

# Data-intensive tools for modeling and visualizing mass reading

...

Mrkq#Vkdqdkdq#2#Uhdglqj#Fklfdjr#Jhdglqj©#Surnfw  
GhSdxq#Kqlyhuvlw|

# PERSONNEL AND SUPPORT

## PRINCIPAL INVESTIGATORS

John Shanahan  
(English)

Robin Burke  
(Computing)

Ana Lucic  
(Library)

## CHICAGO PUBLIC LIBRARY

Jennifer Lizak

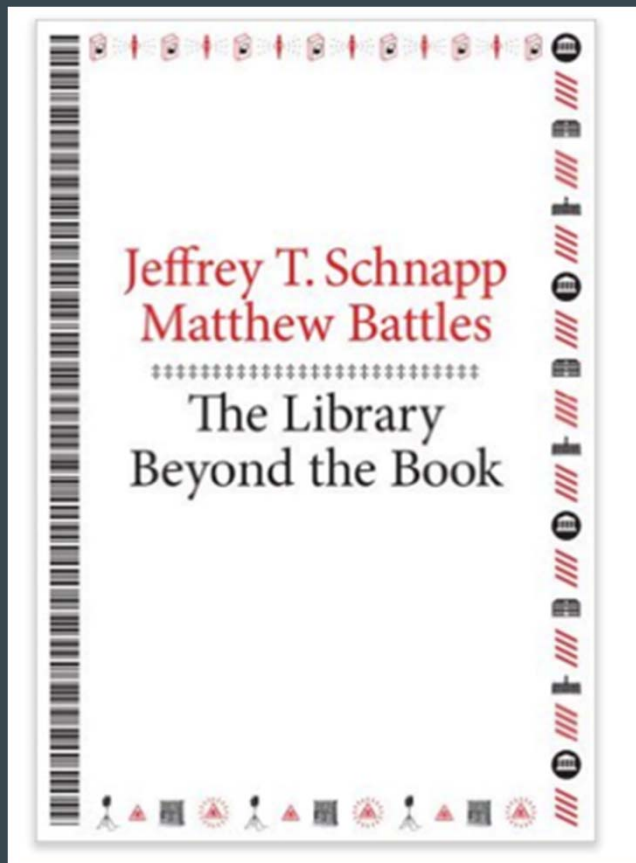
## SSRC SUPPORT

Jessica Bishop-Royse  
Nandhini Gulasingam

## STUDENTS:

Danno Aasland, Hyunyou  
Choi, Yingting Guan,  
Monika Moore, Nasim  
Sonboli, Mihaela Stoica,  
He Yang, Tim Zhang





## Library data

“Every time a book is taken off the shelf, a file is downloaded, or a computer work station is booted up, a story is told, and cataloged, and filed away in a database. In this way, each act of reading in the library broadcasts a handful of seeds, from which new growths of data will either spring—or disappear into a forest of statistical noise” (Schnapp and Battles 2014, 126).



Since September 2001.  
One of the earliest “One  
Book” programs --  
27 events and counting

Circulation data for “OBOC” books,  
since 2011

## DATA SOURCES



Bokeh

CPL ILS → Circulation

US Census → Demographics

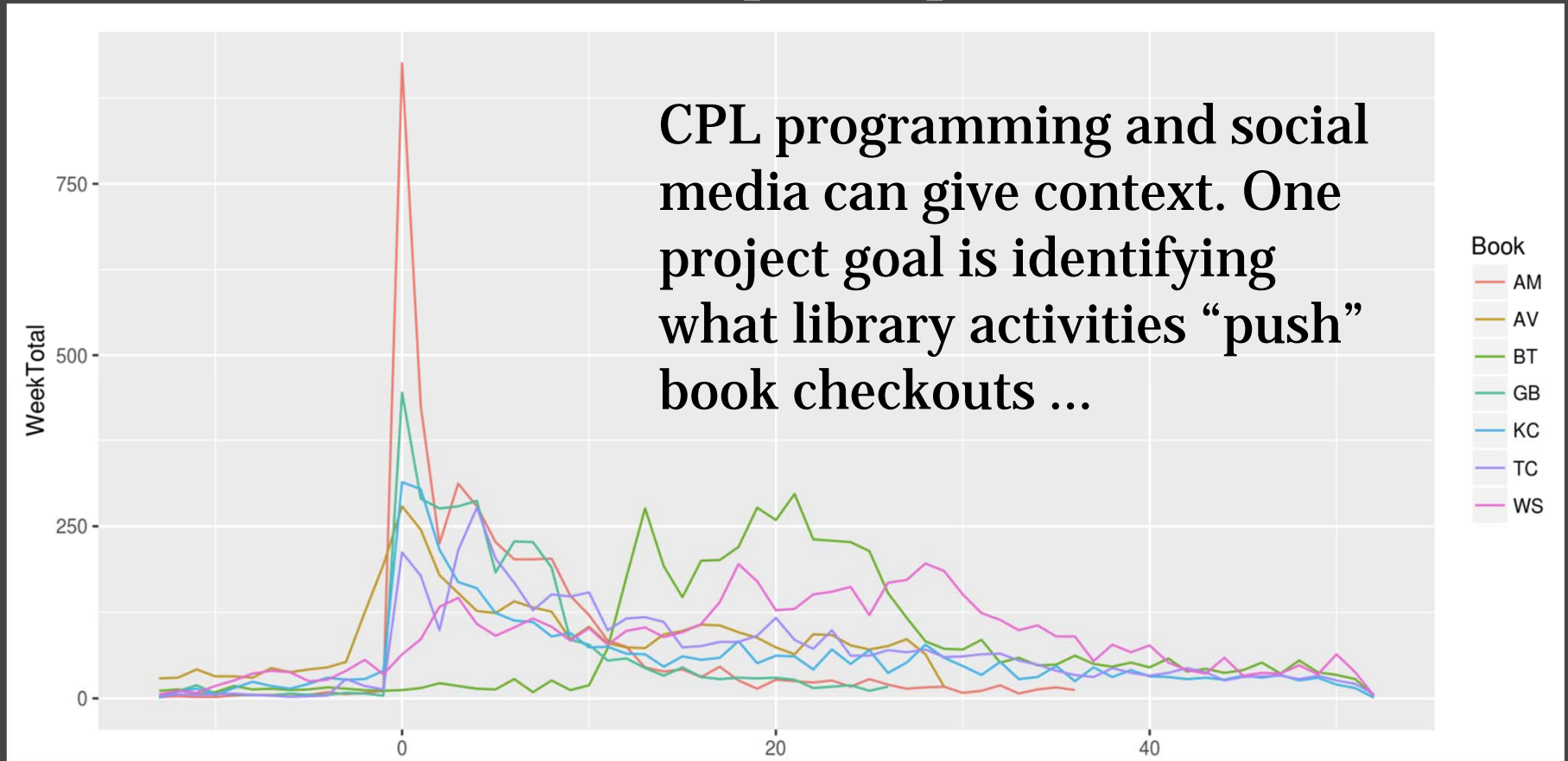
HathiTrust → Text

CPL Archives → Events

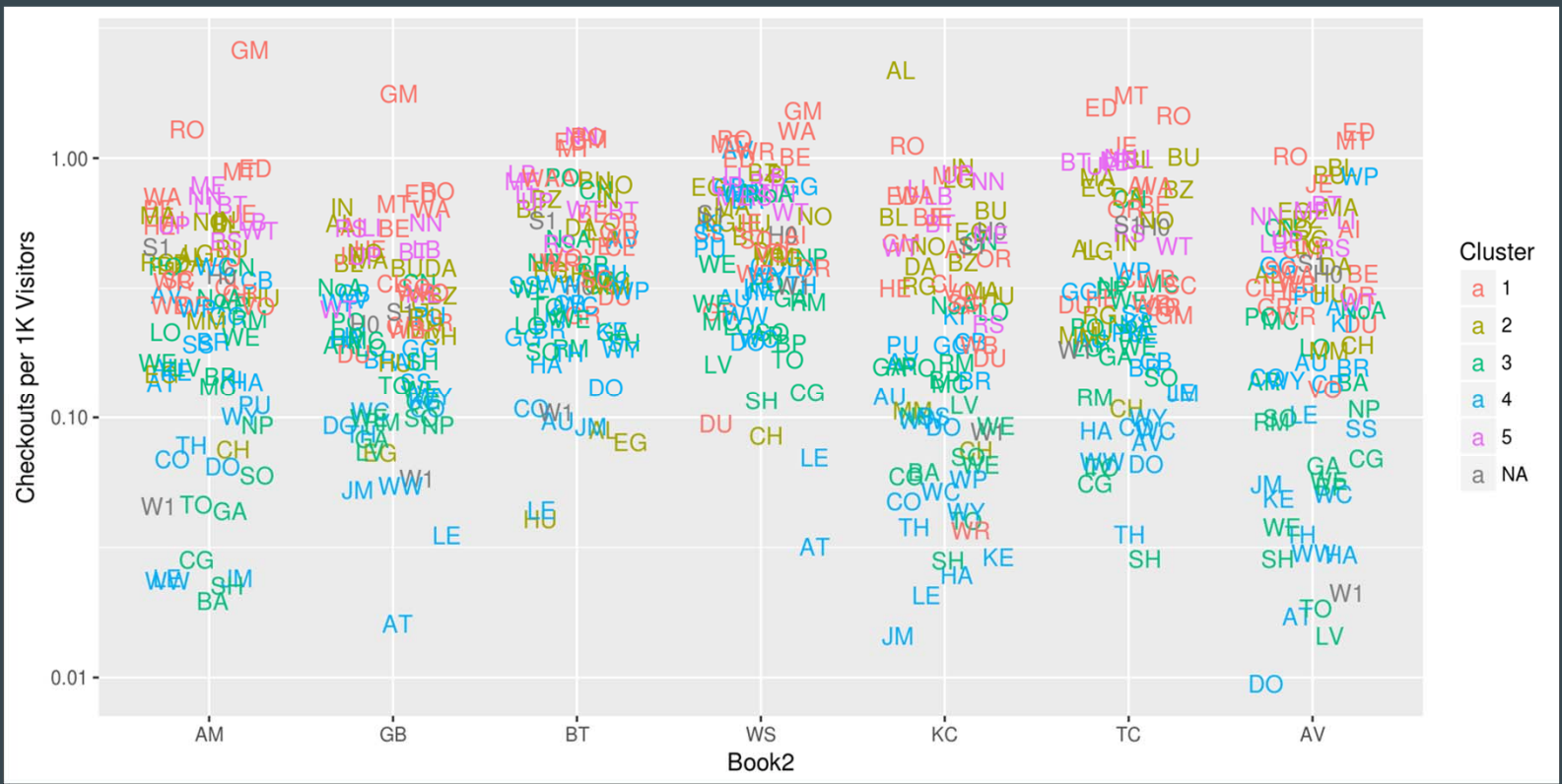
Social Media → Tweets, GR, etc.

Dashboard

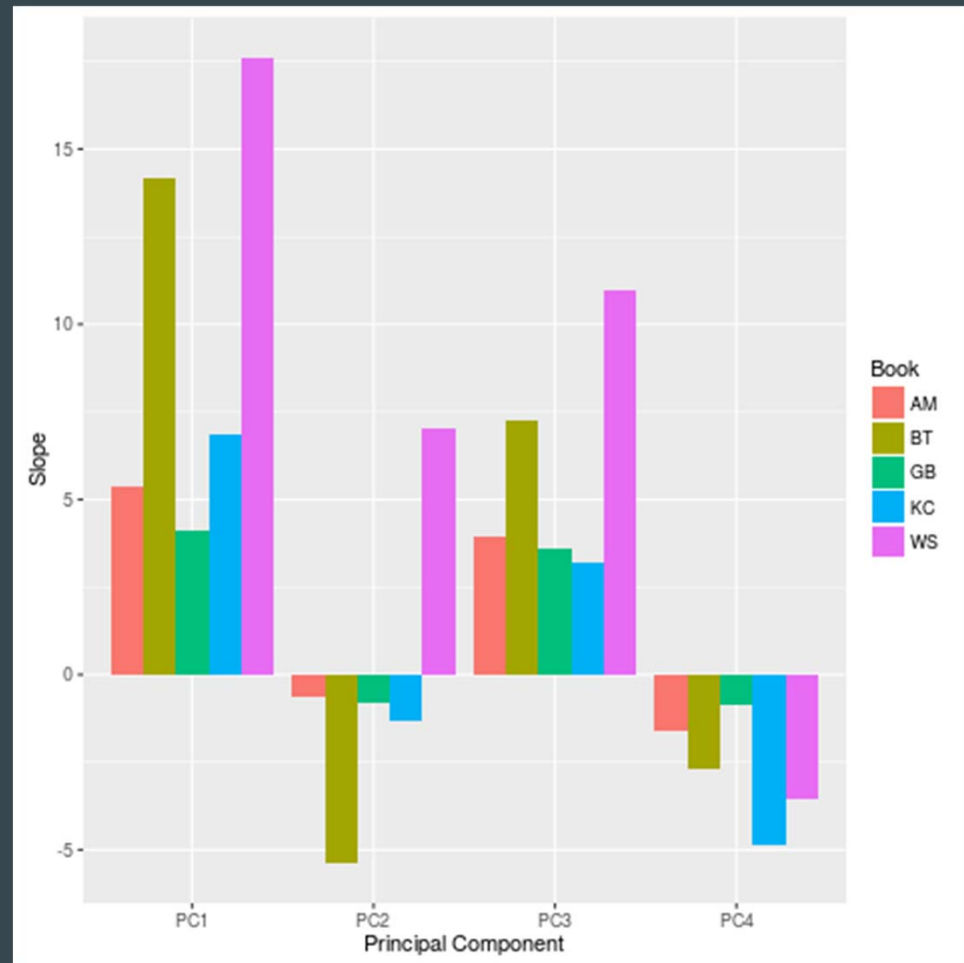
# Seven OBOC seasons superimposed



# What branches / books?

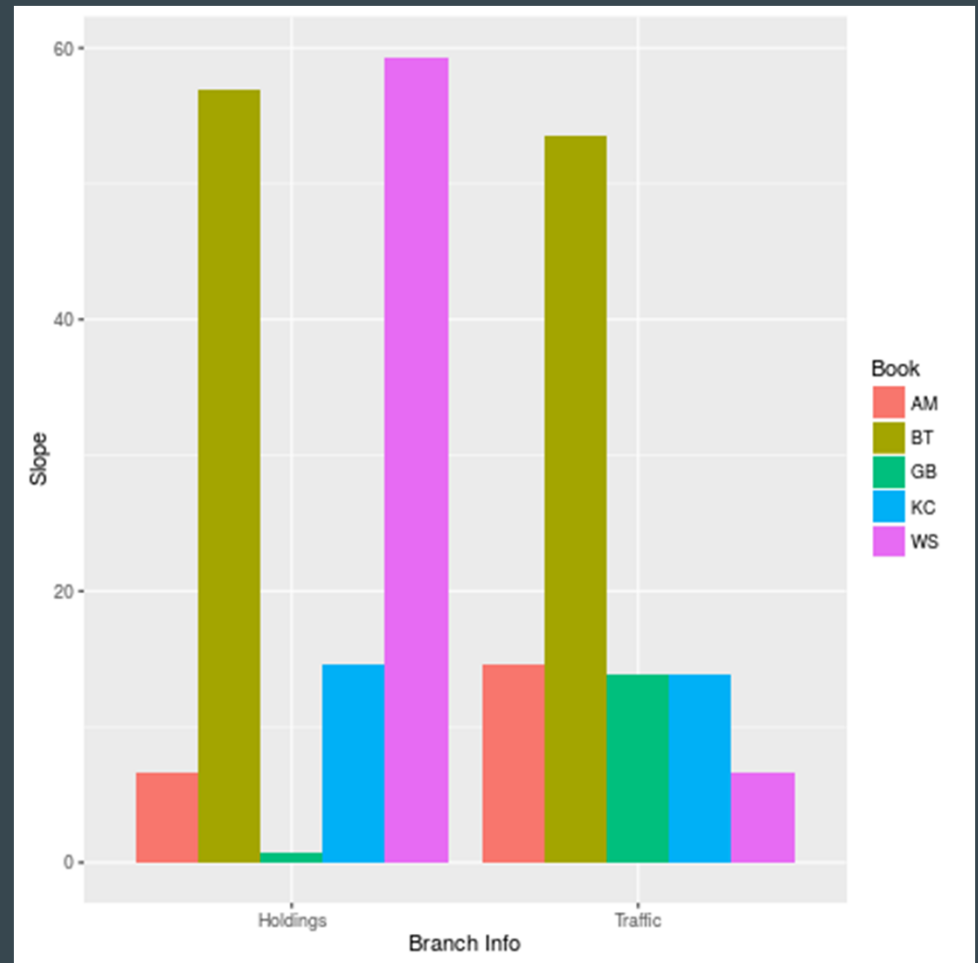


# Who reads?



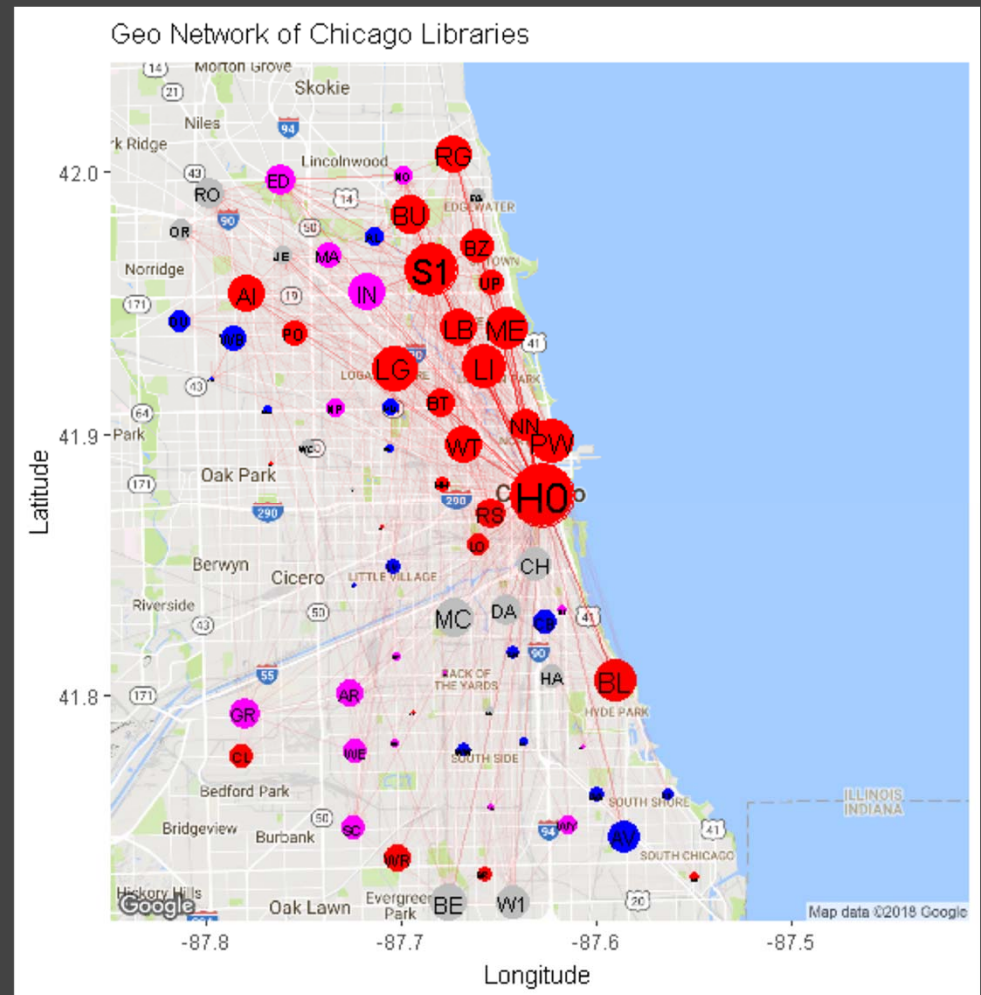


# What drives checkouts?



# How do books move in the system?

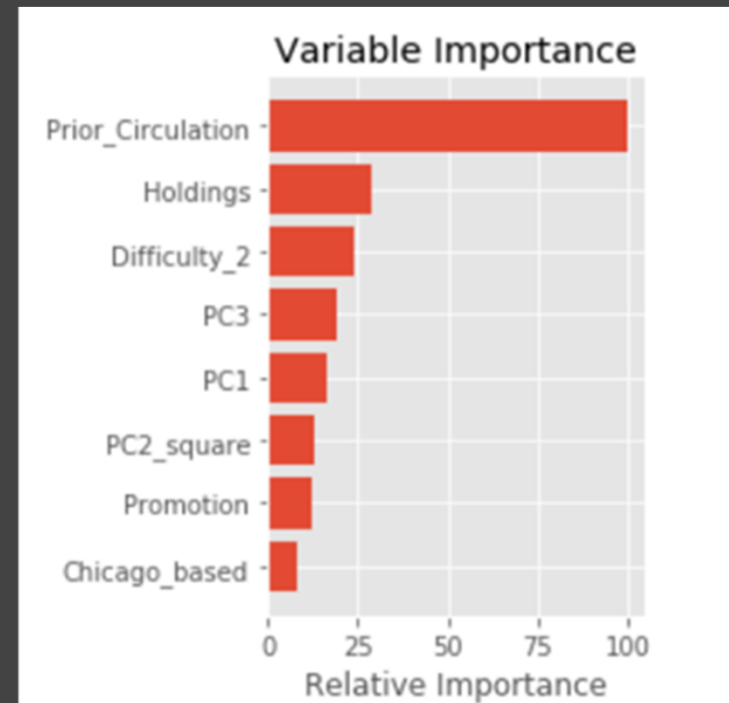
*Augie March (fall 2011)*



# PREDICTIVE MODEL

With 8 features

Able to predict branch  
circulation within  
0.1%



# OUTREACH

Open-source software / dashboard

Presentations and papers

Workshops for librarians