

ISSSV-1337, Group 3 Report UN Association of Norway

Gard Olav Dietrichson, Hannes Brauer, Markus Opheim, Andreas Kroknes, Ingrid Johannesen, Nora Di

2022-07-25

Given Objective

The task provided to us, was to provide the UNA with “quantitative measures” of their work. The wish from our mission provider was to gain knowledge about how much the members of parliament are aware of the activities of the UN

Problem

The central problem at hand is how to measure the impact of lobbying, as direct contact between politicians is difficult to find and to measure.

Our Approach

In order to map the knowledge of the activities of the UNA, we decided to collect all available texts from Stortinget. After gathering these texts, we needed to figure out how to measure relative prevalence in our data. To this we first decided to create a list over relevant key terms that represent the work of the UNA. We divided these into three levels, based on the suggestions provided to us by our mission provider. The top level search word is “FN-sambandet” the UNA in Norwegian. The first thing that struck us was the fact that this yielded very few results, and as such we saw it necessary to focus much of our work outside what UNA has been explicitly mentioned in.

Our mission description also contained more suggestions for keywords, and so we used those to create level 2 and level 3 searchwords. Level 2 consists of a particular goal that our provider explained that they had been working towards. This goal was UN sustainability goal 4.7 Education for Sustainable Development. The keywords here were also used to investigate, but again returned very few results. The problems of this will be elaborated on in a later section.

The final level of search words was the relevant humanitarian sub organisations of UN, such as UNAIDS, WFP, UNDP and UNICEF. These ended up forming the backbone of our approach, as they were far more prevalent, however, all words were maintained in the end, when a complete analysis was performed.

Finally it is worth mentioning that we also foresaw the large amount of documents being being problematic, as such, we early on decided to limit our timeframe. The package stortingsscape informs us that voting data is only available from the 2011-2012 session and onwards, so we viewed that as a natural starting point.

Joint setup code and used packages

Below we include a setup code for the things we do in this script, this primarily includes the relevant packages, but also included is a function for a colourblind safe palette, given that we are interested in accessibility.

```
## -- Attaching packages ----- tidyverse 1.3.1 --  
## v ggplot2 3.3.6      v purrr   0.3.4  
## v tibble  3.1.7      v dplyr  1.0.9
```

```
## v tidyr 1.2.0 v stringr 1.4.0
## v readr 2.1.2 v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

## Package version: 3.2.1
## Unicode version: 13.0
## ICU version: 69.1

## Parallel computing: 8 of 8 threads used.

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

## stm v1.3.6 successfully loaded. See ?stm for help.
## Papers, resources, and other materials at structuraltopicmodel.com

## Loading required package: viridisLite

##
## Attaching package: 'rvest'

## The following object is masked from 'package:readr':
##
## guess_encoding

## [1] "LC_COLLATE=Norwegian Bokmål_Norway.utf8;LC_CTYPE=Norwegian Bokmål_Norway.utf8;LC_MONETARY=Norwegian Bokmål_Norway.utf8"
```

Sectioning

Due to the large corpus that would arise from joining all documents, and the severe lack of coherent genre textuality, and to the unique opportunities afforded by each form of retrievable information, we decided to split up our work, and pursue separate roads of inquiry based on each section.

Questions

The questions were scraped, below is the plot drawing the overview of the spreading of mentions of our relevant keywords over the relevant timeframe.

The questions we ask of the questions

- Who are asking questions about the UN and the UNA?
- What are these questions about?
- What questions mentions the UNA and the associated UN organizations?
- How are these talked about?
- Do these questions touch on issues regarding sustainability and other topics of interest that the UNA might be interested in?
- Other interesting patterns in the questions

In order to answer this, we scrape all the questions from the storing's homepage. The following code scrapes all the questions for our relevant time frame.

```
## For now the sessions we are interested in are limited to 2011-2022

sessions_storting <- get_parlsessions()

sessions_storting <- sessions_storting %>%
  filter(id %in% c("2011-2012", "2012-2013", "2013-2014",
```

```

"2014-2015", "2015-2016", "2016-2017",
"2017-2018", "2018-2019", "2019-2020",
"2020-2021", "2021-2022"))

```

This section retrieves all the sessions that we are interested in, the object `sessions_starting` will be used for most of our code chunks later.

The next step is then to retrieve the metadata on all the questions.

```

## The id's are then entered into this loop, which *should* retrieve
## ALL question id's for our relevant period.

a<-list()
b<-list()
c<-list()

for(x in unique(sessions_starting$id)){
  a[[x]] <- get_session_questions(sessionid = x, q_type = "interpellasjoner", status = NA, good_manners
  b[[x]] <- get_session_questions(sessionid = x, q_type = "sporretimesporsmal", status = NA, good_manne
  c[[x]] <- get_session_questions(sessionid = x, q_type = "skriftligesporsmal", status = NA, good_manne

  #paste0("stortingsporsmal", x) = rbind(a, b, c)
}

## Unlisting all of the resulting lists, they are lists of identically
## sized dataframes.

questionlista <- do.call("rbind", a)
questionlistb <- do.call("rbind", b)
questionlistc <- do.call("rbind", c)

clist <- list(questionlista, questionlistb, questionlistc)

## Combining all for a dataframe we can fetch

questionlist <- do.call("rbind", clist)

```

This has created our full list of questions, but not the text of the question and answer themselves. The next step is retrieving all of those.

```

d <- list()

for(x in unique(questionlist$id)){
  it <- 100*(which(unique(questionlist$id) == x) / length(unique(questionlist$id)))
  cat(paste0(sprintf("Progress: %.4f%%", it), "\r"))

  d[[x]] <- get_question(questionid = x, good_manners = 0)
  #paste0("stortingsporsmal", x) = rbind(a, b, c)
}

questiontext <- do.call("rbind", d)

save(questionlist, file = "Question_Data/MetadataQuestionList.Rdata")
save(questiontext, file = "Question_Data/All_Questions.Rdata")

```

This now gives us a dataframe for analysis. After this we approach the data from two ways, we create a structural topic model, and we use this to analyse the totality of the questions dataframe.

The structural topic model is created with the code below. First we have to include the answers as well, as part of the documents.

```
answerlist <- questiontext %>%
  select(id, text = answer_text, title, type, question_from_id,
         qustion_to_id)

## Dropping the observations with no text, note that no text in
## Stortingscraps is coded as empty string, not as an NA

answerlist <- answerlist %>%
  mutate(notapplicable = ifelse(text == "", NA, 1)) %>%
  drop_na(notapplicable) %>%
  select(-notapplicable)

## Limiting the dataset

questionshort <- questiontext %>%
  select(id, text = question_text, title, type, question_from_id,
         qustion_to_id)

## Binding the sets together, so I have one dataframe with complete
## texts from either answers or questions. **OF NOTE** I should probably
## have added a string to the id's so that i separate answer texts
## from question texts, given their identical id.

assembledtextquestions <- rbind(questionshort, answerlist)

## removing reduntant objects

rm(answerlist, questionlist, questionshort, questiontext)

## First i turn all strings into lower case

assembledtextquestions <- assembledtextquestions %>%
  mutate(text_l = str_to_lower(text))

## Before tokenizing I create a stopwords dataset, I add br, because it
## is not properly removed from some observations.

stop_words <- get_stopwords(language = "no")

removeword <- data.frame(word = "br", lexicon = "own library")

stop_words <- rbind(stop_words, removeword)

## Next I kennize the text, this takes a moment.

questiontokens <- assembledtextquestions %>%
  unnest_tokens(input = text, # Which variable to get the text from
               output = word, # What the new variable should be called
               token = "words") # Type of tokens to split the text into
```

```
## Removing all stopwords from my stop words book

questiontokens <- questiontokens %>%
  anti_join(stop_words, by = "word") # Joining against the stopwords dataframe to get rid of cells with

## Brief category investigation, stock companies are frequent

questiontokens %>%
  count(id, word, sort = TRUE)

## Creating the document feature matrix

questiontokens <- questiontokens %>%
  mutate(stem = wordStem(word, language = "norwegian"))

questiontokens_dfm <- questiontokens %>%
  count(id, stem, name = "count") %>% # By default, the count-variable is called "n". Use name = "" to
  cast_dfm(id, # Specify the documents in your analysis (becoming the rows)
    stem, # Specify the tokens in your analysis (becoming the columns)
    count) # Specify the number of times each token shows up in each document (becoming the cell)

## Saving these objects, ready for analysis, note that the tokens dataframe
## is far too large to be uploaded to GitHub

save(questiontokens_dfm, file = "Question_Data/questionsdfm.Rdata")
save(questiontokens, file = "Question_Data/questiontokens.Rdata")
```

With the document feature matrix, we can apply an stm function to it. We run it with 75 categories, slightly more than what the package itself recommends for a corpus of this size.

```
questiontokens_lda_75 <- stm(questiontokens_dfm,
  init.type = "LDA",
  K = 75,
  seed = 910,
  verbose = TRUE,
  max.em.its = 100, ## You can adjust these numbers
  emtol = 1e-5) ## However, too high numbers will drastically increase compu

save(questiontokens_lda_75, file = "Question_Analysis/QuestionSTM_K75.Rdata")

## Save the object, because the model takes an eternity to run.
```

The model has now been run, and we have a Latent Dirichlect Allocation model with 75 topics and a corpus of 27 792 documents. This can be used to assign certain document topics, based on the models estimation of their Gamma value.

The next step is then to make a grouping of the tokens, and creating a frame of all the documents and their gamma for each topic

```
load("Question_Analysis/QuestionSTM_K75.Rdata")
load("Question_Data/questionsdfm.Rdata")

questiontopics_75 <- tidy(questiontokens_lda_75,
  matrix = "beta")
```

```

## Grouping by topic and getting the highest charging words for each topic

questiontopics_group_75 <- questiontopics_75 %>%
  group_by(topic) %>% # Getting the top term per topic, thus using group_by
  slice_max(beta, n = 10) %>% # Fetching the 10 terms with the highest beta
  ungroup() # Ungrouping to get the dataframe back to normal

## Quite difficult to plot all of them at the same time, so that is not included
## here, but it has been done in order to find some interesting topics.

#### Making the matrix of Gammas for each document

question_doc_prob_75 <- tidy(questiontokens_lda_75, matrix = "gamma", # Calculating document probabilities
                             document_names = rownames(questiontokens_dfm)) # Adding the question id's

### Then we assign each document their top 3 topics

top_docs <- question_doc_prob_75 %>%
  group_by(document) %>% # Find the next statistic per document
  slice_max(gamma, n = 3) # Find the max value

topic_assignment <- top_docs %>%
  group_by(document) %>%
  summarise(TopTopic = first(topic),
            SecTopic = nth(topic, 2),
            ThirdTopic = nth(topic, 3),
            TopGamma = first(gamma),
            SecGamma = nth(gamma, 2),
            ThirdGamma = nth(gamma, 3))

### Renaming document to id, because the name of the documents are used to match
### the full data later

topic_assignment <- topic_assignment %>%
  rename(id = document)

```

The next step is then to join these variables with our full set of questions. Leaving us with a dataset that we can filter out keywords on. The previous frames and objects will still be used later, when we explore alternative topic spreads.

```

load("Question_Data/All_Questions.Rdata")

## Changing it all to lower case.

questiontext <- questiontext %>%
  mutate(text_q_l = str_to_lower(question_text),
         text_a_l = str_to_lower(answer_text),
         justification_l = str_to_lower(justification))

## Filtering out our relevant texts

textrelevance <- questiontext %>%
  filter(str_detect(text_q_l, "fn-sambandet") |

```

```

str_detect(text_q_l, "utdanning for bærekraftig utvikling") |
str_detect(text_q_l, "utdanning for berekraftig utvikling") |
str_detect(text_q_l, "unicef") |
str_detect(text_q_l, "wfp") |
str_detect(text_q_l, "unaids") |
str_detect(text_a_l, "fn-sambandet") |
str_detect(text_a_l, "utdanning for bærekraftig utvikling") |
str_detect(text_a_l, "utdanning for berekraftig utvikling") |
str_detect(text_a_l, "unicef") |
str_detect(text_a_l, "wfp") |
str_detect(text_a_l, "unaids") |
str_detect(justification_l, "fn-sambandet") |
str_detect(justification_l, "utdanning for bærekraftig utvikling") |
str_detect(justification_l, "utdanning for berekraftig utvikling") |
str_detect(justification_l, "unicef") |
str_detect(justification_l, "wfp") |
str_detect(justification_l, "unaids") |
str_detect(text_q_l, "unesco") |
str_detect(text_a_l, "unesco") |
str_detect(justification_l, "unesco") |
str_detect(text_q_l, "undp") |
str_detect(text_a_l, "undp") |
str_detect(justification_l, "undp") |
str_detect(text_q_l, "\\b(ilo)\\b") |
str_detect(text_a_l, "\\b(ilo)\\b") |
str_detect(justification_l, "\\b(ilo)\\b") |
str_detect(text_q_l, "fns matvareprogram") |
str_detect(text_a_l, "fns matvareprogram") |
str_detect(justification_l, "fns matvareprogram") |
str_detect(text_q_l, "who") |
str_detect(text_a_l, "who") |
str_detect(justification_l, "who") |
str_detect(text_q_l, "verdens helseorganisasjon") |
str_detect(text_a_l, "verdens helseorganisasjon") |
str_detect(justification_l, "verdens helseorganisasjon") |
str_detect(text_q_l, "ipcc") |
str_detect(text_a_l, "ipcc") |
str_detect(justification_l, "ipcc") |
str_detect(text_q_l, "fns klimapanel") |
str_detect(text_a_l, "fns klimapanel") |
str_detect(justification_l, "fns klimapanel") |
str_detect(text_q_l, "bærekraftsmål") |
str_detect(text_a_l, "bærekraftsmål") |
str_detect(justification_l, "bærekraftsmål") |
str_detect(text_q_l, "bærekraftsmål") |
str_detect(text_a_l, "bærekraftsmål") |
str_detect(justification_l, "bærekraftsmål") |
str_detect(text_q_l, "2030-agendaen") |
str_detect(text_a_l, "2030-agendaen") |
str_detect(justification_l, "2030-agendaen") |
str_detect(text_q_l, "agenda 2030") |
str_detect(text_a_l, "agenda 2030") |
str_detect(justification_l, "agenda 2030"))

```

```

## Joining them with the top 3 topics for their document.

textrelevance <- left_join(textrelevance, topic_assignment, by = "id")

## lets have a look at the most frequent topics

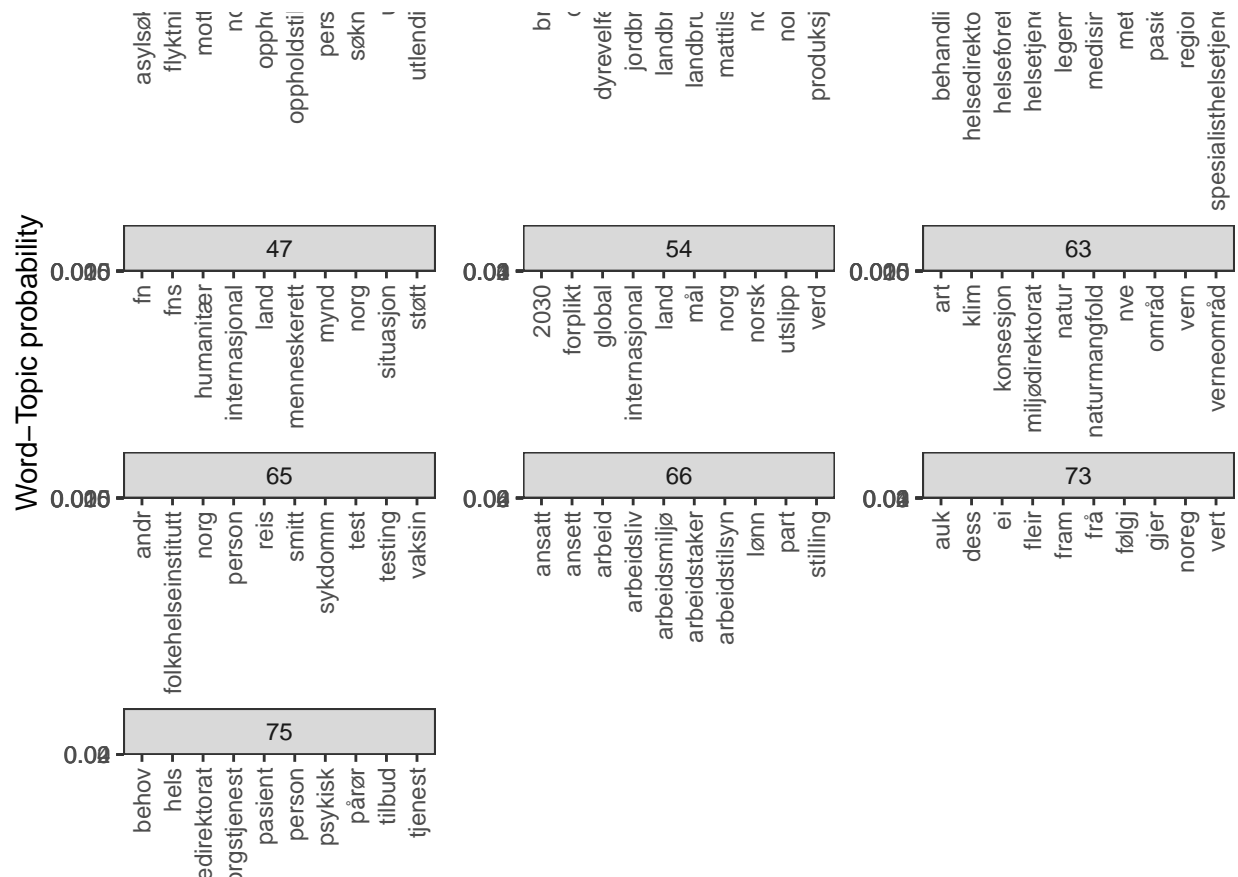
toppingtopics <- textrelevance %>%
  group_by(TopTopic) %>%
  add_count() %>%
  summarise(count = mean(n)) %>%
  slice_max(order_by = count, n = 10)

toppingtopicswords <- questiontopics_group_75 %>%
  filter(topic %in% toppingtopics$TopTopic)

## Lets look at what these topics are about

toppingtopicswords %>%
  ggplot(aes(term, beta, fill = topic)) +
  geom_bar(stat = "identity") +
  facet_wrap( ~ topic,
             ncol = 3,
             scales = "free") +
  labs(x = "", y = "Word-Topic probability") +
  theme_bw() +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

```

Shocking, I know, documents that mention UN organizations are usually about the UN. The interesting thing is that UN organizations are frequently used in talks regarding climate policy and emissions policies. This implies that this sector is more ahead relatively speaking in regards to having a global view on the problems at hand. This is perhaps unsurprising given the global nature of the issue, but perhaps an important path in the future can be to focus on making sure that some other topics, related to global issues are also related more to UN and their organizations.

We should probably investigate the secondary topics, especially since topic 73 is a dedicated nynorsk category. It helps to look at the secondary categories.

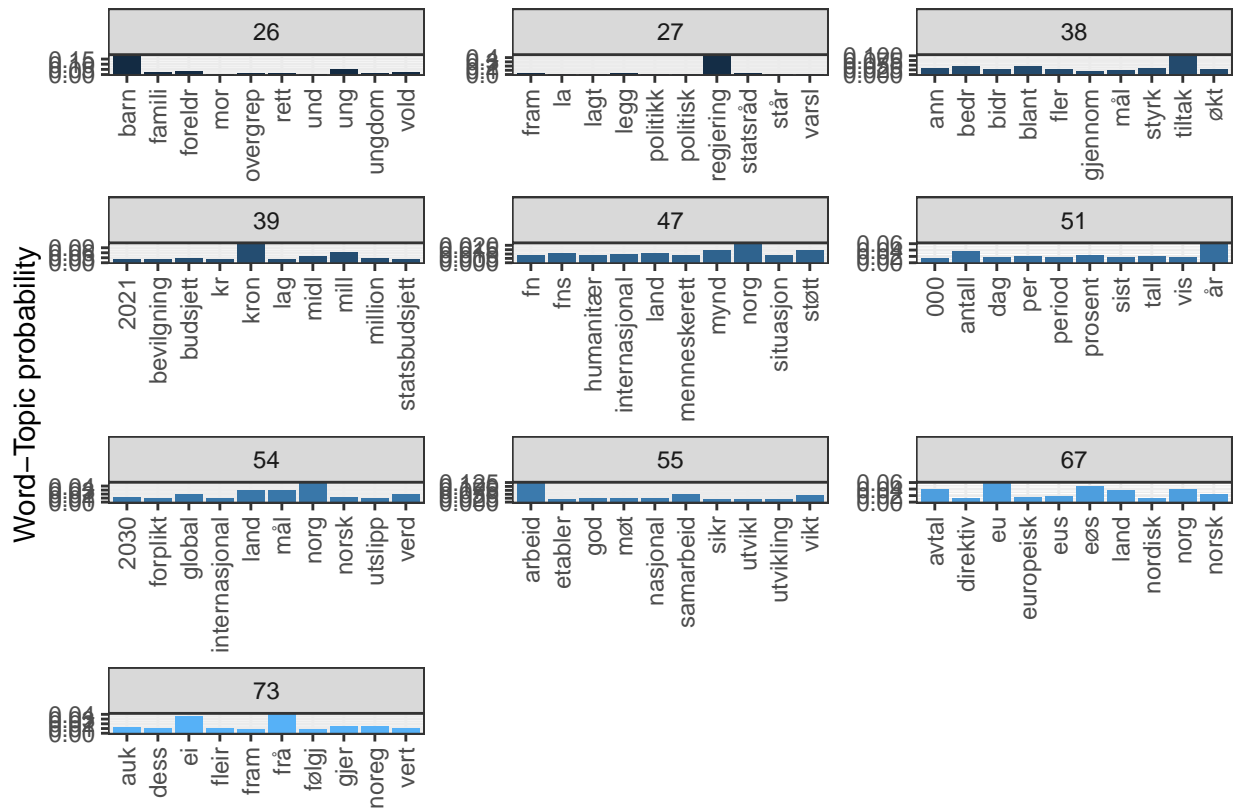
```
secondtopics <- textrelevance %>%
  group_by(SecTopic) %>%
  add_count() %>%
  summarise(count = mean(n)) %>%
  slice_max(order_by = count, n = 10)

secondtopicswords <- questiontopics_group_75 %>%
  filter(topic %in% secondtopics$SecTopic)

## Lets look at what these topics are about

secondtopicswords %>%
  ggplot(aes(term, beta, fill = topic)) +
  geom_bar(stat = "identity") +
  facet_wrap( ~ topic,
             ncol = 3,
             scales = "free") +
```

```
labs(x = "", y = "Word-Topic probability") +
theme_bw() +
theme(legend.position = "none",
      axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```

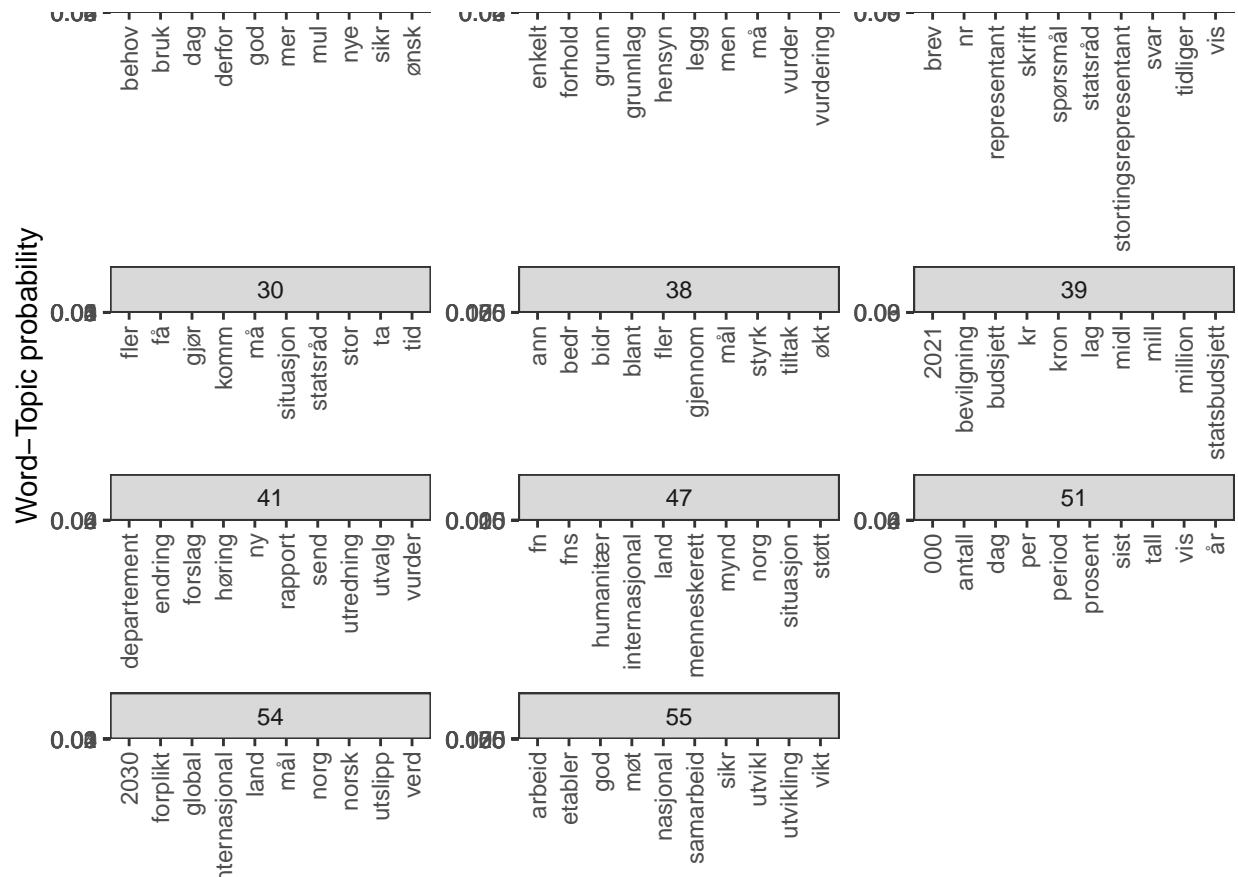


```
thirdtopics <- textrelevance %>%
  group_by(ThirdTopic) %>%
  add_count() %>%
  summarise(count = mean(n)) %>%
  slice_max(order_by = count, n = 10)

thirdtopicswords <- questiontopics_group_75 %>%
  filter(topic %in% thirdtopics$ThirdTopic)

## Lets look at what these topics are about

thirdtopicswords %>%
  ggplot(aes(term, beta, fill = topic)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ topic,
             ncol = 3,
             scales = "free") +
  labs(x = "", y = "Word-Topic probability") +
  theme_bw() +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



Topic 55 charges quite highly in this third topic category. This appears to be a generic “plan of action” category, most likely a common topic for all questions which asks for coherent national policy in fields. The problem here is that we are currently scraping quite low gamma values, so more common topics are brought to the forefront.

So we have talked about what the questions that mention the UN and their associated organisations are talking about, what these questions are substantially about. Unsurprisingly these questions are mostly about climate, health, and the UN itself. Lets now look at some general spread of the questions, when are these questions asked? Do we see a historic pattern?

First lets add some variables of interest, such as mp identities and join it with our priority level.

First we add the priority levels for some of our observations.

```
textrelevance <- textrelevance %>%
  mutate(priority_word_level = case_when(str_detect(text_q_1, "fn-sambandet") ~ "1",
    str_detect(text_q_1, "utdanning for bærekraftig utvikling") ~ "2",
    str_detect(text_q_1, "utdanning for berekraftig utvikling") ~ "2",
    str_detect(text_q_1, "unicef") ~ "3",
    str_detect(text_q_1, "wfp") ~ "3",
    str_detect(text_q_1, "unaids") ~ "3",
    str_detect(text_a_1, "fn-sambandet") ~ "1",
    str_detect(text_a_1, "utdanning for bærekraftig utvikling") ~ "2",
    str_detect(text_a_1, "utdanning for berekraftig utvikling") ~ "2",
    str_detect(text_a_1, "unicef") ~ "3",
    str_detect(text_a_1, "wfp") ~ "3",
    str_detect(text_a_1, "unaids") ~ "3",
    str_detect(justification_1, "fn-sambandet") ~ "1",
```

```

str_detect(justification_1, "utdanning for bærekraftig utvikling") ~ "2",
str_detect(justification_1, "utdanning for bærekraftig utvikling") ~ "2",
str_detect(justification_1, "unicef") ~ "3",
str_detect(justification_1, "wfp") ~ "3",
str_detect(justification_1, "unaids") ~ "3",
str_detect(text_q_1, "unesco") ~ "3",
str_detect(text_a_1, "unesco") ~ "3",
str_detect(justification_1, "unesco") ~ "3",
str_detect(text_q_1, "undp") ~ "3",
str_detect(text_a_1, "undp") ~ "3",
str_detect(justification_1, "undp") ~ "3",
str_detect(text_q_1, "\\b(ilo)\\b") ~ "3",
str_detect(text_a_1, "\\b(ilo)\\b") ~ "3",
str_detect(justification_1, "\\b(ilo)\\b") ~ "3",
str_detect(text_q_1, "fns matvareprogram") ~ "3",
str_detect(text_a_1, "fns matvareprogram") ~ "3",
str_detect(justification_1, "fns matvareprogram") ~ "3",
str_detect(text_q_1, "who") ~ "3",
str_detect(text_a_1, "who") ~ "3",
str_detect(justification_1, "who") ~ "3",
str_detect(text_q_1, "verdens helseorganisasjon") ~ "3",
str_detect(text_a_1, "verdens helseorganisasjon") ~ "3",
str_detect(justification_1, "verdens helseorganisasjon") ~ "3",
str_detect(text_q_1, "ipcc") ~ "3",
str_detect(text_a_1, "ipcc") ~ "3",
str_detect(justification_1, "ipcc") ~ "3",
str_detect(text_q_1, "fns klimapanel") ~ "3",
str_detect(text_a_1, "fns klimapanel") ~ "3",
str_detect(justification_1, "fns klimapanel") ~ "3",
str_detect(text_q_1, "bærekraftsmål") ~ "3",
str_detect(text_a_1, "bærekraftsmål") ~ "3",
str_detect(justification_1, "bærekraftsmål") ~ "3",
str_detect(text_q_1, "bærekraftsmål") ~ "3",
str_detect(text_a_1, "bærekraftsmål") ~ "3",
str_detect(justification_1, "bærekraftsmål") ~ "3",
str_detect(text_q_1, "2030-agendaen") ~ "3",
str_detect(text_a_1, "2030-agendaen") ~ "3",
str_detect(justification_1, "2030-agendaen") ~ "3",
str_detect(text_q_1, "agenda 2030") ~ "3",
str_detect(text_a_1, "agenda 2030") ~ "3",
str_detect(justification_1, "agenda 2030") ~ "3",
str_detect(text_q_1, "bærekraftsmål 4") ~ "2",
str_detect(text_a_1, "bærekraftsmål 4") ~ "2",
str_detect(justification_1, "bærekraftsmål 4") ~ "2"))

```

```

## First we need to get all the mps for the period, then join them with the full
## dataset

```

```

periods_starting <- get_parlperiods()

```

```

periods_starting <- periods_starting %>%
  filter(id %in% c("2021-2025", "2017-2021", "2013-2017", "2009-2013"))

```

```

j <- list()

for (x in periods_starting$id) {
  j[[x]] <- get_parlperiod_mps(periodid = x, substitute = TRUE,
                              good_manners = 0)
}

## 2021-2025 done
## 2017-2021 done
## 2013-2017 done
## 2009-2013 done

allmps <- do.call("rbind", j)

allmpsparties <- allmps %>%
  select(question_from_id = mp_id, gender, party_id, county_id, period_id)

textrelevance <- textrelevance %>%
  mutate(period_id = case_when(session_id == "2021-2022" ~ "2021-2025",
                               session_id == "2020-2021" ~ "2017-2021",
                               session_id == "2019-2020" ~ "2017-2021",
                               session_id == "2018-2019" ~ "2017-2021",
                               session_id == "2017-2018" ~ "2017-2021",
                               session_id == "2016-2017" ~ "2013-2017",
                               session_id == "2015-2016" ~ "2013-2017",
                               session_id == "2014-2015" ~ "2013-2017",
                               session_id == "2013-2014" ~ "2013-2017",
                               session_id == "2012-2013" ~ "2009-2013",
                               session_id == "2011-2012" ~ "2009-2013"))

textrelevanceparty <- left_join(textrelevance, allmpsparties,
                                by = c("question_from_id", "period_id"))

## Finally we assign each question as coming from either a government party,
## an opposition party, or a confidence and supply party

textrelevanceparty <- textrelevanceparty %>%
  mutate(govposition = case_when(party_id == "A" & session_id == "2021-2022" ~ "Government",
                                 party_id == "Sp" & session_id == "2021-2022" ~ "Government",
                                 party_id == "SV" & session_id == "2021-2022" ~ "Confidence and Supply",
                                 party_id == "H" & session_id == "2021-2022" ~ "Opposition",
                                 party_id == "FrP" & session_id == "2021-2022" ~ "Opposition",
                                 party_id == "KrF" & session_id == "2021-2022" ~ "Opposition",
                                 party_id == "MDG" & session_id == "2021-2022" ~ "Opposition",
                                 party_id == "R" & session_id == "2021-2022" ~ "Opposition",
                                 party_id == "V" & session_id == "2021-2022" ~ "Opposition",
                                 party_id == "A" & session_id == "2020-2021" ~ "Opposition",
                                 party_id == "Sp" & session_id == "2020-2021" ~ "Opposition",
                                 party_id == "SV" & session_id == "2020-2021" ~ "Opposition",
                                 party_id == "H" & session_id == "2020-2021" ~ "Government",
                                 party_id == "FrP" & session_id == "2020-2021" ~ "Confidence and Supply",
                                 party_id == "KrF" & session_id == "2020-2021" ~ "Government",
                                 party_id == "MDG" & session_id == "2020-2021" ~ "Opposition",

```

[illegible]

```

party_id == "MDG" & session_id == "2014-2015" ~ "Opposition",
party_id == "R" & session_id == "2014-2015" ~ "Opposition",
party_id == "V" & session_id == "2014-2015" ~ "Confidence and Supply",
party_id == "A" & session_id == "2013-2014" ~ "Opposition",
party_id == "Sp" & session_id == "2013-2014" ~ "Opposition",
party_id == "SV" & session_id == "2013-2014" ~ "Opposition",
party_id == "H" & session_id == "2013-2014" ~ "Government",
party_id == "FrP" & session_id == "2013-2014" ~ "Government",
party_id == "KrF" & session_id == "2013-2014" ~ "Confidence and Supply",
party_id == "MDG" & session_id == "2013-2014" ~ "Opposition",
party_id == "R" & session_id == "2013-2014" ~ "Opposition",
party_id == "V" & session_id == "2013-2014" ~ "Confidence and Supply",
party_id == "A" & session_id == "2012-2013" ~ "Government",
party_id == "Sp" & session_id == "2012-2013" ~ "Government",
party_id == "SV" & session_id == "2012-2013" ~ "Government",
party_id == "H" & session_id == "2012-2013" ~ "Opposition",
party_id == "FrP" & session_id == "2012-2013" ~ "Opposition",
party_id == "KrF" & session_id == "2012-2013" ~ "Opposition",
party_id == "MDG" & session_id == "2012-2013" ~ "Opposition",
party_id == "R" & session_id == "2012-2013" ~ "Opposition",
party_id == "V" & session_id == "2012-2013" ~ "Opposition",
party_id == "A" & session_id == "2011-2012" ~ "Government",
party_id == "Sp" & session_id == "2011-2012" ~ "Government",
party_id == "SV" & session_id == "2011-2012" ~ "Government",
party_id == "H" & session_id == "2011-2012" ~ "Opposition",
party_id == "FrP" & session_id == "2011-2012" ~ "Opposition",
party_id == "KrF" & session_id == "2011-2012" ~ "Opposition",
party_id == "MDG" & session_id == "2011-2012" ~ "Opposition",
party_id == "R" & session_id == "2011-2012" ~ "Opposition",
party_id == "V" & session_id == "2011-2012" ~ "Opposition"
))

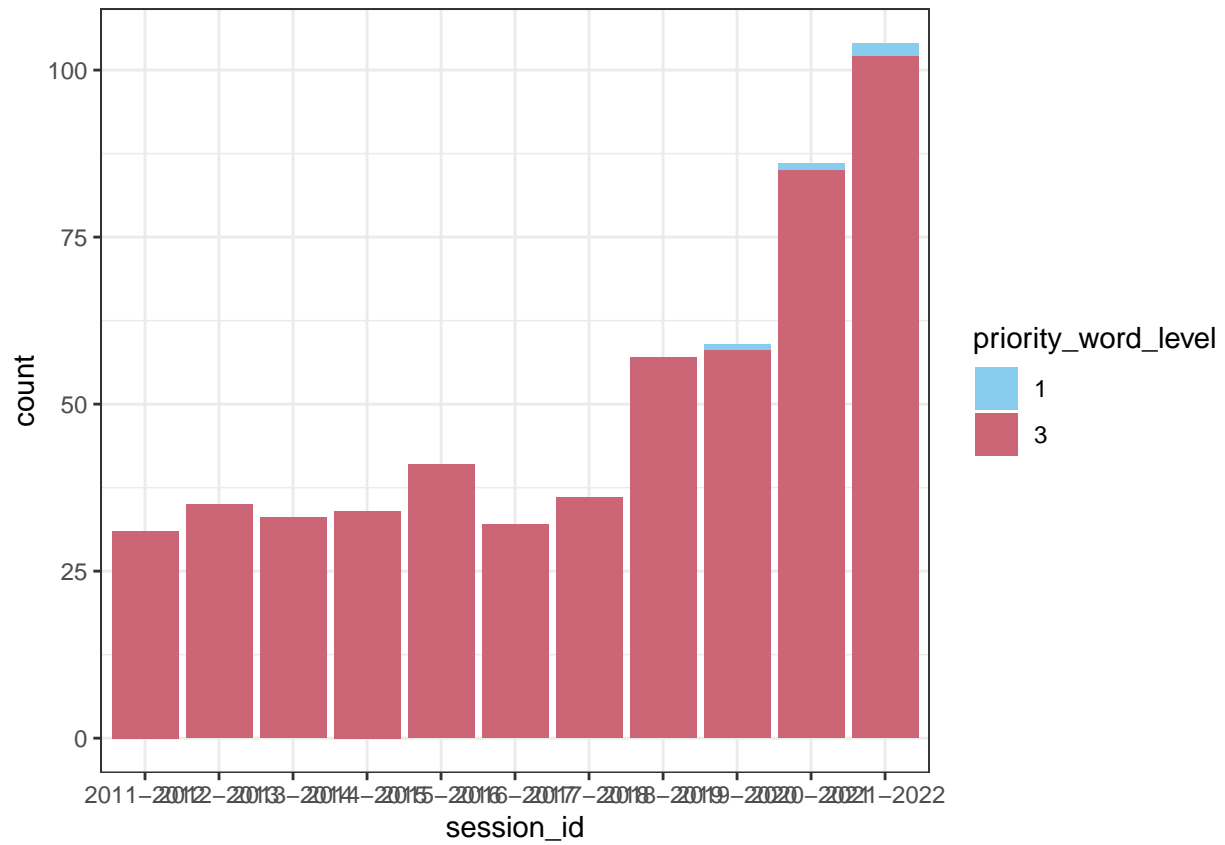
```

Now that we have a dataset with more info, lets look at some of the patterns we can notice over time.

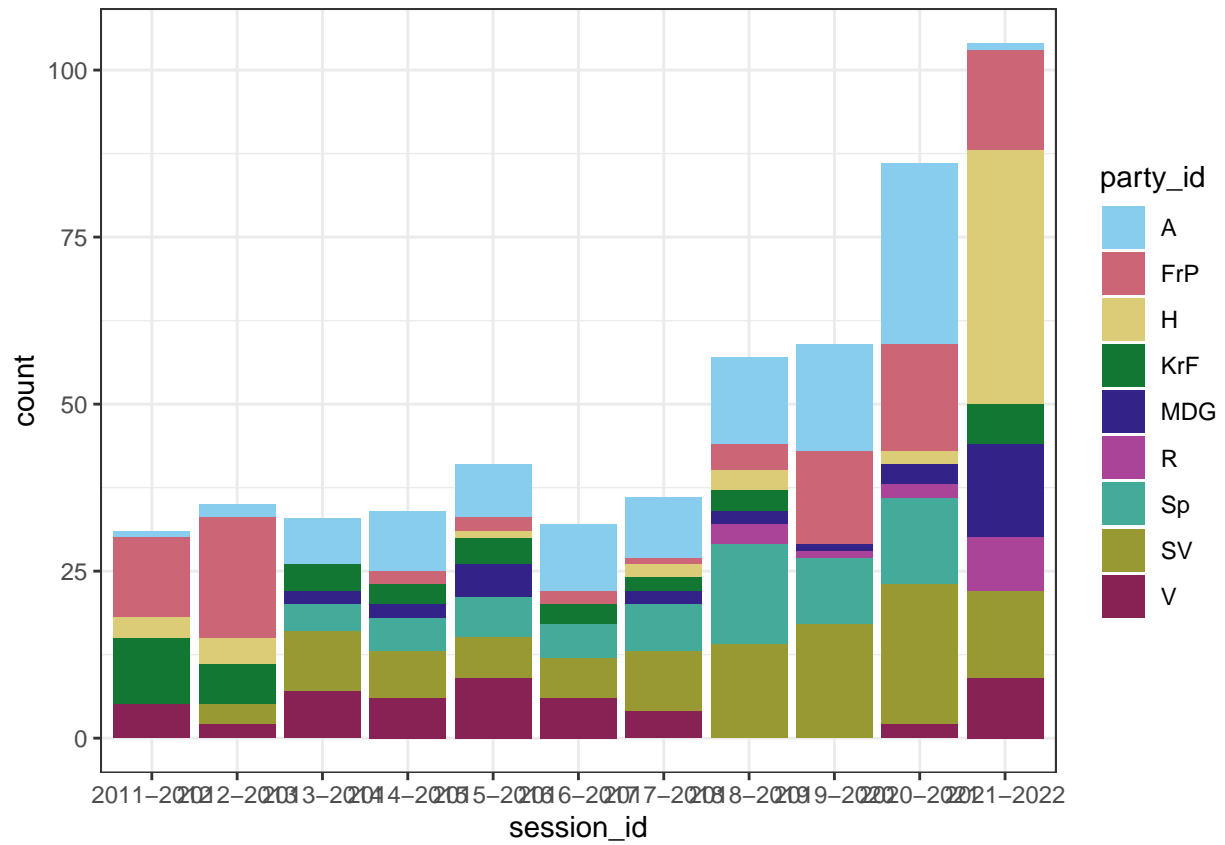
```

ggplot(textrelevanceparty, aes(x = session_id, fill = priority_word_level)) +
  geom_bar() +
  scale_fill_discrete(type = safe_colorblind_palette) +
  theme_bw()

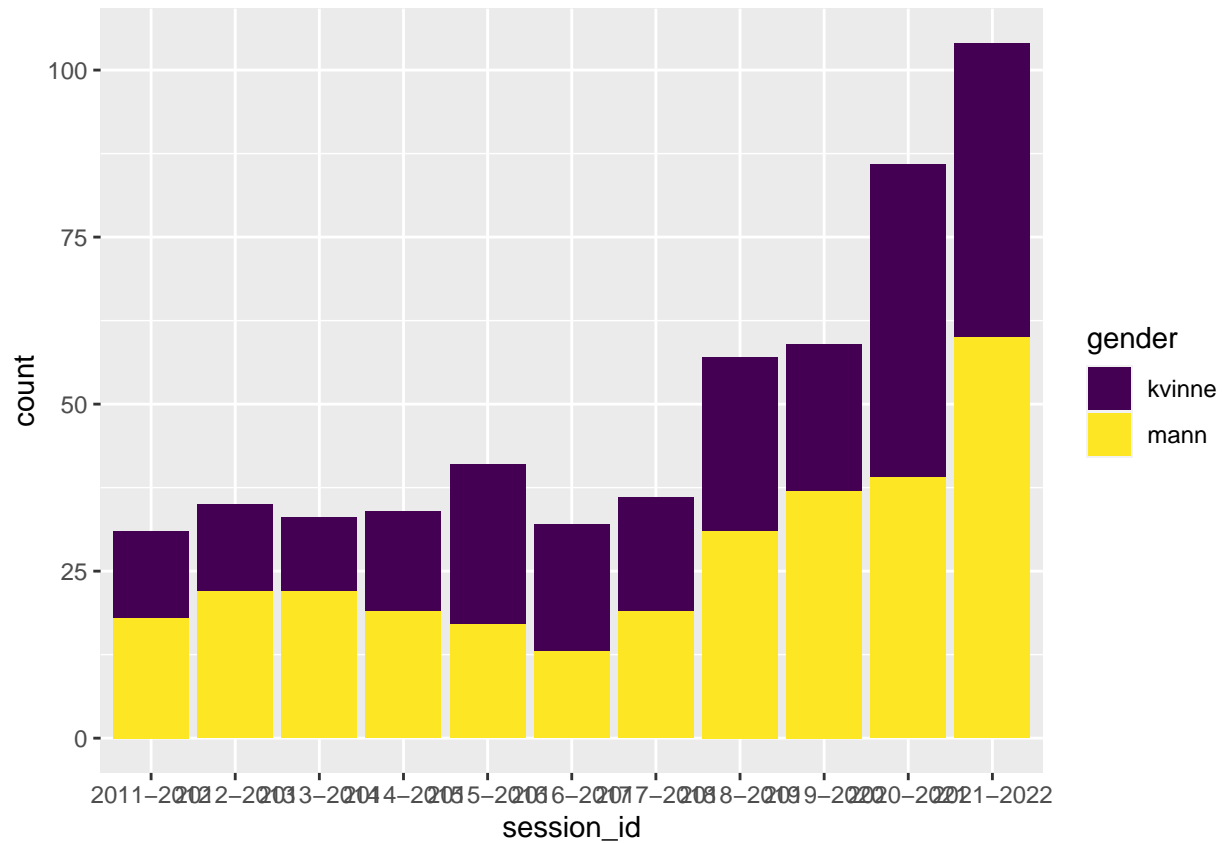
```



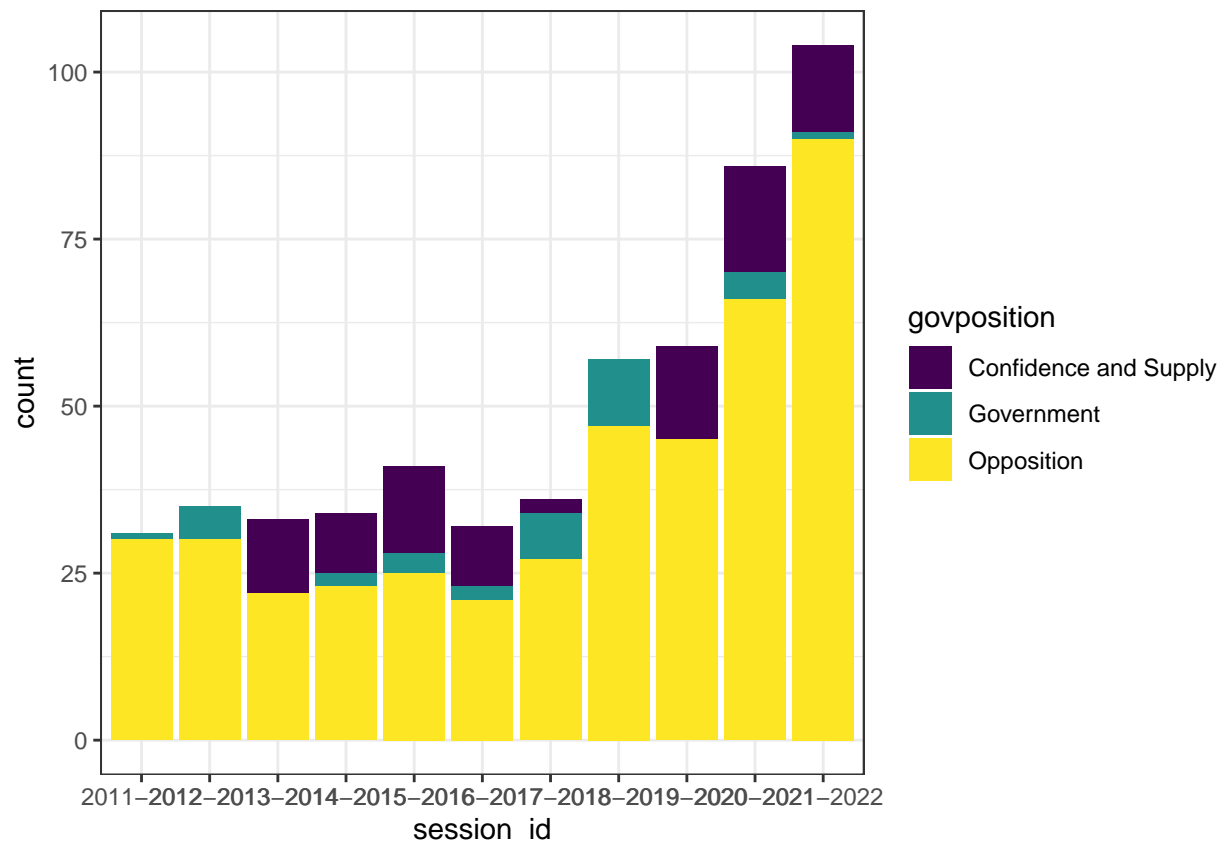
```
ggplot(textrelevanceparty, aes(x = session_id, fill = party_id)) +
  geom_bar() +
  scale_fill_discrete(type = safe_colorblind_palette) +
  theme_bw()
```

```
ggplot(textrelevanceparty, aes(x = session_id, fill = gender)) +
  geom_bar() +
  scale_fill_viridis(discrete = TRUE)
```



```
ggplot(textrelevancexparty, aes(x = session_id, fill = govposition)) +
  geom_bar() +
  scale_fill_viridis(discrete = TRUE) +
  theme_bw()
```



We see a clear spike in mentions in 2021-2022, likely due to the war in Ukraine. In order to further evaluate this patterns, we need to create a comparison to all the questions being asked, that is plotted below.

*## Creating the same variables in the full dataframe (this is in fact a very
overcomplicated way of doing this, but it was done this way and now its legacy
code)*

```
questiontext2 <- questiontext %>%
  mutate(period_id = case_when(session_id == "2021-2022" ~ "2021-2025",
                                session_id == "2020-2021" ~ "2017-2021",
                                session_id == "2019-2020" ~ "2017-2021",
                                session_id == "2018-2019" ~ "2017-2021",
                                session_id == "2017-2018" ~ "2017-2021",
                                session_id == "2016-2017" ~ "2013-2017",
                                session_id == "2015-2016" ~ "2013-2017",
                                session_id == "2014-2015" ~ "2013-2017",
                                session_id == "2013-2014" ~ "2013-2017",
                                session_id == "2012-2013" ~ "2009-2013",
                                session_id == "2011-2012" ~ "2009-2013"))

questiontextwparty <- left_join(questiontext2, allmpsparties,
                                by = c("question_from_id", "period_id"))

questiontextwparty <- questiontextwparty %>%
  mutate(govposition = case_when(party_id == "A" & session_id == "2021-2022" ~ "Government",
                                  party_id == "Sp" & session_id == "2021-2022" ~ "Government",
```



```

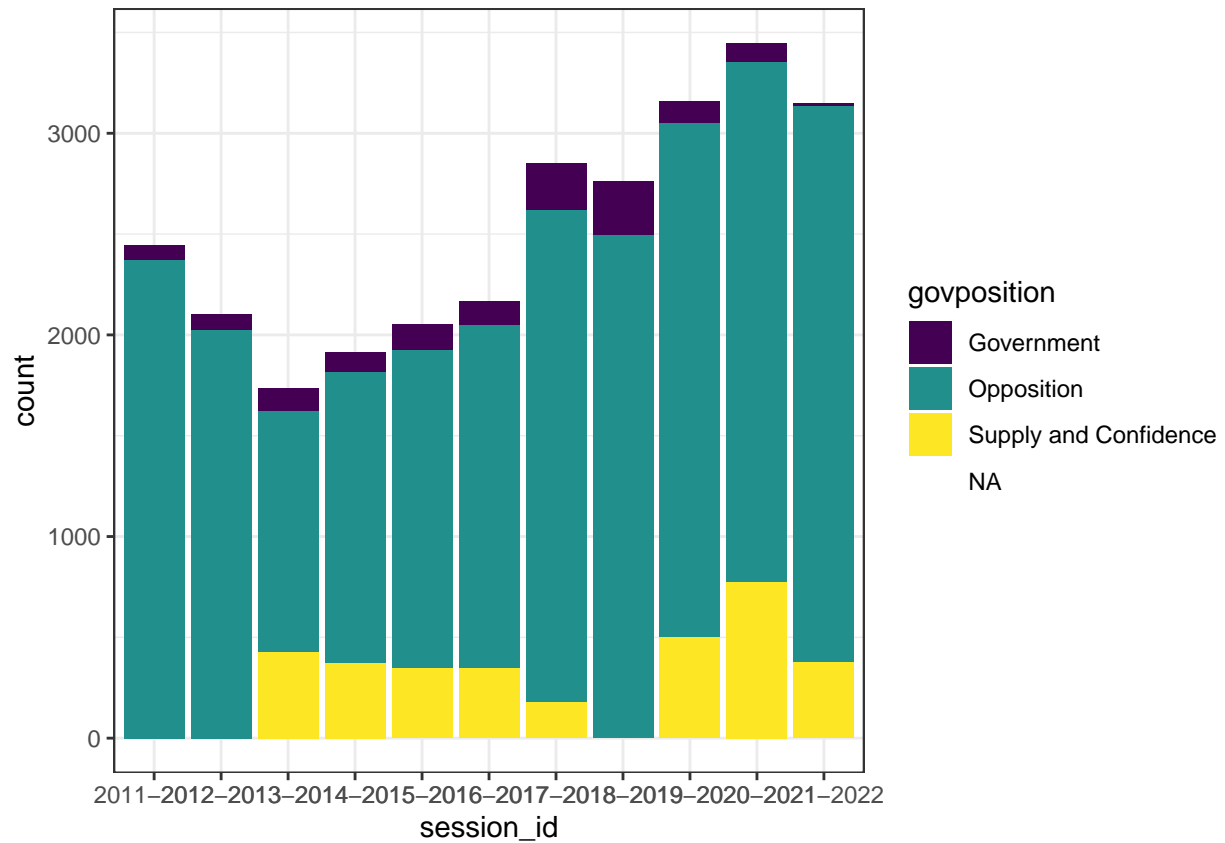
party_id == "Sp" & session_id == "2015-2016" ~ "Opposition",
party_id == "SV" & session_id == "2015-2016" ~ "Opposition",
party_id == "H" & session_id == "2015-2016" ~ "Government",
party_id == "FrP" & session_id == "2015-2016" ~ "Government",
party_id == "KrF" & session_id == "2015-2016" ~ "Supply and Confidence",
party_id == "MDG" & session_id == "2015-2016" ~ "Opposition",
party_id == "R" & session_id == "2015-2016" ~ "Opposition",
party_id == "V" & session_id == "2015-2016" ~ "Supply and Confidence",
party_id == "A" & session_id == "2014-2015" ~ "Opposition",
party_id == "Sp" & session_id == "2014-2015" ~ "Opposition",
party_id == "SV" & session_id == "2014-2015" ~ "Opposition",
party_id == "H" & session_id == "2014-2015" ~ "Government",
party_id == "FrP" & session_id == "2014-2015" ~ "Government",
party_id == "KrF" & session_id == "2014-2015" ~ "Supply and Confidence",
party_id == "MDG" & session_id == "2014-2015" ~ "Opposition",
party_id == "R" & session_id == "2014-2015" ~ "Opposition",
party_id == "V" & session_id == "2014-2015" ~ "Supply and Confidence",
party_id == "A" & session_id == "2013-2014" ~ "Opposition",
party_id == "Sp" & session_id == "2013-2014" ~ "Opposition",
party_id == "SV" & session_id == "2013-2014" ~ "Opposition",
party_id == "H" & session_id == "2013-2014" ~ "Government",
party_id == "FrP" & session_id == "2013-2014" ~ "Government",
party_id == "KrF" & session_id == "2013-2014" ~ "Supply and Confidence",
party_id == "MDG" & session_id == "2013-2014" ~ "Opposition",
party_id == "R" & session_id == "2013-2014" ~ "Opposition",
party_id == "V" & session_id == "2013-2014" ~ "Supply and Confidence",
party_id == "A" & session_id == "2012-2013" ~ "Government",
party_id == "Sp" & session_id == "2012-2013" ~ "Government",
party_id == "SV" & session_id == "2012-2013" ~ "Government",
party_id == "H" & session_id == "2012-2013" ~ "Opposition",
party_id == "FrP" & session_id == "2012-2013" ~ "Opposition",
party_id == "KrF" & session_id == "2012-2013" ~ "Opposition",
party_id == "MDG" & session_id == "2012-2013" ~ "Opposition",
party_id == "R" & session_id == "2012-2013" ~ "Opposition",
party_id == "V" & session_id == "2012-2013" ~ "Opposition",
party_id == "A" & session_id == "2011-2012" ~ "Government",
party_id == "Sp" & session_id == "2011-2012" ~ "Government",
party_id == "SV" & session_id == "2011-2012" ~ "Government",
party_id == "H" & session_id == "2011-2012" ~ "Opposition",
party_id == "FrP" & session_id == "2011-2012" ~ "Opposition",
party_id == "KrF" & session_id == "2011-2012" ~ "Opposition",
party_id == "MDG" & session_id == "2011-2012" ~ "Opposition",
party_id == "R" & session_id == "2011-2012" ~ "Opposition",
party_id == "V" & session_id == "2011-2012" ~ "Opposition"

))

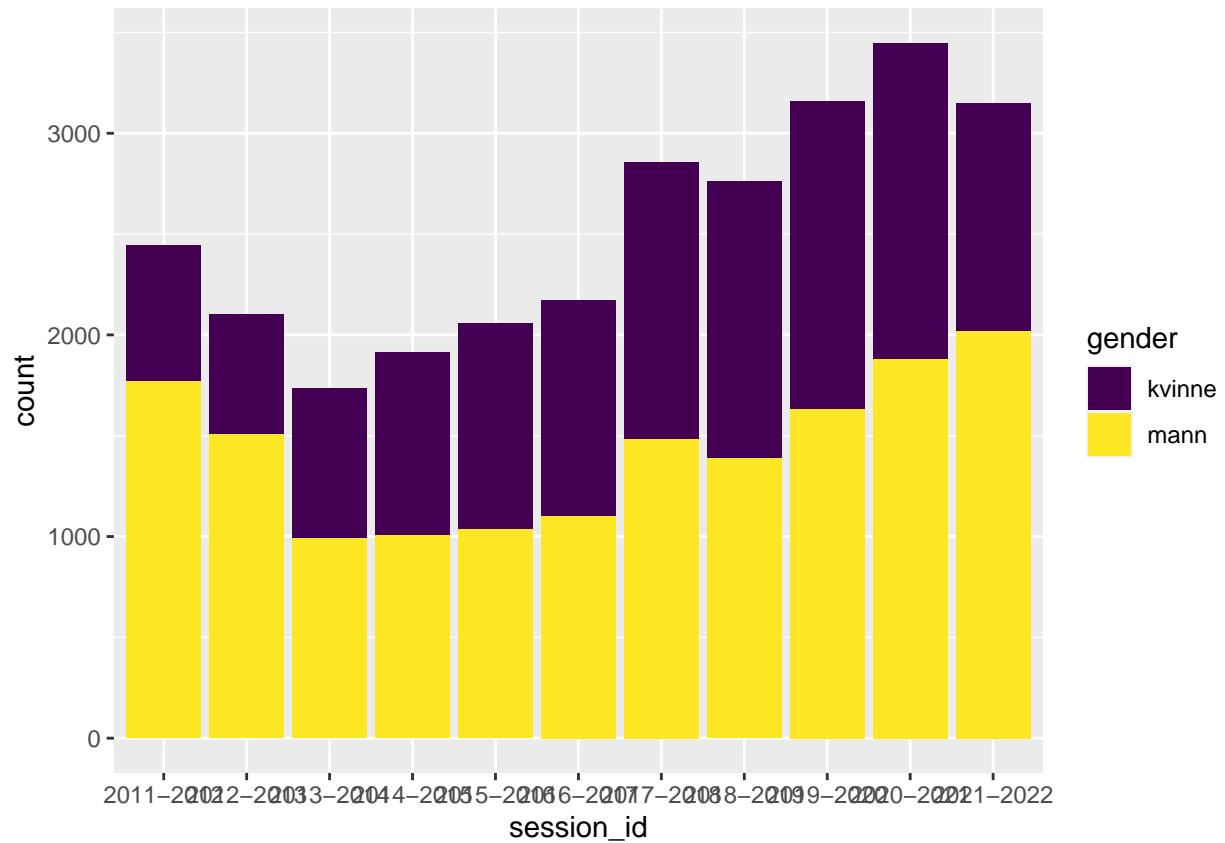
## Plotting this

ggplot(questiontextwparty, aes(x = session_id, fill = govposition)) +
  geom_bar() +
  scale_fill_viridis(discrete = TRUE) +
  theme_bw()

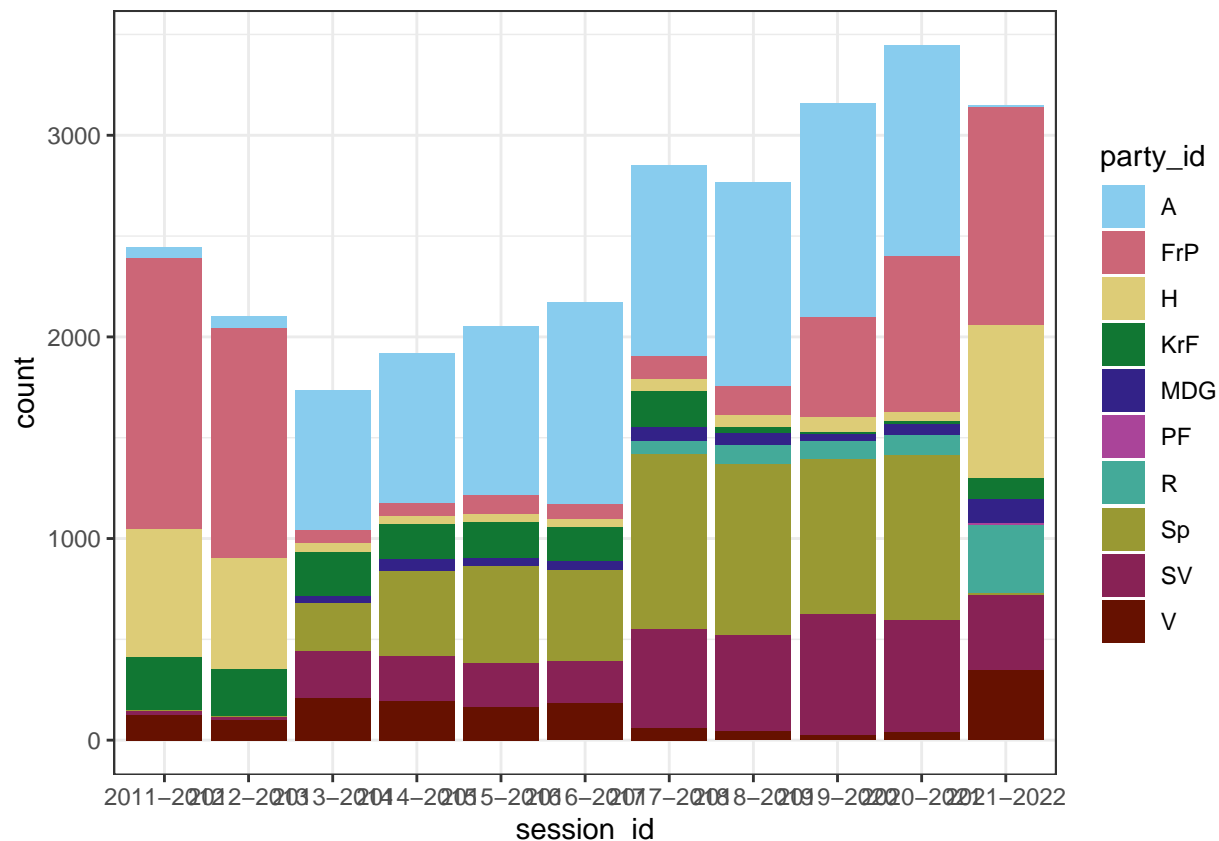
```



```
ggplot(questiontextwparty, aes(x = session_id, fill = govposition)) +
  geom_bar() +
  scale_fill_viridis(discrete = TRUE)
```



```
ggplot(questiontextwparty, aes(x = session_id, fill = party_id)) +  
  geom_bar() +  
  scale_fill_discrete(type = safe_colorblind_palette) +  
  theme_bw()
```



In summary, if we compare these to the results of our keyword containing documents, we see similar patterns in terms of both gender and government position, this is perhaps unsurprising. We do however see some significant differences in terms of parties asking questions. The liberal party (V) and the Socialist Left (SV) are heavily overrepresented. In addition the Progress party (FrP), was much more active during the Stoltenberg 2 Government, than they are against the Støre government. Hinting at abandoning a focus on international aid, perhaps due to the leadership change in 2020.

What else can we use the topic model for?

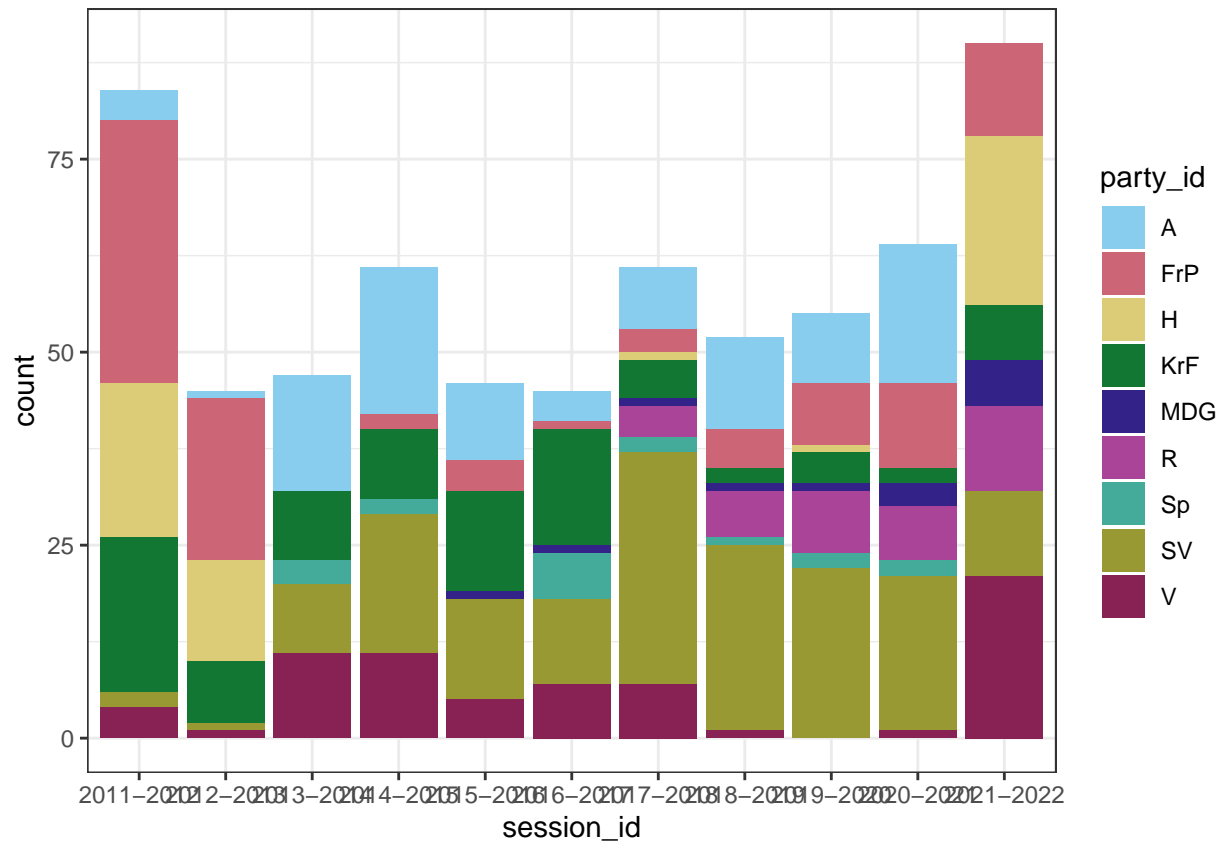
Interestingly the topic 47 seems to be about humanitarian aid and the UN in general, we can use this to find the documents that are most likely to be about UN.

```
## The Gamma for these documents are in fact quite high. But more importantly
## they touch on the UN and their activities in a manner unrelated to security
## policy.
```

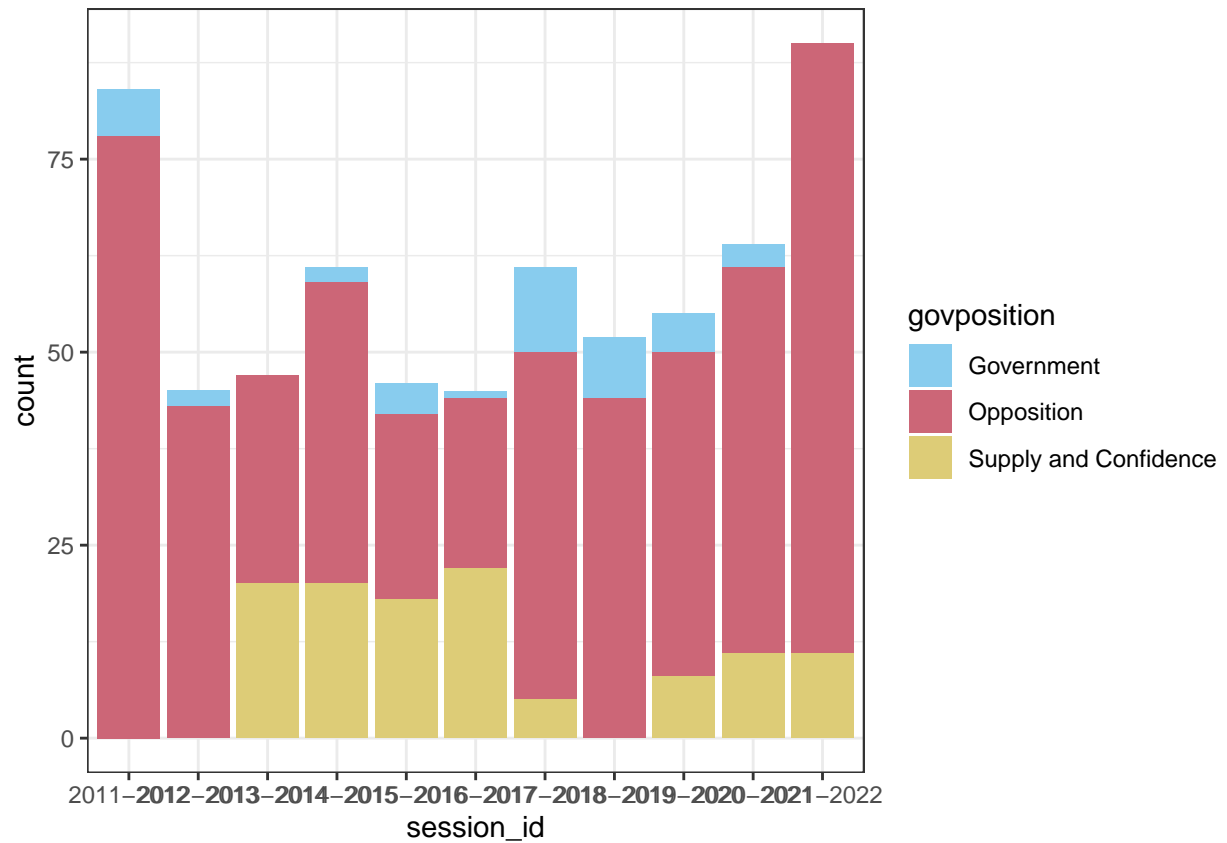
```
topic47docs <- question_doc_prob_75 %>%
  filter(topic == 47) %>%
  slice_max(order_by = gamma, n = 650)

top100question <- questiontextwparty %>%
  filter(id %in% topic47docs$document)

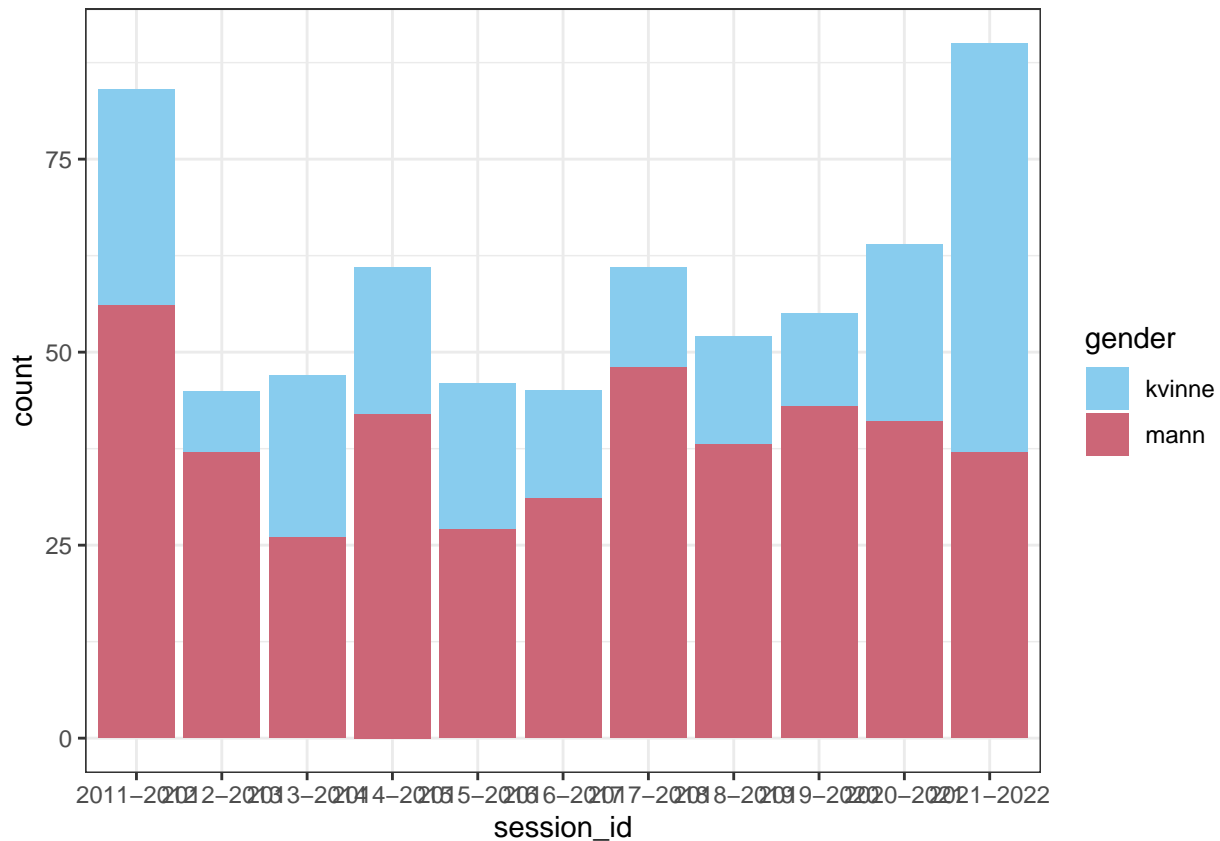
ggplot(top100question, aes(x = session_id, fill = party_id)) +
  geom_bar() +
  scale_fill_discrete(type = safe_colorblind_palette) +
  theme_bw()
```

```
ggplot(top100question, aes(x = session_id, fill = govposition)) +
  geom_bar() +
  scale_fill_discrete(type = safe_colorblind_palette) +
  theme_bw()
```



```
ggplot(top100question, aes(x = session_id, fill = govposition)) +
  geom_bar() +
  scale_fill_discrete(type = safe_colorblind_palette) +
  theme_bw()
```



Similar patterns as in the original keywords.

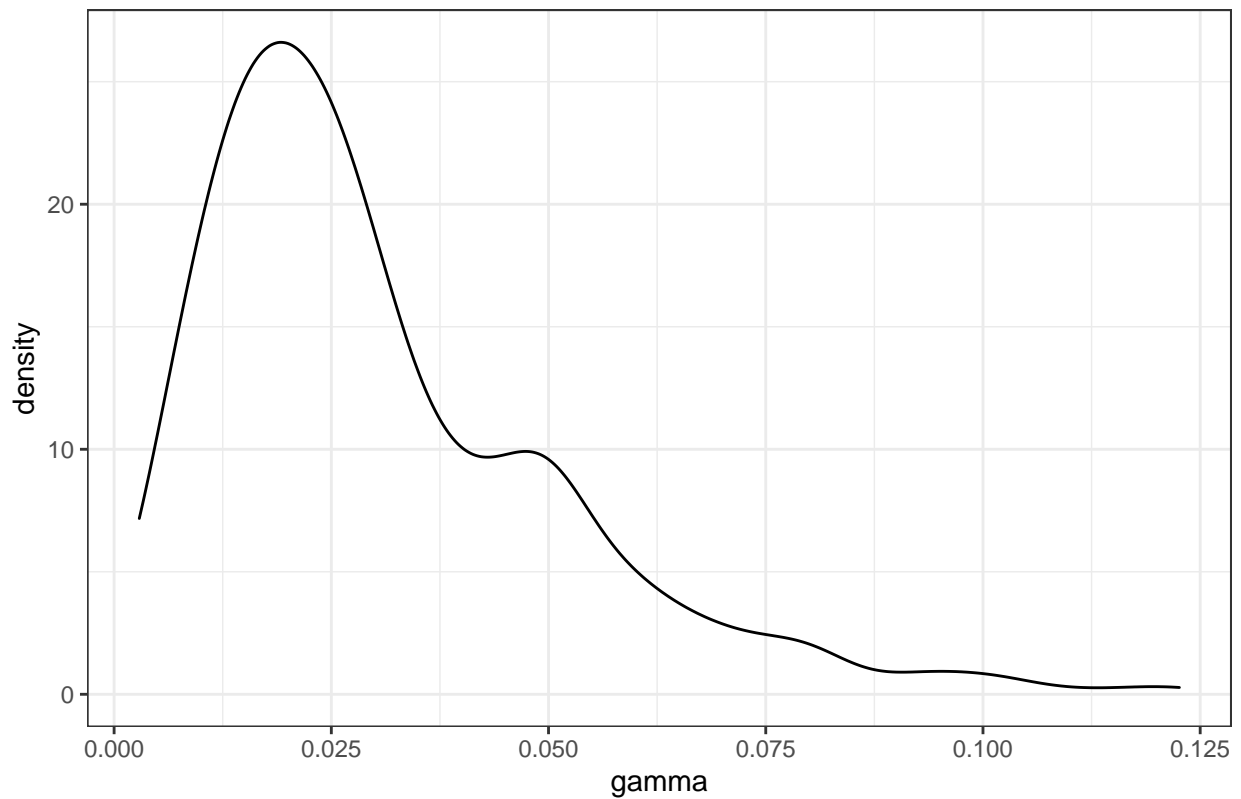
Lets now look at some other categories inside the topic model, and whether they are of interest. By looking closely at the different topics, we find that topics 30, 33, 43, 52, 54, and 63 are of interest. These topics primarily look at climate and conservation topics. However, topic 30 and 52 are interesting in particular. Topic 30 seems to be about making requirements of the minister that is being asked the question, attempting to hold them accountable. A correlation on this topic would imply that MPs are using the UN and their talking points to drive forwards critique of the government, holding them accountable. Topic 52 is also interesting because it is about higher education, and relating that to the UN and their suborganisations would be helpful to see if they connect it to the activities of global development. To explore these topics, we plot the density of the gammas of our top topic 47 documents, to see to what degree they actually measure this.

#Topic 30 is related to questions asked that require the minister in charge to do something.

```
top_docs100topic30 <- question_doc_prob_75 %>%
  group_by(document) %>% # Find the next statistic per document
  filter(topic == 30 & document %in% top100question$id)
```

```
ggplot(top_docs100topic30, aes(x = gamma)) +
  geom_density() +
  theme_bw() +
  labs(title = "Density of Topic 30")
```

Density of Topic 30

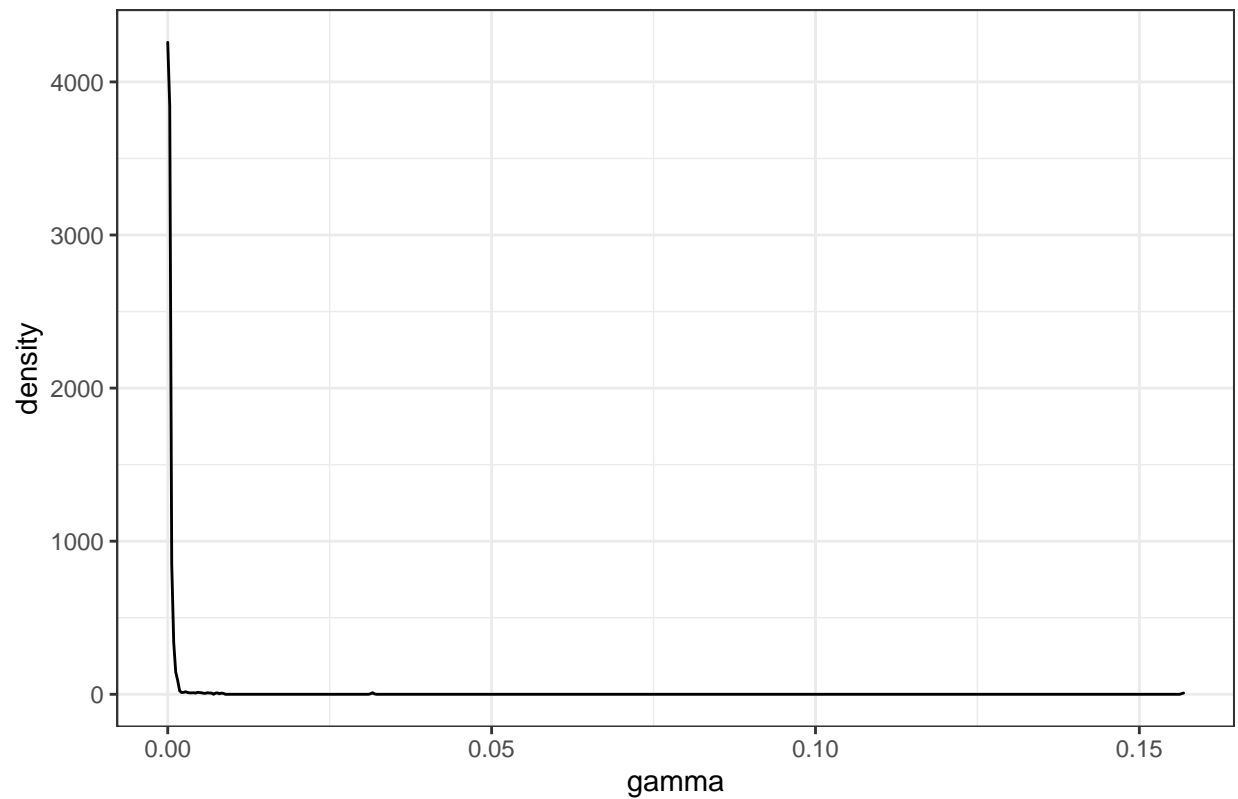


#Topic 33 is about emissions.

```
top_docs100topic33 <- question_doc_prob_75 %>%  
  group_by(document) %>% # Find the next statistic per document  
  filter(topic == 33 & document %in% top100question$id)
```

```
ggplot(top_docs100topic33, aes(x = gamma)) +  
  geom_density() +  
  theme_bw() +  
  labs(title = "Density of Topic 33")
```

Density of Topic 33

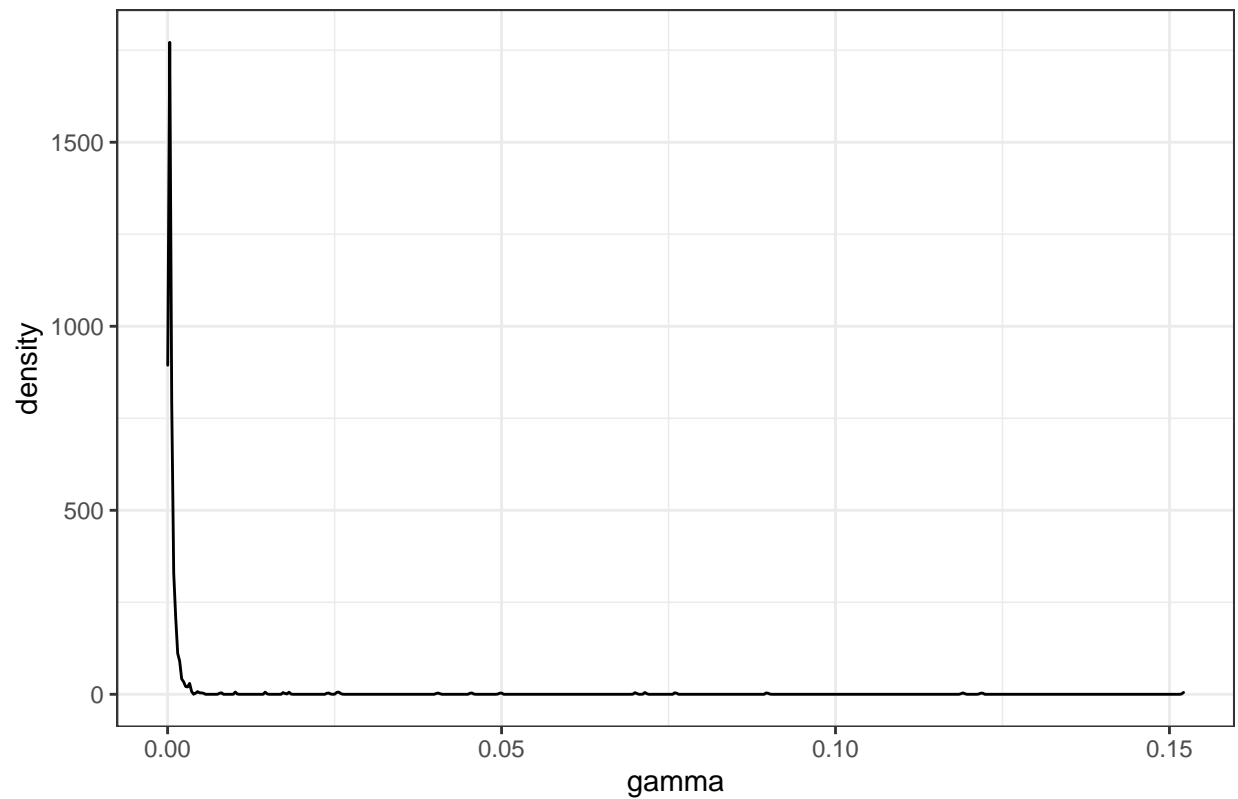


#Topic 43 is about sustainability.

```
top_docs100topic43 <- question_doc_prob_75 %>%  
  group_by(document) %>% # Find the next statistic per document  
  filter(topic == 43 & document %in% top100question$id)
```

```
ggplot(top_docs100topic43, aes(x = gamma)) +  
  geom_density() +  
  theme_bw() +  
  labs(title = "Density of Topic 43")
```

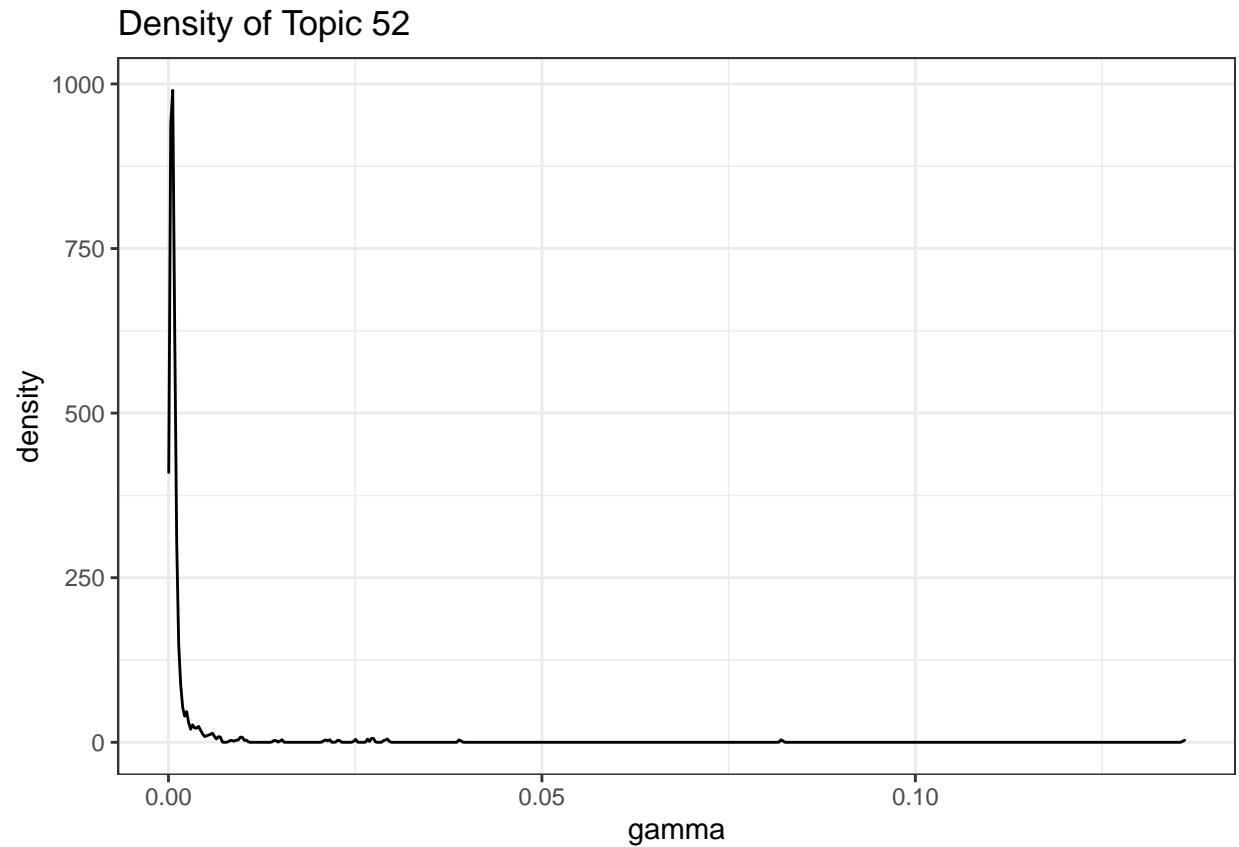
Density of Topic 43



#Topic 52, Higher education

```
top_docs100topic52 <- question_doc_prob_75 %>%  
  group_by(document) %>% # Find the next statistic per document  
  filter(topic == 52 & document %in% top100question$id)
```

```
ggplot(top_docs100topic52, aes(x = gamma)) +  
  geom_density() +  
  theme_bw() +  
  labs(title = "Density of Topic 52")
```

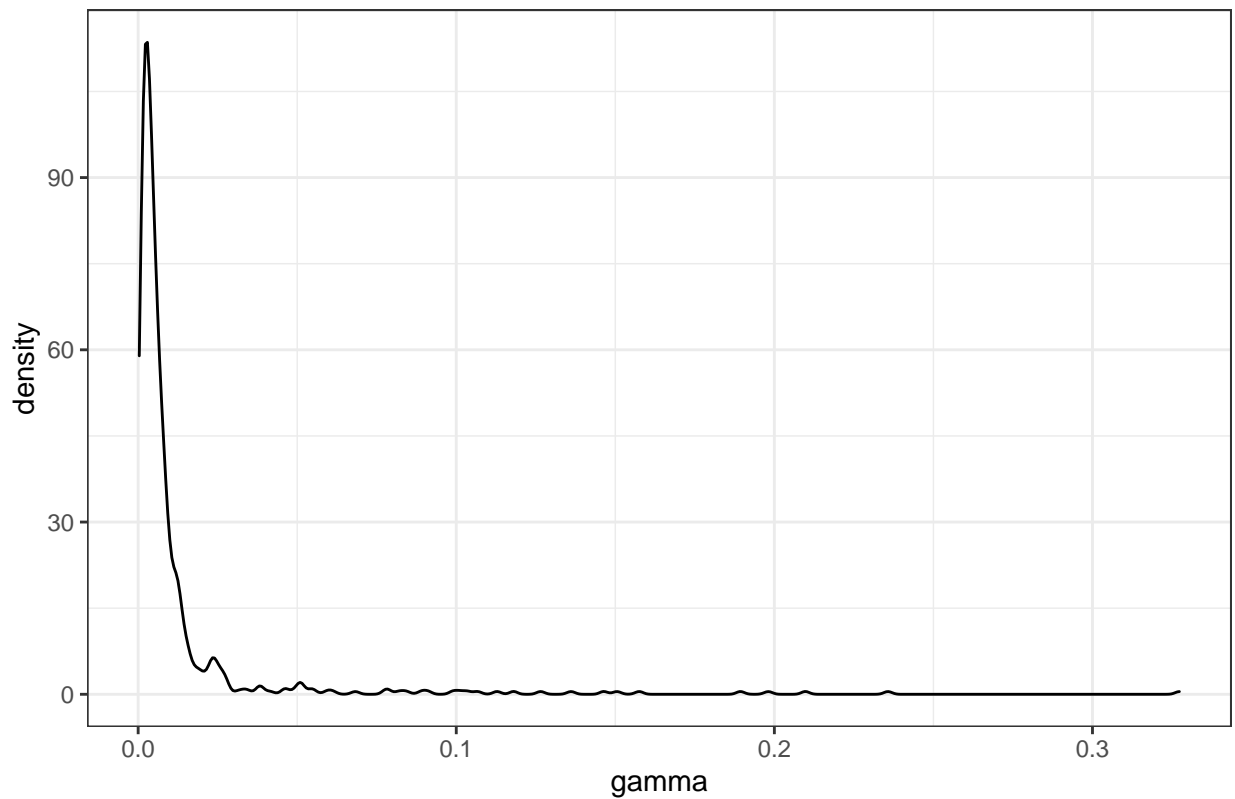


```
#Topic 54, more climate emissions

top_docs100topic54 <- question_doc_prob_75 %>%
  group_by(document) %>% # Find the next statistic per document
  filter(topic == 54 & document %in% top100question$id)

ggplot(top_docs100topic54, aes(x = gamma)) +
  geom_density() +
  theme_bw() +
  labs(title = "Density of Topic 54")
```

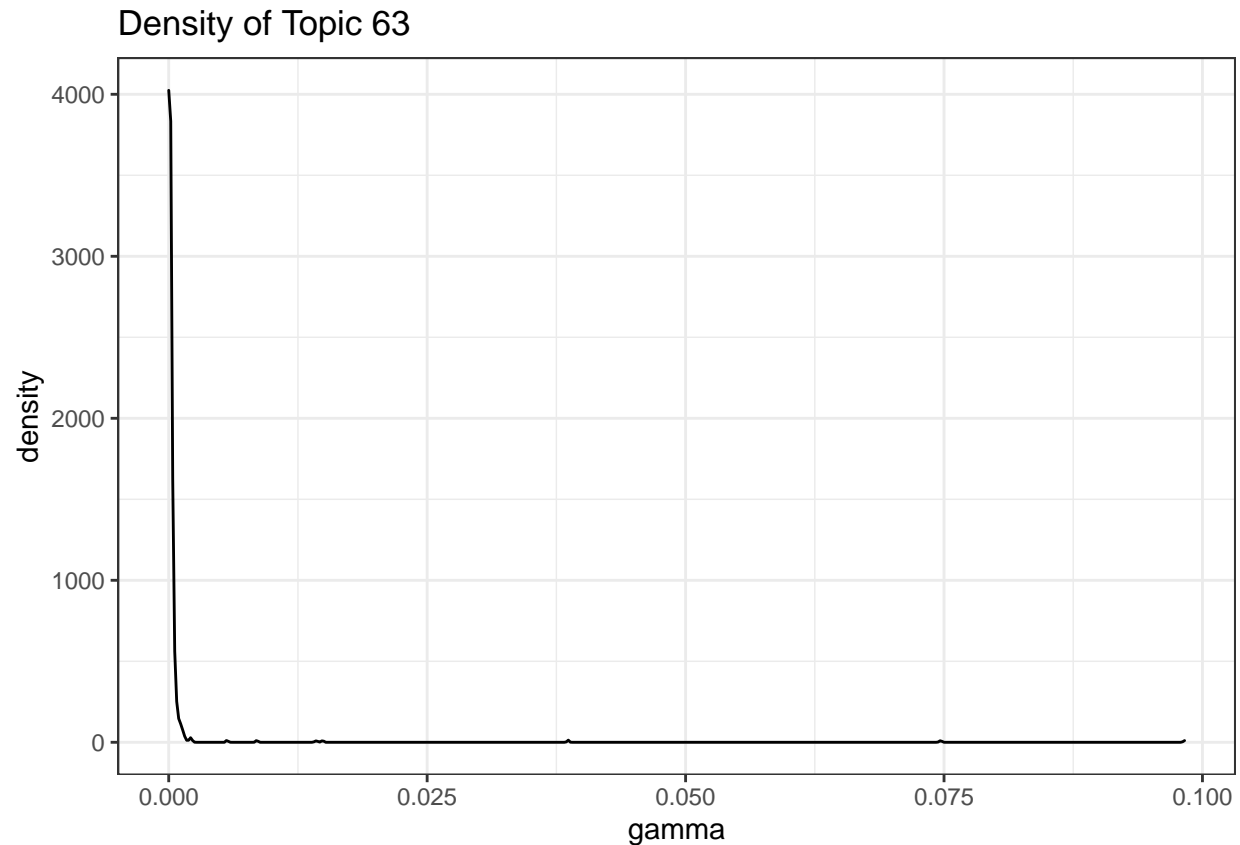
Density of Topic 54



#Topic 63, Environmentalism

```
top_docs100topic63 <- question_doc_prob_75 %>%  
  group_by(document) %>% # Find the next statistic per document  
  filter(topic == 63 & document %in% top100question$id)
```

```
ggplot(top_docs100topic63, aes(x = gamma)) +  
  geom_density() +  
  theme_bw() +  
  labs(title = "Density of Topic 63")
```

Quite low charges on all, save for topic 30, which performs best. The clear thing take away, especially in regards to education is that the MPs are not actively using the UN in talking about development, nor in terms of sustainability. A clear goal for the future lobbying should be to attempt to make sure that 4.7 becomes its own topic, together with keywords like education. Unfortunately, our model does not find this. Primarily because 4.7 appears quite rarely in the questions.

```
table(questiontext$session_id[which(str_detect(questiontext$text_a_1, "delmål 4.7")|
                                         str_detect(questiontext$text_a_1, "bærekraftsmål 4") |
                                         str_detect(questiontext$text_a_1, "utdanning for bærekraftig utvikling")|
                                         str_detect(questiontext$text_q_1, "delmål 4.7")|
                                         str_detect(questiontext$text_q_1, "bærekraftsmål 4") |
                                         str_detect(questiontext$text_q_1, "utdanning for bærekraftig utvikling")|
                                         str_detect(questiontext$justification_1, "delmål 4.7")|
                                         str_detect(questiontext$justification_1, "bærekraftsmål 4") |
                                         str_detect(questiontext$justification_1, "utdanning for bærekraftig utv
```

```
##
## 2018-2019 2020-2021 2021-2022
##          2         1         1
```

As we can see, we see only 4 mentions of sustainable development related keyphrases in our set of questions. This is despite the fact that we see quite a lot more mentions of education in general.

```
table(questiontext$session_id[which(str_detect(questiontext$text_a_1, "utdanning") |
                                         str_detect(questiontext$text_q_1, "utdanning") |
                                         str_detect(questiontext$justification_1, "utdanning"))])
```

```
##
```

##	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019
##	211	204	163	165	217	207	278	248
##	2019-2020	2020-2021	2021-2022					
##	279	299	246					

On average, about 200 questions are asked every session about education, but none of these speak of sustainable development, or about the UNs sustainability goals.

Budget

Before we begin, allow me to paint you the full picture...

In the process of retrieving mentions of some of the suborganisations in the decisions, we quickly found that the decisions that mention this largely consists of budgets. This piqued our interest as we could retrieve the money that is allocated to some of the organisations of the UN over time, thereby meeting a requirement of “quantitative analysis” on the activities of the UNA, given that their main priority is securing funding for the amazing organisations that they work with.

Below is a hidden chunk which creates a new function, because the `get_case` function from `Stortingscrape` did not perform as expected. It creates the function `get_case_fancey`, which does the same as `get_case`, but it actually works.

Lets then get on to getting the decisions that turn out to be about the UN activities.

```
a <- list()

## For loop that fetches all of our fun stuff

for(x in unique(sessions_storting$id)){
  it <- 100*(which(unique(sessions_storting$id) == x) / length(unique(sessions_storting$id)))
  cat(paste0(sprintf("Progress: %.4f%%", it), "\r"))

  a[[x]] <- get_session_decisions(sessionid = x, good_manners = 0)
}

## Progress: 9.0909%           Progress: 18.1818%           Progress: 27.2727%           Progress: 36.3636%
all_decisions <- do.call("rbind", a)

rm(a)

## Lets filter it all out (this is not done properly because this is just
## showing how to accidentally stumble across something usable)

UN_decisions <- all_decisions %>%
  mutate(decision_text_l = str_to_lower(decision_text),
         title_text_l = str_to_lower(decision_title)) %>%
  filter(str_detect(decision_text_l, "fn-sambandet") |
        str_detect(title_text_l, "fn-sambandet") |
        str_detect(decision_text_l, "utdanning for bærekraftig utvikling") |
        str_detect(title_text_l, "utdanning for bærekraftig utvikling") |
        str_detect(decision_text_l, "unicef") |
        str_detect(decision_text_l, "wfp") |
        str_detect(decision_text_l, "unaids") |
        str_detect(title_text_l, "unicef") |
        str_detect(title_text_l, "wfp") |
        str_detect(title_text_l, "unaids"))
```

```
## Omg its like only 15 stuffs, but what are they?? lets get the case

a<-list()

for (x in UN_decisions$case_id) {
  it <- 100*(which(unique(UN_decisions$case_id) == x) / length(unique(UN_decisions$case_id)))
  cat(paste0(sprintf("Progress: %.4f%%", it), "\r"))

  try(a[[x]] <- get_case_fancy(caseid = x, good_manners = 0))
}
```

```
## Progress: 6.6667%      Error in GET(url) : could not find function "GET"
## Progress: 13.3333%     Error in GET(url) : could not find function "GET"
## Progress: 20.0000%     Error in GET(url) : could not find function "GET"
## Progress: 26.6667%     Error in GET(url) : could not find function "GET"
## Progress: 33.3333%     Error in GET(url) : could not find function "GET"
## Progress: 40.0000%     Error in GET(url) : could not find function "GET"
## Progress: 46.6667%     Error in GET(url) : could not find function "GET"
## Progress: 53.3333%     Error in GET(url) : could not find function "GET"
## Progress: 60.0000%     Error in GET(url) : could not find function "GET"
## Progress: 66.6667%     Error in GET(url) : could not find function "GET"
## Progress: 73.3333%     Error in GET(url) : could not find function "GET"
## Progress: 80.0000%     Error in GET(url) : could not find function "GET"
## Progress: 86.6667%     Error in GET(url) : could not find function "GET"
## Progress: 93.3333%     Error in GET(url) : could not find function "GET"
## Progress: 100.0000%    Error in GET(url) : could not find function "GET"
```

```
### we got the decisions, woo.
```

```
### why is this here? who knows? its not used later, do with iteth whateth
### thoueth pleatheth.
```

```
b <- list()

for (x in unique(UN_decisions$case_id)) {
  b[[x]] <- a[[x]][["root"]]
}

### This seems to capture all budgets, which we can still use.

UN_cases <- do.call("rbind", b)
```

Inside the decision text we find the full budgets relevant to our area. By applying the magix of **regex** we can discover hitherto unknown treasures of information. (meaning we can find how much money each org gets every year, which would probably take us about 15 minutes to do manually, but thanks to the magic of automation we can turn a 15 minute task into a 3 day task, because we're programmers)

```
## First i write a new function to retrieve strings next to the target words
## we are interested in.

grab_text <- function(text, target, before, after){
  min <- which(unlist(map(str_split(text, "\\s"), ~grepl(target, .x))))-before
  max <- which(unlist(map(str_split(text, "\\s"), ~grepl(target, .x))))+after

  paste(str_split(text, "\\s")[[1]][min:max], collapse = " ")
}
```

```

}

## The text has been so dirty, we need to give it a good clean ;)

for (i in 1:length(UN_decisions$decision_text_1)) {
  tmp1 <- read_html(UN_decisions$decision_text_1[i])

  tmp2 <- html_text(tmp1)

  UN_decisions$cleanedtext_1[i] <- str_squish(str_remove_all(tmp2, "\r\n"))
}

## Lets make som empty lists, two sets of them even.

unicefbudget <- list()
wfpbudget <- list()
unaidsbudget <- list()
undpbudget <- list()

unicefbudgetn <- list()
wfpbudgetn <- list()
unaidsbudgetn <- list()
undpbudgetn <- list()

## Empty, like my skull

### Lets look at what UNICEF gets

for (x in 1:length(UN_decisions$cleanedtext_1)) {
  if(str_detect(UN_decisions$cleanedtext_1[x], "unicef")){
    unicefbudget[[x]] <- grab_text(text = UN_decisions$cleanedtext_1[x], target = "unicef", before = 1, after = 1)

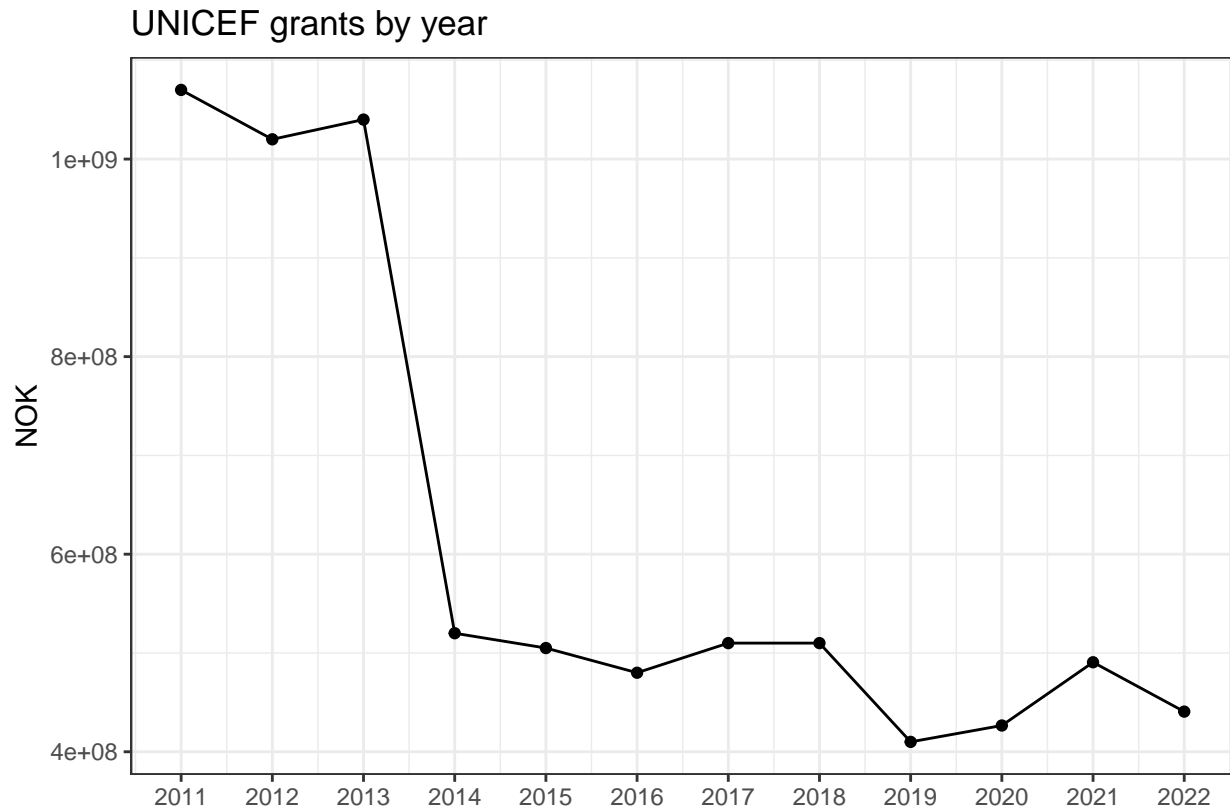
    if(str_detect(unicefbudget[[x]], "reduseres")){
      tmpred <- grab_text(text = unicefbudget[[x]], target = "til", before = 0, after = 4)
      tmp1 <- str_extract(tmpred, "[[:digit:]]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")
      unicefbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
    }else{
      tmp1 <- str_extract(unicefbudget[[x]], "([[:digit:]]{1,3}\\s?)[[:digit:]]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}\\s?")
      unicefbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
    }
  }else{next}
}

unicefbudgets_t <- do.call("rbind", unicefbudgetn)

unicefbudget_f <- data.frame(Year = c(2022,2021,2020,2019,2018,2017,
                                     2016,2015,2014,2013,2012,2011),
                           UnicefMoney = unicefbudgets_t)

```

```
ggplot(unicefbudget_f, aes(x = Year, y = UnicefMoney)) +
  geom_point() +
  geom_line() +
  theme_bw() +
  labs(title = "UNICEF grants by year", y = "NOK", x = "") +
  scale_x_continuous(breaks = c(2011, 2012, 2013, 2014, 2015, 2016,
                                2017, 2018, 2019, 2020, 2021, 2022))
```



By the looks of this, I would say that SV in government is associated with more money for UNICEF.

Lets look at WFP.

```
for (x in 1:length(UN_decisions$cleanedtext_l)) {
  if(str_detect(UN_decisions$cleanedtext_l[x], "wfp")){
    wfpbudget[[x]] <- grab_text(text = UN_decisions$cleanedtext_l[x], target = "wfp", before = 1, after = 1)
    if(str_detect(wfpbudget[[x]], "reduseres")){
      tmpred <- grab_text(text = wfpbudget[[x]], target = "til", before = 0, after = 4)
      tmp1 <- str_extract(tmpred, "[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")
      wfpbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
    }else{
      if(str_detect(wfpbudget[[x]], "auka")){
        tmpinc <- grab_text(text = UN_decisions$cleanedtext_l[x], target = "wfp", before = 1, after = 20)
        tmpinc2 <- grab_text(text = tmpinc, target = "til", before = 0, after = 5)
        tmpinc3 <- str_extract(tmpinc2, "[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")
        wfpbudgetn[[x]] <- (as.numeric(str_remove_all(tmpinc3, " ")))
      }
    }
  }
}
```

```

    }else{

      tmp1 <- str_extract(wfpbudget[[x]], "([:digit:]{1,3}\\s?){:digit:}{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}\\s?")

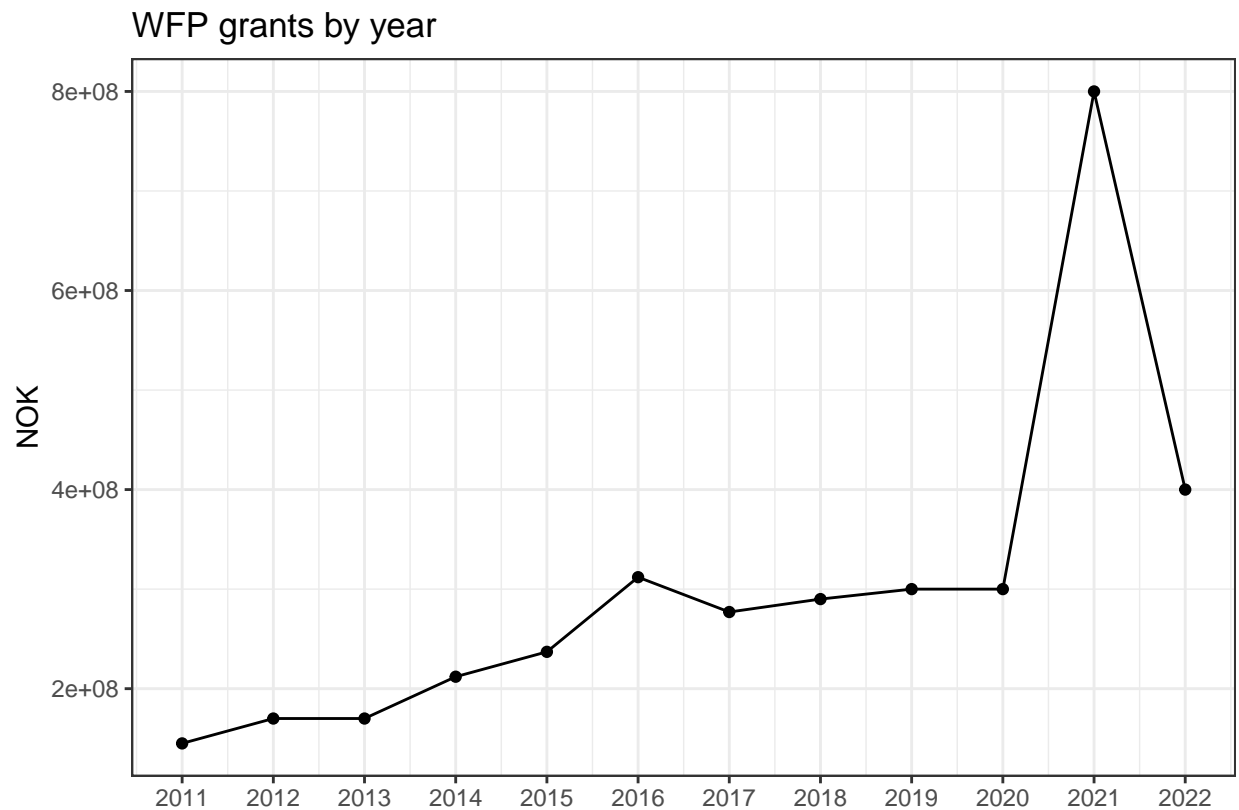
      wfpbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
    }
  }else{next}
}

wfpbudget_t <- do.call("rbind", wfpbudgetn)

wfpbudget_f <- data.frame(Year = c(2022,2021,2020,2019,2018,2017,
                                   2016,2015,2014,2013,2012,2011),
                          WFPMoney = wfpbudget_t)

ggplot(wfpbudget_f, aes(x = Year, y = WFPMoney)) +
  geom_point() +
  geom_line() +
  theme_bw() +
  labs(title = "WFP grants by year", y = "NOK", x = "") +
  scale_x_continuous(breaks = c(2011, 2012, 2013, 2014, 2015, 2016,
                                2017, 2018, 2019, 2020, 2021, 2022))

```



Looks like the peace prize was helpful.

Next we look at UNDP.

```
for (x in 1:length(UN_decisions$cleanedtext_1)) {
  if(str_detect(UN_decisions$cleanedtext_1[x], "undp")){
    undpbudget[[x]] <- grab_text(text = UN_decisions$cleanedtext_1[x], target = "undp", before = 1, after = 1)
    if(str_detect(undpbudget[[x]], "reduseres")){
      tmpred <- grab_text(text = undpbudget[[x]], target = "til", before = 0, after = 4)
      tmp1 <- str_extract(tmpred, "[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")
      undpbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
    }else{

      if(str_detect(undpbudget[[x]], "auka")){
        tmpinc <- grab_text(text = UN_decisions$cleanedtext_1[x], target = "undp", before = 1, after = 1)
        tmpinc2 <- grab_text(text = tmpinc, target = "til", before = 0, after = 5)
        tmpinc3 <- str_extract(tmpinc2, "[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")

        undpbudgetn[[x]] <- (as.numeric(str_remove_all(tmpinc3, " ")))

      }else{

        tmp1 <- str_extract(undpbudget[[x]], "[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")

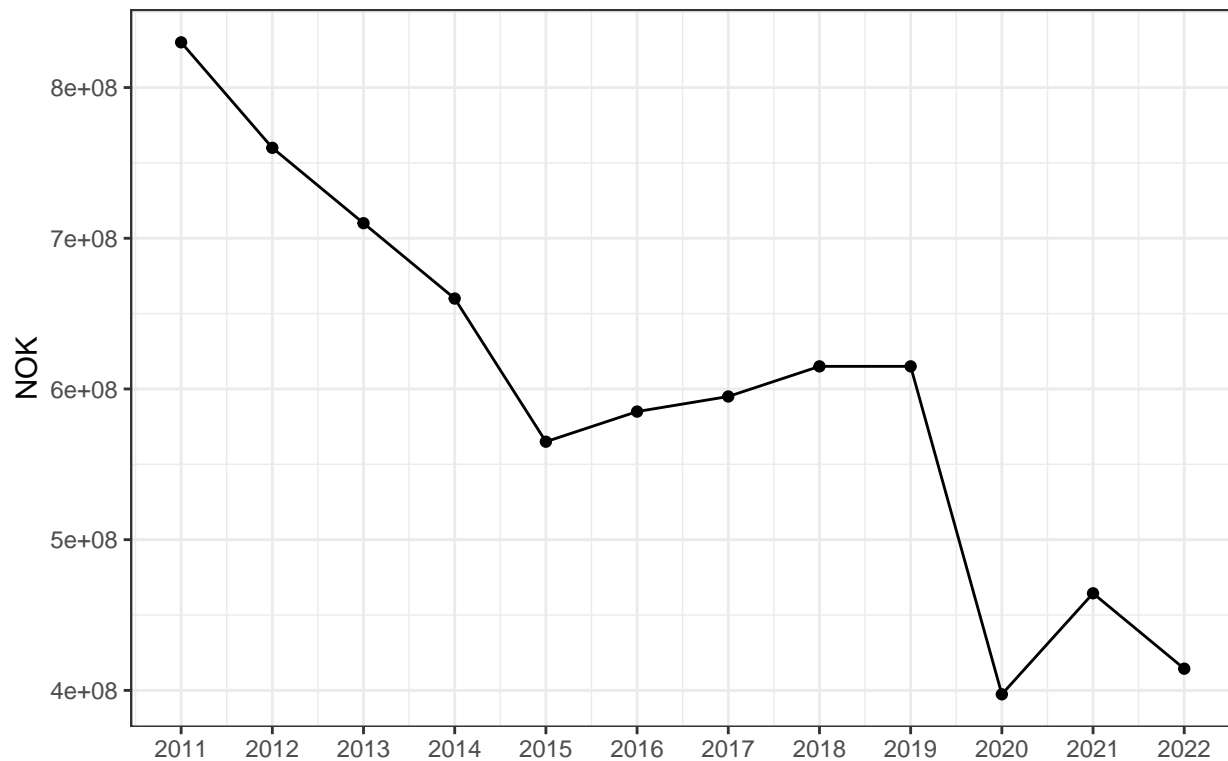
        undpbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
      }
    }else{next}
  }
}

undpbudget_t <- do.call("rbind", undpbudgetn)

undpbudget_f <- data.frame(Year = c(2022,2021,2020,2019,2018,2017,
                                   2016,2015,2014,2013,2012,2011),
                          UNDPMoney = undpbudget_t)

ggplot(undpbudget_f, aes(x = Year, y = UNDPMoney)) +
  geom_point() +
  geom_line() +
  theme_bw() +
  labs(title = "UNDP grants by year", y = "NOK", x = "") +
  scale_x_continuous(breaks = c(2011, 2012, 2013, 2014, 2015, 2016,
                                2017, 2018, 2019, 2020, 2021, 2022))
```

UNDP grants by year



Looks like decreasing the budget is a popular thing. Maybe its not all mean old Støre that wants to reduce it.

Lets finally have a gander at UNAIDS grants for every year.

```
for (x in 1:length(UN_decisions$cleanedtext_1)) {
  if(str_detect(UN_decisions$cleanedtext_1[x], "unaids")){
    unaidsbudget[[x]] <- grab_text(text = UN_decisions$cleanedtext_1[x], target = "unaids", before = 1,
    if(str_detect(unaidsbudget[[x]], "reduseres")){
      tmpred <- grab_text(text = unaidsbudget[[x]], target = "til", before = 0, after = 4)
      tmp1 <- str_extract(tmpred, "[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")
      unaidsbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
    }else{

      if(str_detect(unaidsbudget[[x]], "auka")){
        tmpinc <- grab_text(text = UN_decisions$cleanedtext_1[x], target = "unaids", before = 1, after = 1)
        tmpinc2 <- grab_text(text = tmpinc, target = "til", before = 0, after = 5)
        tmpinc3 <- str_extract(tmpinc2, "[:digit:]{1,3}\\s[:digit:]{1,3}\\s[:digit:]{1,3}")

        unaidsbudgetn[[x]] <- (as.numeric(str_remove_all(tmpinc3, " ")))

      }else{

        tmp1 <- str_extract(unaidsbudget[[x]], "([[:digit:]]{1,3}\\s?)[[:digit:]]{1,3}\\s[:digit:]{1,3}\\s0")

        unaidsbudgetn[[x]] <- (as.numeric(str_remove_all(tmp1, " ")))
      }
    }
  }else{next}
```



```

}
UN_decisions$decision_title

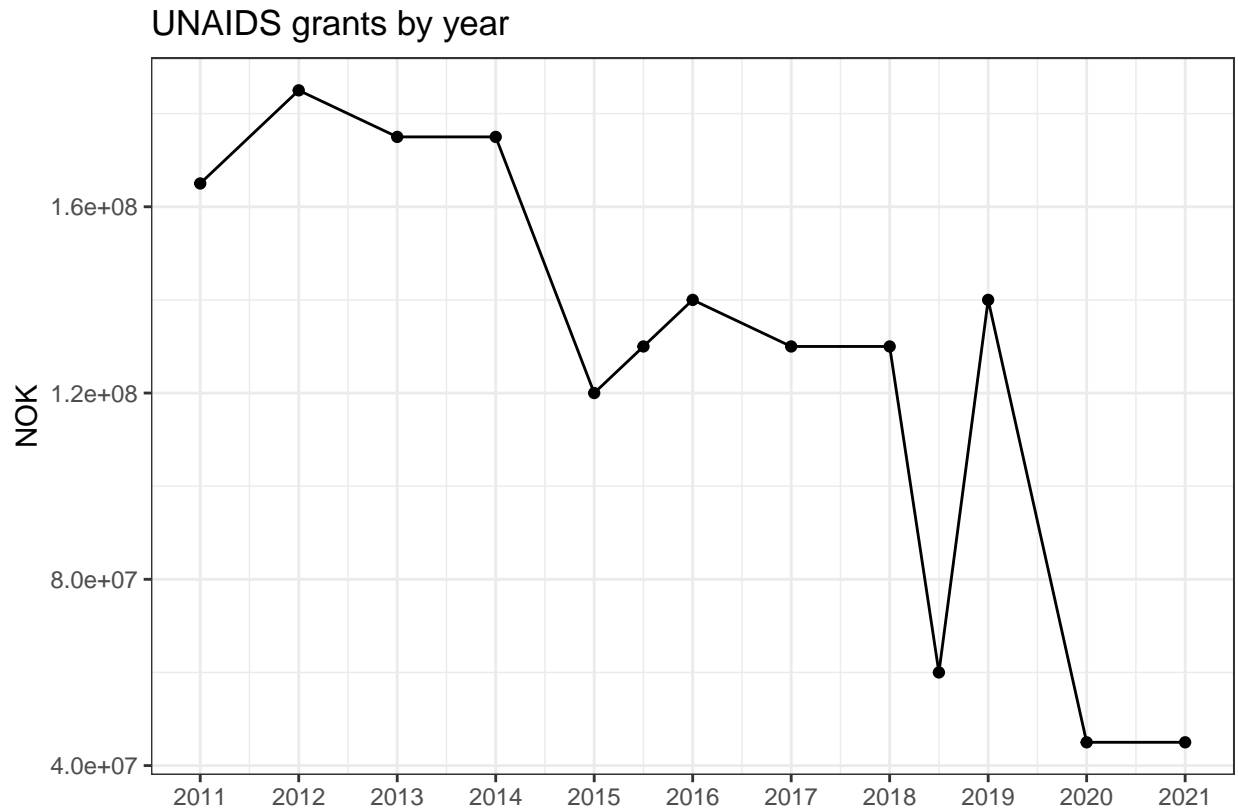
## [1] "Revidert nasjonalbudsjett 2022"
## [2] "Statsbudsjettet 2022 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [3] "Statsbudsjettet 2021 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [4] "Endringer i statsbudsjettet 2019 under Utanriksdepartementet"
## [5] "Statsbudsjettet 2020 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [6] "Statsbudsjettet 2019 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [7] "Statsbudsjettet 2018 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [8] "Endringer i statsbudsjettet 2016 under Utanriksdepartementet"
## [9] "Statsbudsjettet 2017 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [10] "Endringer i statsbudsjettet 2015 under Utanriksdepartementet"
## [11] "Statsbudsjettet 2016 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [12] "Statsbudsjettet 2015 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [13] "Statsbudsjettet 2014 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [14] "Statsbudsjettet 2013 (kapitler fordelt til utenriks- og forsvarskomiteen)"
## [15] "Statsbudsjettet for 2012 (kapitler fordelt til utenriks- og forsvarskomiteen)"

unaidsbudget_t <- do.call("rbind", unaidsbudgetn)

unaidsbudget_f <- data.frame(Year = c(2021,2020, 2019, 2018.5,2018,2017,
                                     2016, 2015.5,2015,2014,2013,2012,2011),
                             UNAIDSMoney = unaidsbudget_t)

ggplot(unaidsbudget_f, aes(x = Year, y = UNAIDSMoney)) +
  geom_point() +
  geom_line() +
  theme_bw() +
  labs(title = "UNAIDS grants by year", y = "NOK", x = "") +
  scale_x_continuous(breaks = c(2011, 2012, 2013, 2014, 2015, 2016,
                                2017, 2018, 2019, 2020, 2021, 2022))

```



Lots of supplementary budgets here. Weird that the negotiations in 2018 lead to a decrease.

Note that for all of these, Støres 2021 budget is coded as 2022, purely out of convenience.

Overall this provides a possibility of overviewing the budgets of the UN organizations, that are gathered in an automated manner. Although, given the detailed regex controls it is arguable whether it is truly “automatic”, nonetheless it can be reused in the future to gather even more data simply by adding the session in the start of the code.