Module 1: Data Collection and Design

Immigrants in the U.S. Gulf Region

Here's how you determine which variable goes in the columns and rows of a bivariate table, along with the principles behind those decisions:

- **General Guidance**
- **Independent vs. Dependent Variable:** The most common convention is to place the independent variable in the rows and the dependent variable in the columns. Recall:
  - **Independent variable:** The variable you believe is influencing or causing a change in another variable.
  - **Dependent Variable:** The variable you believe is being affected or changed by the independent variable.

Example:

If you're studying the relationship between study hours (independent) and test scores (dependent), you would likely structure your table as follows:

Study Hours (Rows)	Test Score: 0-50	Test Score: 51-80	Test Score: 81-100
0-2			
3-5			
6+			

Other considerations

- **No Clear Dependent Variable:** If you're exploring relationships without a predefined dependent variable, the choice is more flexible. Choose an arrangement that makes the most sense for analysis.
- **Focus of Interest:** Place the variable whose distribution you're most interested in analyzing within the columns. This makes it easier to compare percentages across categories.

In-case of percentages

Typical Case:

Column totals usually add up to 100% when the percentages within each column represent the distribution of the dependent variable for different categories of the independent variable.

Study Hours (Rows)	Test Score: 0-50	Test Score: 51-80	Test Score: 81-100
0-2			
3-5			
6+			
Totals	100%	100%	100%

How to develop research questions?

Step 1: Familiarize Yourself with the Data

Understand the Variables:

- Identify the variables included
- Determine the type of data for each variable (categorical, numerical, etc.).
- Note any missing values or inconsistencies.

Step 2: Exploratory Data Analysis (EDA)

Univariate Analysis:

- Examine each variable individually for distributions, frequencies, or summary statistics (mean, median, etc.) depending on the data type.

Bivariate Analysis:

- Create bivariate tables with basic percentages
- Or create visualizations (scatterplots, histograms, box plots) to examine potential relationships between pairs of variables.
- Or calculate correlation coefficients for numerical variables for a basic measure of possible association.

Step 3: Question Generation

Start with the Basics:

- What are the most common values/categories in each variable?
- Are there any unusual findings or outliers?
- What basic trends or patterns can you spot?

Look for Relationships

- Do changes in one variable seem associated with changes in another?
- Are there surprising correlations or a lack of expected correlations?

Challenge Assumptions

- Does the data confirm common beliefs about the topic or does it reveal contradictions?

Step 4: Refining Research Questions

- **Clarity and Focus:** Ensure your questions are clear, specific, and not overly broad.
- **Feasibility:** Can the questions be answered with the available data?
- **Significance:** Do your questions contribute meaningfully to the field of study or have practical implications?

Example

Let's say your dataset contains immigrant demographic information, educational attainment, and poverty status. Here are some potential research questions you could develop:

- What is there a relationship between immigrant status and socioeconomic status or educational attainment?
- Does higher educational attainment correlate with less likelihood of being in poverty among immigrants?
- Are there differences in educational attainment between different immigrant groups?