# Exercise 4

Min Wang

2024-03-05

# Load packages and import data

First, load the packages and import the data.

```
library(kableExtra)
library(dplyr)
```

```
##
## 载入程辑包：'dplyr'
```

```
## The following object is masked from 'package:kableExtra':
##
##     group_rows
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(quanteda) # includes functions to implement Lexicoder
```

```
## Warning: 程辑包'quanteda'是用R版本4.3.2 来建造的
```

```
## Warning in .recacheSubclasses(def@className, def, env):
## "replValueSp"类别的子类别"ndiMatrix"没有定义；因此没有更新
```

```
## Warning in stringi::stri_info(): Your current locale is not in the list of
## available locales. Some functions may not work properly. Refer to
## stri_locale_list() for more details on known locale specifiers.

## Warning in stringi::stri_info(): Your current locale is not in the list of
## available locales. Some functions may not work properly. Refer to
## stri_locale_list() for more details on known locale specifiers.
```

```
## Package version: 3.3.1
## Unicode version: 13.0
## ICU version: 69.1
```

```
## Parallel computing: 16 of 16 threads used.
```

```
## See https://quanteda.io for tutorials and examples.
```

```
library(quanteda.textmodels) # for estimating similarity and complexity measures
```

```
## Warning: 程辑包'quanteda.textmodels'是用R版本4.3.3 来建造的
```

```
library(quanteda.textplots) #for visualizing text modelling results
```

```
## Warning: 程辑包'quanteda.textplots'是用R版本4.3.3 来建造的
```

```
kaneko_dfm  <- readRDS(gzcon(url("https://github.com/cjbarrie/CTA-ED/blob/main/data/wordscalin
g/study1_kaneko.rds?raw=true")))
```

# Estimate wordfish model

Since it is already a document feature matrix, we are able to group and trim the document feature matrix before estimating the wordfish model.

```
table(docvars(kaneko_dfm, "Newspaper"))
```

```
##
##       Asahi      Chugoku     Chunichi     Hokkaido      Kahoku     Mainichi
##          38           24           47           46          18           26
##      Nikkei Nishinippon       Sankei      Yomiuri
##          13           27           14           30
```

```
## prepare the newspaper-level document-feature matrix
# compress the document-feature matrix at the newspaper level
kaneko_dfm_study1 <- dfm_group(kaneko_dfm, groups = Newspaper)
# remove words not used by two or more newspapers
kaneko_dfm_study1 <- dfm_trim(kaneko_dfm_study1, min_docfreq = 2, docfreq_type = "count")

## size of the document-feature matrix
dim(kaneko_dfm_study1)
```

```
## [1]   10 4660
```

Now the Wordfish model can be estimated, using "summary" function to see the results.

```
#### estimate the Wordfish model ####
set.seed(123L)
kaneko_dfm_study1_results <- textmodel_wordfish(kaneko_dfm_study1,
                                                sparse = TRUE)

summary(kaneko_dfm_study1_results)
```

```
##
## Call:
## textmodel_wordfish.dfm(x = kaneko_dfm_study1, sparse = TRUE)
##
## Estimated Document Positions:
##                   theta      se
## Asahi         -1.01534 0.01966
## Chugoku       -0.40012 0.02727
## Chunichi      -1.39207 0.01450
## Hokkaido      -0.46705 0.02093
## Kahoku        -0.38169 0.03194
## Mainichi      -0.06059 0.02770
## Nikkei         1.26617 0.04258
## Nishinippon   -0.29256 0.02949
## Sankei         1.21020 0.03855
## Yomiuri        1.53306 0.02417
##
## Estimated Feature Scores:
##            安保     法制     国会     一線     越え 安倍内閣    新た 安全保障政策
## beta 0.005043 -0.2277 -0.1162 -0.1951 -1.041  -1.6489 0.1173      -0.1650
## psi  3.655225  3.3082  3.6948 -0.1608 -1.612   0.4952 2.1727       0.7975
##        関連法案     閣議     決定   提出 安倍首相     先月       米     議会     演説
## beta   -0.2154 -0.3584 -0.3934 0.323  0.06609 -0.6974 0.3167 -0.9097 -0.37893
## psi     2.3778  0.7088  1.2334 2.491  2.55716 -0.4438 2.8074  0.6683 -0.06753
##        安全保障     戦後   初めて 大改革       夏     成就     約束     通り     合意
## beta 0.0001012 -0.4946 -0.1843 -1.041 -0.7545 -1.859 -0.7678 -0.5569 0.2582
## psi  3.0028167  2.0520  0.9314 -1.612  0.7277 -1.983  0.1928  0.4735 1.8620
##          歴史的     転換 集団的自衛権     行使
## beta -0.8812 0.03341     -0.02387 -0.1965
## psi  -1.7698 1.62121      3.74762  3.6801
```
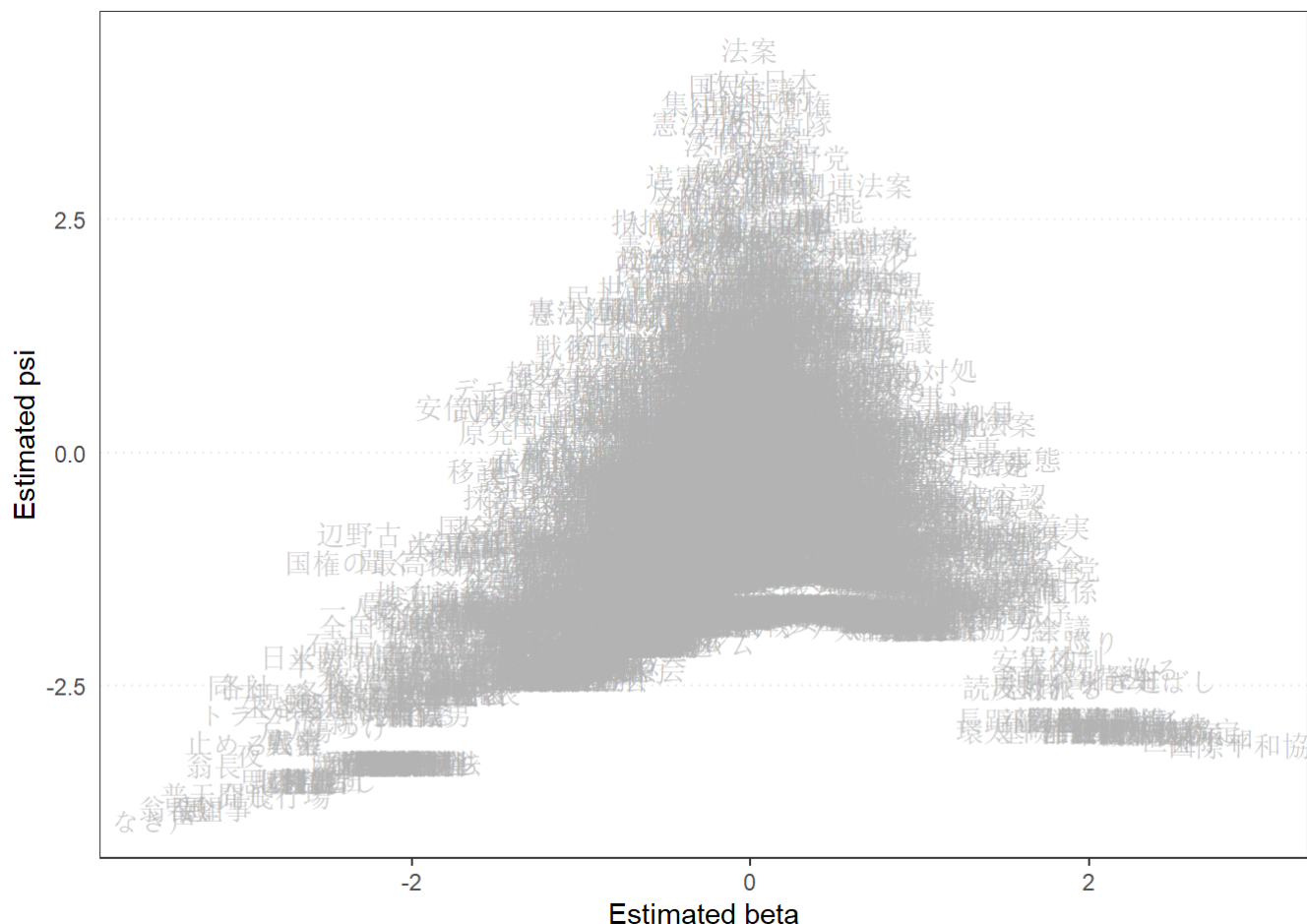
# Visualize the results

We can then plot our estimates of the $\theta$s.

```
textplot_scale1d(kaneko_dfm_study1_results)
```
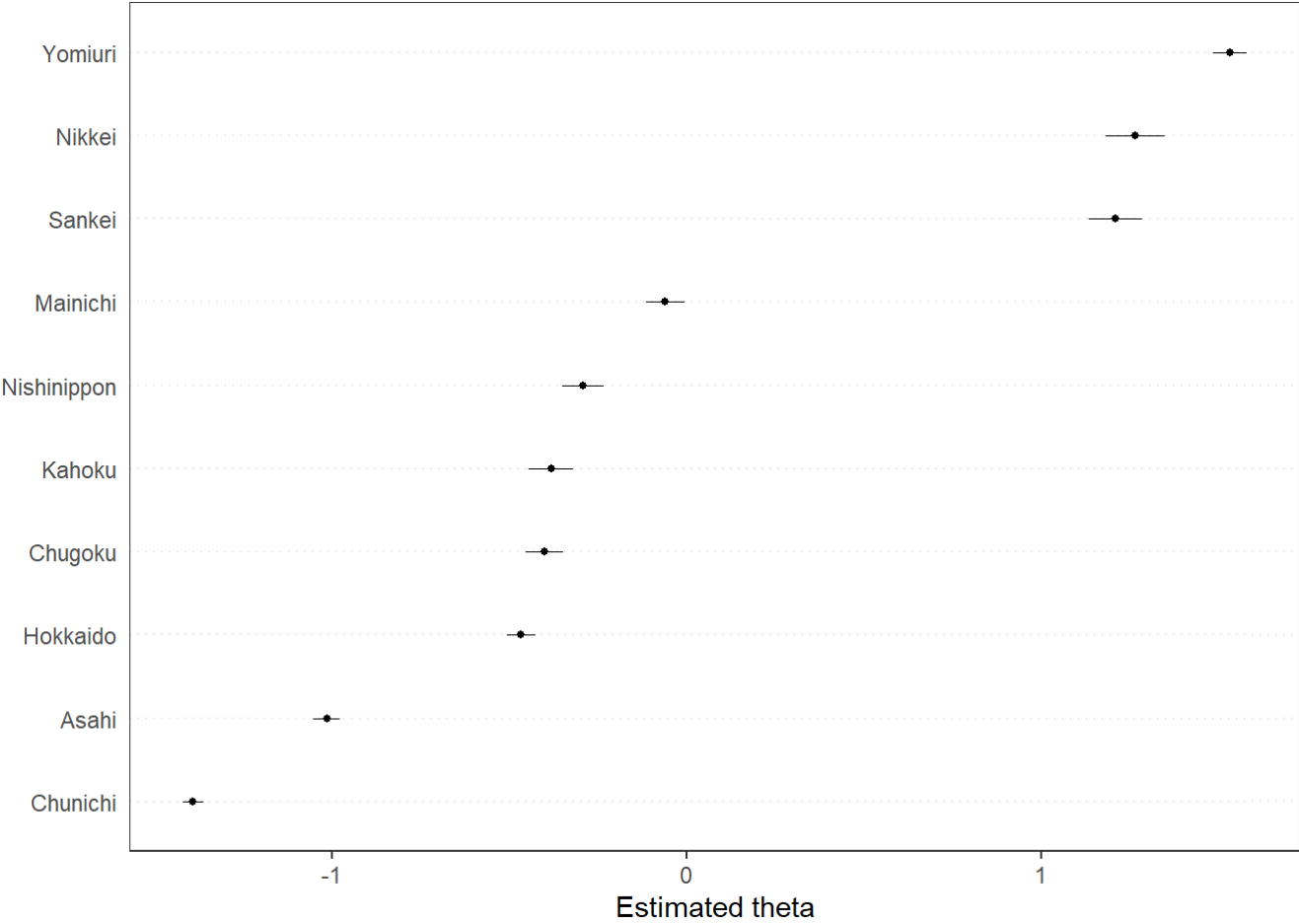
Plotting the "features".

```
textplot_scale1d(kaneko_dfm_study1_results, margin = "features")
```

we can also look at these features.

```
# extract "features" from a data structure("kaneko_dfm_study1_results") and assigns its content
s to a new variable named "features1"
features1 <- kaneko_dfm_study1_results[["features"]]

# the same process as "features1"
betas1 <- kaneko_dfm_study1_results[["beta"]]

# merge "features1" and "betas1" into a new data frame called feat_betas1
feat_betas1 <- as.data.frame(cbind(features1, betas1))

# set the type of "betas1" to numeric
feat_betas1$betas1 <- as.numeric(feat_betas1$betas1)


feat_betas1 %>%
  arrange(desc(betas1)) %>%  # sort by the value of "betas1" in descending order
  top_n(20) %>%  # select the first 20 rows after sorting
  kbl() %>%    # change data frame to a table
  kable_styling(bootstrap_options = "striped") # adds a stripe style that alternates the backgr
ound color of consecutive rows
```

```
## Selecting by betas1
```

| features1 | betas1 |
| --- | --- |
| 国際平和協力 | 2.967613 |
| 極小 | 2.633119 |
| 世界平和 | 2.633119 |
| 勘案 | 2.537293 |
| 巡る | 2.374824 |
| 心がけ | 2.371728 |
| 気がかり | 2.371728 |
| 乱れ | 2.371728 |
| 享受 | 2.371728 |
| 中国軍 | 2.371728 |
| 世界秩序 | 2.371728 |
| 取り組め | 2.371728 |
| 問責決議案 | 2.371728 |
| 引き延ばし | 2.344777 |
| 農協 | 2.283231 |
| 横行 | 2.283231 |
| 収める | 2.283231 |
| 除き | 2.283231 |
| 合戦 | 2.050239 |
| 合区 | 2.050239 |
| 演出 | 2.050239 |
| 具体策 | 2.050239 |
| 騒動 | 2.050239 |
| 離合集散 | 2.050239 |
| 遅滞 | 2.050239 |
| 変調 | 2.050239 |
| 環太平洋経済連携協定 | 2.050239 |

| features1 | betas1 |
|---|---:|
| 活性 | 2.050239 |
| 消費税率 | 2.050239 |
| 基礎的財政収支 | 2.050239 |
| 黒字 | 2.050239 |
| 佐藤正久 | 2.050239 |

| features1 | betas1 |
|---|---:|
| 活性 | 2.050239 |
| 消費税率 | 2.050239 |