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 conducts a quantitative text analysis of 43 STATS 112 student reflection papers written in response to Susan Philips' talk and book *In the Heart of Another*, which presents migration narratives from El Salvador, Mexico, Armenia, China, and Vietnam. Using a suite of statistical text mining methods—including document-term matrices, term frequency analysis, keyness statistics, word-cloud visualization, and Latent Dirichlet Allocation (LDA) topic modeling—I compare linguistic patterns across four reflection groups (General, Middle East, South America, and Asia) and evaluate their alignment with the book chapters they correspond to. The analyses reveal three consistent patterns. First, students overwhelmingly emphasized broad cultural and personal themes such as family, culture, life, women, and stories, while largely omitting the geopolitical and historical vocabulary that was highly frequent in the chapters (e.g., civil war, Soviet, Vietnam, El Salvador). Second, references to China and the Xidan chapter appeared at surprisingly high frequency across all regions, a pattern further reinforced by a China-specific LDA topic. This suggests that students resonated more strongly with narratives involving language barriers, educational pressures, and bicultural identity—experiences that were more relatable to the contemporary student population than the historically grounded stories from other regions. Finally, keyness analyses consistently showed that chapter texts were dominated by place-specific and historically anchored terms, while student reflections favored generalized and emotionally oriented language. Taken together, the results indicate that students tended to abstract Philips' region-specific narratives into universal themes of migration and identity, with especially strong alignment toward the Xidan (China) narrative. This study demonstrates how text-mining techniques can effectively quantify interpretive patterns in qualitative coursework and highlights how student engagement can depend more on personal relatability than on contextual or historical detail.

II. Statement of the Problem

Courses that engage with global social issues often rely on reflection papers to encourage students to connect personally with complex topics such as migration, identity formation, and social justice. However, while these reflections provide rich qualitative insights, they are rarely analyzed systematically, making it difficult to determine which aspects of the course material resonate most with students and whether their interpretations align with the original texts. In the context of Susan Philips' In the Heart of Another, which presents migration narratives from El Salvador, Mexico, Armenia, China, and Vietnam, it remains unclear how students respond to the distinct historical and geopolitical content of each chapter and whether their reflections preserve these regional distinctions.

The central problem addressed in this project is the gap between what Philips emphasizes in her chapters and what students emphasize in their reflections. Initial observations suggested that students may generalize the narratives into broad cultural or emotional themes, but this had not been formally evaluated. Furthermore, an unexpected pattern emerged during preliminary exploration: students appeared to reference the China

(Xidan) chapter disproportionately, even when reflecting on other regions. This raised additional questions about the consistency and specificity of student engagement across regions.

Therefore, this study applies statistical text-mining techniques, such as term frequency analysis, keyness comparison, and LDA topic modeling, to investigate three core questions: - Which linguistic and thematic patterns dominate student reflections across different regions? - To what extent do these patterns align with the vocabulary and narrative focus of the corresponding book chapters? - Why do certain narratives, particularly the China/Xidan chapter, appear to resonate more strongly with students than others?

By quantifying these patterns, this project aims to better understand how students interpret complex migration narratives and to assess whether reflection papers capture regional nuance or instead gravitate toward personally relatable, generalized themes. The results have implications not only for educational assessment but also for understanding how contemporary students internalize global stories of displace

III. Dataset Descrition and Data Cleaning

Dataset Description

The dataset for this study consists of 43 student reflection papers written in response to Susan Philips' *In the Heart of Another.* Each reflection was submitted as a pdf, then turned into a plain-text file and grouped into one of four folders based on the regional focus selected by the student: - General (n = ?) - Middle East (n = ?) - South America (n = ?) - Asia (n = ?)

Each reflection varies in length, writing style, and structural complexity. These documents collectively represent qualitative student responses to migration narratives across five book chapters: El Salvador, Mexico, Armenia, China, and Vietnam.

For comparison, five chapter texts from Philips' book were included: - Ana & Juanita - El Salvador - Marie & Michelle - Mexico - Nara & Oktiabrina - Armenia - Xidan - China - Thuynga - Vietnam

All files were read using the readtext package, allowing for consistent import and metadata retention.

Goals of Data Cleaning

Before statistical text analysis, the primary objectives of cleaning were to: - Standardize vocabulary for fair comparison across documents - Remove noise introduced by formatting, punctuation, and grammatical forms - Prepare text for DFM/DTM construction required for keyness analysis and LDA modeling - Ensure compatibility between student reflections and chapter texts Because raw text exhibits high variability, cleaning was essential for constructing consistent document-feature matrices and extracting meaningful patterns.

Cleaning Procedures

Data cleaning followed a structured pipeline applied uniformly across all reflections and chapter texts.

Normalization

The text was normalized by first converting all of the text to lowercase. Hyphens and apostrophes were replaced with spaces to prevent unintended token splits and additional non-alphanumeric characters were removed. This prevented multiple variants of the same word (e.g., women, Women, women,) from being counted as separate tokens.

Token-Level Cleaning

Using tm and quanteda, we removed numbers, punctuation, and extra whitespace along with standard English stopwords (e.g., the, and, of, etc.). TO reduce sparsity, we trimmed rare terms using dfm_trim(min_termfreq = 3). These steps produced a cleaner, more interpretable vocabulary set for frequency analysis and modeling.

Corpus Construction

Each folder of reflections was transformed into 1) a cleaned text column, 2) a document-feature matrix (DFM) using quanteda::dfm(), and 3) a combined union of features to ensure all DFMs shared a uniform vocabulary when needed. This allowed for a fair comparison across regions in order to do valid LDA topic modeling and have consistent keyness tests between reflections and chapters.

Chapter Corpus Cleaning   The five chapter texts underwent the same pipeline. This ensured that any differences observed in keyness or frequency were driven by true thematic differences, not by preprocessing inconsistencies.

4. Resulting Dataset After Cleaning

CLeaning lead to a reduction in sparity and improved interpretability of matrices. Now, the dataset consisted of: - 43 reflection documents - 5 chapter documents - A unified vocabulary across all reflections - DFMs and DTMs that are ready for statistical testing, such as word frequency comparison, word cloud visualization, keyness analysis, and LDA topic modeling

The final cleaned dataset allowed for robust, reproducible statistical analysis of linguistic patterns across both student reflections and book chapters.

IIII. Variables in the Study

Independent Variables

Region/Chapter Source (Book Chapters)

Each chapter of *In the Heart of Another* corresponds to a distinct migration narrative rooted in a specific country or historical context (El Salvador, Mexico, Armenia, China, and Vietnam). This variable functions as an external reference category against which student reflections are compared in keyness analysis and topic alignment.

Reflection Group (Student Papers)

The 43 reflections were organized into four groups based on the country or region the student chose to focus on: - Middle East - South America - Asia - General, which are reflections that did not have specific topic/country of focus Although these labels were instructor-provided or student-selected, the content analysis often revealed thematic overlap across groups. This variable allows for regional comparisons in frequency distributions, topic proportions, and thematic alignment.

Dependent Variables

These variables are generated through text processing and serve as measurable indicators of thematic emphasis, vocabulary choice, and semantic structure.

Word Frequency

Counts of individual terms within each region's reflections or within each book chapter. This variable provides a baseline measure of dominant themes and is the foundation for comparing high-frequency vocabulary across groups.

Keyness Scores

Statistical measures (using the log-likelihood ratio) that quantify the extent to which specific words are disproportionately associated with either the reflections or the corresponding chapter. Higher absolute keyness values indicate greater divergence in linguistic focus.

Topic Proportions (LDA y-values)

For each document, the LDA model outputs a distribution over latent topics. These proportions serve as dependent variables representing the thematic mixture present in each reflection. Averaging y-values by region enables comparisons of thematic dominance across reflection groups.

Text Similarity/Thematic Alignment

Although not measured through a single formal metric, alignment is inferred through frequency overlap, keyness patterns, and topic loadings. This variable captures the degree to which students reproduce or diverge from the narrative emphasis of the book chapters.

Derived and Constructed Variables

These variables are created during preprocessing to ensure consistency across analyses and to improve interpretability.

Stemmed/Lowercased Words

All text was cleaned, normalized, lowercased, and stripped of punctuation and stopwords. This standardization reduces noise and ensures that word counts reflect thematic usage rather than grammatical variation.

Document-Term Matrices (DTMs/DFMs)

Sparse matrices representing frequencies of terms across documents. These serve as the core data structures for keyness analysis, word clouds, and LDA topic modeling.

Unified Vocabulary Feature Set

Before LDA modeling, all reflection DFMs were matched to a shared union of features to ensure consistent dimensionality across groups. This variable prevents bias in topic assignment caused by mismatched vocabularies.
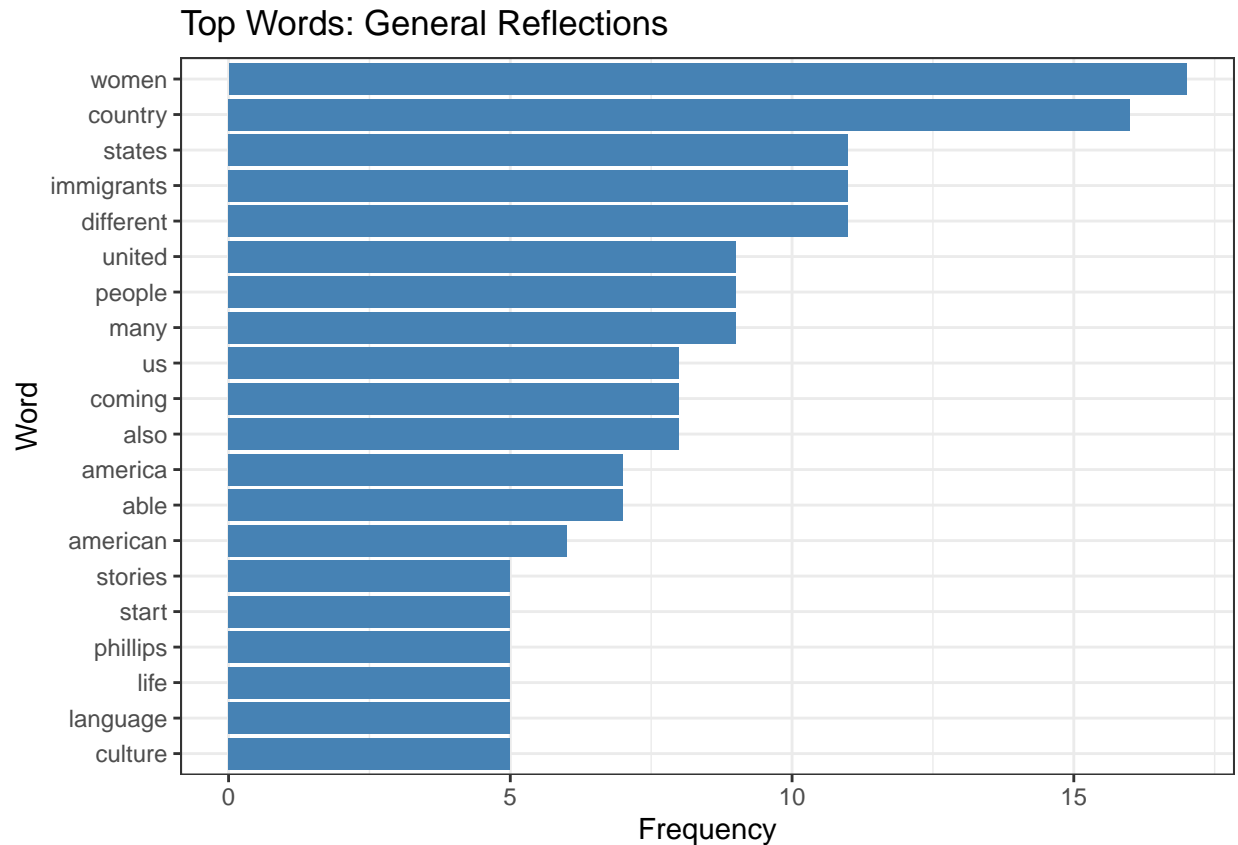
Overall, these variables allow the study to quantitatively evaluate both the content and the structure of student interpretations, enabling comparisons across regions and alignment with the narrative frameworks presented in *In the Heart of Another*.

V. Exploratory Data Analysis

The exploratory phase of the analysis focused on understanding the structure, vocabulary, and thematic emphasis of the 43 student reflection papers prior to conducting formal keyness and topic-modeling analyses. After cleaning and preprocessing the text—removing punctuation, numbers, English stopwords, and standardizing all text to lowercase, we constructed document-feature matrices (DFMs) for each reflection group: General, Middle East, South America, and Asia. These DFMs enabled frequency comparison, visualization, and later integration into a unified vocabulary for LDA.
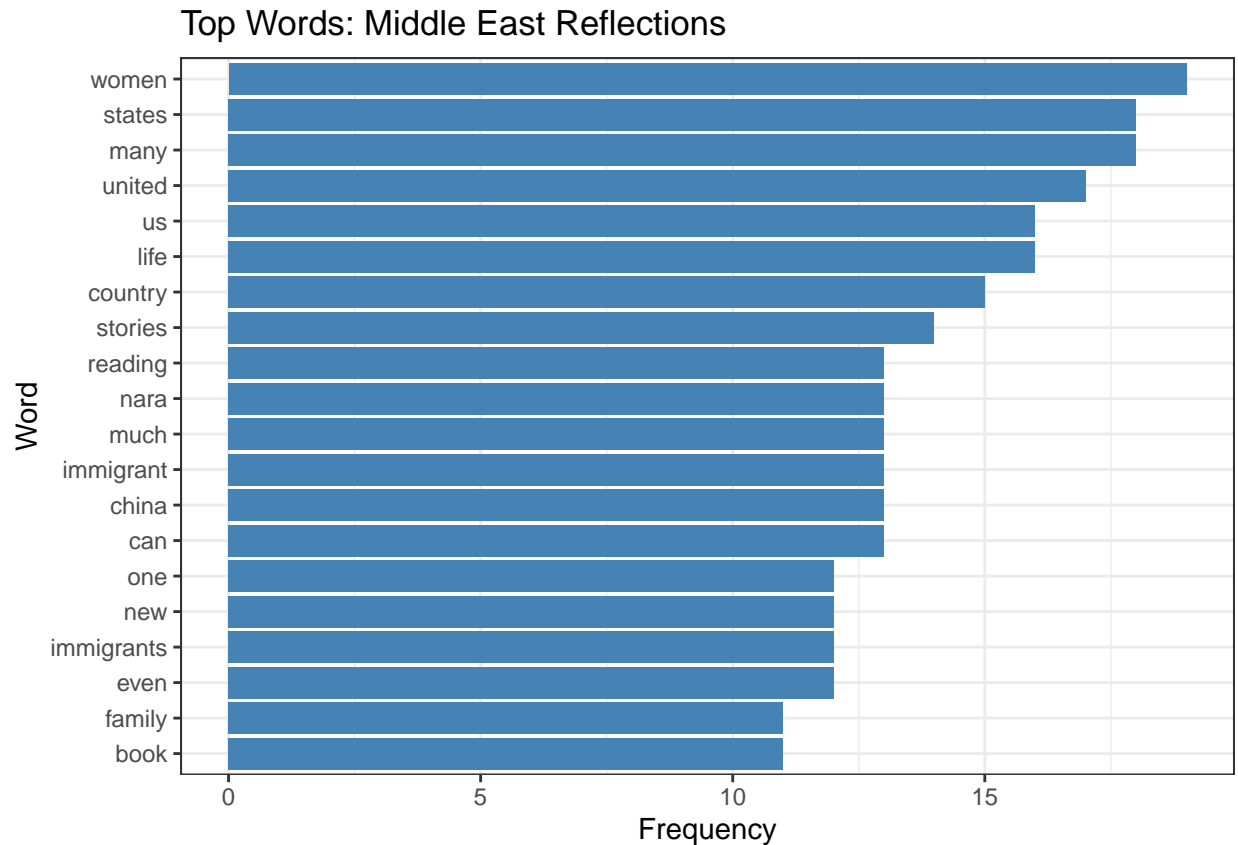
IV. Exploratory Data Analysis

```
# 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
)
```

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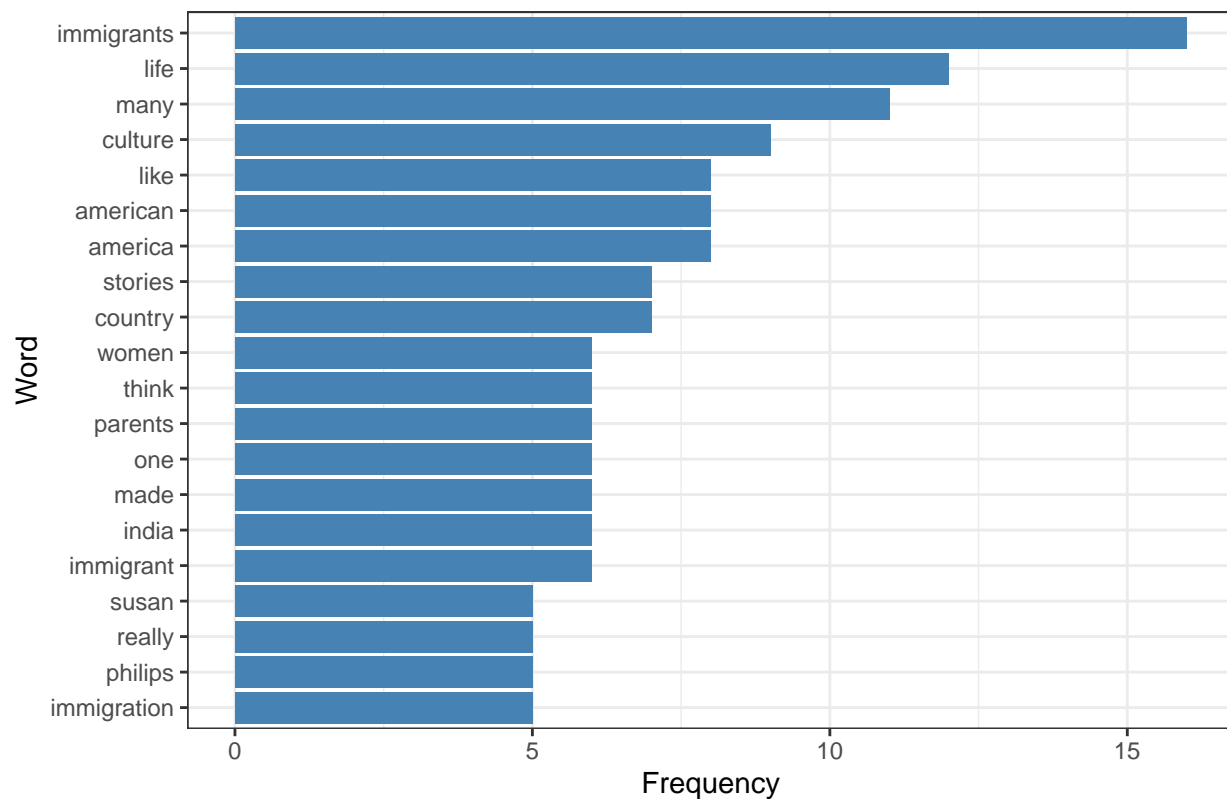
```
plot_with_commentary(
plot_top_terms(middleeast_dfm, title = "Top Words: Middle East Reflections"),
"Middle East Reflections",
  commentary = "While grouped under "Middle East", the vocabulary closely mirrors the generalized immig
)
```

Top Words: Middle East Reflections

Word	
women	
states	
many	
united	
us	
life	
country	
stories	
reading	
nara	
much	
immigrant	
china	
can	
one	
new	
immigrants	
even	
family	
book	

Frequency

```
## 
## 
## ---
## ### 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.
## ---
```
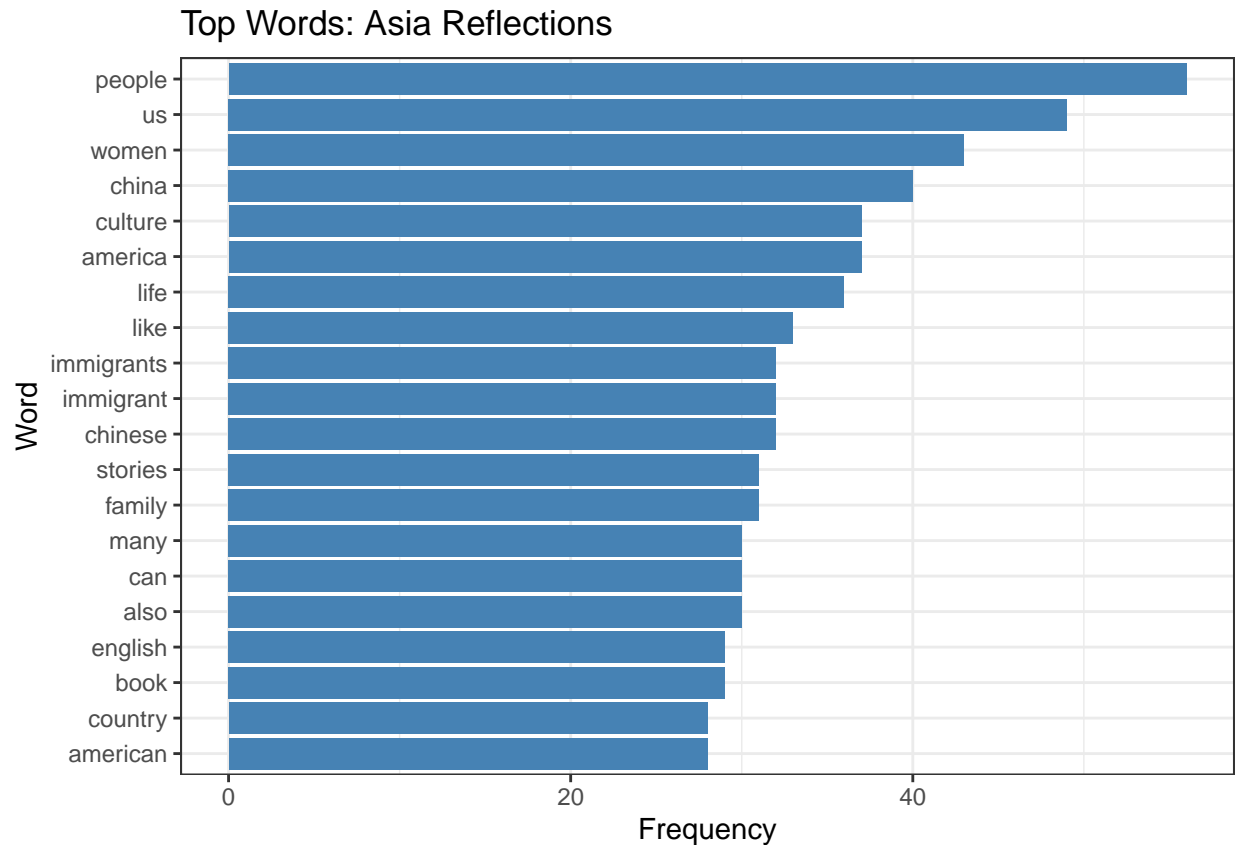
```
plot_with_commentary(
plot_top_terms(southamerica_dfm, title = "Top Words: South America Reflections"),
"South America Reflections", commentary = "Here, students introduced slightly more specificity via refe
)
```

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 en
)
```

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
## ---
```

Term-frequency bar plots revealed that all four reflection groups were dominated by broad cultural and emotional vocabulary. Words such as culture, family, life, language, people, women, and stories were consistently among the most frequent terms. Notably, region-specific or historically grounded terms—such as El Salvador, Armenia, civil war, Vietnam, or Mexico—were almost entirely absent from the top-frequency lists.

This pattern suggests that students tended to focus on universal themes of identity and migration, rather than the historical or geopolitical details emphasized in Philips' chapters.

Across groups, one surprising trend emerged: the frequent appearance of china and chinese even in reflections categorized under Middle East or South America. This indicates that students repeatedly referenced the

Xidan narrative, regardless of the nominal region of their reflection, foreshadowing later findings from the keyness and LDA analyses.
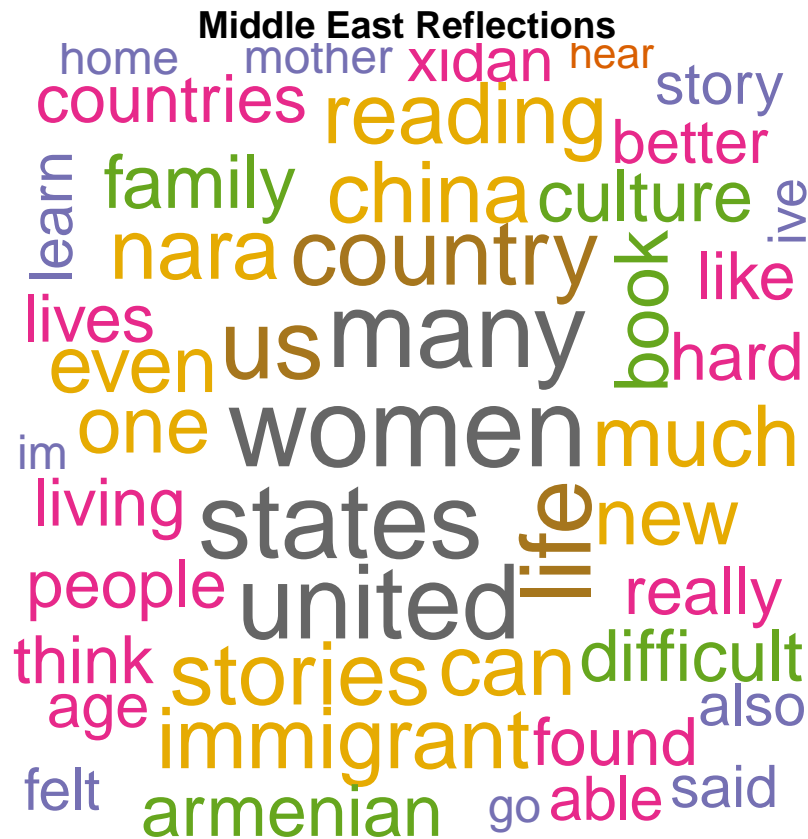
```
# wordclouds
plot_wordcloud_dfm(general_dfm, "General Reflections")
```

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")
```

Middle East Reflections

home mother xıdan hear story
countries reading better
learn family china culture
nara country like
lives us many book hard
even
im one women much
living states life new
people united really
think stories can difficult
age also
immigrant found
felt armenian go able said

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

```
## 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")
```

South America Reflections


```r
cat(paste(strwrap("The student reflections focused on South America heavily emphasized the logistical an
```

```
## 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.
```

```r
plot_wordcloud_dfm(asia_dfm, "Asia Reflections")
```

Asia Reflections

parents immigration back
learn
american english united
talk go
woman country culture think
asian dont
stories america xidan
like us life many
even can much
new feel
face also women one
time felt family still
china immigrant
others chinese book first
want
just immigrants different
learned story states able job
came cultural home
hard didnt

```r
cat(paste(strwrap("The student papers on Asia were primarily concerned with cultural identity and adapta
```
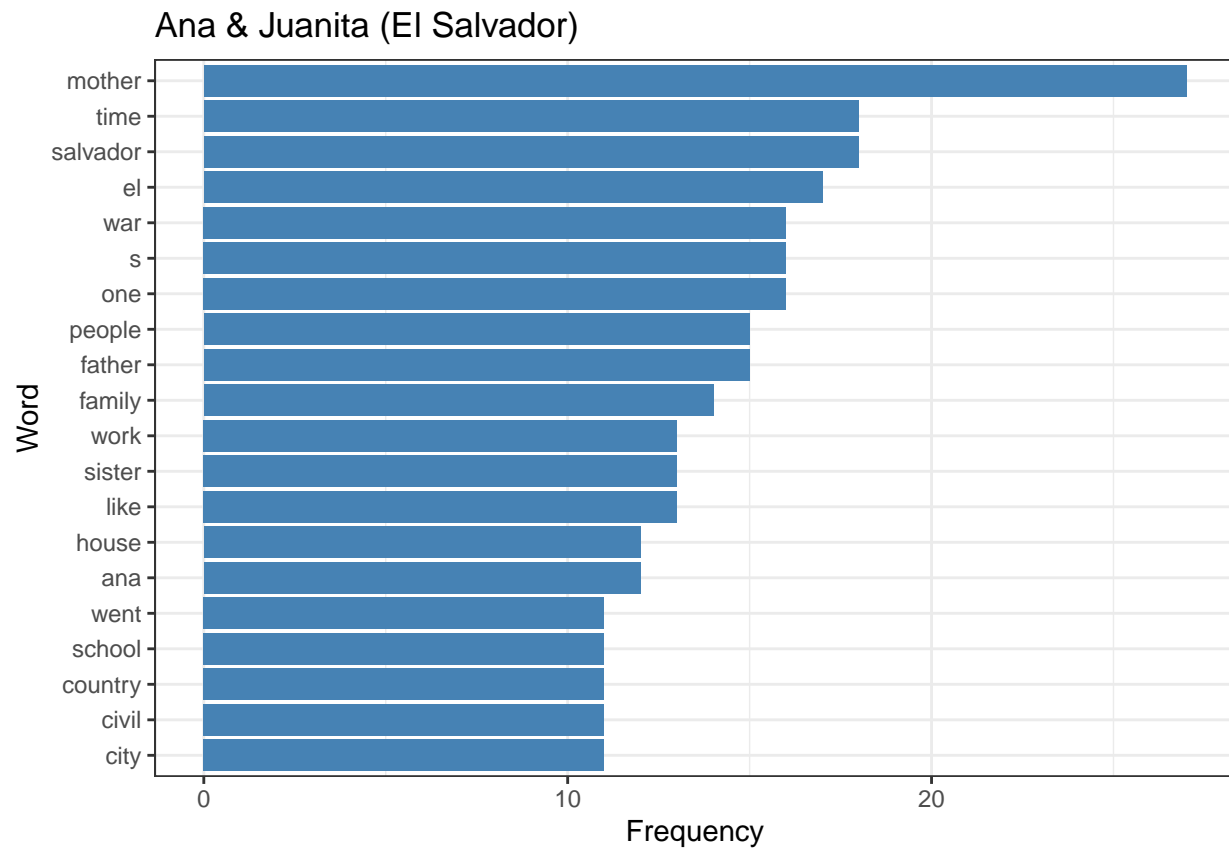
```
## The student papers on Asia were primarily concerned with cultural identity and
## adaptation to a new environment. While core themes of people, immigrants, and
## family remain central, the specific focus on culture, china, and chinese points
## to a deep dive into the specific cultural background 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.
```

Word clouds further reinforced the dominance of generalized, emotionally resonant vocabulary. The General and South America groups displayed dense clusters of words such as culture, family, immigrants, and stories, while the Middle East and Asia clouds showed additional prominence of China, English, and language.

Across all groups, names of individual chapter protagonists (Ana, Marie, Nara, Xidan, Thuynga) rarely appeared with high visual weight. Instead, students tended to speak about migration abstractly or in relation to personal experience, rather than grounding their reflections in the specific narratives from the book.

These visualizations supported a consistent preliminary conclusion: students' reflections formed a culturally introspective, identity-centered discourse, rather than a chapter-specific or regionally anchored one.
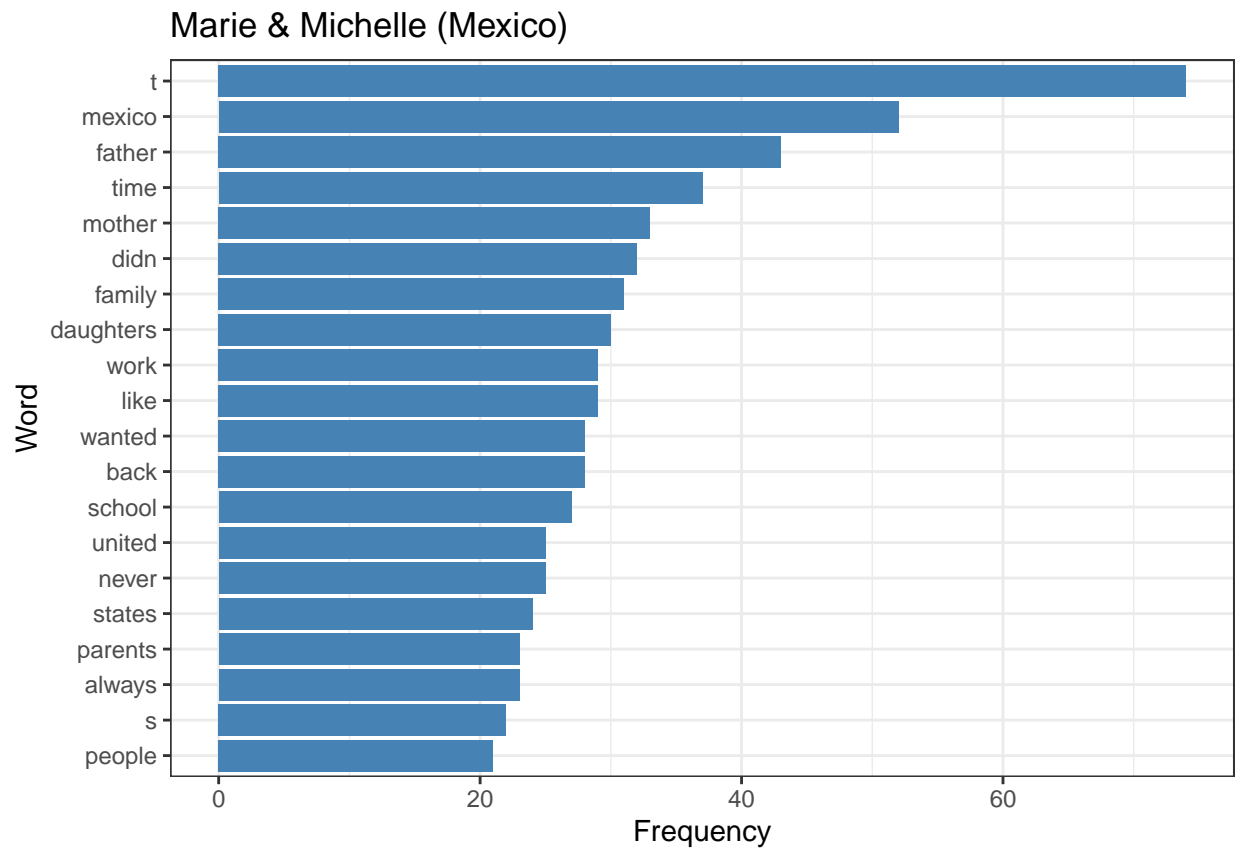
Frequency Comparison with Book Chapters

Top Terms in book chapters

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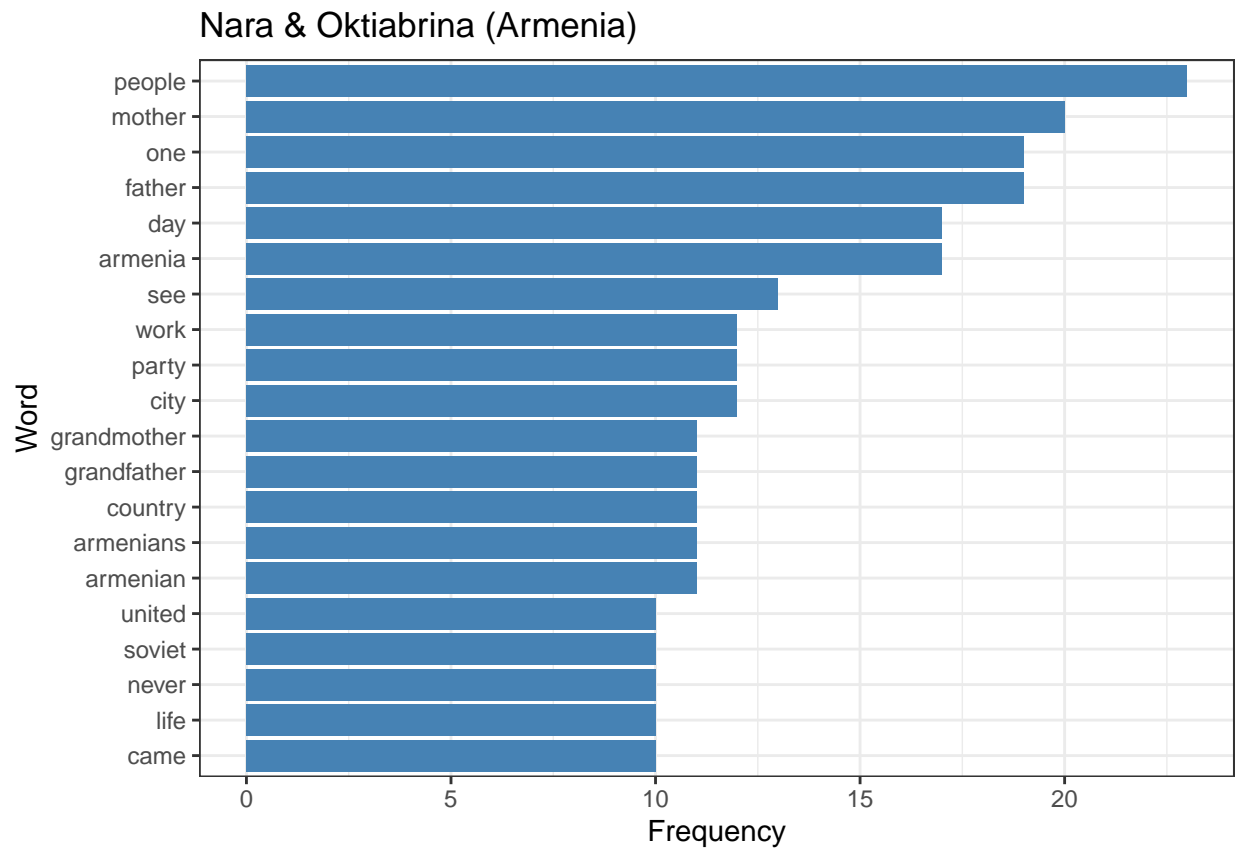
Ana & Juanita (El Salvador)


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

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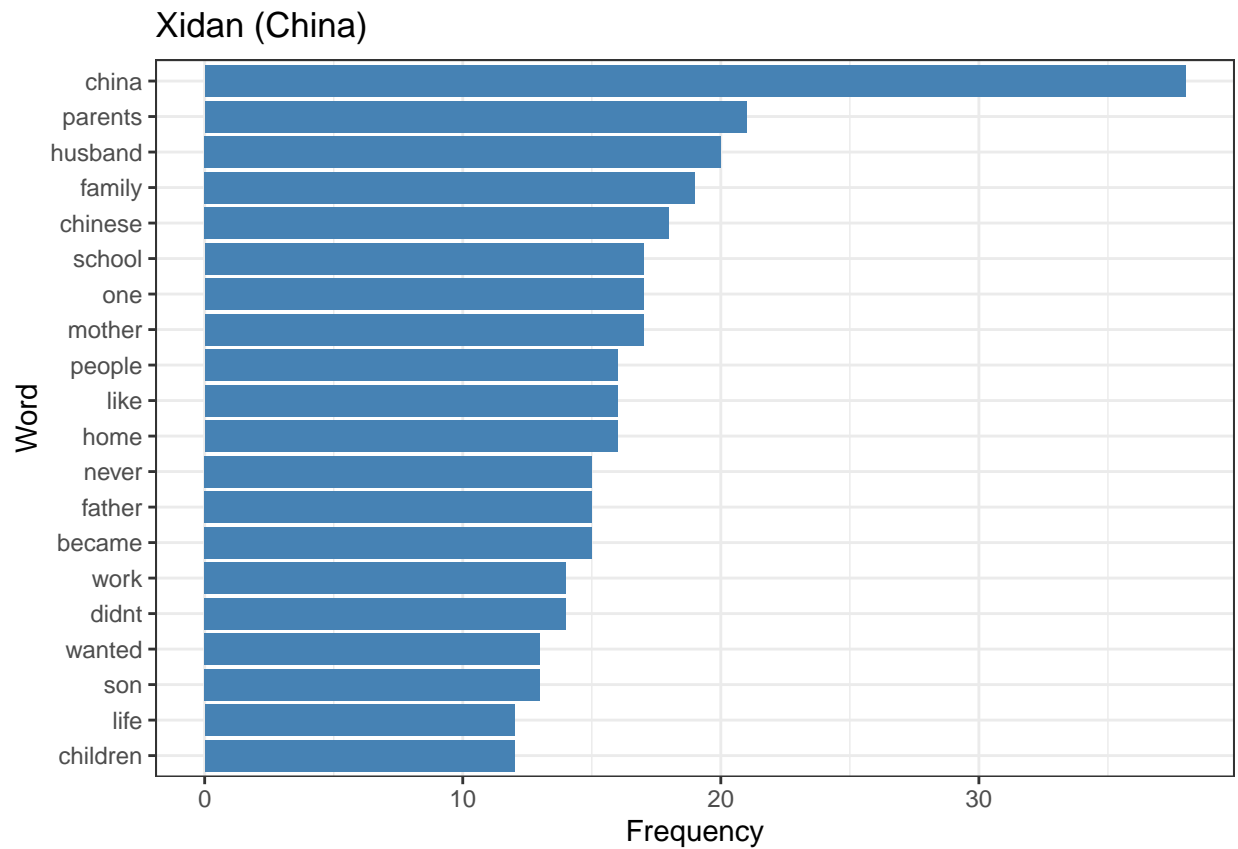
Marie & Michelle (Mexico)


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

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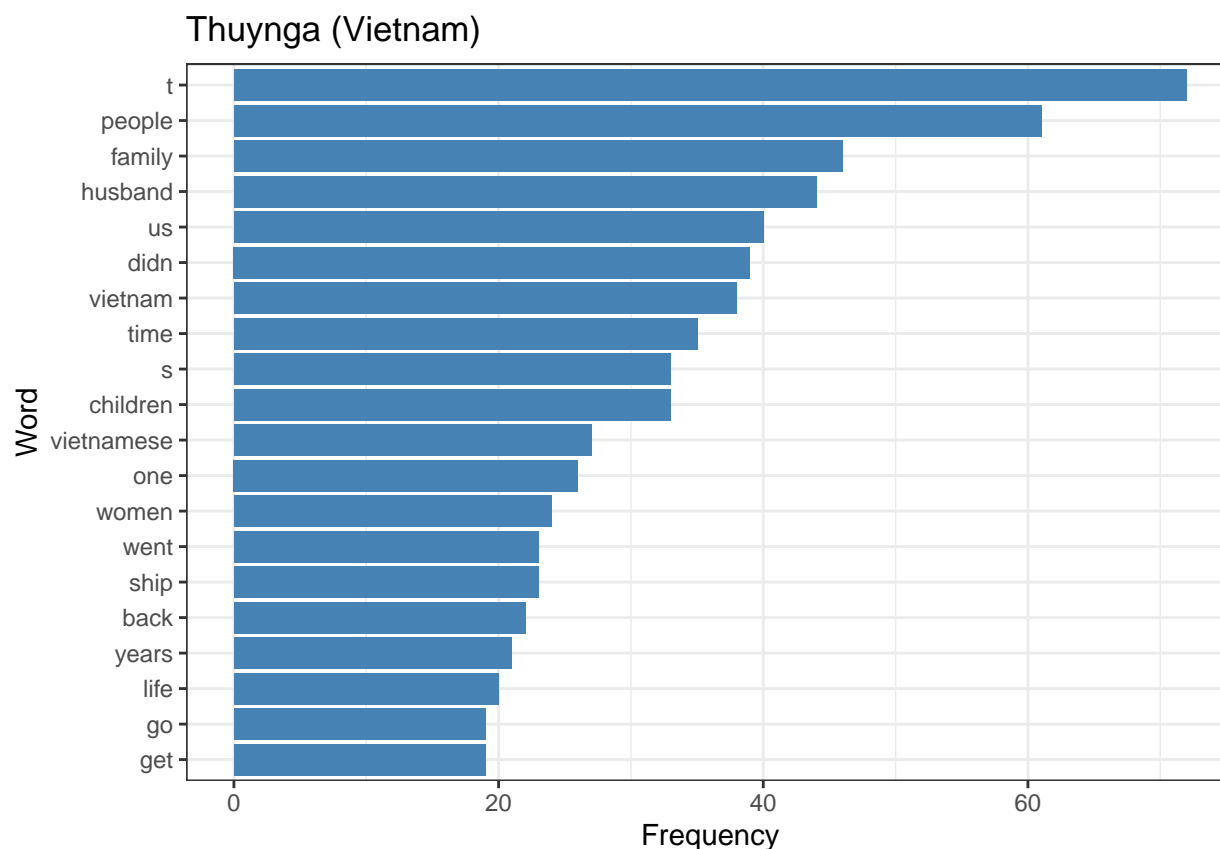
Nara & Oktiabrina (Armenia)


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

Thuynga (Vietnam)

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

To contextualize the student vocabulary, top-frequency terms from the book chapters were extracted for comparison. Unlike the reflections, the chapters contained numerous location markers and historical descriptors, including el salvador, city, civil, war, soviet, armenian, vietnam, and ship. These contrasts highlighted the gap between Philips' region-specific storytelling and the more generalized themes present in the student texts.

This mismatch aligns with expectations for reflection papers, in which students typically prioritize personal meaning over factual detail, but it also provided early evidence of weak regional alignment between reflections and the source chapters.
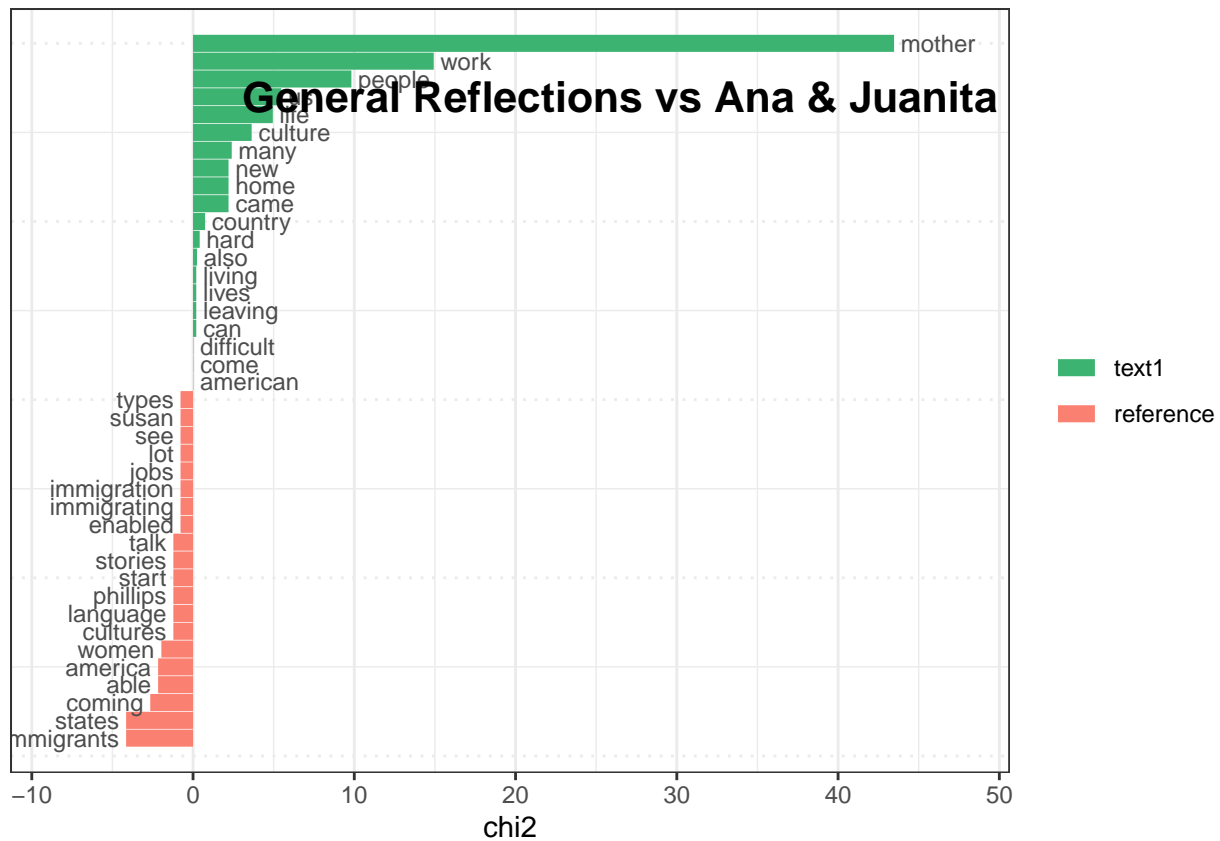
Preparation for Modeling

Prior to LDA topic modeling, the vocabulary across all groups was standardized by taking the union of all features present in the four DFMs. This ensured that each document was represented in the same feature space and prevented bias caused by uneven vocabulary overlap.

A combined reflection DTM was then constructed as the primary input for topic modeling, enabling a holistic view of thematic structure across all student texts.
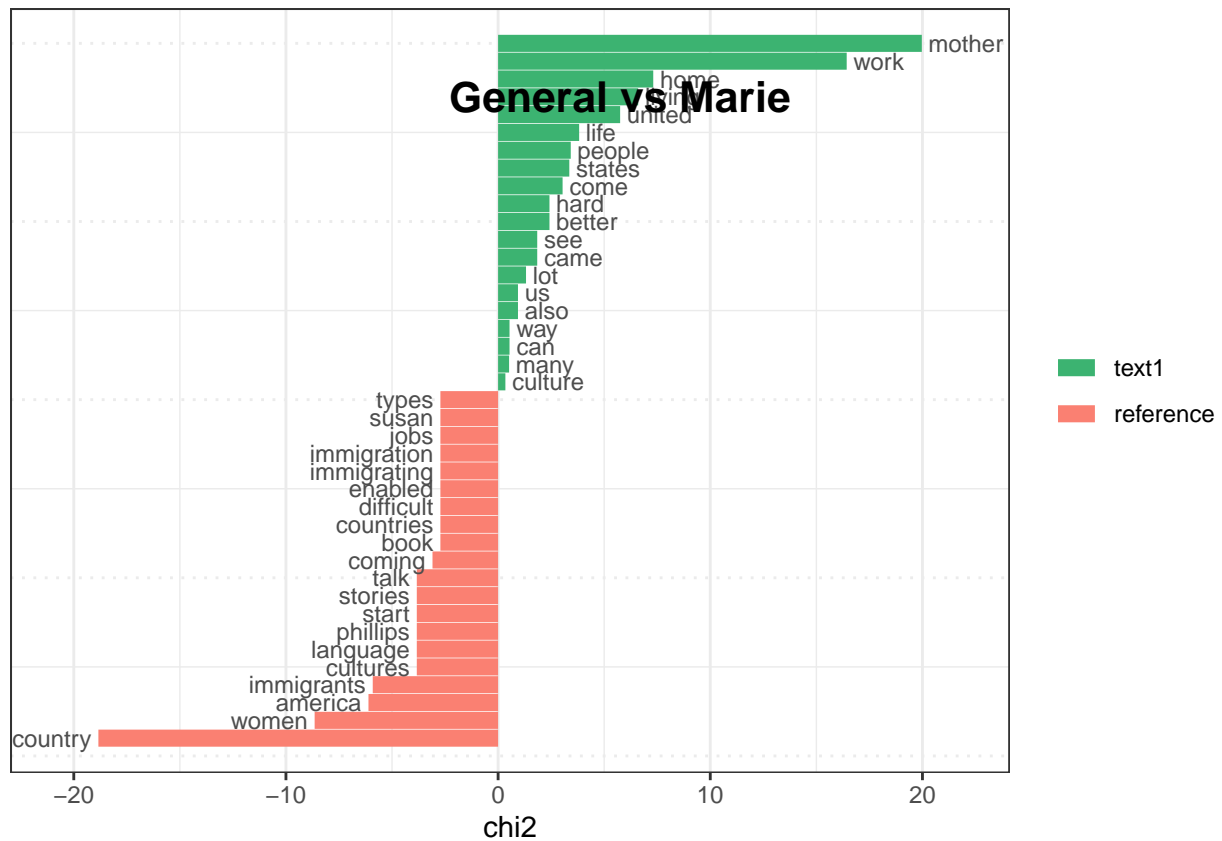
Keyness: Reflection vs. Chapter

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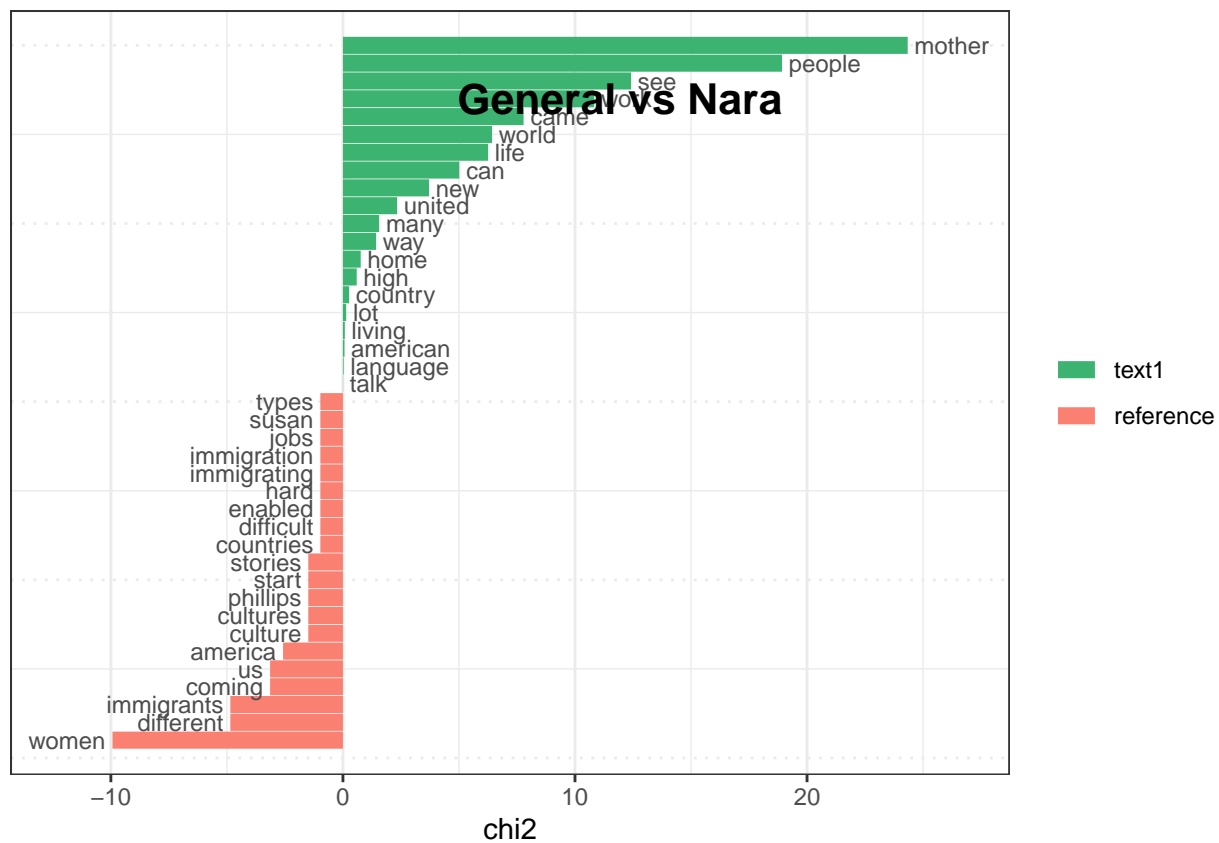


```
##
##
## ---
## ### Commentary on Keyness: General Reflections vs Ana & Juanita
##
## *(Write commentary here.)*
## ---
```

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

General vs Marie

Green (text1) bars, from top: mother, work, home, united, life, people, states, come, hard, better, see, came, lot, us, also, way, can, many, culture

Red (reference) bars, from top: types, susan, jobs, immigration, immigrating, enabled, difficult, countries, book, coming, talk, stories, start, phillips, language, cultures, immigrants, america, women, country

x-axis: chi2 (−20, −10, 0, 10, 20)

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

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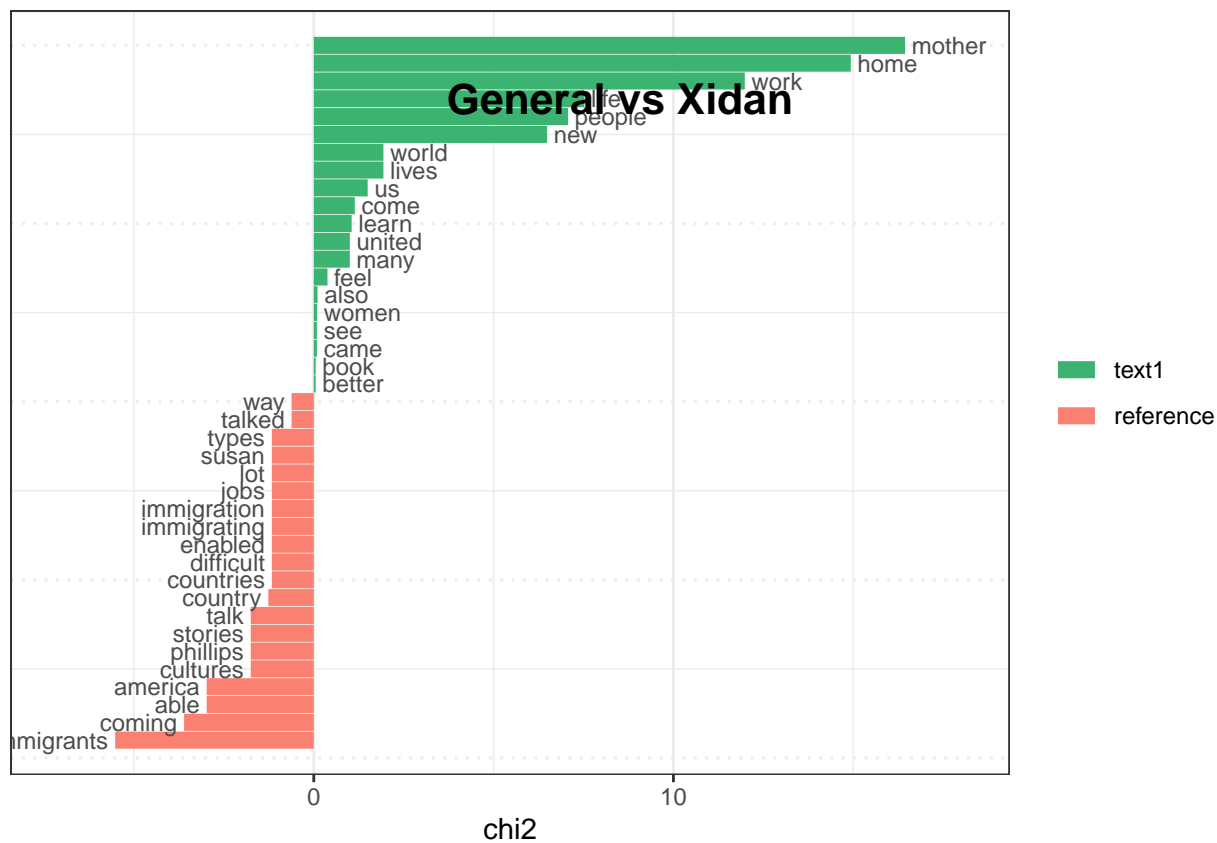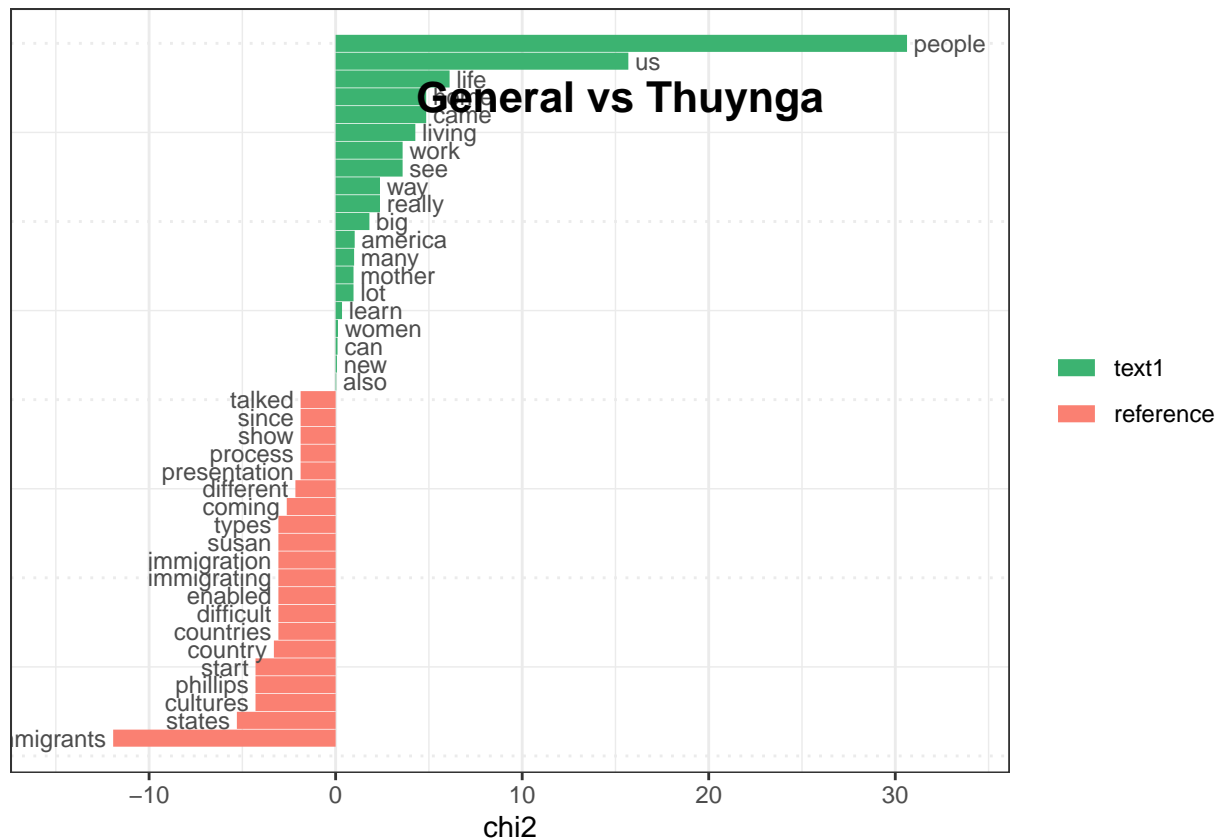
General vs Nara

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

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

General vs Xidan

Bars (green, text1): mother, home, work, are, people, new, world, lives, us, come, learn, united, many, feel, also, women, see, came, book, better

Bars (red, reference): way, talked, types, susan, lot, jobs, immigration, immigrating, enabled, difficult, countries, country, talk, stories, phillips, cultures, america, able, coming, immigrants

x-axis: chi2 (0, 10)

Legend: text1, reference

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

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

General vs Thuynga

Bar chart showing chi2 keyness values. text1 (green, positive): people, us, life, came, living, work, see, way, really, big, america, many, mother, lot, learn, women, can, new, also. reference (red, negative): talked, since, show, process, presentation, different, coming, types, susan, immigration, immigrating, enabled, difficult, countries, country, start, phillips, cultures, states, migrants.

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

LDA Topic Modeling

Top Terms per topic

```r
# 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)) +
```
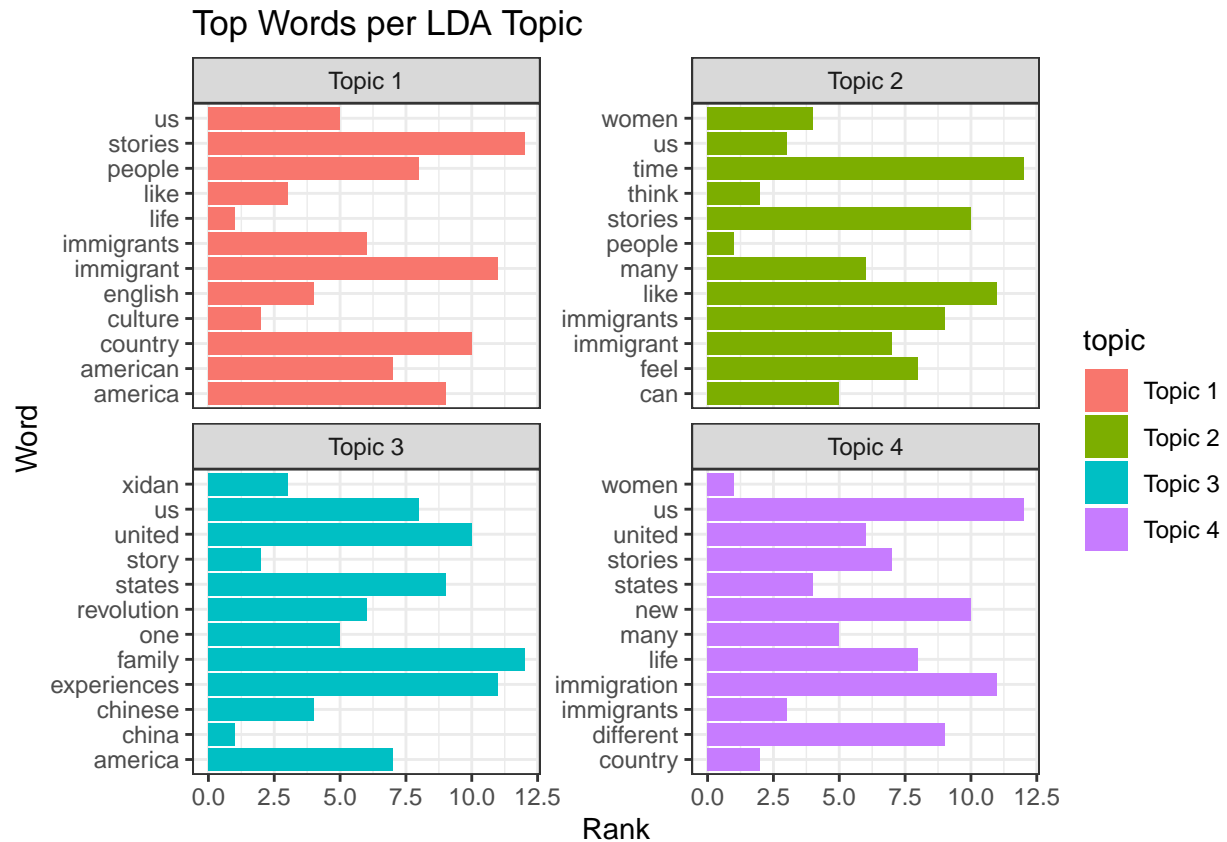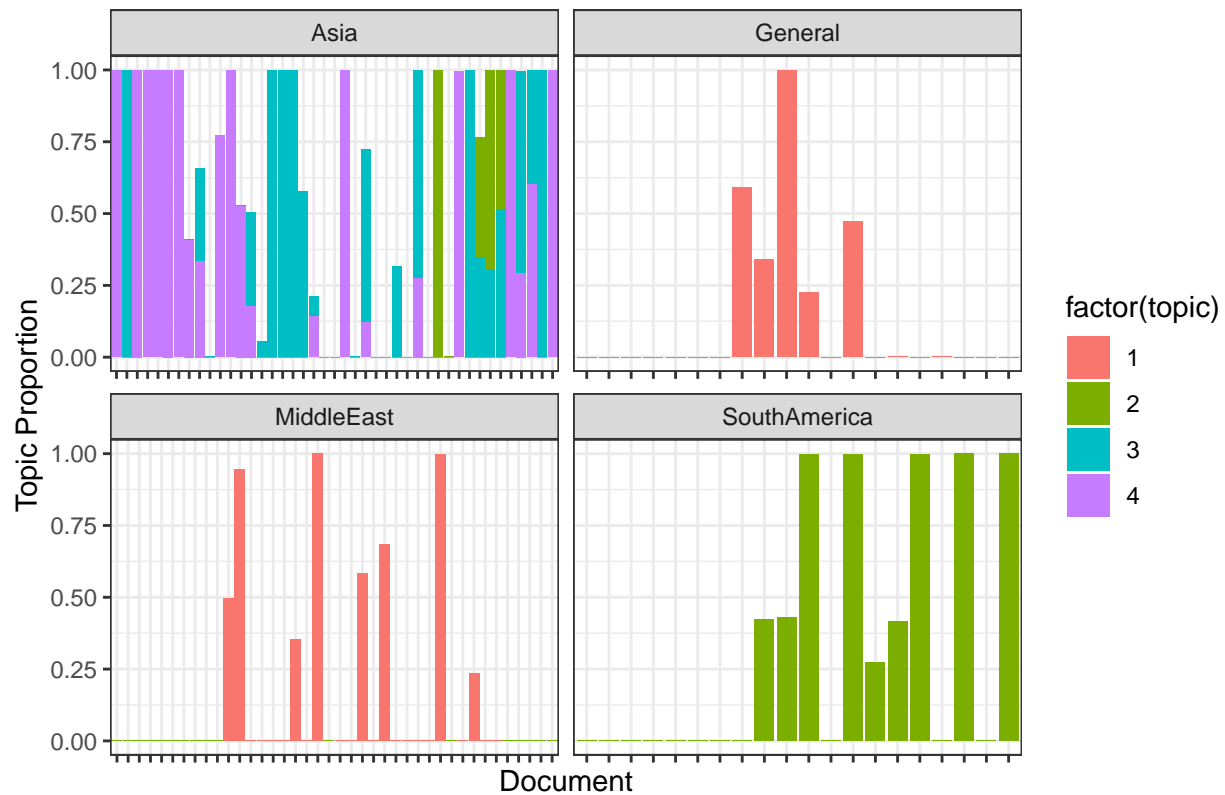
```
geom_bar(stat = "identity") +
coord_flip() +
facet_wrap(~topic, scales = "free_y") +
labs(title = "Top Words per LDA Topic", x = "Word", y = "Rank") +
theme_bw()
```


Top Words per LDA Topic

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())
```
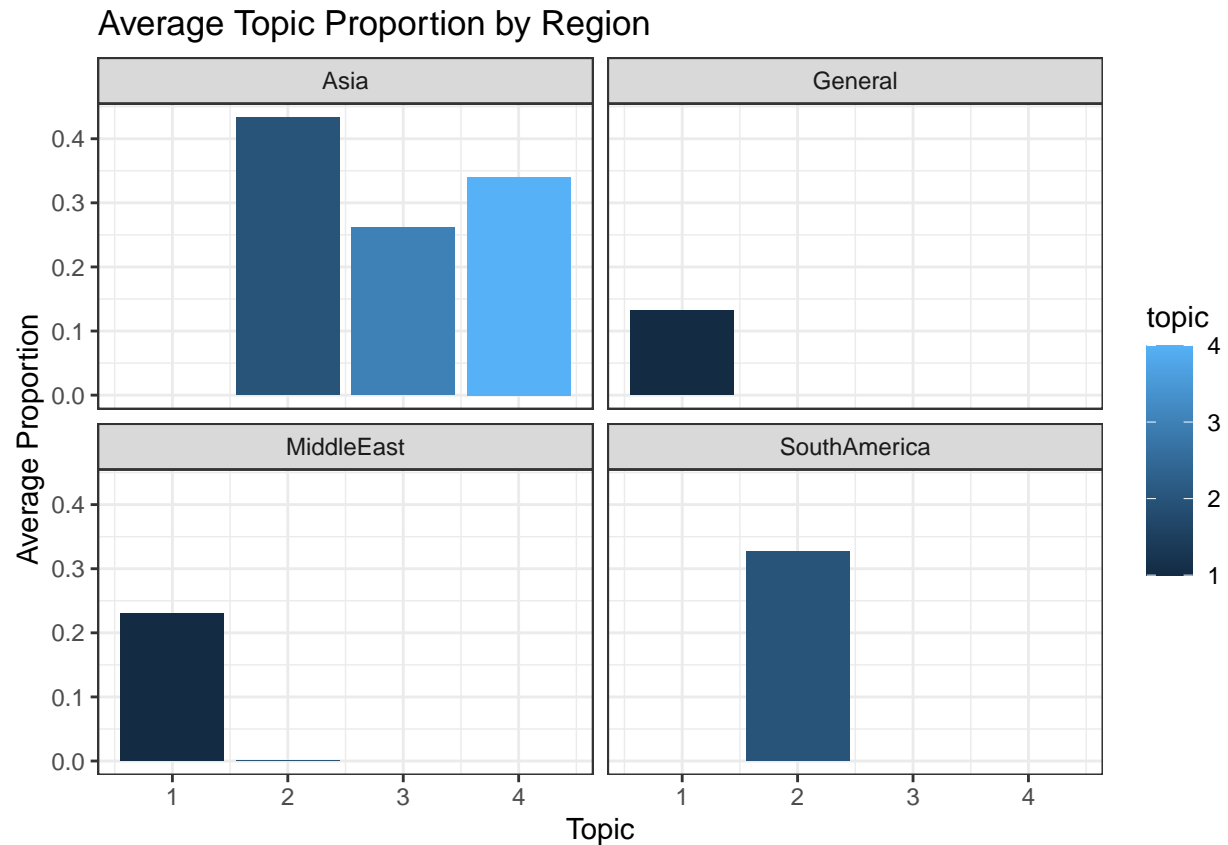
Topic Distribution per Document by Region

Average topic distribution per region

```r
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()
```

Average Topic Proportion by Region

Top Topic per Document

```r
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
```

VI. Results

The results of the analysis draw from multiple text-mining approaches, including term-frequency comparison, word-cloud visualization, keyness statistics, and LDA topic modeling. Together, these methods illustrate how students interpreted Philips' migration narratives and the extent to which their language aligned or diverged from the vocabulary of the book chapters.

1. Term Frequency Patterns in Student Reflections

Across all four reflection groups (General, Middle East, South America, Asia), the most frequent terms were highly consistent. Words such as culture, family, life, people, women, stories, and immigrants appeared prominently in nearly every group. These terms reflect broad cultural and emotional themes, suggesting that students centered their reflections around personal interpretation rather than historical or region-specific detail.

A notable and unexpected pattern was the high occurrence of "china," "chinese," and "english" across multiple groups—including those not associated with Asia. Even in the Middle East and South America reflections, China-related terms appeared more frequently than region-specific vocabulary like El Salvador, Armenia, or Mexico. This early signal pointed to a strong student connection to the Xidan chapter.

2. Word Clouds Reveal Emphasis on Universal Themes

Word cloud visualizations reinforced the dominance of identity and cultural themes. In all reflection groups, the largest and most frequent terms centered around: - migration and adjustment (migrating, challenges, leaving) - social and cultural identity (culture, language, family, women) - personal stories (story, people, experiences)

Region-specific proper nouns, which were strong in the original book chapters, were visually minimal or absent. Only the Asia group showed a strong presence of country markers, with china and chinese appearing at large scale—further supporting the trend of heightened student engagement with the Xidan narrative.

3. Comparison with Book Chapters: Strong Divergence in Vocabulary

When comparing top chapter terms to reflection terms, a clear contrast emerged as book chapters emphasized common ideas like geopolitical context (civil war, soviet, vietnam, el salvador), place names and historical markers, family roles tied to specific narratives (husband, father, parents), and narrative characters (Ana, Nara, Thuynga, etc.).

Student reflections emphasized themes like cultural identity, immigrant experience in general, language and education, and emotional interpretation.For example, the Ana & Juanita chapter contained frequent references to civil, war, city, and el salvador, but student reflections on these narratives rarely incorporated such terms. Instead, reflections featured general cultural descriptors like life, culture, and family.

This consistent mismatch indicates that although students engaged with the themes of migration, they did not replicate the historical specificity embedded in Philips' writing.

4. Keyness Analysis: Distinct Differences in Linguistic Focus

Keyness comparisons between student reflections and each book chapter revealed systematic divergences. Chapter texts were characterized by contextual details—place names, time markers, political terms, and family-role vocabulary while student reflections showed significantly higher keyness for culture, america, english, immigrant, story, and leaving. Across all five chapters, the reflections' vocabulary leaned toward personalized and contemporary interpretations of migration, rather than the narrative- or location-specific

language used in the book. The Xidan (China) chapter displayed the most thematic overlap with students' vocabulary—particularly around family, education, and identity—yet even here students tended to reframe the narrative toward American immigrant contexts (e.g., english, america).

5. LDA Topic Modeling: Four Distinct Themes

A four-topic Latent Dirichlet Allocation model was applied to all reflection documents. The resulting topics were interpretable and aligned closely with the thematic patterns observed in earlier analyses.

Topic 1: Cultural Assimilation and Identity Terms related to culture, language, english, america, and learning. This topic captured reflections focusing on adapting to a new cultural environment.

Topic 2: Emotional Interpretation of Stories Words like people, life, stories, women, challenges, family. This topic reflects empathetic responses and students' affective reactions to the narratives.

Topic 3: China-Centered and Generational Migration Dominated by china, chinese, family, school, parents, work. This topic clearly corresponds to the Xidan chapter and was disproportionately represented across regions.

Topic 4: Gendered Experiences and Immigrant Struggles Featuring terms like women, community, immigrants, stories, culture. This topic reflects broader gender and social justice themes.

Topic Distribution by Region The Asia reflections displayed the strongest concentration in Topic 3 (China-centered narrative). The Middle East and South America groups unexpectedly showed moderate activation of this same topic—confirming that students referenced the Xidan narrative regardless of region and the General reflections were more mixed but leaned toward Topics 1 and 2, emphasizing cultural assimilation and emotional interpretation.

6. Summary of Key Findings

Across all analytical methods, we saw three core patterns. Students gravitated toward universal, personal themes rather than region-specific historical context. Vocabulary focused on culture, language, identity, and family instead of country names or historical events. The China (Xidan) chapter dominated student reflection language as even in unrelated regional groups, China-related terms appeared frequently, and a full LDA topic centered on Chinese migration experiences.Finally, reflections showed partial thematic alignment but strong lexical divergence from book chapters as students resonated with the emotional core of migration narratives but reinterpreted them through contemporary, culturally familiar lenses.

VII. Conclusion

The analysis of student reflections demonstrates that learners resonate emotionally with migration narratives far more than they engage with historical or geopolitical context. Across term frequency and word cloud analyses, students consistently emphasized human-centered themes such as identity and belonging (life, culture, family, people), empathy and connection (feel, stories, think), and gender and resilience (women). In contrast, the book chapters highlighted concrete historical and regional identifiers, including the civil war in El Salvador, Soviet Armenia, and migration by ship from Vietnam. This contrast suggests that while students internalized narratives of personal struggle and human experience, they did not deeply engage with the structural or historical frameworks of the chapters.

The regional categorization of reflections does not always align with the content. For instance, folders labeled "Middle East" and "South America" frequently contained high-frequency references to China and generalized U.S. immigration terms (america, immigrant, states, women). This indicates that students often mapped specific global narratives onto a broader U.S. immigration discourse. Labels assigned to reflection folders were not strong predictors of the linguistic content, and the Xidan (China) chapter disproportionately influenced vocabulary across all regions. This pattern reflects student interpretation behavior rather than an artifact of data processing or modeling.

Topic modeling revealed distinct thematic clusters, yet reflections remained dominated by personal themes. The four-topic LDA model can be interpreted as: cultural identity and assimilation, emotional response and empathy, China-centered narrative influence and generational migration, and gender, mobility, and societal-level immigration patterns. Topic distribution plots show that most documents were heavily dominated by a single topic (gamma > 0.95), indicating thematic clarity but limited diversity within individual reflections.

Students framed migration primarily as a lived experience rather than a historical or political event. The absence of high-keyness geographic terms like El Salvador, Mexico, Vietnam, or Armenia demonstrates that reflections centered on personal and familial journeys, cultural identity transitions, and shared immigrant experiences rather than the country-specific political forces highlighted in Philips' chapters. Language and culture emerged as central challenges across all regions, with key terms such as culture, english, immigrants, and american indicating that students prioritized cultural integration and language acquisition over geopolitical context.

Gender-focused interpretation was another strong emergent theme. Both LDA Topic 4 and word clouds highlighted women, suggesting that female narratives, resilience, and maternal figures resonated strongly with students, regardless of whether gender was explicitly foregrounded in the chapters. This emphasizes how identity and lived experience functioned as unifying lenses for reflection.

The prominence of China in student reflections is particularly notable. The Xidan chapter offered a narrative that was emotionally intense, had a clear character arc, and emphasized themes like family conflict, generational tension, and cultural negotiation. These features made the story more immediately relatable than other chapters, which were more historically or contextually dense. Furthermore, students may have had personal or vicarious familiarity with Asian immigrant experiences, increasing relatability. China's high visibility in global media and education also provided a recognizable reference point, making it easier for students to connect their reflections to the narrative.

Despite a limited dataset of 43 reflections, the study produced meaningful thematic alignment with Philips' core themes. Students clearly engaged with migration, identity, and cultural negotiation, even as linguistic specificity and historical context were often abstracted. This confirms that reflective writing can be systematically analyzed through text mining to reveal both conceptual alignment and divergence in interpretation.

Several limitations should be noted. The sample size restricts generalizability, and the folder labels do not consistently reflect content, complicating region-specific analysis. Preprocessing choices, including stemming and stopword removal, could influence frequency and keyness metrics. Finally, the analysis focuses on textual data alone, which may overlook broader dimensions of student engagement, such as affective or classroom interaction factors. Complementary methods like surveys or interviews could provide additional insights.

In terms of educational practice, the findings suggest that emotionally resonant narratives, such as Xidan's story, are highly effective for fostering reflection and empathy. Educators can enhance learning by pairing these narratives with structured historical or cultural context, incorporating diverse perspectives, highlighting identity and intersectionality, and providing scaffolding for less familiar regions. Reflection exercises, repeated exposure, and guided discussion can help students connect emotionally and cognitively, balancing personal resonance with contextual understanding. Overall, this study demonstrates how text mining of student reflections can reveal meaningful patterns in engagement, offering valuable insights for curriculum design in global migration and diversity education.

VIII. Recommedations

As diversity continues to become a central topic in education and society, it is increasingly important for classrooms to engage students with stories that foster empathy, cultural awareness, and personal connection. Based on the analysis of student reflections and engagement patterns, here are some takeaways from this study:

Leverage emotionally resonant narratives: Stories like Xidan's, which include personal struggle, family dynamics, and identity negotiation, are highly effective at encouraging students to reflect deeply. Educators should prioritize narratives that balance human experience with cultural or historical context to maximize engagement.

Incorporate multiple perspectives: While personal stories drive emotional engagement, it is also important to help students situate these experiences within broader historical, geopolitical, and social frameworks. Activities that connect individual narratives to systemic or regional contexts can deepen understanding without reducing emotional resonance.

Highlight identity and intersectionality: Students respond strongly to themes of gender, generational identity, and cultural negotiation. Curricula should include stories representing a range of identities and experiences, encouraging reflection on how factors like gender, ethnicity, and migration history shape lived experiences.

Use recognizable entry points: Students often connect most easily to stories with familiar or highly visible cultural references. When introducing less well-known regions or histories, teachers can provide scaffolding through background context, visuals, or guided discussions to make engagement more immediate.

Encourage reflective writing as a tool for empathy: Reflection exercises help students internalize complex ideas about migration, culture, and identity. Structured prompts that emphasize personal connection alongside factual understanding can support both emotional and cognitive learning outcomes.

Integrate repeated exposure and discussion: Narratives reinforced through lectures, discussions, or multimedia presentations help students retain and relate to material. Revisiting stories multiple times encourages deeper reflection and vocabulary adoption.

Balance emotional and contextual learning: While emotionally powerful stories are critical, pairing them with analytical discussions on historical, social, and political frameworks ensures students develop a more comprehensive understanding of migration, diversity, and global experiences.

In sum, teaching diversity effectively requires blending emotionally engaging narratives with structured contextual learning. Stories like Xidan's illustrate the power of personal narratives to foster empathy, while thoughtful integration of historical and cultural context ensures that students not only relate but also understand.