

# Are women underrepresented in Scottish Politics?

Alanah Sarginson, Noémie Clare, Len Metson

Our final product

1. [Tableau Public story - final product](#)
2. [Tableau Public summary dashboard](#)

Our code

1. [GitHub repository](#)

## Executive Summary

In this project we analysed whether women are underrepresented in Scottish politics. Most studies of representation assume that representation is about the number of women there are in Parliament (descriptive representation). By this view, representation is achieved when the number of women in parliament is proportional to the number of women in the population. In this project we extended what is meant by representation. As well as the pure number of women there are in parliament, we were interested in whether women get the same chance to speak in Parliamentary debates.

Our results confirmed that women are underrepresented in Scottish Parliament. Across the history of the Scottish Parliament, 34% of MSPs were women. Across the same period, 29% of speeches were made by women. The proportion of speeches by women increased over time. When excluding MSPs with positions, such as ministers, women made up 36% of MSPs and made 31% of speeches. Therefore, the overall results are similar whether we include ministers or not. The only difference is that, when excluding for MSPs with more senior positions, there is no increase over time in the ratio of proportion of speeches by women to proportion of women there are. This suggests that the improvement in women's representation over time has been driven primarily by more ministers being women. These results highlight the importance of increasing both the overall number of women MSPs as well as ensuring that women are able to contribute in ministerial positions. In our data visualisations we show how gender interacts with age and geographical variables in terms of speech time.

Representation can also be defined by what the government does for women (substantive representation). To address this perspective, we analysed how much men and women in parliament use their speeches to talk about women's issues. Across the whole dataset, 5.48% of speeches by women were about women's issues but only 2.32% by men were about women's issues. This shows that having more women in parliament actually improves how well women are substantively represented in parliament.

## Motivation for our project

We wanted to use the data science tools we have learned about to quantify the extent to which women are represented in parliamentary politics. The case we looked at was the Scottish Parliament.

There are three aspects to representation that we explored:

1. Whether women have an equal amount of MSPs to men
2. Whether women in parliament speak as much as men.
3. Whether women in parliament use their position to represent women's issues

## Justification for our project

- Fourth-wave feminism's advocacy for greater representation of women in politics is premised on the notion that society would be more equitable if policies incorporate perspectives of all people
- Inspired by this argument for the equitable representation of gendered perspectives in political settings, we were interested in exploring women's representation in Scottish parliament across time to see the progress Scotland has made in not only including more women in parliamentary debates, but by including more contributions by women to the debates themselves
- Contextualising our data:
  - ["Scottish election 2021: Record number of women elected" - BBC](#)
  - ["The Scottish Parliament's record on women's representation is in the balance" - Democratic Audit](#)
  - ["Backwards for gender equality in the new Scottish parliament? Or a new Scottish velvet triangle?" - Center for Constitutional Change](#)

## Aim

We had three aims:

1. Use parliamentary transcripts to quantify how much women and men speak and relate our findings to patterns in other demographic variables such as age and region.
2. Run a content analysis to see whether gender correlates with whether MSPs talk about women's issues.
3. Visualise the data we find to tell a story about women's representation in the Scottish Parliament

## Data

We used a pre-scraped dataset of speeches by Scottish Members of Parliament (MSPs). The dataset was downloaded by HTTP request using an R wrapper. The data has entries from the opening of the Scottish Parliament so we were able to analyse all speeches made in the Scottish Parliament.

[Link to Harvard Dataverse Page](#)

### Descriptive statistics on Raw Data

Total	Female	% Female
-------	--------	----------

	Total	Female	% Female
Speeches	430,577	125,078	29.05%
MSPs	306	105	34.31%

Additional data on age was acquired from WikiData. Using python, code was written to connect to WikiData's API and pull each MSPs birthdate.

## Methodology

We made some key assumptions that had to be checked. Whilst the aim of our project was not to make inferential conclusions, it is important to discuss potential limitations in our data.

Why syllables:

- The assumption is that the more syllables a speech is, the more time it takes to say.
- Having time on the floor of parliament is a precious resource for politicians. They use it to signal to voters, bring policies forward and influence legislation.
- We used it instead of raw number of speeches because syllables are more accurately aligned with units of time (i.e., roughly people can say 5 syllables per second) compared to words / speeches, and we assumed that many short speeches are less impactful than an MSP given space to make longer speeches. Length of speech is also defined by an MSP's relevance to a piece of legislation or debate, so MSPs who give longer speeches are the most influential. Therefore, we choose it as a more sensitive proxy for measuring influence, parliamentary power, and representation.
- For purposes of validation, we did however, replicate all analysis using number of speeches and found similar results.

Ministers and Sturgeon:

- Ministers are likely to have more fixed opportunities to speak in Parliament. For example, they are required to make speeches in answer to speeches.
- The election of Nicola Sturgeon, a woman, as First Minister meant that she was speaking for a very significant amount of time. In the most recent parliamentary session, Sturgeon accounted for almost 10% of all speeches.
- We ran the analysis with no ministers and found there was no upward trend in the amount that women spoke. Therefore, the improving picture we saw is most likely due to women entering more ministerial positions.

## Steps:

### 1. Data collection and cleaning

- We collected data from the [Harvard Dataverse](#). The dataset has been scraped from HTML files containing transcripts of parliamentary speech published on the Scottish Parliament Website.
- Please refer to `r-analysis/clean-data.R`, `r-analysis/y-year.R`, `python-analysis/format_dataset.py`, `python-analysis/wikipedia_infobox_scrape.py`, `python-analysis/wikidata_page_scrape.py`, `python-analysis/prelim-analysis-jupyter.py`.

### 2. Syllable analysis

- Our main analysis was how much time women spoke in the Scottish Parliament.
- We decided to explore using syllable count instead of number of speeches because we thought it relevant the *time* that women get to speak. Ultimately the results were similar but it was good to establish syllables as a way of measuring speech time.
- Please refer to `r-analysis/syl-count.R`, `python-analysis/syllable_count_by_speech.py`.

### 3. Content analysis

- As well as analysing how much women speak, we wanted to look at what men and women talk *about*.
- We created a dictionary with words associated with women's rights issues.
- To include this in our analysis, we based our algorithm on the NLP workshop given in DS105, and included an "if then" look to increase the score if more than one word from the dictionary is included in the speech.
- The resulting speech scores, which are essentially weighted averages of dictionary words that are in the speeches, adjusted for the increased correlation from the presence of multiple words, are stored in the original data frame as an additional series.
- We also reiterated the analysis in R with a slightly different methodology: we used boolean indexing instead of weights, and this allowed us to detect pairs of words such as "sexual violence" instead of individually flagging them, and separately accounting for the presence of both of them within the speech.
- Please refer to `r-analysis/content-analysis.R`, `r-analysis/agg-speech-count.R`, `python-analysis/speech_content_analysis.py`.

### 4. Visualisation

- We used Tableau to visualise the data once we had them all in CSV form, to eventually build the final product as a narrative of how women's representation in Scottish parliament has evolved throughout time
- This also involved adjusting certain CSVs on the local machine for certain visual results; this was done intentionally to produce quicker results in data structure, which worked with the fact that the connections made from Tableau were also public. These CSVs are uploaded to the `output_data` folder
- Here, we used the worksheets feature for each visualisation, the dashboards feature for the summary dashboard and each page of the story, and then finally the story feature for the final product, where we collated everything built to construct the narrative of the project, ultimately forming our final product

## Results

- We found some interesting descriptive results about the role of women in Scottish Parliamentary proceedings.
- Looking at the number of seats taken up, the average number of seats taken up by men across the history of the Scottish Parliament is 86, whereas this number is only 48 for women. Furthermore, there has been little variation on both sides of these means, suggesting little progress in getting more women taking up seats.
- Initially, female MSPs spoke disproportionately less than there were women in the Scottish Parliament. However, over time, this rose so that women were speaking roughly as much as there were women in Parliament (i.e., the ratio of speech by women to amount of women taking up seats has increased towards 1:1). That being said, much of this trend has been driven by women entering more senior positions.
- We also found there to be an upwards trend for the amount of speech by both genders which had mentions of women's issues in it. This growth has been steadily increasing since 2012, suggesting that for the past decade, there have been important and sustaining shifts in the structure of Scottish Politics that

have driven conversation around women's issues. With that being said, women have been the significant drivers of this change. For example, in 2019, when conversation around women's issues were at its peak, the proportion of this conversation by women were 64%, opposed to just 36% of men speaking about women's issues.

## Conclusion

---

Conclusions:

- In this project we leveraged skills taught in DS105: using GitHub, using the Terminal, interacting with APIs - https protocol, SQL and SPARQL - and cleaning, analysing and visualizing structured data. We adapted NLP methodology taught through DS105 workshops to support our exploration of the content of speeches. These tools enabled us to look for markers of gender discrimination through three different prisms.
- Firstly, we found that disproportionately less women are elected to the Scottish Parliament. There has been little change in this since 1999. Thus our analysis confirmed a well-known fact that there exists a persisting gender disparity in parliament in terms of descriptive representation.
- Secondly, we explored whether women who are in parliament speak more or less than men. We found that in the first and second parliament, that women's speaking time was disproportionately lower than the amount of women MSPs there were. This shows that women are underrepresented in terms of the time that women are speaking in Parliament. Not only did women have disproportionately less parliamentary seats than women in the population, this gap was aggravated the gap in the amount they speak.
- A positive takeaway is that this gap decreased over time. Women are now speaking close to proportionally to the amount of seats they have. However, because women still do not make up 50% of seats, there is still not equal representation.
- Finally, we looked at whether women talked more about women's rights issues than men. Analysis using NLP revealed that, despite the perceptible trend towards more equal involvement of men and women in gender issues, a higher proportion of speeches that mention women's issues are delivered by women. This allows us to conclude that there exists a substantive gender representation gap: the onus of fighting for women's rights issues is born more by women than by men. This tend also shows that that descriptive representation (the amount of women and how much they contribute) is highly related to substantive representation in this context (what the parliament does for women).
- We discuss further conclusions as well as opportunities of further research and brief recommendations for women's representation in parliament in the last slide of our [Tableau Public](#) story.

## Appendix

---

1. [Harvard Dataverse](#)
2. [Data.parliament.scot](#)
3. [Parliamentary Seat Map Maker](#)
4. [Tableau Public summary viz dashboard](#)
5. [Tableau Public story - final product](#)
6. [Our GitHub repository](#)