Computational approaches to large-scale literary geography

Matthew Wilkens

Cornell University

Department of Information Science

Outline

1. The Textual Geographies project

2. American literary provincialism

3. Other things we can do with literary-geographic data

What is Textual Geographies?

What is Textual Geographies?

A large-scale project to:

1. identify geographic references in digitized texts,

2. associate those references with geographic data, and

3. make the results available to researchers

# How does the project work?

Textual Geographies

- HathiTrust corpus. ACLS & NEH support.

- 10 million volumes in 4 languages

- Stanford CRF-NER, Google geocoding

- 16.5 billion place name mentions

- Logistic regression for error correction; geographic LDA for disambiguation

- Design and build fast, flexible query and visualization tools

What can you do with it?

"The U.S. is too isolated, too insular. They don't
translate enough and don't really participate in the big
dialogue of literature. That ignorance is restraining."

–*Horace Engdahl, Permanent Secretary of the Swedish Academy*
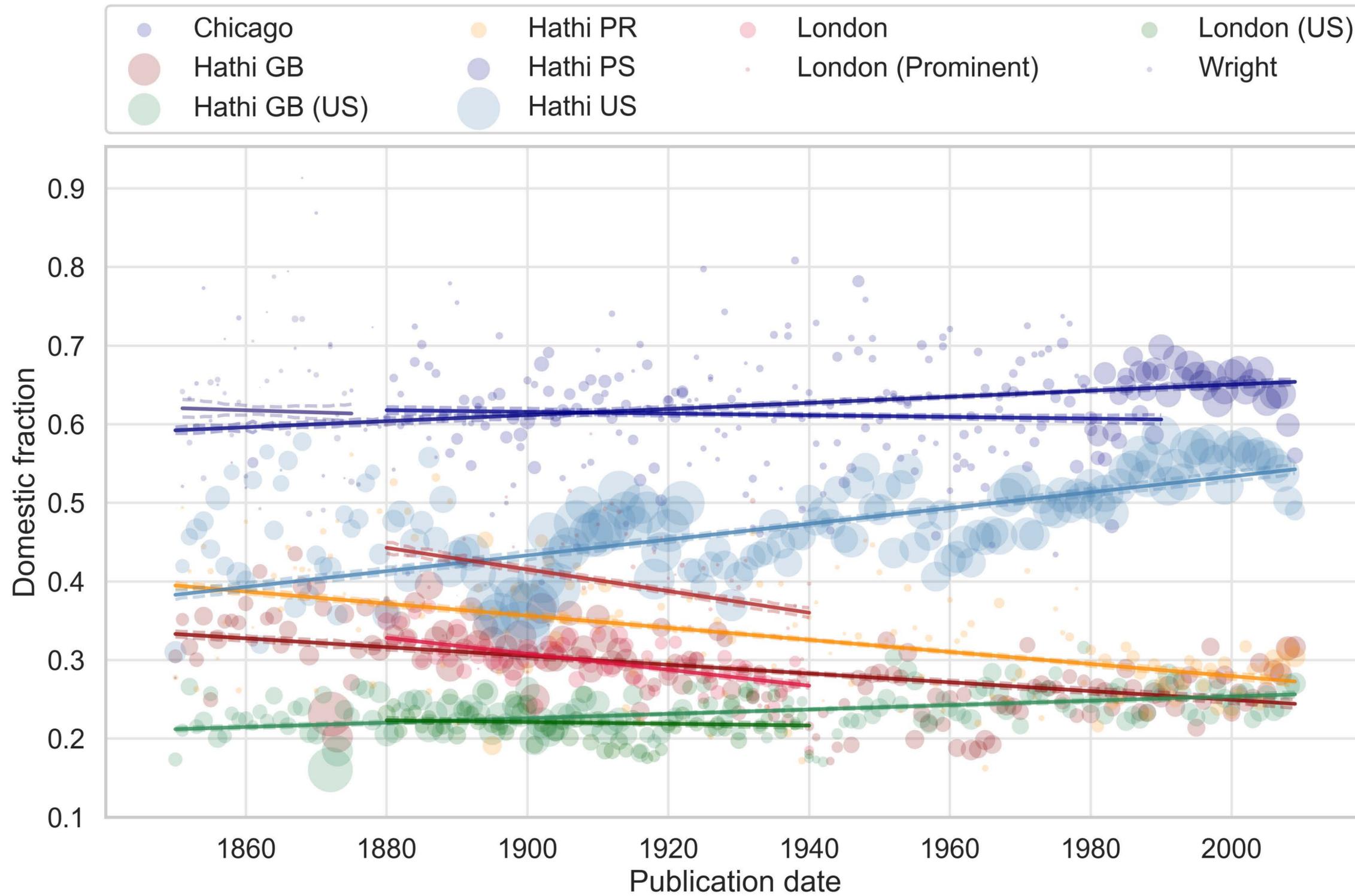
Domestic attention in fiction by American and British authors
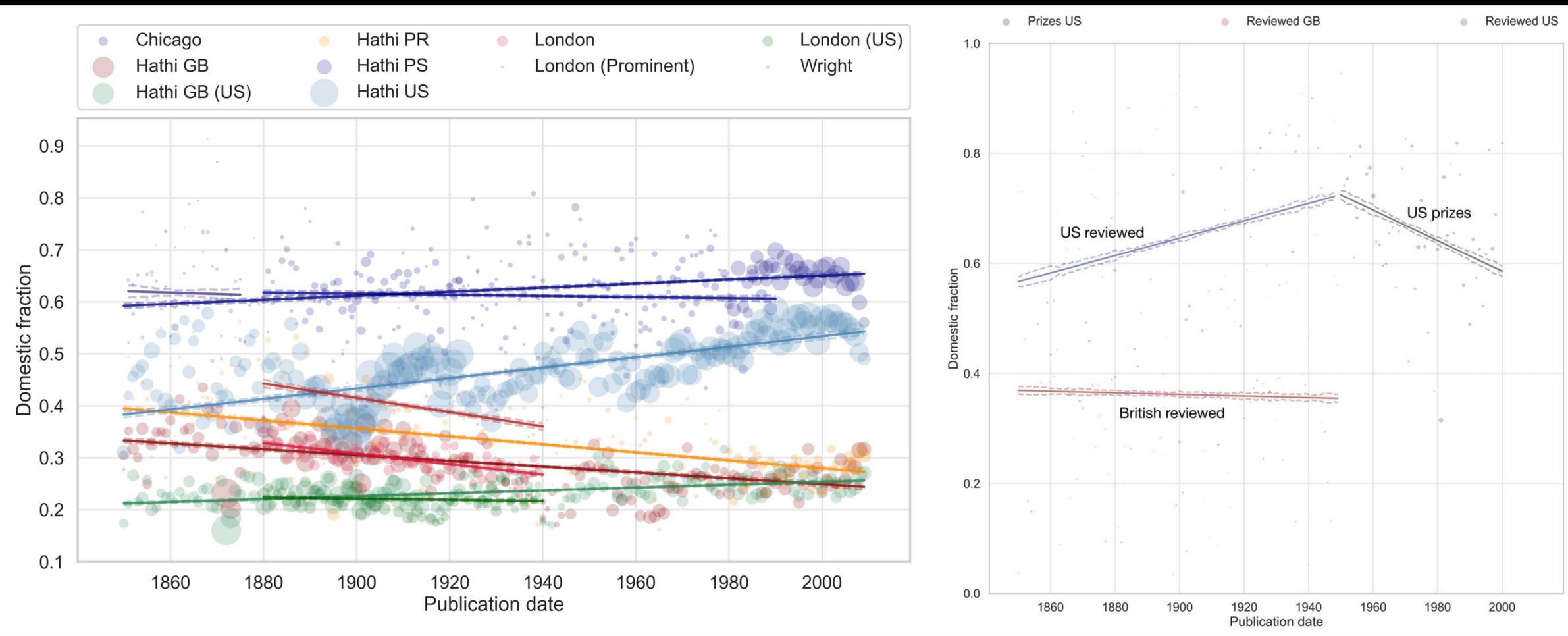
+

Hathi PR and PS

+

Hathi US

Hathi GB

+

London corpus

+

US fraction in Hathi GB

Domestic attention in fiction by American and British authors
+
Hathi PR and PS
+
Hathi US
Hathi GB
+
London corpus
+
US fraction in Hathi GB

"Prestige" fiction

Biggest movers across WWII

- Hathi US, 1900-45 vs. 1945-2010

  - Increased attention: South Korea, Vietnam, Japan, Iran, Mexico, *and US*

  - Decreased: UK, France, Belgium

- Hathi GB

  - Increased: Nigeria, Kenya, Hong Kong, Brazil, Japan, *and US*
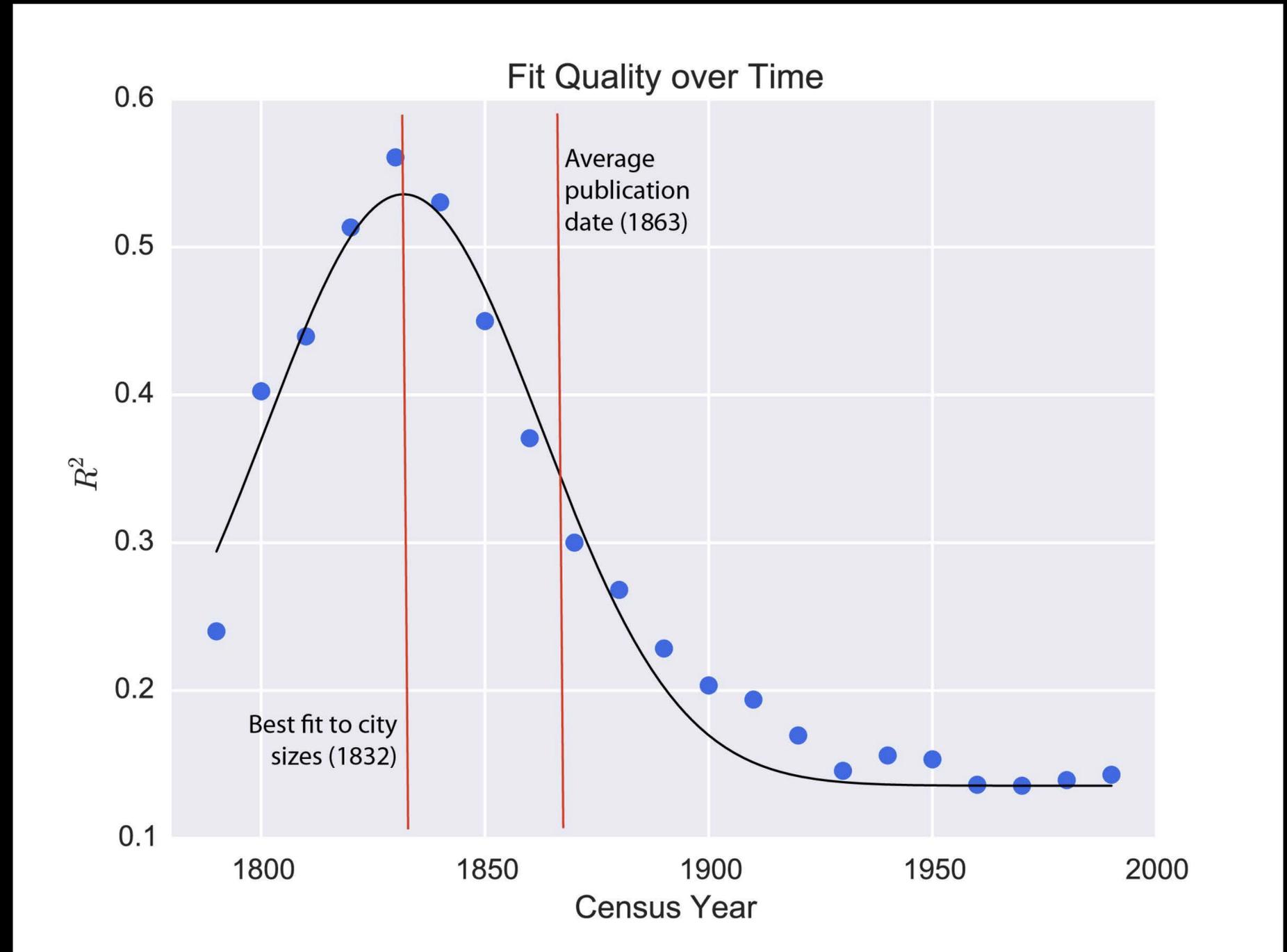
  - Decreased: France, Belgium, Spain, *and the UK*

What *else* can you do with it?

Regionalism and the Civil War

Conflict drives literary attention

Attention vs.
Population
1850-1875

Change after
Civil War

Literature and demographic change

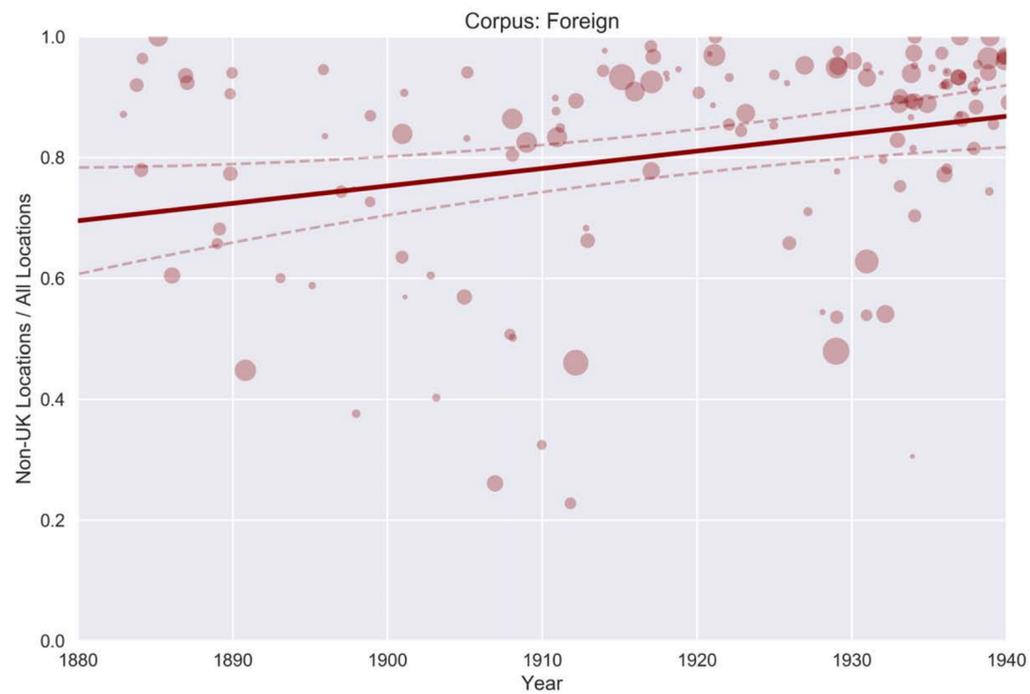Populations change over time. How does literature respond to (or anticipate) these shifts?

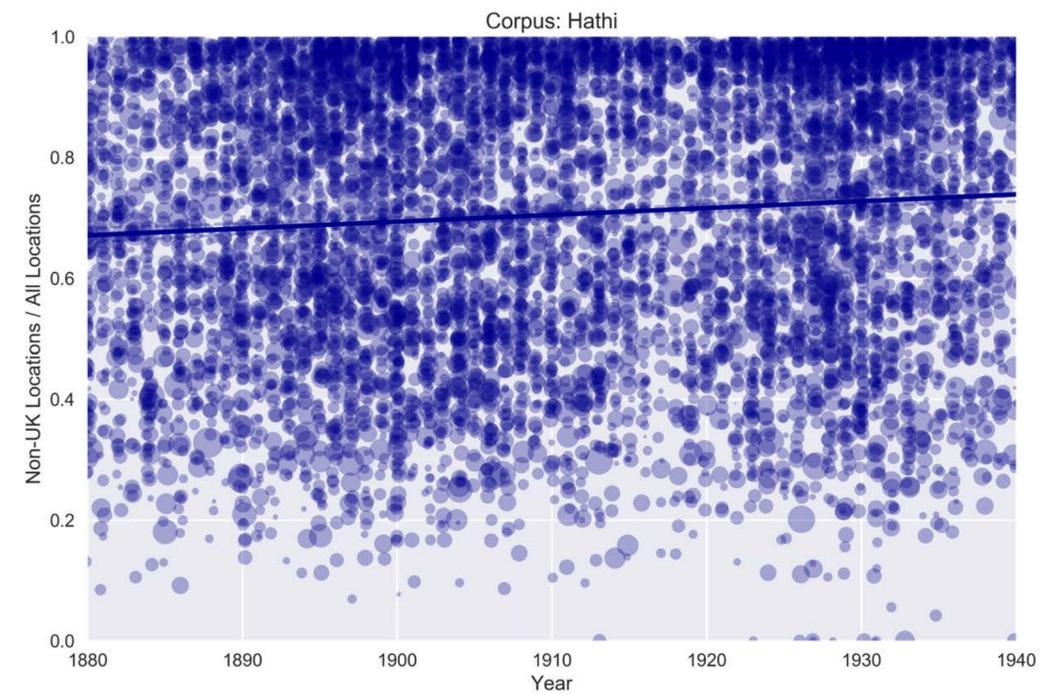With a lag of about 30 years, in the case of nineteenth-century US fiction.

Fit Quality over Time

R^2

Average publication date (1863)

Best fit to city sizes (1832)

Census Year

Nation and ethnicity in the UK before WWII

Fraction non-UK

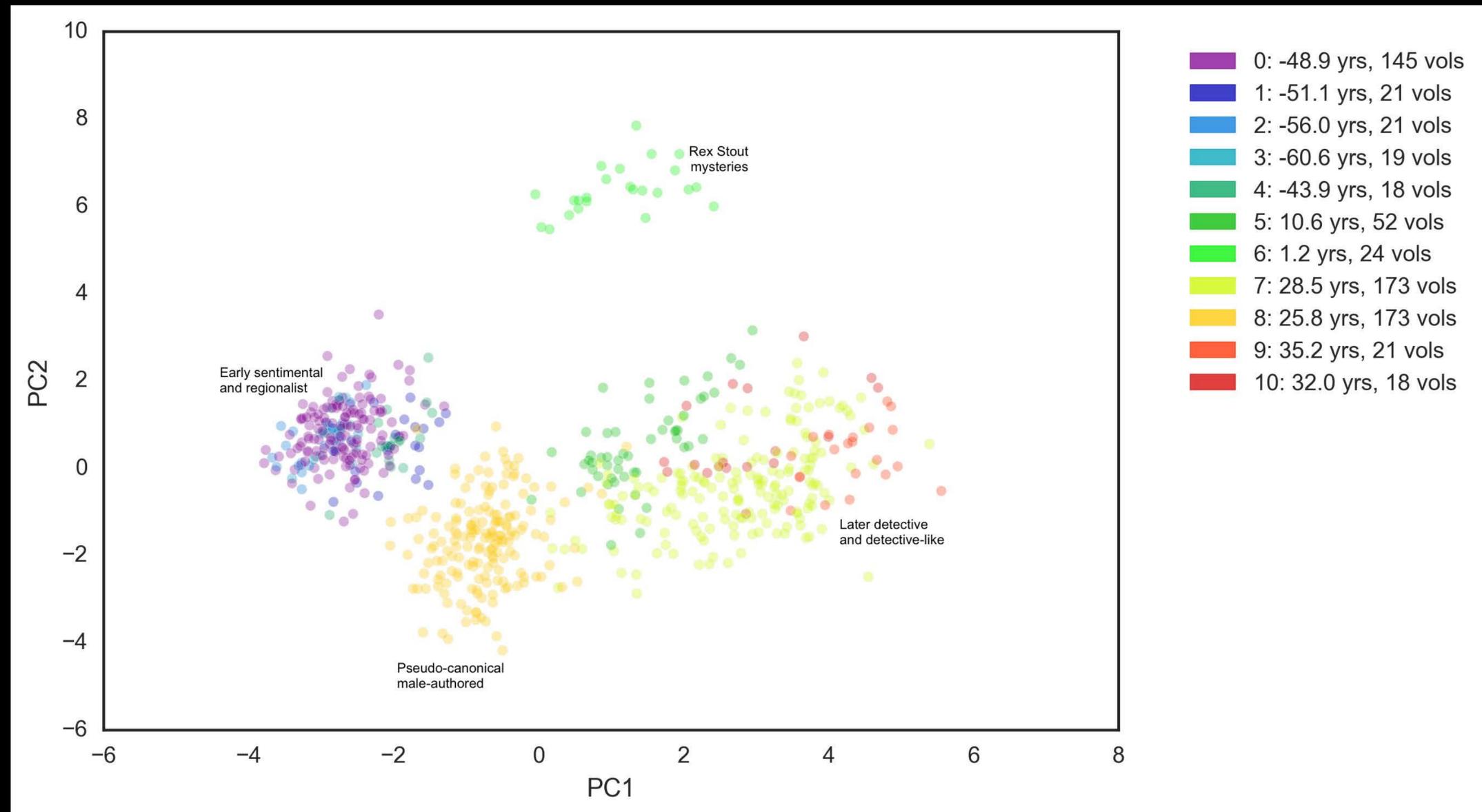# London centers of gravity by era
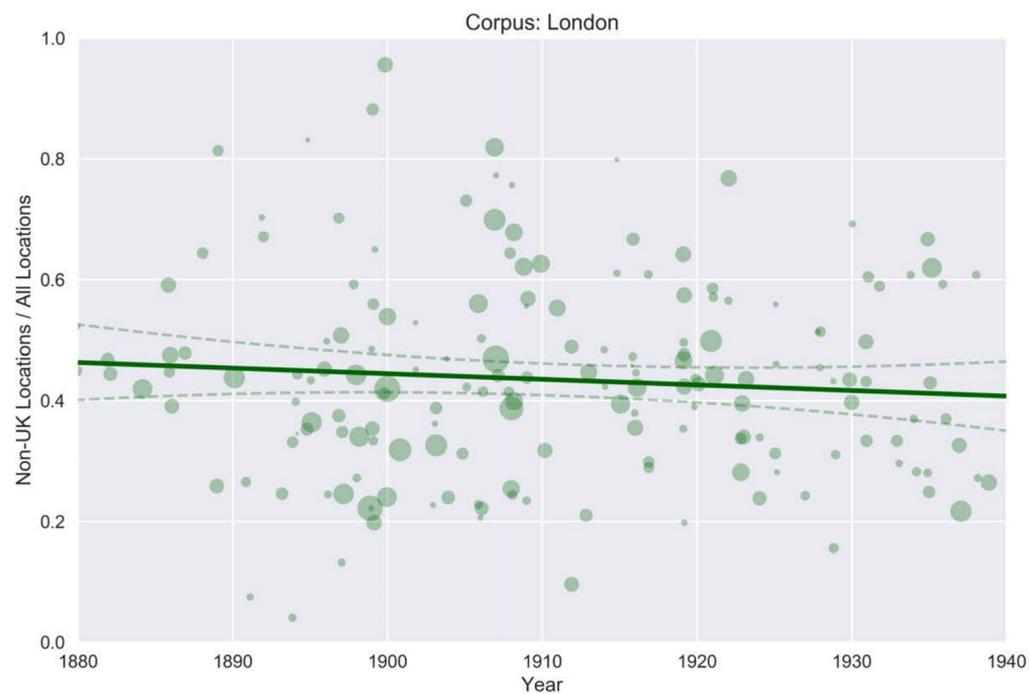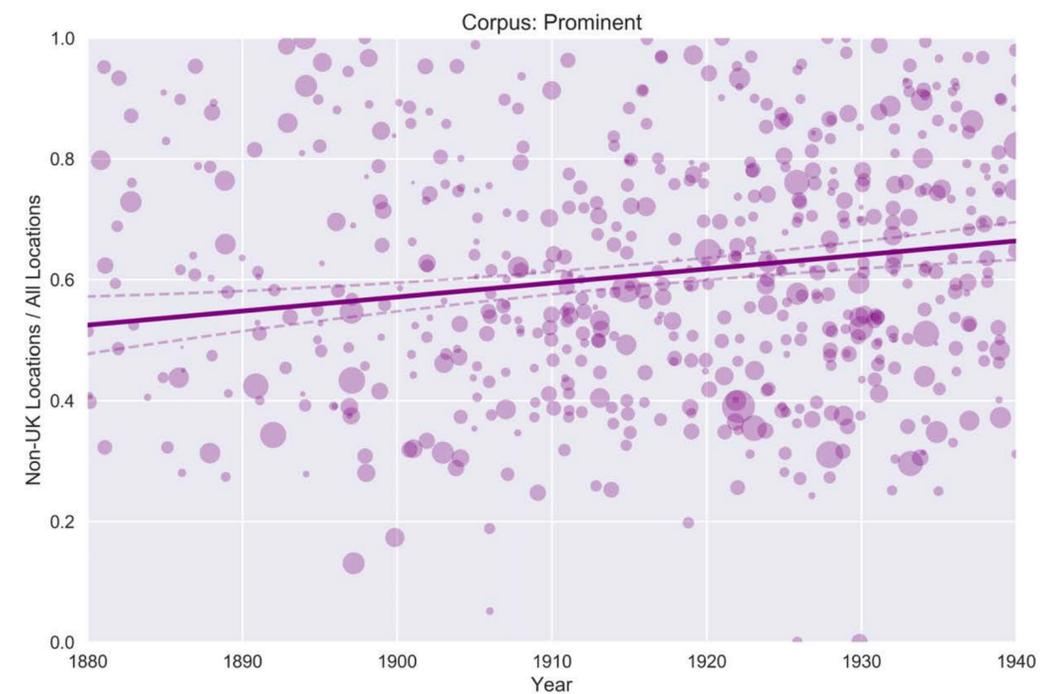
Genre

Genres of 20th-century US novel

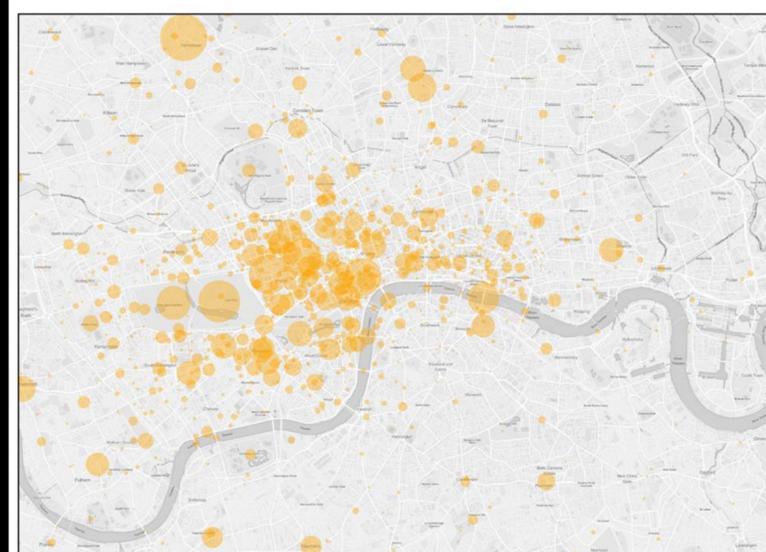Do we know the system of genres in US novels of the 20th century? Is "literary fiction" a genre?

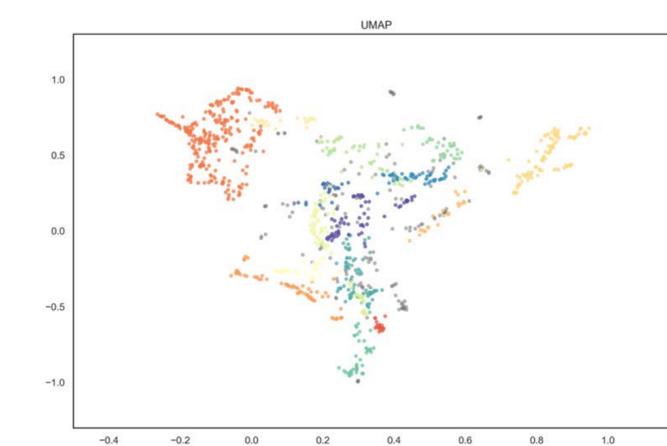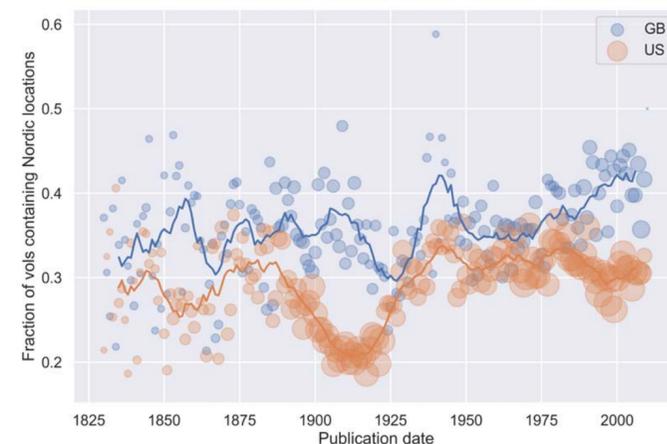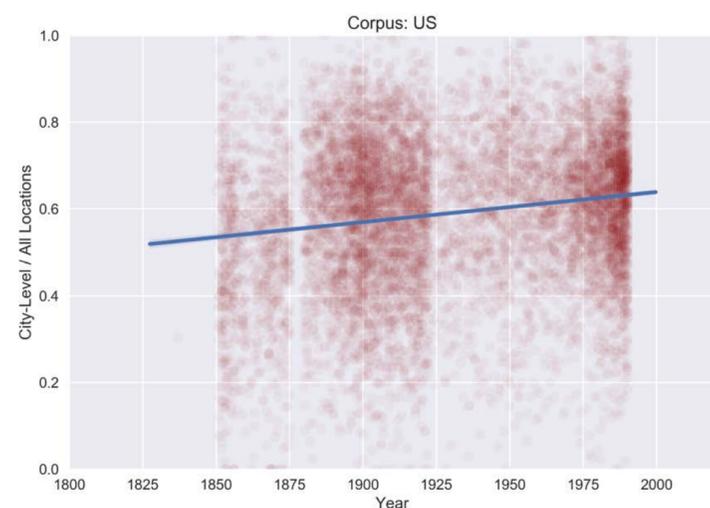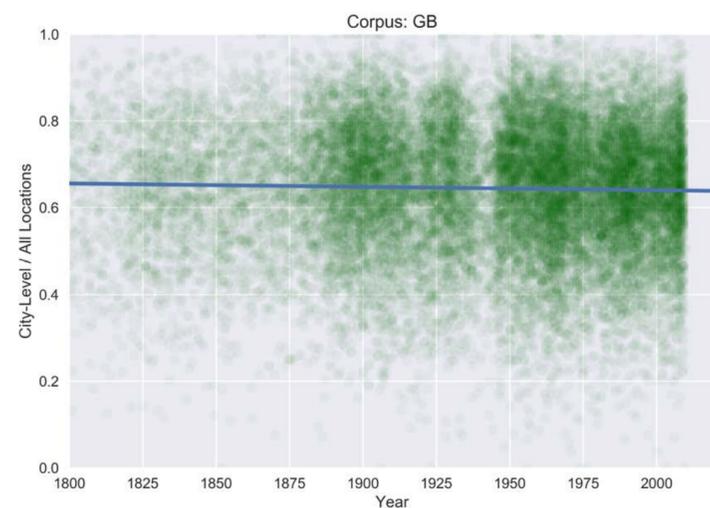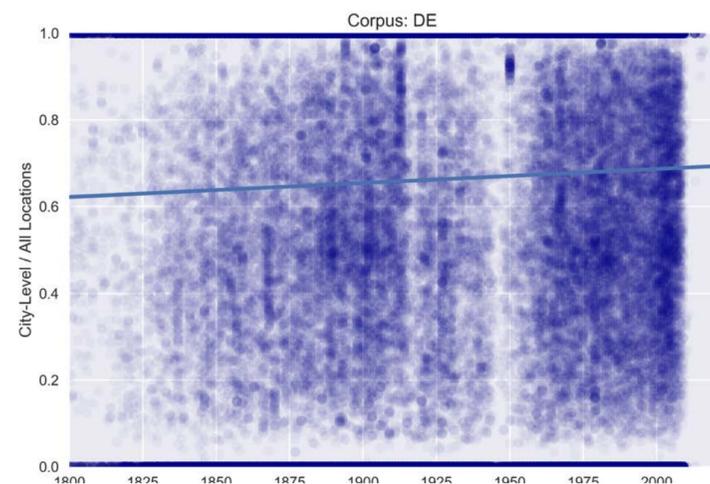Kind of and yes. But there are some surprises.

- Wilkens, Matthew. "Genre, Computation, and the Varieties of Twentieth-Century US Fiction." *CA* (2016).

Other current projects

Urbanization

Nordic geography

Publishing conglomeration

Gender and geography

In sum

- The books we read and study are imperfect proxies for a wider range of literary production

- We have the ability to work with thousands or millions of digital books to address cultural questions

- Geography and social data help us reassess central claims in cultural studies

- Computation helps to connect literary criticism, cultural studies, and the social sciences in ways that illustrate a developing trend in humanities scholarship

- Existing digital libraries and archives have significant gaps. Need targeted work to avoid and address collection bias.

Special thanks to
Prof. Elizabeth Evans, Dr. Daniel Sinykin
Kara Mlynski, Patrick Evans, John Villaflor, Megan Kollitz
Dr. Melissa Dinsman, Erik-John Fuhrer

Thanks!