**Please provide a general description and justification for your project.**

The Data is Plural archive includes a dataset on social scientists testifying in congress. People who have very extensive experience in a certain topic could provide an excellent resource in the consultation of legislation. Because the individuals in the dataset are social scientists, one can assume that they would prioritize social benefits of whatever legislative process the congressional committee is going through. For example, an economist might make suggestions on the budget based on the financial needs of the country. The presence of social scientists in congress could have positive effects overall. Understanding what makes a solid testimony could be important for those interested in speaking to congress and making a strong impact.

**Who is the end user for this project? What problem does it solve for them? Is the problem urgent and pervasive? Is someone willing to pay for a solution?**

The end user theoretically would be those giving testimonies to congress and organizations that employ them. The problem that this solves is avoiding weak statements to congress. Testimonies can be the most memorable portions of a hearing- speeches are sometimes circulated through media and have the ability to even sway public opinion. Optimizing testimonies could save time for those delivering speeches. The most urgent problem is lack of representation of social scientists other than economists. Other social scientists can provide valuable insight on issues of structural inequity, mental health, etc. Providing social scientists with consultation on how to develop an effective testimony backed by data is crucial.

**What are the minimum requirements for this project to provide value in solving the end user's problem? How much can be accomplished in 8 weeks?**

My goal for a minimum viable product would be creating a report with all my findings based on text analysis of past congressional testimonies and the main takeaways of this analysis. At the minimum, the project can showcase some interesting observations and common themes from exploratory analysis of testimonies given by social scientists. In 8 weeks, I could spend my time producing concrete suggestions that are tied to desired outcomes for those delivering speeches to legislative bodies. For example, looking more deeply at the nature of these hearings and investigating how testimonies influence the passage of bills or other pieces of legislation could provide insight into how researchers can provide convincing arguments to their audiences.

**What data would you ideally have for this project? What data can you reasonably access? How do you reconcile any differences between these sets of resources?**

I already have the dataset from Data is Plural on social science testimony that provides an interesting starting point for analysis. Ideally, I would have a large corpus of speeches available to analyze, but this has proven to be difficult to access. The ProQuest Congressional Documents site has no way currently of downloading transcripts in bulk and the process of searching for different speeches is a bit tedious. I have so far manually copy-and-pasted documents into text files, but I would like to explore a better, more efficient way of collecting data, perhaps through web scraping. I would also prefer having variables in the original dataset with additional pieces of information about the hearing, such as the bills that were proposed.

**What, if any, solutions already exist for this problem? What additional value does your solution provide?**

There are plenty of professional speech writers and experts in the field of writing congressional testimony. However, it is unclear to me how much “coaching” goes into preparing professors and researchers who otherwise do not have the opportunity to speak in front of a congressional audience often. This research can provide a great resource for those wanting to make their empirically sound arguments as effective as possible when delivering speeches to congress. This research into congressional testimony will be as in-depth and thorough as their own research in their field so they can rely on data-backed analysis to make their speeches truly shine.

**How do you plan to validate the success of this project beyond measuring model performance? For example, how if your model "improves accuracy by 10%", how would you contextualize the effect of this improvement to a nontechnical stakeholder?**

I am curious about the tangible effect that congressional testimonies could have on actual legislative change. My current research question is what factors of a social science testimony make the most impact in congressional hearings? By impact, I decided somewhat arbitrarily on a few factors: the total presence of social scientists at hearings, the ability to get bills passed during related hearings, as well as the ability to incite change in specific committees. The long-term effects of this project could be measured by observing future trends in social science congressional testimony, increasing diversity in those speaking to congress, and evaluating if expert opinion is taken into account when making legislative decisions.