

Course Project

In this course, you will develop a research project that uses textual data to answer a research question in social science. There is a fair amount of flexibility in the project, meant to allow you to adapt the project to suit your personal goals. This flexibility exists in two primary domains: the research content (what you study) and how you deliver the research. Your guidelines in these domains are as follows.

Research Content

Fundamentally, text as data concerns a method of inquiry. Ultimately, the methods you learn in this course are only valuable if you can apply them to answer questions that you and others (your employer, other social scientists, etc.) care about. Because of this, the course is oriented around your application of these methods in order for you to develop a research project over the course of the semester. Thus, what you take away from this course will, in large part, depend on what you put into the research project.

You're encouraged to research a question that you care about and which helps you move your academic and career goals forward. You are encouraged to select a topic that you might use for your capstone project so that you can leverage the content of this course toward the completion of it. If you are presently employed in some research capacity, you might also wish to choose a topic relevant to your work if doing so will help you professionally.

With that in mind, you don't have to choose a topic that is obviously related to political science. However, the question should be **interesting**. Asking interesting questions is perhaps the most important quality in a researcher. Good data leveraged at a good question will almost always be interesting. A bad question, no matter how well it is answered, is never interesting. Good questions have the following characteristics, which are more important than the exact substantive content of your project.

- **Good questions do not have obvious answers.** The answer to your question should not be obvious until after you've conducted the research. For example, a question like, "Is college tuition more affordable in countries with higher tax rates?" is not a good question, because somewhat obviously we'd expect the answer to be "yes." However, a better question might be, "Why are Americans who have already completed a university education more willing to pay taxes that subsidize university costs than those who haven't yet earned a college degree?" Why is the latter more interesting? It implicitly notes a *puzzle*, or a curious empirical observation without an obvious explanation. In this case, the puzzle is that people who no longer benefit from college subsidies (because they've completed their degree) are willing to make a costly contribution to the education of others, but those who still might benefit don't support the subsidy.

Now, using the stylized example above, we can easily think of a few possible explanations: for instance, support for subsidies might be endogenous to interest in completing a college degree. However, the possibility of generating explanations is precisely why the example question is a good one – it generates testable explanations for a clearly stated empirical phenomenon. Interesting questions generally either (a) note a clearly puzzling empirical phenomenon or (2) explain how two established literatures generate competing explanations about a question and ask, which is correct?

- **Good questions are answerable.** A common temptation is to ask questions that are interesting but unanswerable. For example, questions like, “How will China’s foreign policy behavior change when it reaches a level of international power equal to that of the USA?” This is an interesting question, in that the answer is not obvious and we’d all like to know it. However, it is not an answerable question – there exists no available data in the history of the world that would easily resolve this question.

We could, of course, look at the behavior of previous rising powers and attempt to extrapolate from those cases to China. However, such an answer isn’t vulnerable to being proven wrong – that is, because China’s behavior in a yet-to-occur scenario is not knowable, ultimately the question of how China might behave is not testable. A better question would be, “How have rising powers behaved upon reaching parity with the status-quo hegemon?” This question makes explicit the implicit “answerable” question in the first formulation – namely, what can we learn about previous cases similar to the foreseen-but-not-yet-real case of China.

If you meet both of the above objectives and ask a question about social and/or political behavior, your question is appropriate for this course, assuming you can get relevant textual data.

How you deliver your project

As noted, the intention in this course is for you to learn skills that move your professional ambitions forward. Similarly, you’re given flexibility in this project so that you can optimize it to serve your long-term goals. With this in mind, upon selecting a research question strategically, you’re similarly encouraged to deliver your project in a manner that best supports your goals. Your options are as follows.

- **A research paper.** If you do not have a strong sense of which option to select, you should write a paper. Papers are the bread and butter of academic research, and if you’re so inclined, you’re encouraged to work with me to write a publishable paper. If you hope to continue your project, perhaps as a capstone topic or in your career, you should certainly write a paper.

- **An online post.** If you're presently employed in a research position or would like to be, it might be advantageous for you to present your research online as a (long) blog post, to serve as a part of a portfolio in an online web presence. You will never get your research published, but the skills you develop in order to present your research online (R Markdown, HTML, creating a personal website, etc.) are also skills you'd use as a data scientist at the New York Times, for example. If these skills serve you, you should develop them in this course. Compared to a research paper, this approach will require considerably less literature review but considerably more effort on graphics, visualization, and web design.

- **A presentation.** Last, you're also welcome to deliver your research through a slide deck with an oral presentation. Presentations are common forms of communicating research in academia, but even more so in consulting and business. If your career ambitions are more likely to situate you in front of a board room than in front of three peer reviewers, you may prepare a slide deck and record a corresponding oral presentation, and can deliver both. Functionally, this approach will require the same research required by a research paper (for instance, you must include a substantial literature review), but in comparison, you will spend more time learning how to present your research quickly and how to present it orally, and less writing.

Grade

The course project will serve as the primary source of variation in your grade (so you're **strongly** encouraged to take it seriously). Your weekly assignments are oriented toward moving your project forward and will be graded leniently. Your research project will not be, and must meet the following expectations.

[1] Your research must direction address the question you identify. Many research projects fail on this goal. In short, the data will not actually have much to say about the research question you attempt to answer. You should think carefully about your question and the empirical implications of it when selecting the data you collect and analyze.

[2] You must analyze the text you collect correctly and carefully. We will learn many methods for answering different questions about textual data in this course. However, you may need to do independent research into the method you select for your research, or may have to use techniques we did not cover in the course to test the success of your method, preprocess as necessary, etc.

You're encouraged to discuss your project with me and with your peers on the Discussion Board to ensure that you're doing everything you can to answer the question well with the method you select. If this seems a bit "open-ended," it is meant to. Academic research often leads the researcher to problems slightly beyond those which they learned in class; learning how to identify and solve these problems is what separates "students" from "researchers" This course

is designed to teach you to be researchers who employ text rather than students who were taught some things about text.

[3] Your project must identify a good question and treat it fairly. This is the least “text as data” part of the project but it is very important. As mentioned above, a good question is critical to a successful research project. Regarding “fair treatment,” you’re expected to fairly and comprehensively review the literature on the question you identify, to motivate your question around this review, and to attempt to fill a gap in knowledge about the topic you research.

Conclusion

My intention with this course and this assignment in particular is that you develop applicable skills that you can use to answer questions you care about, for either personal or professional reasons. This intention is reflected in this assignment, and because of its centrality in the course design, you’re strongly encouraged to select your topic and question carefully. I’ll be available throughout the course to answer questions you might have about the research process and to hopefully guide you through conducting research with text as data.