

Backwards Design Template

October 24, 2024

1 Topic:

What problem are you (or your stakeholder) trying to address?

2 Project Question

What specific question are you seeking to answer with this project?

3 Project Hypothesis

What is your hypothesized answer to your question?

4 Model Results

One of the hardest parts of developing a good data science project is developing a question that is actually answerable. Perhaps the best way to figure out if your question is answerable is to see if you can imagine what an answer to your question would look like. Below, draw the graph, regression table, etc. that you would consider to be an answer to your question. Then draw it

again, so you have a model result for if your hypothesized answer is true, and a model result for if your hypothesized answer is false. (If the answer to your question is continuous, not discrete (e.g. “what is the level of inequality in the United States?”), draw it for high values (high inequality) and low values (low inequality)).

Result if your hypothesis is true

Result if your hypothesis is false

5 Final Variables Required

Now that you’ve specified what an answer to your question looks like, what data do you need to generate that answer?

You don’t have to know the exact variable and dataset (“I need *ANNPOP2008* from the *NHGIS 2019* census 1% sample release”), but you should be specific enough that all properties that are critical to your analysis are fully specified (e.g., I need individual-level annual income data for a nationally representative sample of all working US citizens).

Basically, this should be a list of all the variables (not in terms of specific variable names, but the substantive content of each variable) you want in your final dataset **and** an explanation of the population represented in that data (i.e., what is represented by each row of your data, and what is the group of entities included in the data).

6 Data Sources

Given the variables you need for your analysis, what actual data sources do you think will have the data you need?

*In specifying the datasets you need, if you list more than one **also** indicate how you think you can relate these datasets (i.e. if you're gonna merge datasets, what variables do you think those datasets will provide that will allow you merge them? There's no use saying "I'll merge this political survey with medical records of who has received bad care" if the political survey doesn't provide identifying information you can use to link survey respondents to medical records, even if you have both the survey and medical records!)*

If you are familiar with Entity Relational Diagrams for databases, something similar is a great way of diagramming how you expect datasets you plan to use will relate to one another.

7 Division of Labor

Now that you have identified what needs to be done, how do you plan to divide those tasks among team members?