

# STATS 112 — Text Analysis of Student Reflections on *In the Heart of Another*

Wonhyeok Choi, Ryan Chu, Jessica Nguyen

2025-12-02

## I. Abstract

This study analyzes 43 student reflection papers in response to Susan Philips' talk and her book *In the Heart of Another*, which explores migration, identity, and social justice across multiple global contexts. Student reflections, categorized by regions (Middle East, South America, Asia, and a general group) were compared to the text of the book chapters corresponding to El Salvador, Mexico, Armenia, China, and Vietnam. Using text mining techniques, including term frequency analysis, keyness analysis, word clouds, and Latent Dirichlet Allocation (LDA) topic modeling, the study identifies the primary themes emphasized by students and their alignment with the book's content. Results indicate that students consistently highlight themes of migration, identity, family, community, and social justice, with notable regional differences in emphasis. The findings demonstrate that while student reflections resonate with the central topics of Philips' work, they often emphasize personal and social dimensions over historical or contextual details. The study underscores the utility of text mining for examining qualitative student data and suggests that larger datasets and complementary surveys could provide deeper insights into student engagement with global societal issues.

## II. Statement of the Problem

Understanding how students engage with complex global issues, such as migration, identity, and social justice, is a central goal of courses like Stats 112, which incorporate reflection papers as a pedagogical tool. While Susan Philips' book *In the Heart of Another* provides a rich exploration of these topics through personal narratives from multiple regions, there is limited systematic analysis of how students interpret and respond to this material. Specifically, it is unclear which themes resonate most strongly with students, whether these responses vary by regional context, and how closely student reflections align with the book's content. This project addresses this gap by applying text mining and topic modeling techniques to 43 student reflection papers, comparing them with the corresponding chapters of the book, to uncover patterns in student engagement, thematic emphasis, and regional differences in interpretation.

## III. Variables in the Study

### 1. Independent variables

- Chapter/Region of the book: each chapter of *In the Heart of Another* focuses on a different region or set of narratives (El Salvador, Mexico, Armenia, China, Vietnam). This variable represents the source material that students responded to in their reflections.

- Student Focus/Folder: the reflections are organized by the main country of focus in each paper (Middle East, South America, Asia, and General). This variable allows for comparison of student responses across different region contexts.

## 2. Dependent variables

- Word frequency: The count of each word appearing in student reflections and book chapters, used to identify which terms or themes are most common.
- Keyness / Text Similarity: the statistical significance of word usage differences between reflections and corresponding book chapters.
- Topic Distribution: The mixture of topics present in each reflection, as identified by Latent Dirichlet Allocation (LDA) topic modeling.
- Thematic Alignment: The degree to which student reflections align with the major themes and narratives presented in the book.

## 3. Derived Variables

- Stemmed Words: Words reduced to their root form to consolidate variants (e.g., migrate, migrating → migr).
- Document-Term Matrices (DTM): Matrices representing the frequency of terms across reflections or book chapters, which serve as the basis for keyness analysis and topic modeling.

## IV. Exploratory Data Analysis

```
# plot top terms
plot_top_terms <- function(dfm_obj, n = 20, title = "") {
  tf <- topfeatures(dfm_obj, n)
  df <- data.frame(word = names(tf), freq = as.numeric(tf))

  ggplot(df, aes(x = reorder(word, freq), y = freq)) +
    geom_col(fill = "steelblue") +
    coord_flip() +
    theme_bw() +
    labs(title = title, x = "Word", y = "Frequency")
}

# wordclouds
plot_wordcloud_dfm <- function(dfm_obj, title = "") {
  freq <- sort(colSums(dfm_obj), decreasing = TRUE)
  wordcloud(
    names(freq), freq,
    min.freq = 2,
    max.words = 100,
    colors = brewer.pal(8, "Dark2")
  )
  title(main = title)
}

# plot keyness comparison
plot_keyness_comparison <- function(ref_dfm, chap_dfm, plot_title) {
```

```

# align dfms
chap_dfm <- dfm_match(chap_dfm, features = featnames(ref_dfm))

# combine
combined <- rbind(ref_dfm, chap_dfm)

# want chapter rows
target <- c(rep(FALSE, nrow(ref_dfm)), rep(TRUE, nrow(chap_dfm)))

# compute keyness
key <- textstat_keyness(combined, target = target)

# create plot object
p <- textplot_keyness(key, color = c("darkblue", "firebrick3"), labelsize = 3)

# print
print(p)

# add title
grid::grid.text(plot_title,
                y = unit(1, "npc") - unit(2, "lines"),
                gp = grid::gpar(fontsize = 16, fontface = "bold"))

# analysis
cat("\n\n---\n### Commentary on Keyness:", plot_title, "\n\n")
cat("*Write commentary here.)*\n---\n\n")
}

# load data
general_ref <- readtext("Reflections/General/*.txt")
middleeast_ref <- readtext("Reflections/Middle East/*.txt")
southamerica_ref <- readtext("Reflections/South America/*.txt")
asia_ref <- readtext("Reflections/Asia/*.txt")

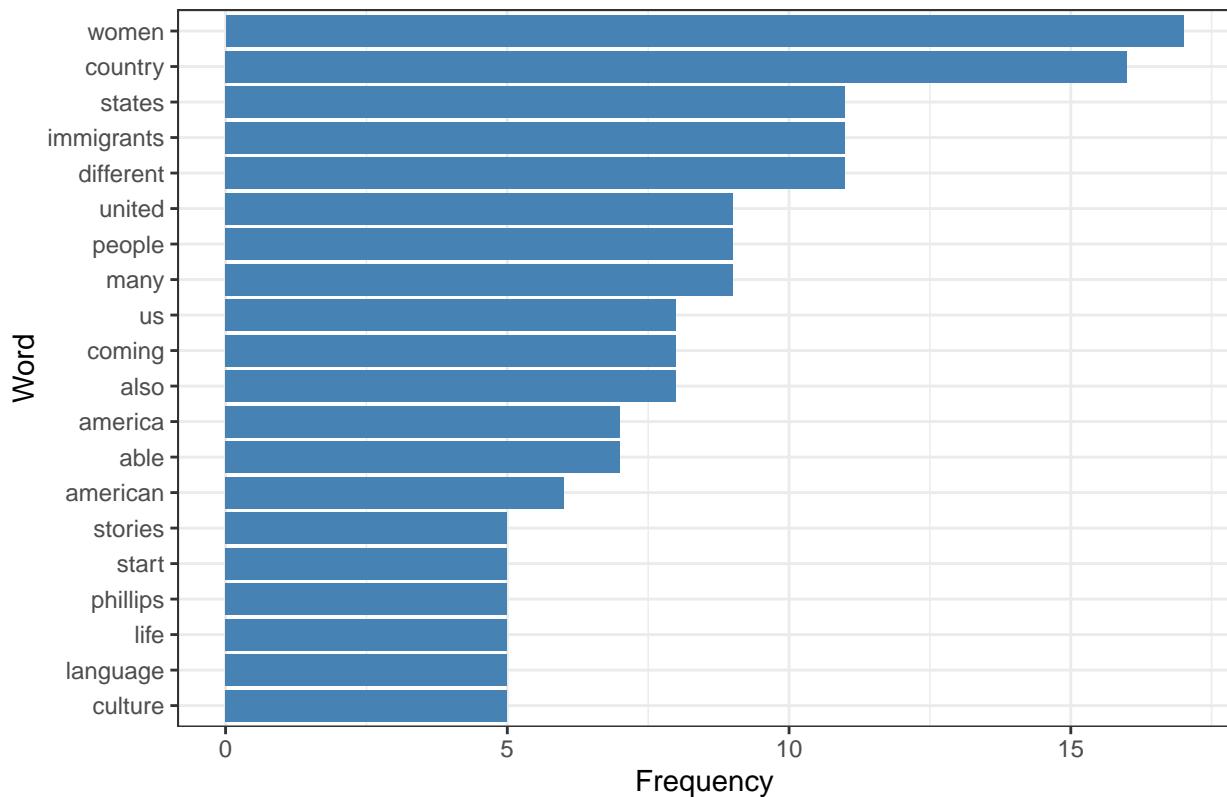
# clean data
general_ref$text <- clean_text_simple(general_ref$text)
middleeast_ref$text <- clean_text_simple(middleeast_ref$text)
southamerica_ref$text <- clean_text_simple(southamerica_ref$text)
asia_ref$text <- clean_text_simple(asia_ref$text)

# by region
general_dfm      <- create_dfm(general_ref$text)
middleeast_dfm   <- create_dfm(middleeast_ref$text)
southamerica_dfm <- create_dfm(southamerica_ref$text)
asia_dfm        <- create_dfm(asia_ref$text)

# frequency tables
plot_with_commentary(
  plot_top_terms(general_dfm, title = "Top Words: General Reflections",
                 title = "General Reflections",
                 commentary = "These reflections emphasize personal dimensions of global themes. Terms like culture, la
)

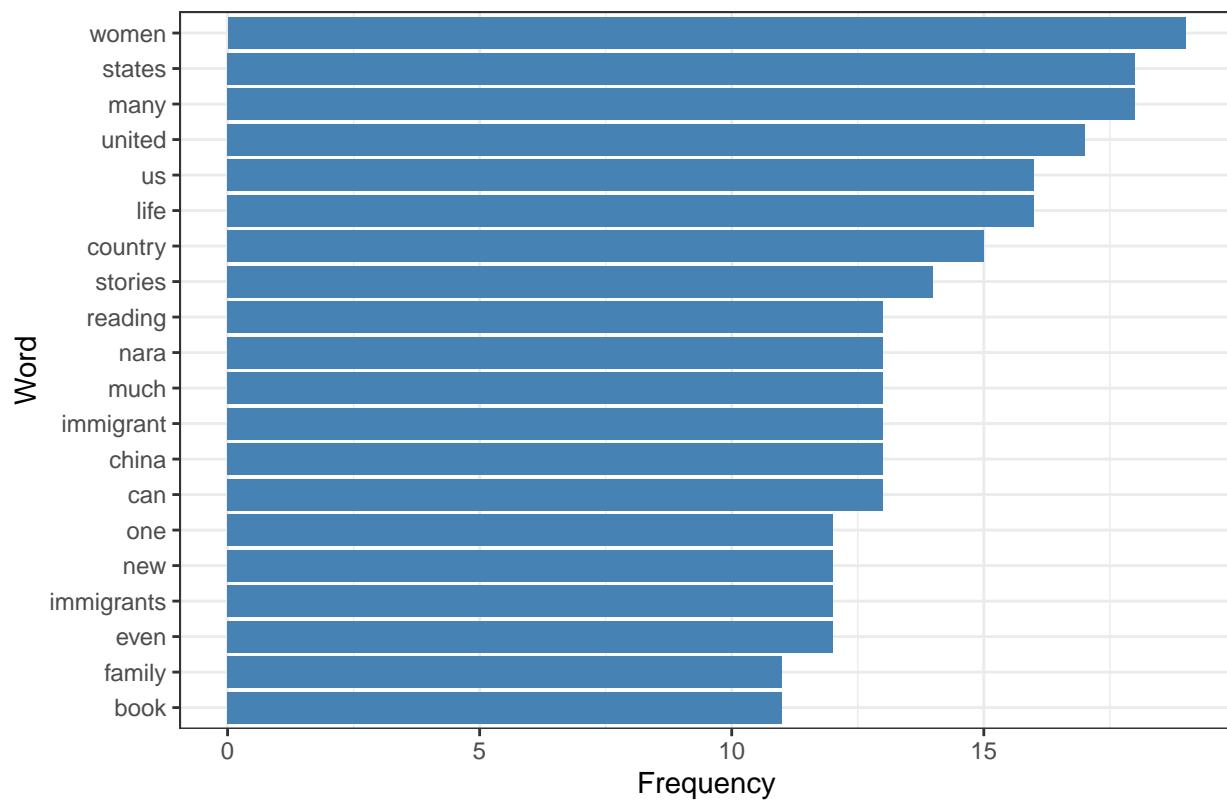
```

## Top Words: General Reflections



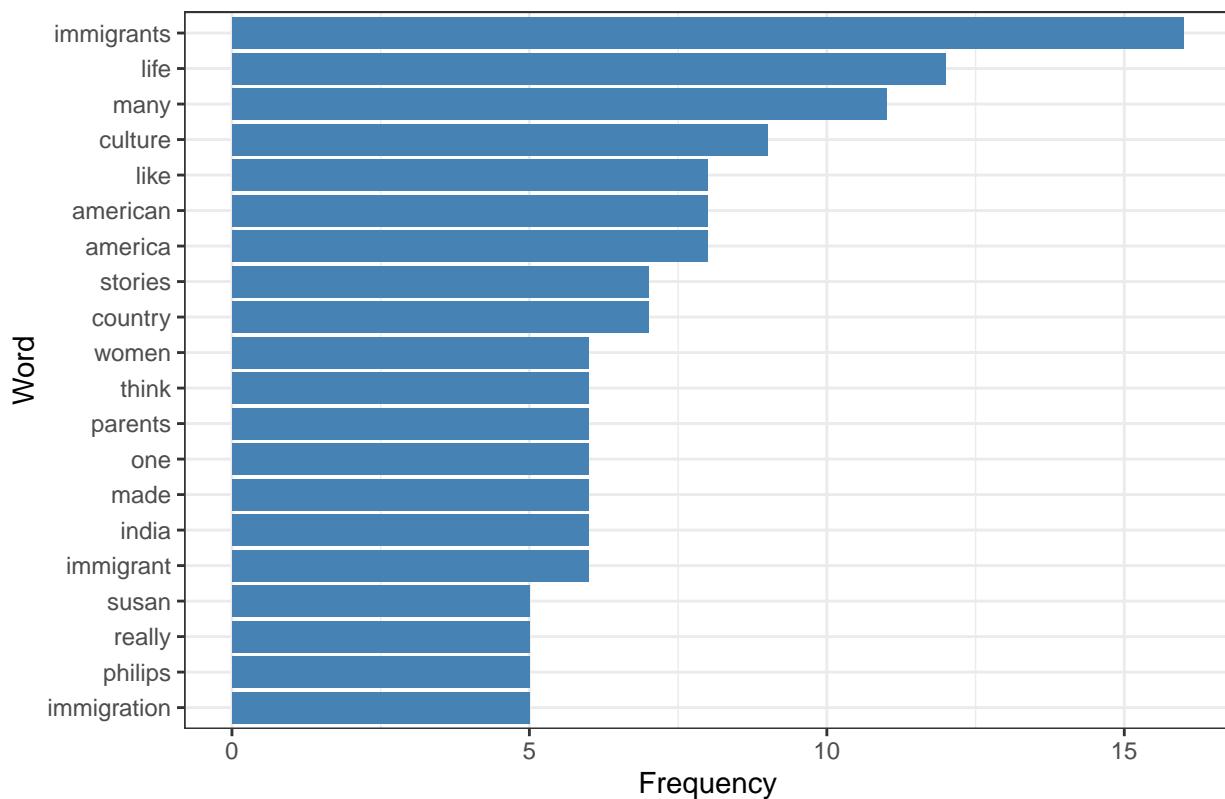
```
##  
## ---  
## #### Commentary on: General Reflections  
##  
## These reflections emphasize personal dimensions of global themes. Terms like  
## culture, language, life, and people dominate, while region-specific or  
## historical markers are absent. This suggests that many students interpreted the  
## material through a general U.S. immigrant lens rather than a Middle-East or  
## South-American frame. Notably, references to women and states imply that  
## identity and mobility are discussed at a societal level, but lack the  
## geopolitical specificity present in Philips' chapters. Overall, the vocabulary  
## reflects emotional resonance, personal relatability, and cultural reflection  
## rather than historical analysis.  
## ---  
  
plot_with_commentary(  
  plot_top_terms(middleeast_dfm, title = "Top Words: Middle East Reflections"),  
  "Middle East Reflections",  
  commentary = "While grouped under \"Middle East\", the vocabulary closely mirrors the generalized immigr")
```

## Top Words: Middle East Reflections



```
##  
##  
## ---  
## ### Commentary on: Middle East Reflections  
##  
## While grouped under "Middle East", the vocabulary closely mirrors the  
## generalized immigrant discourse. Words such as china, immigrants, family, and  
## stories appear far more frequently than Middle-Eastern named entities. This  
## highlights a potential disparity between folder labeling and thematic content,  
## implying that many students connected more strongly with East-Asian migration  
## narratives (e.g., Xidan's story). The consistent presence of family and women  
## aligns with Philips' interpersonal storytelling style, but contextual anchors  
## specific to the Middle East are missing, indicating focus on universal  
## migration struggles over localized frameworks.  
## ---  
  
plot_with_commentary(  
  plot_top_terms(southamerica_dfm, title = "Top Words: South America Reflections"),  
  "South America Reflections", commentary = "Here, students introduced slightly more specificity via ref")
```

## Top Words: South America Reflections



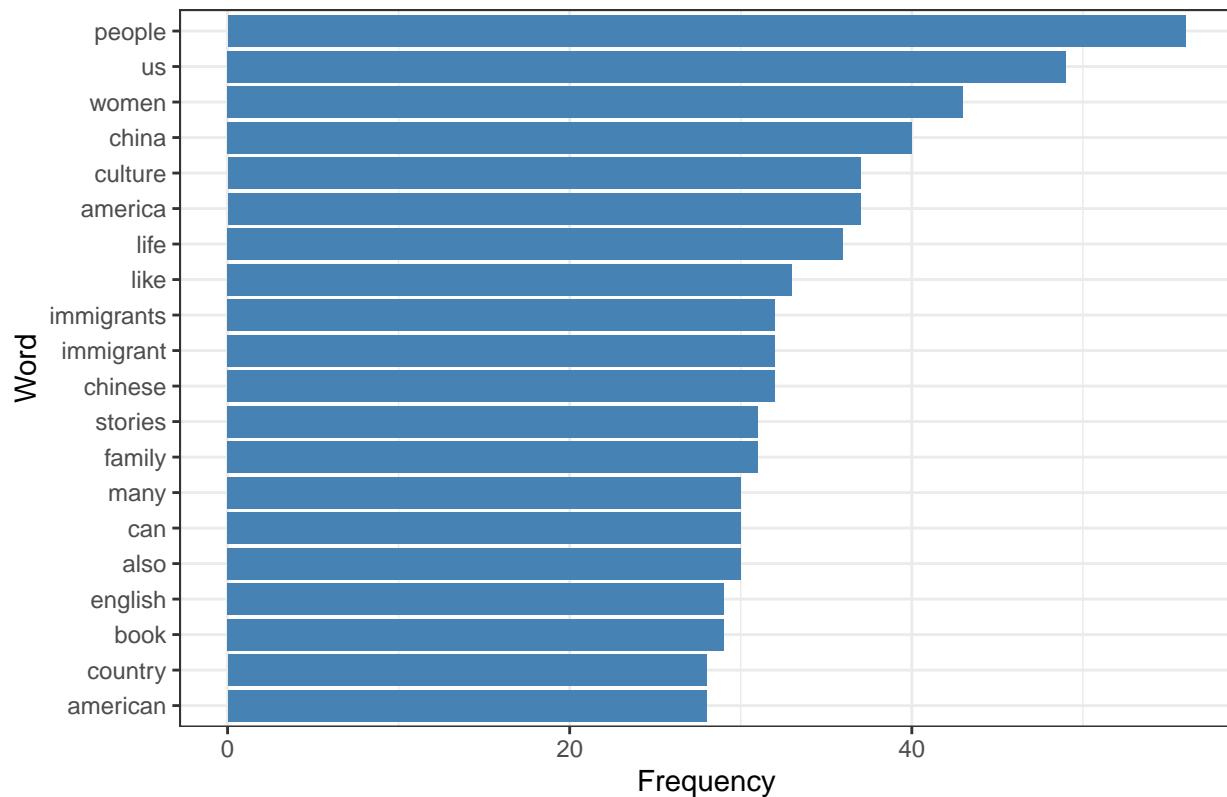
```

## 
## 
## --- 
## ### Commentary on: South America Reflections
## 
## Here, students introduced slightly more specificity via references to susan,
## philips, and america, demonstrating stronger engagement with the speaker than
## with the region's historical background. The prominence of culture, life, and
## many signals broad identity discourse, while immigrants appears more frequently
## than El Salvador, Mexico, or other exact regional terms. This indicates that
## students framed South-American migration as part of a collective American
## immigrant experience rather than anchoring their analysis in regional
## sociopolitical details. Their responses highlight cultural integration and
## identity formation, reinforcing emotional alignment with Philips' themes but
## diverging in depth of historical context.
## --- 

plot_with_commentary(
  plot_top_terms(asia_dfm, title = "Top Words: Asia Reflections"),
  "Asia Reflections", commentary = "This region shows the strongest presence of country-specific named entities"
)

```

### Top Words: Asia Reflections



```
##  
##  
## ---  
## #### Commentary on: Asia Reflections  
##  
## This region shows the strongest presence of country-specific named entities,  
## particularly China and terms like chinese and english, reflecting both cultural  
## identity and language barriers. Unlike other folders, china appears with high  
## frequency, implying that Asia-focused students cited concrete migration  
## contexts more often. However, the recurring terms women, culture, family, and  
## stories still indicate that personal narratives outweighed broader political or  
## historical discourse. Students therefore aligned strongly with Philips on core  
## migration themes, but centered their interpretation on individual experience  
## and cultural identity, especially linguistic challenges, rather than structural  
## migration forces  
## ---
```

```
# wordclouds  
plot_with_commentary(  
  plot_wordcloud_dfm(general_dfm, "General Reflections"),  
  "Word Cloud - General"  
)
```



```

## NULL
##
##
## ---
## #### Commentary on: Word Cloud - General
##
##
## ---
## ---


plot_with_commentary(
  plot_wordcloud_dfm(middleeast_dfm, "Middle East Reflections"),
  "Word Cloud - Middle East"
)

```

## Middle East Reflections



```
## NULL
##
##
## ---
## #### Commentary on: Word Cloud - Middle East
##
##
## ---
## ---
```

```
plot_with_commentary(
  plot_wordcloud_dfm(southamerica_dfm, "South America Reflections"),
  "Word Cloud - South America"
)
```

## **South America Reflections**

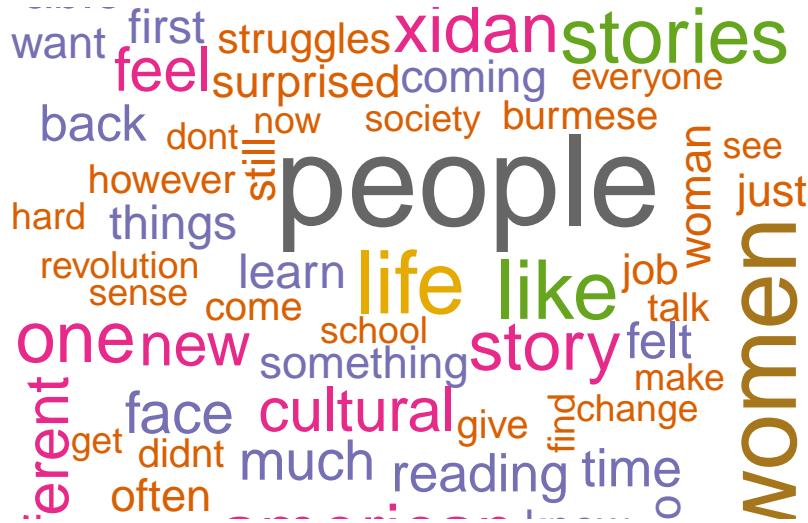


```
## NULL
##
##
## ---
## ### Commentary on: Word Cloud - South America
##
##
## ---
## ---
```

```
plot_with_commentary(
  plot_wordcloud_dfm(asia_dfm, "Asia Reflections")
  "Word Cloud - Asia"
)
```

## Asia Reflections



```
## NULL
##
##
## ---
## #### Commentary on: Word Cloud - Asia
##
##
## ---
```

Chapter Analysis

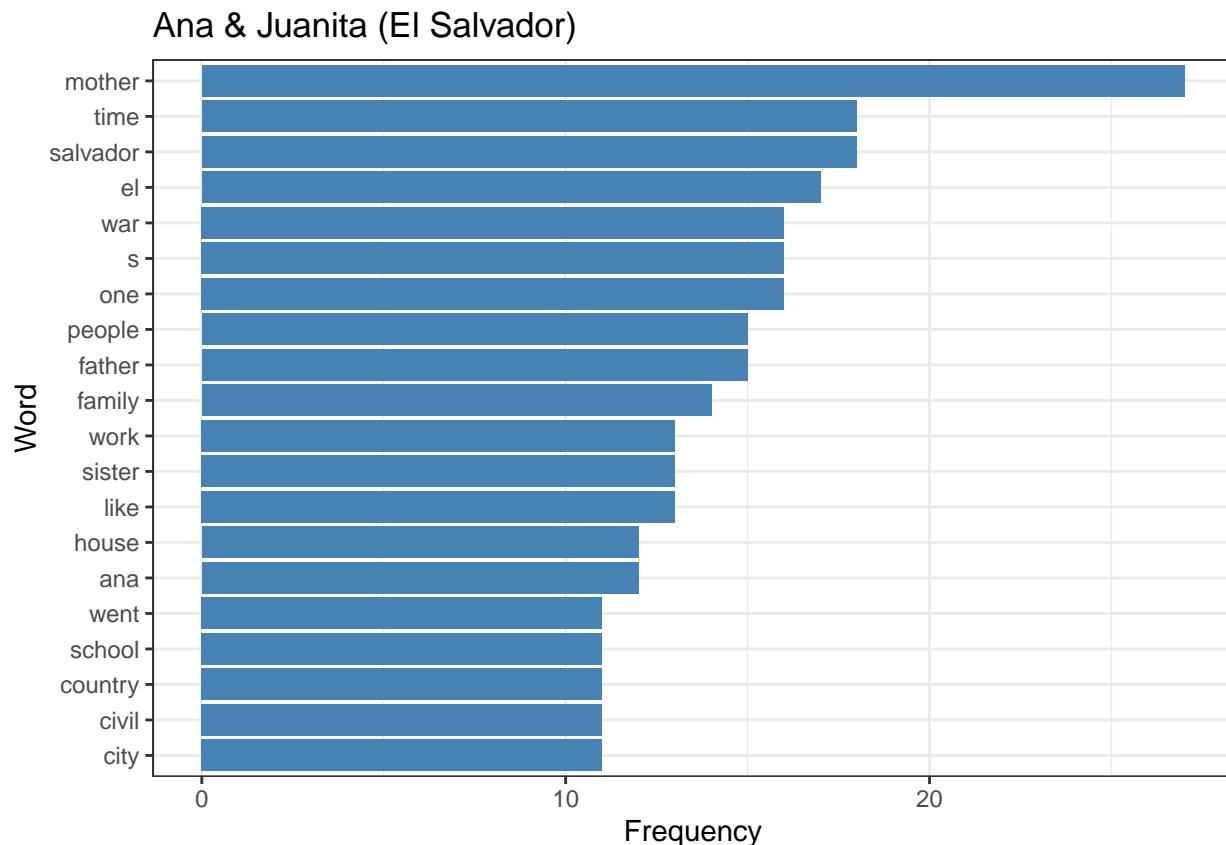
```
chapters <- list(
  ana      = readtext("In the Heart of Another/ana and juanita.txt"),
  marie    = readtext("In the Heart of Another/marie and michelle.txt"),
  nara     = readtext("In the Heart of Another/nara and oktiabrina.txt"),
  xidan    = readtext("In the Heart of Another/xidan.txt"),
  thuynga  = readtext("In the Heart of Another/thuynga.txt")
)

chapters <- lapply(chapters, function(x) {
  x$text <- clean_text_simple(x$text)
  x
})

chap_dfms <- lapply(chapters, function(df) create_dfm(df$text))
```

Top Terms in book chapters

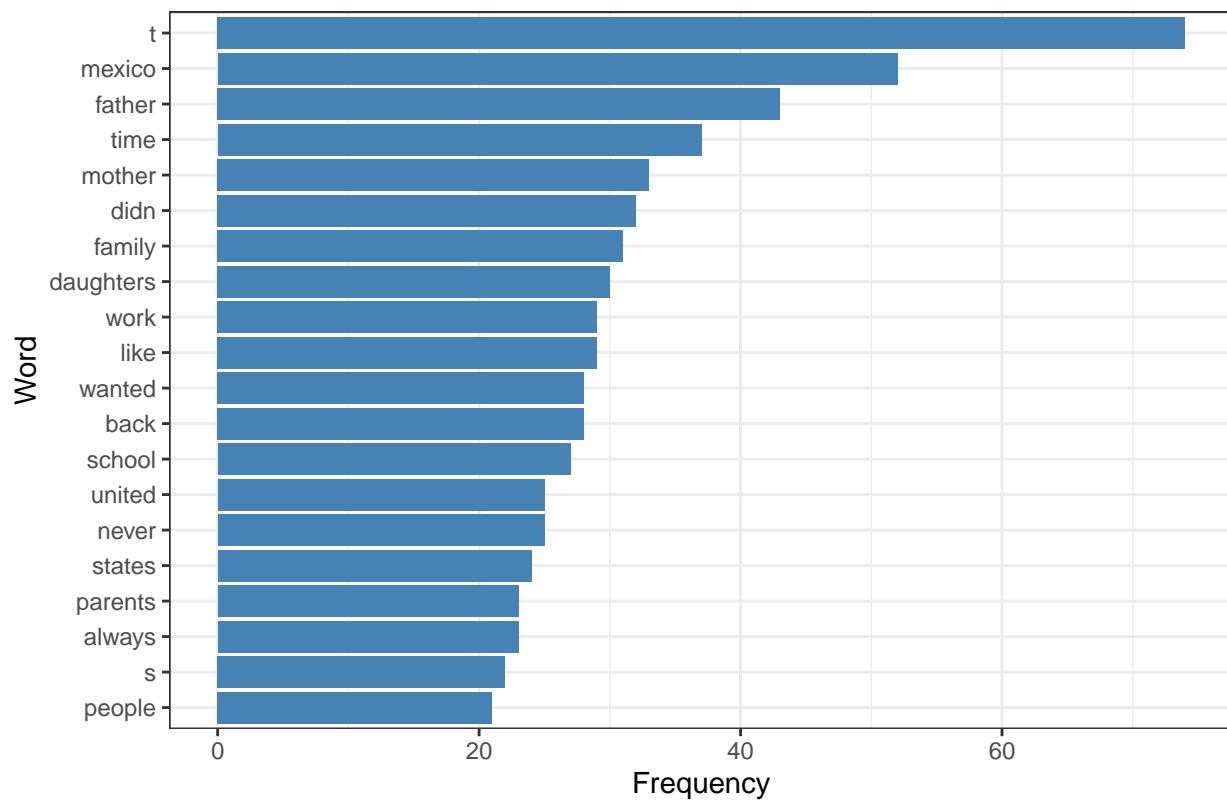
```
plot_with_commentary(  
  plot_top_terms(chap_dfms$ana, title = "Ana & Juanita (El Salvador)",  
  "Ana & Juanita"  
)
```



```
##  
##  
## ---  
## ### Commentary on: Ana & Juanita  
##  
##  
## ---
```

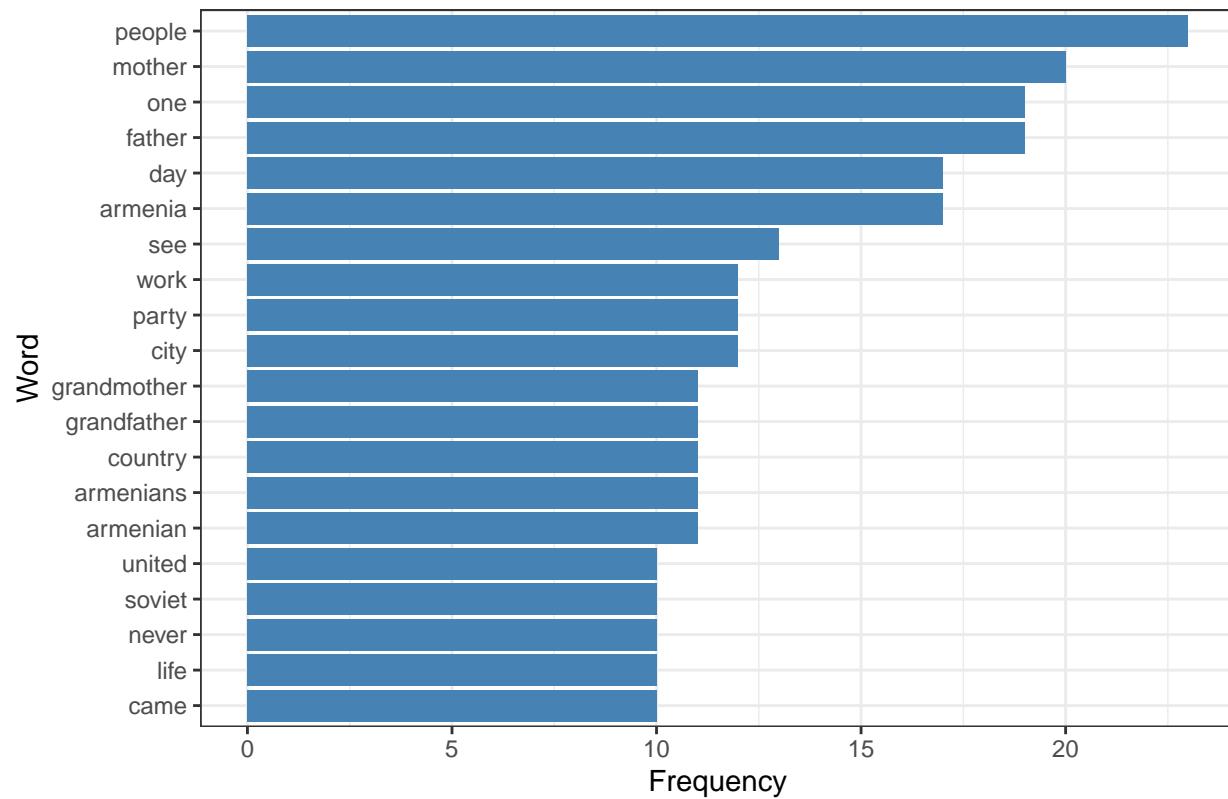
```
plot_with_commentary(  
  plot_top_terms(chap_dfms$marie, title = "Marie & Michelle (Mexico)",  
  "Marie & Michelle")
```

Marie & Michelle (Mexico)



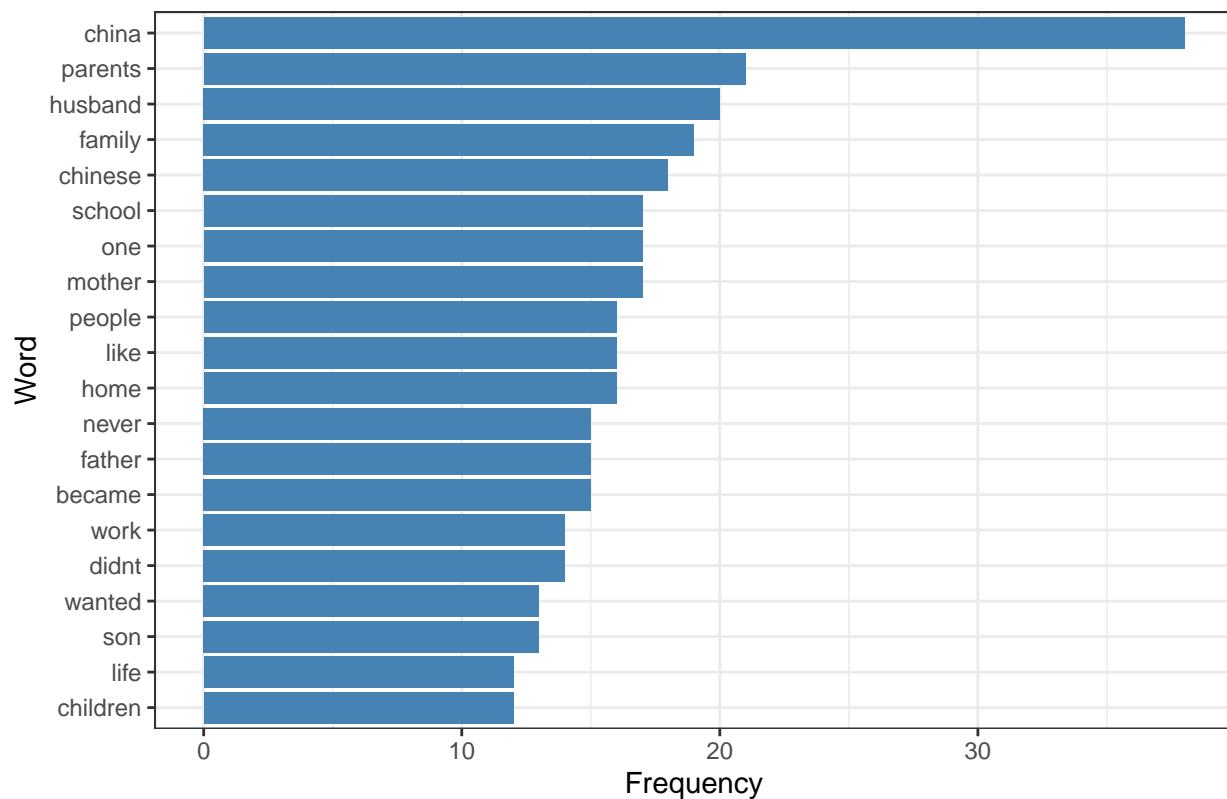
```
##  
##  
## ---  
## #### Commentary on: Marie & Michelle  
##  
##  
## ---  
  
plot_with_commentary(  
  plot_top_terms(chap_dfms$nara, title = "Nara & Oktiabrina (Armenia)",  
  "Nara & Oktiabrina"  
)
```

Nara & Oktiabrina (Armenia)

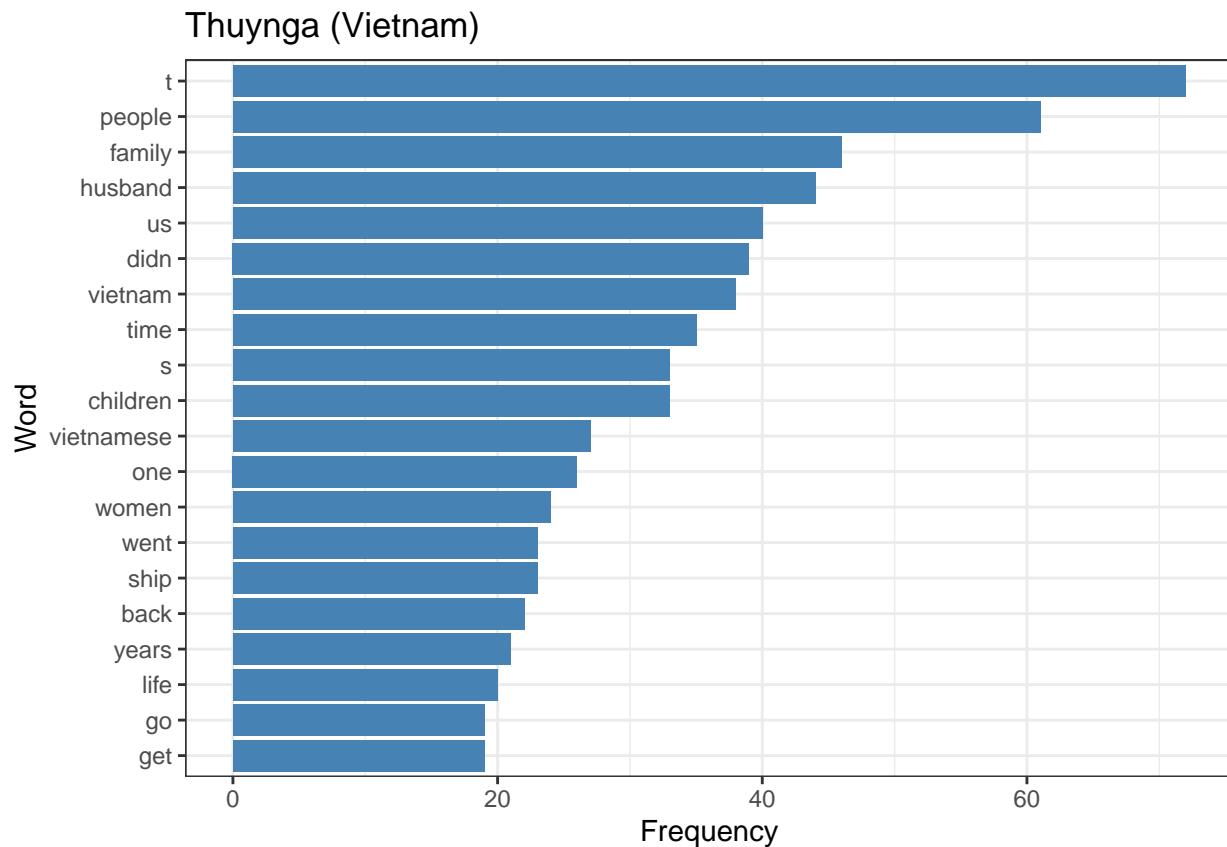


```
##  
##  
## ---  
## #### Commentary on: Nara & Oktiabrina  
##  
##  
## ---  
  
plot_with_commentary(  
  plot_top_terms(chap_dfms$xidan, title = "Xidan (China)",  
  "Xidan"  
)
```

Xidan (China)



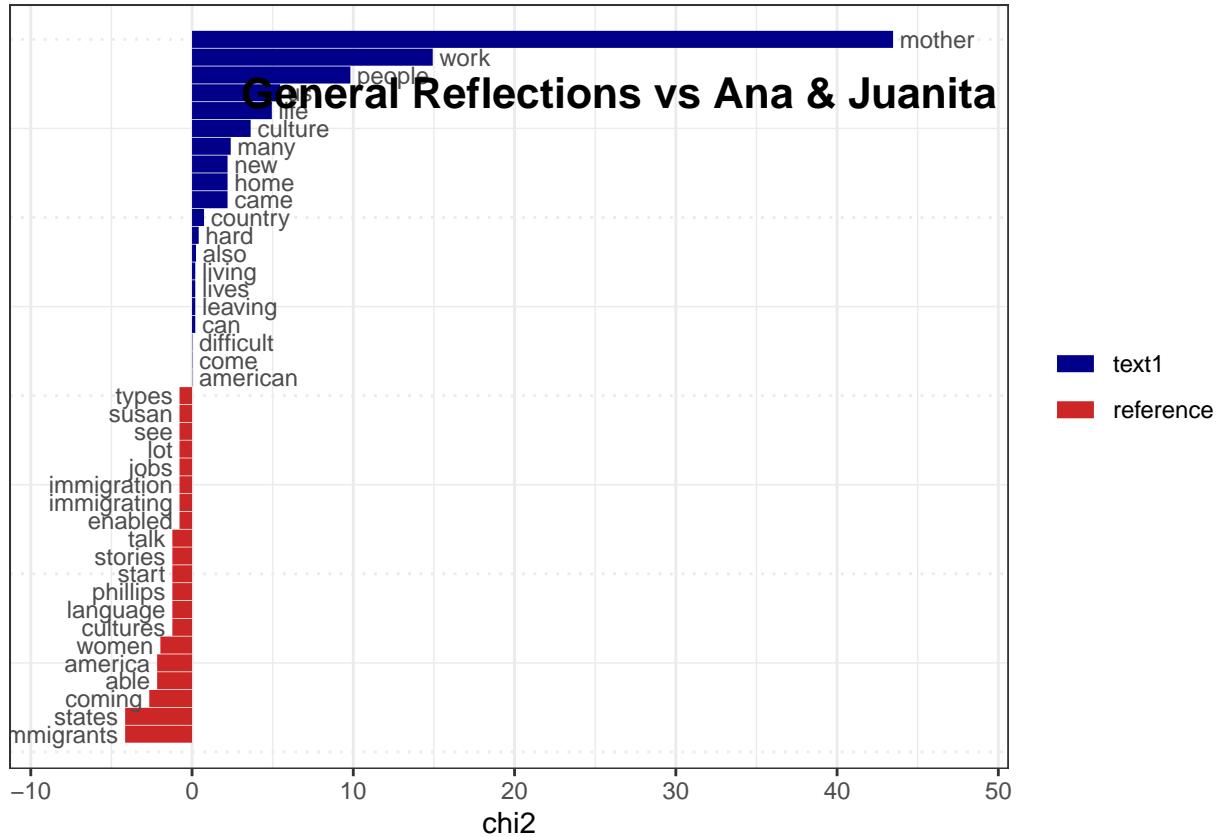
```
##  
##  
## ---  
## #### Commentary on: Xidan  
##  
##  
## ---  
  
plot_with_commentary(  
  plot_top_terms(chap_dfms$thuynga, title = "Thuynga (Vietnam)",  
  "Thuynga"  
)
```



```
##  
##  
## ---  
## #### Commentary on: Thuynga  
##  
##  
## ---
```

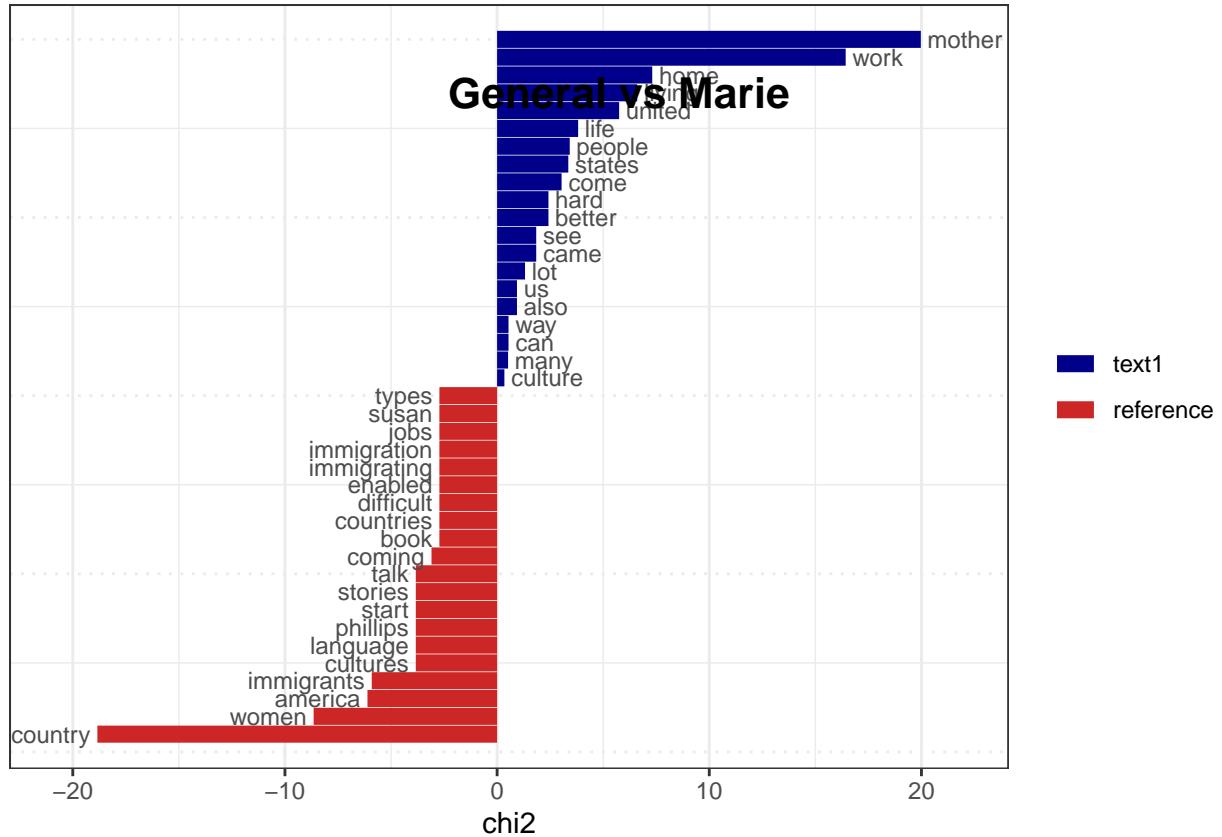
Keyness: Reflection vs. Chapter

```
plot_keyness_comparison(  
general_dfm,  
chap_dfms$ana,  
"General Reflections vs Ana & Juanita"  
)
```



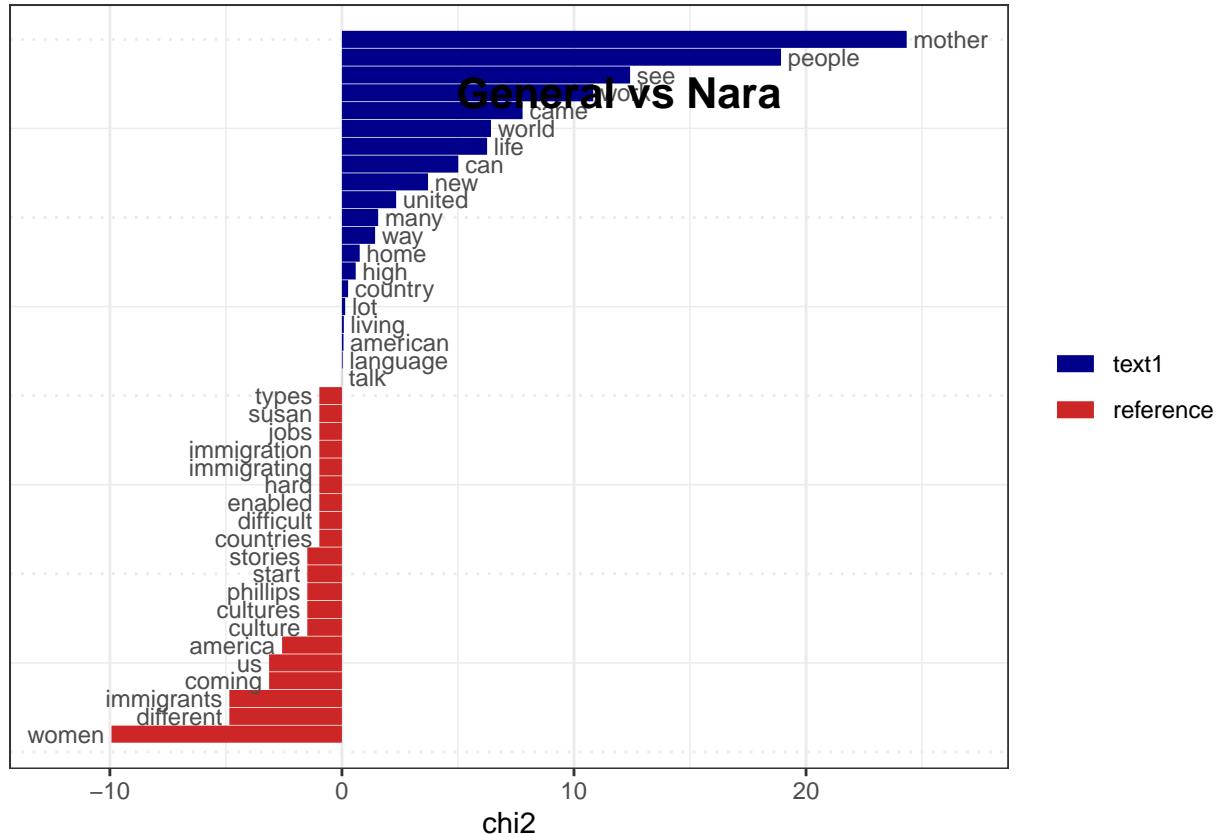
```
##  
##  
## ---  
## #### Commentary on Keynes: General Reflections vs Ana & Juanita  
##  
## *(Write commentary here.)*  
## ---
```

```
plot_keyness_comparison(general_dfm, chap_dfms$marie, "General vs Marie")
```

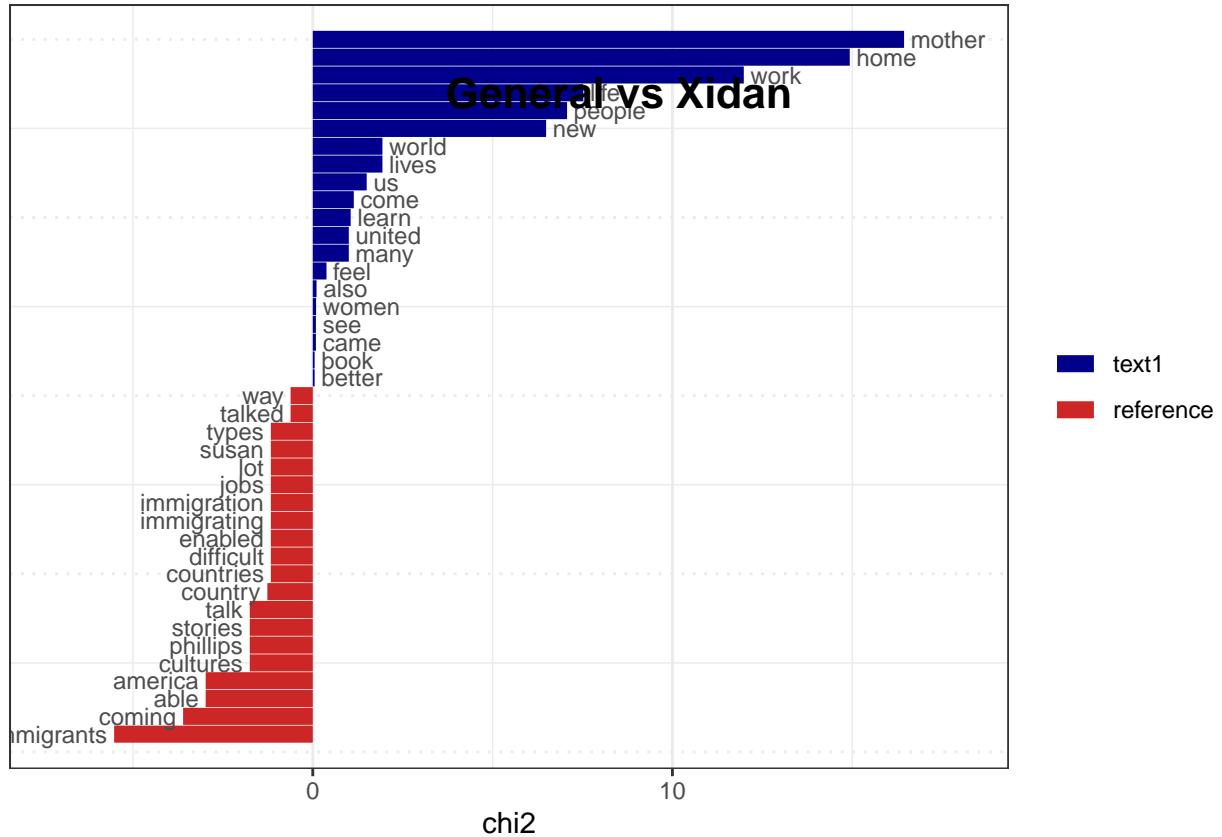


```
## 
## --- 
## #### Commentary on Keynes: General vs Marie 
## *(Write commentary here.)* 
## --- 

plot_keyness_comparison(general_dfm, chap_dfms$nara, "General vs Nara")
```

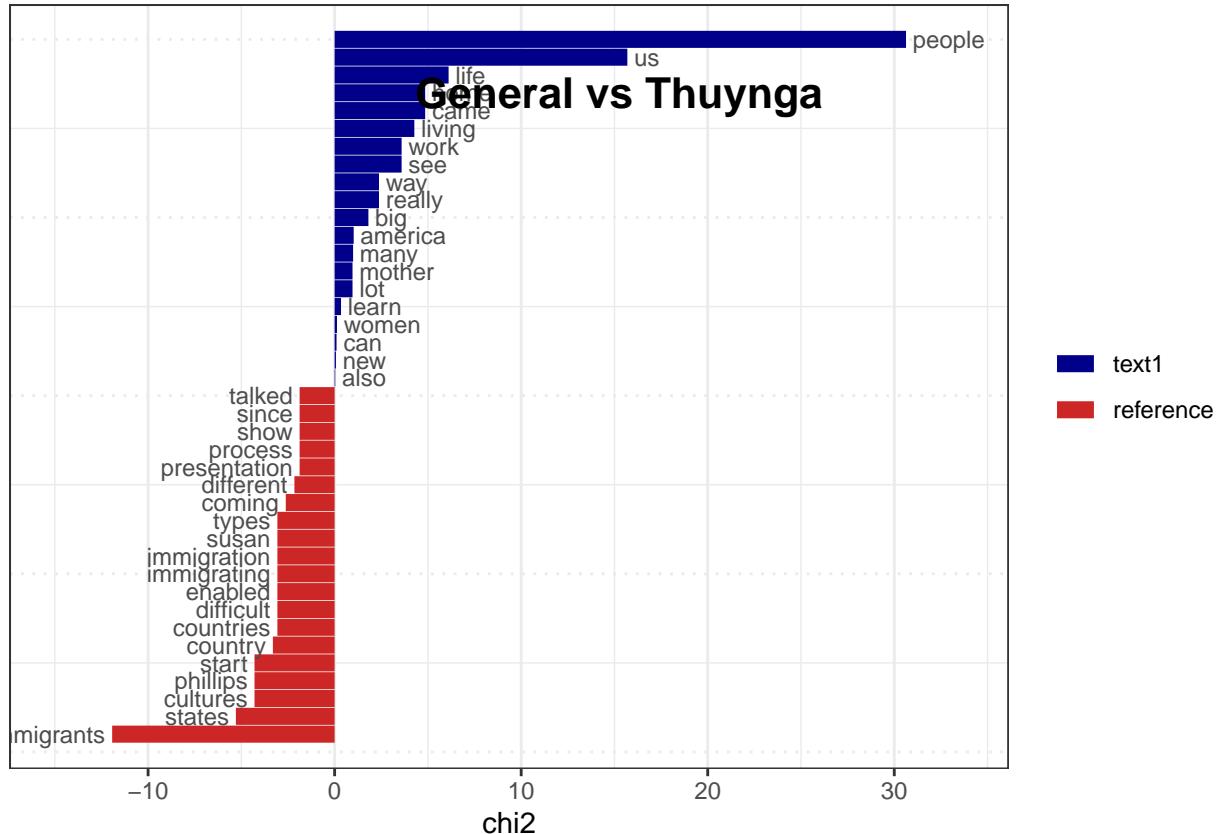


```
##  
## ---  
## #### Commentary on Keyness: General vs Nara  
## *(Write commentary here.)*  
## ---  
  
plot_keyness_comparison(general_dfm, chap_dfms$xidan, "General vs Xidan")
```



```
## 
## --- 
## #### Commentary on Keyness: General vs Xidan 
## *(Write commentary here.)* 
## --- 

plot_keyness_comparison(general_dfm, chap_dfms$thuynga, "General vs Thuynga")
```



```
##
##
## ---
## #### Commentary on Keyness: General vs Thuynga
##
## *(Write commentary here.)*
## ---
```

### LDA Topic Modeling

```
# lda modeling
library(topicmodels)
library(slam)

convert_for_lda <- function(dfm_obj){
  # convert to regular matrix
  m <- as.matrix(dfm_obj)
  # construct dtm manually
  dtm <- tm::DocumentTermMatrix(
    x = SimpleCorpus(VectorSource(rep(" ", nrow(m)))),
    control = list(
      weighting = weightTf,
      dictionary = colnames(m)
    )
  )
```

```

# insert counts into the dtm
dtm$v <- as.numeric(m[m > 0])
dtm$i <- as.integer(row(m)[m > 0])
dtm$j <- as.integer(col(m)[m > 0])

# remove empty rows
nonempty <- slam::row_sums(dtm) > 0
dtm <- dtm[nonempty, ]

return(dtm)
}

# ensure consistent vocabulary across dfms
features_union <- unique(
  c(
    featnames(general_dfm),
    featnames(middleeast_dfm),
    featnames(southamerica_dfm),
    featnames(asia_dfm)
  )
)

general_dfm      <- dfm_match(general_dfm, features = features_union)
middleeast_dfm   <- dfm_match(middleeast_dfm, features = features_union)
southamerica_dfm <- dfm_match(southamerica_dfm, features = features_union)
asia_dfm         <- dfm_match(asia_dfm, features = features_union)

# combined reflection dfm
all_ref_dfm <- rbind(
  general_dfm,
  middleeast_dfm,
  southamerica_dfm,
  asia_dfm
)
all_ref_dtm <- convert_for_lda(all_ref_dfm)

lda_model <- LDA(all_ref_dtm, k = 4, control = list(seed = 1234))

topic_terms <- terms(lda_model, 12) # top 12 words per topic
topic_terms

```

```

##          Topic 1        Topic 2        Topic 3        Topic 4
## [1,] "life"       "people"      "china"       "women"
## [2,] "culture"     "think"       "story"        "country"
## [3,] "like"        "us"          "xidan"        "immigrants"
## [4,] "english"      "women"       "chinese"      "states"
## [5,] "us"           "can"         "one"          "many"
## [6,] "immigrants"   "many"        "revolution"   "united"
## [7,] "american"     "immigrant"   "america"      "stories"
## [8,] "people"       "feel"        "us"           "life"
## [9,] "america"      "immigrants"  "states"       "different"

```

```

## [10,] "country"      "stories"      "united"       "new"
## [11,] "immigrant"    "like"        "experiences"  "immigration"
## [12,] "stories"      "time"        "family"       "us"

```

Top Terms per topic

```

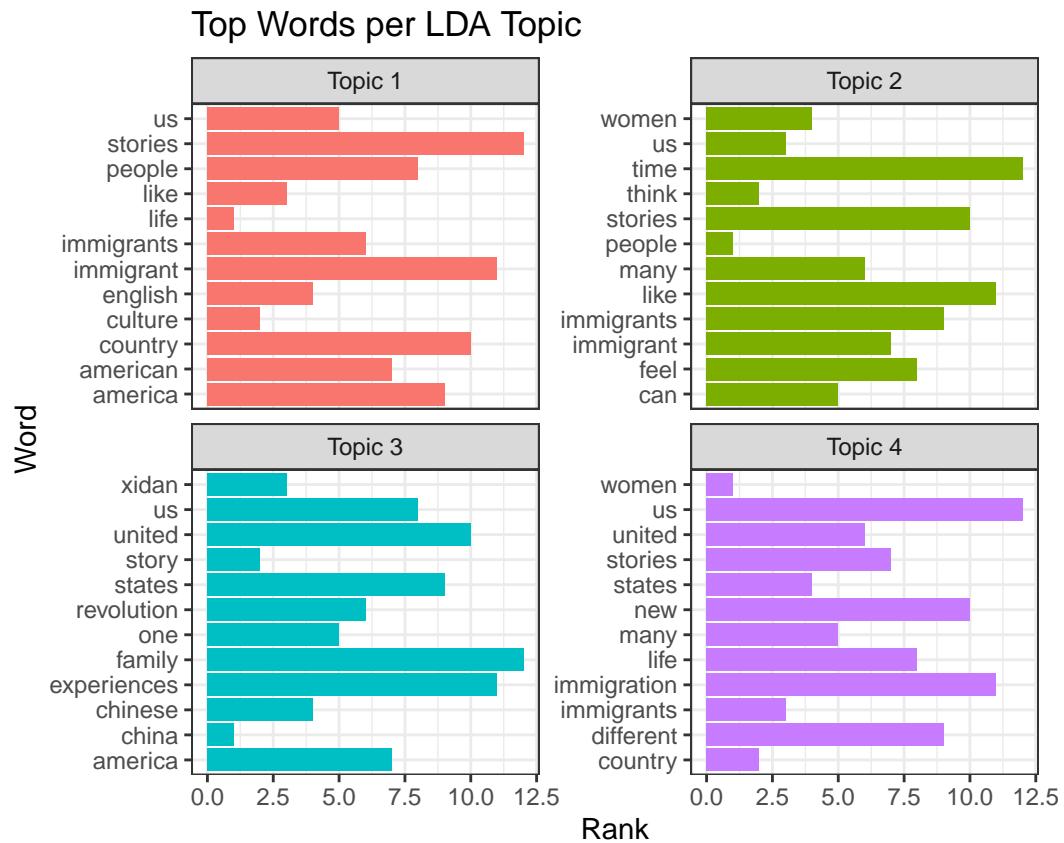
# top 12 terms per topic
topic_terms <- terms(lda_model, 12)

topic_terms_df <- as.data.frame(topic_terms) %>%
  tibble::rownames_to_column("rank") %>%
  pivot_longer(-rank, names_to = "topic", values_to = "term")

topic_terms_df$topic <- factor(topic_terms_df$topic)

# plot top words per topic
ggplot(topic_terms_df, aes(x = reorder(term, rank), y = as.numeric(rank), fill = topic)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  facet_wrap(~topic, scales = "free_y") +
  labs(title = "Top Words per LDA Topic", x = "Word", y = "Rank") +
  theme_bw()

```



Topic Distribution per document

```

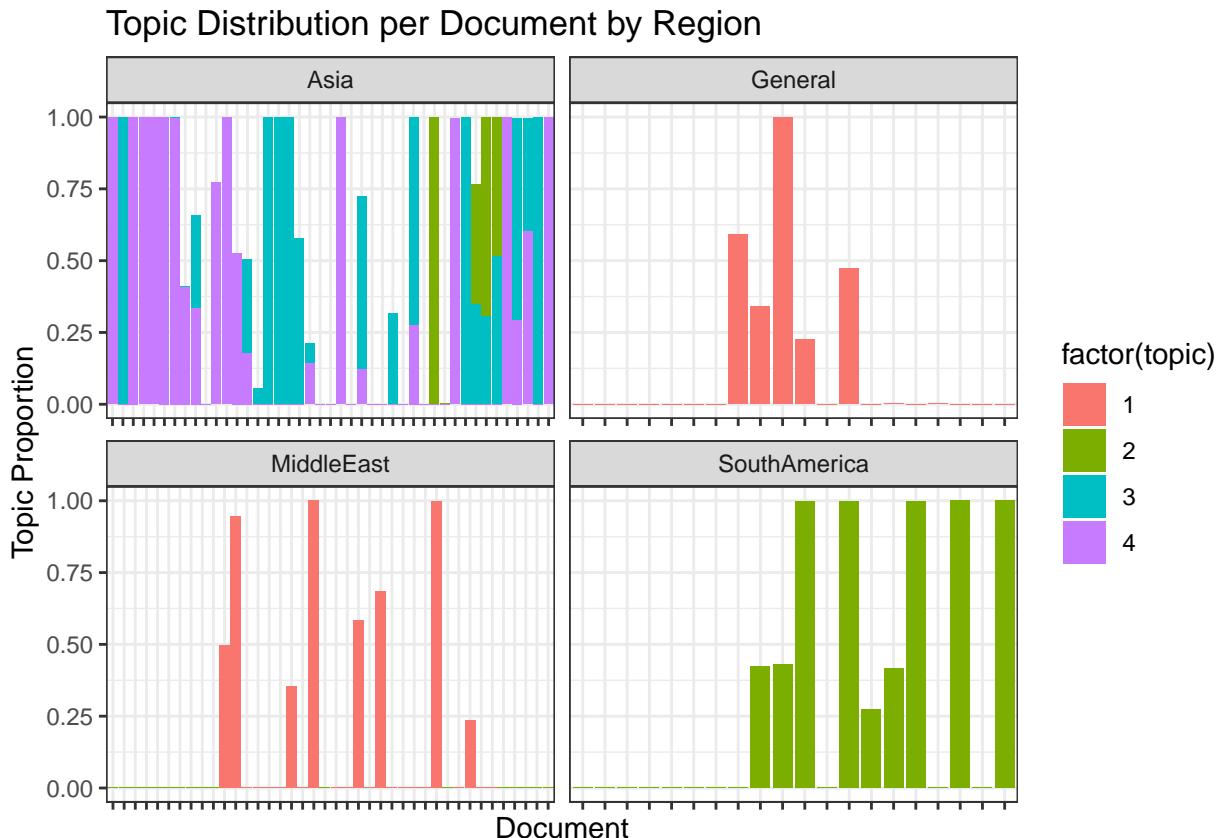
k <- lda_model@k # number of topics
D <- length(all_ref_dtm$dimnames$Docs) # number of documents

doc_topics <- tidy(lda_model, matrix = "gamma")

# repeat each region label for each topic
doc_topics$region <- rep(
  c(rep("General", nrow(general_dfm)),
    rep("MiddleEast", nrow(middleeast_dfm)),
    rep("SouthAmerica", nrow(southamerica_dfm)),
    rep("Asia", nrow(asia_dfm))
  ),
  each = k
)

ggplot(doc_topics, aes(x = factor(document), y = gamma, fill = factor(topic))) +
  geom_bar(stat = "identity") +
  facet_wrap(~region, scales = "free_x") +
  labs(title = "Topic Distribution per Document by Region",
       x = "Document", y = "Topic Proportion") +
  theme_bw() +
  theme(axis.text.x = element_blank())

```



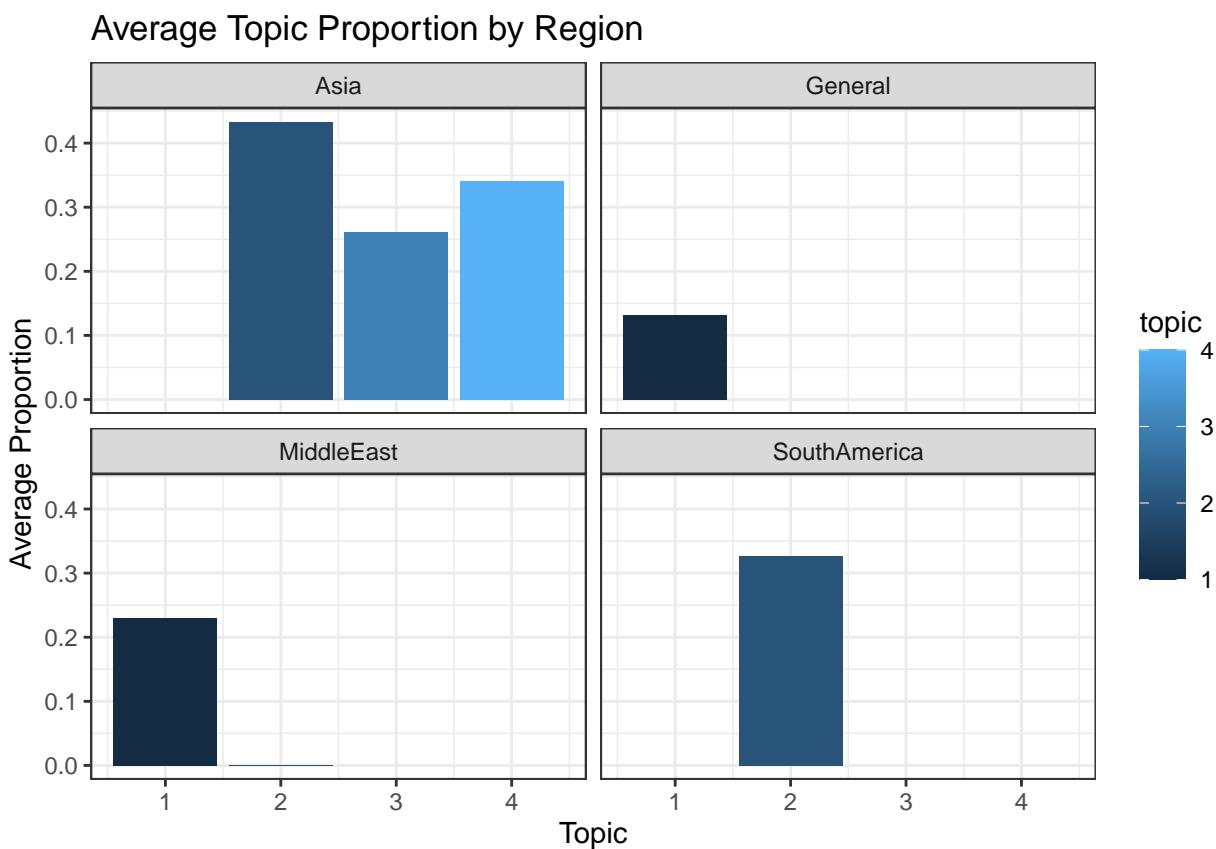
Average topic distribution per region

```

avg_topic_region <- doc_topics %>%
  group_by(region, topic) %>%
  summarise(avg_gamma = mean(gamma), .groups = "drop")

ggplot(avg_topic_region, aes(x = topic, y = avg_gamma, fill = topic)) +
  geom_col() +
  facet_wrap(~region) +
  labs(
    title = "Average Topic Proportion by Region",
    x = "Topic",
    y = "Average Proportion"
  ) +
  theme_bw()

```



Top Topic per Document

```

top_topic_doc <- doc_topics %>%
  group_by(document) %>%
  slice_max(gamma, n = 1) %>%
  select(document, topic, gamma, region)

top_topic_doc

```

```

## # A tibble: 43 x 4
## # Groups:   document [43]

```

```
##      document topic gamma region
##      <chr>    <int> <dbl> <chr>
## 1 1          4 0.999 Asia
## 2 10         3 0.998 Asia
## 3 11         4 0.999 Asia
## 4 12         4 0.999 Asia
## 5 13         4 0.998 Asia
## 6 14         4 0.999 Asia
## 7 15         4 0.997 Asia
## 8 16         1 0.591 General
## 9 17         1 0.341 General
## 10 18        1 0.999 General
## # i 33 more rows
```