Trends in Children's Book Reviews: Language and Sentiments

Ekin Ozince, Ruonan Sun, Vivian Sui

Client: Rebekah Fitzsimmons
Advisor: Zachary Branson

Introduction

- The *Bulletin of the Center for Children's Books* is a monthly journal that reviews children's literature, acting as a resource for educators and libraries.
- We analyze 32,235 reviews of books published between 1937 and 1992.
- We use visualizations and natural language processing methods to address the following research goals:
  - Create an **interactive dashboard** to summarize the data and extract insights
  - Analyze how the **discussion of key social and historical topics** has evolved over time
  - Analyze how **sentiment** in book reviews changes across time and book recommendation codes

Data

Our data was obtained from digitized scans of issues from the *Bulletin of the Center for Children's Books*. Metadata and our primary variable of interest, Review, were extracted using optical character recognition, as follows:

Review Example

Author	Title	Illustrator

Manning-Sanders, Ruth. A Book of Dwarfs; drawings by Robin Jacques. Dutton, 1964. 128p. Trade ed. $3.50; Library ed. $3.39 net.

Year	Pages	Price	Review

Code R / 4-6 Grade

A very nice collection of stories, most of them from the north of Europe. The seventeen tales are told with complete fidelity to the genre, and the very attractive illustrations are—again—completely appropriate. A lovely book to have, to read aloud, or to use as a source for storytelling.

Metadata

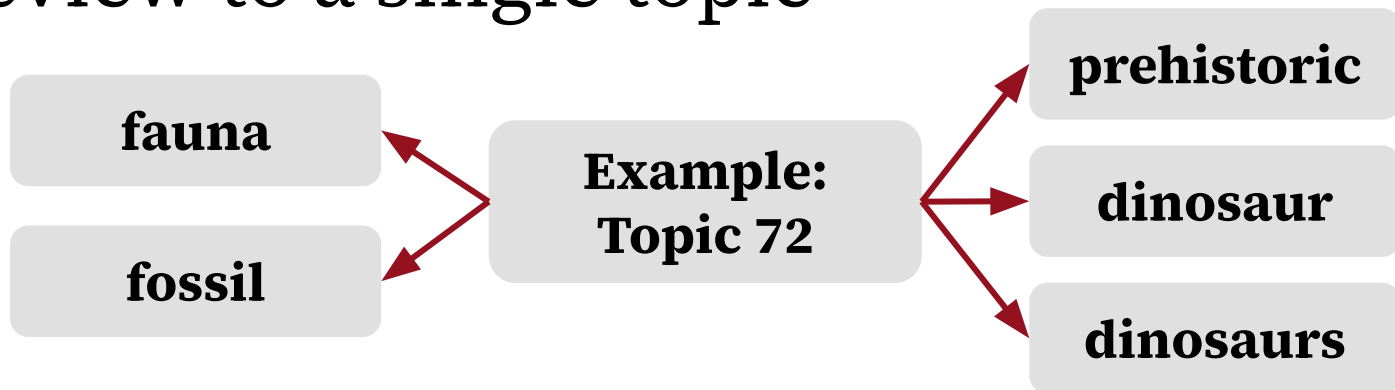|			
ID	Journal issue identifier	Reviewer	Reviewer initials
Year	Year the book was published	Code	Rating given to the book (e.g., R/R* = recommended, and NR = not recommended)
Title	Title of the book		
Author	Author of the book	Grade	Recommended age or grade range for the book
Publisher	Publisher of the book	Illustrator	Illustrator of the book
Price	Listed price(s) of the book	Pages	Number of pages in the book

Methodology

Data Preprocessing

- Cleaned text by removing punctuation, extra spaces, and standardizing case
- Clustered similar publisher names using string matching techniques, reducing unique publisher names from 1501 to 649
- Created a mapping logic to standardize grade ranges to consistent age categories

Topic Modeling

- Used Top2Vec to identify semantic topics from the book reviews, assigning each review to a single topic

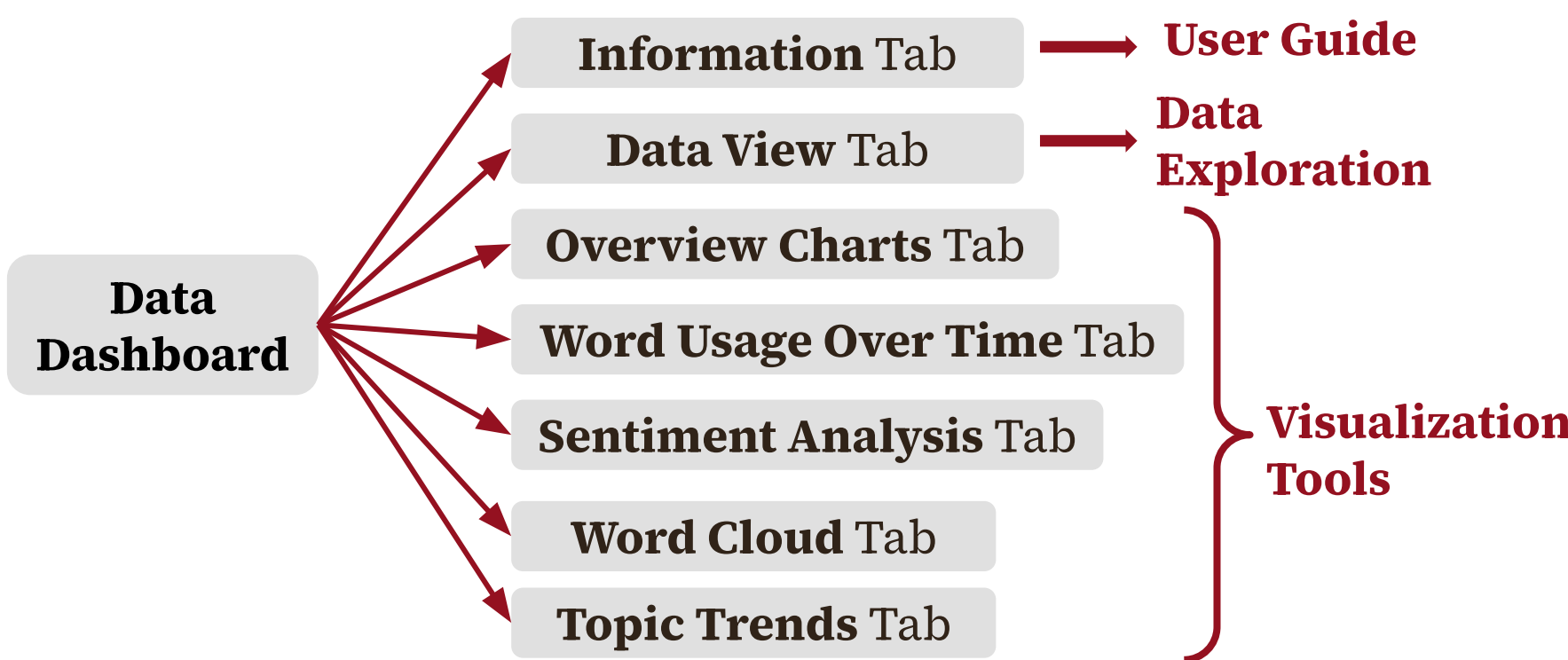fauna, fossil → Example: Topic 72 → prehistoric, dinosaur, dinosaurs

Sentiment Analysis

- Used an existing sentiment dictionary (Bing lexicon) to label words in the reviews as positive, negative, or neither
- Calculated sentiment scores per review:

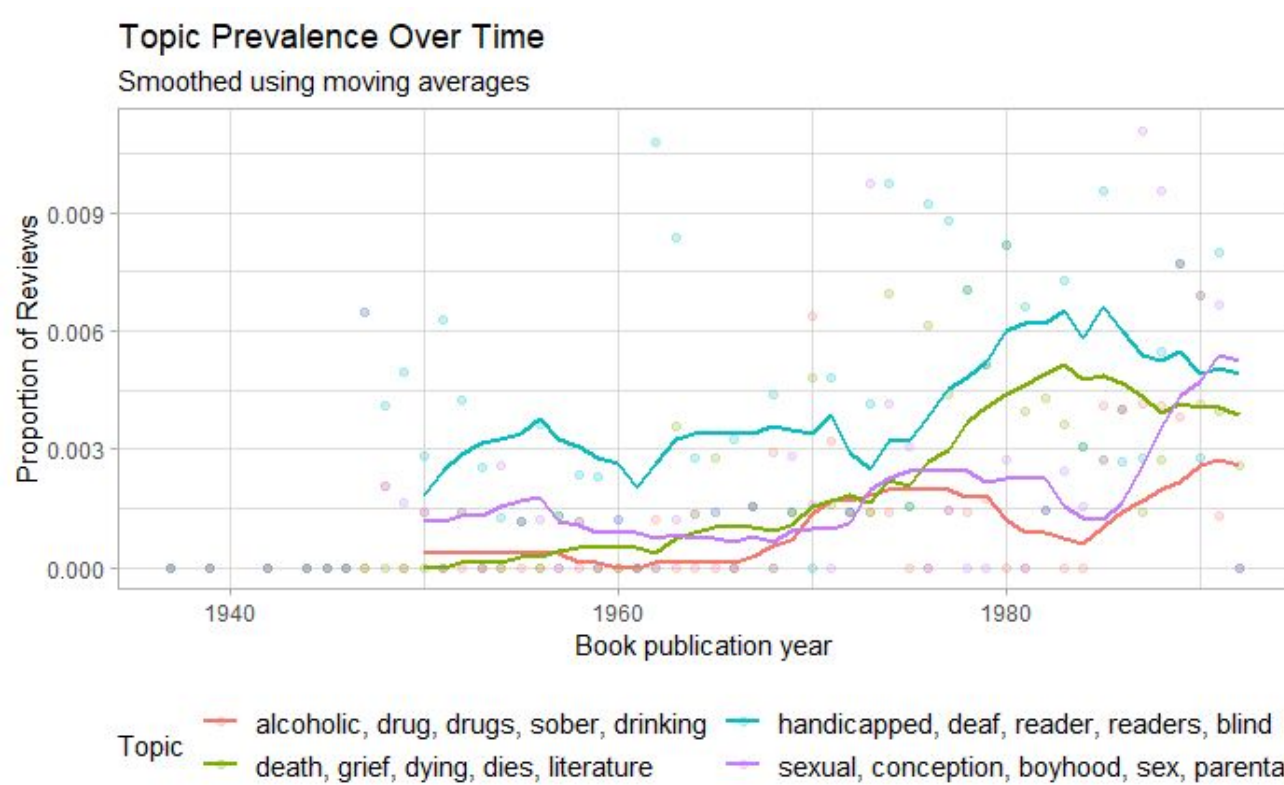$$\frac{\text{\# of positive or negative words}}{\text{total \# of words in review}}$$

Results

Data Dashboard

Our interactive data dashboard enables analysis of the dataset through various visualizations, helping users explore metadata, language use, sentiment trends, and topic shifts across time.
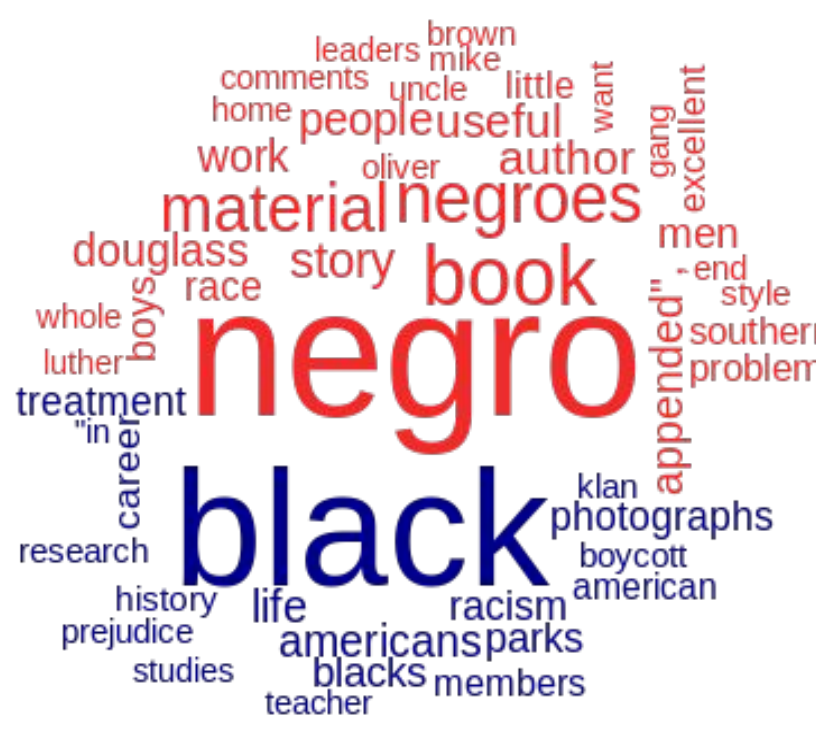
Data Dashboard →
- Information Tab → User Guide
- Data View Tab → Data Exploration
- Overview Charts Tab
- Word Usage Over Time Tab
- Sentiment Analysis Tab → Visualization Tools
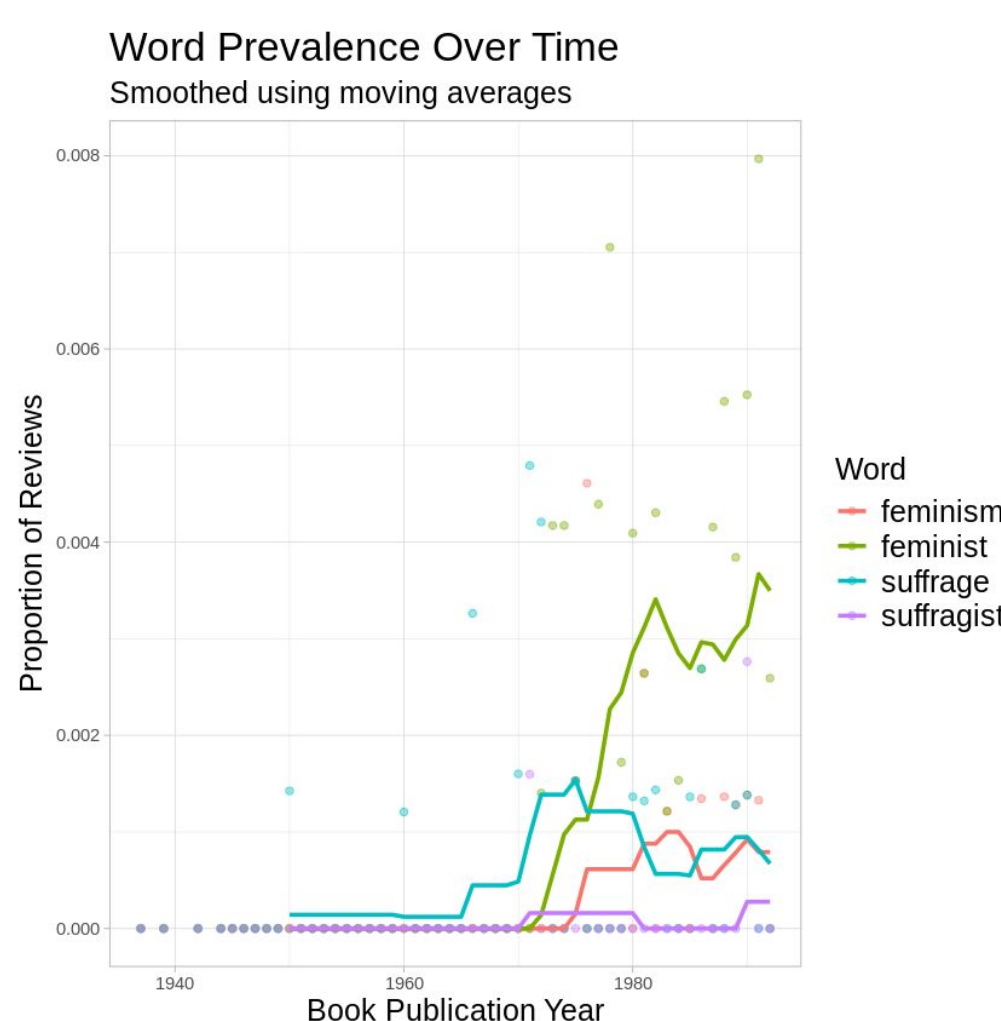- Word Cloud Tab
- Topic Trends Tab

Word Usage & Topic Prevalence

Topic modeling revealed themes of disability, substance use, death, and sexuality, all showing increased prevalence over time.

Topic Prevalence Over Time
Smoothed using moving averages

Topic: alcoholic, drug, drugs, sober, drinking; handicapped, deaf, reader, readers, blind; death, grief, dying, dies, literature; sexual, conception, boyhood, sex, parental

Language used in Black history-related reviews (according to the topic model) differed substantially between the periods before (red) and after (blue) 1970.
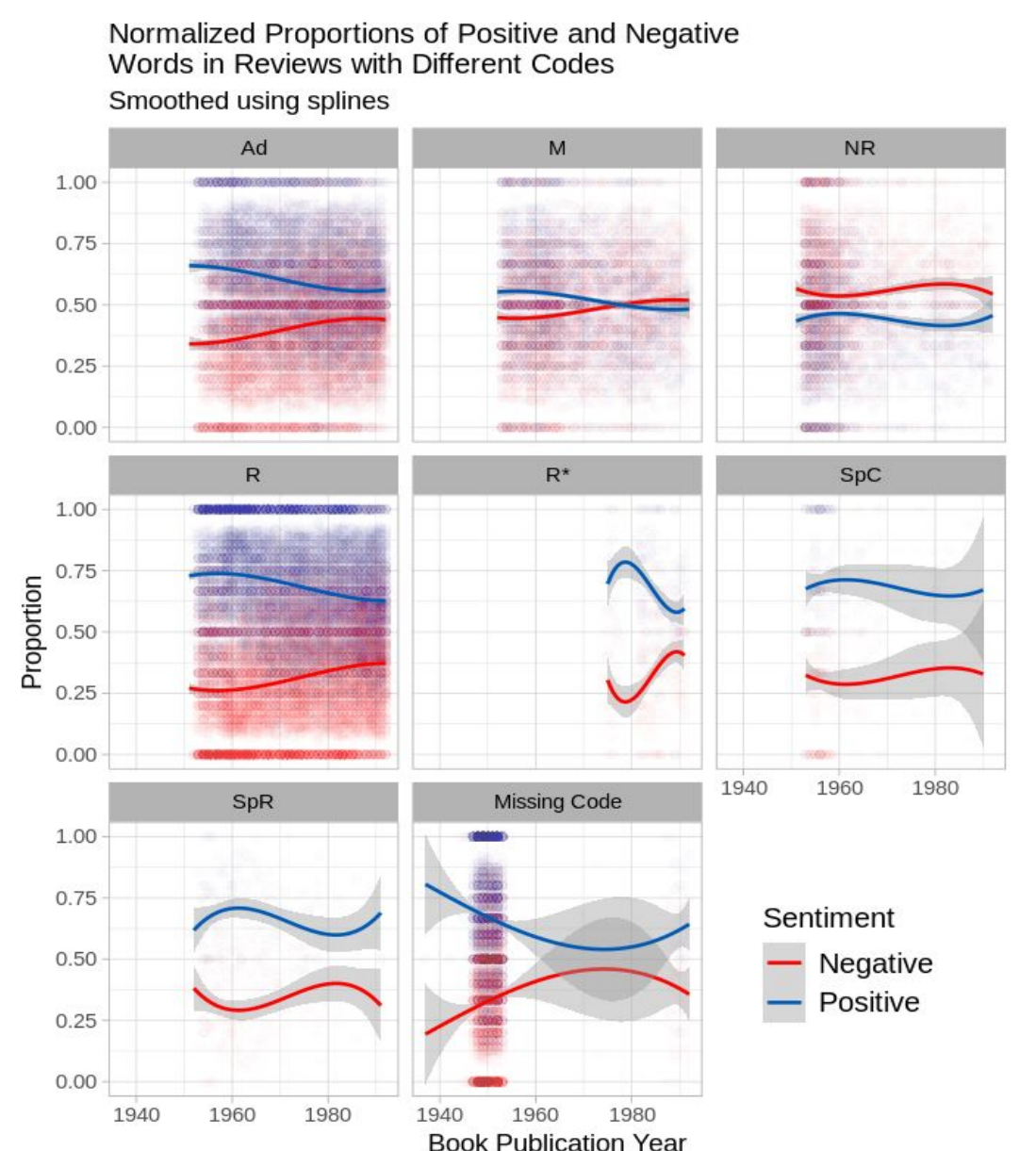

We tracked word usage for topics not captured by the model, such as feminism.

Word Prevalence Over Time
Smoothed using moving averages

Word: feminism, feminist, suffrage, suffragist

Sentiment Analysis

While some review codes (R, R*, NR) have clear meanings, others are more ambiguous. Plotting sentiment suggests that most of these codes lean more positive, but reviews rated M become more negative over time.

Normalized Proportions of Positive and Negative Words in Reviews with Different Codes
Smoothed using splines

Sentiment: Negative, Positive

Conclusion

- Modeling uncovered topics related to **substance abuse**, **disability**, **death**, and **reproduction**, with **increasing prominence over time**.
- **Topic and word trends** often align with **historical events** and **social changes**, such as the rise of feminism and evolving racial terminology.
- Review sentiment is **largely positive**, even for those with ambiguous codes, though **positivity levels shift over time**, especially for reviews with code M.

References

- Angelov, Dimo. Top2Vec: Distributed Representations of Topics. 2020. top2vec.readthedocs.io.
- The Bulletin of the Center for Children's Books. IDEALS Repository, University of Illinois. ideals.illinois.edu/units/119.
- The Bulletin of the Center for Children's Books. "About the Journal." University of Illinois. bccb.ischool.illinois.edu.
- The Bulletin of the Center for Children's Books. Vol. 19, no. 4, 1965. ideals.illinois.edu/items/12628.