

Making Libraries for
Everyone: *Building a
Reproducible Workflow to
Assess Classification*

Slides: <https://go.uncg.edu/r8jdpz>



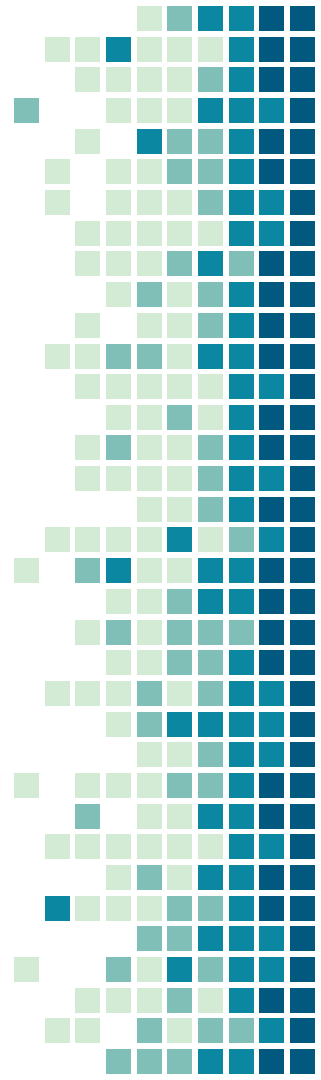
HELLO!


I am Lindsay Gypin she/her

Data Services Librarian
University of North Carolina
Greensboro

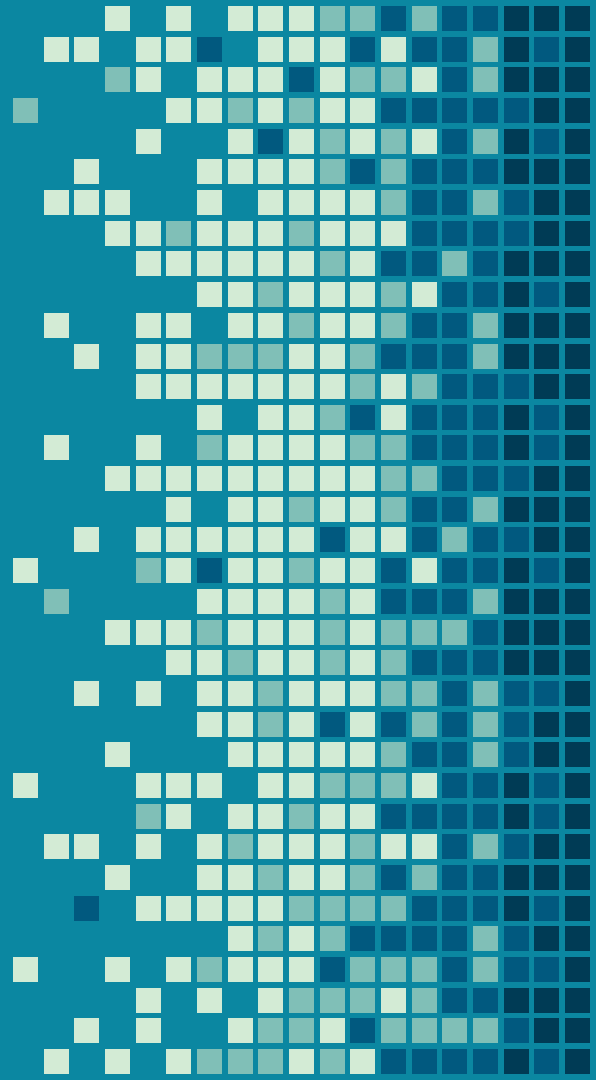
Life goal: end systemic
racism in libraries

Find me: @menacegypin
lagypin@uncg.edu





“ *Frictionless Data is a set of specifications for data and metadata interoperability, accompanied by a collection of software libraries that implement these specifications, and a range of best practices for data management.* ”





Frictionless Data for Reproducible Research Fellows Programme

fellows.frictionless.io





Got Data?

300 __ |a 152 pages : |b illustrations ; |c 20 cm.

336 __ |a text |2 rdacontent.

337 __ |a unmediated |2 rdamedia.

338 __ |a volume |2 rdacarrier.

520 __ |a "For Ta-Nehisi Coates, history has always been personal. At every stage of his life, he's sought in his explorations of history answers to the mysteries that surrounded him -- most urgently, why he, and other black people he knew, seemed to live in fear. What were they afraid of? In Tremble for My Country, Coates takes readers along on his journey through America's history of race and its contemporary resonances through a series of awakenings -- moments when he discovered some new truth about our long, tangled history of race, whether through his myth-busting professors at Howard University, a trip to a Civil War battlefield with a rogue historian, a journey to Chicago's South Side to visit aging survivors of 20th century America's 'long war on black people,' or a visit with the mother of a beloved friend who was shot down by the police. In his trademark style -- a mix of lyrical personal narrative, reimagined history, essayistic argument, and reportage -- Coates provides readers a thrillingly illuminating new framework for understanding race: its history, our contemporary dilemma, and where we go from here"-- |c Provided by publisher.

505 0_ |a Prologue : the talk -- PART 1. BETWEEN THE WORLD AND ME. The changes -- The second change : Malcolm and the body -- The third change : Mecca and the death of mythology -- PART 2. THE SOOTY DETAILS OF THE SCENE. The fourth change : New York and the death of mercy -- The fifth change : Gettysburg and the long war -- The sixth change : Chicago and the streets -- PART 3. A GRASSY CLEARING. The seventh change : eyes open to the world -- The eighth change : the blast -- Epilogue : into the world.

651 _0 |a United States |x Race relations.

650 _0 |a Race discrimination |z United States.

650 _0 |a African Americans |x Social conditions.

650 _0 |a African Americans |x Public opinion.

650 _0 |a White people |z United States |x Attitudes.



Lindsay Gypin ▸ Troublesome Catalogers and Magical
Metadata Fairies



January 7 · 🌐

Hello fabulous cataloguers and metadata experts! I have a question/favor. I am a Data Services Librarian at an Academic Library. I am participating in a Data Science Fellowship (Frictionless Data for Reproducible Research Fellowship) and want to use catalog metadata / MARC records to explore systemic racism from a data science/text mining lens.

I'm wondering... do any of you have suggestions on where to look within a MARC record (I'm thinking subject headings and/or general descriptions) and which subset of records within an academic institution's catalog to examine (I realize this will be different depending on institution. I'm a bit overwhelmed on this aspect.)

THANK YOU in advance!



4

16 Comments



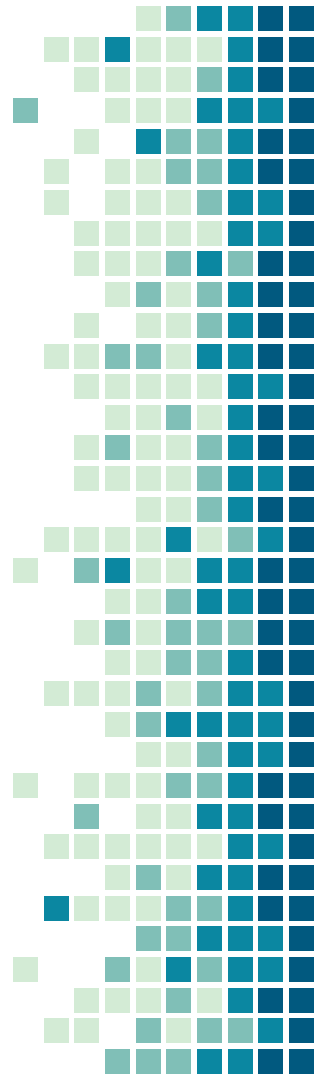
Like



Comment



Share



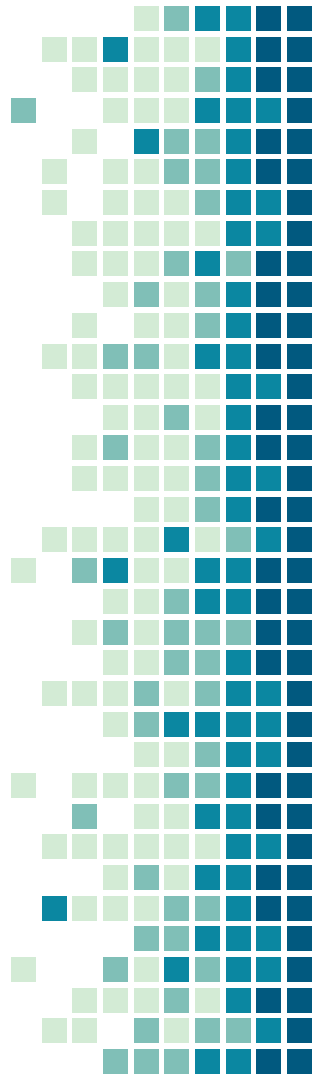


The Data

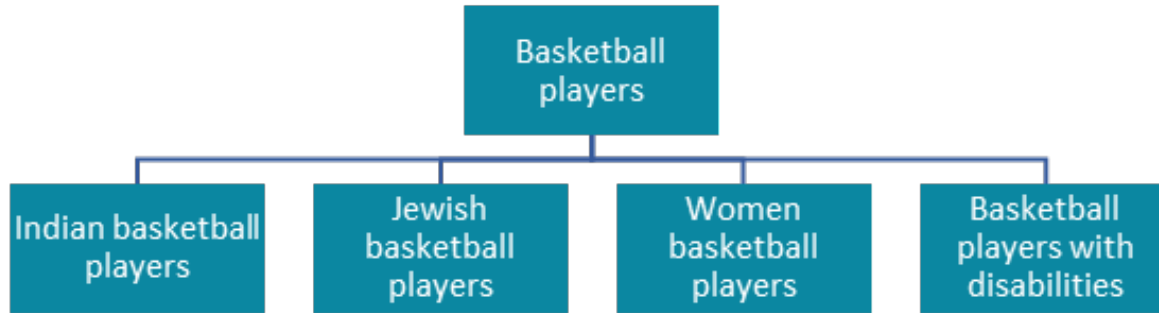
HathiTrust Digital Library - downloadable MARC records do not include these fields

UNCG Library Catalog

- Results from "basketball" keyword search

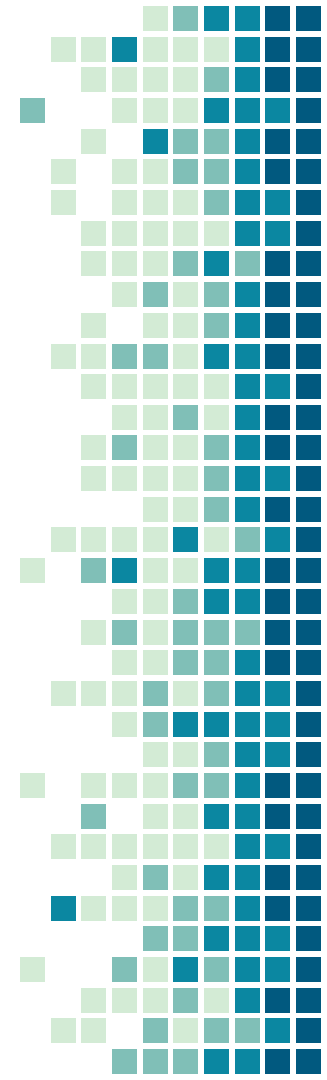


Why “basketball” ?



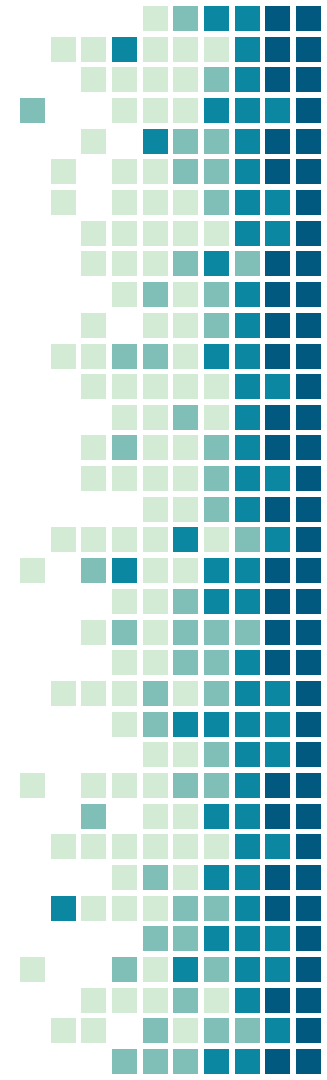
LCSH

The bias hiding in your library by Amanda Ros



Frictionless Tools

- Data Package Creator
 - Data Package = tabular data file + metadata schema JSON file
- Goodtables
 - Continuous validation of data using schema

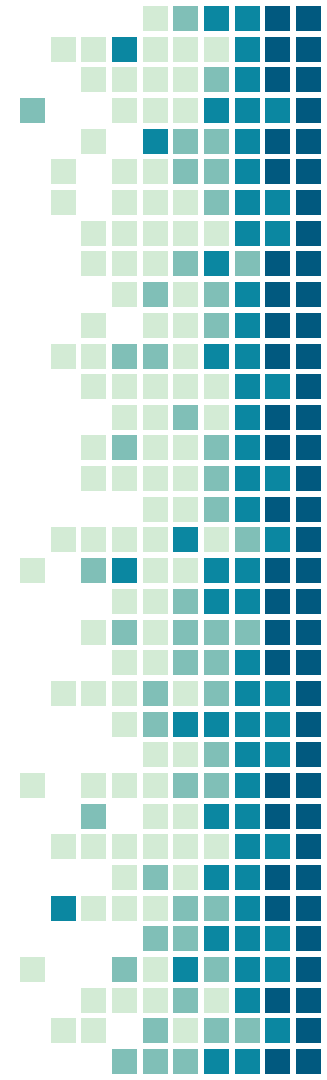




Data Export

- BasketballMarcRecordsRawMarc.mrc
- BasketballMarcRecordsRawMarc.txt
- BasketballMarcRecordsForOpenRefine.json
- BasketballMarcRecordsMarcEditFormat.mrk

Solution: MarcEdit 7.5



100a	245a	245b	500a	505a	650a
Aamidor, Abraham.	Chuck Taylor, Converse all star :	the true story of the man behind the most famous athletic shoe in history /	N/A	Hall of fame -- Non-skids -- Salesman -- The invisible pass -- Special service -- Air-tecs -- World tourney -- "Me" -- Glory.	Basketball players;Basketball players.
Aamidor, Abraham.	Chuck Taylor, Converse all star :	the true story of the man behind the most famous athletic shoe in history /	N/A	Hall of fame -- Non-skids -- Salesman -- The invisible pass -- Special service -- Air-tecs -- World tourney -- "Me" -- Glory.	Basketball players;BIOGRAPHY & AUTOBIOGRAPHY;SPORTS & RECREATION;SPORTS & RECREATION;Basketball players.
Aardahl, Anya.	The influence of height on body image, self-confidence, and performance of female basketball players /	N/A	N/A	N/A	Women basketball players;Stature;Self-perception in women.;Self-confidence.;Body image in women.;Competition (Psychology);Body image in women.;Competition (Psychology);Self-confidence.;Self-perception in women.;Stature;Women basketball players
Abdul-Jabbar, Kareem,	A season on the reservation :	my soujourn with the White Mountain Apache /	N/A	N/A	Indians;Basketball;Basketball;White Mountain Apache Indians.
Abdul-Jabbar, Kareem,	On the shoulders of giants :	my journey through the Harlem Renaissance /	N/A	Our future or our fate -- "Some technicolor bazaar": how Harlem became the center of the universe -- "Mad medley": how Harlem influenced my life -- "Master intellects and creative giants": the "talented tenth" paint the world black -- "The gifts that my ancestors gave": how Harlem writers influenced my life -- "Fairness creeps out of the soul": basketball comes to Harlem -- "Hoping against hope": how the Rens basketball team influenced my life -- "Musical fireworks": jazz lights up the heavens of Harlem -- "Everything was mostly fun": how jazz influenced my life.	Harlem Renaissance.;Harlem Renaissance;Harlem Renaissance.;Harlem Renaissance.;African Americans;Basketball players.



DATA PACKAGE CREATOR

Upload

Validate

Download

Metadata

Name

my-data-package

Title

My Data Package

Profile

Tabular Data Package

Description

Describe your dataset

Home Page

https://example.com

Version

Resources

Name resource1

Path BasketballMarcRecordsRawMarc.csv

Load



100a

1 Aamidor, Abraham

Title

Description

Data Type

string

Data Format

default

245a

1 Chuck Taylor, Converse all star

Title

Description

Data Type

string

Data Format

default

245b

1 the true story of the man behind

Title

Description

Data Type

string

Data Format

default

500a

1 N/A

Title

Description

Data Type

string

Data Format

default

505a

1 Hall of fame -- Non-skids --

Title

Description

Data Type

string

650a

1 Basketball players:Basketball players

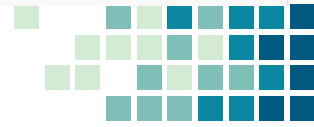
Title

Description

Data Type

string

Add field



Source [\[Provide Link\]](#)

BasketballMarcRecordsRawMarc.csv

[REQUIRED] Add a data table to validate.

Schema [\[Provide Link\]](#)

BasketballMARCSchema.json

Format

Auto

Encoding

Auto

[OPTIONAL] Select to validate this data against a Table Schema ([What is it?](#)).

Ignore blank rows

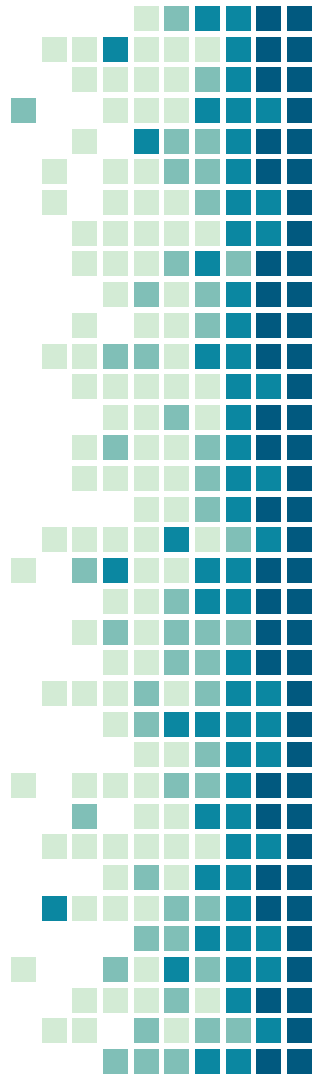
Ignore duplicate rows

Permalink: <http://try.goodtables.io/?checks%5Bduplicate-row%5D=false&checks%5Bblank-row%5D=false&apiJobId=08e31f9c-98aa-11ec-9125-0eccce673dfd>


Warnings

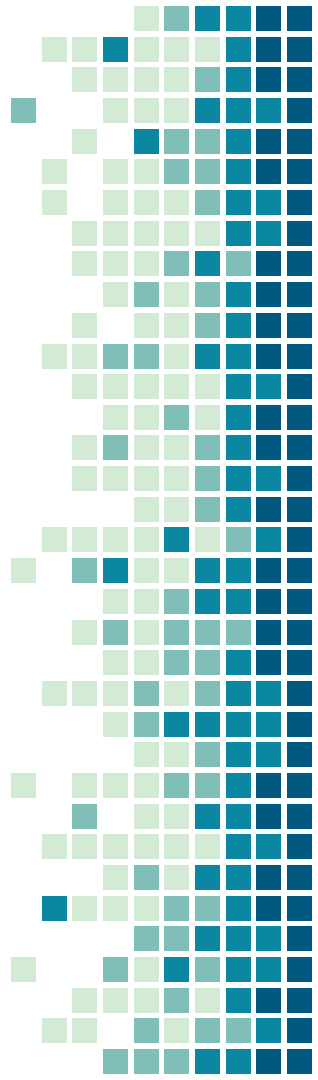
Table "BasketballMarcRecordsRawMarc.csv" inspection has reached 1000 row(s) limit

BasketballMarcRecordsRawMarc.csv



Reproducible Workflow Takeaways

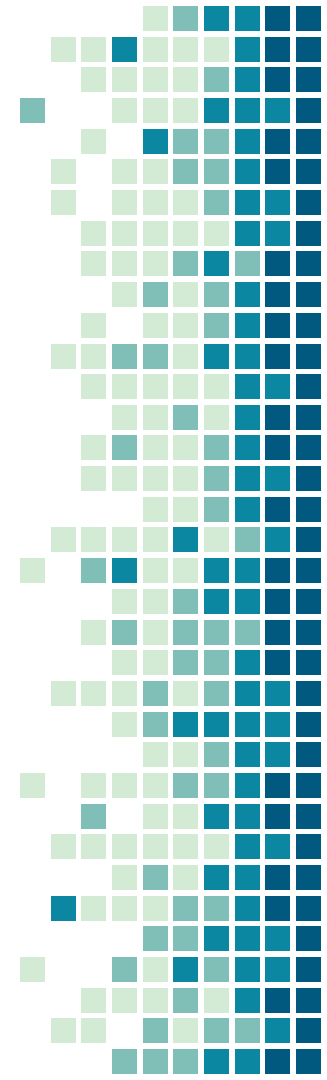
- Frictionless Tools = 
 - Tabular data management
 - Reproducible Research
 - Shallow learning curve





Future Ideas

- “Find and Replace” outdated LCSH in library collections using machine learning
- Machine learning to predict problematic LCSH or descriptive language?



THANKS!

Any questions?

Find me:

@menacegypin

lagypin@uncg.edu

**Read about my experience as a
Frictionless Fellow on our Blog:**

<https://fellows.frictionlessdata.io/blog/>