Week 8: Critique a Research Design

8.1 Critique It

8.1.1 Why This Approach?

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

Here we want to practice how to apply the principles of research design that we've talked about up until this point. In this section, we will talk about what you should look for when you either develop your own design or when a colleague comes to you to ask for feedback on their project proposal.

Think about what should be included, and think about the pitfalls that we should try to avoid. Overall, it's important to practice how to provide critical feedback with concrete recommendations on how one can improve their project.

8.2 Overview of Research Project

8.2.1 Goal and Context

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

The headline takeaway here is to link the project to the ask. What's the objective of the project? What's the goal? A brief overview of the project will include necessary background information on the domain and on the problem. This will include the justification for the problem-- for the project.

This is the so what or this is why this matters. And this is a necessary step of a successful project. You might think to yourself, well, of course, everyone knows the topic. And of course, everyone knows why it's important. But guess what? The room is often full of skeptics.

Here you want to articulate the intended audience. This explicit articulation of the intended audience may or may not make it in the final report. But you should at

minimum use this as a guiding heuristic. That is, what language and what design would make the most sense to this audience? Let's go through an example.

8.2.2 Research Question

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

Here you'll include your research question. It's possible that you'll have an overarching question. And then you may have smaller subquestions. If you plan to test hypothesis, each of the question may translate into a separate hypothesis. Here it's also really important to define any key terms.

Even if your intended audience knows the subject matter pretty well, it's important that your audience is familiar with how you use particular terms. We often use terms in technical ways but forget that there's other colloquial uses of the terms. Also, your audience may not be willing to admit that they don't know what a particular term means. So help them out and spend a little time here defining any particular key terms that you'll use throughout the project.

For example, the research question is do letters that encourage individuals to turn out increase the vote in the California general election? Here we're talking about old school letters via the United States Postal Service. We intend to explore if certain messages turn out voters more than other messages. And to be clear, we will only know if someone turned out to vote. We will not know who an individual voted for.

8.3 Research Question Critique

8.3.1 Research Question Critique

In this and subsequent discussion questions, you will be asked to comment on the various aspects of a research design.

Spend five minutes on the following prompt:

Imagine you work for an airline company. You are on a team that is tasked with how to improve customer experience. Your team has proposed the following research question:

- How can we improve customer experience?

Your team wants to know your first impressions of this question. Provide feedback to your team.

Highlight the strengths of this research question and how you might improve it. If applicable, propose a revised version of the research question. Explain why your suggestions improve the research question.

8.4 Your Prior Expectations

8.4.1 Theory

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

In this section, you'll want to express your prior expectations. Articulate what you expect to find and why. Describe the existing work in this area and how it relates to your proposed study.

Now, in some domains, you may not have best practices for the type of project you want to develop. In that case, you may have to build from related work. Some projects may not have strong theories because of the approach to the problem. You might be theory agnostic, or you may not have strong priors. That's OK too. Make that point explicit.

For example, your theory statement might go something like this. Our theory is that letters sent to voters will prime social norms of electoral participation and will increase the probability that a voter cast a ballot. Existing studies suggest that one piece of mail can increase turnout by about 0.5 percentage points compared to those who do not receive a letter. We expect a similar result in this study.

8.4.2 Hypotheses

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

If you're hypothesis testing, make it very explicit what you want to test. Here you want to articulate your directional or nondirectional hypothesis. An explicit articulation of your tests helps avoid p-hacking. P-hacking is when you run a lot of analysis and only report what is statistically significant. But remember, just because something is statistically significant does not mean that it's substantively interesting.

In academia, it's common to preregister a study. When you preregister a study, you are explicit about the planned analysis. And you commit to report on all of the analysis, even aspects of the

project that did not work out as planned. You can do something pretty similar in industry. Describe to your audience what you plan to test, and tell them that you will report on all the findings.

This is a form of commitment mechanism. Let's go through an example. My hypothesis is the group of voters who receive a letter that encourages them to vote will turn out at higher rates than the group of voters who did not receive a letter.

Let's talk about preregistration. Here we'll test the hypothesis listed above. Depending on our final budget, we may send out additional variations of letters to test whether or not the effect of the letter is due to the message contained in the letter or whether or not the effect is present regardless of the content of the letter.

If we have sufficient budget, we will test and report on the following hypothesis. Letters that prime the social norm of participation will have a greater effect on turnout than letters that only provide information about the upcoming election.

8.5 Hypotheses Critique

Spend five minutes on the following task:
Provide one or two positive comments on the current set of hypotheses and provide one or two comments on how the set of hypotheses may be improved.
Remember the goal of this research is to improve the customer experience.
Hypotheses:

1. The season will impact customer experience.
2. Customers will experience more stress during the U.S. Thanksgiving holiday than at other times of the year.
3. Customers that belong to our point rewards program will expect more quality attention from the customer service team than customers who do not belong to the rewards program.
4. Free drinks on the flight will increase customer satisfaction.

8.6 Data and Sample

8.6.1 Data

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO3: Assess and select specific data and the data collection methods that best fit a specific outcome or need.

- LO7: Imagine, plan, and design a data science project.

Describe the data you anticipate you will need. You can propose the collection of new data, or you can use existing sources. To help narrow down what existing data sources you may use, think about one, open source data, and two, prospective audience.

Think about the data source you already have access to. If you're going to encourage your audience to collect additional data, make it clear what the value added is. In other words, you have to convince them that the pain of collecting new data is worth the gain.

Now, you could list multiple sources of data because you think it will produce the most comprehensive answer to your question, or you may list multiple sources because you are unsure about the availability of data. In this section, you may want to talk about how you conceptualized and operationalized the concepts of the study. Let's go through an example.

8.6.2 Sample

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO3: Assess and select specific data and the data collection methods that best fit a specific outcome or need.
- LO7: Imagine, plan, and design a data science project.

Describe your sampling frame. In other words, what's the list you will sample from? Are you using probability or nonprobability sampling? Give a rough estimate of the number of subjects or observations you will need. Give your best guess based on existing work. You may also want to conduct a statistical power calculation. And you'll go into this in other classes.

But in brief, a statistical power calculation will provide the-- will give you information about the minimum detectable effect given the design and the hypothetical sample size. And you do this because you want to understand what's the minimum number of observations you can collect and still have a chance of detecting an effect. Let's go through an example.

The sampling frame is the voter file from the California Secretary of State. The sample will include an appropriate number of votes that reflects the budget of the organization. We anticipate that we'll be able to include 600,000 subjects in our study. After we touch base again with key stakeholders, we will determine if we will use sampling techniques other than simple random sampling to make sure our sample reflects the demographics of California.

8.7 Method and Analysis

8.7.1 Method

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO4: Justify an analytic approach that informs decision making.
- LO7: Imagine, plan, and design a data science project.

Here you will articulate the design of your proposal. If you're doing observational work, describe if you'll gather data from a single point in time, i.e. a cross-section, or whether or not you'll gather data over time, time series or panel data. If you're doing an experiment, describe the intervention. Are you using machine learning or some other form of artificial intelligence? Provide some context here.

Overall, describe the method and the tools you will use. Describe why the method is appropriate and what the trade-offs are of this method compared to the use of other methods. Let's go through an example.

The experimental method is ideal to assess the causal impact of receiving a letter on one's decision to turn out. Critical to the experimental design is random assignment to treatment. Assignment to the treatment and control in this study will be determined in the following way. We will split the sample into equal-sized control and treatment groups.

8.7.2 Analysis

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO4: Justify an analytic approach that informs decision making.
- LO7: Imagine, plan, and design a data science project.

In the analysis section, you describe the comparisons you plan to make. What relationships are you looking for? Would you describe, predict, or explain? For example, here we'll use conventional political science statistical methods to compare participation rates of voters who received the letters, i.e. the treatment condition, with voters who did not receive the letters, i.e. the control condition. In order to more precisely assess the effect of the treatment, we will account for demographic factors that have been shown in the political science literature to influence turnout.

In the analysis section, you may also identify potential risks in your approach and how you may overcome them. Examples of these risks include but are not limited to potential legal or ethical issues with how you acquire the data or how you use the results.

There might be concern that you might be unable to get the data. You might be concerned about costs or political or organizational pushback. You might be concerned that the idea is not in line with the organization's vision. Let's go through an example.

Here we think individuals are used to receiving unsolicited mail. However, it's possible that this intervention may cause stress or anxiety for the voters who receive our letters. It's also possible that we will not have enough statistical power to detect an effect-- to detect the effect that this intervention intended to cause given the sample size, design, and anticipated effect size. Existing literature finds that the effect size of a single piece of mail is about one half of a percentage point. And we'll continue to consult with experts on how to best think about statistical power.

8.8 Which Method?

Spend five minutes of the following prompt:

Remember the goal of this research is to improve the customer experience.

The company wants to focus on how the in-flight experience contributes to one's overall satisfaction with the airline. More specifically, the company wants to test the hypothesis: customers who receive free drinks in-flight will be more satisfied with the airline than customers who do not receive free drinks.

The team will consider two approaches to capture customer sentiment:

Approach 1: An Observational Study
The plan is to ask customers the following questions:

- How was your experience during the flight?
- Did you receive any complimentary alcoholic beverages?
- What is the quality of your overall experience with the airline?

Approach 2: An Experiment Study
The plan would be to assign certain flights to distribute free drinks and then ask customers the following questions:

- How was your experience during the flight?
- What is the quality of your overall experience with the airline?

Provide one or two comments on suggested Approach 1 and provide one or two comments on suggested Approach 2.

8.9 Deliverables

8.9.1 How Will You Communicate Your Findings?

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO5: Identify the audience and the most effective method to communicate a persuasive argument.
- LO6: Navigate organizational, personal, legal, and ethical constraints to facilitate better decision making and improve communication.
- LO7: Imagine, plan, and design a data science project.

Articulate the anticipated timeline to execute the research that you've developed. Also articulate what final deliverable the audience can expect. Your final deliverable will depend, or will vary, based on the audience and the research task. It could be a slide deck and a presentation, a paper or a report, a proof of concept, or a minimum viable product, or it could be a code-based deliverable. Maybe you're expected to produce a new model.

Let's go through an example that has kind of month-by-month outline of what's going to happen and what the deliverable will be. Here's an example. In May, we will consult with election experts to determine the best treatment language, and to further assess the ideal sample size. In June, we'll reassess the necessary sample size and the organizational budget.

In July, we'll make the go or no-go decision on the project. In August and September, we'll finalize the treatment language and the sampling strategy. And importantly, we'll consult with the United States Postal Service regarding the best date to send the letters. In October, we'll print and merge all the material.

In November, we'll send out the treatment right before the election. In December, after the election, we'll acquire the post-election voter file from the California Secretary of State to obtain turnout information. And finally in January of the year following the election, we will conduct the analysis as described above, and we'll provide a report that outlines our findings. We will also prepare a brief slide deck that outlines the report.

8.10 How to Provide Critical Feedback

8.10.1 Not All Designs Look the Same

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO5: Identify the audience and the most effective method to communicate a persuasive argument.
- LO6: Navigate organizational, personal, legal, and ethical constraints to facilitate better decision making and improve communication.

- LO7: Imagine, plan, and design a data science project.

Here we present a broad outline of how to design a project. But just remember there's going to be quite a bit of variation in the design you develop. This variation could come from the domain that you're in or your approach. Is the research qualitative, quantitative, a little bit of both? And some of this variation is going to come from the goal of the project. Is your goal to describe, predict, explain? Are you expected to provide explicit recommendations?

One thing that should be common among these designs is the following guiding heuristic-- your audience. Perhaps at this point, you've noticed that this is a major theme in the course.

Is this an external or internal-facing audience? Have you already received permission to do the study, and you need to provide an outline of the plan? Or are you asking for permission? Is your audience a group of venture capitalists? Remember, you want to design a project in a way that will be recognizable to your audience.

8.10.2 Focus on Actionable Feedback

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO5: Identify the audience and the most effective method to communicate a persuasive argument.
- LO6: Navigate organizational, personal, legal, and ethical constraints to facilitate better decision making and improve communication.
- LO7: Imagine, plan, and design a data science project.

Let's talk specifically about how to provide feedback on a colleague's design. The most important thing is to focus on actionable feedback. In other words, give feedback that will allow them to make changes within their particular time frame.

It's helpful to identify which aspects of the design are problematic. But it's even more helpful to identify what's wrong and why, and then provide a recommendation for how to move forward and how to improve the project. Let's go through an example.

Let's say you work for a watch company. And you want to get a sense of how popular a new line of watches could be. Your colleague proposes a survey of individuals who have previously purchased watches from your company to see if they would be interested in a new watch with a certain set of features.

Your colleague asked for feedback because they know that you know something about research design. You may recommend to your colleague that they use observational data to see what some of the common features are for some of the most popular watches purchased.

Rather than start with a particular watch design and survey customers if they'd be into it, instead you recommend that your colleague looks for patterns to get a sense of what the popular watch design features are to help inform the design of the new watch. You may also recommend that your colleague surveys those who purchase watches from other companies. Maybe there's a potential to gain new customers.

8.10.3 The Compliment Sandwich

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO5: Identify the audience and the most effective method to communicate a persuasive argument.
- LO6: Navigate organizational, personal, legal, and ethical constraints to facilitate better decision making and improve communication.
- LO7: Imagine, plan, and design a data science project.

Let's talk about some of the mechanics of how to provide helpful feedback. In response to your colleague about their project, you may want to provide section-by-section feedback. Now, if they ask specifically for you to point out only the areas that need improvement, I recommend you follow their request. Otherwise, I recommend that you point out both what they did well and what they can improve upon.

There's two reasons for this. One is relationship building. A lot of what we do is relationship maintenance. And I want you to give your colleague praise for what they did well. Otherwise, your feedback may be perceived as just a laundry list of negative comments. And two, it's important to reinforce good habits. If you point out what your colleague did well, you'll help reinforce best practices.

In addition to section-by-section feedback, I recommend you give a few sentences that provide an overview of your overall impression of the proposal. Start with a positive comment. Provide an outline of the critical feedback, and then end with a positive. This will help convey an encouraging tone. Let's go through an example.

You might tell your colleague, overall, I think your design is well crafted. Your research question is clear. And you provide sufficient context around the project for nonexperts to understand why this is important. You can find more comments in each section below.

My headline comment is that you should focus on clarifying the data source you intend to use. It sounds like it will take a lot of time and money to collect that new data. A few more sentences on how this new data will help you better understand the question compared to existing data would be super helpful. Again, overall great work. I think this project has promise. So this kind of positive-critical-positive sandwich can help smooth over critical feedback.