My title*

My subtitle if needed

First author   Another author

April 10, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

2 Data

This paper uses the ParlEE plenary speeches data set to conduct its analysis. This dataset specifically has been obtained from version 1, and uses the UK speech data. The dataset has information on speeches from 2009 to 2019.

For their data collection, the authors chose the government body bearing the most responsibility for legislation. For the United Kingdom they chose the House of Commons. As a starting point, the authors used the The Rauh et al ParlSpeech dataset set since it already contained a large collection of raw speech data for many countries of interest, especially the UK. There was some data missing towards the end of the timeframe of interest (towards 2019), so the authors went through the speeches from (https://hansard.parliament.uk/commons) and filled in the missing parts themselves using an API provided to them.

The authors ironed out typical wrinkles with government related datasets - removing boilerplate information, page numbers, eliminating corrupted social characters and introducing consistent naming conventions for party and speaker alike. This made the data uniform across countries, improving readability.

---

*Code and data are available at: LINK.

The dataset has variables explained below :

- **Date** : The date on which a specific sentence from the speech has been said is recorded in the 'date' variable. This is recorded in the 'DD/MM/YYYY' format, with duplicated values for multiple sentences spoken in the same speech as well as if there were multiple speeches on the same day.

- **Agenda** : The variable 'agenda' is the title given to a set of speeches given pertaining to the issue at hand in the House of Commons at the time. This is provided by official sources.

- **Speechnumber** : The variable 'speechnumber' gives a unique identifier to all the sentences spoken in the same speech. So if the first speech in the data set has 5 recorded sentences, they will all be allotted the number 1.

- **Sentencenumber** : The variable 'sentencenumber' assigns a unique identifier to every sentence spoken within the same speech. Continuing the example above, every sentence in speech number 1 is allotted a sentence number 1 through 5. This variable resets from speech to speech, such that the first sentence in every speech will always have number 1.

- **Speaker** : The variable 'speaker' indicates the name of the member making the speech. Measures have been taken to eliminate inconsistencies, such as removing any nicknames, variations in spelling or honorary titles and using only one standard naming convention. The only exception is the usage of ministerial titles in the speaker column, since it is worth indicating when the same speaker is speaking as 'the Minister' versus just an individual. This is because that being in that position may lead to influence in the opinion of the speaker as demanded by the given title.

- **Party** : The 'party' variable records the alignment of every speaker. The parties recorded are APNI, Conservative Party, DUP, GPEW, Labour Party, LibDem, other, Plaid Cymru, SDLP, SNP, UKIP and UUP.

- **Text** : The 'text' variable records the raw text of sentences part of a speech given in parliament. This has already been cleaned as stated previously, and therefore is ready to be studied across the other variables like party, agenda, etc.

- **Parliament**The 'parliament' variable stores the name of the legislative body. For the UK and for the purposes of this paper, this variable is 'UK- HouseofCommons' everywhere.

- **iso3country** : Similarly, the variable 'iso3country' records the country in question, which for this paper is the United Kingdom throughout, recorded as 'GBR'.

The authors augmented the collected data so as to make it enable analysis beyond what was already possible. These variables are described below:

- **EU** : The variable 'eu' is a dummy variable introduced by the authors that determines whether the given text or sentence discusses anything related to the European Union. The dummy is '1' if it does discuss the European Union and '0' if it doesn't.

- **Policyarea** The variable 'policyarea' assigns a number in accordance with the specific policy topic of a text. The numbers used are from the CAP major categories coding scheme.

The CAP coding scheme is shown in the table below :

The following graph shows the occurrence of the given policy types over the years:
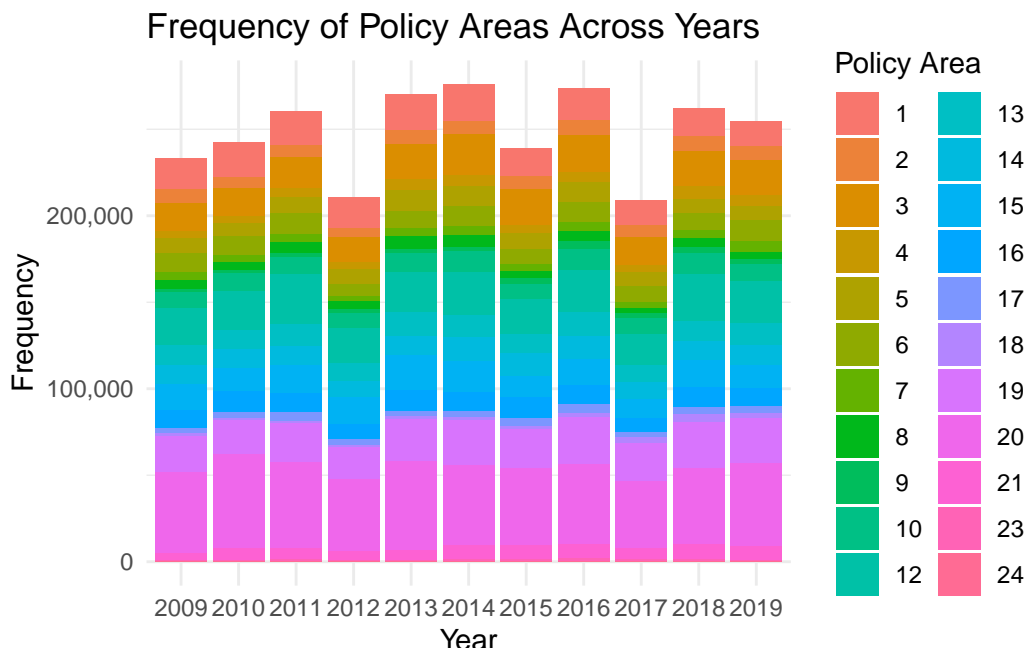
Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table 1.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Table 1: Explanatory models of flight time based on wing width and wing length

	First model
(Intercept)	1.12
	(1.70)
length	0.01
	(0.01)
width	−0.01
	(0.02)
Num.Obs.	19
R2	0.320
R2 Adj.	0.019
Log.Lik.	−18.128
ELPD	−21.6
ELPD s.e.	2.1
LOOIC	43.2
LOOIC s.e.	4.3
WAIC	42.7
RMSE	0.60

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

B.2 Diagnostics

Figure 1a is a trace plot. It shows... This suggests...

Figure 1b is a Rhat plot. It shows... This suggests...
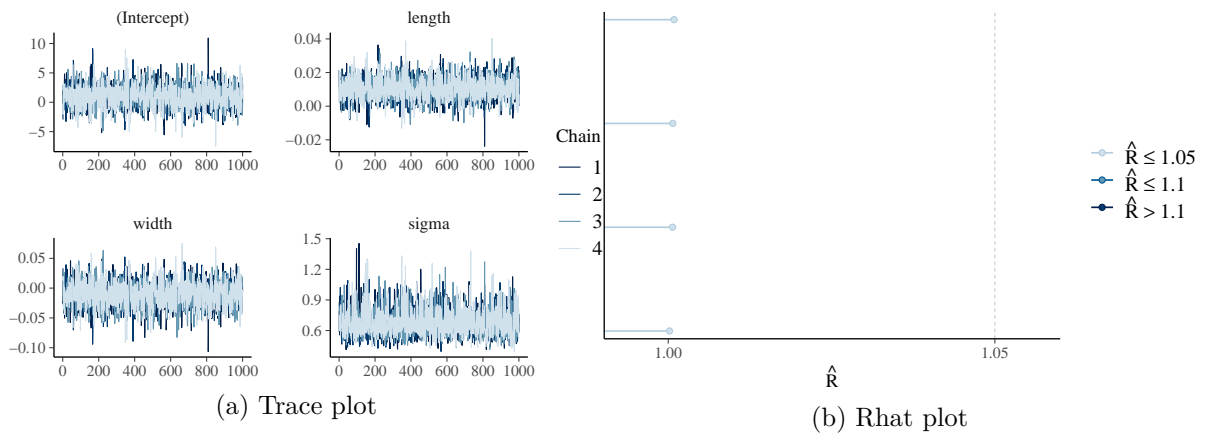

(a) Trace plot

(b) Rhat plot

Figure 1: Checking the convergence of the MCMC algorithm

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.