

# 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

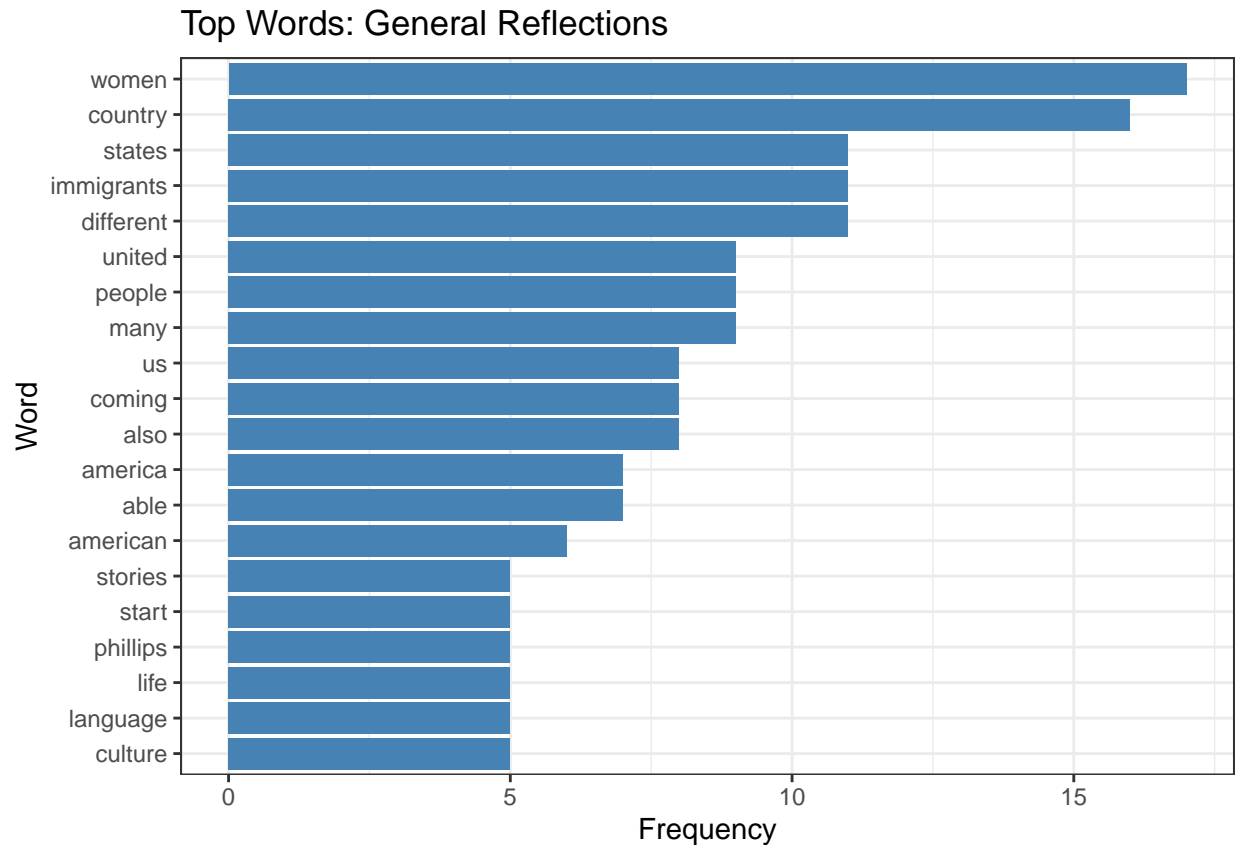
- Word frequency: The count of each word appearing in student reflections and book chapters, used to identify which terms or themes are most common.
- Keyness / Text Similarity: the statistical significance of word usage differences between reflections and corresponding book chapters.
- Topic Distribution: 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.

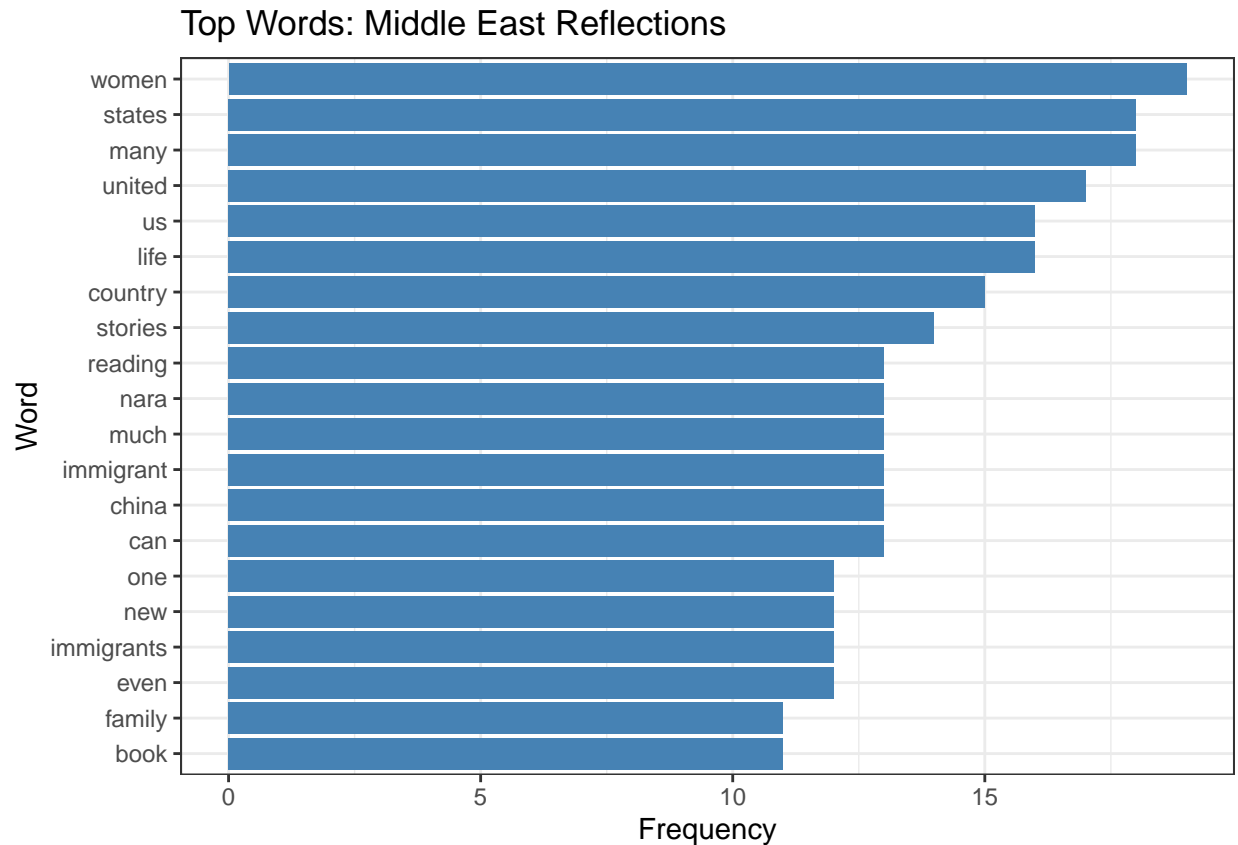
# IV. Exploratory Data Analysis

```
# 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, l
)
```



```
##
##
## ---
## ### 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.
## ---
```

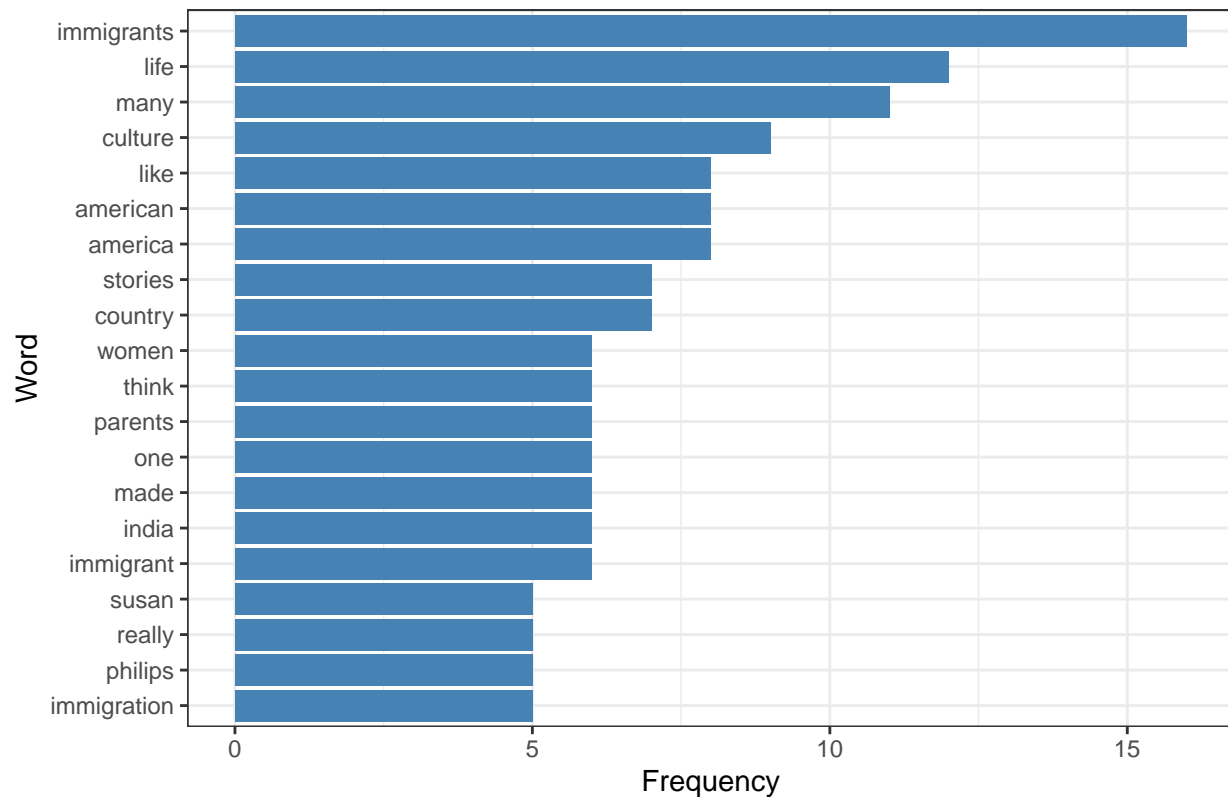
```
plot_with_commentary(
  plot_top_terms(middleeast_dfm, title = "Top Words: Middle East Reflections"),
  "Middle East Reflections",
  commentary = "While grouped under \"Middle East\", the vocabulary closely mirrors the generalized immigrant experience."
)
```



```
##
##
## ---
## ### 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
)
```

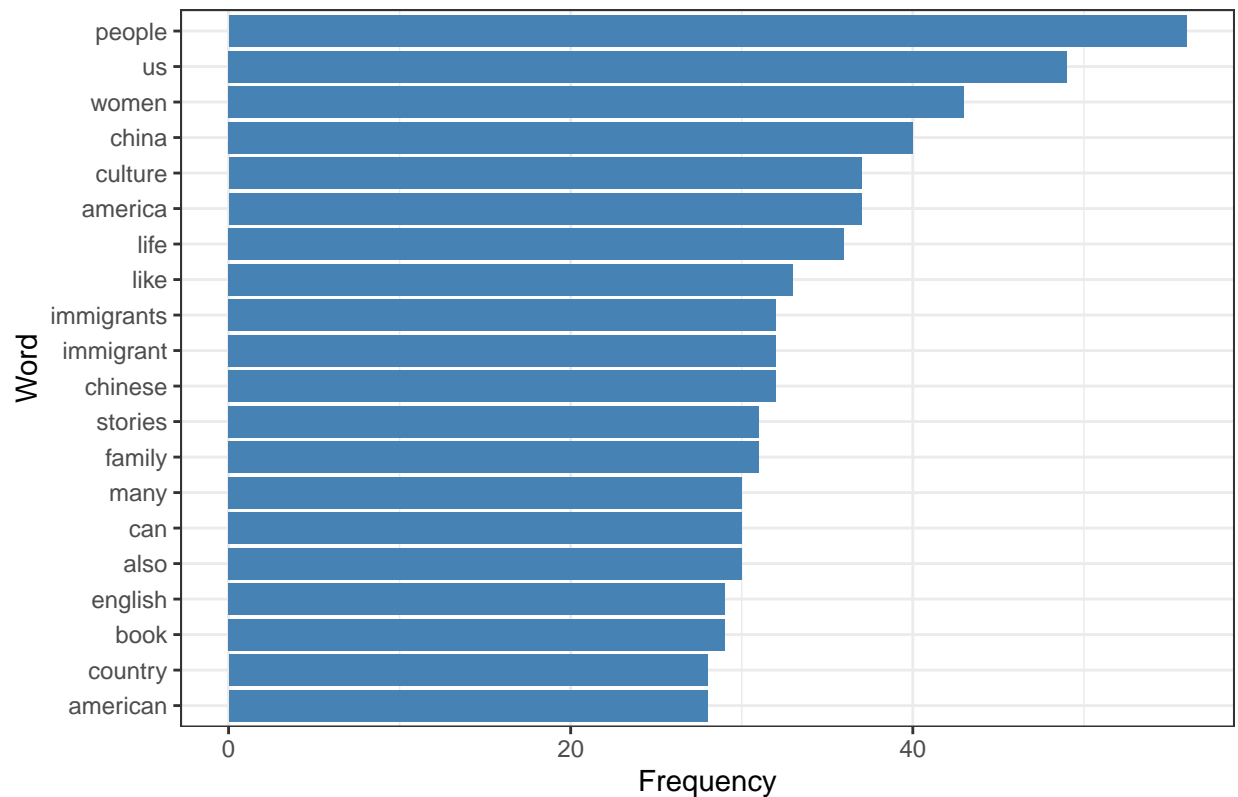
### Top Words: South America Reflections



```
##
##
## ---
## ### 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 e
)
```

## 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_wordcloud_dfm(general_dfm, "General Reflections")
```

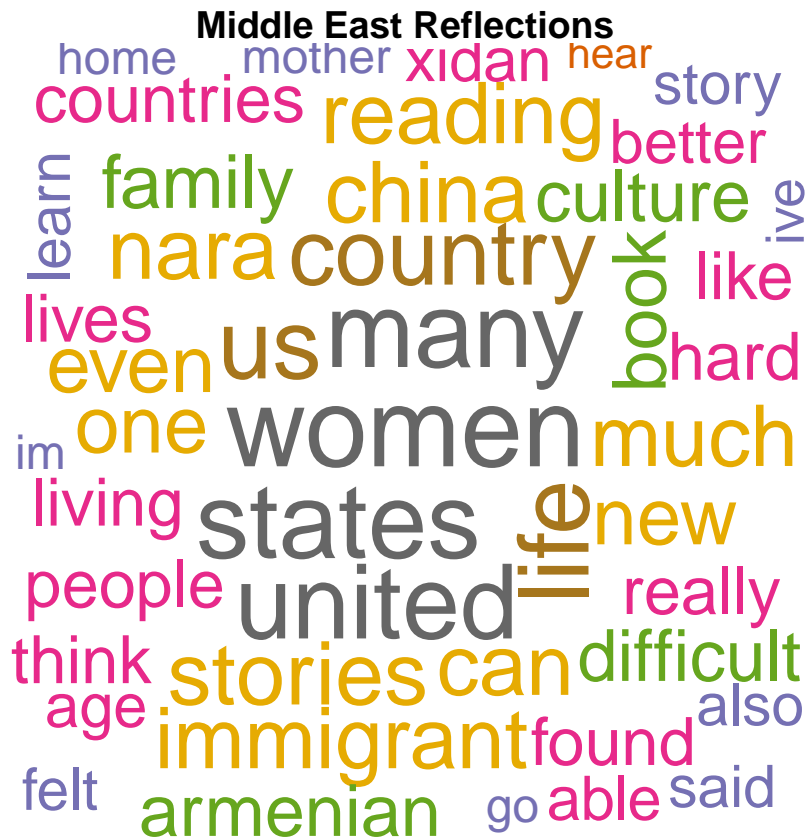
## General Reflections



```
cat(paste(strwrap("Overall, the word cloud illustrates that students gravitate toward themes of migrati
```

## Overall, the word cloud illustrates that students gravitate toward themes of  
## migration, national identity, gender, cultural diversity, and personal  
## experience. These themes strongly resonate with the central concerns of *In the*  
## *Heart of Another*, confirming the study's finding that student reflections  
## meaningfully engage with Philips' material.

```
plot_wordcloud_dfm(middleeast_dfm, "Middle East Reflections")
```



```
cat(paste(strwrap("The Middle East reflections (Armenia) show a pronounced focus on the personal, and en
```

```
## The Middle East reflections (Armenia) show a pronounced focus on the personal,
## and emotional dimensions of the material. Dominant words like life, women, and
## stories indicate a deep engagement with the individual narratives and
## experiences, particularly those of female immigrants. Uniquely, the large size
## of reading and book suggests students were highly conscious of the source
## material and the act of learning from it. The frequent appearance of terms like
## difficult suggests they reflected on the emotional weight and challenges
## presented in these stories.
```

```
plot_wordcloud_dfm(southamerica_dfm, "South America Reflections")
```



[illegible]

```
## The student reflections focused on South America heavily emphasized the
## logistical and legal challenges of migration to the United States. The
## prominence of words like immigrants, life, american, united, and states clearly
## frames the experience within the context of the destination country.
## Furthermore, key terms such as illegal and struggles highlight the reflections'
## attention to social justice issues and the difficult journey, often centered
## around immigrant parents and the impact on family.
```

9

# Asia Reflections



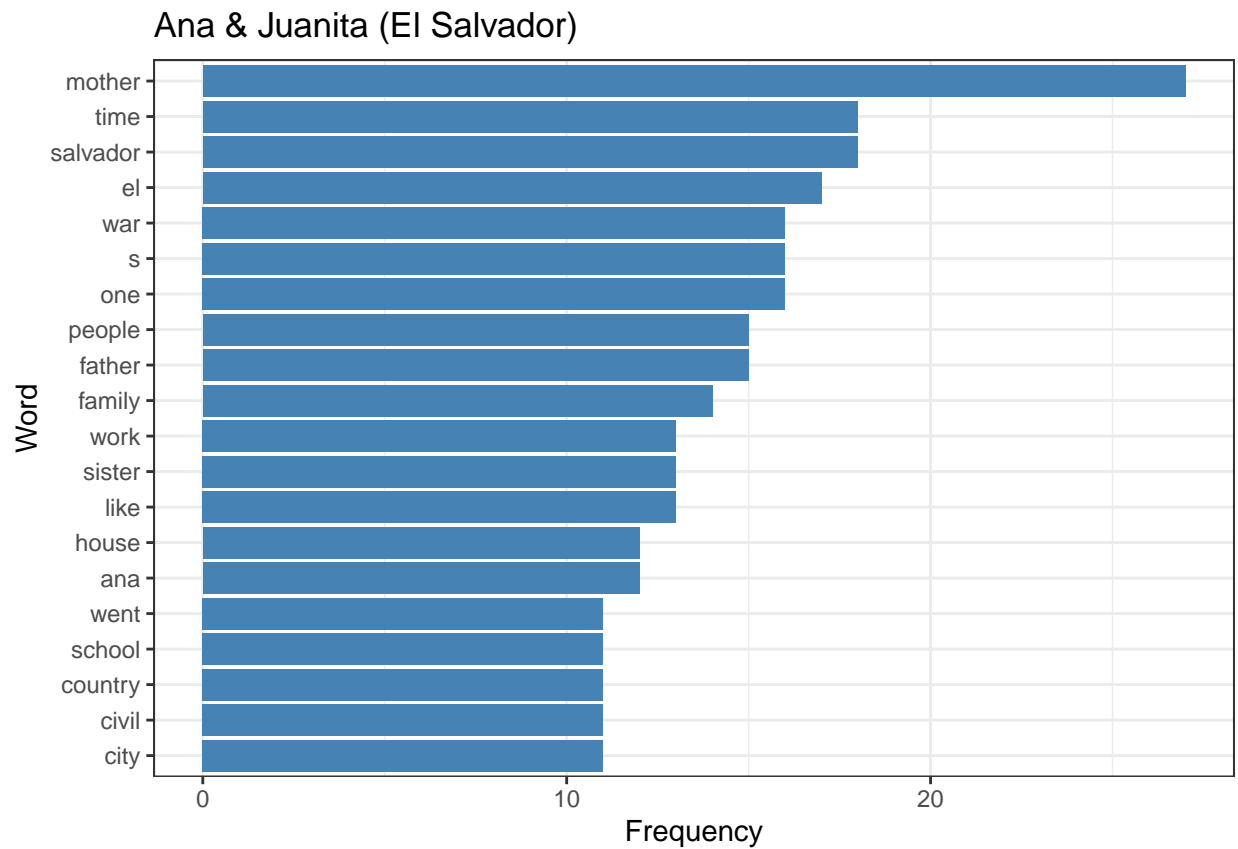
```
cat(paste(strwrap("The student papers on Asia were primarily concerned with cultural identity and adapt
```

```
## The student papers on Asia were primarily concerned with cultural identity and
## adaptation to a new environment. While core themes of people, immigrants, and
## family remain central, the specific focus on culture, china, and chinese points
## to a deep dive into the specific cultural background and adjustment process.
## The inclusion of terms like woman and english suggests reflections on how
## gender roles and linguistic barriers affect assimilation and the construction
## of a new life in the receiving country.
```

## Chapter Analysis

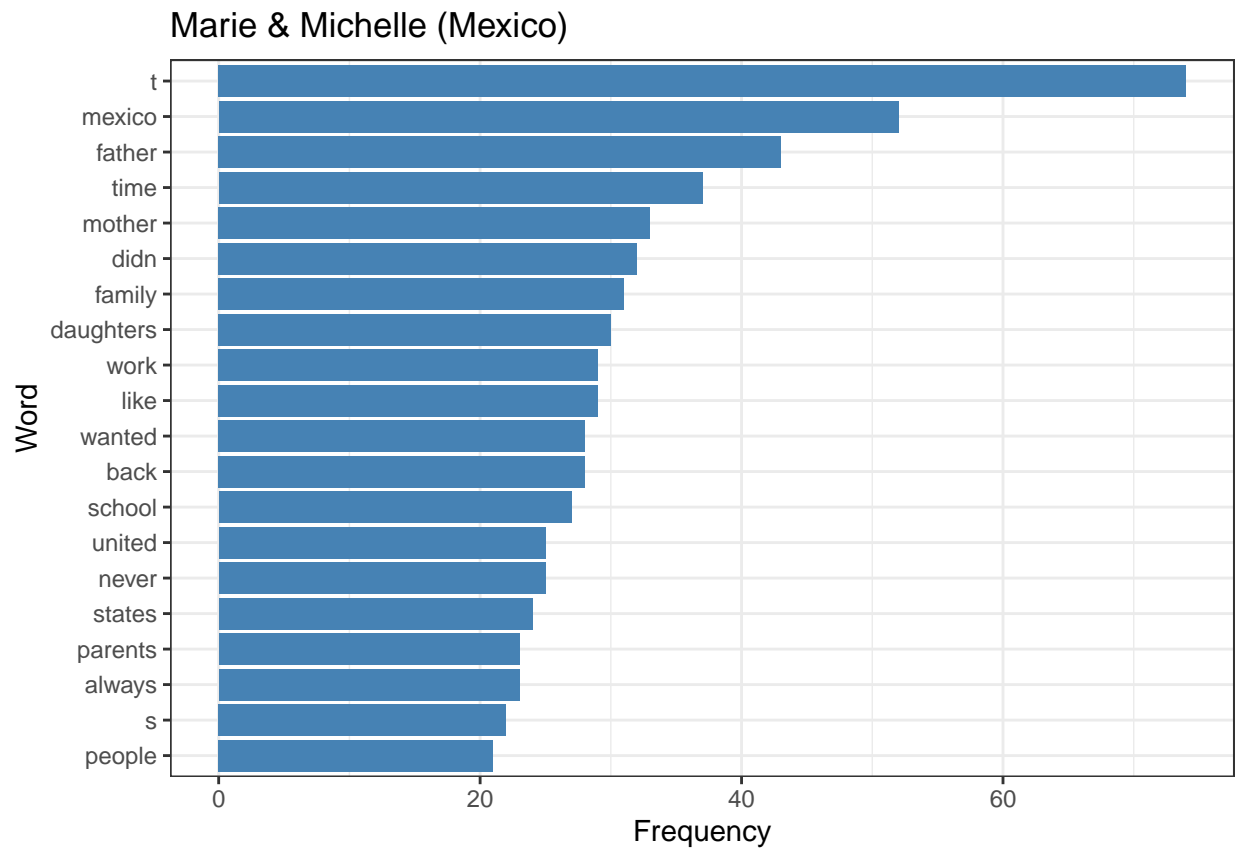
## Top Terms in book chapters

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



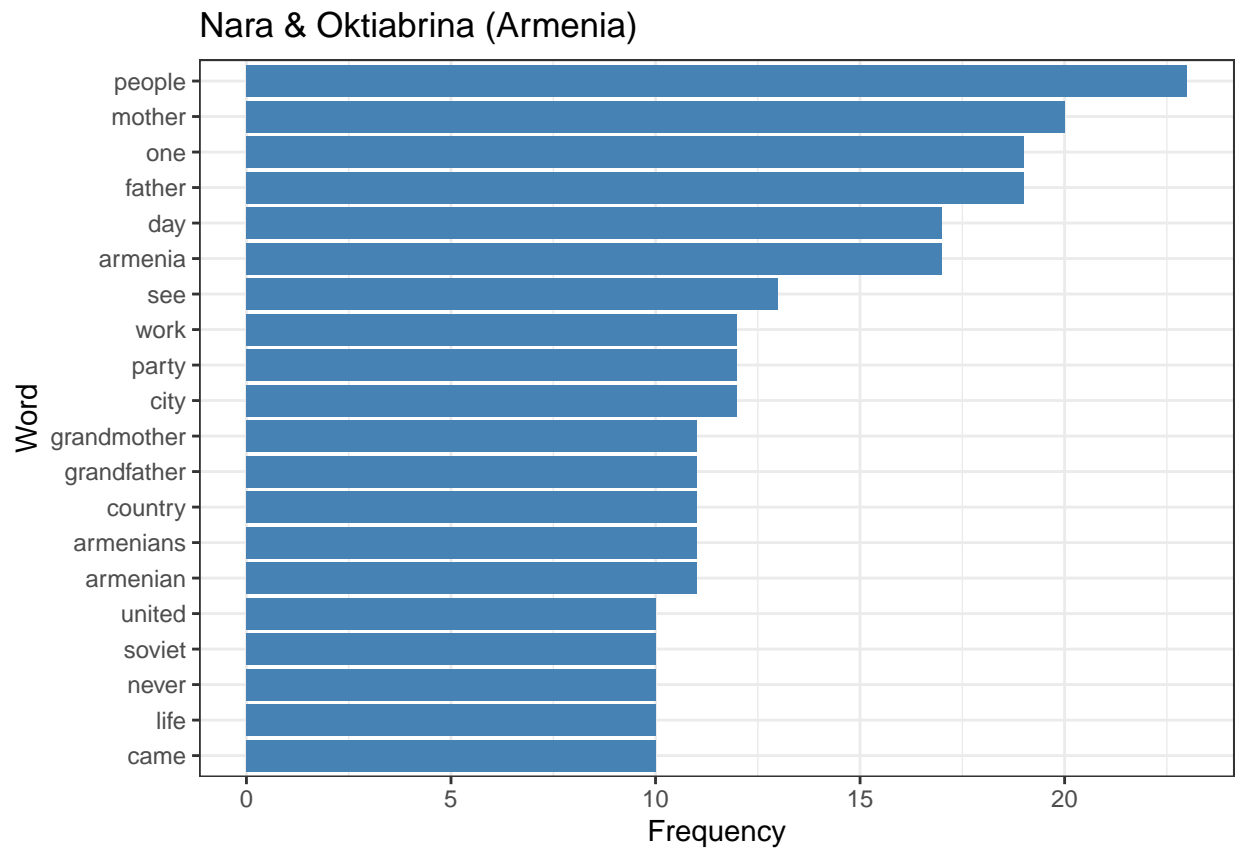
```
##
##
## ---
## ### Commentary on: Ana & Juanita
##
##
## ---
```

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



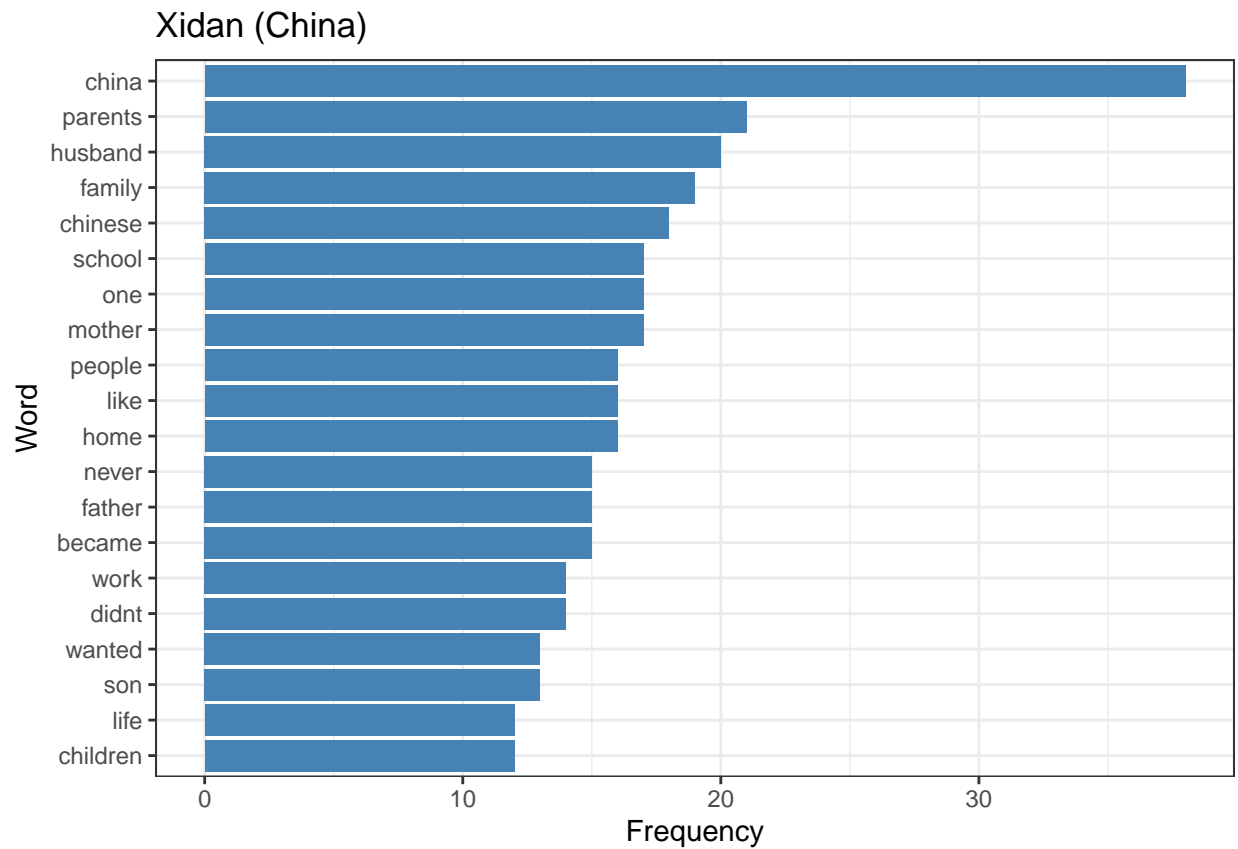
```
##
##
## ---
## ### Commentary on: Marie & Michelle
##
##
## ---
```

```
plot_with_commentary(
plot_top_terms(chap_dfms$nara, title = "Nara & Oktiabrina (Armenia)",
"Nara & Oktiabrina"
)
```



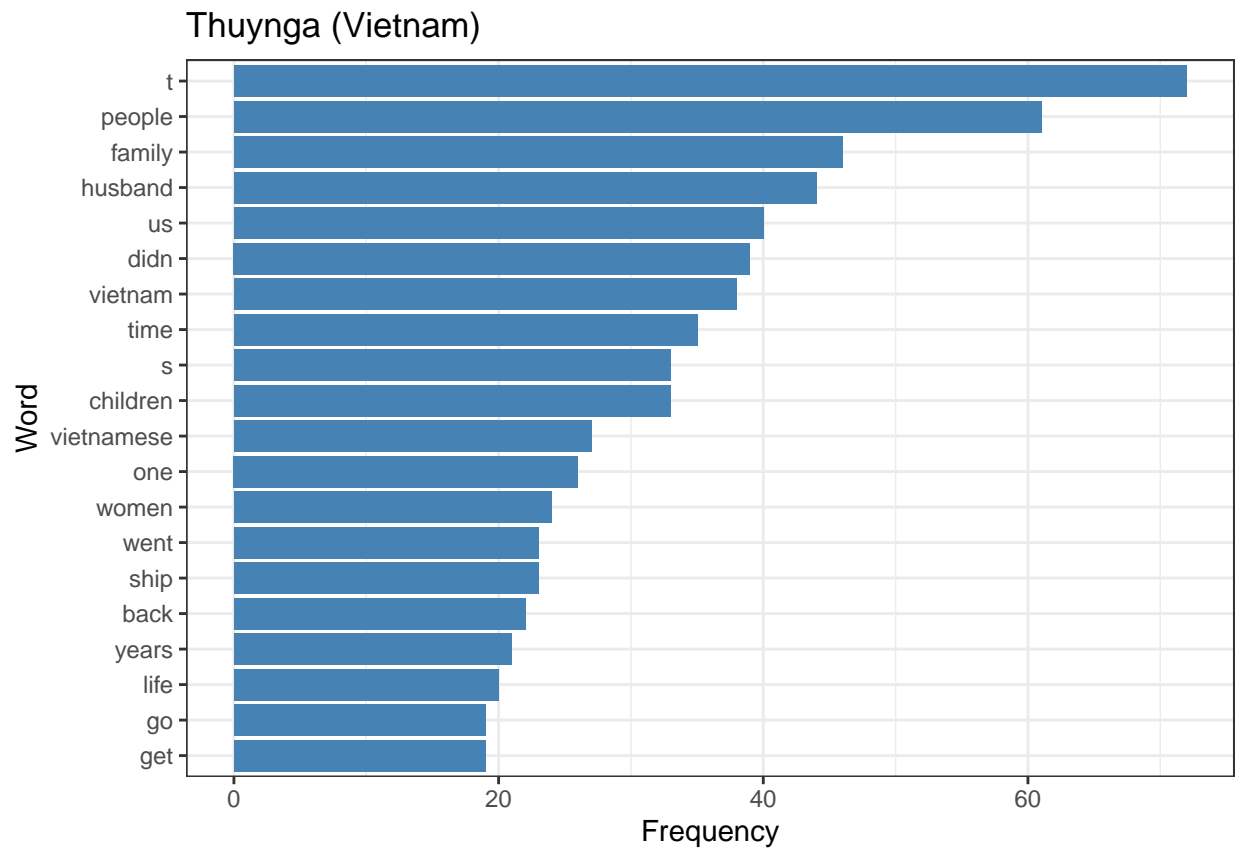
```
##
##
## ---
## ### Commentary on: Nara & Oktiabrina
##
##
## ---
```

```
plot_with_commentary(
plot_top_terms(chap_dfms$xidan, title = "Xidan (China)",
"Xidan"
)
```



```
##
##
## ---
## ### Commentary on: Xidan
##
##
## ---
```

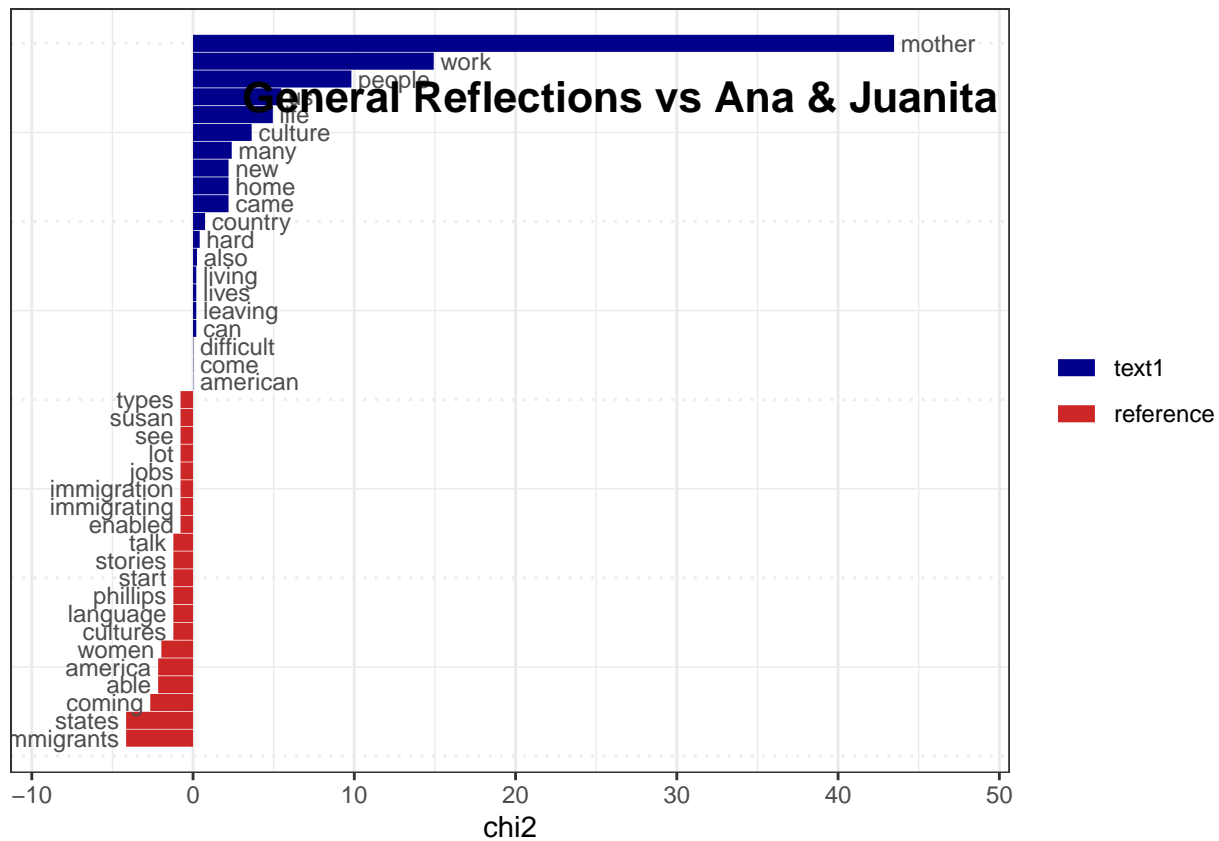
```
plot_with_commentary(
plot_top_terms(chap_dfms$thuynga, title = "Thuynga (Vietnam)",
"Thuynga"
)
```



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

Keyness: Reflection vs. Chapter

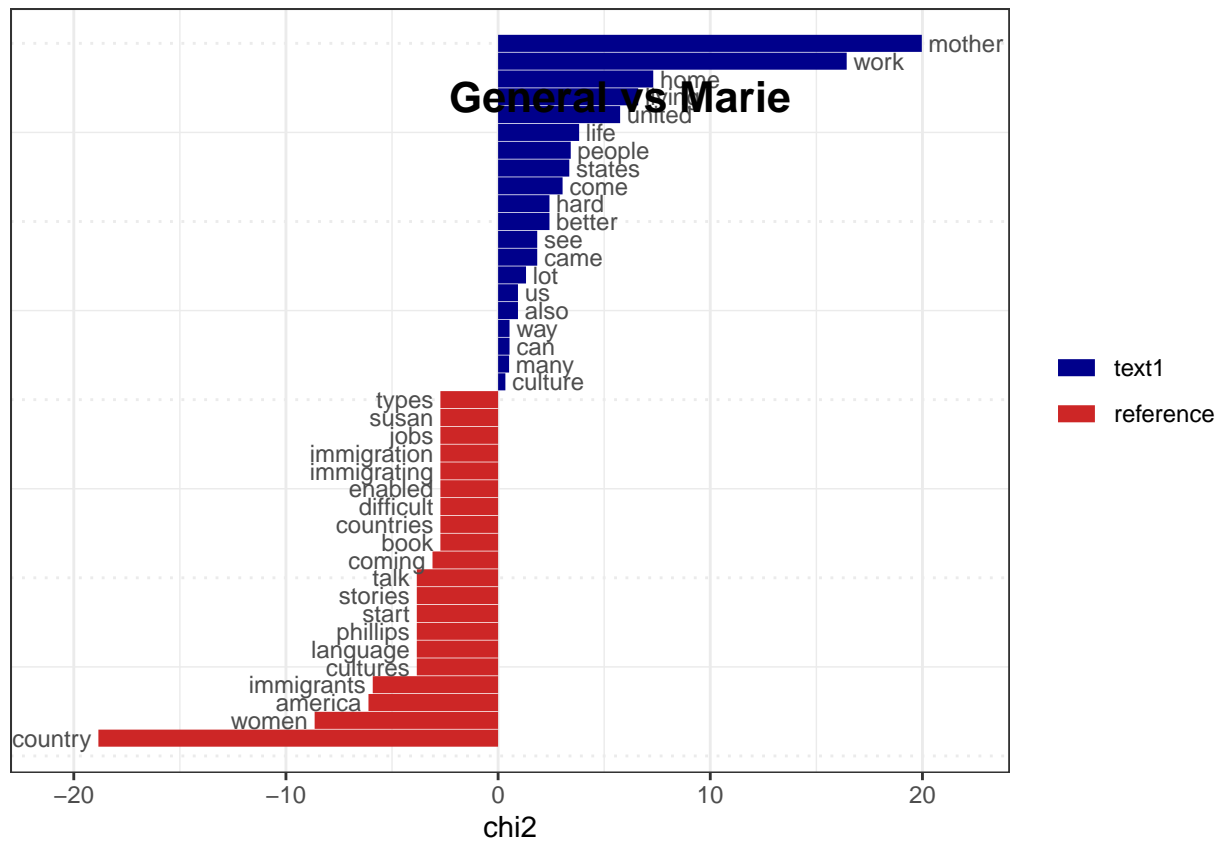
```
plot_keyness_comparison(
  general_dfm,
  chap_dfms$ana,
  "General Reflections vs Ana & Juanita"
)
```



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

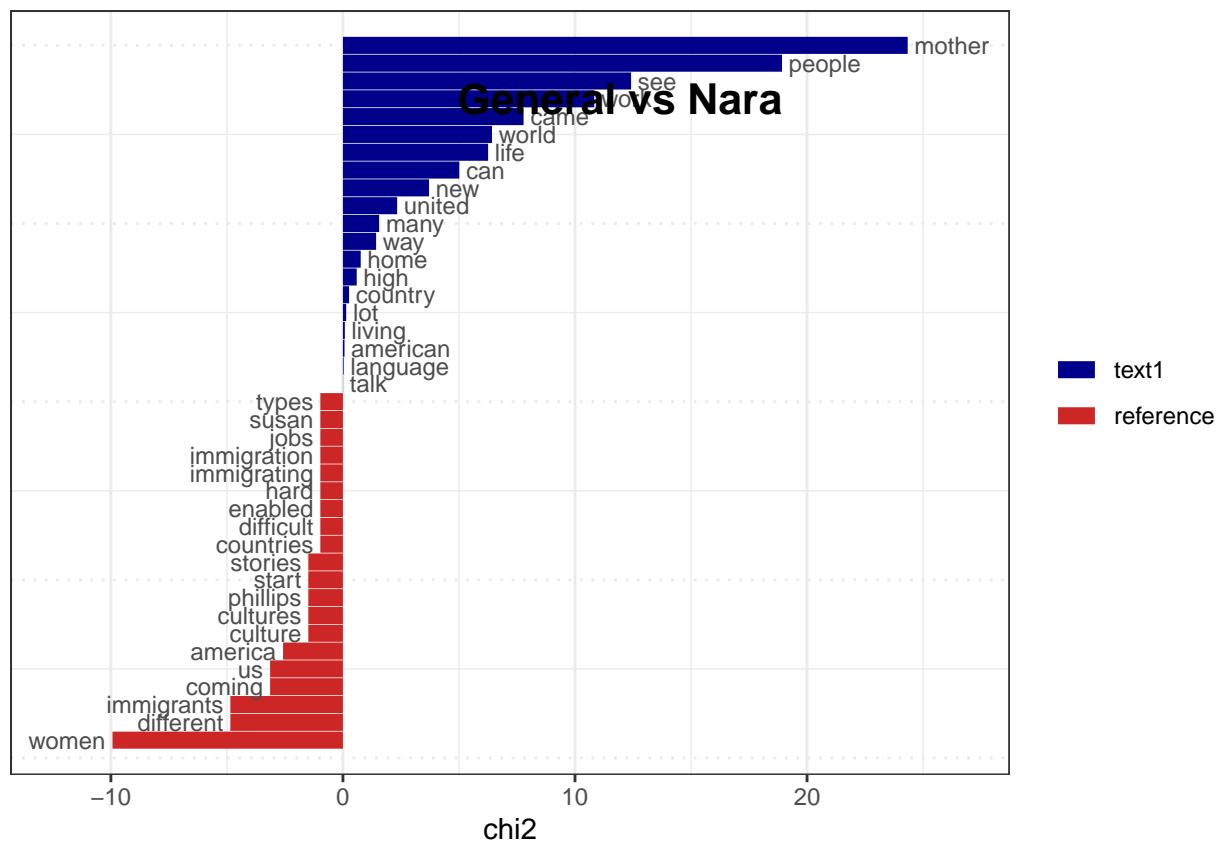
```
plot_keyness_comparison(general_dfm, chap_dfms$marie, "General vs Marie")
```





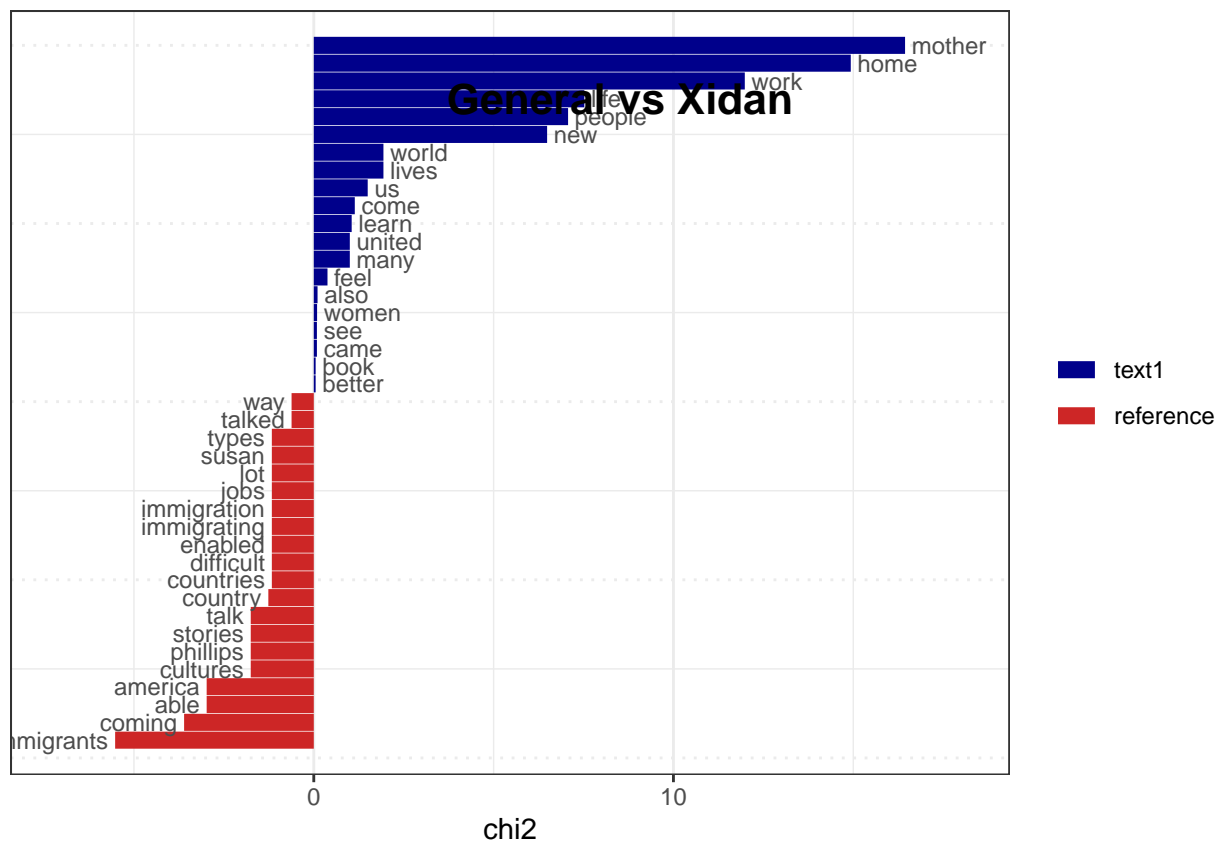
```
##
##
## ---
## ### Commentary on Keynes: General vs Marie
##
## *(Write commentary here.)*
## ---
```

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



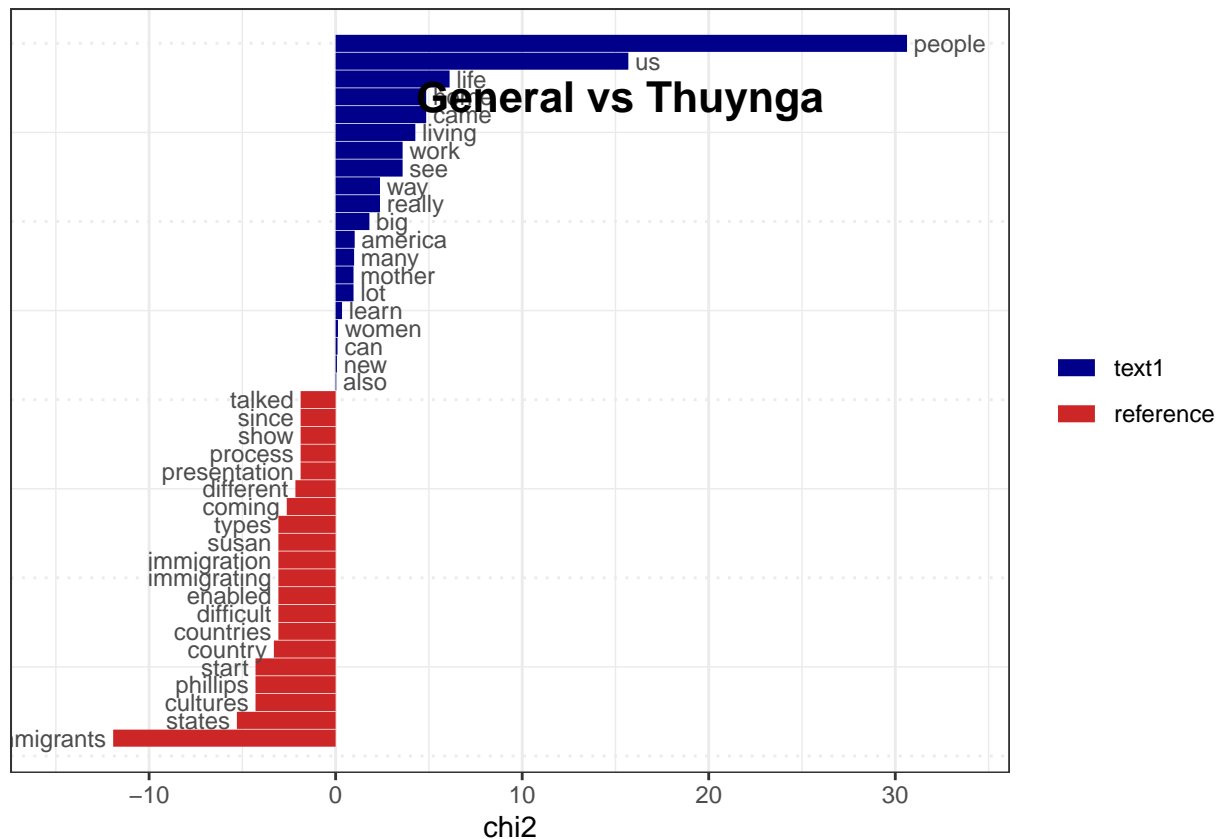
```
##
##
## ---
## ### Commentary on Keynes: General vs Nara
##
## *(Write commentary here.)*
## ---
```

```
plot_keyness_comparison(general_dfm, chap_dfms$xidan, "General vs Xidan")
```



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

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



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

LDA Topic Modeling

Top Terms per topic

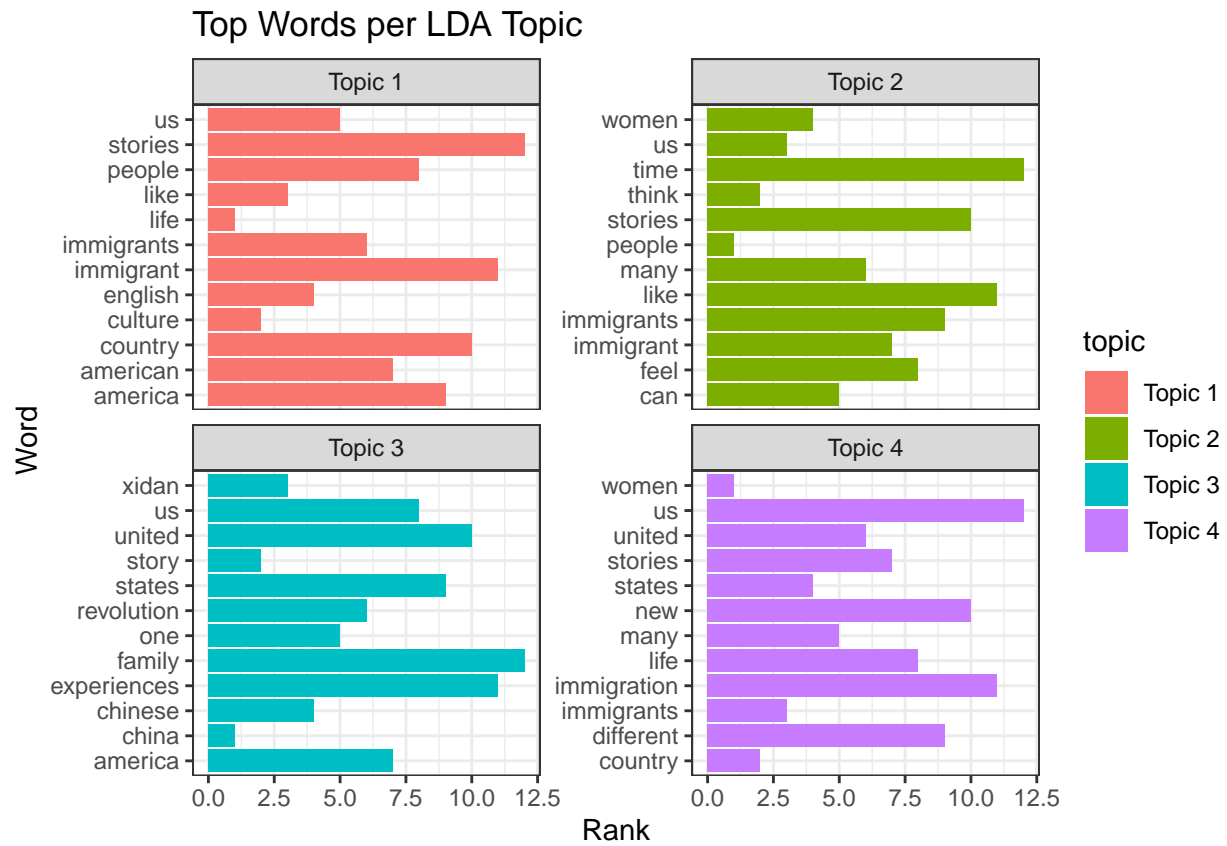
```
# top 12 terms per topic
topic_terms <- terms(lda_model, 12)

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") +
```

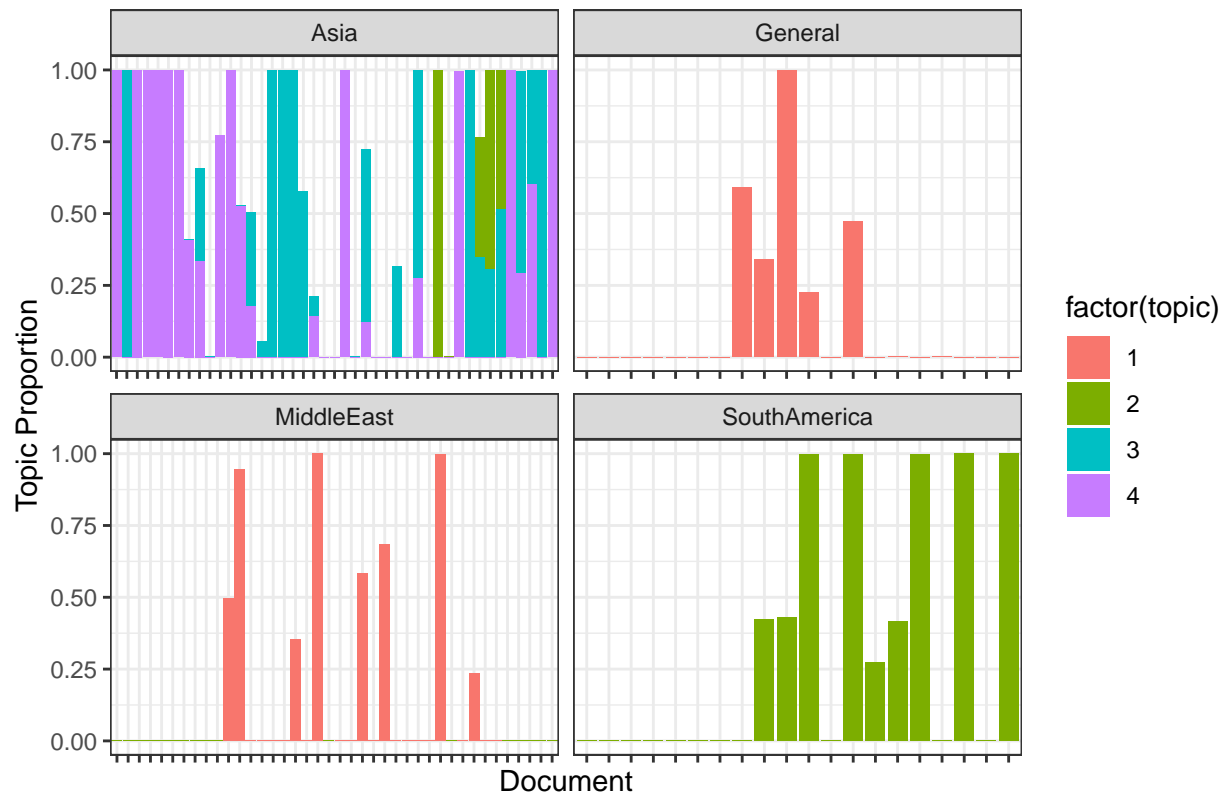
```
labs(title = "Top Words per LDA Topic", x = "Word", y = "Rank") +  
theme_bw()
```



Topic Distribution per document

```
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())
```

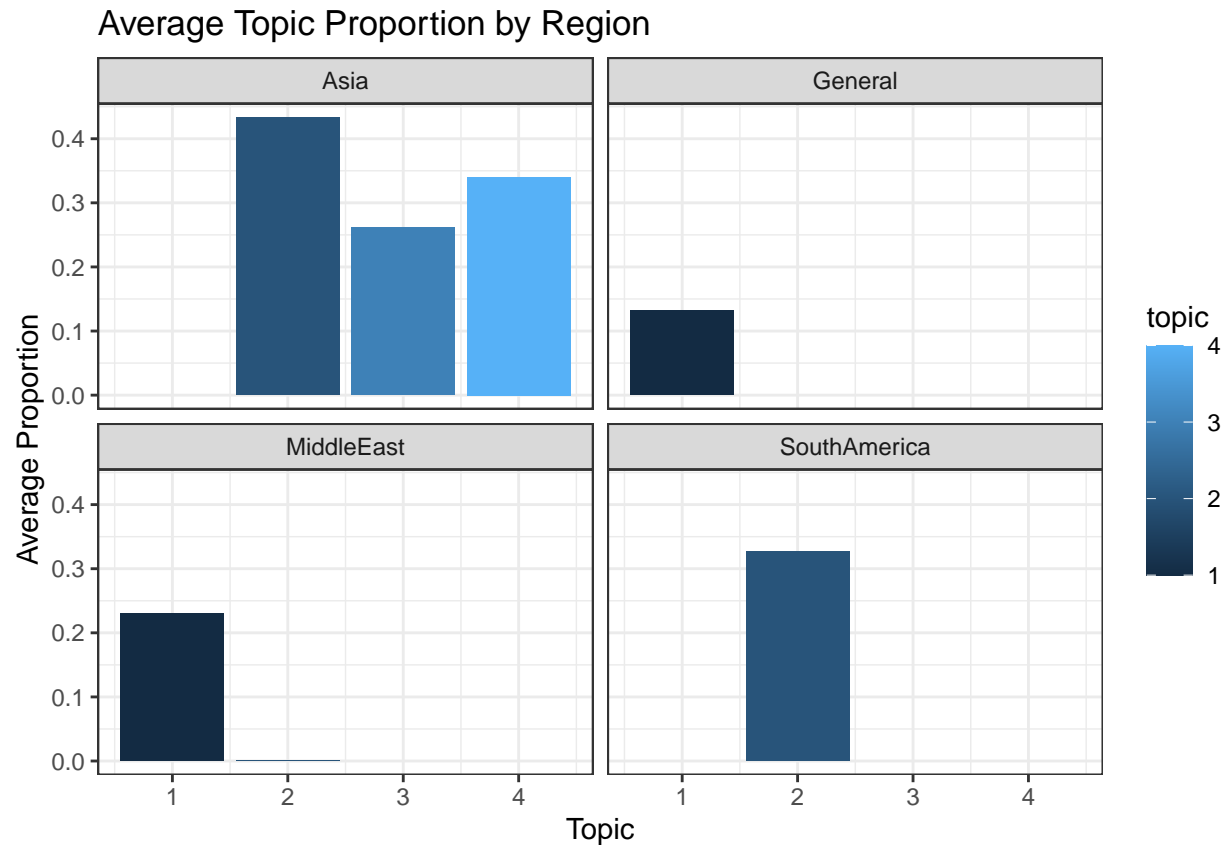
Topic Distribution per Document by Region



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
```

## Conclusion

1. Students resonate emotionally more than contextually Across TF and word clouds, students emphasize human-centered themes like: identity and belonging (life, culture, family, people), empathy and connection (feel, stories, think), and gender and resilience (women). In contrast, the book chapters surface historical & geopolitical identifiers (e.g., civil war in El Salvador, Soviet Armenia, migration by ship from Vietnam), showing that students internalized personal struggle narratives but did not engage as deeply with regional historical framing.
2. Regional categorization of reflections does not always match content The “Middle East” and “South America” folders still contain high-frequency references to China and generalized U.S. immigration terms (america, immigrant, states, women). This suggests that students often mapped specific global narratives onto broader U.S. immigration discourse. Reflection labels were not strong predictors of linguistic content and the Xidan (China) story disproportionately shaped student vocabulary across all regions. This is an important conclusion about student interpretation behavior, not modeling error.
3. Topic modeling reveals distinct thematic clusters, but personal themes dominate The 4-topic LDA model can be interpreted as: 1) cultural identity and assimilation, 2) emotional response and empathy, 3) China-centered narrative influence and generational migration, and 4) gender, mobility, and societal-level immigration patterns. The topic distribution plots show that most documents were almost entirely dominated by a single topic ( $\gamma > 0.95$ ), meaning student reflections were thematically clear but narrow in focus.
4. Students engage with migration as a lived experience, not a historical event The lack of high-keyness geographic terms like El Salvador, Mexico, Vietnam, or Armenia in student text shows that students framed migration as a personal/familial journey, a cultural identity transition, and part of a shared immigrant condition rather than grounding it in country-specific political forces highlighted in Philips’ chapters.
5. Language and culture barriers are perceived as central to migration Across all reflection regions, students elevate words like culture, english, immigrants, and american as significantly key, implying that students saw cultural integration and language acquisition as core challenges. Even when chapters didn’t prioritize those terms (e.g., Mexico or Armenia), students still foregrounded them. Linguistic difficulty and cultural identity were more salient to students than to the chapter texts themselves.
6. Gender-focused interpretation is a strong emergent theme Topic 4 and word clouds both strongly highlight women, supporting the conclusion that students perceived female narratives, resilience, and mother-figures as particularly impactful and gender was one of the strongest unifying takeaways, despite not being the controlling variable.
7. Limited dataset still produced meaningful thematic alignment Despite only 43 reflections, students and chapters shared migration/identity themes in LDA and frequent shared terms, showing clear conceptual connection to Philips’ core themes, divergence in linguistic specificity, and proof of concept that reflection papers can be systematically analyzed with text mining.

### Why China?

1. Students may have related to the story as it is the most dramatic & narrative-driven. The Xidan chapter in the book includes personal transformation, family conflict, generational tension, escaping a restrictive society, and emotional struggle + resilience. Stories with clear protagonists, conflict, risk, and resolution are always more memorable and relatable, even if students aren’t from China.
2. More students in the class may personally identify as Asian or know Asian immigrant experiences. Even if students aren’t Chinese, many Asian-American students can easily connect to themes like parents having high expectations and cultural pressure around success. Many students can also relate to negotiating two identities (home culture vs America) and the challenges like language barriers and



intergenerational sacrifice. So the China narrative became a stand-in for the broader “first-generation Asian immigrant experience.”

3. China is a highly recognizable country compared to other countries since it has massive global visibility in media, politics, and education. Students already have mental associations they can attach the story to, making it easier to write and reflect on.
4. The protagonist challenges Western/US identity directly. Xidan’s narrative frequently involves moving to America and contrasting Chinese vs American culture. She goes over common challenges that other immigrants also face like learning English, finding work in the U.S., and motherhood in displacement. This naturally overlaps with the dominant vocabulary students used in reflections, reinforcing the signal in TF, keyness, and LDA.
5. Other chapters are more historical or context-heavy. Ana and Juanita’s story focuses on civil war and schooling during conflict, which may be harder to relate to. Marie and Michelle looks into poverty and leaving family behind and Nara goes into Soviet politics, genocide, and ancestry history. Thugnya shared her experience to America by boat and her displacement timeline. These chapters require historical knowledge or geopolitical framing for deeper engagement, but students tend to focus on personal feelings, culture, family, and gender. In contrast, the China chapter didn’t require much outside context to emotionally understand, so engagement was immediate and strong.
6. “China” appeared often during the talk too Since Philips discussed Xidan during her presentation, those ideas may have acted as a second exposure, making it even more salient to students.

In conclusion,

People didn’t necessarily like China, they related most to a specific story from China (Xidan’s) because it is emotionally intense and has a clear character arc. It is also less historical and requires less background, making it easier to empathize with and many Asian-American identity themes overlap with Xidan’s story. Her story also attaches easily to a country everyone already recognizes. Seeing that, it was the narrative and identity themes that happened to be tied to China, which many students already subconsciously connect to their own or their parents’ experiences.