

25 May 2021

Visually Exploring Current Events Patterns in WorldCat Records and Holdings Data

DevConnect 2021

Visually exploring current events + social justice patterns in WorldCat data

Alexandria Kalika

Senior Software Engineer

DATA

Purpose

How trends in holdings data reflect best seller book trends

Helping libraries serve their communities by highlighting trends in popular/relevant materials

Using different visualization techniques to spot patterns

Current Events and Books

Berkeley Library NEWS
UNIVERSITY OF CALIFORNIA

[More News](#) [Library Stories](#) [Li](#)

Black Lives Matter: Librarians share books, films that shine light on race, injustice

[Boston](#)

Boston Public Library spent \$75K to increase anti-racism books, libraries across Massachusetts are looking to do the same

Updated Nov 16, 2020. Posted Nov 16, 2020

COMMUNITY | GOVERNMENT

Rural library reckons with racial tensions

By  Tabitha Mueller

September 6th, 2020 - 2:00am

RACE

How Libraries Are Supporting The Black Lives Matter Movement

June 13, 2020 · 7:59 AM ET

Heard on [Weekend Edition Saturday](#)

People Are Marching Against Racism. They're Also Reading About It.

Books on the subject have soared up best-seller lists as protests continue across the country.

Business

Demand for anti-racist literature is up. These black bookstore owners hope it lasts.

Library and book sales data show how interest in anti-racist and social justice titles exploded after George Floyd's death

Popular Authors

Michelle Alexander
Ijeoma Oluo
Layla F. Saad
Ibram X Kendi
Te-Nehisi Coates

NY Times Best Seller lists

1
↑
8 WEEKS ON THE LIST
HOW TO BE AN ANTIRACIST
by Ibram X. Kendi | One World
A primer for creating a more just and equitable society through identifying and opposing racism.

4
↑
3 WEEKS ON THE LIST
ME AND WHITE SUPREMACY
by Layla F. Saad | Sourcebooks
Ways to understand and possibly counteract white privilege.

5
↑
3 WEEKS ON THE LIST
THE NEW JIM CROW
by Michelle Alexander | New Press
A law professor takes aim at the “war on drugs,” mass incarceration and their impact on black men.

6
↑
67 WEEKS ON THE LIST
BETWEEN THE WORLD AND ME
by Ta-Nehisi Coates | Spiegel & Grau
Winner of the 2015 National Book Award for nonfiction. A meditation on race in America as well as a personal story, framed as a letter to the author's teenage son.

2
↑
3 WEEKS ON THE LIST
SO YOU WANT TO TALK ABOUT RACE
by Ijeoma Oluo | Seal
A look at the contemporary racial landscape of the United States.

9
NEW THIS WEEK
STAMPED FROM THE BEGINNING
by Ibram X. Kendi | Bold Type
Winner of the 2016 National Book Award for nonfiction. A look at anti-black racist ideas and their effect on the course of American history.

Transforming Data for Analysis

Data starts in XML tags, moved to structured tables, easier to manipulate, use and examine

Multiple transformation along the way – data cleaning, etc.



Data used

- Subject headings in subject field 650– “Racism”, “Black Lives Matter”, “Police Brutality”
- Selected popular writers from best seller lists and news
- Popular titles, types of libraries, material types

All numbers are provisional – this is a proof of concept

Subject heading fields

010			2020034614	
040			DLC t eng t e rda t c DLC 1 d OCLCO 1 d OCLCF 1 d YDX 1 d OCLCO	
020			9780691172361 \$q hardcover \$q alk, paperback	
020			0691172366 \$q hardcover \$q alk, paperback	
020			\$z 9780691200330 \$q electronic book	
042			pcc	
050			AZ186 \$b .K37 2021	
082			001.3072/1 \$2 23	
090			\$b	
049			OCLC	
100			Karsdorp, Folgert, t e author.	
245			Humanities data analysis : \$b case studies with Python / \$c Folgert Karsdorp, Mike Kestemont & Allen Riddell.	
264			Princeton, New Jersey : \$b Princeton University Press, \$c [2021]	
300			xi, 337 pages : \$b illustrations (some color), maps (some color) ; \$c 27 cm	
336			text \$b txt \$2 rdacontent	
337			unmediated \$b n \$2 rdamedia	
338			volume \$b nc \$2 rdacarrier	
504			Includes bibliographical references and index.	
520			"The use of quantitative methods in the humanities and related social sciences has increased considerably in recent years, allowing researchers to discover patterns in a vast range of source materials. Despite this growth, there are few resources addressed to students and scholars who wish to take advantage of these powerful tools. Humanities Data Analysis offers the first intermediate-level guide to quantitative data analysis for	
650			0 Humanities t x Research t x Methodology.	
650			0 Social sciences t x Research t x Methodology.	
650			0 Quantitative research t x Data processing.	
650			0 Python (Computer program language)	
650			7 Humanities \$x Research \$x Methodology. \$2 fast \$0 (OCoLC)fst00963624	
650			7 Python (Computer program language) \$2 fast \$0 (OCoLC)fst01084736	
650			7 Social sciences \$x Research \$x Methodology. \$2 fast \$0 (OCoLC)fst01122961	
700			1 Kestemont, Mike, 1 d 1985-; t e author.	
700			1 Riddell, Allen, t e author.	
938			YBP Library Services \$b YANK \$n 16733110	
938			Brodart \$b BROD \$n 126983887	
029			1 DKDLA \$b 800010-katalog:99123538354105763	

650		0	Humanities t x Research t x Methodology.
650		0	Social sciences t x Research t x Methodology.
650		0	Quantitative research t x Data processing.
650		0	Python (Computer program language)
650		7	Humanities \$x Research \$x Methodology. \$2 fast \$0 (OCoLC)fst00963624
650		7	Python (Computer program language) \$2 fast \$0 (OCoLC)fst01084736
650		7	Social sciences \$x Research \$x Methodology. \$2 fast \$0 (OCoLC)fst01122961

Subject heading fields

040		LMJ 1 eng 1 e rda 1 c LMJ 1 d OCLCO 1 d MDR 1 d BDX 1 d SMDRL 1 d INR 1 d OCLCF 1 d ZP7 1 d MJS 1 d OCLCO 1 d JTH 1 d OCL 1 d OCLCO
020		9780593396803 \$q (pbk. ; \$q lg. print)
020		0593396804 \$q (pbk. ; \$q lg. print)
043		n-us---
050	4	E184 A1 \$b K344 2020
082	0 4	305.800973 \$2 23
090		\$b
049		OCLC
100	1	Kendi, Ibram X., 1 e author.
245	1 0	How to be an antiracist / \$c Ibram X. Kendi.
246	3	How to be an anti-racist
250		First large print edition.
264	1	New York : \$b Random House Large Print, \$c [2020]
264	4	\$c ©2019
300		viii, 610 pages (large print) ; \$c 21 cm
336		text \$2 rdacontent
337		unmediated \$2 rdamedia
338		volume \$2 rdacarrier
340		\$n large type. \$2 rda
520		"Kendi takes readers through a widening circle of antiracist ideas--from the most basic concepts to visionary possibilities--that will help readers see all forms of racism clearly, understand their poisonous consequences, and work to oppose them in our systems and in ourselves. [He] weaves [a] ... combination of ethics, history, law, and science with his own personal story of awakening to antiracism. This is an essential work for anyone who wants to go beyond the
504		Includes bibliographical references and index.
505	0	Definitions -- Dueling consciousness -- Pover -- Biology -- Ethnicity -- Body -- Culture -- Behavior -- Color -- White -- Black -- Class -- Space -- Gender -- Sexuality -- Failure -- Success -- Survival.
650	0	Anti-racism 1z United States.
650	0	Racism 1x Psychological aspects.
650	0	Large type books.
651	0	United States 1x Race relations.
600	10	Kendi, Ibram X.
650	7	Anti-racism \$2 fast \$0 (OCoLC)fst01742067
650	7	Large type books \$2 fast \$0 (OCoLC)fst00992678
650	7	Race relations \$2 fast \$0 (OCoLC)fst01086509
650	7	Racism \$x Psychological aspects \$2 fast \$0 (OCoLC)fst01086630
651	7	United States \$2 fast \$0 (OCoLC)fst01204155
655	7	Autobiographies. \$2 fast \$0 (OCoLC)fst01919894
655	0	Large type books.
655	4	Large type books.
655	7	Autobiographies. 1 z lcgrr
938		Brodart \$b BROD \$n 127349510

650	0	Anti-racism 1z United States.
650	0	Racism 1x Psychological aspects.
650	0	Large type books.
651	0	United States 1x Race relations.
600	10	Kendi, Ibram X.
650	7	Anti-racism \$2 fast \$0 (OCoLC)fst01742067
650	7	Large type books \$2 fast \$0 (OCoLC)fst00992678
650	7	Race relations \$2 fast \$0 (OCoLC)fst01086509
650	7	Racism \$x Psychological aspects \$2 fast \$0 (OCoLC)fst01086630

Looking for interesting trends

- Material types – Digital vs. Print
- Library types – Academic, Public, School and Community College
- Authors, titles and holdings – 2019 vs. 2020
- Ethics – historical data, controversial authors, topic authority

MACHINE LEARNING AND DATA

Challenges

- Bibliographic data is not 'traditional' data.
- Cleansing/finding/refining data requires substantial, specialized knowledge of meta data.
- Is there biases in the data?
- Work in progress

Machine learning

- Complexity of data = many factors in the data
- Reducing factors means making decisions on what's "important" which can become a technological /accuracy dilemma
- Not reducing factors = hundreds of columns of data
- Example of classification tree – can show non-linear trends in data

Machine learning and library data

Over representation in data?

Purpose of Machine Learning

Effort of ETL vs. ML benefits

What is an appropriate question for ML vs. other data techniques?

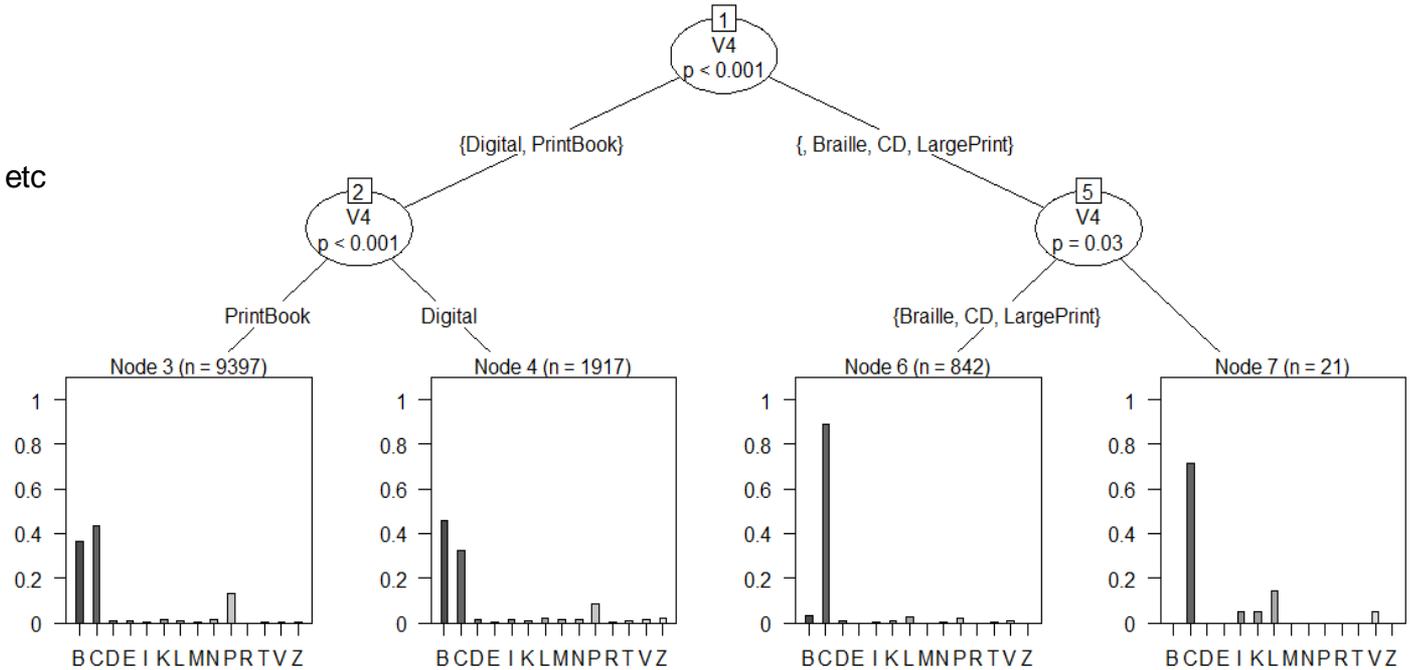


Decision tree example

Library types

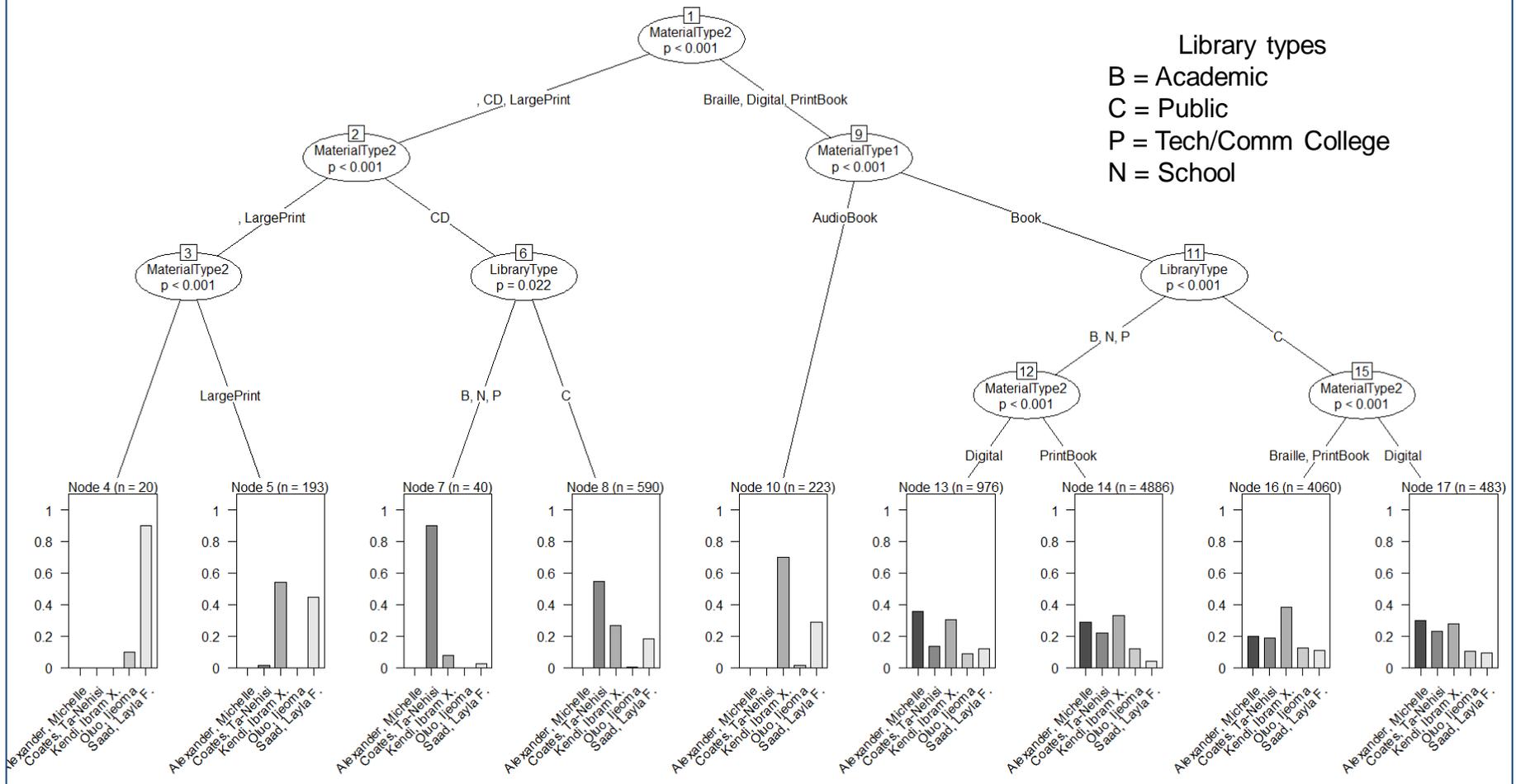
B = Academic
 C = Public
 P = Tech/Comm College etc

library("partykit")



Library types

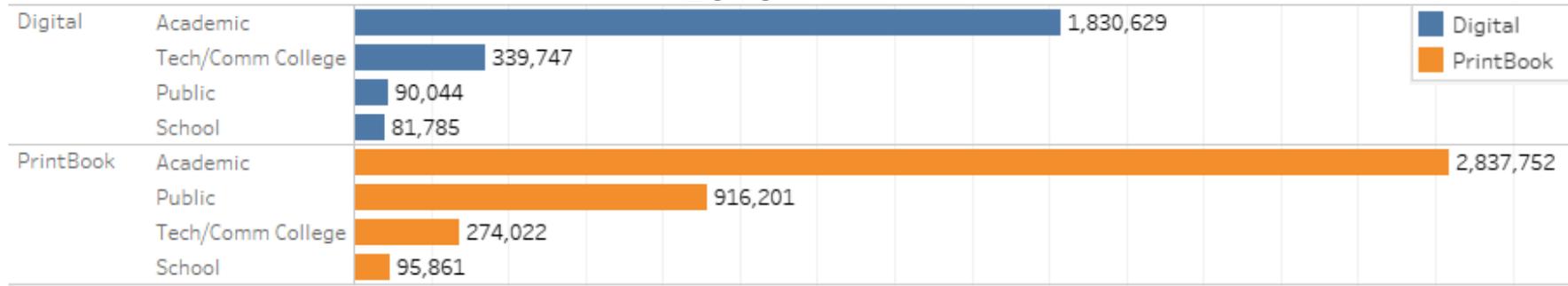
- B = Academic
- C = Public
- P = Tech/Comm College
- N = School



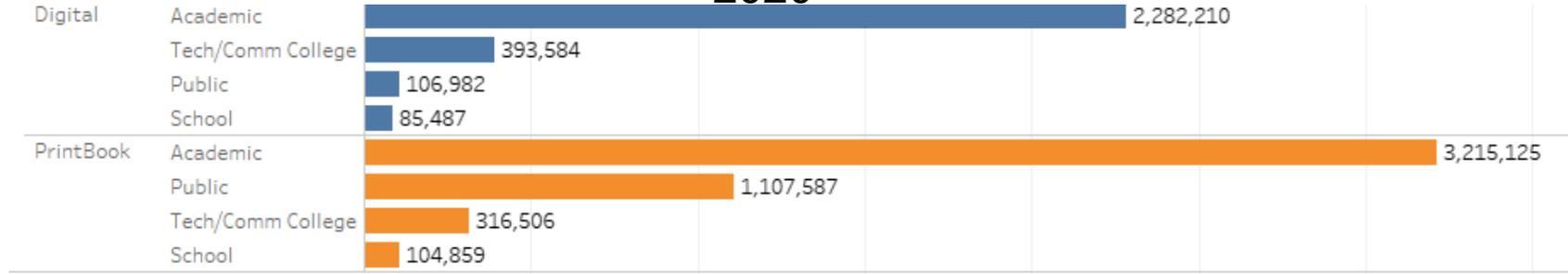
VISUALIZING DATA

Subject heading 'Racism'

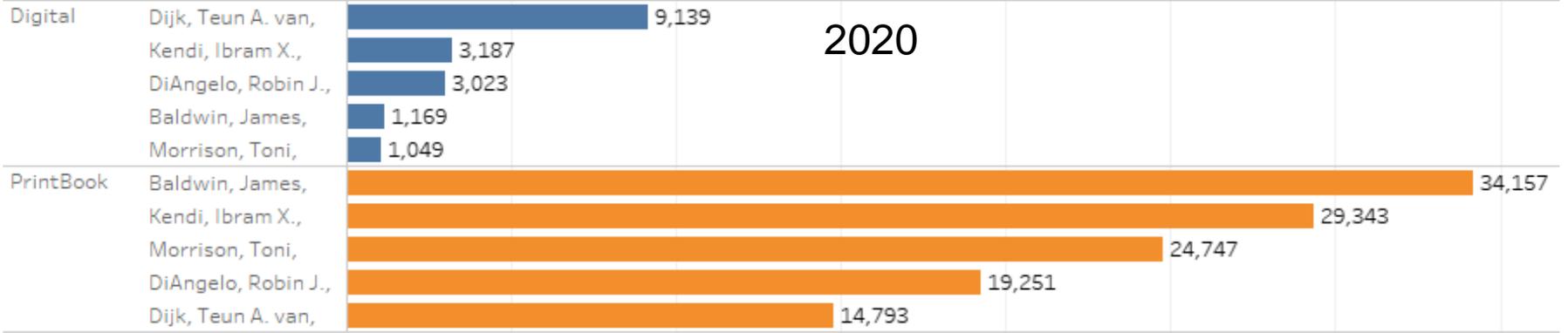
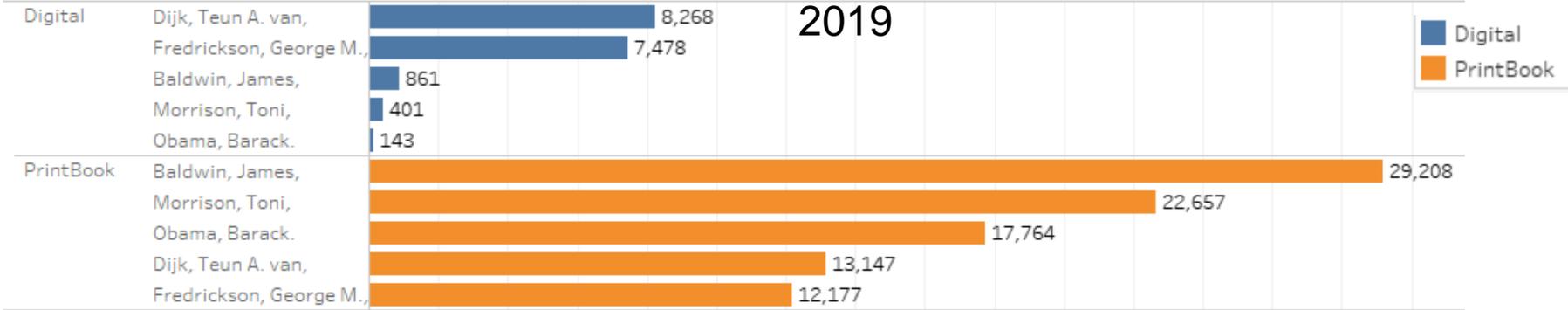
2019



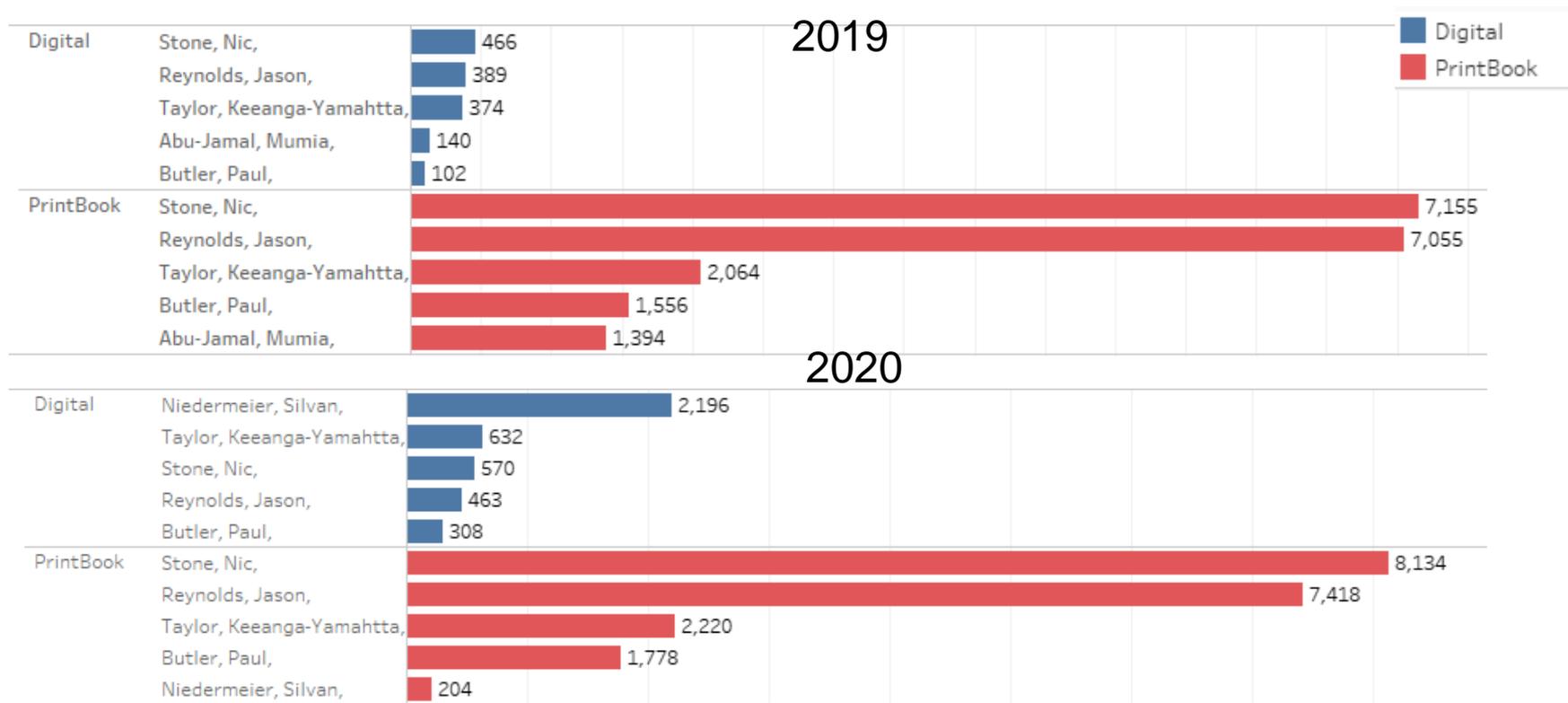
2020



Subject heading 'Racism'

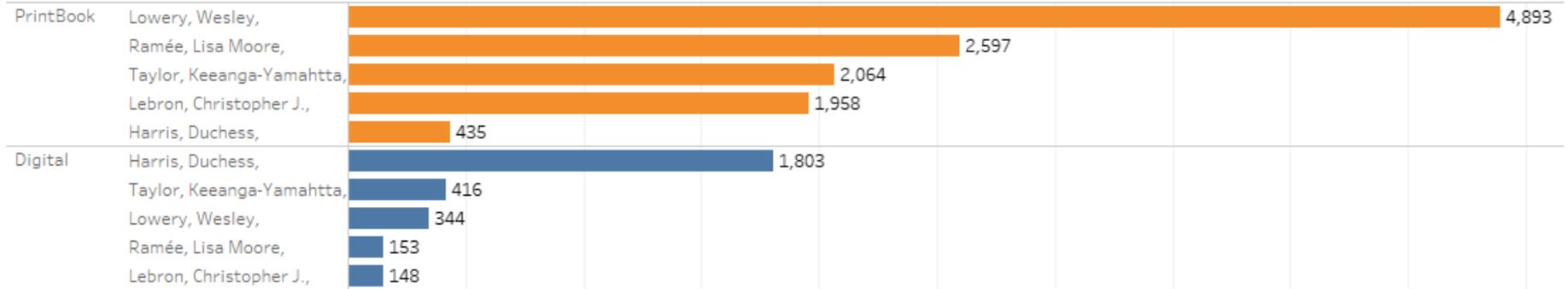


Subject heading 'Police Brutality'

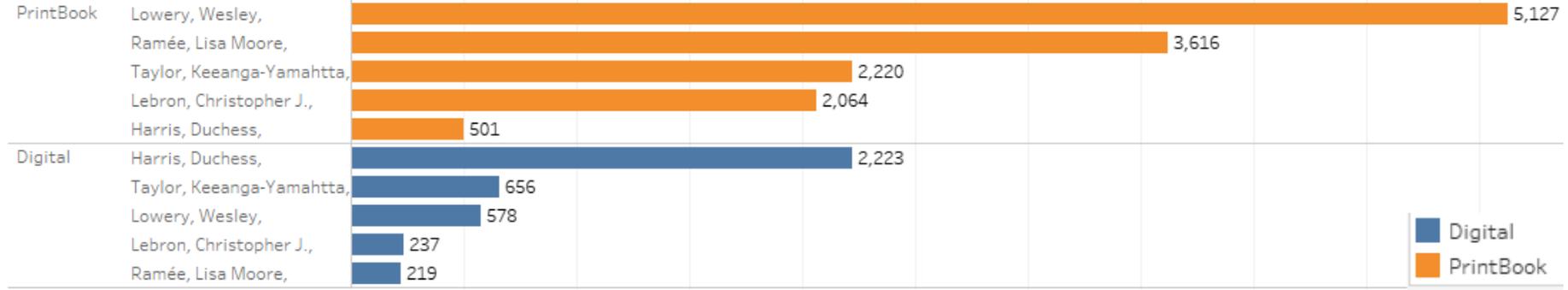


Subject heading 'Black Lives Matter'

2019



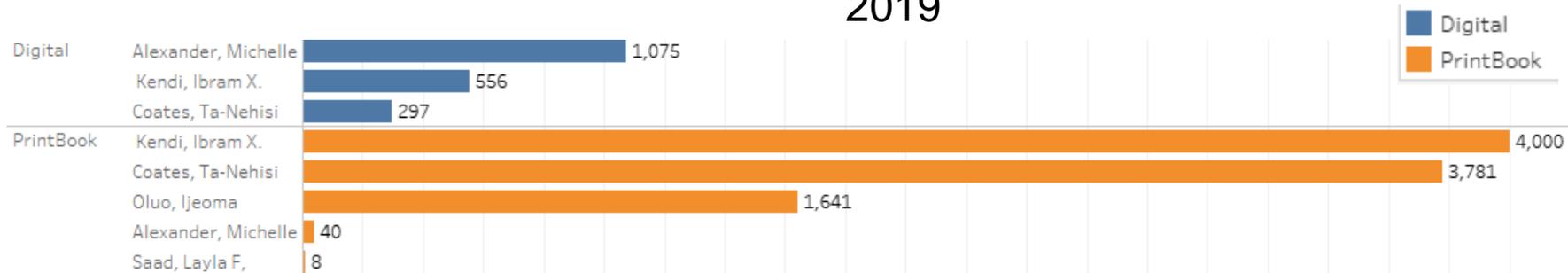
2020



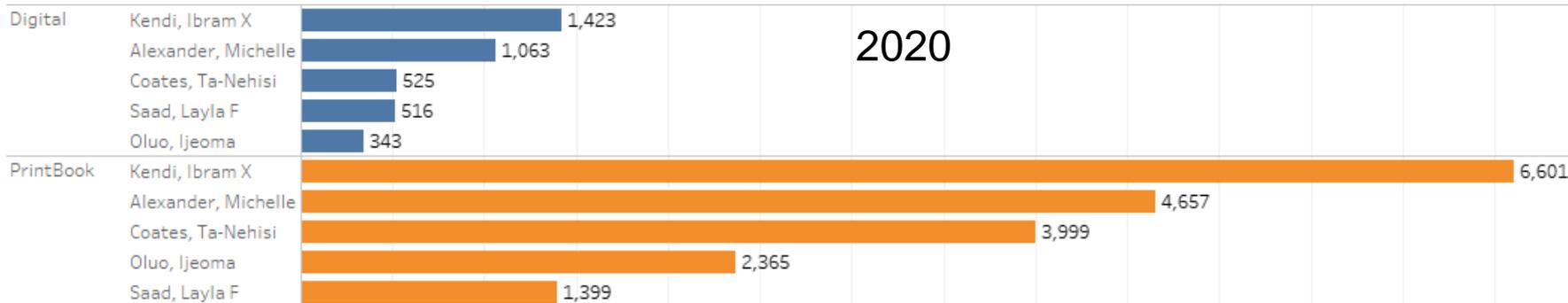
Digital
PrintBook

Popular creators

2019



2020



Holdings 2019 vs. 2020

2019

Materialtype2	Author (group)	School	Tech/Comm College	Public	Academic
Digital	Alexander, Michelle	16	154	202	641
	Kendi, Ibram X.	13	35	276	143
	Coates, Ta-Nehisi	15	6	200	39
PrintBook	Kendi, Ibram X.	57	544	1,602	1,625
	Coates, Ta-Nehisi	95	523	1,475	1,516
	Oluo, Ijeoma	30	226	707	597
	Alexander, Michelle		1	10	23
	Saad, Layla F,			3	1

2020

Materialtype2	Author (group)	School	Comm/Tech College	Academic	Public
Digital	Kendi, Ibram X	26	120	586	500
	Alexander, Michelle	18	115	557	279
	Coates, Ta-Nehisi	15	38	204	218
	Saad, Layla F	6	21	260	178
	Oluo, Ijeoma		25	179	98
PrintBook	Kendi, Ibram X	87	798	2,295	3,074
	Alexander, Michelle	67	704	2,002	1,666
	Coates, Ta-Nehisi	101	553	1,569	1,577
	Oluo, Ijeoma	40	295	853	1,033
	Saad, Layla F	12	117	265	903

Ethics of popularity trends

- Who can classify “popularity”
- What about controversial authors?
- Rise of digital content
- How to objectively show popularity trends

CONCLUSION

There are patterns in holdings data based on subject holdings and authors.

Machine learning has potential to find trends and patterns to help libraries in highlighting in-demand works – if used properly, with deep understanding of data.

Analysis of print vs. digital works shows trends that can reveal accessibility of materials.

FUTURE DATA WORK

Data Opportunities

- More understanding as to where geographically a holding is located
- More refined visual representation of titles based on current events – through best seller lists, etc.
- Seeing change over time in topic interest based on holdings data

Thank you for your time!

Alexandria Kalika

Senior Software Engineer

kalikaa@oclc.org

**Because
what is
known must
be shared.®**