

# Assignment 4

Steve Kerr (211924)

2023-10-23

## Introduction

In this assignment, you are asked to use topic modelling to investigate manifestos from the manifesto project maintained by [WZB](#). You can either use the UK manifestos we looked at together in class, or collect your own set of manifestos by choosing the country/countries, year/years and party/parties you are interested in. You should produce a report which includes your code, that addresses the following aspects of creating a topic model, making sure to answer the questions below.

This time, you will be assessed not only on whether the code gets the right result, but on how you understand and communicate your understanding of the modelling process and how this can answer your research question. The best research question is one that is interesting and answerable, but the most important thing is that the research question is answerable with the methods you choose.

You will also be assessed on the presentation of your results, and on the concision and readability of your code.

```
# load packages
pacman::p_load(
  here,
  lubridate,
  manifestoR,
  readr
)
```

## 1. Data acquisition, description, and preparation

Bring together a dataset from the WZB.

```
# define file path for data
data_path <- here("Assignments_SK/Assignment-4/data/corpus.rds")

# download data if not yet downloaded
if (!file.exists(data_path)) {

  # set API key
  mp_setapikey(here("Assignments_SK/Assignment-4/manifesto_apikey.txt"))

  # query data
  corpus <- mp_corpus(countryname == "United States" & edate > as_date("2000-01-01"))

  # save data
```

```

saveRDS(corpus, here("Assignments_SK/Assignment-4/data/corpus.rds"))

} else {

  # load data
  corpus <- read_rds(data_path)

}

# inspect corpus
corpus

```

```

## <<ManifestoCorpus>>
## Metadata:  corpus specific: 0, document level (indexed): 0
## Content:   documents: 12

```

```

# explore data
head(content(corpus[[1]])) # view beginning of text of first manifesto

```

```

## [1] "The 2000 Democratic National Platform: Prosperity, Progress, and Peace as adopted by the 2000

```

```

table(codes(corpus)) # count codes of all manifestos

```

```

##
##   000   101   102   103 103.1 103.2   104   105   106   107   108   109   110
##   218   319    66     6     4     4  927   203   132  695    4   196    2
##   201 201.1 201.2   202 202.1 202.2 202.3   203   204   301   302   303   304
##   180   310   194   178   302     4    74   187    11  281    5  255    94
##   305 305.1 305.2 305.3   401   402   403   404   405   406   407   408   409
##   364   101    86     2  638   403   434    69    21    67    81   150    7
##   410   411   412   413   414   415   416 416.2   501   502   503   504   505
##   285   640    20     6  132     1    33    63   412    37  967   832   146
##   506   507   601 601.1 601.2 602.1 602.2   603   604   605 605.1 605.2   606
##   505    48   206   169    58    31   107   698    62  490   229   134    39
## 606.1 606.2   607 607.1 607.2 607.3   608 608.1 608.2   701   702   703 703.1
##     50     8    56    32    12  152    14     2     5  490    61    83   163
## 703.2   704   705   706     H
##     2     80    92   289   380

```

What years, countries and parties are included in the dataset? How many texts do you have for each of these?

Prepare your data for topic modelling by creating a document feature matrix. Describe the choices you make here, and comment on how these might affect your final result.

## 2. Research question

Describe a research question you want to explore with topic modelling. Comment on how answerable this is with the methods and data at your disposal.

### 3. Topic model development

Create a topic model using your data. Explain to a non-specialist what the topic model does. Comment on the choices you make here in terms of hyperparameter selection and model choice. How might these affect your results and the ability to answer your research question?

### 4. Topic model description

Describe the topic model. What topics does it contain? How are these distributed across the data?

### 5. Answering your research question

Use your topic model to answer your research question by showing plots or statistical results. Discuss the implications of what you find, and any limitations inherent in your approach. Discuss how the work could be improved upon in future research.

## Sources

- Lehmann, Pola / Franzmann, Simon / Burst, Tobias / Regel, Sven / Riethmüller, Felicia / Volken, Andrea / Weßels, Bernhard / Zehnter, Lisa (2023): The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2023a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem). <https://doi.org/10.25522/manifesto.mpds.2023a>

## Resources

- [manifestoR vignette](#)