shape (vs) was related to mpg. Using **an ANOVA,** I discovered that there was a statistically significant relationship between having a v-shaped engine and mpg. Cars without a v-shaped engine (1) had a significantly higher mpg compared to those with v-shaped engines. If someone wanted a car with good mpg, they should choose a straight (rather than v-shaped) engine.

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
vs           1  496.5   496.5   23.66 3.42e-05 ***
Residuals   30  629.5    21.0                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



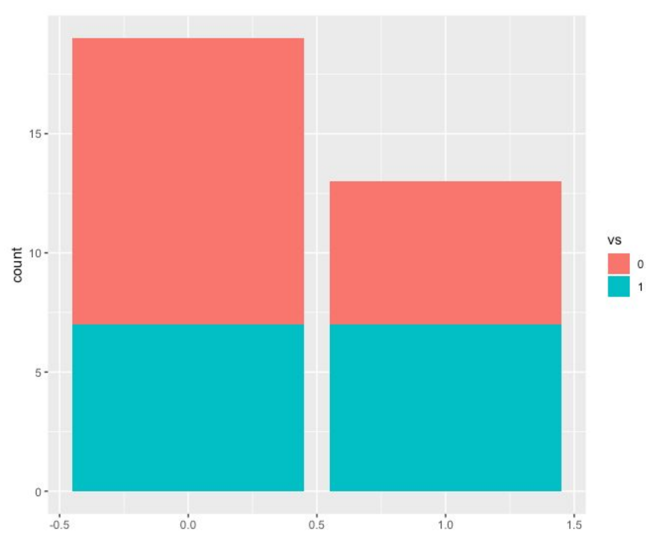
Next, I used **k-means** clustering with k = 2 to cluster cars into two groups based on mpg and qsec. There was one group with higher mpg and slightly higher qsec, and one with lower mpg and lower qsec. This makes sense to me as a car expert, because the better cars which can have a faster 1/4 mile time, also have good engines. Good engines also often lead to better mpg.

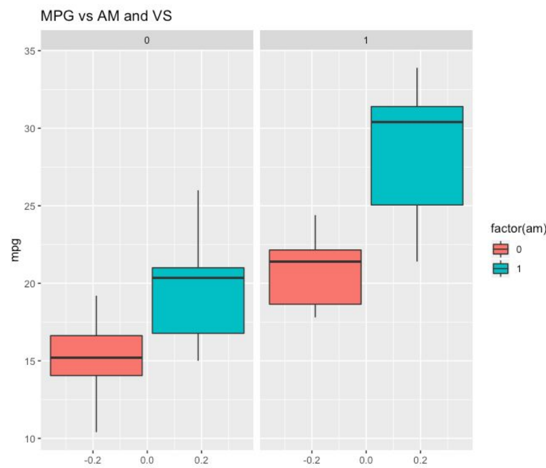


I also built a **logistic regression** model to predict whether a car had an automatic or manual transmission using mpg, the number of cylinders and weight. Overall, this model performed relatively well. Using cylinders, weight and mpg, the model was able to predict 29 out of 32 cars in the dataset correctly. That's an accuracy of ~90%. To improve this model, I might want to add some other variables like year manufactured.

```
Call:
glm(formula = am ~ mpg + cyl + wt, family = "binomial", data = mtc
ars)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.97676  -0.25913  -0.03874   0.21014   1.92638

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 18.37047   14.50717   1.266   0.2054
mpg         -0.07499    0.36600  -0.205   0.8377
cyl          1.23755    0.88856   1.393   0.1637
wt          -8.07603    3.30246  -2.445   0.0145 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.23  on 31  degrees of freedom
Residual deviance: 14.69  on 28  degrees of freedom
AIC: 22.69
```

Lastly, I used a chi square test to determine if engine shape was independent of transmission type. Having a v-shaped engine(vs) is not significantly related to a car being automatic (am) (p > 0.05). Since each engine shape can be used with either automatic or manual cars.





	FALSE	TRUE
0	18	1
1	2	11

```
        Pearson's Chi-squared test with Yates' continuity
        correction

data:  t
X-squared = 0.34754, df = 1, p-value = 0.5555
```