The dataset **ironman\_lake\_placid\_female.csv** contains data on female finishers of the Lake Placid Ironman Triathlon from 2002 to 2022. The motivation for this data analysis is to explore the relationship between swim, bike, and run times (in minutes) in order to gain insights into the performance patterns of the athletes. For this activity, we will specifically focus on times from 2022 Canadian Finishers. Open the file listed above and use it to produce output to help answer the following questions.

Model 1: Bike Times

1. Fit and report the least squares regression equation for predicting **Run Time** (the last stage of the event) using **Bike Time** (in minutes).
2. Examine residual plots for this model. Do you have any concerns about the appropriateness of this linear model?
3. Test (include all steps) if there is evidence that **Bike Time** (in minutes) is a useful predictor of **Run Time**?

Model 2: Swim Times

1. Fit and report the least squares regression equation for predicting **Run Time** using **Swim Time** (in minutes).
2. Examine residual plots for this model. Do you have any concerns about the appropriateness of this linear model?
3. Construct and interpret a 95% confidence interval for the population slope relating **Run Time** and **Swim Time**.

1. Based on your confidence interval, is there evidence that **Swim Time** is a useful predictor of **Run Time**? Explain briefly.

Model 3: Both

1. Now put both **Bike Time** and **Swim Time.** (in minutes) in the model as predictors of **Run Time**. Report the resulting equation below. This is a *multiple linear regression model*.
2. Choose a triathlete from the dataset and use this model to predict their **Run Time**.
3. Compute the residual for your triathlete.
4. Contrast the output from this multiple linear regression model with the output from Models 1 and 2. What differences do you notice?
5. Which model would you recommend using and why?