

# IPLC's Collection Analysis with Gold Rush

**Galadriel Chilton**  
Ivy Plus Libraries  
Confederation



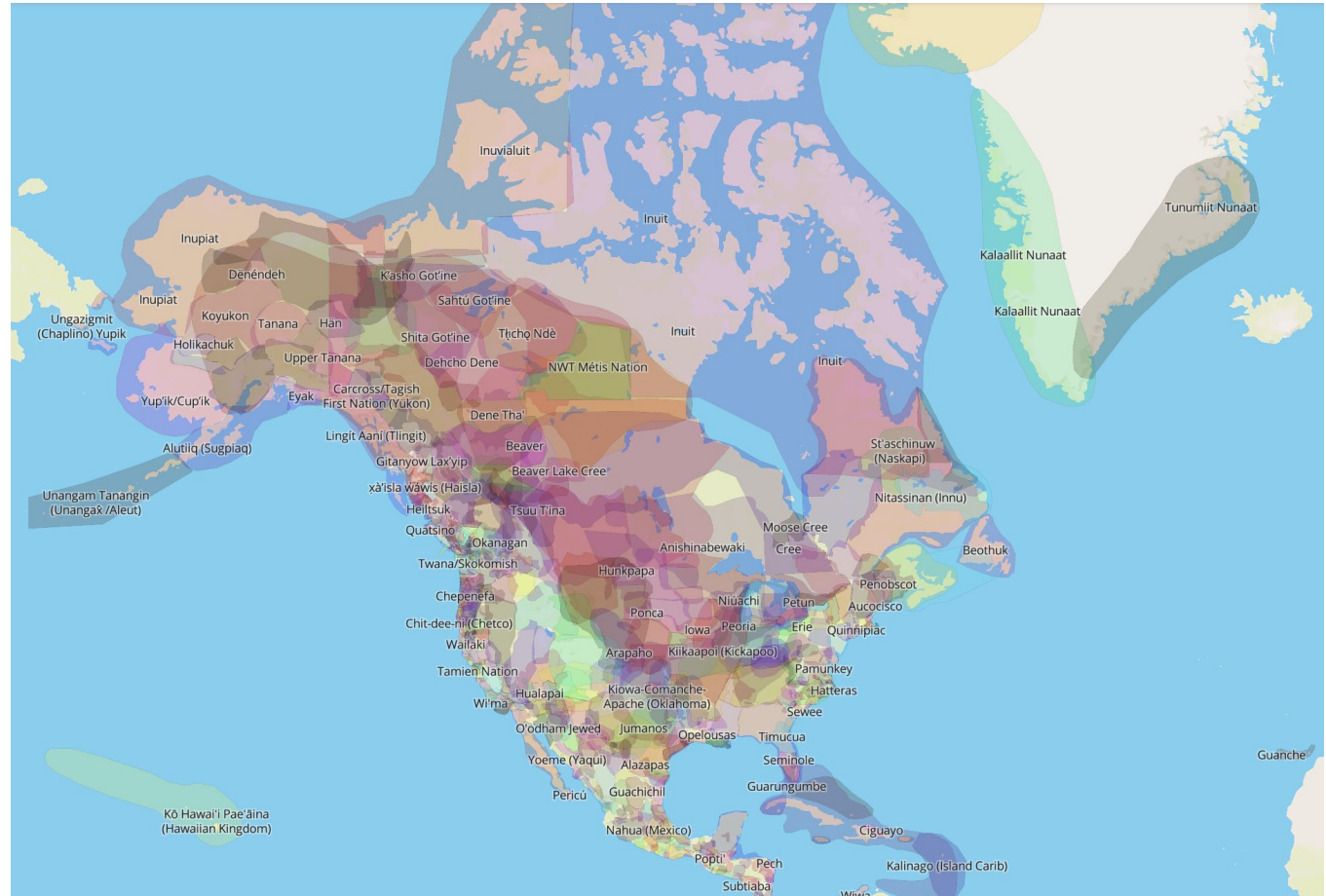
**Jeanette Norris**  
Yale University Library



**Sarah Tudesco**  
Yale University Library

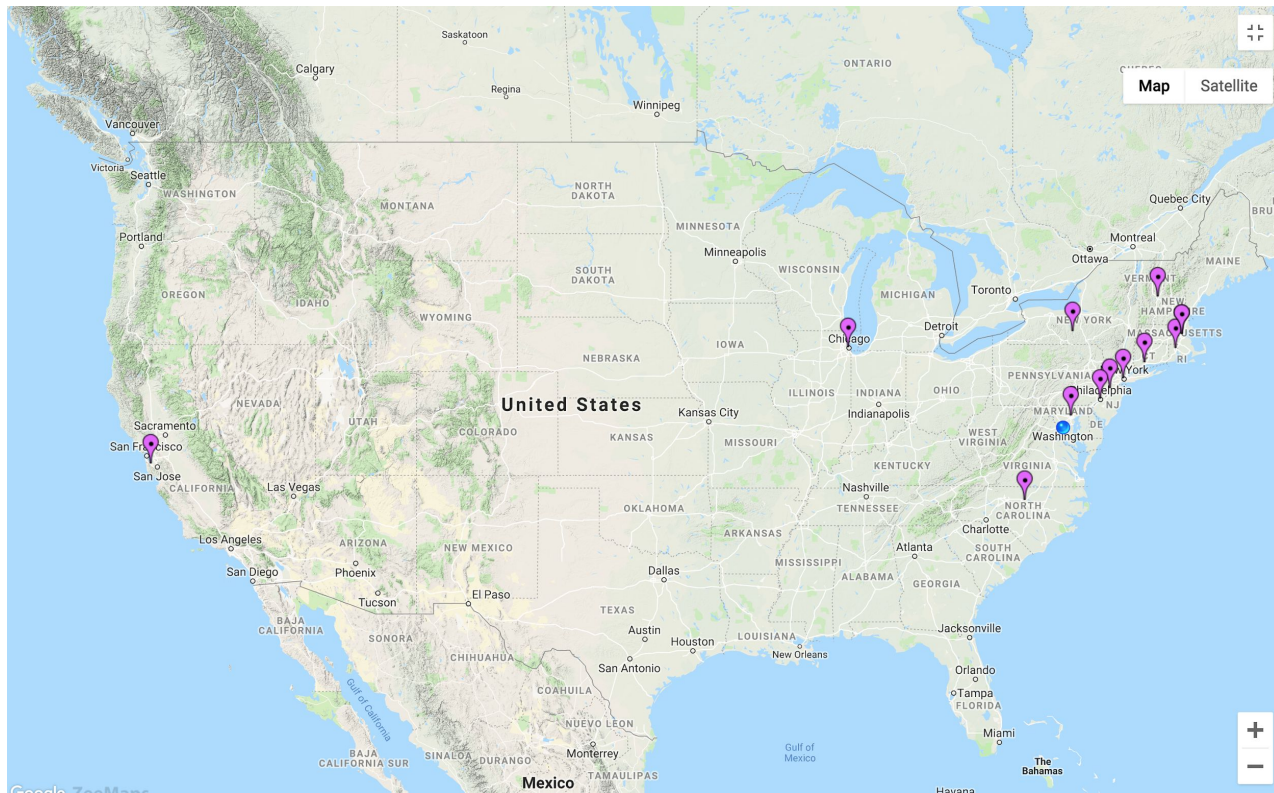


Before proceeding further, we would like to pause for a moment to acknowledge and honor the Native Americans and all of the Indigenous peoples of the land upon which we are all gathered today.



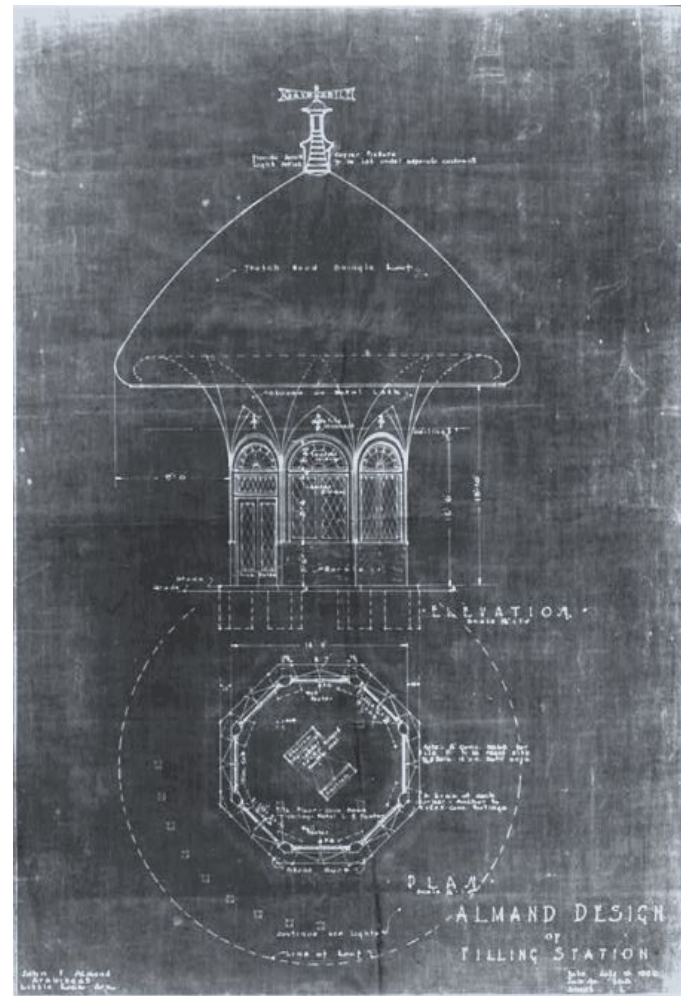


Brown University  
Columbia University  
Cornell University  
Dartmouth College  
Duke University  
Harvard University  
Massachusetts Institute  
of Technology (MIT)  
University of Pennsylvania  
Princeton University  
Stanford University  
The University of Chicago  
The Johns Hopkins University  
Yale University



# Part 1.

## Two Projects

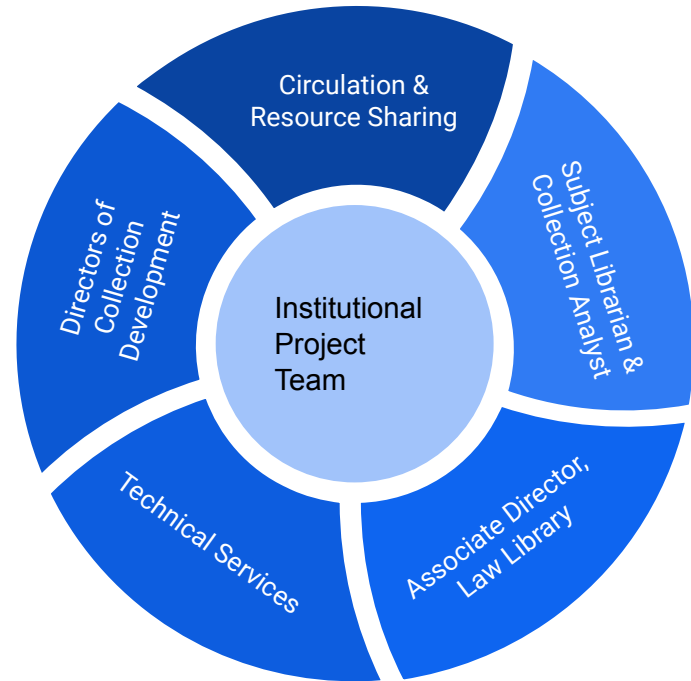


# 1. Collection Analysis Dataset Feasibility Study

(October 2018 - April 2019)

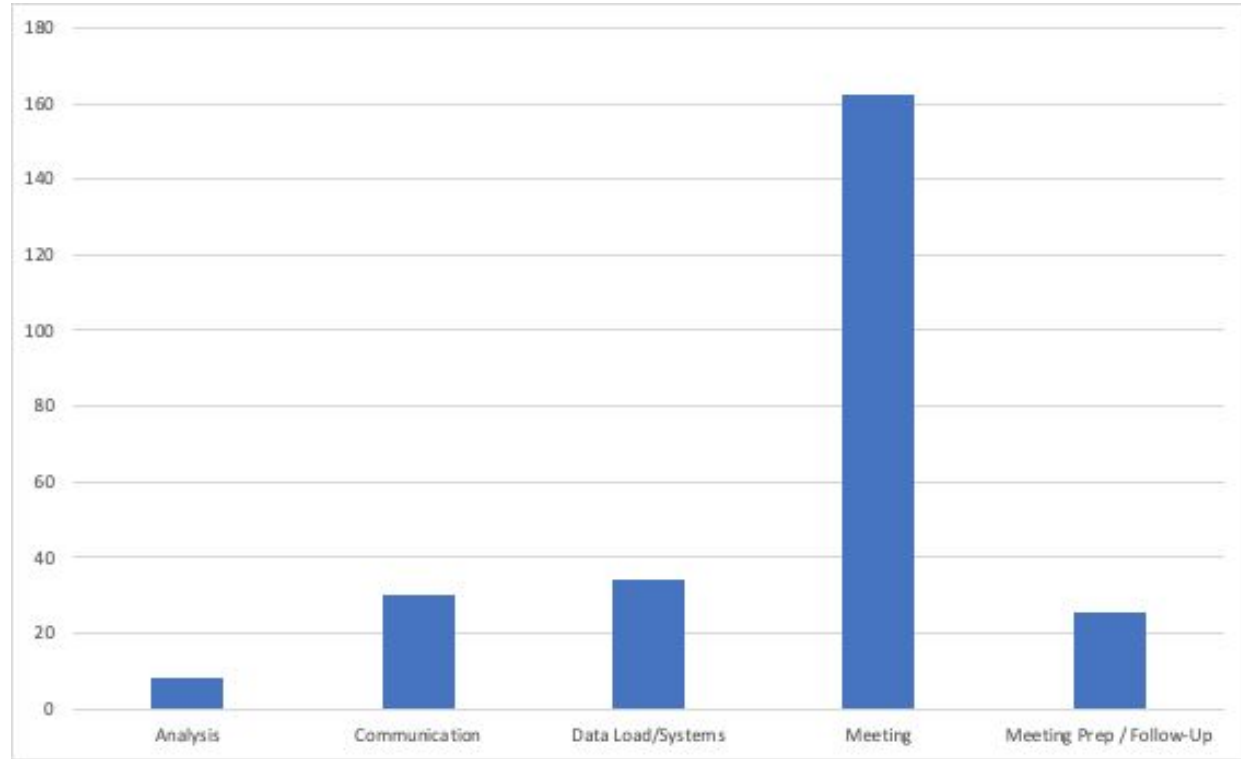
## Project Goals:

- ▣ Test methods of collaboration
- ▣ Document time and resources needed
- ▣ Explore:
  - ▣ How messy is the data?
  - ▣ Is Gold Rush's Library Content Comparison Tool a viable option for IPLC's collection analysis goals?



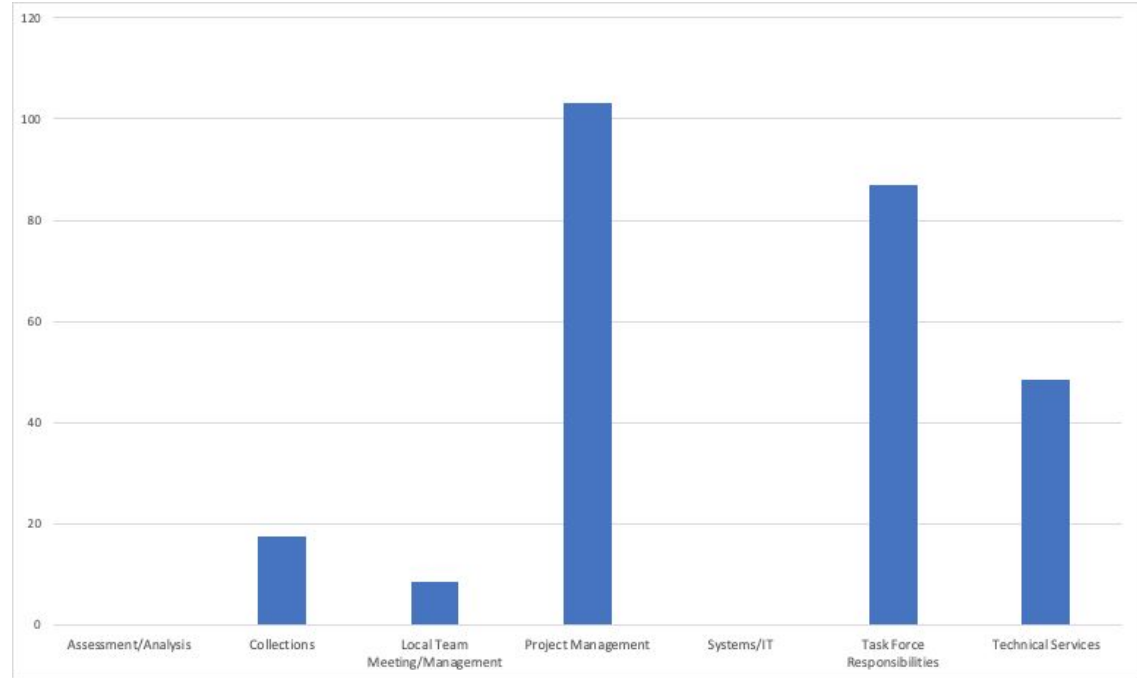
## Outcome:

Collaboration requires work, synchronous communication, and learning together.



*Project Team Time in Hours by Category*

**Outcome:**  
Collaboration  
requires time  
and project  
management.

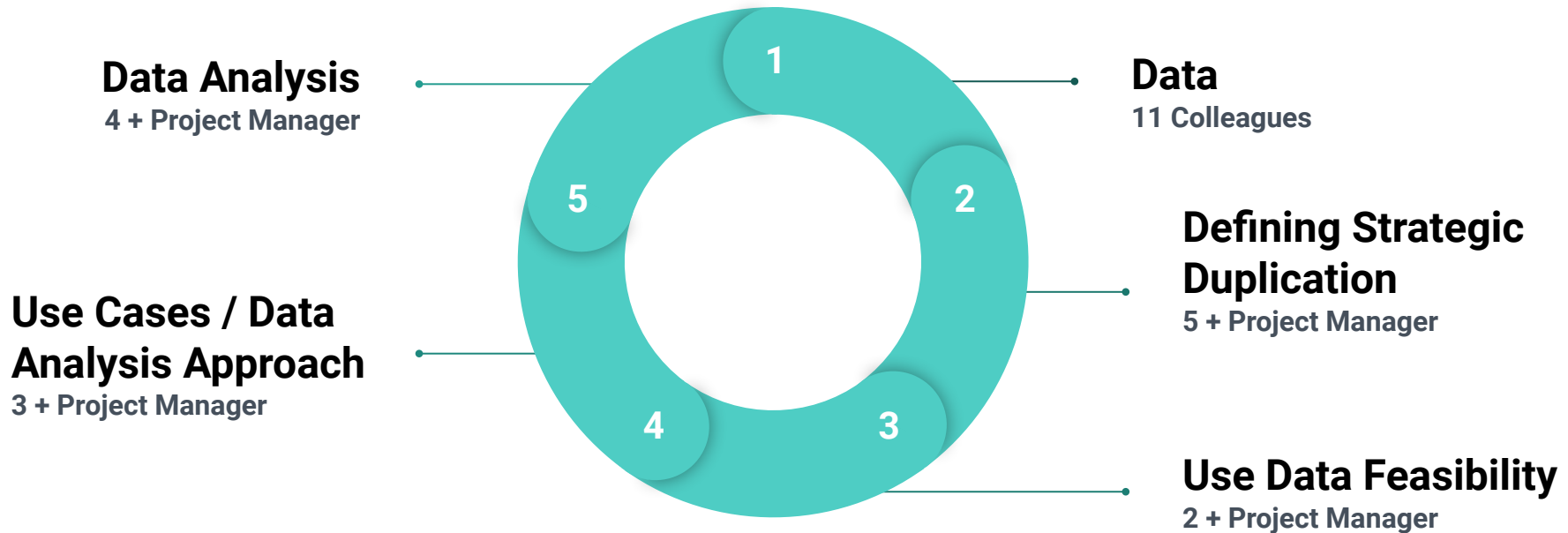


*Project Team Time in Hours by Functional Area.*



## 2. Analysis for Collaborative Collection Development

5-Part Project, 26 People (October 2019 - June 2020)





# Part 2.

## Gold Rush & Data Analysis



# Data Analysis Deliverables

---

## Project 1: Feasibility Study

1. Evaluation of data analysis methods internal & external to Gold Rush
2. Analysis of Gold Rush Match Key effectiveness for identifying duplicates
3. Recommendations for future analysis

## Project 2: Collaborative Collection Development

1. Datasets - normalized and enhanced
2. Tableau Dashboard
3. R Scripts
4. Advice on when to go back to GoldRush
5. Recommendations for future analysis

# Data Analysis Group (Project 2)

---

Our group was charged to develop a methodology to analyze the data and answer the questions outlined by the use cases.

## Goals:

- Evaluate GoldRush Library Content Comparison Tool
- Develop methodology for analyzing holdings vs. usage
- Provide information to inform a collaborative collection development pilot

## Process:

- Meet regularly
- Work collaboratively
- Utilize diverse tools: R, Tableau, MarcEdit, and OpenRefine
- Normalize and enhance data to make them more useful

# Data Overview

---

## Library Holdings

### Gold Rush Content Comparison Tool

- ▣ Single part print monographs
- ▣ Publication date between 2013-2017
- ▣ Holdings data from 13 institutions
- ▣ Numbers...
  - 3,921,524 records
  - ~1,678,579 distinct titles

## Use Transactions

### BorrowDirect (Metridoc System)

- ▣ Filled and unfilled requests
- ▣ Transactions from 2012-2019
- ▣ Requests from 13 institutions\*
- ▣ Numbers...
  - 1,872,741 records
  - ~844,506 distinct titles

*\*Data from when institutions joined BD program.*

# Comparison Dataset

---

We could not compare at title level, so we looked at data points that are common in both datasets.

1. Top Publishers
2. LC Call Number (Class & Subclass)
3. Publication Date 2013-2017

What do they have in common? How do they differ?

Demonstration dashboard: Top publishers with over 10,000 holdings → 18 publishers in the dashboard.

# Interactive Data Dashboard

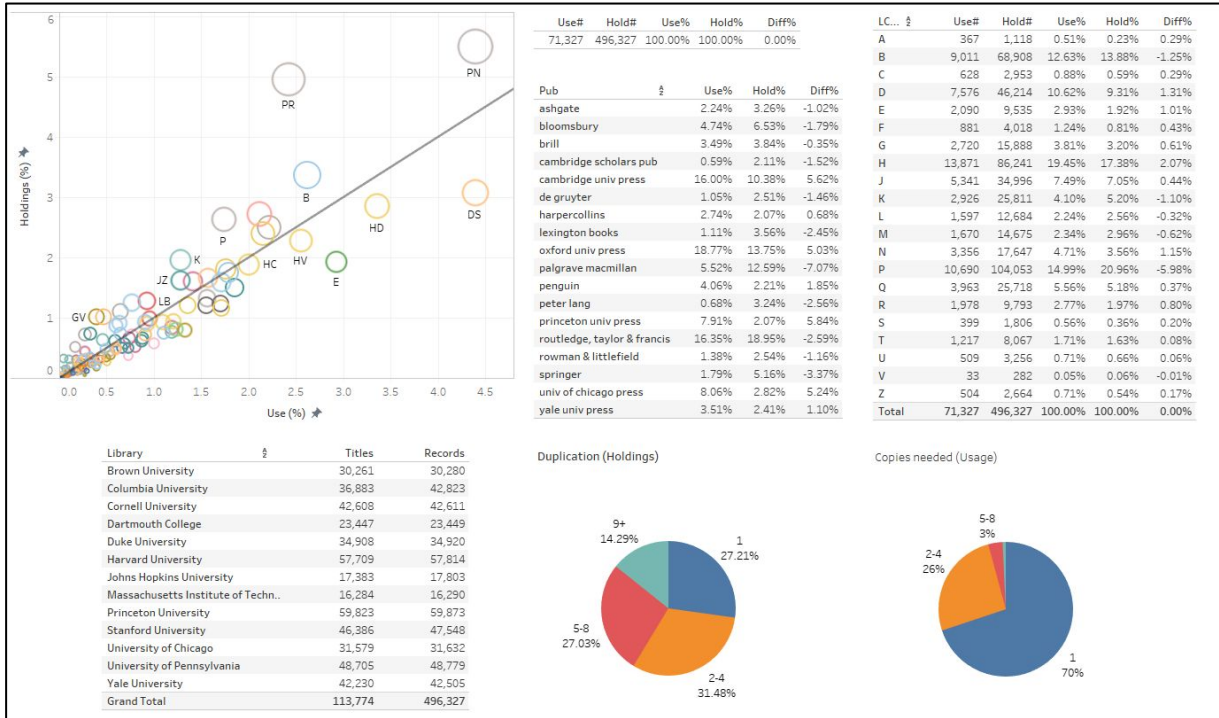


Tableau Public Link:  
<http://bit.ly/IPLC-ERL>

# Ivy Plus Collaborative Collections Data Analysis



Pub	Rank	Use%	Hold%	Diff%
ashgate		2.24%	3.26%	-1.02%
bloomsbury		4.74%	6.53%	-1.79%
brill		3.49%	3.84%	-0.35%
cambridge scholars pub		0.59%	2.11%	-1.52%
cambridge univ press		16.00%	10.38%	5.62%
de gruyter		1.05%	2.51%	-1.46%
harpercollins		2.74%	2.07%	0.68%
lexington books		1.11%	3.56%	-2.45%
oxford univ press		18.77%	13.75%	5.03%
palgrave macmillan		5.52%	12.59%	-7.07%
penguin		4.06%	2.21%	1.85%
peter lang		0.68%	3.24%	-2.56%
princeton univ press		7.91%	2.07%	5.84%
routledge, taylor & francis		16.35%	18.95%	-2.59%
rowman & littlefield		1.38%	2.54%	-1.16%
springer		1.79%	5.16%	-3.37%
univ of chicago press		8.06%	2.82%	5.24%
yale univ press		3.51%	2.41%	1.10%

The dataset focuses on top publishers.

Use = Borrow Direct transactions (usage data)

Hold = GoldRush holdings data (single-part, monographs, published between 2012-2017)

Use%: percentage of transactions in Borrow Direct. (e.g. Princeton Univ Press represents 7.91% of the Borrow Direct transactions in this dataset).

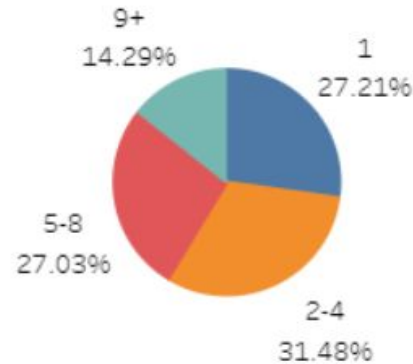
Hold%: percentage of holdings in GoldRush. (Princeton Univ Press represents 2.07% of records in this dataset).

Diff%: Formula  $[Use\%] - [Hold\%]$ . If the value is positive, demonstrates more usage relative to holdings. If the value is negative, demonstrates more holdings relative to usage.



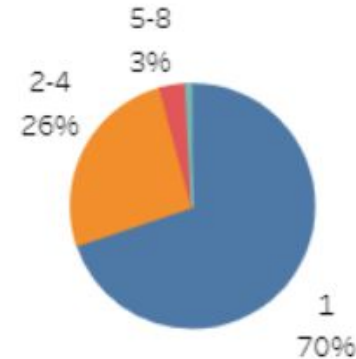
# Duplication of Holdings vs. Copies Needed

Duplication (Holdings)

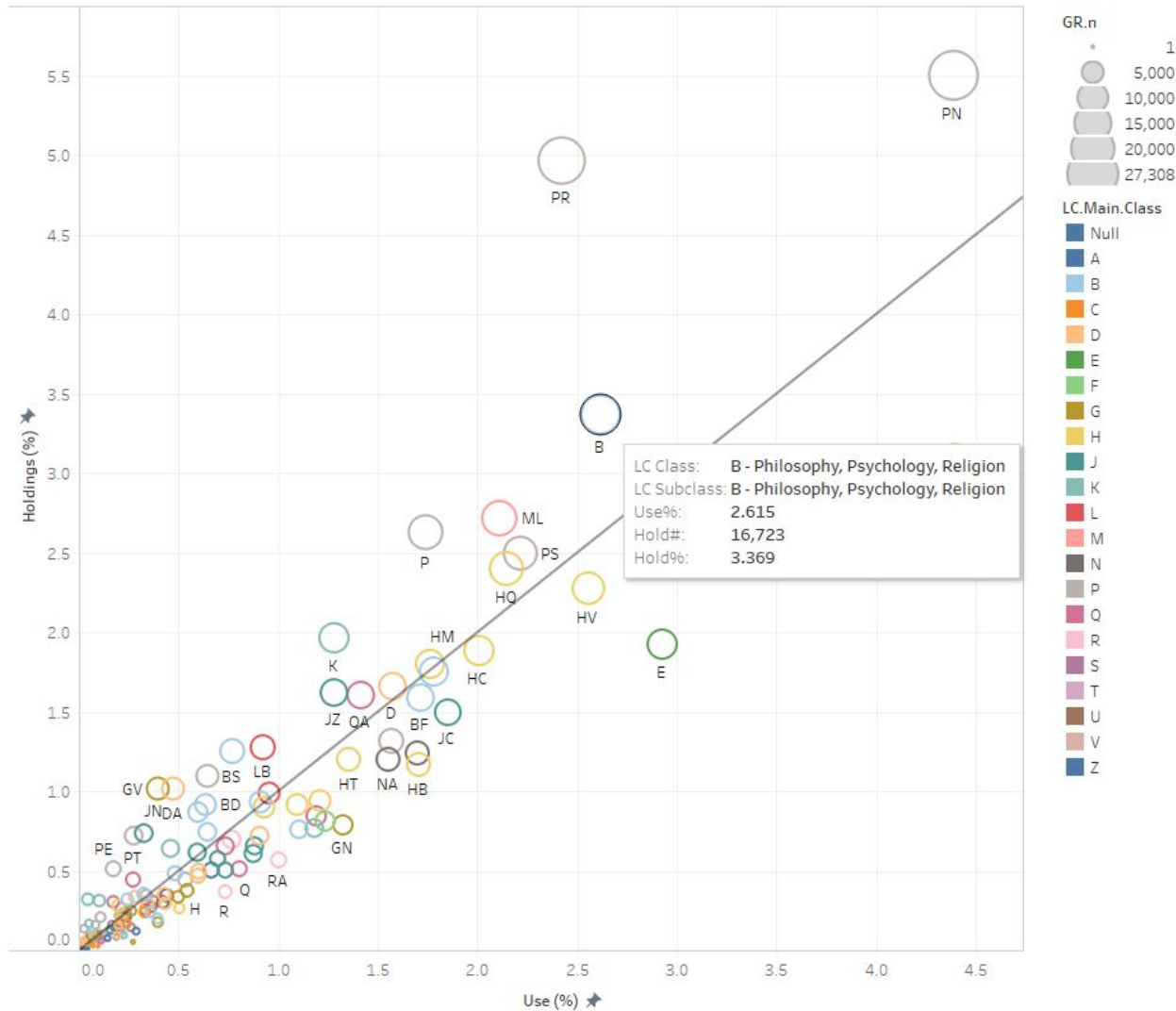


**Duplication (Holdings)**  
chart shows holdings  
overlap between 13 Ivy  
Plus libraries

Copies needed (Usage)



**Copies Needed (Usage)** chart shows how  
many copies were needed to fill demand  
from our resource sharing transactions  
dataset



## Holdings (%) v. Use (%)

Key:

1. Color = LC Class
2. Size of Bubble = #Holdings (larger bubbles = more titles in dataset)

# Data Analysis: Defining and Evaluating Duplicates

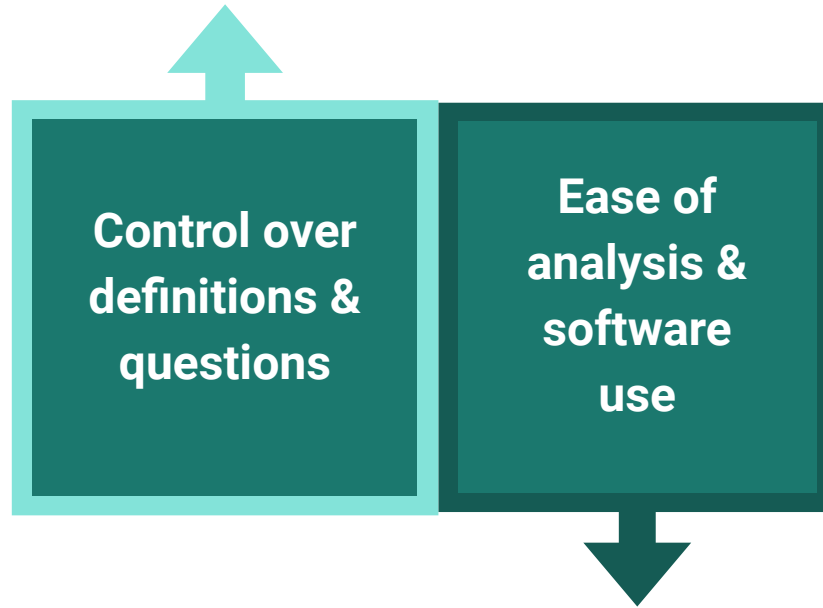
FDR on his houseboat : ^ the Larooco log, 1924-1926 /		9781438462271	State University of New York Press,	(MH)014830896HVD01-Aleph	E807	2016	English
FDR on his houseboat, 1924-1926 : ^ the Larooco log /		9781438462271	State University of New York Press,	(OCoLC)ocn939911243	E807	2016	English
Failed state : ^ dysfunction and corruption in an American statehouse /	Lachman, Seymour,	9781438465739	State University of New York Press,	(OCoLC)953823604	JK3474.7	2017	English
Failed state : ^ dysfunction and corruption in an American statehouse /	Lachman, Seymour,	9781438465739	Excelsior Editions, an imprint of State University of New York Press,	(OCoLC)953823604	JK3474.7	2017	English
Ernest L. Boyer : ^ hope for today's university /		9781438455655 (hardcover : alk. paper)	State University of New York Press,	(OCoLC)906760525	LA227.4	2015	English
Ernest L. Boyer : ^ hope for today's universities /		9781438455655 (hardcover : alk. paper)	State University of New York Press,	(OCoLC-M)906760525	LA227.4	2015	English
Doing time : ^ temporality, hermeneutics, and contemporary cinema /	Carruthers, Lee,	9781438460857	State University of New York Press,	9897880	PN1995.9.T55	2016	English
Doing time : ^ temporality, hermeneutics, and contemporary cinema /	Carruthers, Lee,	9781438460857 (hardcover : alk. paper)	State University of New York Press,	NHCCYBP	PN1995.9.T55	2015	English

# Data Analysis: Normalizing Publishers

Original Publisher Name	Normalized Publisher
routledge	routledge, taylor & francis
routledge, taylor & francis group	
crc press, taylor & francis group	
routledge, an imprint of the taylor & francis group	
taylor & francis	
routledge is an imprint of the taylor & francis group, an informa business	
crc press/taylor & francis	
routledge, taylor & francis	
sage	sage
sage publications	
corwin, a sage company	
sage publications india pvt ltd	
sagep editori	
sage/cq press	
springer	springer
springer vs	
springer verlag	
this palgrave macmillan imprint is published by springer nature	
springer international publishing	
springer publishing company	
palgrave macmillan imprint is published by springer nature	
springer gabler	
springer nature	university of nebraska press
university of nebraska press	
potomac books, an imprint of the university of nebraska press	wiley
john wiley & sons, inc.	
wiley	
wiley blackwell	
wiley-vch	
wiley-vch verlag gmbh & co. kgaa	
jossey-bass, a wiley brand	

# Data Analysis: Current Tradeoffs

---



# Data Analysis: Takeaways

---

- Data normalization and analysis are tricky
  - Develop methods to keep track of your work
  - Document your code
  - Determine when it is 'good enough'
- Diverse skill sets and tools needed to interrogate this data effectively
- Copy analysis exercise demonstrated that 4 or fewer copies meets 97.8% of the distinct titles needed for Borrow Direct.

# Part 3.

## Application



Library of Congress, Prints & Photographs  
Division, Detroit Publishing Company Collection



# In Progress...

---

DEI Definitions for  
Collaborative Collections



Coordinated Print  
Approval Plan



# Potential...

---



- Expand the data in Gold Rush to enable:
  - Shared Print Retention Commitment Management & Analysis
  - Data source to identify commonly held and rare holdings within IPLC
  - Identifying digitized collections across IPLC to inform digitization workflows
- Use the data to help inform strategic acquisition of e-books at local institutions.

# Thanks!

## Any questions?

Galadriel Chilton  
galadriel.chilton@yale.edu

Jeanette Norris  
jeanette.norris@yale.edu

Sarah Tudesco  
sarah.tudesco@yale.edu