Making Sense of Your Dataset

Will your dataset allow you to fully address your research question?

How do your variables relate to each other? Do they correlate?

- Does it make sense for your variables to relate or correlate? How could they relate?

- Is one variable likely to modify another variable? Set your dependent and independent variable.

- Could any other related variables provide better insight into the relation of choice?
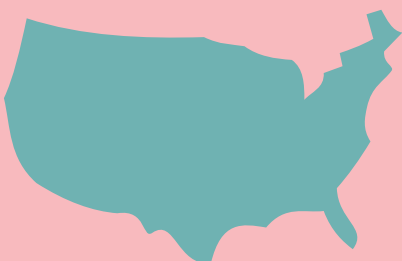
- If so, consider adjusting your question and dataset accordingly. Or, keep these in mind for your conclusions and future directions.

- Explain how you thought your variables would relate to each other at the beginning of your

Does your dataset reflect the scope of your research question?

- With what scope is your question best answered?
  - Broader trends: larger datasets (country or state) and longer time periods.
  - Localized trends: regional datasets and shorter time periods.

PENNSYLVANIA

BROAD → NARROW

- Is this scope manageable?
  - Is there too much or too little data to answer your research question? Do you have the tools to evaluate the dataset based on size?

Is there missing data? What are the implications?

- Is the missing data from a particular type of region?
  - Urban? Rural?
  - Political background?

- Does that region have a distinct diverse community?
  - Low income?
  - Minority racial groups?
  - High cost of living?

- Is this missing data indicative of something about these regions and the topic of your question?

- Can you still answer your question based on the scope with this potentially indicative missing data?

- Note: You may have to do additional research to understand why some data may be missing.

Making Sense of Your Data Set

Will your dataset allow you to fully address your research question?

What filters, if any, do you apply?

- Is the dataset too large? Try selecting a random representative sample.

- Can scope be maintained by with your sample? Would reducing scope help?
  - Trying changing the region or time frame you're looking at.

- Is the data still representative? Does your data reflect the diversity of your sample?

- Note, how much data you sample is based of your research question. You may have to modify your question if the dataset is unable to answer it.
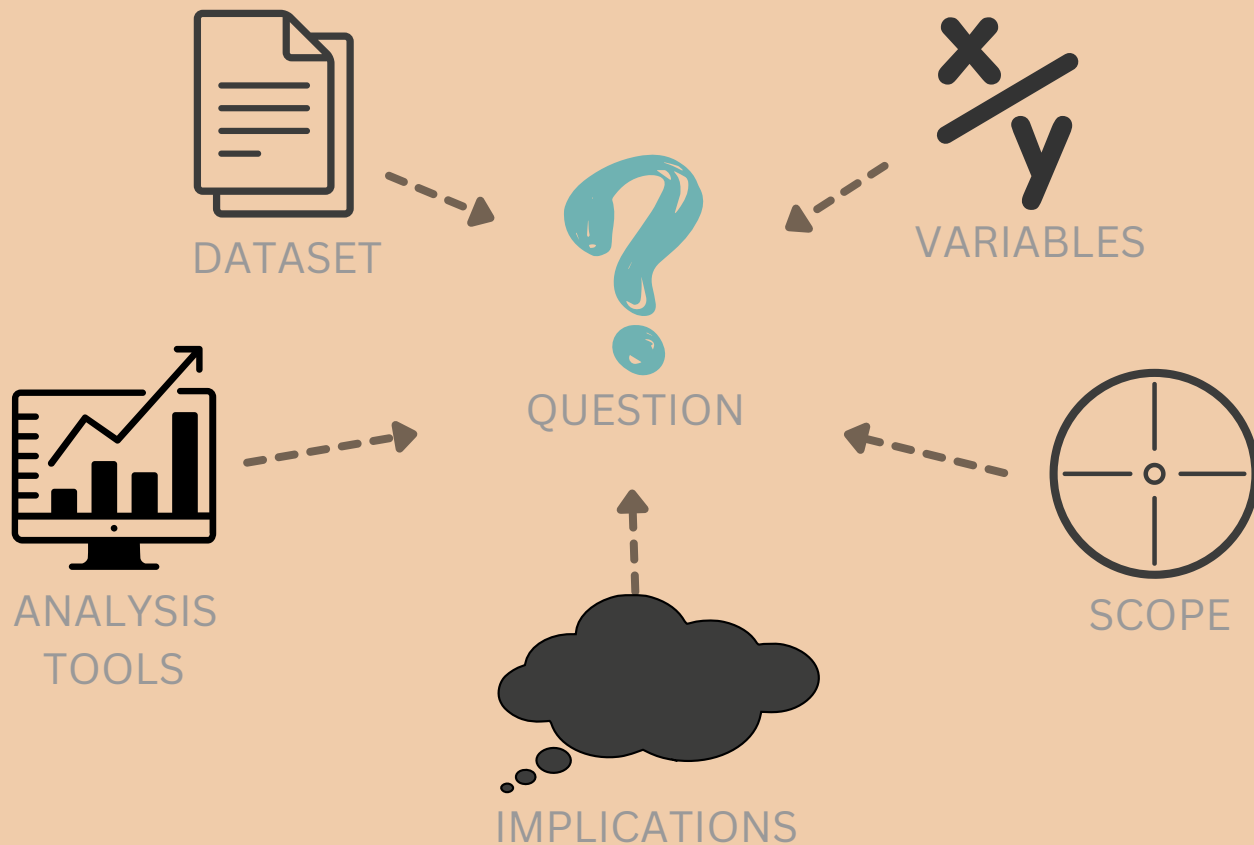
Is your question still answerable from the dataset?

- With the dataset filtering you did, does your dataset still answer your original question?

- Would modifying the question in terms of scope and variables align with your dataset?

- If your dataset is too limiting or small, you might need to change your topic and question!

- If you're really stumped, trying working backwards from datasets that address your general idea and then forming a question.

Final Check!

DATASET → QUESTION ← VARIABLES

ANALYSIS TOOLS → QUESTION ← SCOPE

IMPLICATIONS → QUESTION

If everything has been checked against your question, you're ready to analyze!