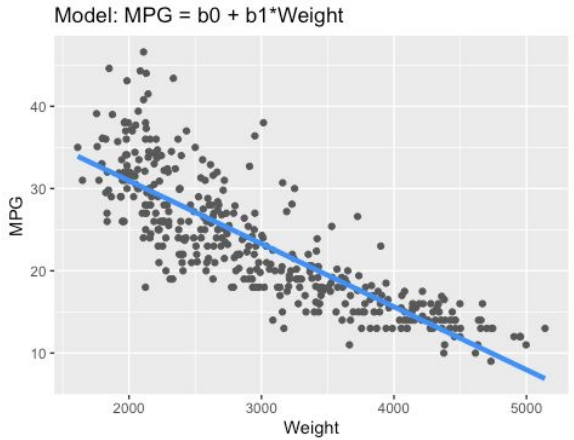
# Unit10: For Live Session

DS6306
Garrity

# PART 1 - Fit model

MPG = β0 + β1*Weight + ε




```
Hypothesis test:
Step 1: H0: β1 is equal to 0, Ha: β1 is not equal to 0
Step 2: Calculate t-crit
qt(0.975, dim(cars)[1]-2)
[1] 1.966034
Step 3: Calculate t-stat  (-29.73)
Step 4: Calculate p-value
pt(-29.73, dim(cars)[1]-2)
[1] 8.492343e-103
Step 5: Reject H0
Step 6: Conclusions and confidence interval:
confint(mpg_model, level=0.95)
                   2.5 %       97.5 %
(Intercept) 44.705532760 47.841351974
Weight      -0.008168061 -0.007154609
```

```
# sanity check the interval
mpg_model$coefficients[[2]] + (qt(0.975,
dim(cars)[1]-2)*0.0002577)
[1] -0.007154688
mpg_model$coefficients[[2]] - (qt(0.975,
dim(cars)[1]-2)*0.0002577)
[1] -0.008167982
```

**We have sufficient statistical evidence (p < 0.0001) that car weight is linearly related to MPG (i.e., slope parameter (β1) is not equal to one). For each additional pound of weight we expect a decrease in MPG of 0.0076. We are 95% confident that the true decrease in MPG per pound of weight is in the interval (-0.0082 MPG, -0.0072 MPG).**

```
mpg_model <- lm(MPG~Weight, data=cars)
> summary(mpg_model)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 46.2734424  0.7974987   58.02   <2e-16 ***
Weight      -0.0076613  0.0002577  -29.73   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1

Residual standard error: 4.332 on 392 degrees of freedom
Multiple R-squared:  0.6927,  Adjusted R-squared:  0.6919
F-statistic: 883.6 on 1 and 392 DF,  p-value: < 2.2e-16
```

# PART 2 - Fit two models and predict MPG

## Fit both models using LOOCV

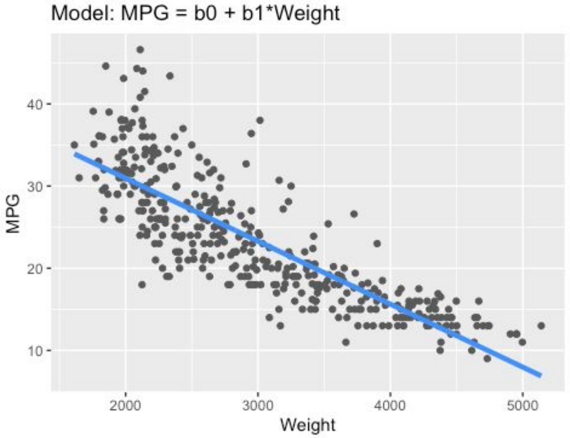
```
iterations = dim(cars)[[1]]
for (i in 1:iterations)
{
  carsTrain = carsNfold[-i,]
  carsTest = carsNfold[i,]
  Model1_fit = lm(MPG ~ Weight, data = carsTrain)
  Model1_Preds = predict(Model1_fit, newdata = carsTest)

  MSPE = mean((carsTest$MPG - Model1_Preds)^2)
  MSPEHolderModel1[i] = MSPE

  Model2_fit = lm(MPG ~ Weight + Weight2, data =
carsTrain)
  Model2_Preds = predict(Model2_fit,newdata = carsTest)
  MSPE = mean((carsTest$MPG - Model2_Preds)^2)
  MSPEHolderModel2[i] = MSPE
}

> mean(MSPEHolderModel1)
[1] 18.84765
> mean(MSPEHolderModel2)
[1] 17.53124
```

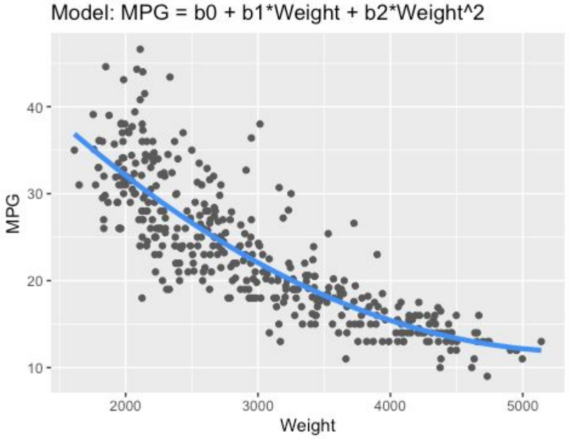
**Model 2 has the lowest MSPE**

## Predict MPG for Weight = 2000

```
pred2000 <- data.frame(MPG=NA, Weight=2000,
Weight2=2000^2)
predict(Model2_fit, newdata = pred2000)
32.06937
```

**A 2000 lb. vehicle is predicted to get 32 MPG.**



# PART 3 - Impute Missing Horsepower Values

EDA

# PART 3 - Impute Missing Horsepower Values

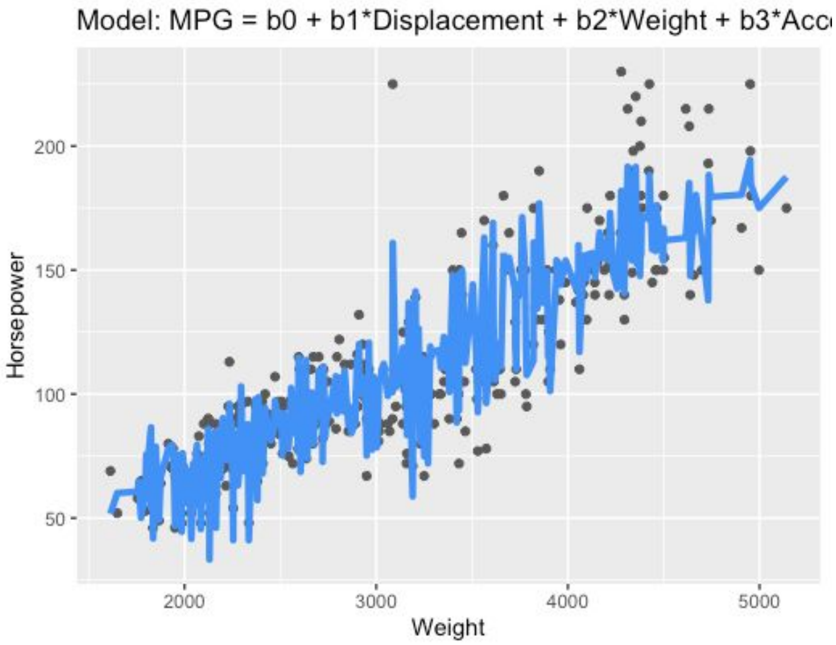
Fit model and fill missing values

**Fit model:**

```
> hp_model <-
lm(Horsepower~Displacement+Weight+Acceleration, data=carsHP)
> summary(hp_model)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  97.412459   5.669484  17.182  < 2e-16 ***
Displacement  0.106176   0.019832   5.354 1.48e-07 ***
Weight        0.020478   0.002256   9.078  < 2e-16 ***
Acceleration -4.797457   0.297833 -16.108  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 13.01 on 388 degrees of freedom
Multiple R-squared:  0.8867,  Adjusted R-squared:  0.8858
F-statistic:  1012 on 3 and 388 DF,  p-value: < 2.2e-16
```

**Impute missing values with model from above:**

```
cars_missing <- cars[which(is.na(cars$Horsepower)),]

filled_HP <- predict(hp_model, newdata=cars_missing)

cars$Horsepower[which(is.na(cars$Horsepower))] <- filled_HP
```



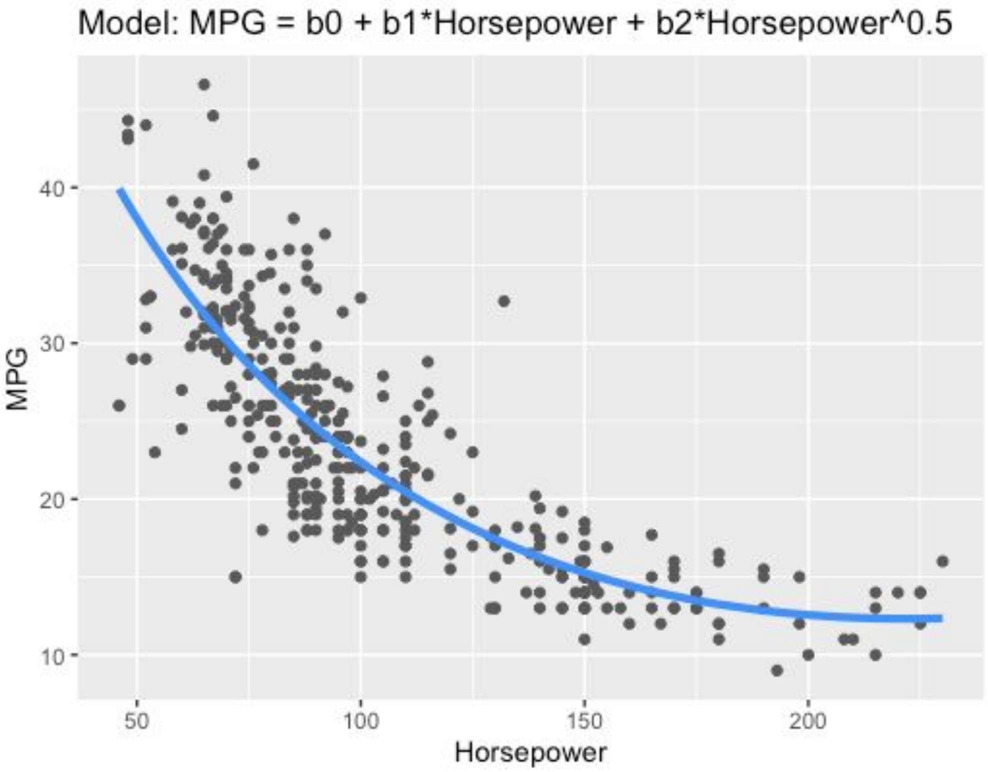

Struggled getting the 3D plot to work. argh!

# PART 3 - Impute Missing Horsepower Values

Model MPG with Horsepower

```
cars$Horsepower05 <- cars$Horsepower^0.5

hp2_model <- lm(MPG~Horsepower+Horsepower05, data=cars)
summary(hp2_model)

Call:
lm(formula = MPG ~ Horsepower + Horsepower05, data =
cars)

Residuals:
     Min       1Q   Median       3Q      Max
-14.5552  -2.5756  -0.2696   2.3272  15.5042

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  105.26637    6.66324  15.798  < 2e-16 ***
Horsepower     0.41849    0.05881   7.115 5.36e-12 ***
Horsepower05 -12.47366    1.26658  -9.848  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Residual standard error: 4.403 on 391 degrees of freedom
Multiple R-squared:  0.6834,  Adjusted R-squared:  0.6818
F-statistic:   422 on 2 and 391 DF,  p-value: < 2.2e-16
```

Why is this a positive slope???



## PART 3 - Impute Missing Horsepower Values

Predict MPG for car with 250 Horsepower

```
pred250 <- data.frame(MPG=NA, Horsepower=250, Horsepower05=250^0.5)
predict(hp2_model, newdata = pred250)
```

**12.66335**

## Takeaways & Questions

Not my best work.

For imputing missing Horsepower, it appears that there are predictor variables that are correlated. I believe this is multicollinearity, which is something that we want to avoid (?). Hopefully you'll have some time to cover this during live session.