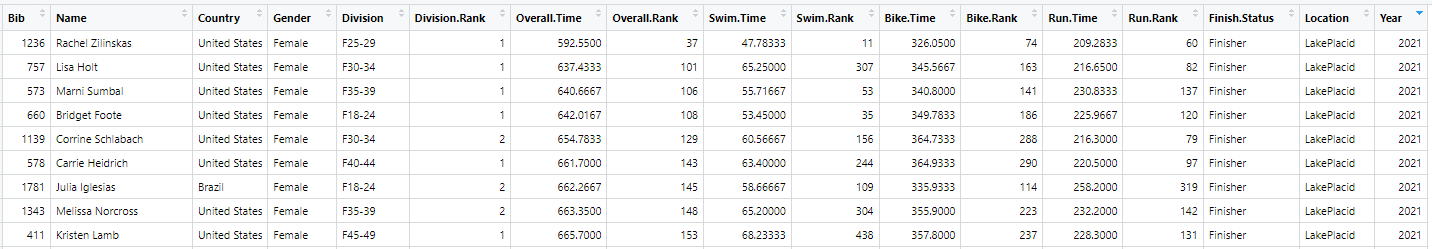
The dataset triDataLakePlacidFinal.json contains data on female finishers of the Lake Placid Ironman Triathlon from 2002 to 2021. For this exploratory analysis, we will be looking at several variables that are tracked each year for the race.



1. What are the cases?

Individual Triathletes

1. Identify some categorical variables:

Country, Gender, Name, Division, Finish Status, Location, Year

1. Identify some numerical variables:

Overall Time, Swim Time, Bike Time, Run Time

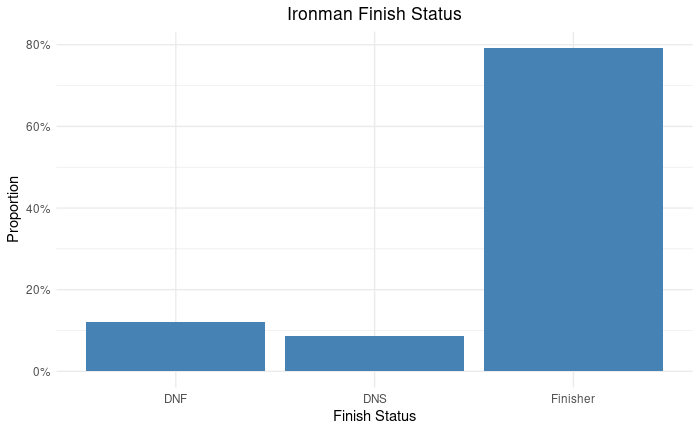
1. Pose two research questions involving a single variable:

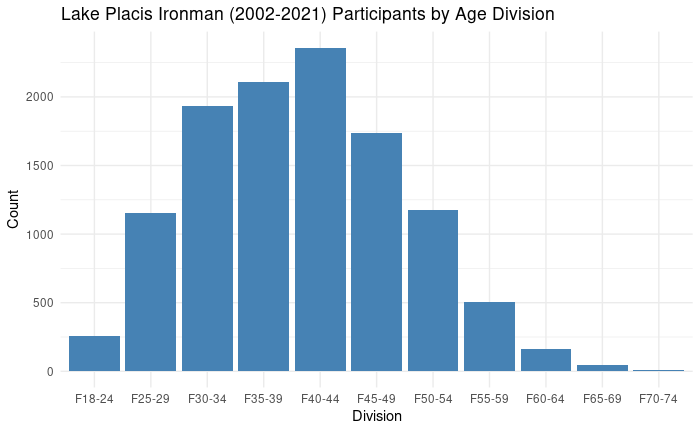
Answers may vary. Sample Responses:  
  
What percentage of participants are from the United States? Canada?  
What is the average overall finish time?

1. Pose two research questions about the relationship between two variables:

Answers may vary. Sample Responses:  
  
Are run times faster by participants from different countries?  
Do overall times change by year?  
Does division affect swim times?

**Bar Charts (One Categorical Variable)**

Make some observations about the Bar Charts below:

****

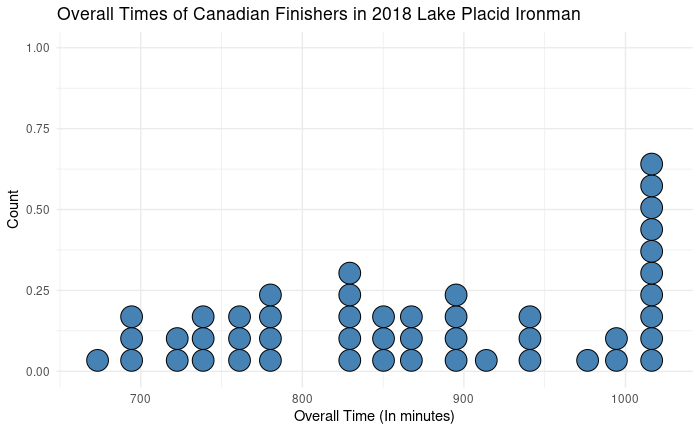
Observations: (Answers may vary)

Most people who participate in the Ironman are finishers  
About 2,500 total people have not finished the LP Ironman

Observations: (Answers may vary)

The biggest age group is 40-44  
Looks like a bell curve  
There are not many participants about 60

**Dot Plots (One Numerical Variable)**



Mean = 866.48 minutes  
N = 48

What percentage of Canadian Female Finishers had an overall time below 800 minutes?

\*\*Count the dots\*\*

Below 800 🡪 17 dots

17/48 = .3542 or 35.42% of Canadian Finishers had an overall time below 800 minutes