

# Noxious Weeding

Using Data Visualization to Revitalize Your Collection

**Accacia Mullen**

Reference Librarian  
Virginia Military Institute

**Rachel Maderik**

Systems and Technology Librarian  
Virginia Military Institute

VLACRL Conference-Within-A-Conference  
October 28, 2016

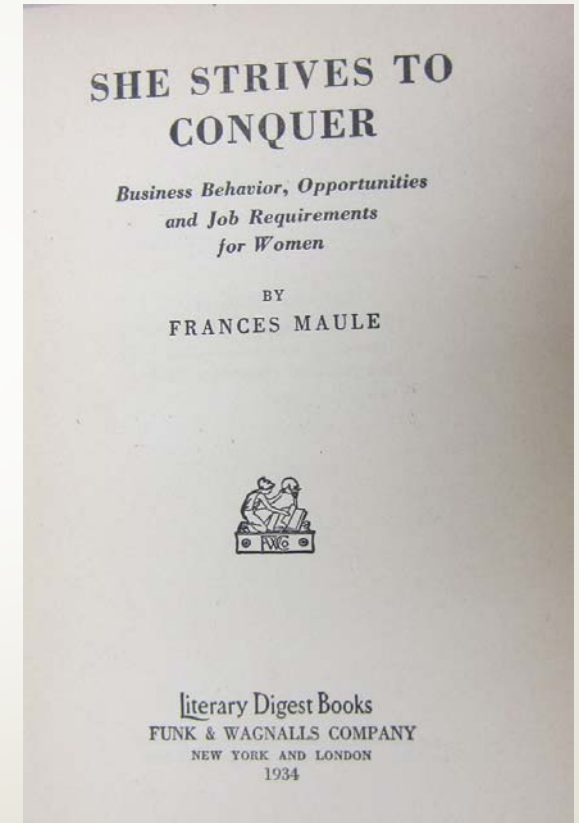
# Introduction

- ThistleCAT Collection Visualization Tool
  - Learn how we developed a collection visualization tool in-house
  - View a live demonstration of the tool
- Weeding Our Collection
  - Take an in-depth look at two weeding projects we undertook using ThistleCAT to guide us:
    - Literature
    - Medicine

# Motivations

## Why Take on a Weeding Project Now?

- Motivations are internal
  - Professional opinions
  - Space concerns
- No external mandate
  - Recent library administration discouraged weeding
  - Building renovation in future, but no definite plans yet



*"The only really safe course for the business woman is to be absolutely odorless." -- p. 146*

# Getting Started

## Beginnings

- Collection Development Task Force formed in October 2015
  - 8 members, all volunteers
- Initial goal was to develop a data-driven weeding plan
- Performed literature review and collected resources for staff
  - <http://libguides.vmi.edu/Collection-development>
- Additional projects were quickly added, such as reviewing our collection development policy

# Visualizing the Monograph Collection

- Began by doing manual reports of selected sections
  - Imported data into Excel
  - Slow and clunky
  - Most staff couldn't make reports themselves
- VIVA Monographic Collection Analysis Project was participating in a collection analysis project using GreenGlass
  - GreenGlass provided just what we needed, but
  - It was prohibitively expensive
- We realized that we could build a similar visualization tool ourselves

1	Category	LCClass	NumberItems	AverageDate	AverageCheckouts
2	Italian	PC1	10	1944	14.1
3	Portuguese	PC5	11	1947	5.82
4	Celtic language and literature	PB1	22	1940	5.18
5	Classical and Greek language	PA1	3	1955	3.33
6	Hyperborean and Native American	PM0	5	1974	3.2
7	Dardic	PK7	1	1969	3
8	Greek literature	PA4	255	1950	2.88
9	Native American	PM3	6	1993	2.6
10	Egyptology	PJ1	15	1968	2.36
11	Spanish	PC4	105	1964	2.33
12	Celtic language and literature	PB2	13	1951	2.23
13	Spanish American literature	PQ8	356	1971	2.19
14	American 19th century literature	PS1	1176	1942	2.12
15	English Renaissance literature	PR2	1818	1947	2.1
16	Flemish literature	PT6	2	1996	2
17	Canadian literature	PS8	2	2008	2
18	American 19th and 20th century literature	PS3	7545	1968	1.97
19	Japanese, Korean	PL0	103	1974	1.86
20	Roman literature	PA6	398	1943	1.79

# Introducing ThistleCAT

**ThistleCAT**

## Thistle Collection Analysis Tool

ThistleCAT assists in analyzing Preston Library's collection by providing data based on various parameters including copyright date, total checkouts, last checkout date, and state and national holdings.

[Collection visualizations »](#) [Foreign language module»](#) [Literature module»](#)

### About this Tool

Please note that for performance reasons, data is not displayed in real time:

- Total circulation history covers 1991 to present.
- Last checkout date covers 2003 to 2015.
- Holdings data is current as of Jan 4, 2016.

### Miscellaneous Views

- [Total Checkouts Graph](#)
- [Ratio of Checkouts to Collection Size](#)

### Barcode List Analysis

Upload a file of scanned barcodes to analyze them.

No file selected.

### Query Builder

Call Number starts with...

Copyright Date  
Before

Total Checkouts  
Fewer than

Last Checkout Date  
Before

Holdings in Other Libraries  
Fewer than  # Holdings in:

Virginia Libraries  U.S. Libraries

Select Language...

VIVA Protected Status

Has Possible Duplicates

### Quick Item Lookup

Get data for an individual item, including:

- Total checkouts
- Last checkout date (2003 or later)
- Possible duplicates
- Holdings in Virginia and U.S. libraries

### Alternate Editions Lookup

**New!!!** View similar and alternate editions for items by LC class.

# Introducing ThistleCAT

- Thistle Collection Analysis Tool
  - What it can't do
    - Benchmarking/peer comparisons
    - HathiTrust comparisons
    - Analyze collections using Dewey (coming soon)
  - What it can do
    - Everything else!
  - Other similar tools may exist, but we did not look any further because we had ThistleCAT up and running within a few days, and it provided what we needed at an acceptable cost.

# Introducing ThistleCAT

## Technical Overview

- Hosted locally on our own web server
- Back end
  - MySQL database that contains item information exported from ILS
  - Does not display live data from the ILS for several reasons:
    - Speed
    - Can integrate data from different sources
    - Flexibility (as long as your ILS can create a comma-delimited report of selected item data, you can use ThistleCAT)
  - Static data allows for weeding to occur without throwing off the visualizations
- Front end
  - PHP to query the database
  - JavaScript to display the charts
- Source code will be available for download



# Introducing ThistleCAT

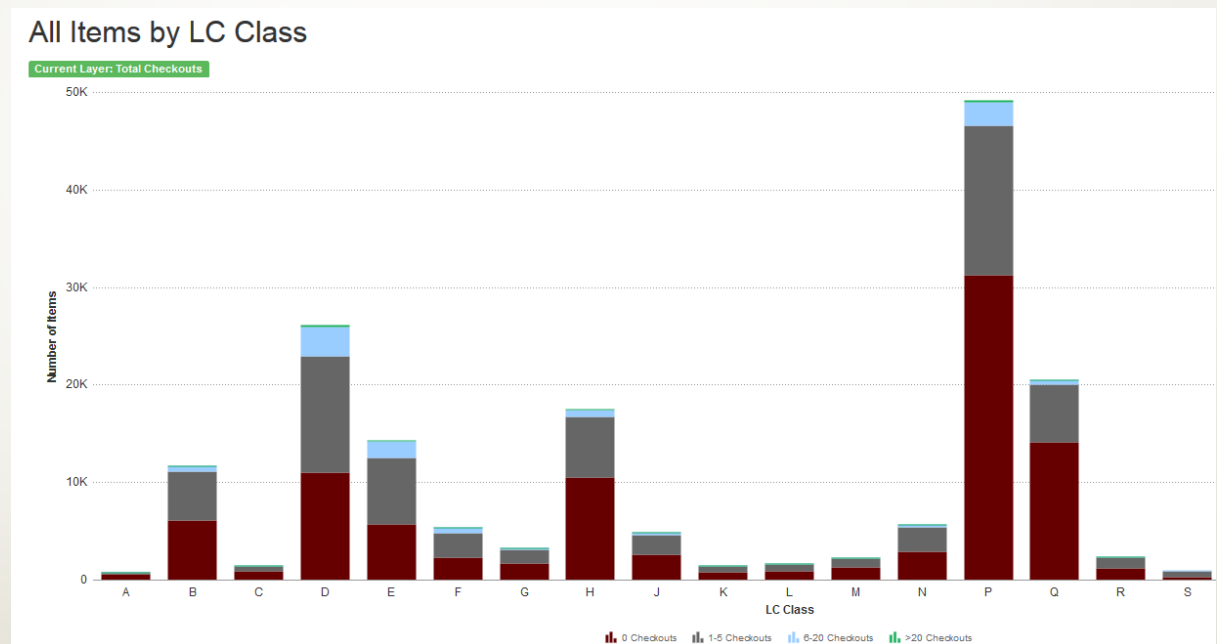
## Demonstration

- Demo!
  - <http://archivesweb.vmi.edu/thistlecat>

# Introducing ThistleCAT

## Standard Visualizations

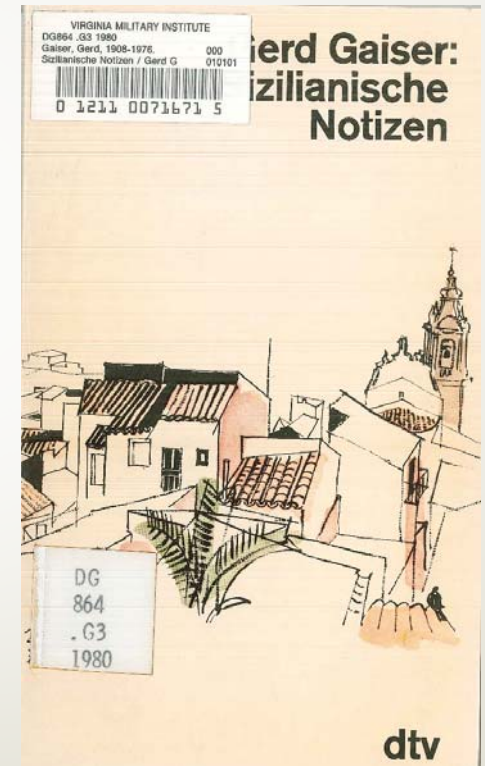
- View items by:
  - LC class or subclass
  - Copyright date
- Additional layers show:
  - Total Checkouts
  - Last Checkout Date
  - Language



# Introducing ThistleCAT

## Foreign Language Module

- Collection development policy says we only acquire foreign language items to support the teaching of that language
  - Also, relevant languages can change over time
- ThistleCAT provides visualizations by LC class or language
  - See which classes have foreign language items
    - Do we need Italian travel books in German?
    - Or engineering books in French?
  - Easily view all items for a specific language



Travelling to Italy? Speak German?  
*Then we've got the book for you!*

# Introducing ThistleCAT

## Quick Item Lookup

- Get data for a single item
- Pulls all relevant data into one place
  - Can include information from multiple ILSes
  - Other data is only available in multiple locations or can be difficult to acquire
    - OCLC # (hard to find in our ILS)
    - VIVA Monographic Collection Analysis Project [Protected Title List](#)

The screenshot displays the ThistleCAT interface for a single item lookup. At the top, the title "Harry Potter and the deathly hallows /" is shown. In the top right corner, there are logos for "koha FREE LIBRARY SYSTEM" and "WorldCat". The main content area is divided into two columns of metadata. The left column includes: "Copyright Date: 2007", "Call Number: PR6068.O93 H39", "Total Checkouts: 16", and "Last Checkout Date: 2015-11-06". The right column includes: "Item Barcode: 0021102169840" and "OCLC Number: 155131850". Below this metadata, there is a green bar with the text "Holdings in Virginia Libraries:" followed by a green button labeled "VIVA Protected Title". At the bottom right of the interface, there is a red button with a white 'X' icon and the text "Close".

<b>Copyright Date:</b> 2007	<b>Item Barcode:</b> 0021102169840
<b>Call Number:</b> PR6068.O93 H39	<b>OCLC Number:</b> 155131850
<b>Total Checkouts:</b> 16	
<b>Last Checkout Date:</b> 2015-11-06	

**Holdings in Virginia Libraries:** VIVA Protected Title

# Introducing ThistleCAT

## Alternate Editions

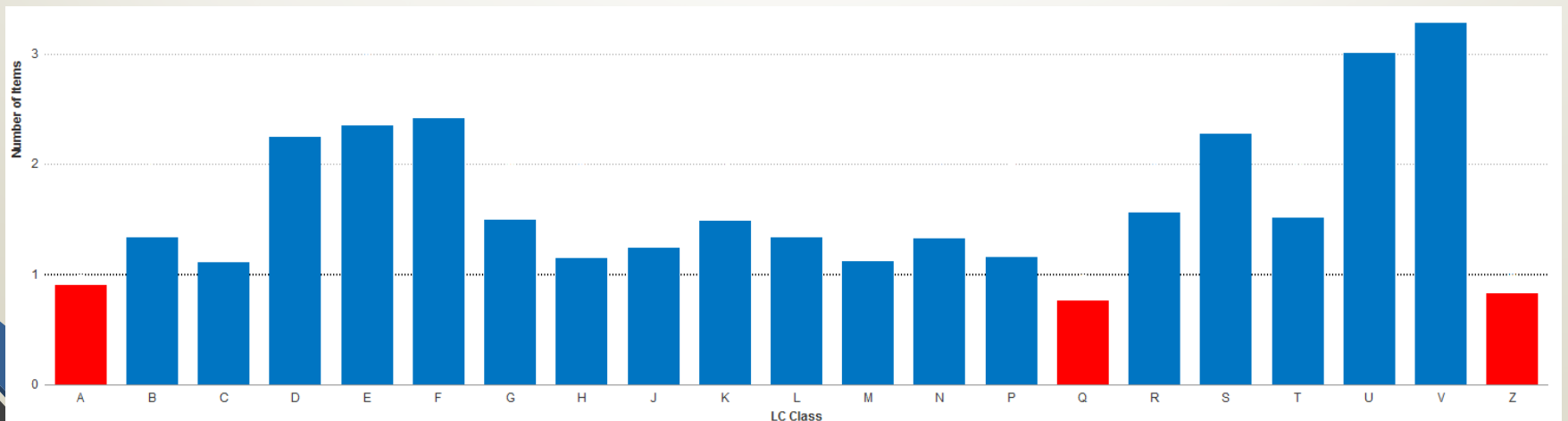
- Awesome feature that is dependent on a deprecated OCLC service: xID API
  - API that provides OCLC numbers for past/future editions of an item
  - Immensely helpful for identifying superseded editions in our collection
- Other sources of this information may be available
  - LibraryThing's "ThingISBN" API

QH325 .J34	0121400328690	Red giants and white dwarfs;	Jastrow, Robert,	1967	0
QB44.2 .J34 1979	0121101236549	Red giants and white dwarfs /	Jastrow, Robert,	1979	1
QB51.3.E43 M6613 1990	0021100213285	Astronomy on the personal computer /	Montenbruck, Oliver,	1991	2
QB51.3.E43 M6613 1994	0021100830633	Astronomy on the personal computer /	Montenbruck, Oliver,	1994	0
QB51.3.E43 M6613 2000	0021101098867	Astronomy on the personal computer /	Montenbruck, Oliver,	2000	1

# Introducing ThistleCAT

## Miscellaneous Views

- Total Checkouts Graph
  - Useful when first setting up ThistleCAT to determine good checkout thresholds
- Ratio of Checkouts to Collection Size
  - Useful to quickly identify classes that have unusually low checkout ratios



# Applying the Data

## Styles of Weeding

- Hands-on (high faculty involvement)
  - Knowledge gaps in library staff
  - Enthusiastic faculty
  - Strengthen relationships with academic departments
  - Good for larger collections
- Hands-off (low faculty involvement)
  - Time limitations (quicker when you don't have to deal with external people)
  - Knowledge gaps in faculty
    - No stakeholders in a particular subject area
    - Low emphasis in the curriculum
  - Adequate for smaller collections

# Applying the Data

Literature: Hands-On

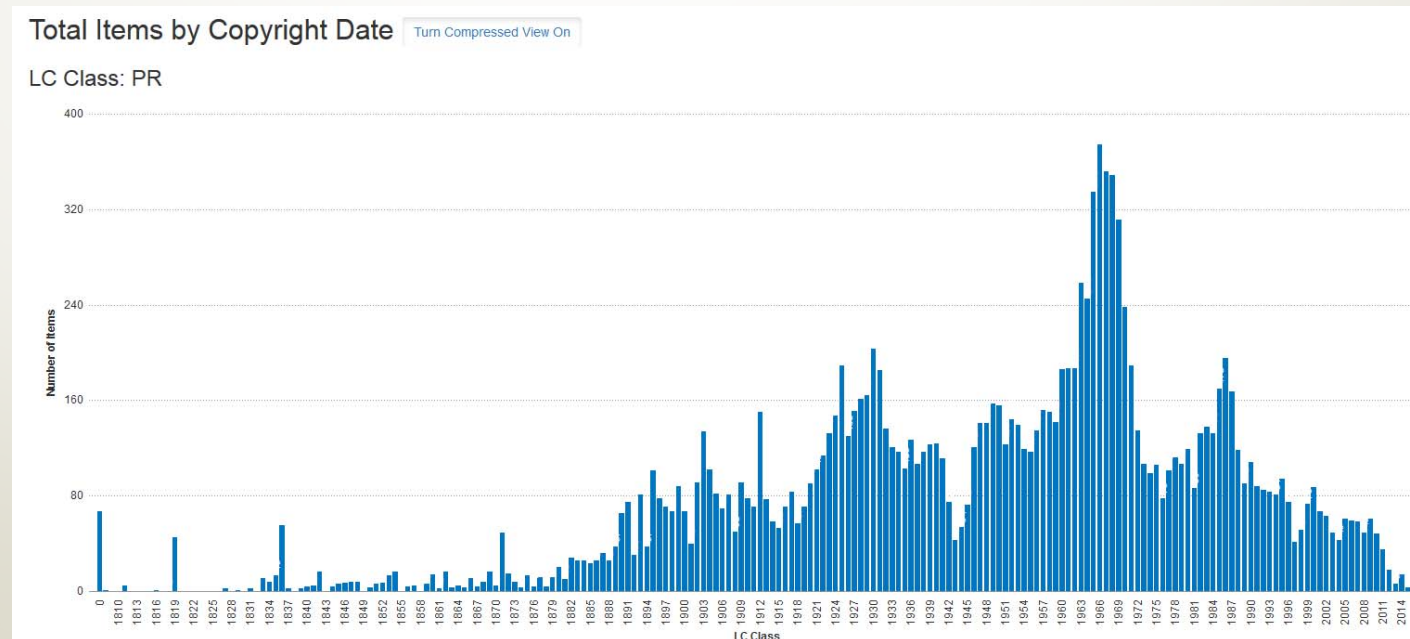
- Largest portion of our collection by total volumes
- Good relationship with department's liaison to library
- Recent helpful feedback on collection development questions
- Department was in the process of weeding their own library



# Weeding Literature

## Getting Started

- ThistleCAT's user-friendly interface allows faculty to interact directly with the data
- Unfortunately, traditional visualizations are not helpful when analyzing literature



# Weeding Literature

## Problems with Visualizing Literature

- No easy way to visualize particular authors or eras
  - Inconsistent call number patterns makes this difficult to assess systematically.
    - Beowulf: PR1580-1588
    - John Milton: PR3550-3598
    - Virginia Woolf: PR6045.O72
    - James Joyce: PR6019.O9
- A new module was developed to allow for author-based visualizations

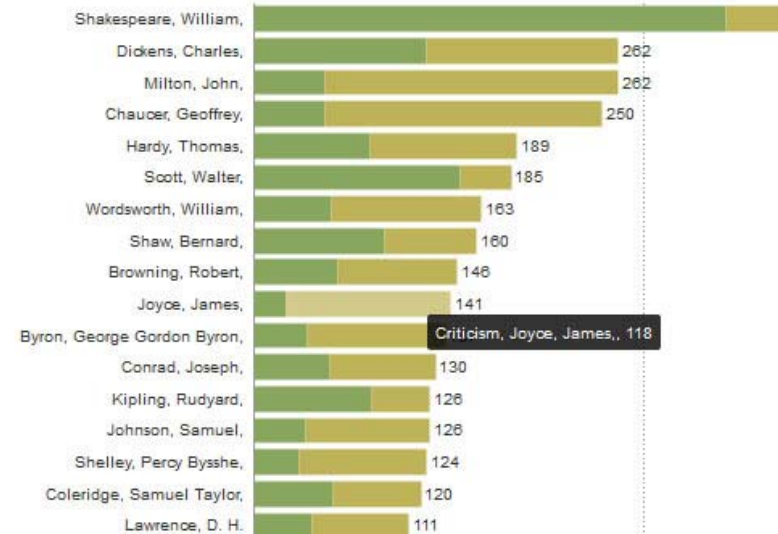
# Weeding Literature

## Tool Development

- Visualizes selected class using the LC outline's categories
- Provides a ranking of authors by number of items in our collection
- Indicates how many are primary vs. secondary literature
  - i.e., original works vs. criticism or biographies

### Author Overview

Current Layer: Original Works vs. Criticism



# Weeding Literature

What Can We Learn From These New Visualizations?

- Does the author ranking match their importance in the literary canon and/or our curriculum?
- Have we been acquiring them consistently over the years?
- Is there a balanced ratio of primary to secondary literature for each author?
  - Too much primary literature may indicate a formerly popular writer of no academic interest
  - Too much secondary literature might mean we need to acquire some of their original works



cf. Raphael, Laura. "[Killing Sir Walter Scott: a philosophical exploration of weeding.](#)" *In the Library with the Lead Pipe* (Jul 24, 2013).

# Weeding Literature

## Involving the Faculty

- ThistleCAT helped us prioritize which authors to start with
- Worked with liaison to identify faculty specialists for the authors we were interested in weeding
- Faculty varied in how much they were able to help
  - Most were willing be consulted with questions as we weeded their sections
  - Some preferred to weed their entire sections themselves

# Weeding Literature

## Involving the Faculty

- Faculty had the option to view author visualizations themselves and make weeding decisions at their computer
- Some instead preferred a hands-on approach and wanted to review at the shelf
  - They were given flags to mark items for removal or replacement
  - ThistleCAT was modified to allow for a barcode file of flagged items to be uploaded and visualized



# Weeding Literature

## What We Learned

- Top authors in the collection are not the ones most heavily taught
- Library staff learned more about the current teaching focuses
- We need to make our collection align with our 21<sup>st</sup> century curriculum

# Applying the Data

## Medicine: Hands-Off

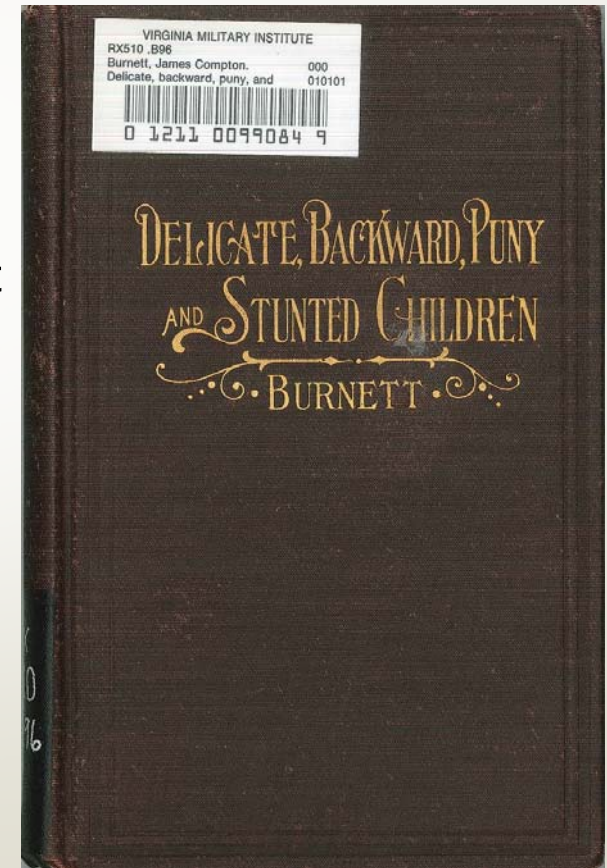
- Relatively small collection (2,400 items)
- Subject that we know needs a lot of specialized help (i.e., librarians)
- Not a big part of our curriculum
- Faculty characteristics
  - No dedicated medical faculty (no one had “ownership” of the collection)
  - Relevant faculty are scattered across departments (psychology, PE, and biology)
  - Limited availability



# Weeding Medicine

## Factors to Consider

- “The American Medical Association [recommends](#) that all books on medicine and health, including diet and nutrition, be no more than **three to five years old.**”
- In our collection, there are only 36 items in R published within the last 5 years (1.5%)
- Need a way to determine a lower threshold



# Weeding Medicine

## Determining Thresholds

- Age
  - Because our entire collection is so old, relying on this too heavily would remove too much
    - < 15 years old: 242 items (10%)
    - < 25 years old: 432 items (18%)
- Circulation
  - In our situation, just because something checked out a lot doesn't mean it's good
  - It could be because it's all we have on that topic
- Need to carefully construct weeding criteria
  - Another reason why it's better to have librarians weed this section

# Weeding Medicine

## Tool Development

- Query Builder
  - Allows for user-friendly query building
  - Can set a desired percentage for weeding to determine a date cutoff
  - Helpful to determine possible weeding thresholds and lists

### Query Builder

Call Number starts with...

Copyright Date  
Before

Total Checkouts  
Fewer than

Last Checkout Date  
Before

Holdings in Other Libraries  
Fewer than  in:  
 Virginia Libraries  U.S. Libraries

Select Language...

VIVA Protected Status

Has Possible Duplicates

**Percentage of Class to Weed** ⓘ

# Weeding Medicine

## Weeding Process

- Identified **high-circulating** old items to determine which subject areas needed to be refreshed
- Identified **low-circulating** old items for withdrawal
  - Ideal
    - 0 checkouts, older than 1990 (42%)
  - Realistic
    - 0 checkouts, older than 1970 (25%)
  - Narrow
    - VIVA protected titles, older than 1990, < 3 checkouts (8%)

# Weeding Medicine

## Evaluating the Narrow List

- Reviewed within ThistleCAT, not at the shelf
- Allowed us to easily search for:
  - Book reviews
  - Biographies of the authors
  - E-versions in our catalog
  - Availability in HathiTrust or Archive.org
- A status report is exported when review is complete
  - Items marked “weed” were flagged by a student worker
  - Items marked “check” were reviewed by librarians in hand

Weeding Status	Call Number	Barcode	Title
<input type="radio"/> Keep <input type="radio"/> Weed <input type="radio"/> Check	R707 .J4 1861	0121100789688	A book about doctors.
<input checked="" type="radio"/> Keep <input type="radio"/> Weed <input type="radio"/> Check	RD151 .C57 c.2	0021100944566	A manual of military surgery, for the use of surgeons in the Confederate States army;
<input type="radio"/> Keep <input type="radio"/> Weed <input type="radio"/> Check	RA1211 .W88	0121101088700	Micro-chemistry of poisons,
<input type="radio"/> Keep <input type="radio"/> Weed <input type="radio"/> Check	RJ45 .W5 1874	0121100654478	Lectures on the diseases of infancy and childhood.
<input type="radio"/> Keep <input checked="" type="radio"/> Weed <input type="radio"/> Check	RC602 .M4 1875	0121100702426	Responsibility in mental disease.
<input type="radio"/> Keep <input type="radio"/> Weed <input type="radio"/> Check	R117 .H79 1887	0121101136830	Medical essays, 1842-1882 / Possible duplicates: <span style="background-color: red; color: white;">5091379</span>
<input type="radio"/> Keep <input type="radio"/> Weed <input type="radio"/> Check	RC606 .A6	0021102047822	From under the cloud;

# Conclusions

- Weeding is often avoided, but collection visualization can make the process manageable.
- In-house tool development allowed for ongoing enhancements as we became aware of our needs.
- Explore or download ThistleCAT:
  - Demo: <http://archivesweb.vmi.edu/thistlecat>
  - GitHub: <https://github.com/thistlecat>

For questions or assistance, please contact us!

Accacia Mullen  
[mullenam@vmi.edu](mailto:mullenam@vmi.edu)

Rachel Maderik  
[maderikra@vmi.edu](mailto:maderikra@vmi.edu)