1.

a. Generate scatter plots and correlations for the variables **EngDispl, NumCyl, ExhaustValvesPerCyl** and the **VarValveTiming** versus the target variable, **FE**.

- Can linear relationships adequately describe these relationships?
  FE with EngDispl and FE with NumCyl appear to have a linear relationship. ExhaustValvesPerCyc and VarValveTiming only have a few values, but potentially could be linear.

- Are there any outliers that you should investigate?
  From the scatterplots, there does not appear to be any outliers.

- What variable has the highest correlation with **FE**?
  EngDispl has the highest correlation with FE.

- What is the p-value for that correlation coefficient? Is it statistically significant at the 0.05 level? What can you conclude?
  The p-value is less than 2.2×10^{-16}, which is significant. We can conclude that there appears to be significant linear relationship between EngDispl and FE.

b. Generate correlations among all of the variables in the previously mentioned variables, minus the target, **FE**. Are there any notable relationships?
Yes! The EngDisp and the NumCyl have a high linear correlation.

c. Fit a simple linear regression model with **FE** as the response variable and **EngDispl** as the predictor.

- What is the value of the F Statistic and the associated p-value? How would you interpret this with regard to the null hypothesis?
  The F-statistic is 1803 with a p-value less than 2.2×10^{-16}. We can conclude that there is a significant linear relationship between EngDispl and FE.

- Write the predicted regression equation.
  $\hat{Y} = 50.56 - 4.52 EngDispl$

- What is the value of R-square? How would you interpret this?
  $R^2 = 0.62$. This means that approximately 62% of the variation in FE can be explained by EngDispl.

2. The IceCream dataset has two columns, **sales** which gives the total daily sales of a local ice cream shop in hundreds of dollars, and **temperature** which reflects the daily high temperature.

a. Run a regression analysis predicting daily sales from temperature.

- Are the errors of your model normally distributed? What evidence would you cite here?

YES, The QQ plot + histogram look fine.

- Do you see evidence of any relationship between temperature and sales? What statistical evidence (think: p-value) would you cite here?

YES, we'd cite the p-value for temperature which is <0.05 and thus we can reject the hypothesis that there is no relationship.

- What is the parameter estimate for temperature in the model equation? Interpret this parameter using a sentence.

1.08. For each degree increase in temperature, the expected value of sales increases by $108.

3. The dataset *MinnTemp* has information for the daily average temperature for a weather station in Minneapolis. The variables **temp** and **time** provide the temperature and time measurements respectively. Time is measured in hours since the study began.

b. Perform a regression analysis predicting **temperature** using the **time** variable.

- Are the errors of your model normally distributed? What evidence would you cite here?

YES, The QQ plot and histogram look fine.

- Do you see violations of our assumptions for simple linear regression? If so, what problems do you see?

YES, there is pattern in the residual plot consistent with a quadratic relationship which means our model is misspecified (i.e. violates linearity of the mean).

- Is there statistical evidence that **time** is related to **temperature** at the confidence level of 0.05? If so, describe the relationship in a sentence, if not discuss what your next steps in this analysis might be.

There is no statistical evidence of a relationship in the output, but the statistical evidence in the output is only testing for a LINEAR relationship. There appears to be quadratic relationship looking at the residuals so our next step might be to try a quadratic term in the model.

There is also evidence of an independence violation because of the striation in the residual plot. We'll get more information on how to deal with this in future classes.