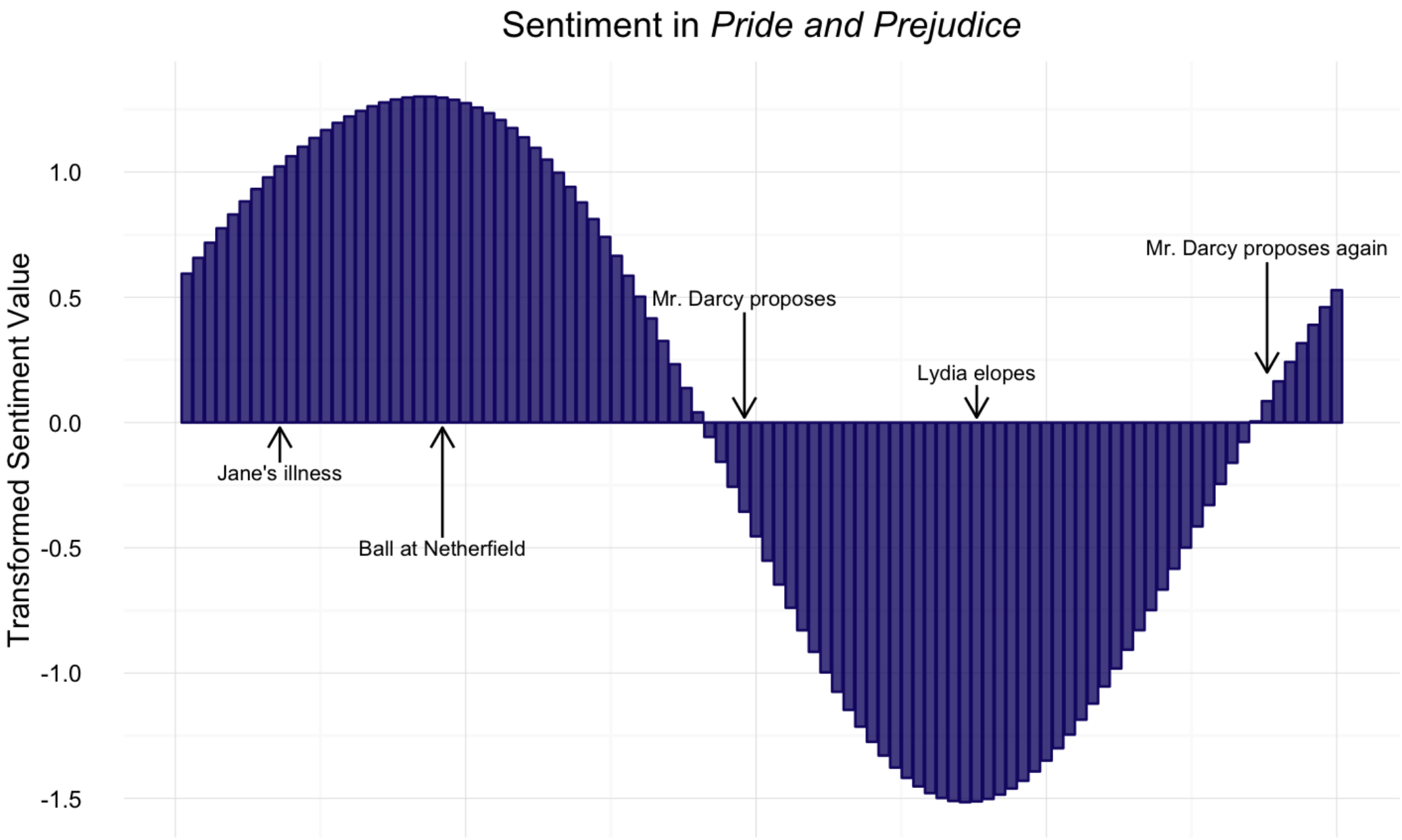
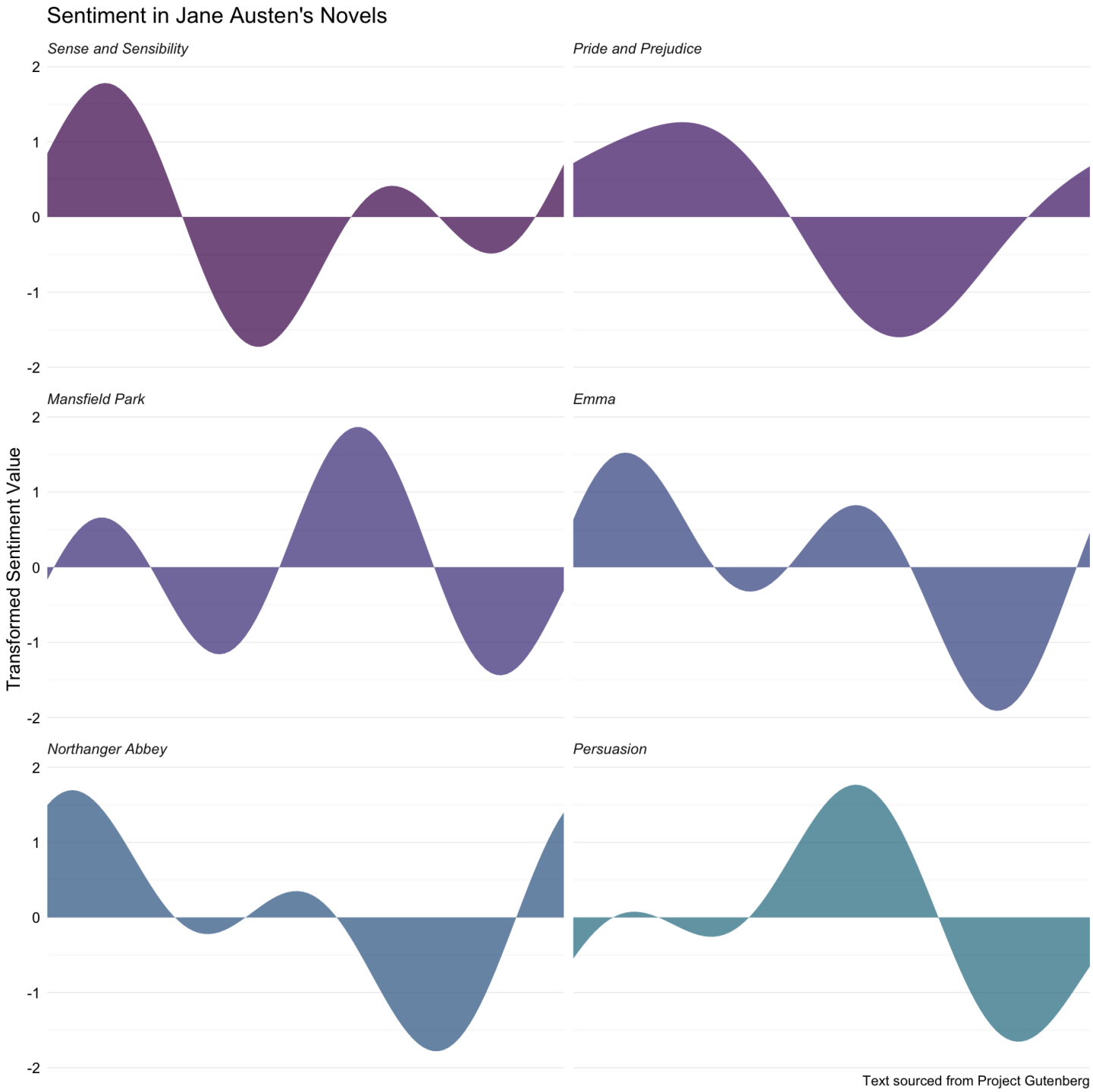
Data on words in books



https://juliasilge.com/blog/you-must-allow-me/



https://juliasilge.com/blog/if-i-loved-nlp-less/

Sentiment in Harry Potter

https://rbind.github.io/crossprod/post/2016-07-13-harry-potter-sentiment/

Sentiments in all books

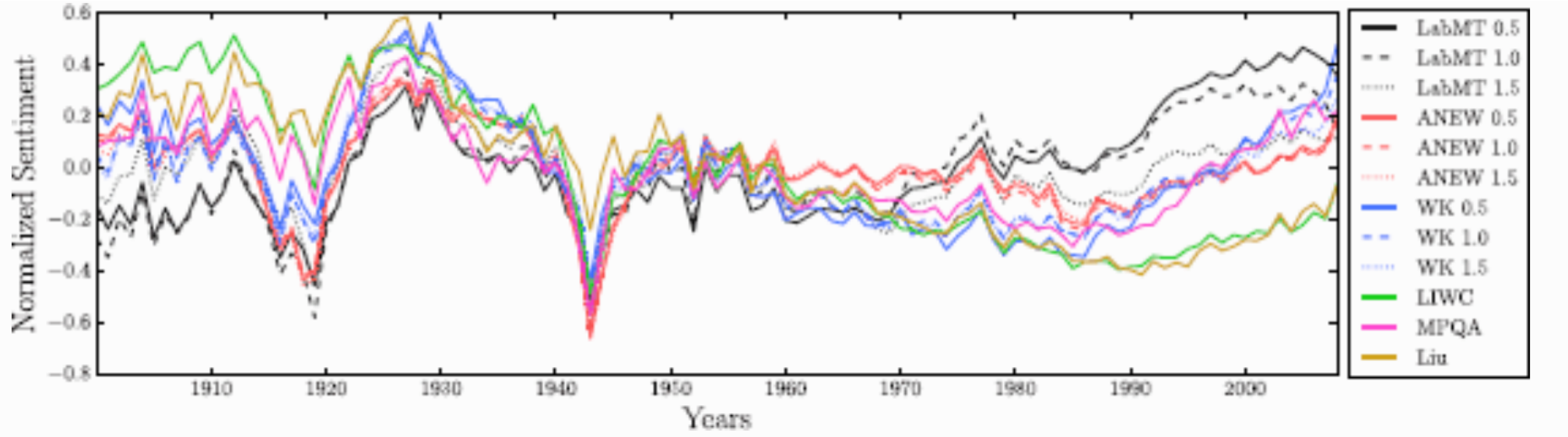


Figure 7

Google Books sentiment time series from each sentiment dictionary, with stop values of 0.5, 1.0, and 1.5 from the dictionaries with word scores in the 1-9 range. To normalize the sentiment score, we subtract the mean and divide by the absolute range. We observe that each time series has increased variance, with a few pronounced negative time periods, and trending positive towards the end of the corpus. The score of labMT varies substantially with different stop values, although remaining highly correlated, and finds absolute lows near the World Wars. The LIWC and OL dictionaries trend down towards 1990, dipping as low as the war periods.

Reagan et al. 2017. Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. EPJ Data Science 6:28

Are more kids books about boys?

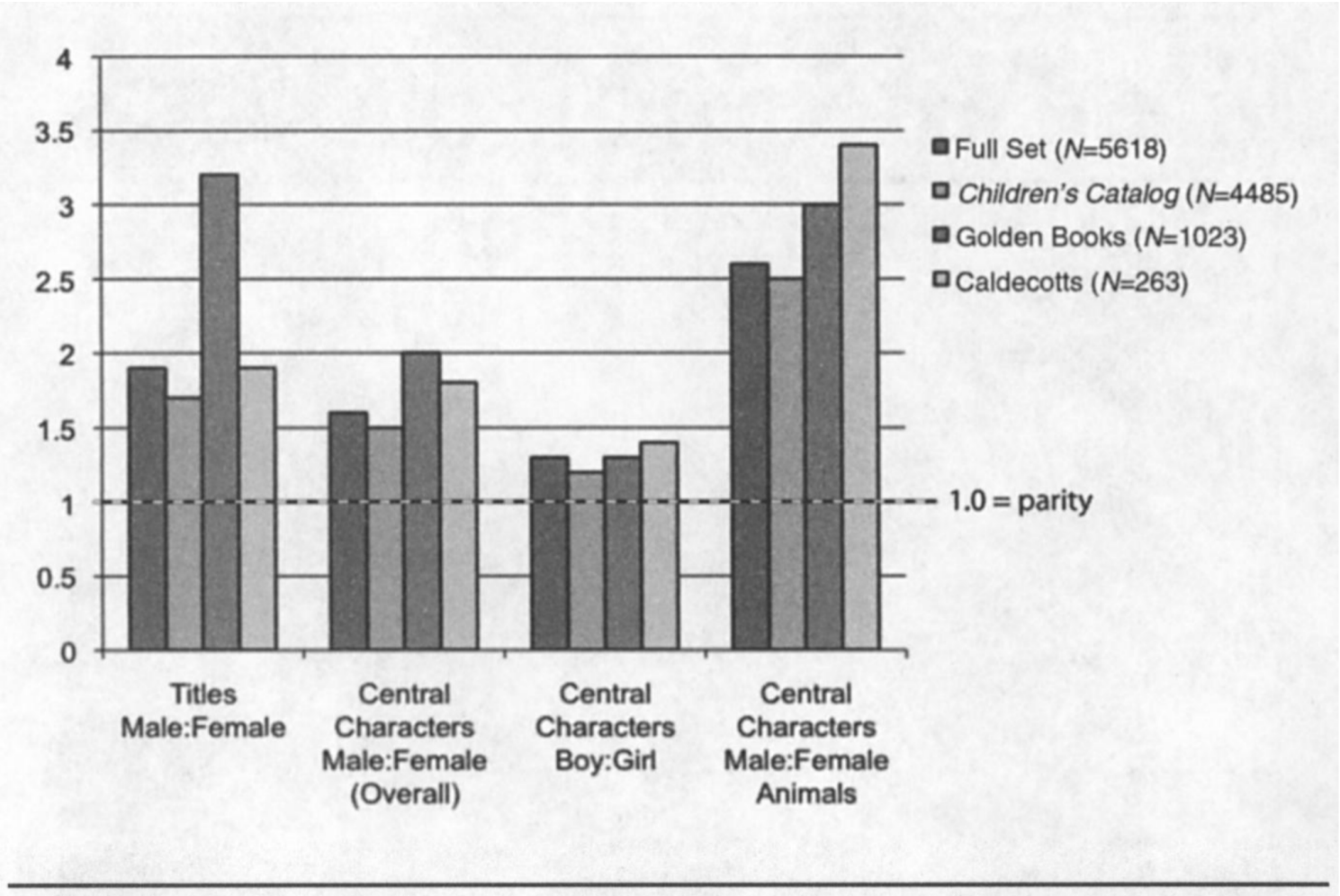


Figure 1: Ratios of Males to Females in Titles and Central Roles, 1900-2000: Full Set of Books (1900-2000), *Children's Catalog* (1900-2000), Little Golden Books (1942-1993), and Caldecotts (1938-2000)

McCabe et al. 2011. Gender in Twentieth-Century Children's Books. Gender and Society, 25: 197-226

Are more kids books about boys?

1. Tag 200 Goodreads books as having a “Male” or “Female” central character
2. Train a ML model using this sample
    1. Given a set of “features” which distinguish these two categories
3. Test the model by predicting the central character on additional books
4. Use the model to predict a class on the entire 3600 Goodreads books list

1. Fast, but maybe not the most accurate thing to do
2. Only for estimation and only for a blog

http://giorasimchoni.com/2017/09/24/2017-09-24-where-my-girls-at/

Features to distinguish books

1. she/her or he/him in book description
2. proportion of girls:boys name in description
3. gender of name in book title
4. gender of author

http://giorasimchoni.com/2017/09/24/2017-09-24-where-my-girls-at/

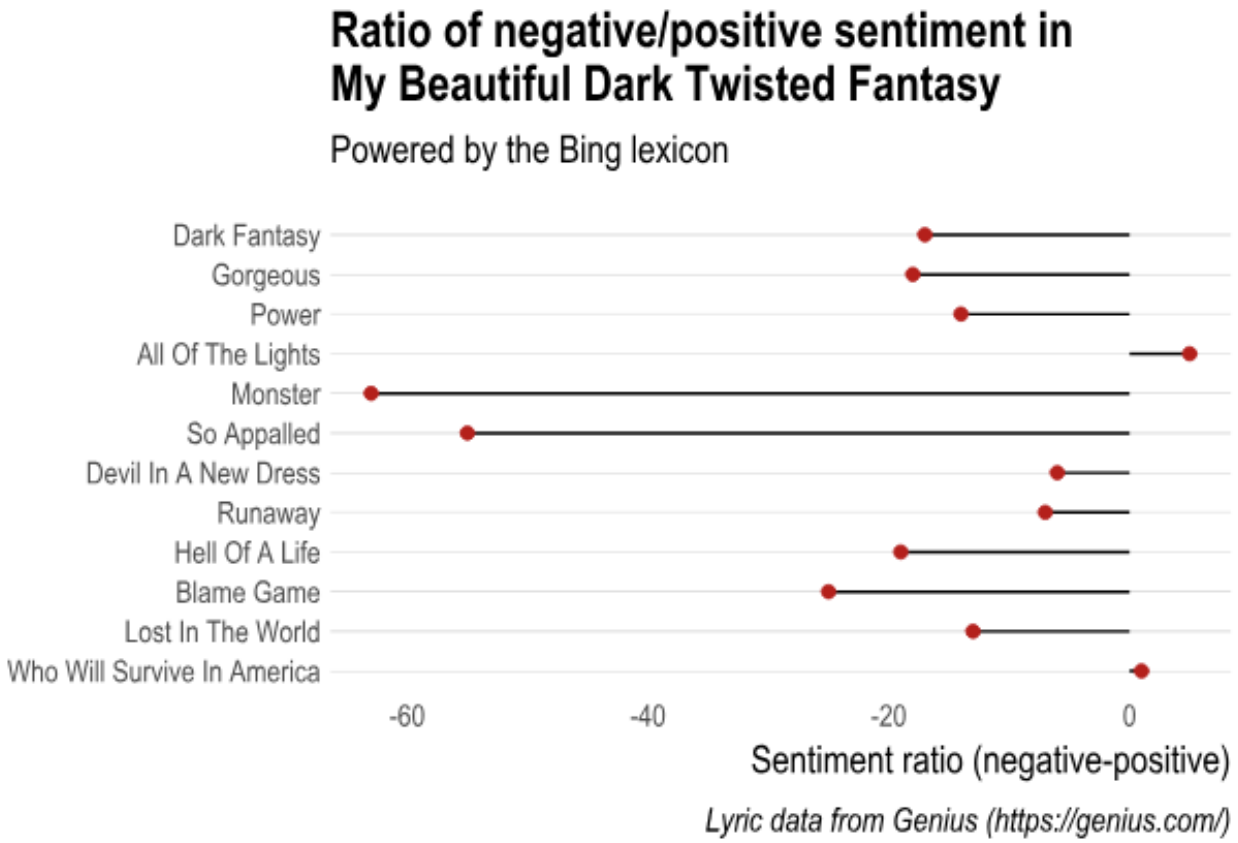
Are more kids books about boys?

# of books	Ratio
200	1.3:1
500	1.25:1
1000	1.25:1
2488	1.1:1

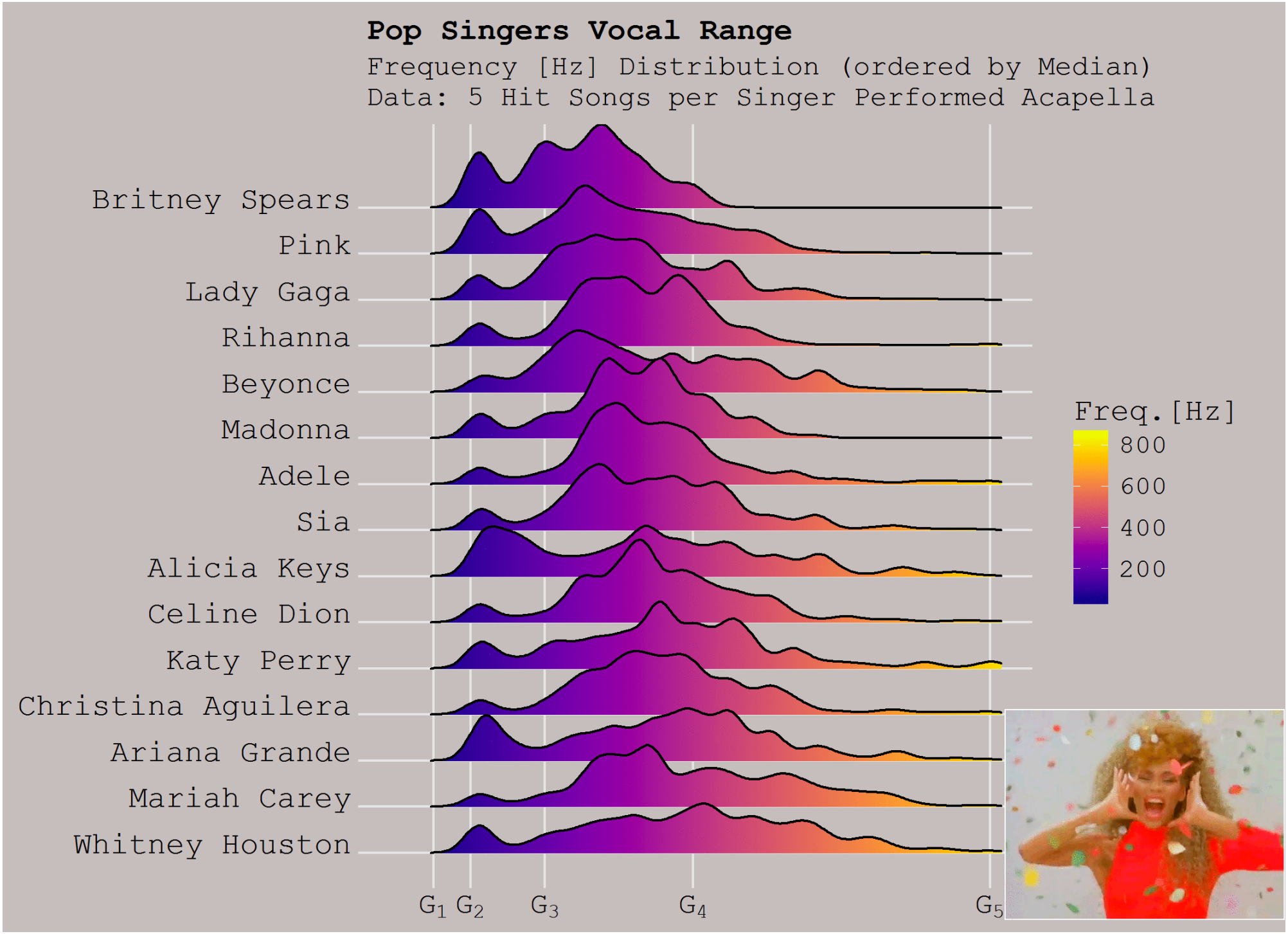
http://giorasimchoni.com/2017/09/24/2017-09-24-where-my-girls-at/

Words in music

A Sentiment Analysis of Kanye West Records



https://ewen.io/2017/03/20/a-sentiment-analysis-of-kanye-west-records/



http://giorasimchoni.com/2017/12/10/2017-12-10-ave-mariah/

““The data can find what a human would never have time to collect, but the human can make subjective and cultural judgments that the machines can’t.” ”

–Glenn McDonald, Data Alchemist, Spotify