

Data Sources for Modeling Library Texts and Readership



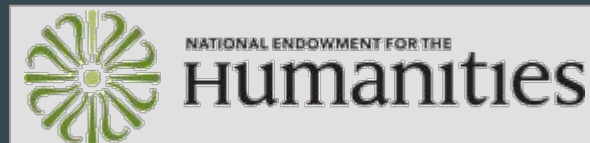
Megan Bernal, Jessica Bishop-Royse,
Robin Burke, John Shanahan
DePaul University

With

- Zachary Budde
- Sam Okrasinski
- Nandhini Gulasingam
- Jorge Staudt
- Antonio Ceraso
- Tim Zhang
- Mihaela Stoica
- Nicolas Nascimento

Thanks to

- National Endowment for the Humanities
- Microsoft Azure for Research
- Social Science Research Center at DePaul University





ONE BOOK
ONE CHICAGO

The book is just
the beginning . . .

‘Just the beginning’ because...

Each “One Book” event is a natural experiment.

- Same patrons (more or less)
- Same branches (more or less)
- Different texts
- Different promotional profiles

—→ Different participation

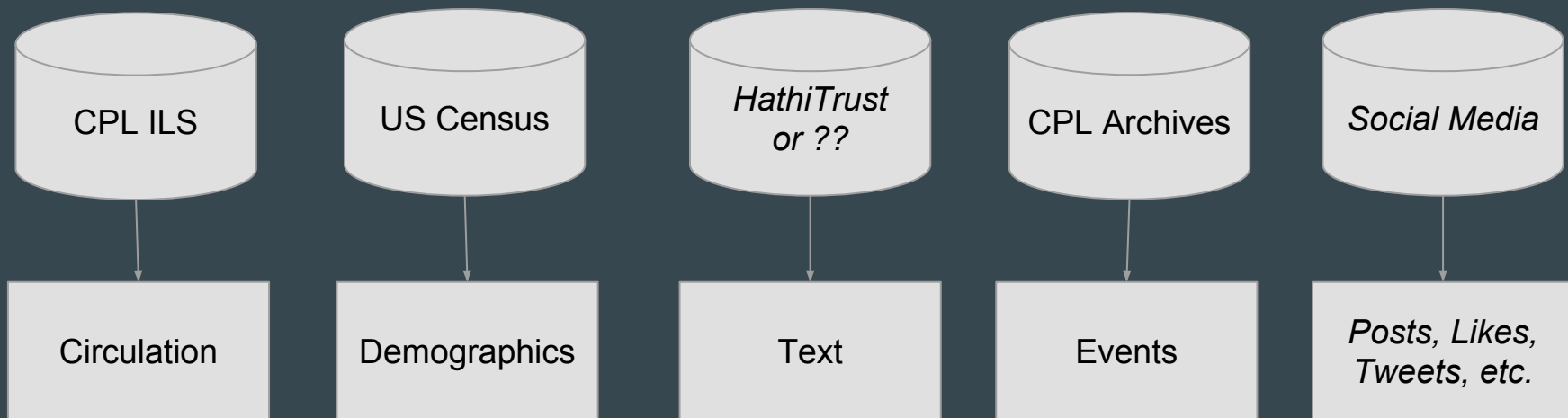
Research Questions

- Can we predict circulation patterns of a One Book offering using a new text?
- How do sponsored events and social media influence One Book reception? Over what time scale?
- Bigger questions
 - How do literary events (like One Book) shape reading at urban scale?
 - Can city-scale elective reading enhance democracy (B. Barber, *If Mayors ...*)?

Books since Fall 2011

- (AM) Saul Bellow, *The Adventures of Augie March*
 - Fall 2011 selection
 - Sprawling classic work of modernist fiction, set (mostly) in Chicago
- (GB) Yiyun Li, *Gold Boy, Emerald Girl*
 - Spring 2012 selection
 - Short stories set in China
- (BT) Markus Zusak, *The Book Thief*
 - Fall 2012 selection
 - Young adult fiction set in WW II Germany
- (WS) Isabel Wilkerson, *The Warmth of Other Suns*
 - Fall 2013 selection
 - Non-fiction history of the Great Migration
- (KC) Michael Chabon, *The Amazing Adventures of Kavalier & Clay*
 - Fall 2014
 - Fictional account of the early days of the comic book industry
- (TC) Thomas Dyja, *The Third Coast: When Chicago Built the American Dream*
 - Fall 2015
 - History of Chicago's rise and impact on America

Data sources



Tools



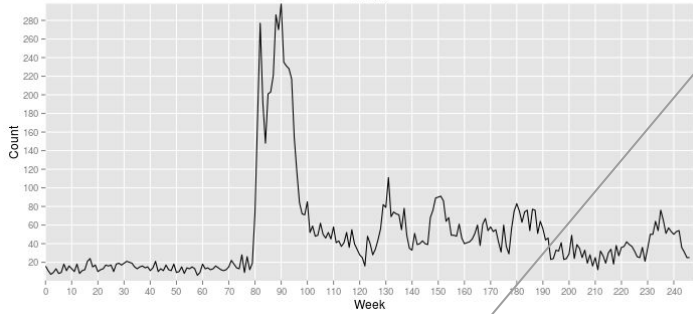
+ STACK (Social Media Tracker, Analyzer and Collector Toolkit)

Circulation

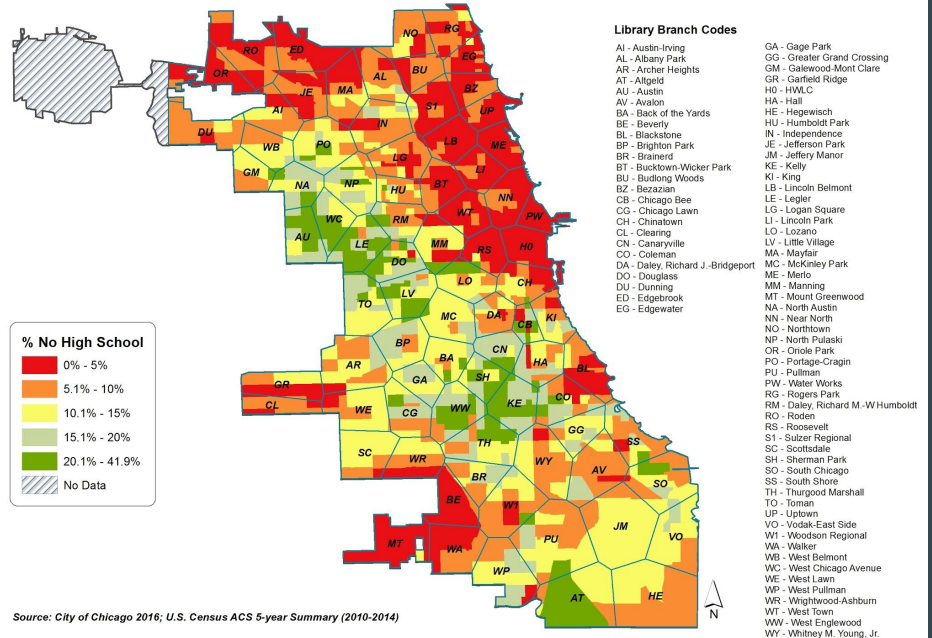
Text

Demographics

Counts by Week for All Branches for BT



Percent Population without a High School Diploma

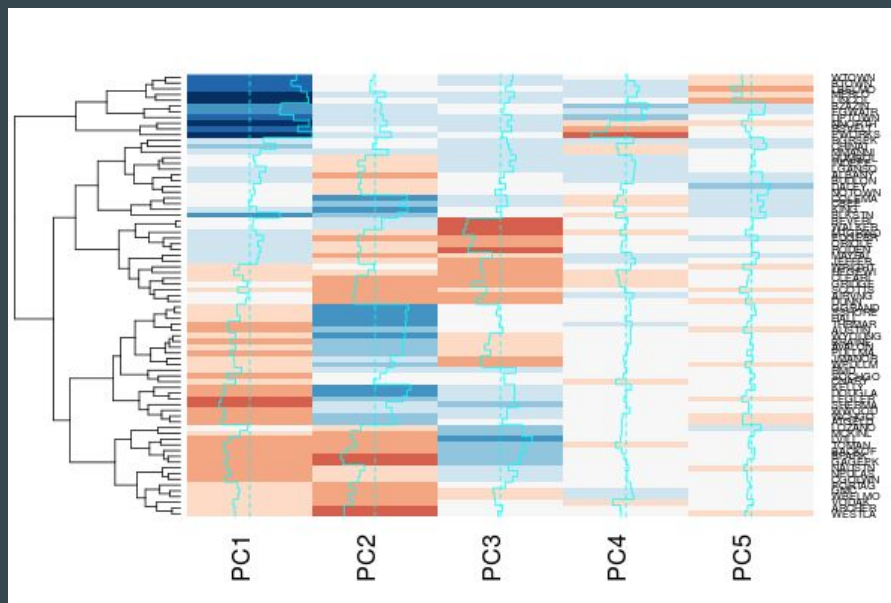
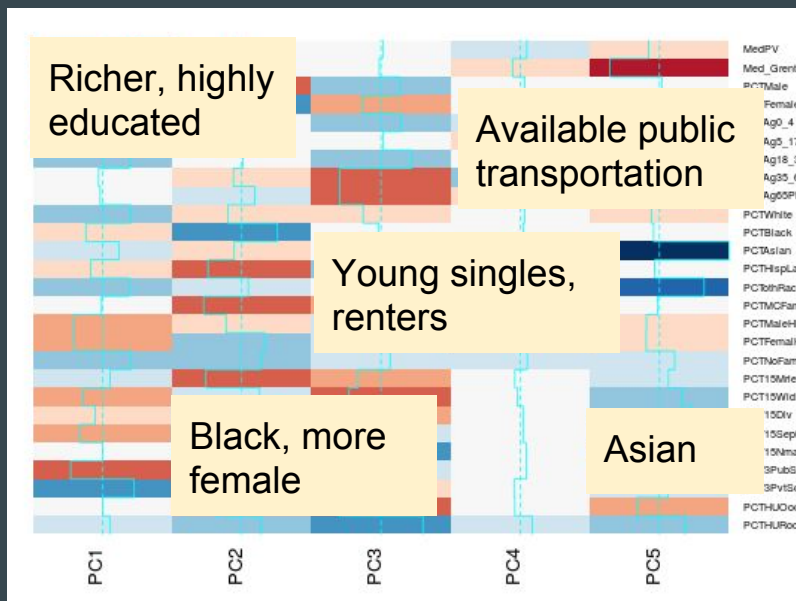


<h1 class="chapter" id="c01">DEATH AND CHOCOLATE</h1>
 <p class="nonindent">First the colors.</p>
 <p class="indent">Then the humans.</p>
 <p class="indent">That's usually how I see things.</p>
 <p class="indent">Or at least, how I try.</p>

...

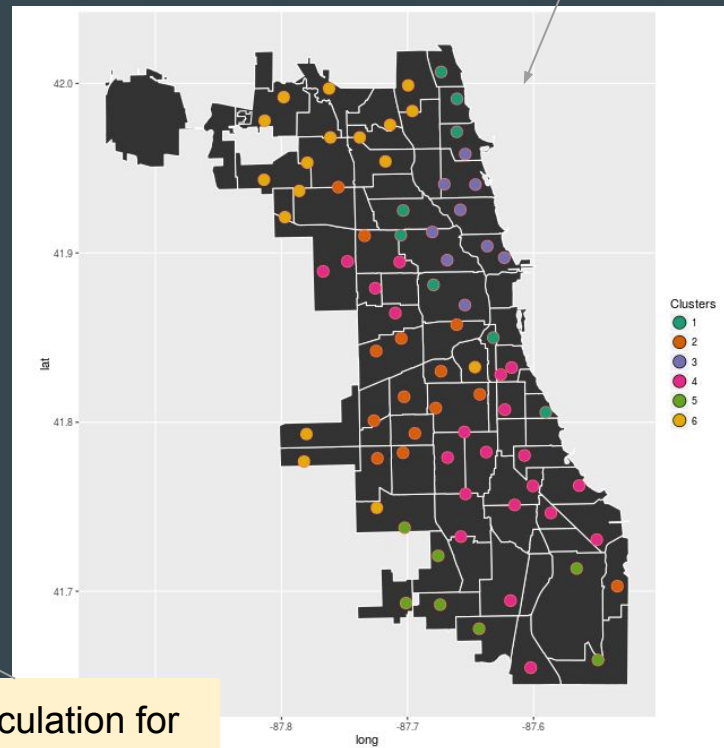
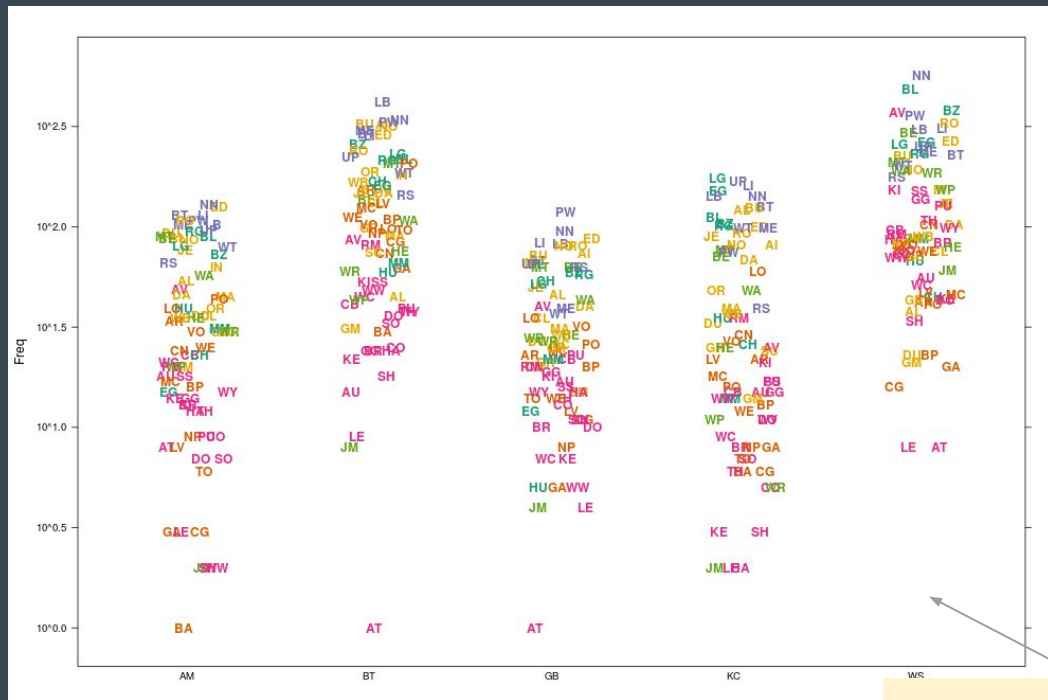
Modeling branch circulation by demographics

First 5 principal components \approx 80% of variance



Branch clusters

Branches clustered using PCA



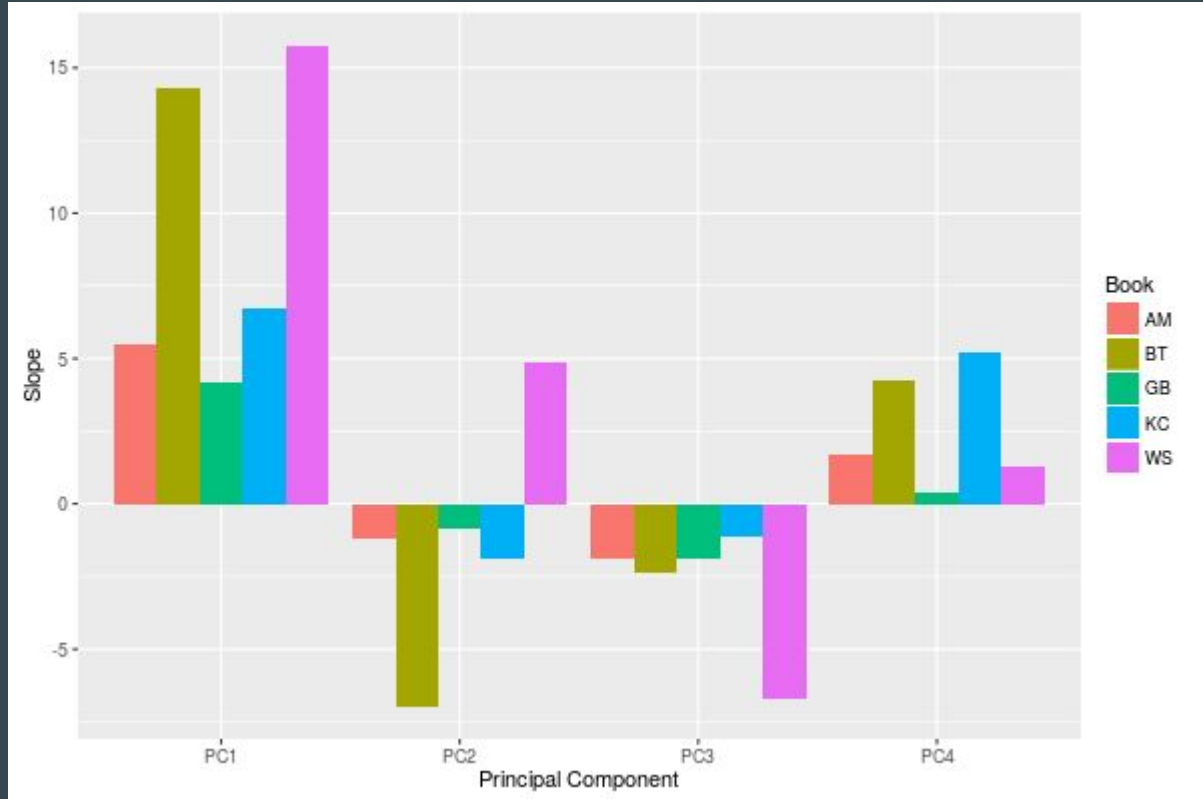
Branch circulation for each book (log scale)

Multi-level model

Circulation as a function
of principal components
grouped by book

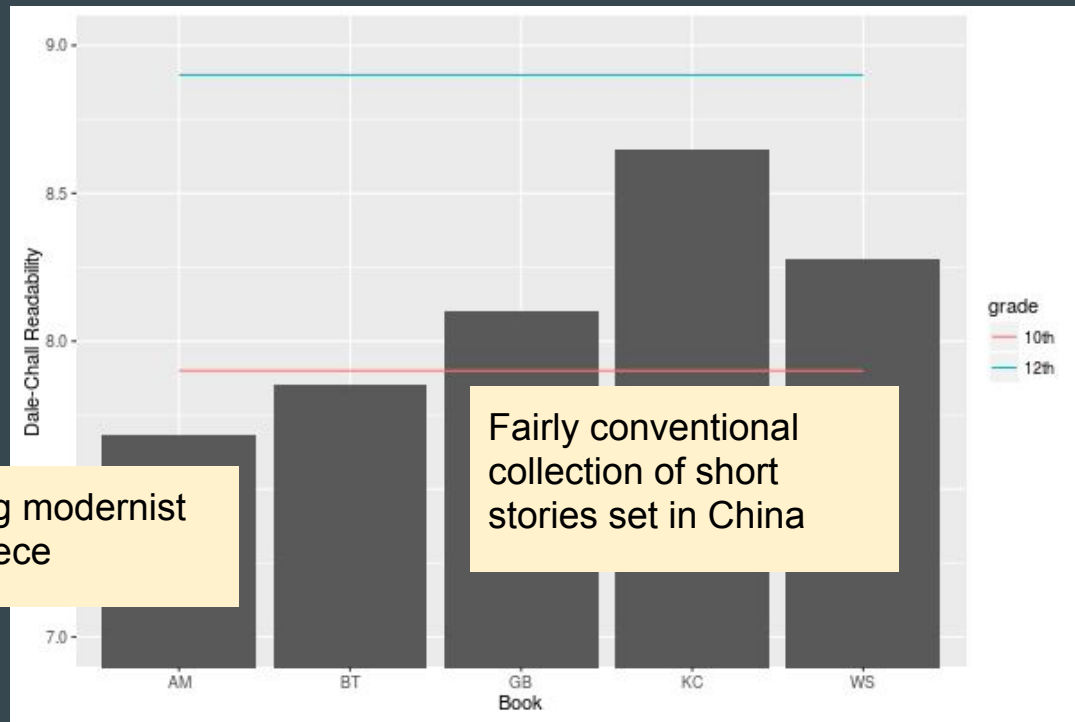
Accounts for 65% of variance

For a better model,
(and for predictions)
we need features from books



Text features

- Only 5 texts
- Very heterogeneous
 - Genre, length, audience
- Most text analysis assumes comparable texts
 - Counter-intuitive results



Sprawling modernist masterpiece

Fairly conventional collection of short stories set in China

Circulation ~ Demographics + Text

- Structural equation modeling (?)
- Some hypothesized constructs from the multi-level model
 - Literacy
 - Group identification
- Ideas? Better modeling approaches?

Learn more at dh.depaul.press/reading-chicago/

Questions? Ideas?