COMPUTATIONAL ANALYSIS OF SPATIAL AND TEMPORAL VARIATION IN LITERARY SENTIMENT

OLGA REDKO '21, CHLOE GJOKA '23, AND PROF. JONATHAN GORDON, COMPUTER SCIENCE

INTRODUCTION

Using the data of tens of thousands of eBooks from the Project Gutenberg corpus, we analyzed differences in sentiments found in texts over various locations and years.

We hoped to help answer historical questions like *"What sentiments tend to increase during wars, revolutions, and epidemics?"*

Scan this code for an audio recording!

METHODOLOGY

1 Measured the distributions of sentiments across texts.

We examined sentiment from two perspectives:

- Using the NRC Emotional Intensity Lexicon (NRC-EIL), we measured the distribution of eight basic emotions as described by psychologist Robert Plutchik, including *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, and *trust*.

Table 1. A sample of NRC-EIL. Words for each emotion are assigned a value between 0 and 1. The closer a value is to 1, the more representative it is of a basic emotion.

Word	Fear	Word	Joy
horror	0.923	*sohappy*	0.868
horrified	0.922	*superb*	0.864
hellish	0.828	*cheered*	0.773
grenade	0.828	*positivity*	0.773
tragedies	0.750	*bestfeeling*	0.712
anguish	0.703	*complement*	0.647
grisly	0.703	*affection*	0.647
cutthroat	0.664	*exalted*	0.591
pandemic	0.664	*woot*	0.588

- Using the Valence Aware Dictionary and sEntiment Reasoner (VADER), we measured the polarity between positive and negative. sentiments.

2 Extracted information about locations and time periods associated with texts from a metadata file.

A partial entry from the metadata file used. First publish or edition dates were used to associate texts with time periods.

```
"edition date": 1909,
"first publish date": 1909,
"subjects": [
  "Vassar College"
],
"title": "Earliest Years at Vassar: Personal
  Recollections"
```

3 Used Wikidata to obtain information about given locations on different levels, such as on the city level, county level, state level, and country level.

4 Plotly for Python used to create maps and graphs depicting sentiment trends throughout time and in different locations.

FUTURE WORK

- Counting locations for each book according to more lenient guidelines (locations mentioned in the book as well as author Wikidata)
- Looking into more comprehensive sentiment lexicons
- Studying more books from countries not in Europe or North America, and in other languages

DATA VISUALIZATION

Figure 1. A map visualizing joy in literature from different countries. Warmer colors signify greater values, and cooler colors signify lower values.
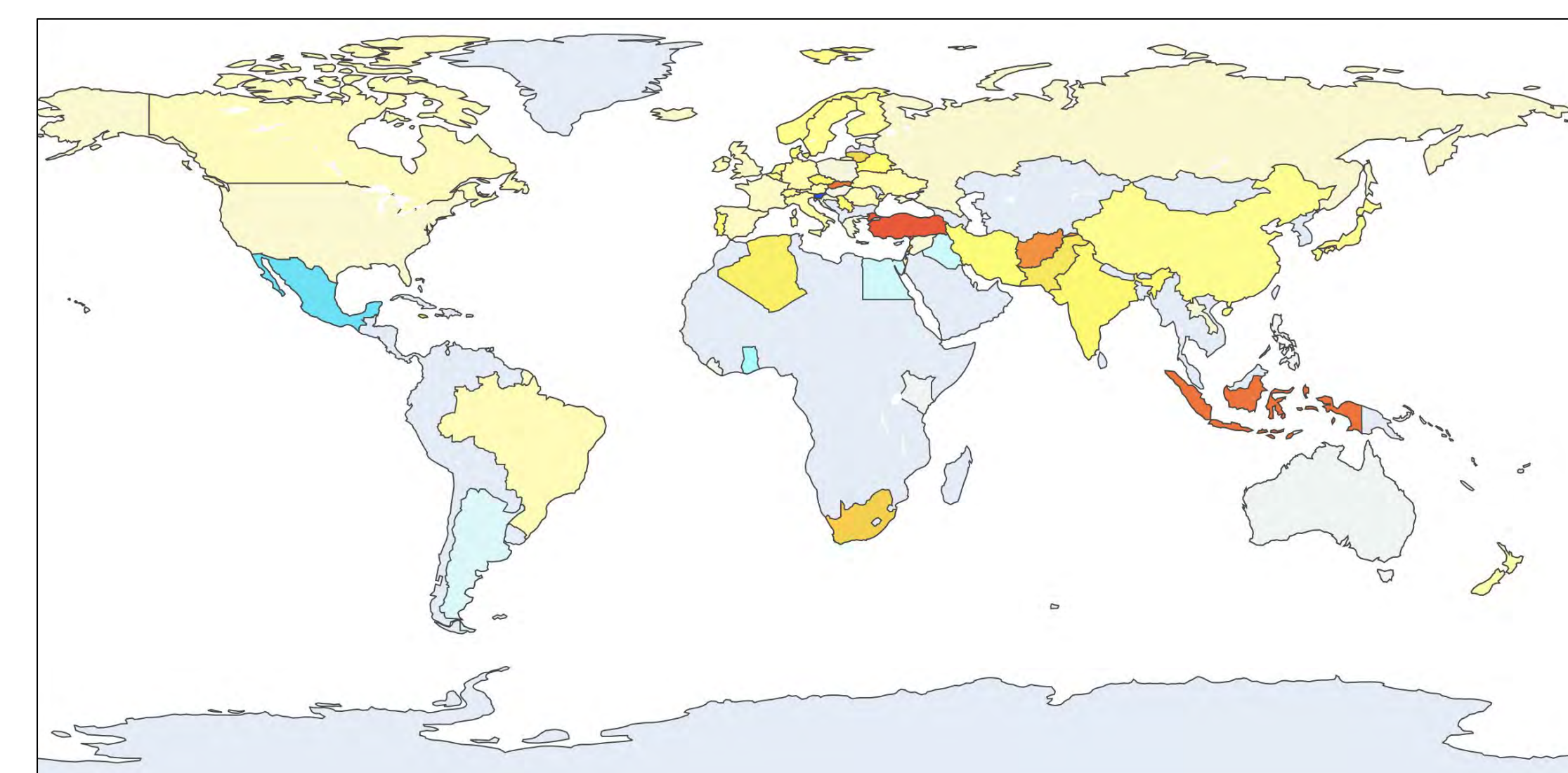

Figure 2. A map visualizing sentiment from authors in different cities, specifically in the year 1852. Again, warmer colors signify greater values, and cooler colors signify lower values.
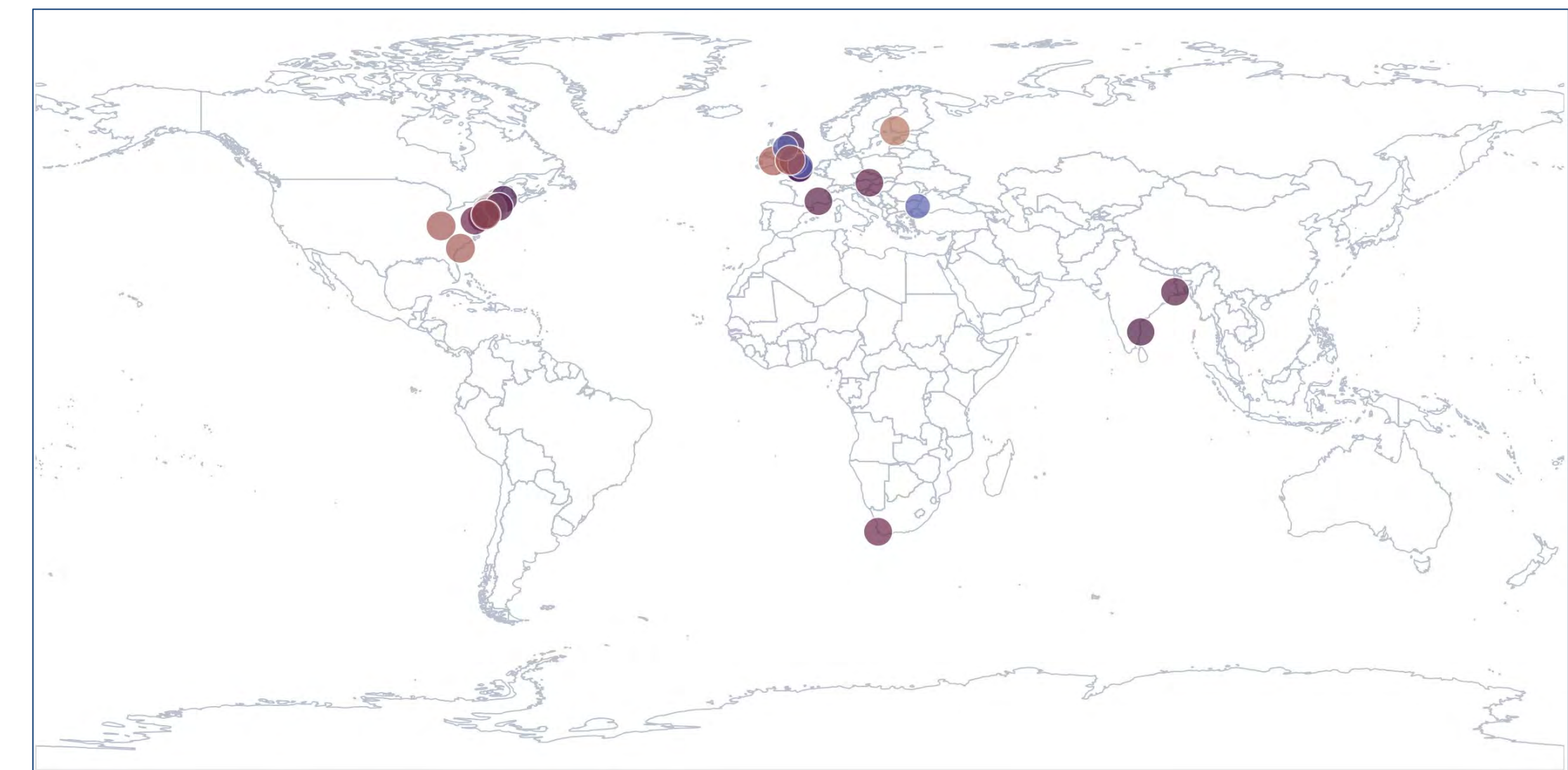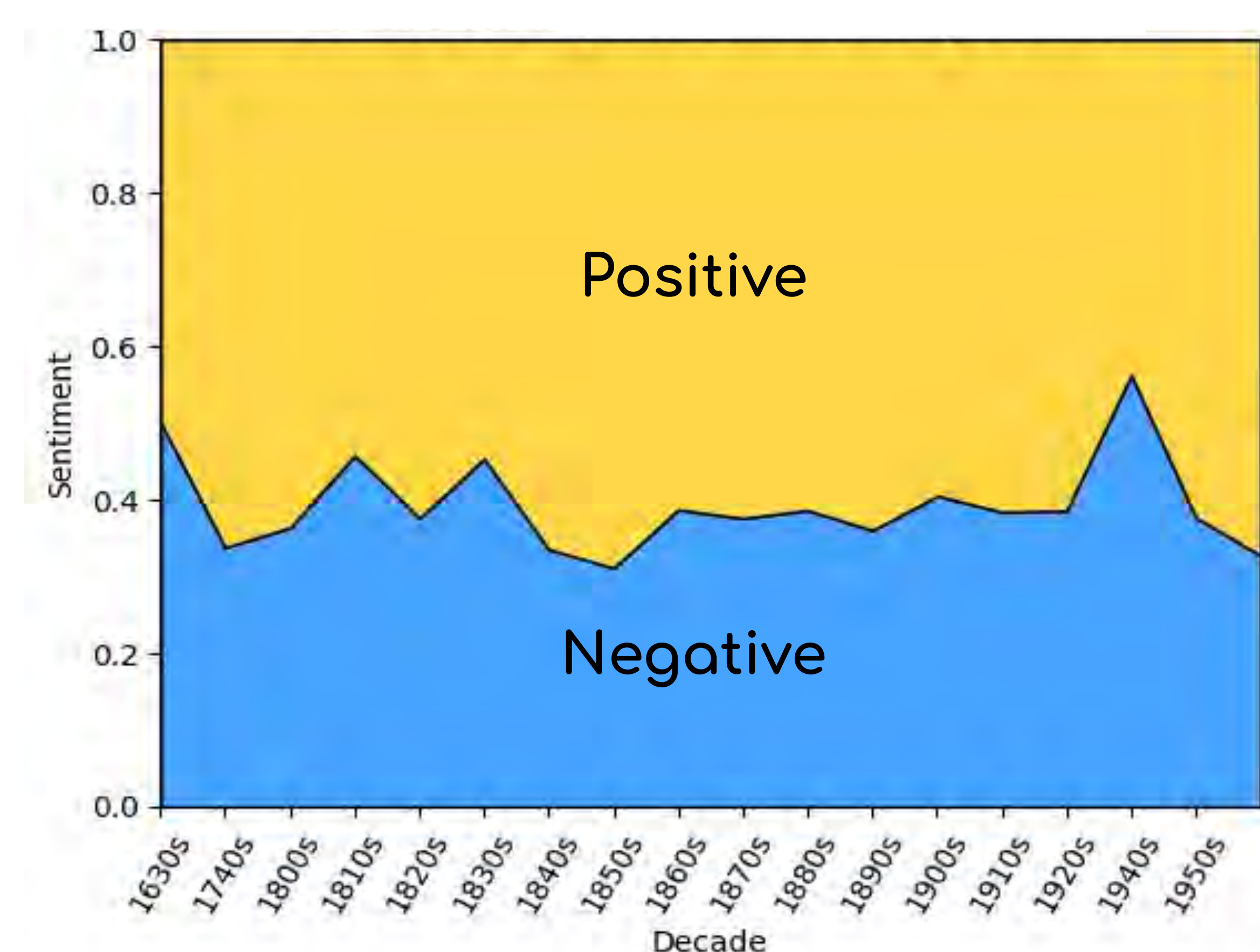

Figure 3. A stacked area plot showing positive and negative sentiment in literature in decades ranging from the 1630s to the 1950s.

REFERENCES

Eric Gilbert and C.J. Hutto. 2017. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Eighth International Conference on Weblogs and Social Media* (*ICWSM-14*), Ann Arbor, Michigan.

Project Gutenberg. 2020. https://gutenberg.org.

Saif M. Mohammad. 2017. Word affect intensities. *arXiv preprint arXiv:1704.08798*.

"Plotly Python Graphing Library." *Plotly Graphing Libraries*, Plotly, plotly.com/python.

ACKNOWLEDGMENTS