

Assignment 1: Collecting and Structuring Data

Gard Olav Dietrichson

September 13, 2022

1 Formulate a hypothesis based on existing theories

The literature on political representation, and particularly that of women's representation make a distinction between what is known as descriptive and substantive representation (Wangnerud, 2009). While there are other types of representation discussed in political science (Pitkin, 1967), these are the ones that capture most of the active research currently. The distinction between these forms can be summed up as focusing on who legislators *are* and what they *do*. The highly active field is also ripe with the debate on the relation between these two, as some claim that descriptive representation is in fact a necessity in order to ensure substantive representation of minorities' interests (Phillips, 1995).

With my semester task I wish to test whether women representatives in the Norwegian parliament are the central actors that bring about the substantive representation of women, or whether they are simply in line with the party focus. To that end I wish to study the individual speech acts of members of the Norwegian parliament, and discuss how patterns of questions regarding women's issues, and issues that affect women, come about. Who asks these questions? When and Why? And how does the government constellation affect these patterns, given that around 90% of all questions are asked by opposition politicians?

My central assumption is that women will be more dedicated to holding the government accountable over women's issues. In addition I believe that this tendency could be augmented by party ideology and the responsible minister the question is directed to.

2 Find a datasource that could answer the hypothesis

Fortunately, these data are easy to find, as the storting has a publicly available API for usage. Even more fortunately, the eminent political scientist Dr. Martin Søyland has created a R-package for easier use of that API. I take advantage of those functions in order to retrieve all the questions that I can get my hands on. However, it should be noted that the API only has data back to 1996, leaving us somewhat wanting in terms of data in the early days of the storting, when a more male dominated storting was present, and women's issues were likely to be viewed with a more conservative bent. However, the data still covers a series of periods coinciding with a different set of government coalitions. It covers 9 different government coalitions, including Jagland, Bondevik I, Stoltenberg I, Bondevik II, Stoltenberg II, Solberg I, Solberg II, Solberg III, and Solberg IV, and Støre I. This includes then several periods of both socialist block government, and socialist block government, and will hopefully help illuminate some of the differences that party and incumbent government can have on MP's furtherance of women's issues in parliament.

3 Retrieve and Structure the Data

The full documentation for how the data was retrieved can be found in the submitted R-script. But the general outline of this phase of the project can be described as such

1. Collecting data on the sessions and their id numbers
2. Collecting the meta information on all questions in the sessions
3. Using this meta data to scrape the actual text of each question
4. Separately, the relevant MP's for the sessions must be found
5. When found, their information must be joined with the dataframe containing all the questions.
6. At this point we have our structured Data.

The full processes of this can be viewed in the R-script, but a fair warning, some of the scraping procedures take an estimated 14 hours to run.

4 Give a brief description of how the data was captured and how they were structured

The data ended up being a large assortment of questions asked in the Norwegian parliament. In fact, every question asked in the past 26 years is included. It bears mentioning that most of these questions are not relevant to the research question itself, but will be useful in order to construct structural topic modelling, and conduct secondary research into what sort of parliamentarians are most attentive towards women's issues.

The data is for now structured into rows of individual observations of questions. This means that a single row is an asked question, along with a justification for the question, and a response from the relevant minister. There is also a discussion to be had in regards to separating the answer from the question, but given that they are, almost by definition, probably the same topic, I see little reason to do so, but would appreciate feedback on this decision. In addition using the gender of the relevant minister could also be an interesting explanatory variable, to see if women ministers encourage other women to speak up on women's issues relating to that department, such as recently done in Blumenau (2021).

As for the challenges that I faced in gathering the information, the packages `stortingscrape` was very useful, and the fact that I have extensive experience using it was also helpful. However, owing to a need to give the storing's database some rest, I made sure to include good manners integer, to set a pause between each scrape. While the database engineers might thank me for that, it also leads to the script requiring 1sec x 51 000 observations to scrape all questions, taking roughly 14-15 hours.

5 Bibliography

Blumenau, J. (2019). The Effects of Female Leadership on Women's Voice in Political Debate. *British Journal of Political Science*, 51(2), pp. 750-771. doi:10.1017/S0007123419000334

Phillips, A. (1995). *The Politics of Presence*. Oxford: Clarendon Press

Pitkin, H. (1967). *The Concept of Representation*. Berkeley: University of California Press