Practicum
Predictions

Data about the nutritional components of popular cereal brands has been collected and made available for your team's use. We are interested in determining which other nutritional component is most closely associated with the amount of sugar contained in a cereal.

Your team will use the data to make predictions using linear models and compare the accuracy of your model to the rest of your classmates. Finally, the class will determine which team had the best prediction. Follow the directions below to explore and analyze the data:

1. You will have two datasets: one for training named **cereal** and one for test named **cereal_test**. Load both datasets. Write down the code you used.

2. Explore the training data. Which variable do you think is the best predictor of sugar? Choose at least 3 variables, make a plot for each one, and fit a linear regression line through each of them. Select the model that you think makes the best prediction.

3. For the linear model your team selected:

   a. Describe what the plot shows.
   b. Explain why you selected that particular model.
   c. Compute the mean squared error of your model using your test data.
   d. Now make a set of predictions with your test data. Calculate the mean squared error for the test data. Is it better or worse than for the training data, or about the same?

4. Present your teams' linear model to the class. Explain why you chose your model and the typical amount of error in its predictions.

5. Give an example of a prediction for one value of x. State that value, give the predicted sugar, and describe, based on the test data, how far off your prediction might actually be.