The dataset **ironman\_lake\_placid\_female.csv** contains data on female finishers of the Lake Placid Ironman Triathlon from 2002 to 2022. An ironman triathlon is comprised of 3 events. he motivation for this data analysis is to explore the relationship between 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.

A picture containing text, screenshot, diagram, line

Description automatically generated

1. What is the explanatory variable in this situation? What is the response variable? What type are both variables? What are the units?
2. What does each point in the scatterplot represent?
3. Report the least squares regression equation for predicting run time from bike time.
4. Suppose that you were interested in using this regression model to predict the run time for Sarah True, the USA Olympic Triathlete. What assumption about your data would you need to make?
5. Sarah True had a 295.5 bike time. What is her predicted run time? Regardless of your answer to the previous question, continue to use the equation from question 3.
6. Sarah’s run time was 184.1 minutes. How far off was the model prediction? Explain why we might see this observation.
7. Interpret the slope of the model in the context of the application. Be sure to be mindful of the units.
8. Interpret the intercept of the model in the context of the application.
9. Is the intercept interpretation meaningful? Explain.
10. What percent of variation in Run Times is explained by the model using Bike Time? What is the sample correlation?
11. Based on your comprehensive analysis, describe the relationship between run times and bike times using multiple pieces of information from your findings.
12. Given the structure of the Ironman race, where participants complete a 2.4-mile swim, a 112-mile bicycle ride, and a marathon 26.22-mile run in that order, explain how this information can be used to explore the relationship between the average speed at which a triathlete completes the bike and run portions of the race.
13. If you have the technology to do so, fit the least squares regression model predicting running speed from biking speed. Record the equation and interpret the slope coefficient in the context of the application.