Step 2: Preparing Our Data

In Step 2, we **prepare** our data. To start, we'll want to select and import the data we intend to use. We might do this by asking a series of questions:

- What data are provided, if any?
- Does this data describe the problems we're trying to solve?
- Are there enough data points to adequately explore the issue?
- Is this data set trustworthy? How was it collected?
- Do we need to go out and collect more data?
- If so, where should we collect additional data?

Solving problems with data may require you to work with a number of different data sources. These include:

- Structured and unstructured data from the web (e.g., Google Analytics, HTML).
- Data from files (e.g., CSV, XML, TXT, JSON).
- Data from pre-existing databases (e.g., SQL, NoSQL, AWS).