

**Digging Deeper,
Reaching Further**

Module 5: Visualizing Textual Data

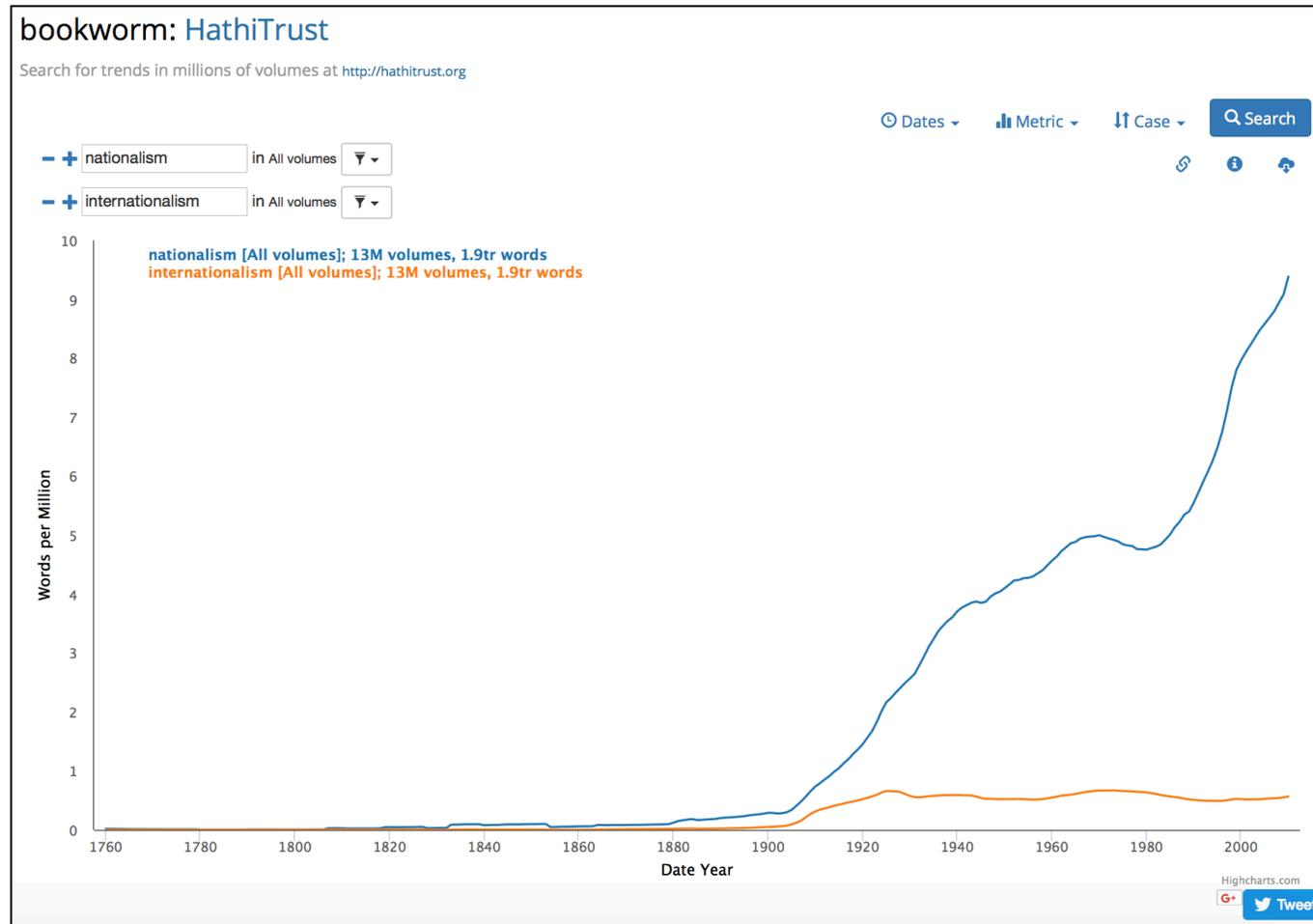
An Introduction

In this module we'll...

- Introduce common visualization strategies for text data
 - *Communicate with researchers about their options*
- Use a web-based visualization tool, HathiTrust+Bookworm
 - *Gain experience creating and reading data visualizations*
- See how Sam used HathiTrust+Bookworm for his project
 - *Learn how HT+BW was utilized in research*



Where we'll end up



Create
visualization of
word usage
trends across the
HathiTrust
corpus.



Data visualization

- Data visualization is the process of converting data sources into a visual representation
- Visualizations present particular ways of interpreting data
- Data visualization is an entire field of study; we're barely scratching the surface



Why visualize text data?

- Understand broader themes of a dataset
- Explore patterns in the data
- Cluster texts for overview or classification
- Compare data to other data (e.g., correlating with social networks)



Place in research process

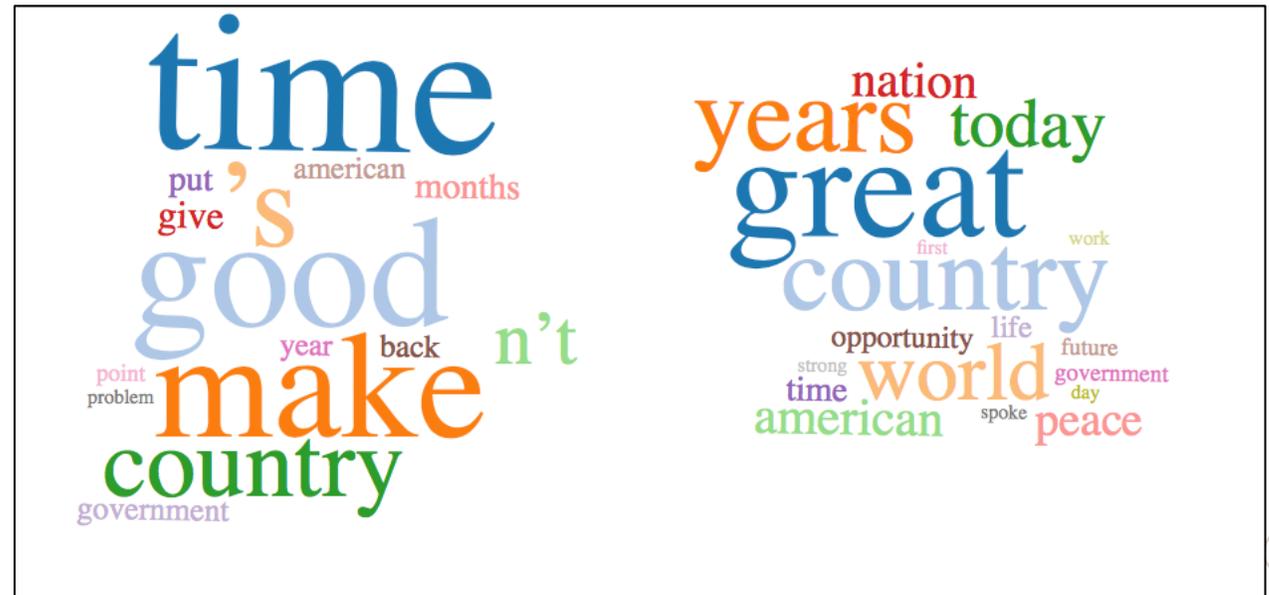
- In the earlier exploration stage of a project:
 - Explore full range of data
 - Discover characteristics and themes in data
- In the later explanation stage of a project:
 - Communicate findings to others in a clearer and more efficient way



Common text data visualizations

Word cloud

- Relatively unsophisticated, but effective
- Size of word relates to prominence or salience



Topic models from
HTRC Algorithms

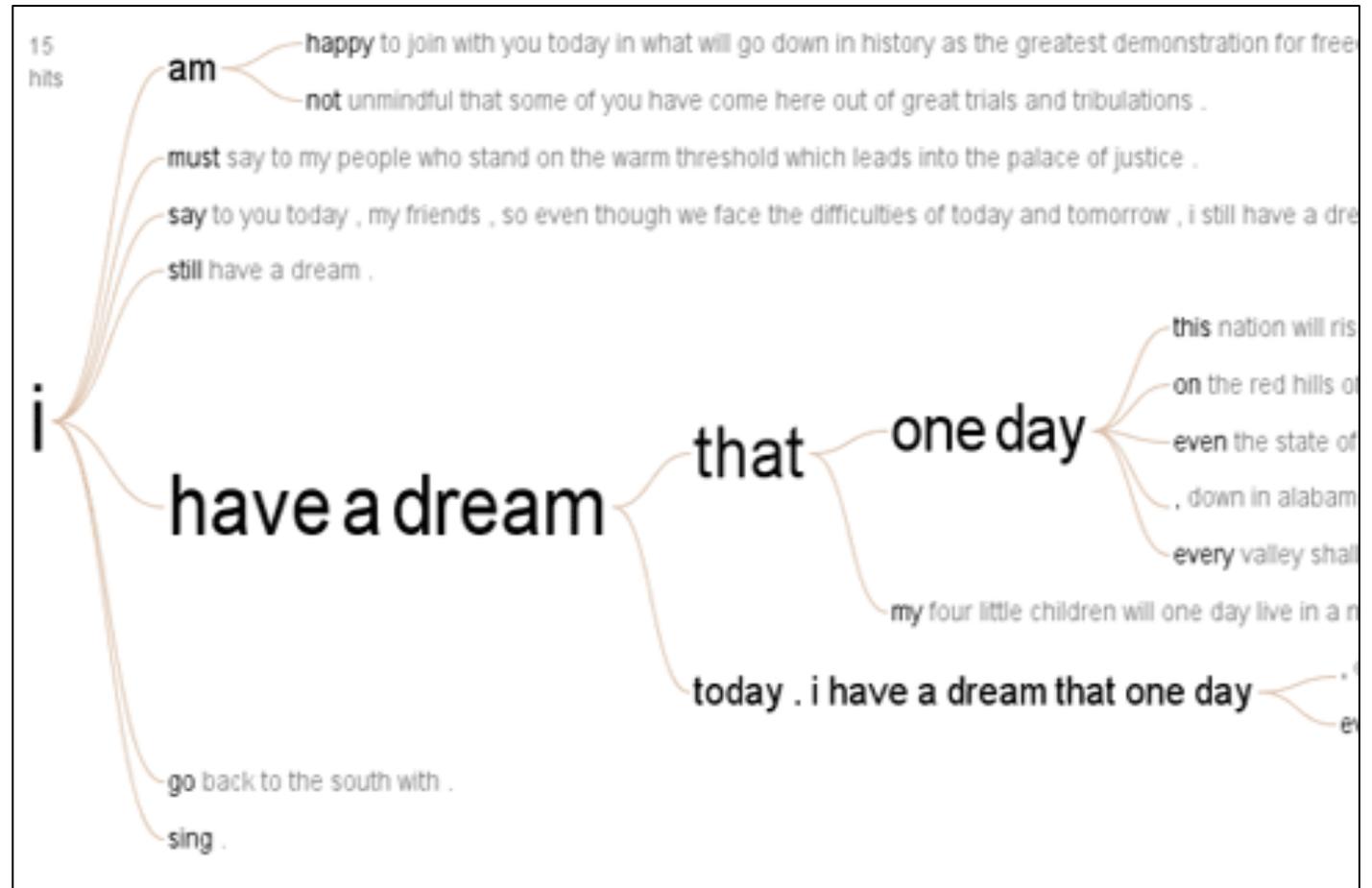


Common text data visualizations

Trees or hierarchies

■ Word trees

Occurrences of “I have a dream” in
Martin Luther King’s historical
speech.
(Wattenberg and Viégas, 2008)



Common text data visualizations

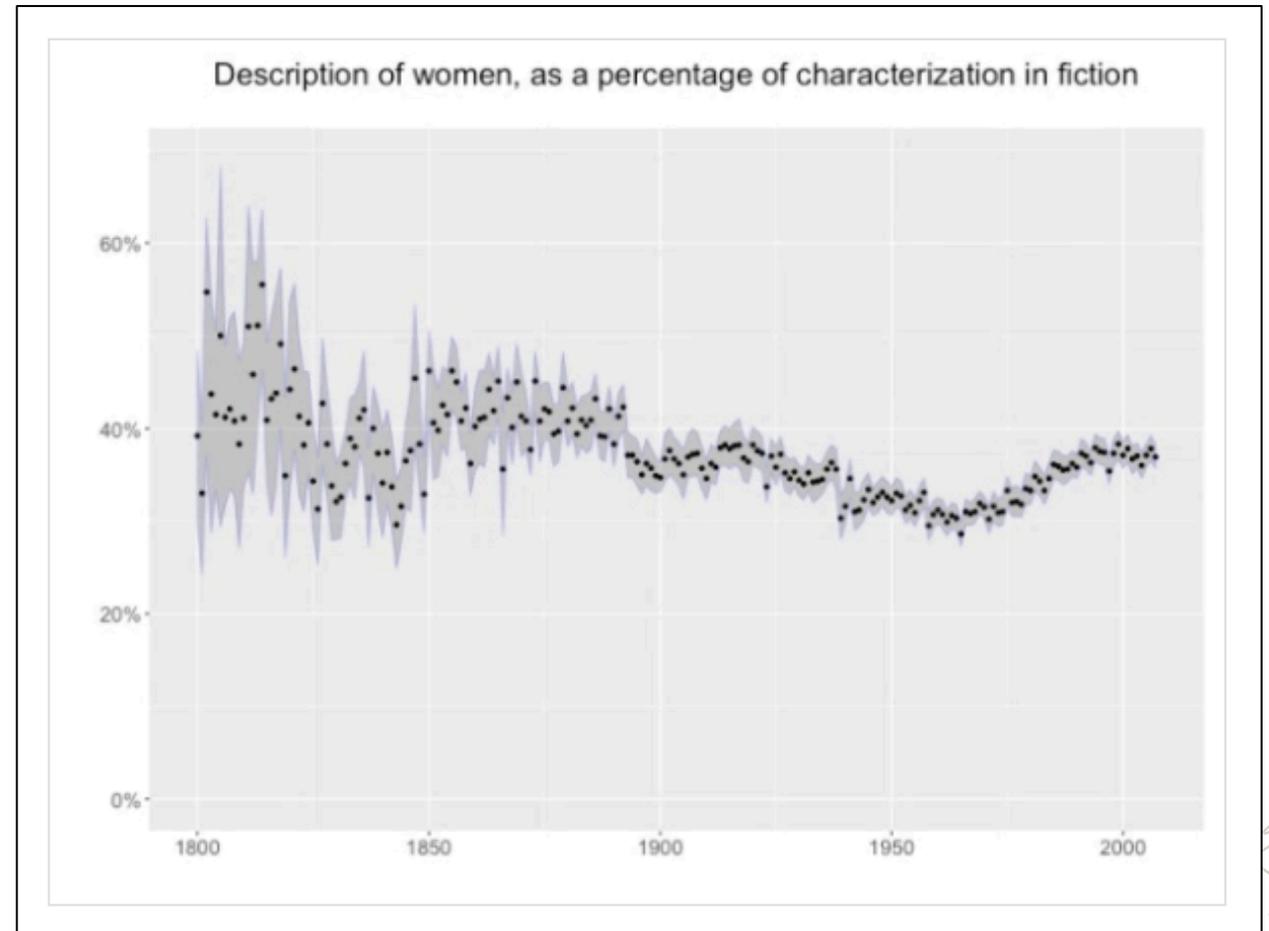
Temporal- or spatial-based visualizations

- Temporal visualizations

Percent representation of female characters
in English literature

(Underwood and Bamman, 2016)

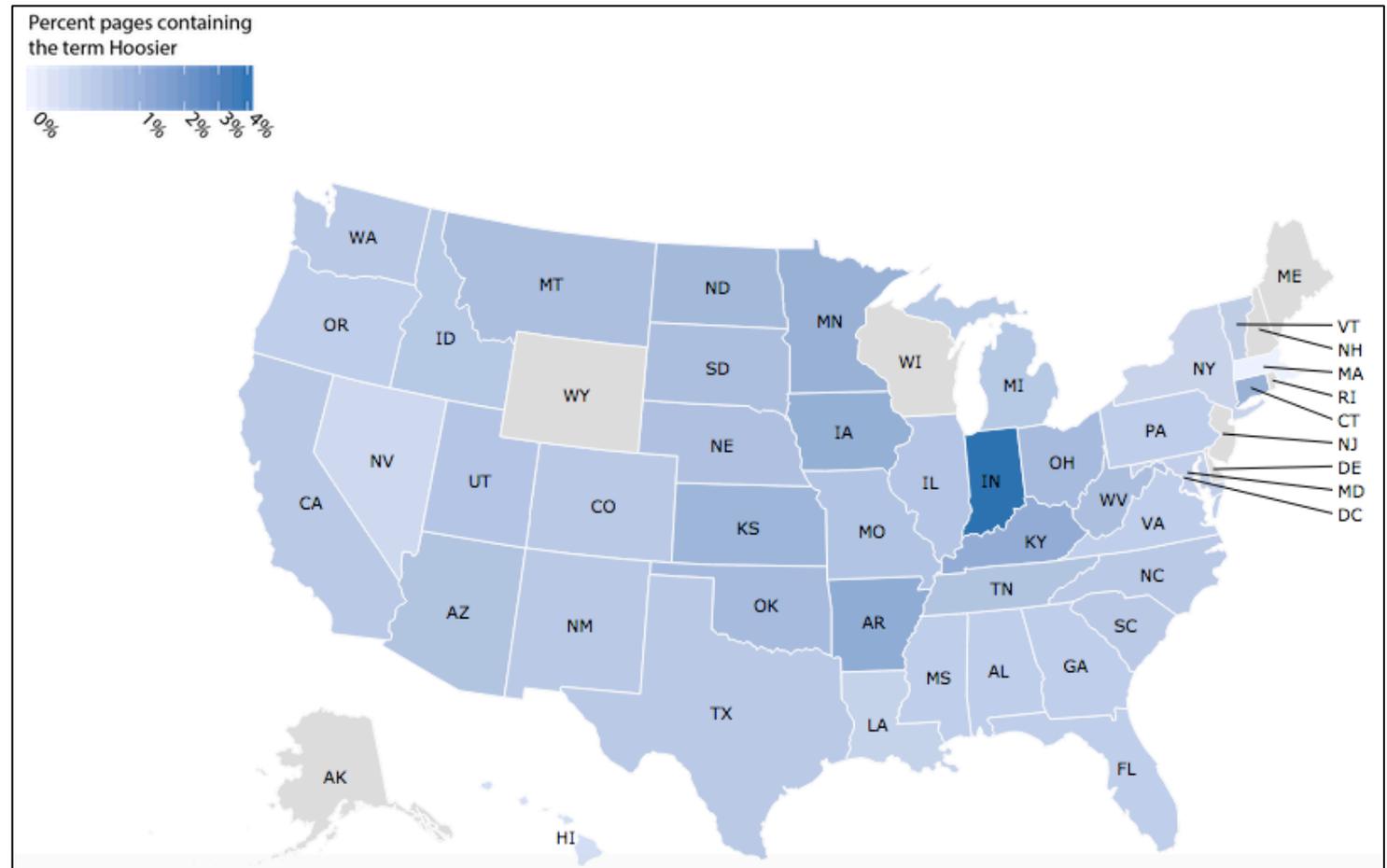
<https://tedunderwood.com/2016/12/28/the-gender-balance-of-fiction-1800-2007/>



Common text data visualizations

Temporal or spatial visualizations

- Maps



Percent of newspaper pages containing the term “hoosier” (Palmer, Polley, & Pollock, n.d.)

<http://centerfordigschol.github.io/chroniclinghoosier/map1.html>

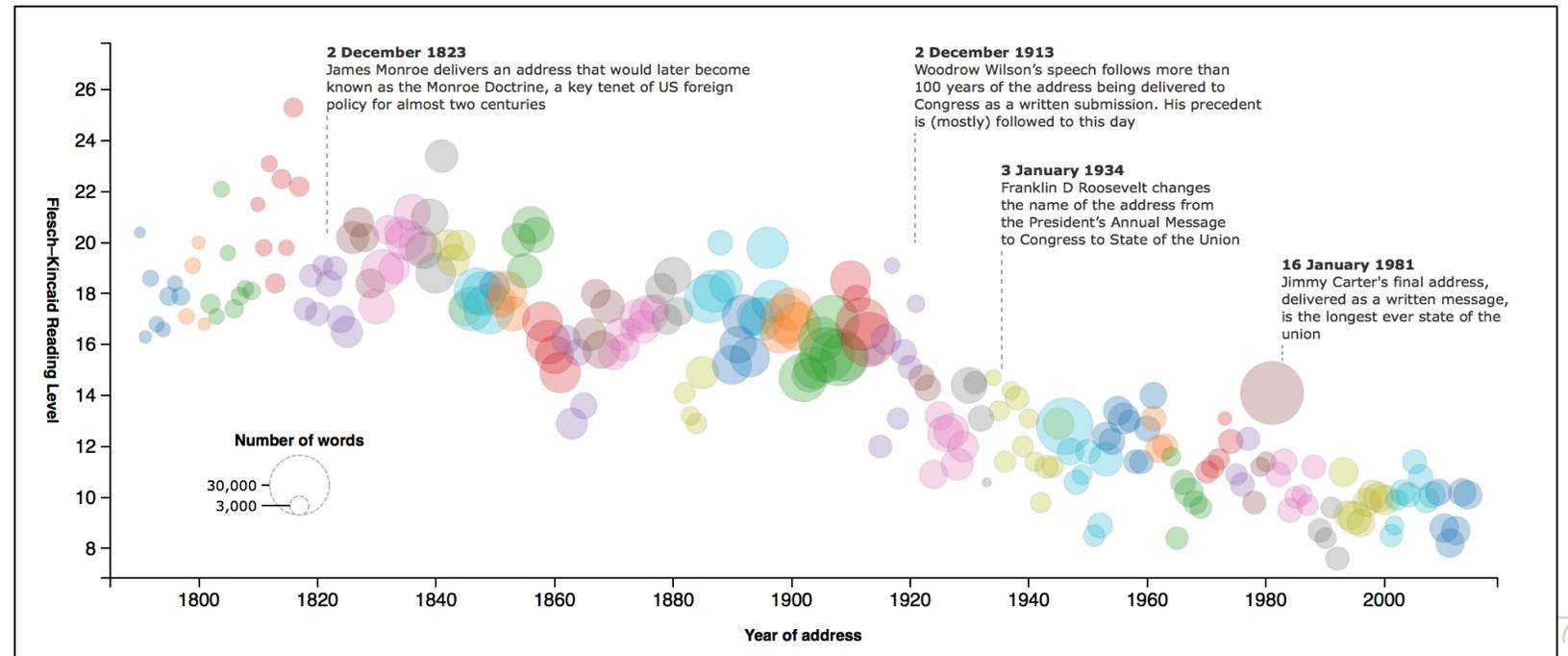


Common text data visualizations

Other “multi-dimensional” visualizations

- Bubble charts
- Heat maps

Bubble chart: readability of U.S. presidential speeches (The Guardian, 2013)



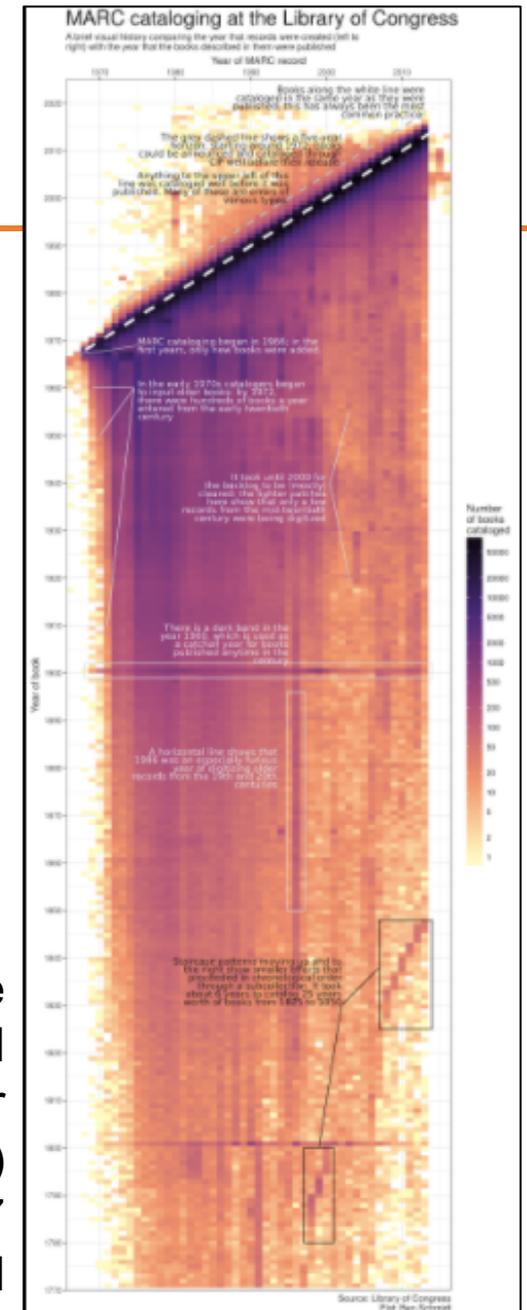
Common text data visualizations

Other “multi-dimensional” visualizations

- Heat maps

Heatmap of MARC cataloging at the Library of Congress by book year and cataloging year (Schmidt, 2017)

<http://sappingattention.blogspot.com/2017/05/a-brief-visual-history-of-marc.html>

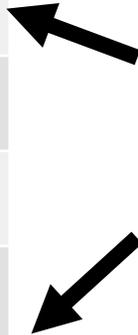


Activity

 See Handout p. 1

Match type of use to the type of visualization:

Visualization	What would it be good for?	Uses
Word cloud		Change over time
Trees or hierarchies		Spatial
Networks		Topical density
Timeline		Relationships
Map		Word distribution
Bubble chart		
Heatmap		



**** Bonus: what kinds of variables (i.e. data points) you would need for each visualization?**



Common visualization tools

- **Word clouds**

- Voyant
- Wordle

- **Word use trends**

- Google Books Ngram Viewer
- HathiTrust+Bookworm

- **Tabular data visualization**

- Tableau

- **Mapping**

- ArcGIS Online with StoryMaps
- Tableau

- **Network graphs**

- Gephi
- NodeXL
- DH Press



Common visualization libraries

- **Python**

- matplotlib, pyplot
- ggplot library

- **R**

- ggplot2

- **D3.js**

- Javascript library for visualizations



Review: key terms in text analysis

N-gram

A contiguous chain of n items from a sequence of text where n is the number of items. Example: Bigram.

four score, score and, and seven, seven years, years ago, ago our, our fathers, fathers brought, brought forth, forth on, on this, this continent, continent a, a new, new nation, nation conceived, conceived in, in liberty, liberty and...



N-gram visualization: HathiTrust + Bookworm

Brings together:

- Text data (unigrams)
 - Bibliographic metadata
- HathiTrust
- Visualization tool
 - Track trends in a repository
- Bookworm

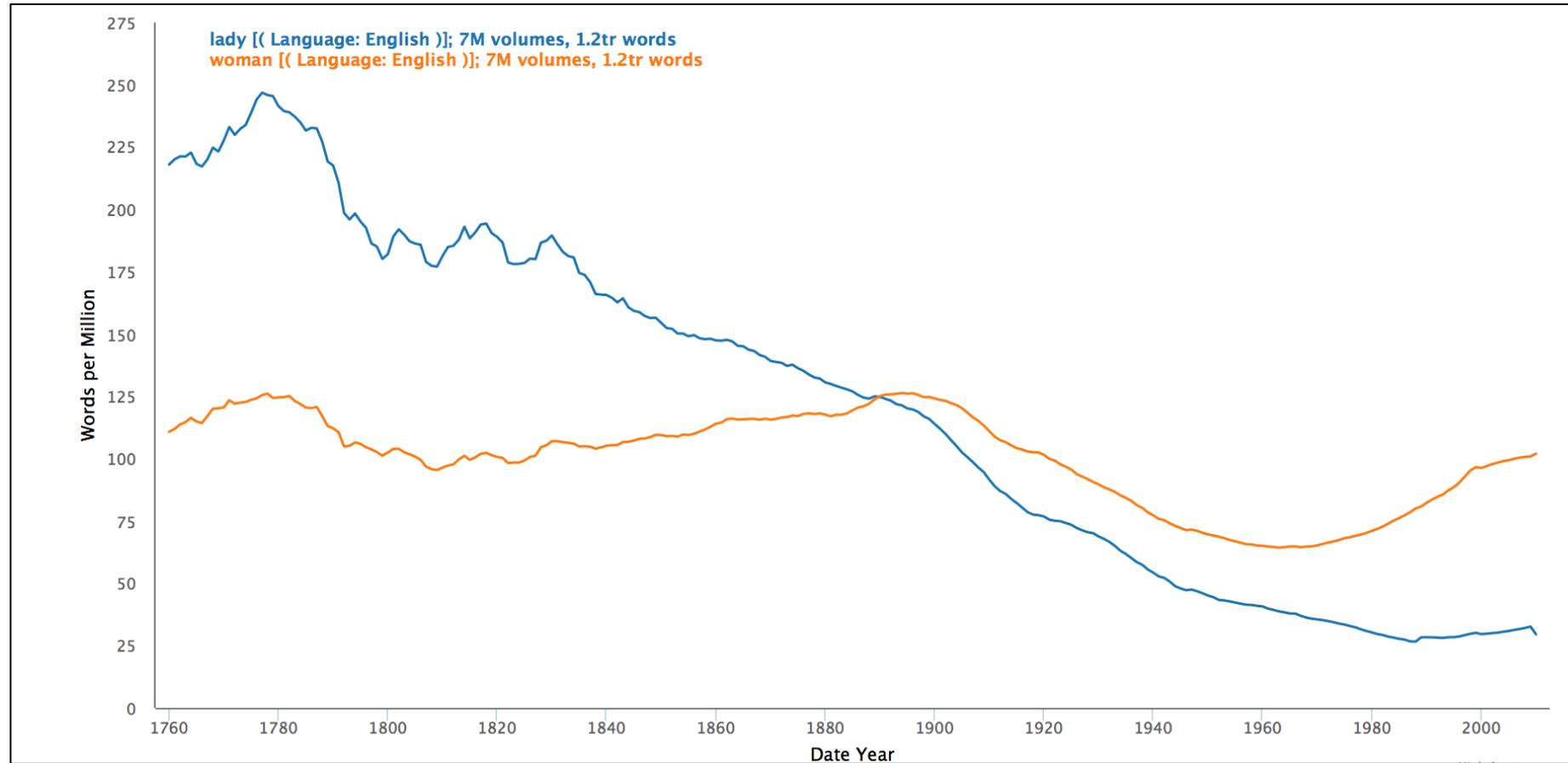


Bookworm framework

- Visualizes categories
- The category is plotted along the x-axis
 - Often plot years along the x-axis
 - Can plot other things!
- HathiTrust+Bookworm is just one implementation of the framework



Example HT+Bookworm view

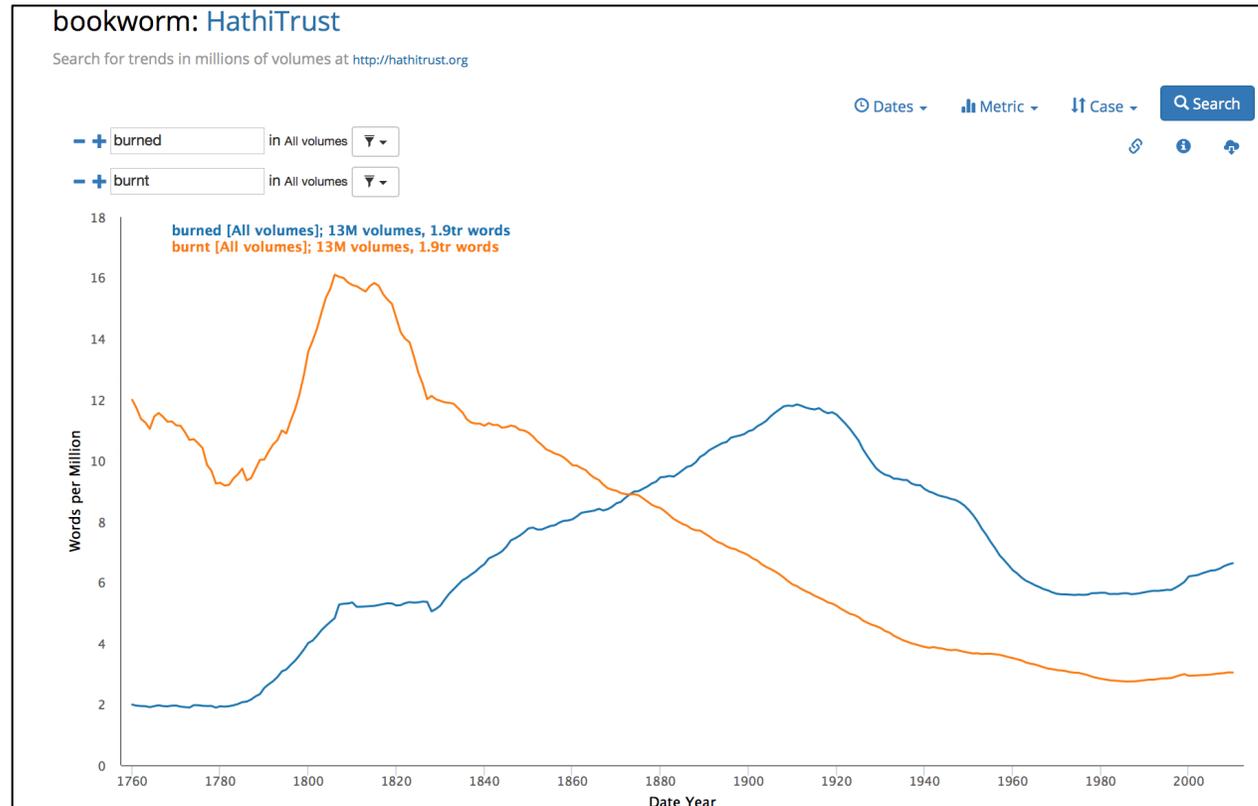


Track social change: lady vs. woman over time



Reading an HT+BW graph

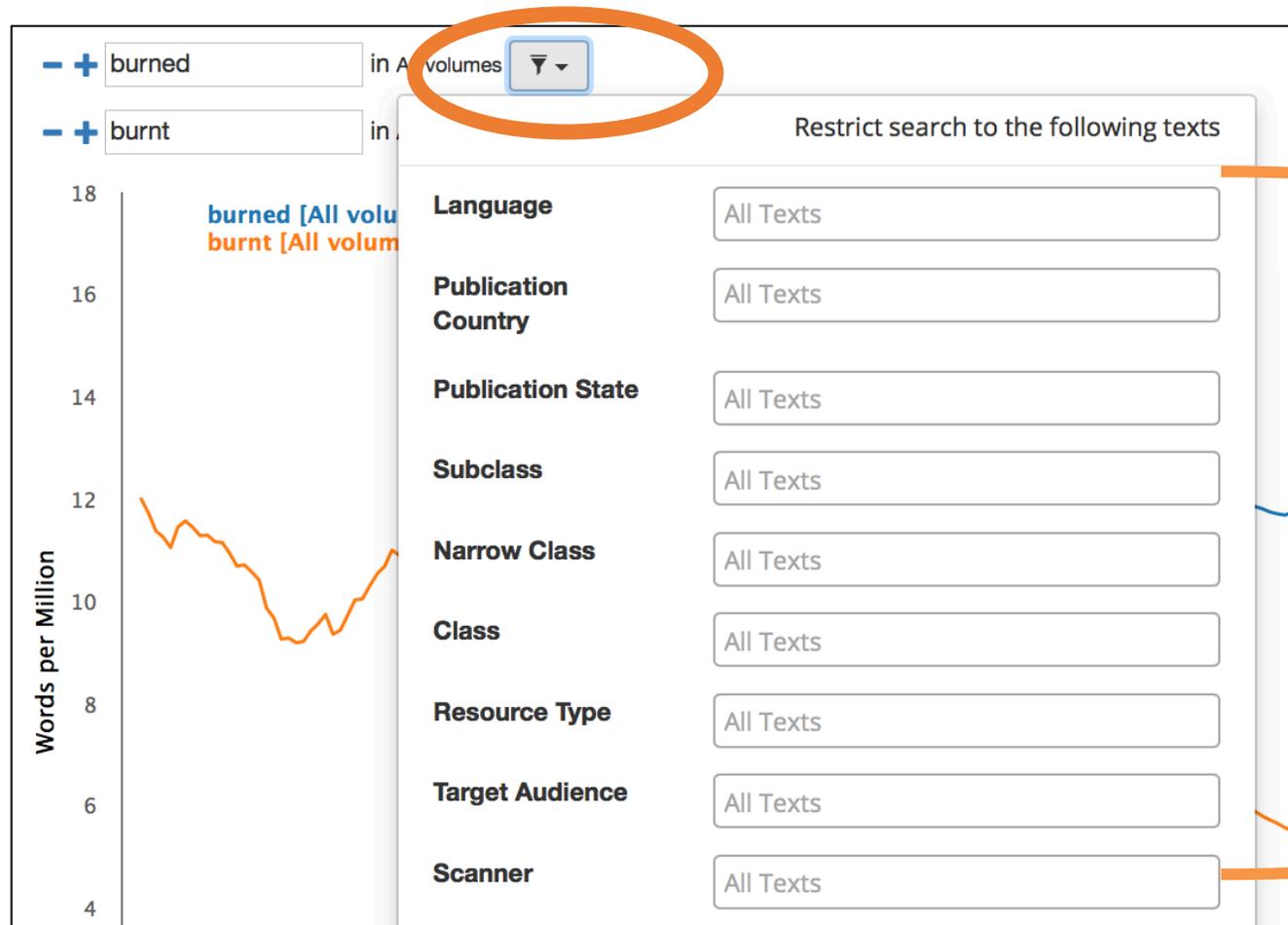
- Let's look at how verbs change over time
 - Eg. Burned vs. burnt



Do you see any trends?



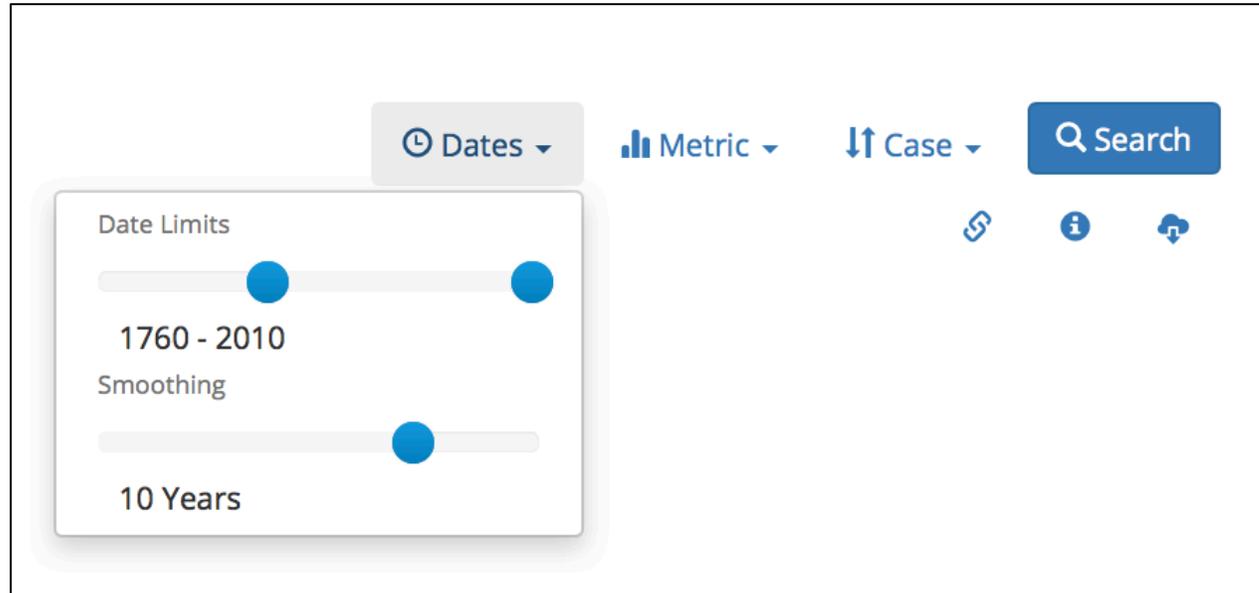
Bookworm interface



Limit
your
search
with
facets



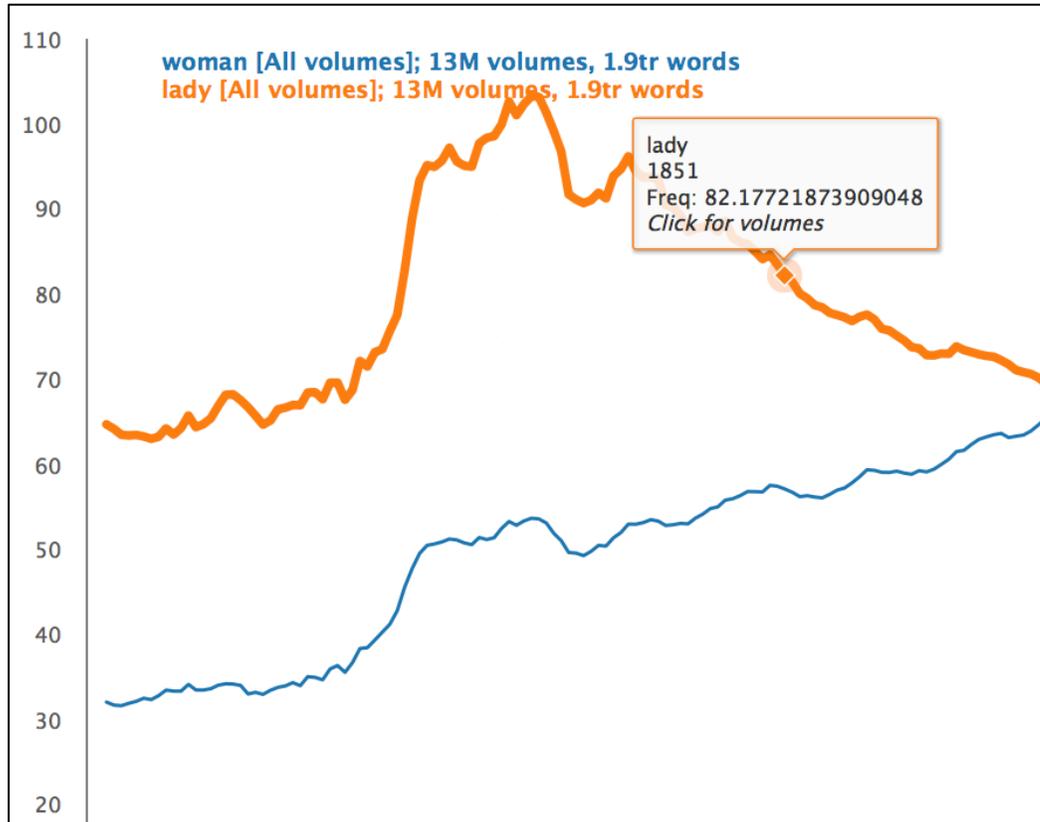
Bookworm interface



Fine-tune
your
results



Bookworm interface



About Collections Help Feedback

Search words about the items

Jump to 3 Go

Search in this text Find

8.22
C19412

LADY DI'S MINUET.

Enter Sir JOHN WILDUCK, speaking as he enters.

Sir J. Tell Lord Mulligatawney that Sir John Wilduck is in the drawing-room. Come, that's settled! I must make an end of it to-day. I don't know what to make of this Mulligatawney—a man that takes such a desperate fancy to one of a sudden, all about a shooting adventure—and will have one marry his daughter, whether one likes it or not. Every morning I come here, with my mind made up to break the thing off; but the moment Mulligatawney sees me, he rushes at me, seizes me by the hand, and calls me his “Dear Sir John—his good Sir John!” I should like to know how I'm to tell such a father as that—“Your daughter's not the thing for me; look out for another son-in-law.” Accordingly, I hesitate—I put it off to the next time. The day's gone by, and if this goes on, I shall find myself a married man before I know where I am; not that there is anything to be said against Lady Diana—she's pretty, witty, rich! Yes, by-the-bye, she has one fault—she's too short—not like my cousin Louisa, with her five feet eight. I forgot my rule—I never fall in love under my own height. How can two people step well together in harness, if one's a foot taller than the other? And then, they call it a good match. But Louisa's up to my shoulder already, and growing visibly—and the taller she grows, the better I like her; besides, our marriage is settled between the families. Well, I'm very sorry for Lady Diana, but I must tell her father to-day.

Mavis Sam

About this Book

Lady Di's minuet : a comedietta : in one act : adapted from the ... Carlingford, Chichester Samuel Parkinson Fortescue, baron, 1823-1893.

View full catalog record

Rights: Public Domain, Google-digitized.

Get this Book

Find in a library

Download this page (PDF)

Download whole book (PDF)

Partner login required

Text Only Views

Go to the text-only view of this item.

See the HathiTrust Accessibility page for more information.

Add to Collection

Login to make your personal collections permanent

Select Collection

Add

Share

f t p

Permanent link to this book

Links directly to texts in the HTDL



Sample Reference Question

I'm a student in history who would like to incorporate digital methods into my research. I study American politics, and in particular I'd like to examine how concepts such as liberty change over time.

Approach:

Explore word usage trends of political concepts within the HathiTrust using HT+BW



Hands-on activity

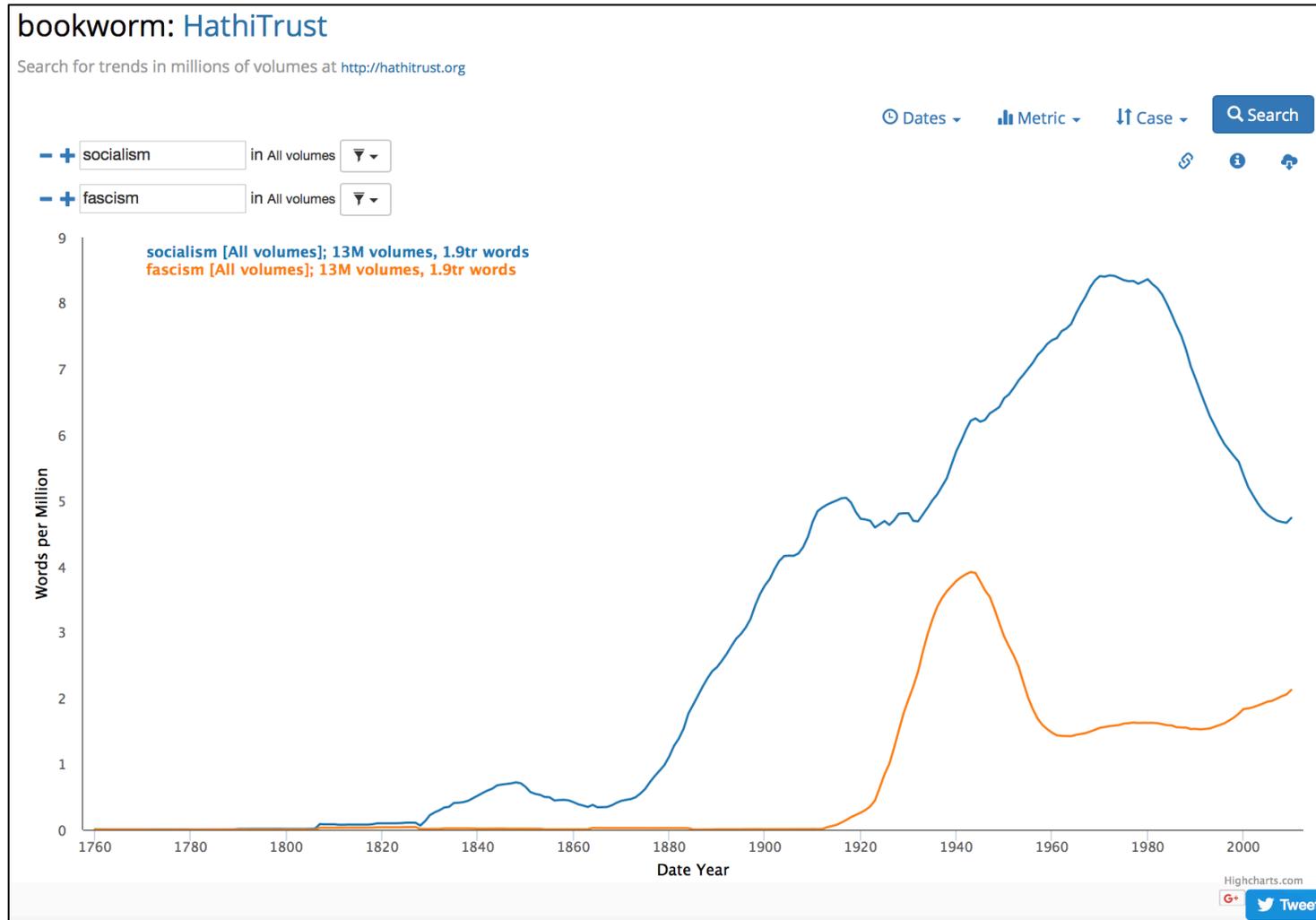
 *See Handout p. 1*

- In this activity, you will use HT+BW to explore lexical trends

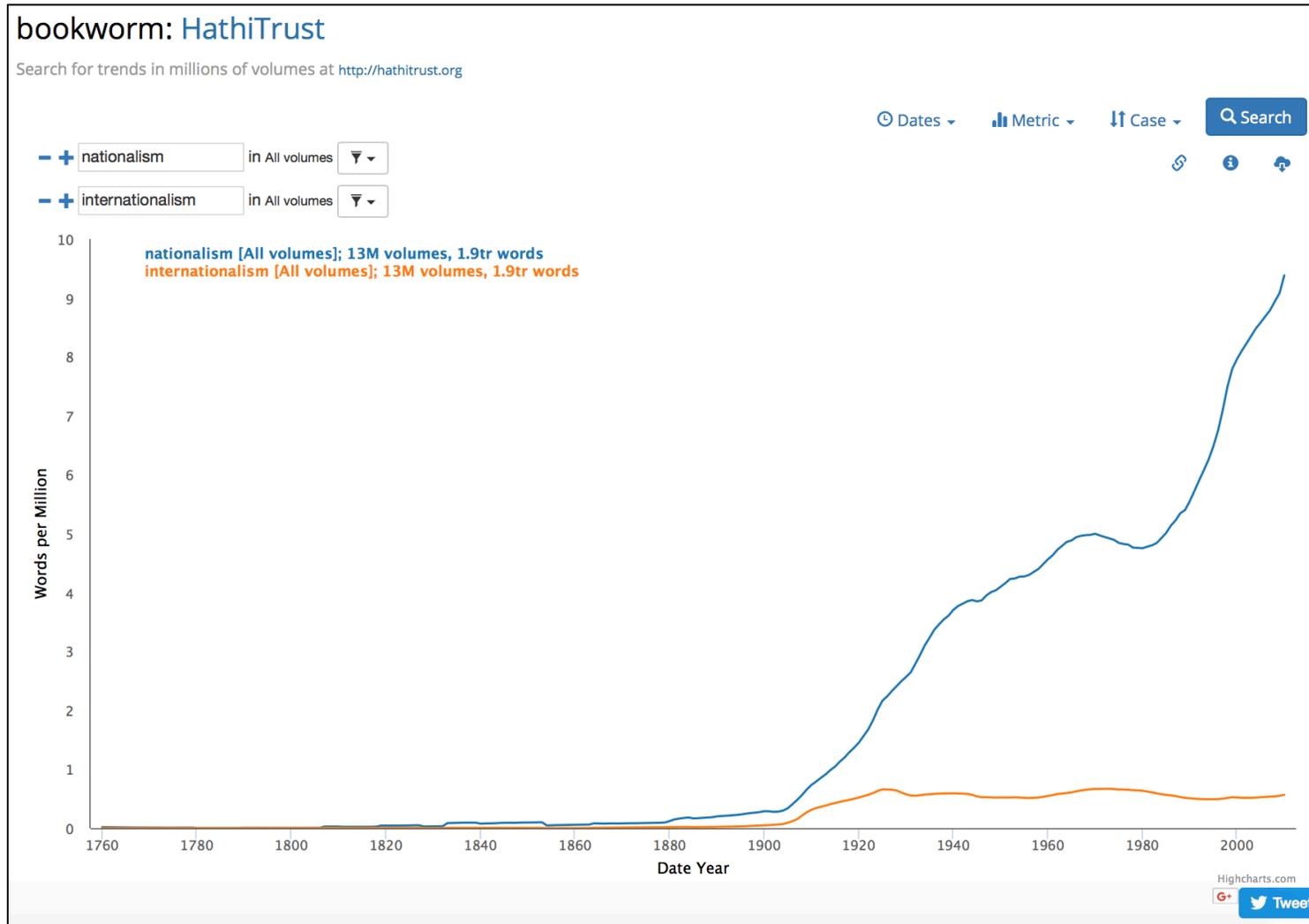
Website: <https://bookworm.htrc.illinois.edu/develop>



Examples



Examples



Discussion

- *What trends did you discover?*

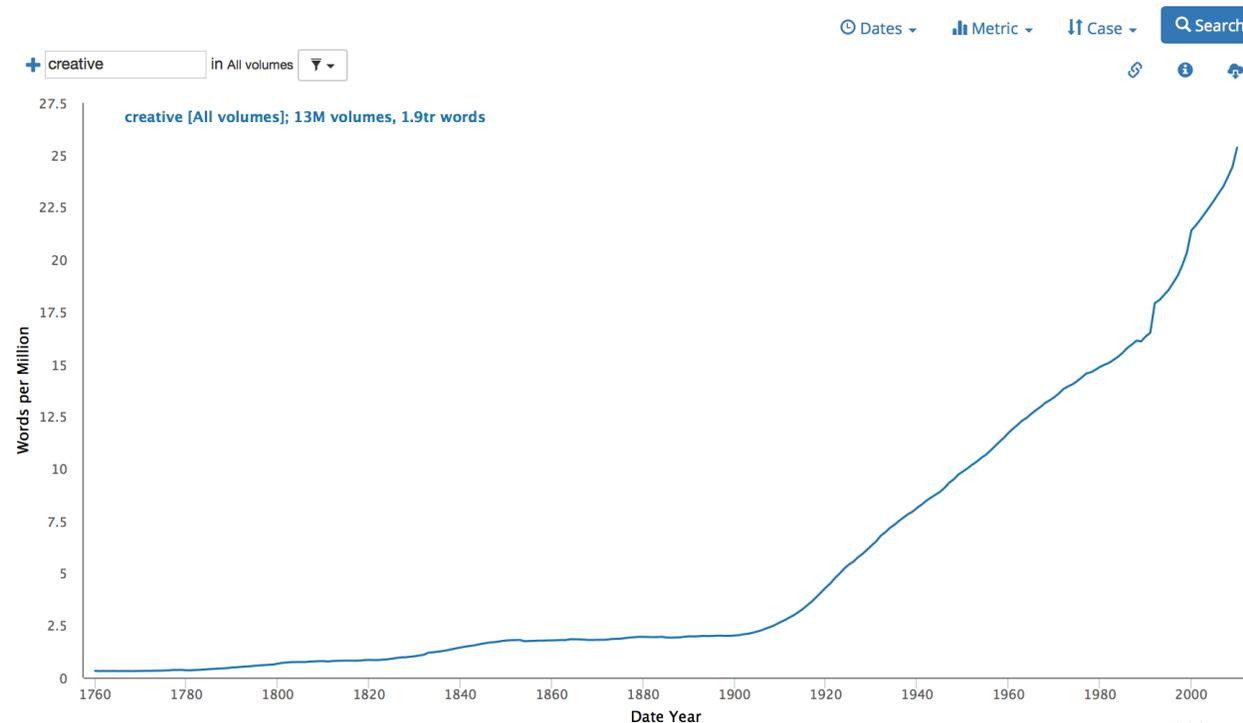


Case Study: *Inside the Creativity Boom*

- Sam used HT+Bookworm to visualize the use of “creative” in the HTDL over time

bookworm: [HathiTrust](#)

Search for trends in millions of volumes at <http://hathitrust.org>



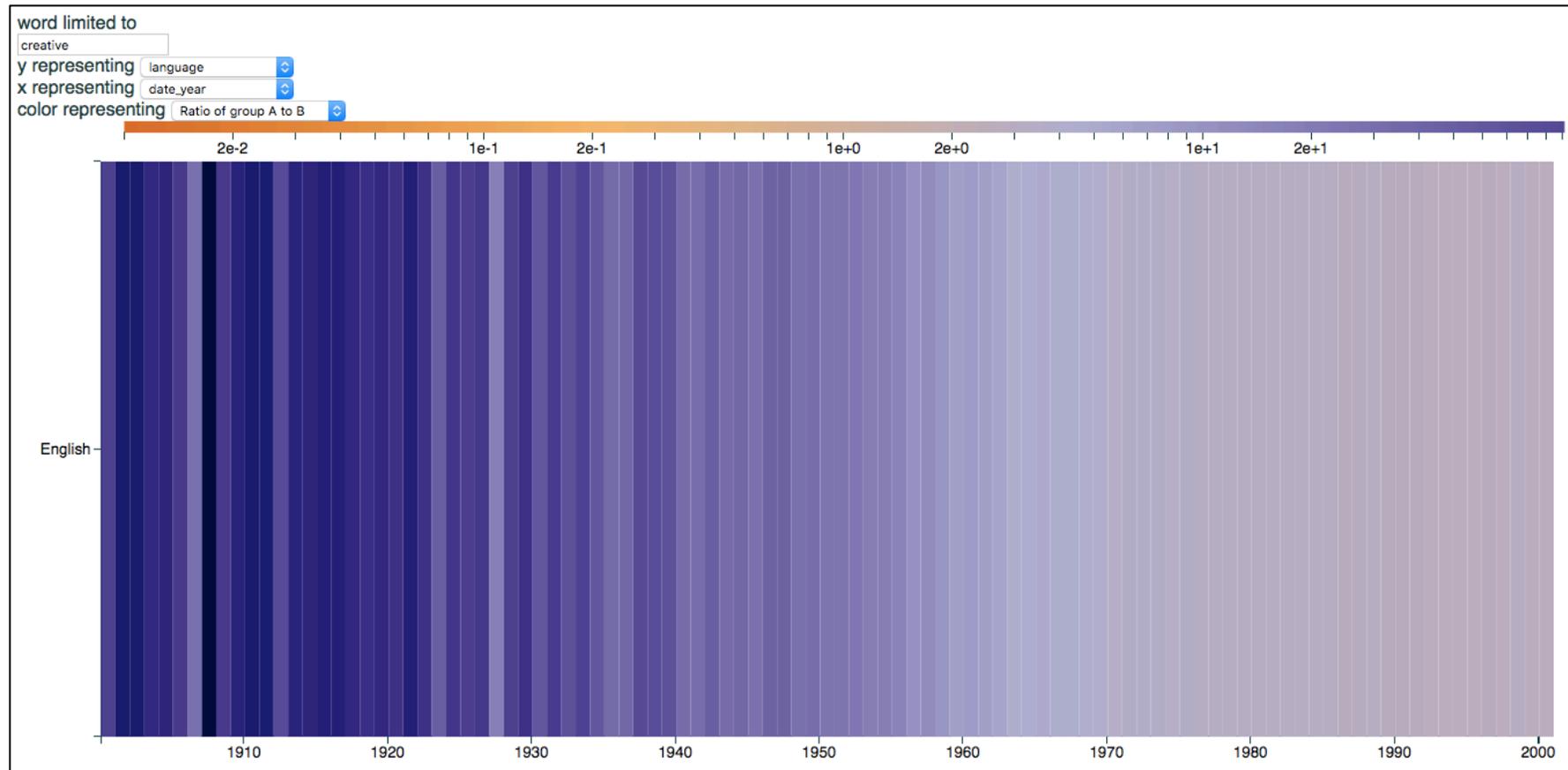
Case Study: *Inside the Creativity Boom*

- Sam also used an experimental HT+BW interface to create different kinds of visualizations...



Case Study: *Inside the Creativity Boom*

- “Creative” by language and year



Case Study: *Inside the Creativity Boom*

- “Creativity” by library classification and year



Discussion

- *Where does visual literacy fit into data literacy overall?*
- *What would it mean to be visually literate, particularly with regard to text analysis?*



Questions?



References

- Chuang, J. (2011). Text Visualization. November 2011. Retrieved January 25, 2017, from <http://hci.stanford.edu/courses/cs448b/f11/lectures/CS448B-20111117-Text.pdf> .
- Palmer K., Polley T., & Pollock C. (n.d.). Chronicling Hoosier. Retrieved August 16, 2017, from <http://centerfordigschol.github.io/chroniclinghoosier/map1.html> .
- Roskey Legal Education Blog. (2011, July 15). Martin Luther King, Jr.'s "I have a dream" speech as a word tree. Retrieved August 16, 2017, from <http://roskeylegaled.com/blog/post/martin-luther-king-jr-s-i-have-a-dream-speech-as-a/> .
- Schmidt, B. (2017, May 16). A brief visual history of MARC cataloging at the Library of Congress. Retrieved August 16, 2017, from <http://sappingattention.blogspot.com/2017/05/a-brief-visual-history-of-marc.html> .
- Schmidt, B. (n.d.). API Philosophy | Bookworm. Retrieved August 16, 2017, from https://bookworm-project.github.io/Docs/api_philosophy.html .



References

- Theguardian.com. (2013, February 12). The state of our union is ... dumber: How the linguistic standard of the presidential address has declined. Retrieved August 16, 2017, from <https://www.theguardian.com/world/interactive/2013/feb/12/state-of-the-union-reading-level>
- Underwood, T., & Bamman, D. (2016, November 28). The Gender Balance of Fiction, 1800-2007 | The Stone and the Shell. Retrieved August 16, 2017, from <https://tedunderwood.com/2016/12/28/the-gender-balance-of-fiction-1800-2007/> .
- Underwood, T. (2012, November 11). Visualizing topic models. | The Stone and the Shell. Retrieved August 16, 2017, from <https://tedunderwood.com/2012/11/11/visualizing-topic-models/> .
- Wattenberg, M., & Viégas, F. B. (2008). The word tree, an interactive visual concordance. *IEEE transactions on visualization and computer graphics*, 14(6). [10.1109/TVCG.2008.172](https://doi.org/10.1109/TVCG.2008.172) .

