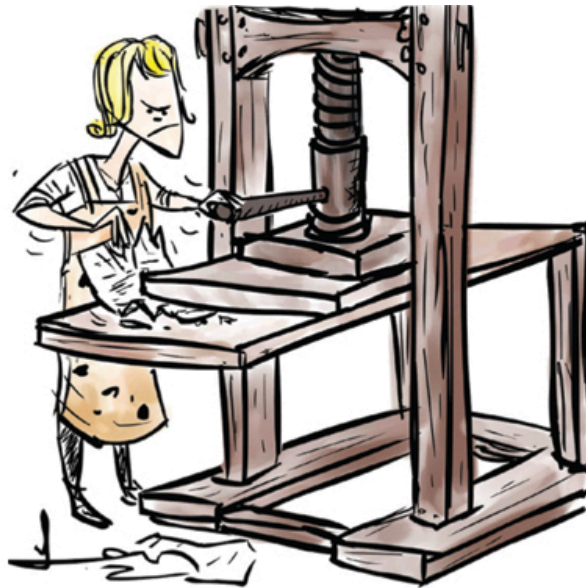


Gutenberg Library Analysis





search for books

- Browse Catalog
- Bookshelves
- Main Page
- Categories
- Contact Info

Project Gutenberg appreciates your donation!

[Donate](#)

- Why donate?

in other languages

- Português
- Deutsch
- Français



Free ebooks - Project Gutenberg

[Book search](#) · [Book categories](#) · [Browse catalog](#) · [Mobile site](#) · [Report errors](#) · [Terms of use](#)

Some of the Latest Books



Welcome

Project Gutenberg offers over 57,000 free eBooks. Choose among free epub books, free kindle books, download them or read them online. You will find the world's great literature here, with focus on older works for which copyright has expired. Thousands of volunteers digitized and diligently proofread the eBooks, for enjoyment and education.

No fee or registration is required. If you find Project Gutenberg useful, please consider a small [donation](#), to help Project Gutenberg digitize more books, maintain our online presence, and improve Project Gutenberg programs and offerings. Other ways to help include [digitizing more books](#) 📄, [recording audio books](#) 🎧, or [reporting errors](#).

News



[Project Gutenberg](#)
[Mobile Site](#)



Project Gutenberg offers 57,692 free ebooks to download.

[Donate](#)

[Flattr this!](#)

[Search](#) [Latest](#) [Terms of Use](#) [Donate?](#) [Mobile](#)

[Help](#)

[Project Gutenberg](#) > [57,692 free ebooks](#) > [115 by Robert Louis Stevenson](#)

The Strange Case of Dr. Jekyll and Mr. Hyde by Robert Louis Stevenson

No cover available



Download

Bibrec

Download This eBook

Format ?	Size	?	?	?
Read this book online: Generated HTML (with images)	174 kB			
Read this book online: Generated HTML (no images)	174 kB			
EPUB (with images)	71 kB			
EPUB (no images)	71 kB			
Kindle (with images)	296 kB			
Kindle (no images)	296 kB			
Plain Text UTF-8	159 kB			
More Files...				

gutenbergr: Search and download public domain texts from Project Gutenberg

David Robinson

2018-01-26

The gutenbergr package helps you download and process public domain works from the [Project Gutenberg](#) collection. This includes both tools for downloading books (and stripping header/footer information), and a complete dataset of Project Gutenberg metadata that can be used to find words of interest. Includes:

- A function `gutenberg_download()` that downloads one or more works from Project Gutenberg by ID: e.g., `gutenberg_download(84)` downloads the text of Frankenstein.
- Metadata for all Project Gutenberg works as R datasets, so that they can be searched and filtered:
- `gutenberg_metadata` contains information about each work, pairing Gutenberg ID with title, author, language, etc
- `gutenberg_authors` contains information about each author, such as aliases and birth/death year
- `gutenberg_subjects` contains pairings of works with Library of Congress subjects and topics

Project Gutenberg Metadata

This package contains metadata for all Project Gutenberg works as R datasets, so that you can search and filter for particular works before downloading.

The dataset `gutenberg_metadata` contains information about each work, pairing Gutenberg ID with title, author, language, etc:

```
library(gutenbergr)
gutenberg_metadata
```

```
## # A tibble: 51,997 x 8
##   gutenberg_id title          author    guten... lang... gutenber... rights has_...
##           <int> <chr>          <chr>    <int> <chr> <chr>    <chr> <lgl>
```

Goals

Topics → Subgenres

Sentiment → Plot

Recommendations

Topics → Subgenres

Clean

- Fantasy, Adventure, Fiction, Children, Myth, Folklore
- ~6% of Gutenberg Library

Lemmatize

- The cars are different → the car be differ

TFIDF

- 12,000 term limit

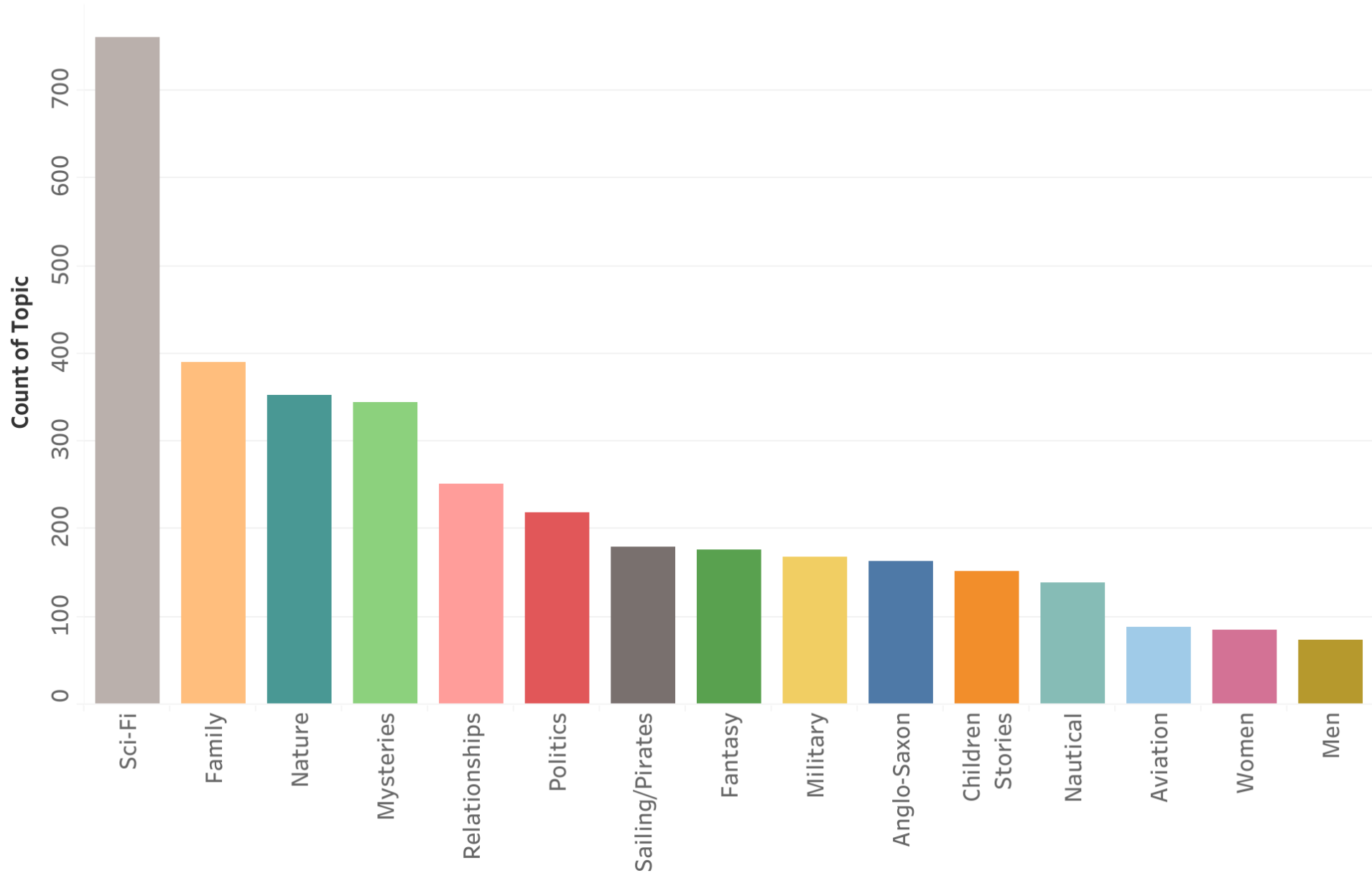
NMF

- ~~LDA~~
- 15 topics

Clusters

- K-means
- T-SNE

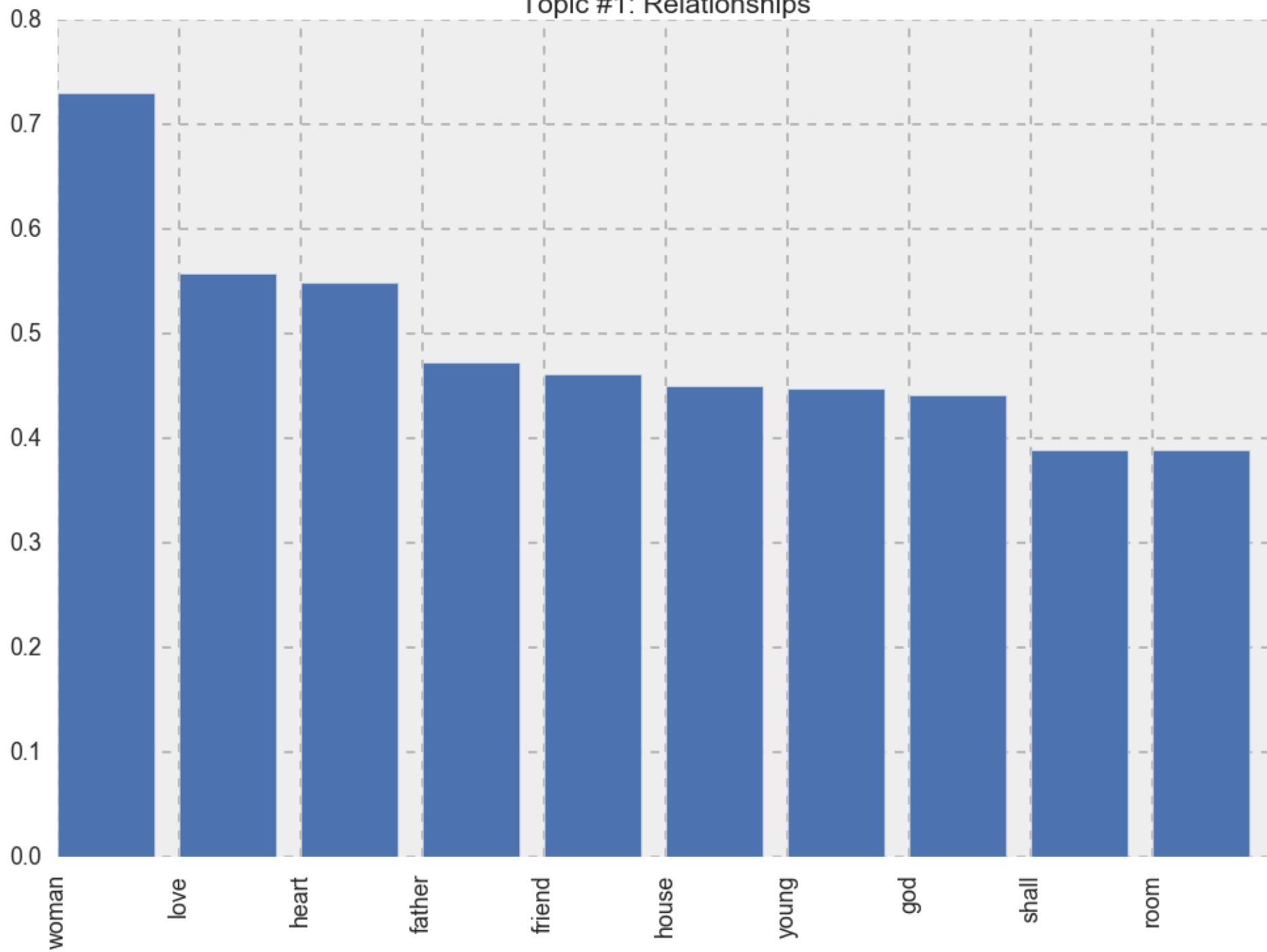
Distribution of Topics/Clusters



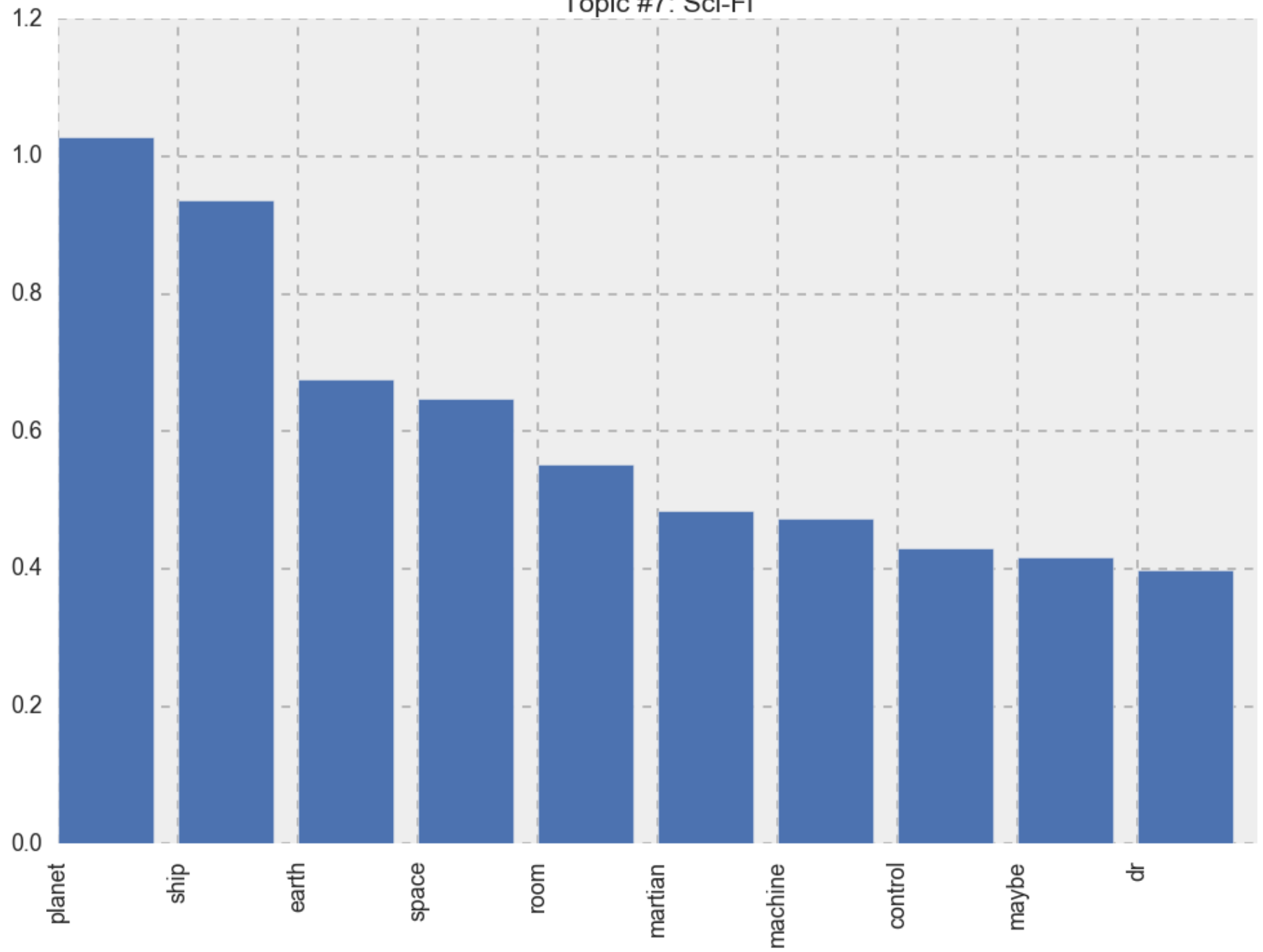
●



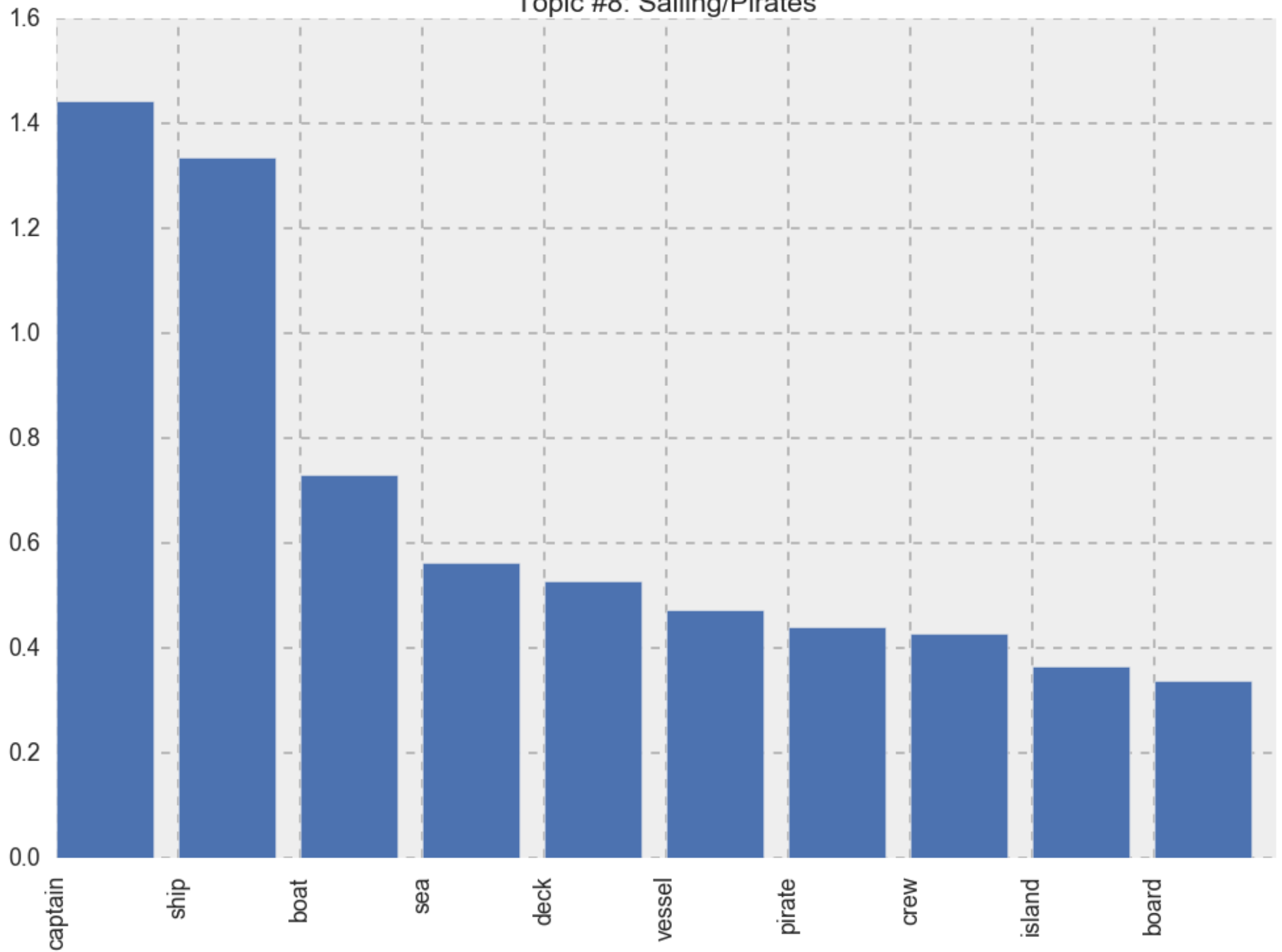
Topic #1: Relationships



Topic #7: Sci-Fi



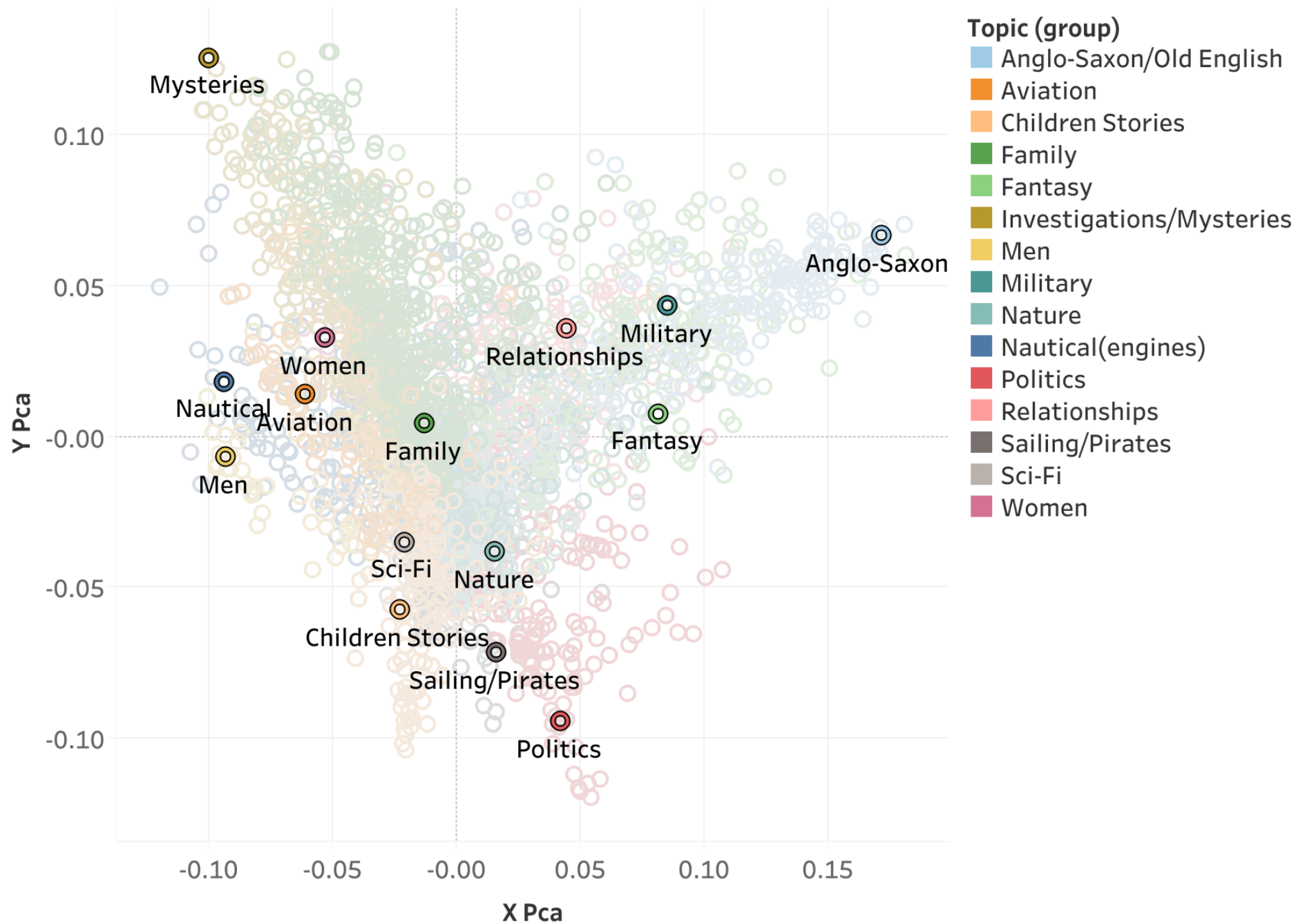
Topic #8: Sailing/Pirates



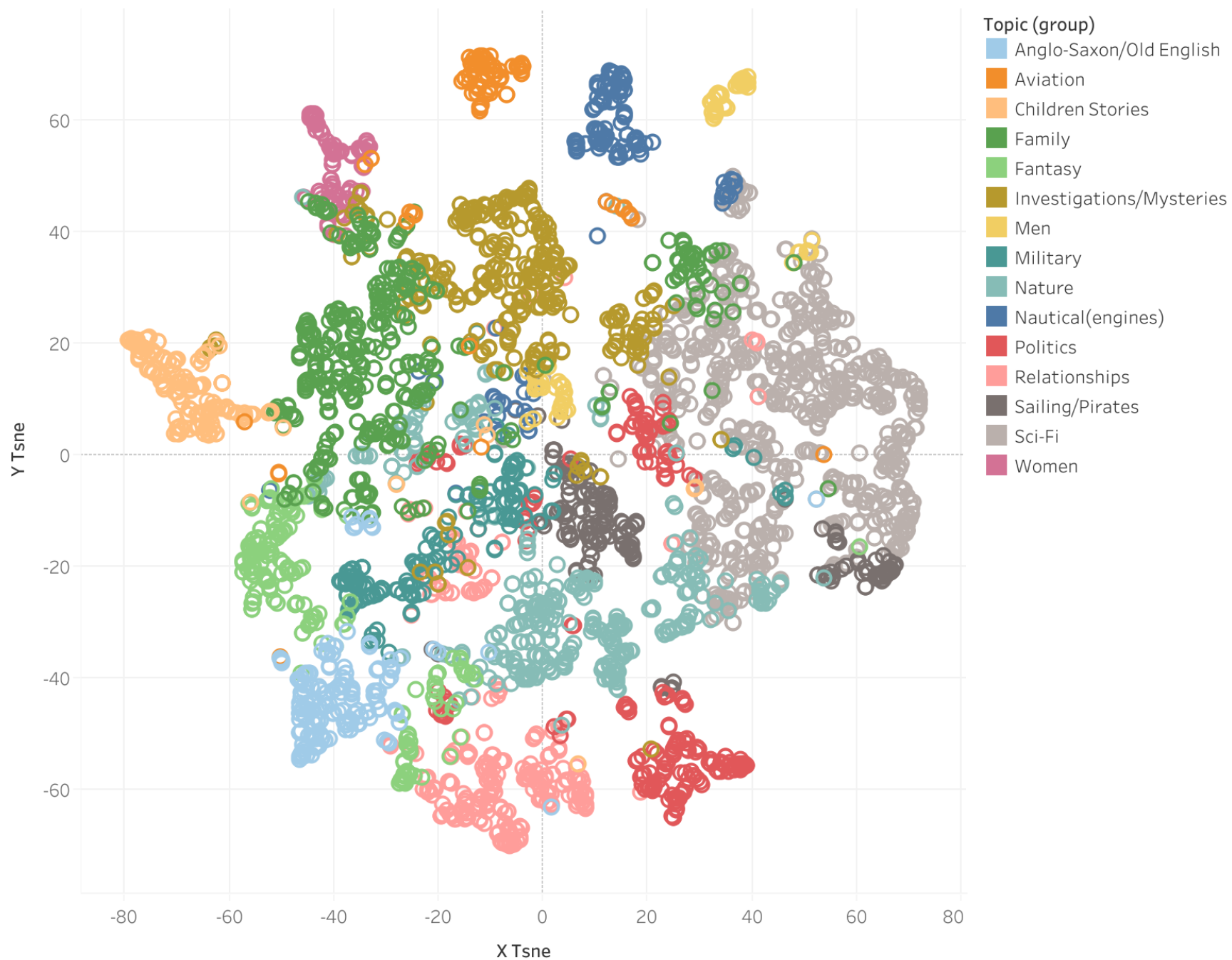
PCA clustering



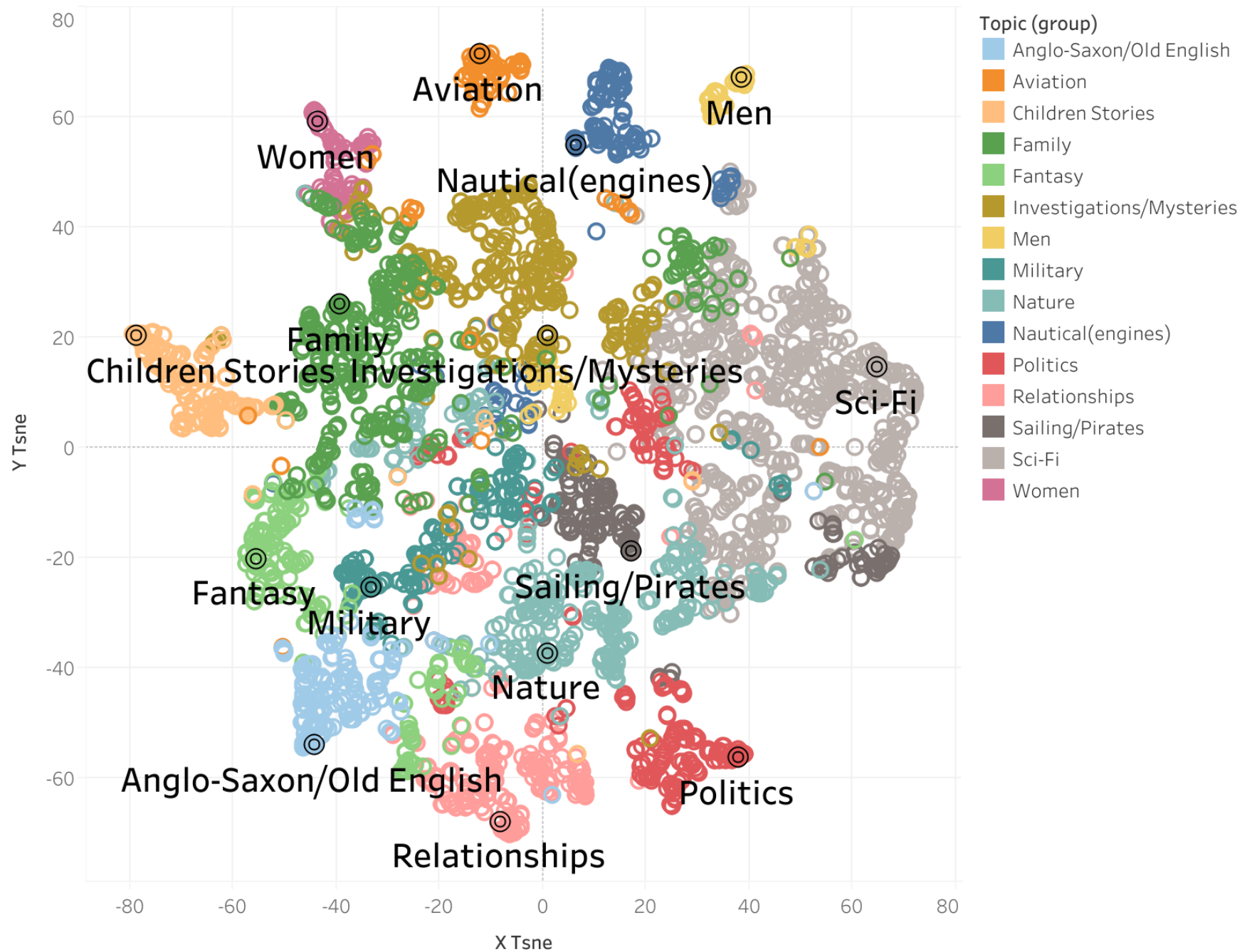
PCA clustering



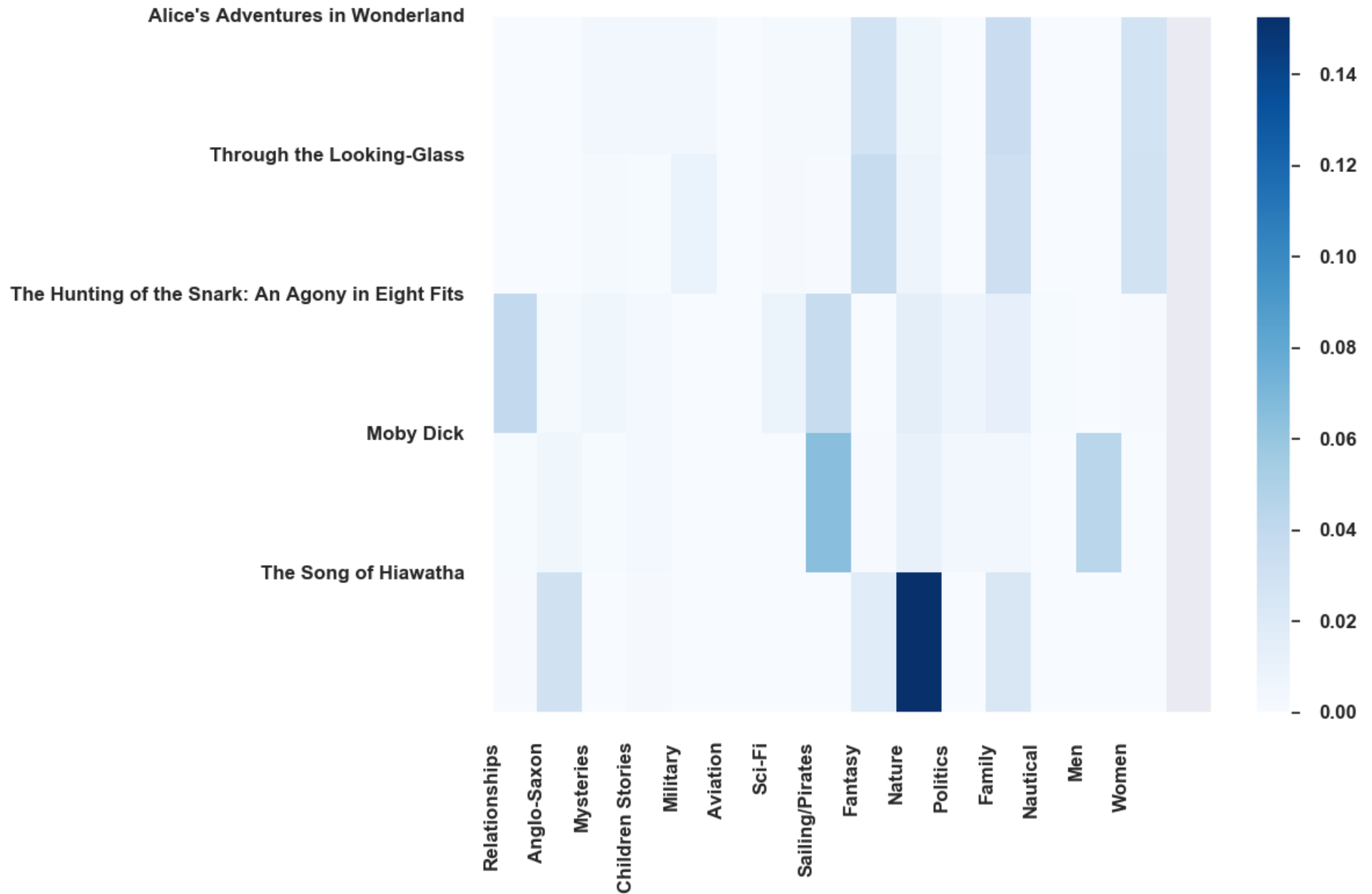
TSNE clustering



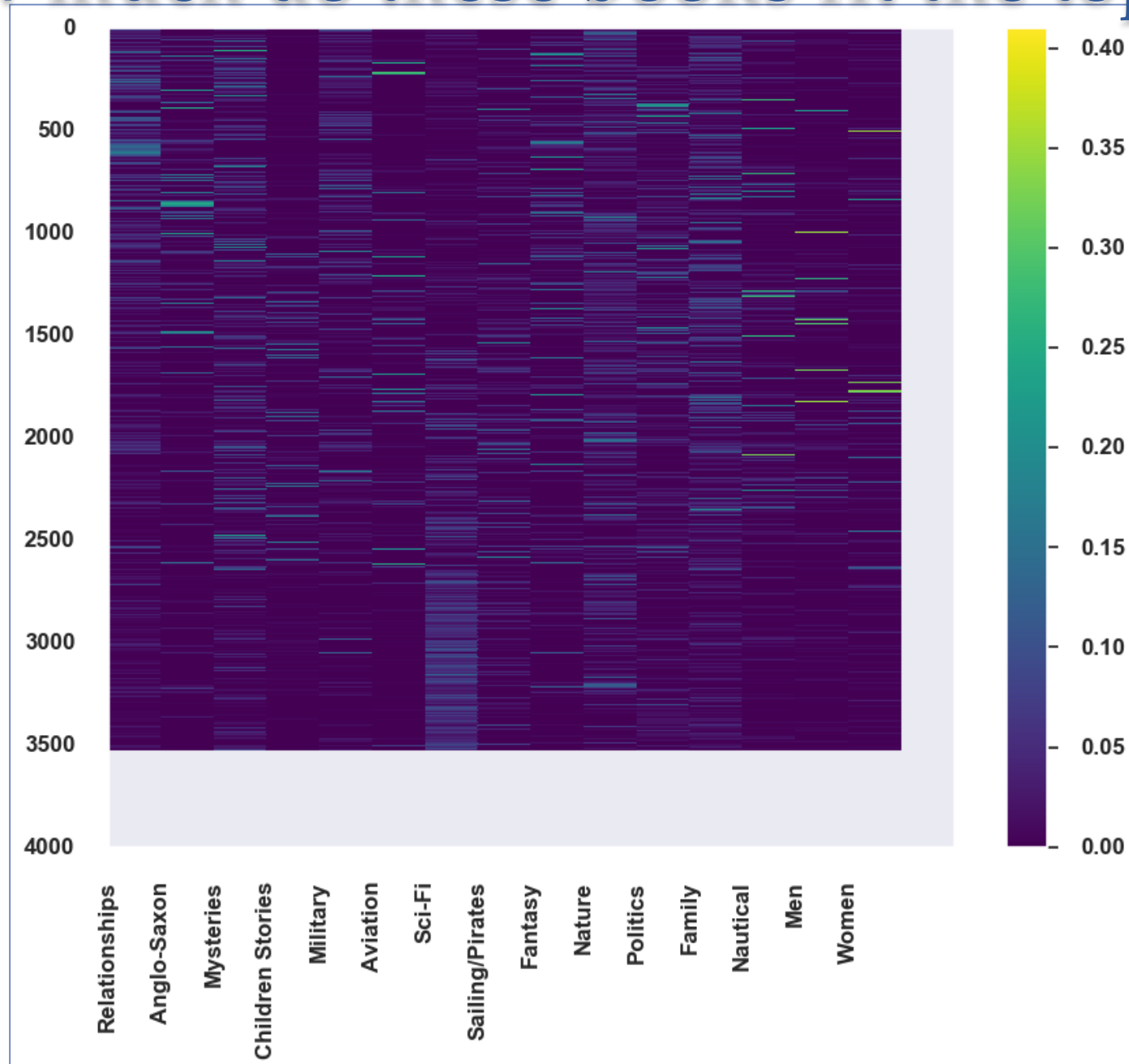
TSNE clustering



How much do these books fit the topics?

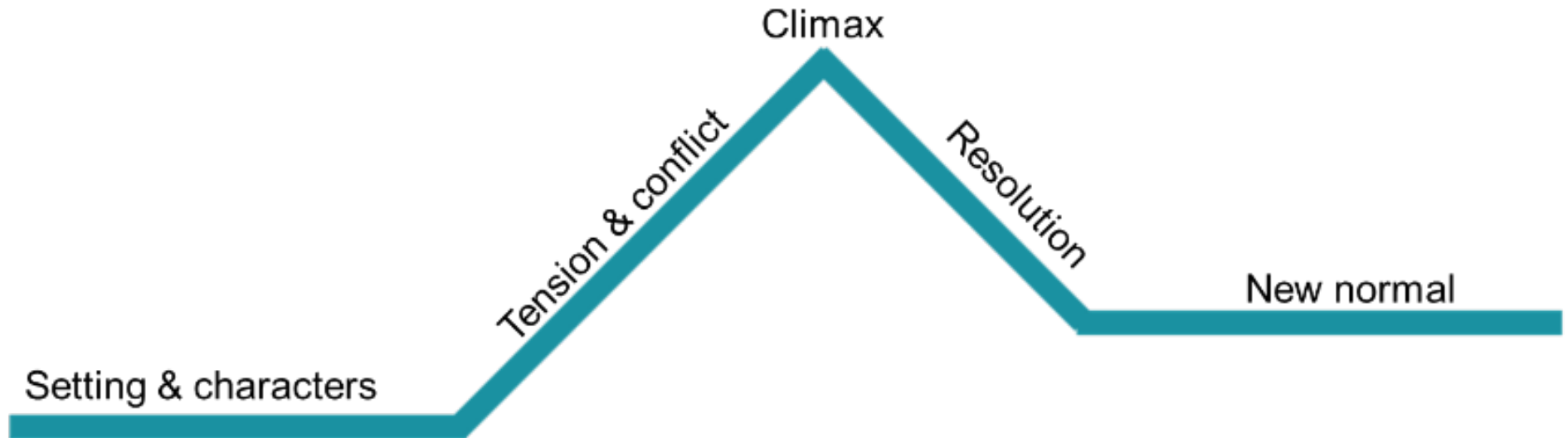


How much do these books fit the topics?



Sentiment → Plot

Classic Narrative Arc

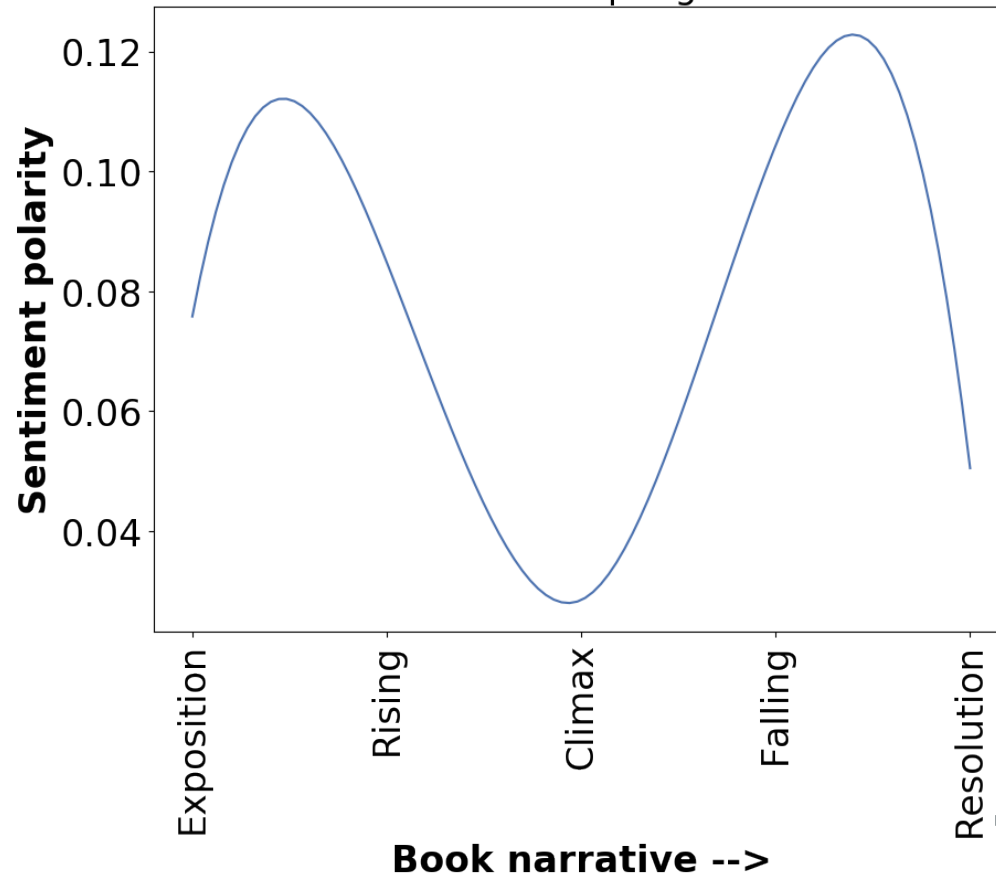


**“THERE’S NO
REASON WHY
THE SIMPLE
SHAPES OF
STORIES CAN’T
BE FED INTO
COMPUTERS.”**

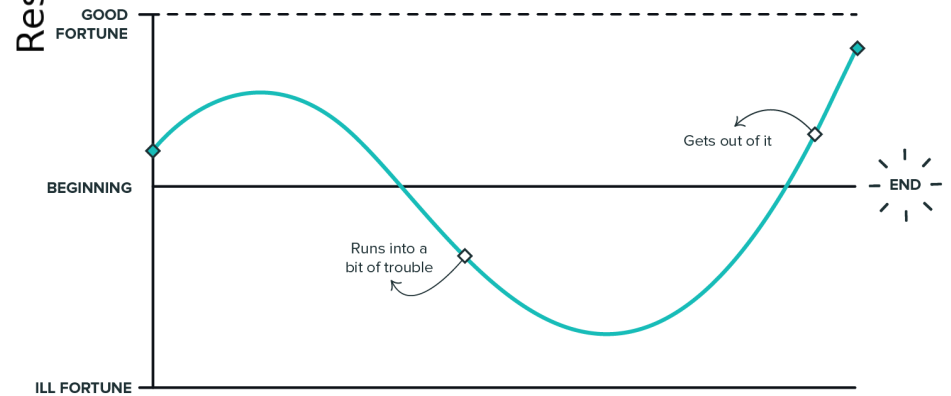
- KURT VONNEGUT

Plot shapes

Sentiment: His Last Bow: An Epilogue of Sherlock Holmes



MAN IN HOLE



Topic Shift → Complexity

His Last Bow: An Epilogue of Sherlock Holmes (score:2.59)

