Supporting Carolina’s Research

Michelle Cawley
Head of Clinical, Academic, and Research Engagement

Jennifer Bissram
Health Sciences Librarian
Liaison to Adams School of Dentistry

Adam Dodd
Data/Applications Analyst
Co-Lead Impact Measurement and Visualization (IMV) Team

UNC | UNIVERSITY LIBRARIES
Health Sciences Library

New Faculty
August 2021

Session Road Map

- Evaluate baseline knowledge
- Resources for literature searches
  - Managing search results
  - Screening literature
  - AI and Machine Learning for prioritizing search results
- Impact Measurement & Visualization
- Q&A

Go To: www.menti.com
Enter Code: 3304 9766

Rate each statement as it applies to you.

Mentimeter

Strongly disagree — Strongly agree

- I am experienced with Systematic Reviews.
- I am experienced using reference managers.
- I am familiar with online tools for screening search results.
- I am interested in using AI to reduce manual screening of search results.
- I am familiar with bibliometric analysis.

Go To: www.menti.com
Enter Code: 3304 9766

Which reference manger do you prefer? Select the best option.

Mentimeter

0	0	0
1. I don't use a reference manager.	2. I prefer EndNote.	3. I prefer SciWheel (formerly F1000).
0	0	
4. I prefer Zotero.	5. Other, not listed (e.g., Mendeley).	

Reference Managers & Screening Software

Jennifer Bissram

Health Sciences Librarian
Liaison to Adams School of Dentistry

UNIVERSITY LIBRARIES
Health Sciences Library

Managing search results

Citation / reference managers

- EndNote (recommended)
- SciWheel
- Zotero

Need a citation manager recommendation?

https://guides.lib.unc.edu/compare-citation

Managing search results

How do they help?

- Store references from multiple sources
- Duplicate removal
- Creates file for importing into Covidence

For more information on managing citations for systematic reviews…

https://guides.lib.unc.edu/systematic-reviews/citations

Screening the literature with Covidence

- Covidence is tool for title/abstract screening, full-text screening, data abstraction, and quality assessment.
- Provided free from HSL & its partners
- Unlimited reviews

Title / Abstract Screening

All | Filter | Tags | Show criteria | Hide highlights | Hide abstracts | Display: 25 | Relevancy

#20 - Liu 2015

Liu, A. G.; Puyau, R. S.; Han, H.; Johnson, W. D.; Greenway, F. L.; Dhurandhar, N. V.

The effect of an egg breakfast on satiety in children and adolescents: a randomized crossover trial

J Am Coll Nutr 2015;34(3):185-90
2015

Hide Abstract & IDs

DOI: 10.1080/07315724.2014.942471

OBJECTIVE: To evaluate the effects of an egg breakfast on lunchtime energy intake in children (age 4-6 years) and adolescents (age 14-17 years). METHODS: In 2 randomized crossover trials, participants received either an egg breakfast or an isocaloric bagel breakfast. In both trials, subsequent lunchtime energy intake was the primary outcome. The trial with adolescents also measured each participant's serum ghrelin, serum peptide YY (PYY), and self-assessment of appetite rated using a visual analog scale. RESULTS: Lunchtime food intakes after egg and bagel breakfasts were not significantly different for either age group. Visual analog scale ratings of hunger and satiety were also not different between the 2 treatments in adolescents. Consumption of the egg breakfast led to a significant increase in serum PYY levels (p = 0.0001) in adolescents. However, increased levels of PYY were not correlated with reduced food intake. CONCLUSION: Short-term food intake in children and adolescents is not differentially altered by an egg breakfast compared to a bagel breakfast.

View history | Add a note

No | Maybe | Yes

#24 - Lee 2015

Lee, H. J.; Ryu, D.

Significance of Ochratoxin A in Breakfast Cereals from the United States

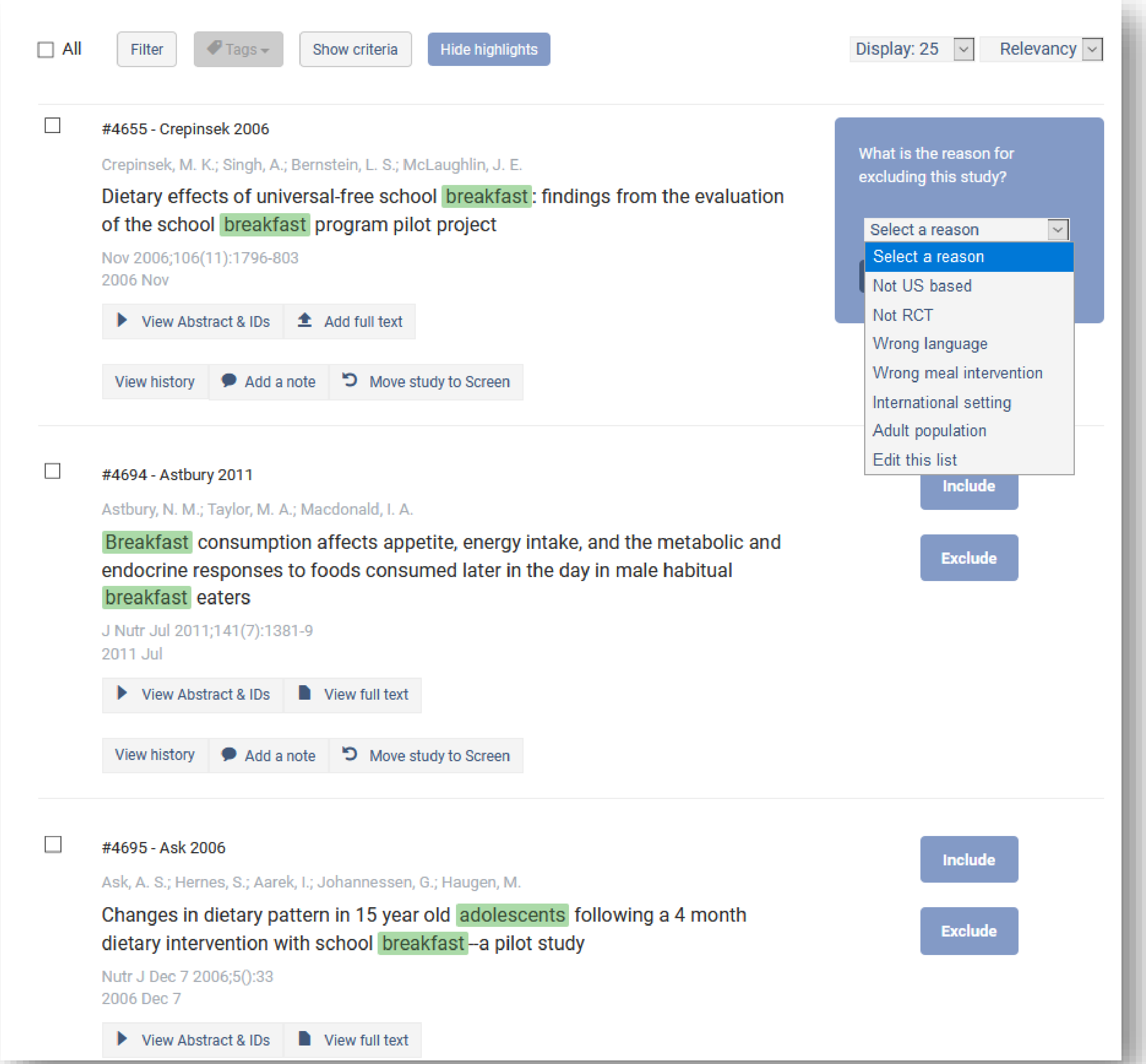
J Agric Food Chem Nov 04 2015;63(43):9404-9
2015 Nov 04

Hide Abstract & IDs

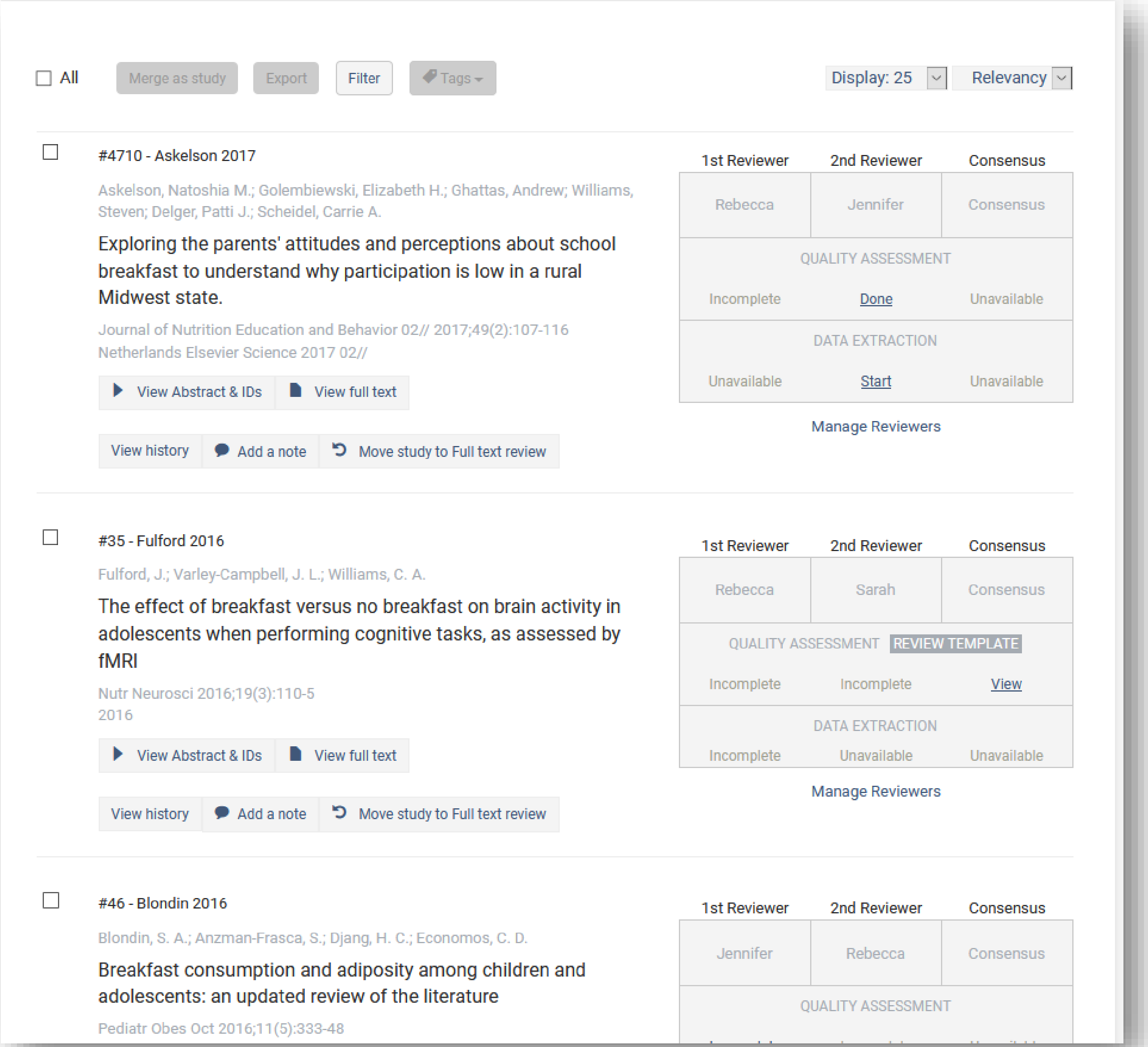
DOI: 10.1021/jf505674v

Ochratoxin A (OTA) has been found in all major cereal grains including oat, wheat, and barley worldwide and considered as a potential concern in food safety. A total of 489 samples of corn-, rice-, wheat-, and oat-based breakfast cereal were collected from U.S. retail marketplaces over a two-year period, and OTA was determined by high-performance liquid chromatography. Overall,

No | Maybe | Yes



Full Text Screening


All
Merge as study
Export
Filter
Tags
Display: 25
Relevancy
#4710 - Askelson 2017
Askelson, Natoshia M.; Golembiewski, Elizabeth H.; Ghattas, Andrew; Williams, Steven; Delger, Patti J.; Scheidel, Carrie A.
Exploring the parents' attitudes and perceptions about school breakfast to understand why participation is low in a rural Midwest state.
Journal of Nutrition Education and Behavior 02// 2017;49(2):107-116
Netherlands Elsevier Science 2017 02//
View Abstract & IDs
View full text
View history
Add a note
Move study to Full text review
1st Reviewer
2nd Reviewer
Consensus
Rebecca
Jennifer
Consensus
QUALITY ASSESSMENT
Incomplete
Done
Unavailable
DATA EXTRACTION
Unavailable
Start
Unavailable
Manage Reviewers
#35 - Fulford 2016
Fulford, J.; Varley-Campbell, J. L.; Williams, C. A.
The effect of breakfast versus no breakfast on brain activity in adolescents when performing cognitive tasks, as assessed by fMRI
Nutr Neurosci 2016;19(3):110-5
2016
View Abstract & IDs
View full text
View history
Add a note
Move study to Full text review
1st Reviewer
2nd Reviewer
Consensus
Rebecca
Sarah
Consensus
QUALITY ASSESSMENT
REVIEW TEMPLATE
Incomplete
Incomplete
View
DATA EXTRACTION
Incomplete
Unavailable
Unavailable
Manage Reviewers
#46 - Blondin 2016
Blondin, S. A.; Anzman-Frasca, S.; Djang, H. C.; Economos, C. D.
Breakfast consumption and adiposity among children and adolescents: an updated review of the literature
Pediatr Obes Oct 2016;11(5):333-48
1st Reviewer
2nd Reviewer
Consensus
Jennifer
Rebecca
Consensus
QUALITY ASSESSMENT

Extraction

Want more information?

Systematic Review Guide

https://guides.lib.unc.edu/systematic-reviews

Covidence Guide

https://guides.lib.unc.edu/Covidence

Upcoming Classes

https://hsl.lib.unc.edu/

Automation Approaches for Literature Searches

Michelle Cawley

Head of Clinical, Academic, and Research Engagement

UNC | UNIVERSITY LIBRARIES
Health Sciences Library

An Automation Approach for Every Step

BUILD SEARCH

Tinker with and compile search terms.

ANALYZE KEYWORDS

Keyword prevalence within a set of search results.

REMOVE DUPLICATES

Two-phase process with AI to predict likely duplicates.

PRIORITIZE RESULTS

Clustering and ML to prioritize records for review.

Keyword Analysis Tool (KAT)

Analyzing Keywords

bisexual
bi-sexual
citizenship status
disabilities
disability
disabled
disadvantage
disadvantaged
discrimination
discriminatory
disparities
disparity
diverse
economic status
elderly
equitable
equities
ethnic
ethnicity
gender

geographically isolated
HIV/AIDS
homeless
homelessness
illegal alien
illegal aliens
immigrant
immigrants
immigration status
indigent
inequalities
inequality
inequity
institutionalized
lesbian
LGBT
LGBTQ
low income
low population density
marginal
marginalization
marginalized

medicaid recipient
medicaid recipients
medically complex
mental illness
mentally disabled
mentally ill
migrant
migrants
minorities
non-English speaking
oppressed
oppression
people of color
peripheralized
power (psychology)
pregnant women
prisoners
refugee
rural
sexual orientation
sexuality

social exclusion
social isolation
social status
socioeconomic
stigmatized
stigmatizing
susceptible
transgender
transgendered
transient
underinsured
underpopulated
underrepresented
under-represented
underserved
undocumented
unemployed
uninsured
veteran
veterans
vulnerable

KAT Output

- Term “gender” appears in 3,503 records or 13.53% of results.
- If term “gender” is removed, results will drop by approximately 1500 results.

Keyword	Percentage Occurrence	Number of Documents with Keyword	Reduction in results if keyword removed
gender	**13.62**	**3503**	**1569**
diverse	12.58	3234	1705
disability	6.82	1753	845
diversity	5.88	1511	863
disparities	5.82	1496	334
socioeconomic	5.71	1467	500
vulnerable	5.6	1440	534
stigma	5.47	1407	421
ethnic	5.36	1377	205
low income	5.32	1367	481
elderly	5.31	1365	846
minority	5.21	1340	255
ethnicity	3.73	958	231
disabilities	3.16	813	314
mental illness	3.1	797	357
pregnant women	3.08	793	451
discrimination	3.04	782	195
equity	2.47	634	194
underserved	2.41	620	143
vulnerability	2.31	595	241
gay	2.18	560	69

Prioritizing Search Results Using Machine Learning

Experience and Interest in Machine Learning (ML)

★ Add annotation where applicable.

No experience using ML on *bibliographic data*.

Some experience using ML on *bibliographic data*.

Some experience with ML on *other data* (not bibliographic).

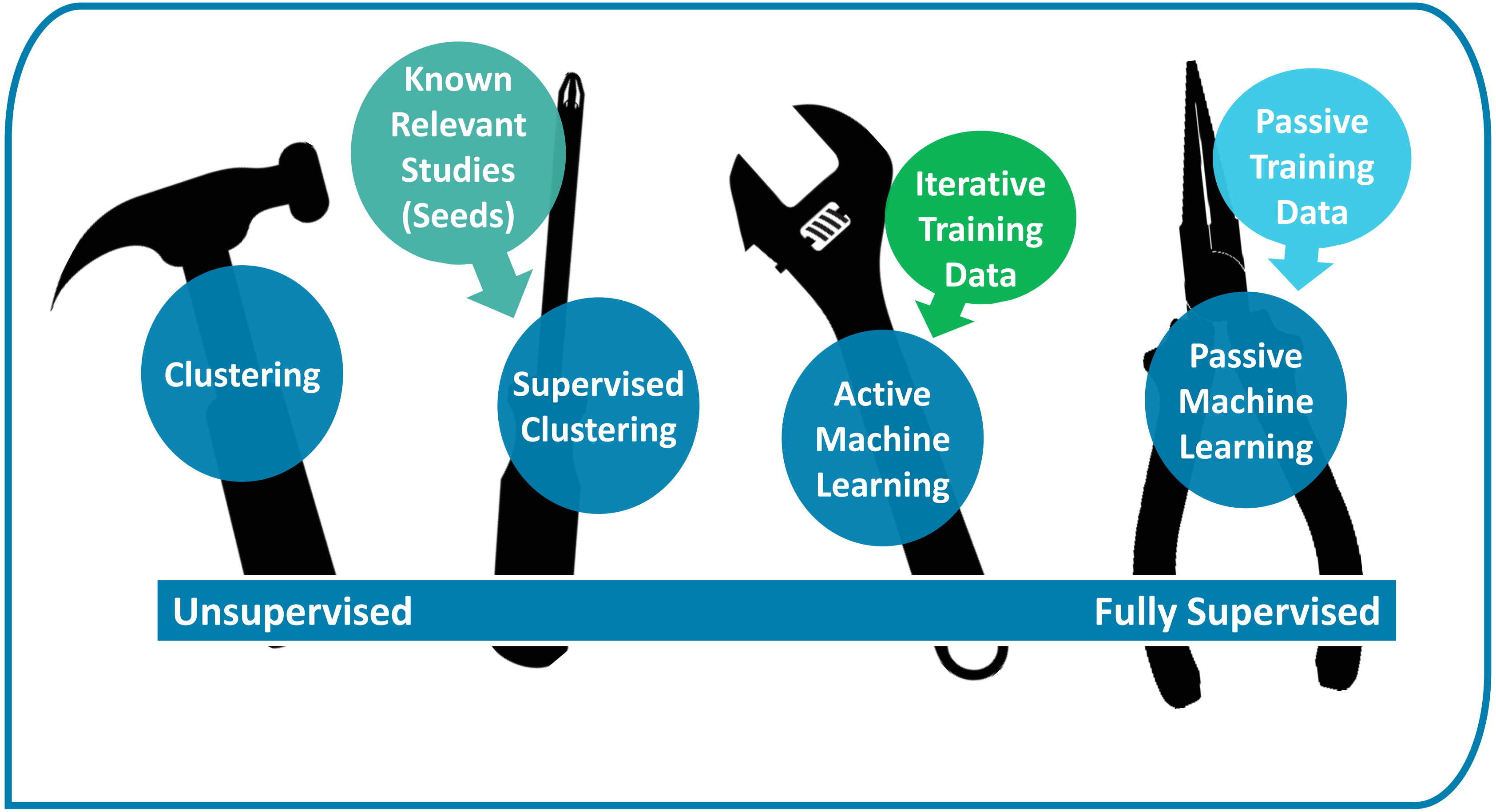
Interested in applying ML to *bibliographic data*.

(Some) Machine Learning Terms

★ Add annotation to terms with which you are familiar.

Artificial Intelligence	Supervised Machine Learning	Active Machine Learning	Deep Learning
Unsupervised Machine Learning	Clustering	Seed Studies	Training Data

Overview of Text Analytics Approaches



Why Use Machine Learning

- Save time.
  - Team members can begin extracting data sooner.
  - Screen fewer studies.
- Expand the scope of your question.

Image by [Jörg Peter]() from [Pixabay]()

When to Use Machine Learning to Reduce Manual Screening

Large Set of Search Results

Generally, we recommend searches with 3,000 or more results to get a measurable benefit.

Various Publication Types

- Scoping reviews
- Rapid reviews
- Systematic and systematized reviews
- Bibliometric analysis

Health Related Topics

Evidence of efficacy is primarily in the areas of health and medicine. Much of the peer-reviewed literature in this domain also.

Low Precision Topics

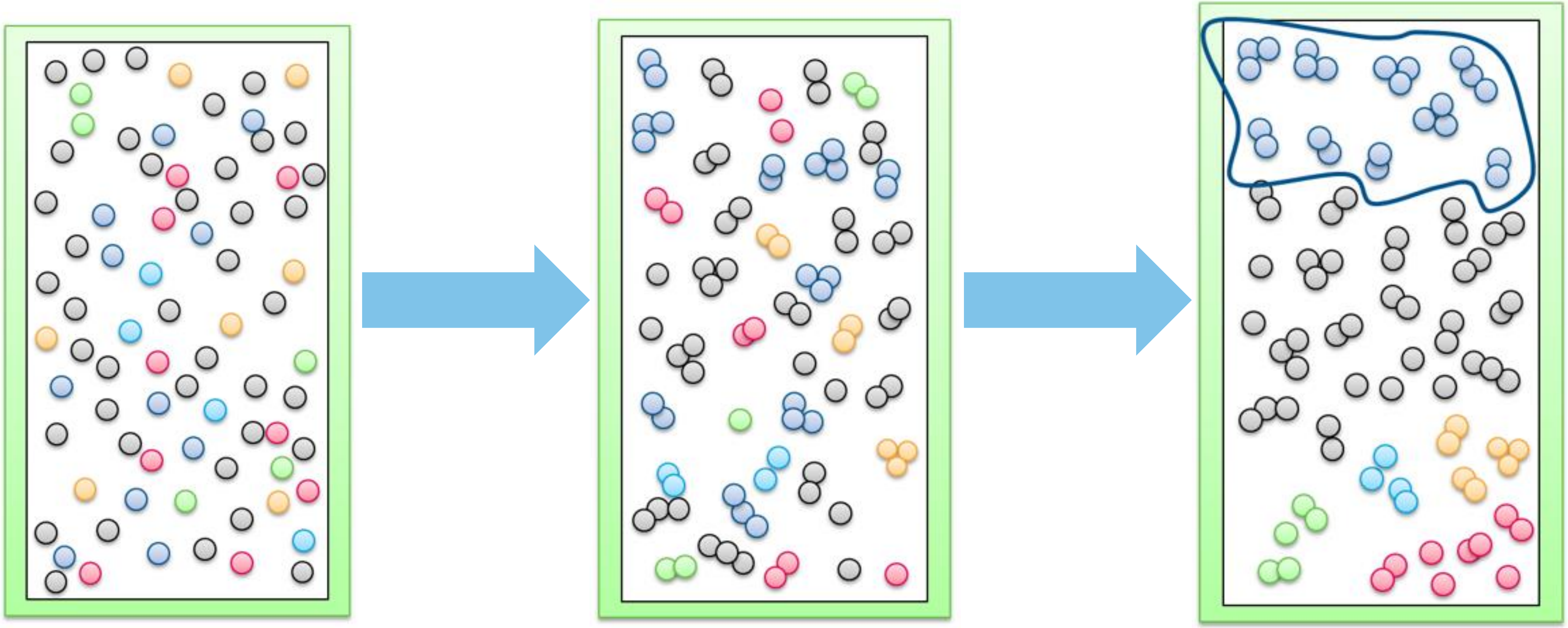
“Messy” topics that require imprecise search terms leading to significant false positives are good candidates.

How it Works

Clustering At a Glance

- **Unsupervised machine learning** on a set of search results.
  - Potential algorithms: K-Means, Nonnegative Matrix Factorization (NMF), Latent Dirichlet Allocation (LDA).
- **“Refine or Shine”**
  - Refine search strategy.
  - Identify a pocket of results to look at more closely.
- **Benefits**
  - No training data necessary.
  - Simple, quick, data-driven approach.
- **Limitations**
  - Requires subject matter knowledge.
  - No quantifiable predictions of recall.

Visualizing Clustering



Compile search results

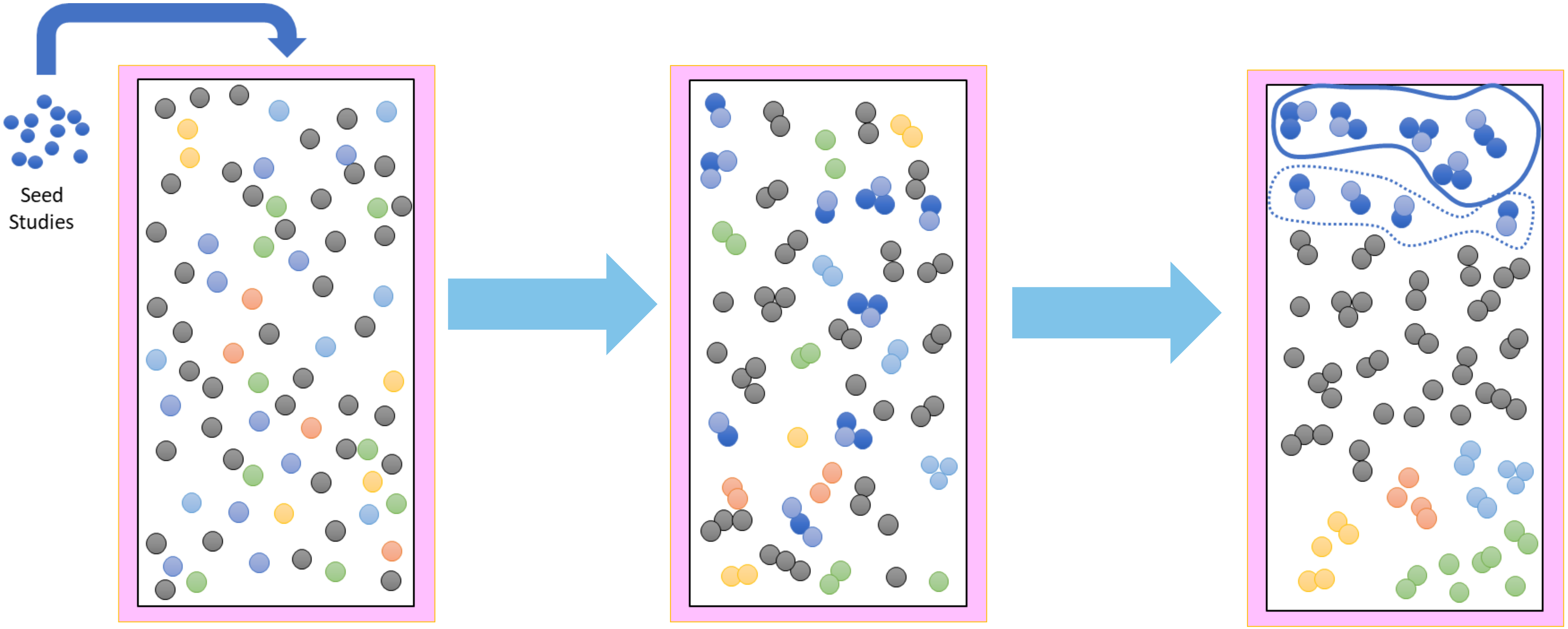
Cluster based on text similarities in titles and abstracts

Review high priority clusters based on keyword summary

Supervised Clustering At a Glance

- **Semi-supervised machine learning** on a set of search results
  - Underlying approach is the same as clustering.
- **Seeds are required**
  - Follow the seeds to prioritize a batch of records to review.
- **Benefits**
  - Relatively little training data are required.
  - Quantitative approach that takes the guess work out of which clusters to review.
  - Recall can be predicted.

Visualizing Supervised Clustering



Compile search results + known relevant studies or seeds.

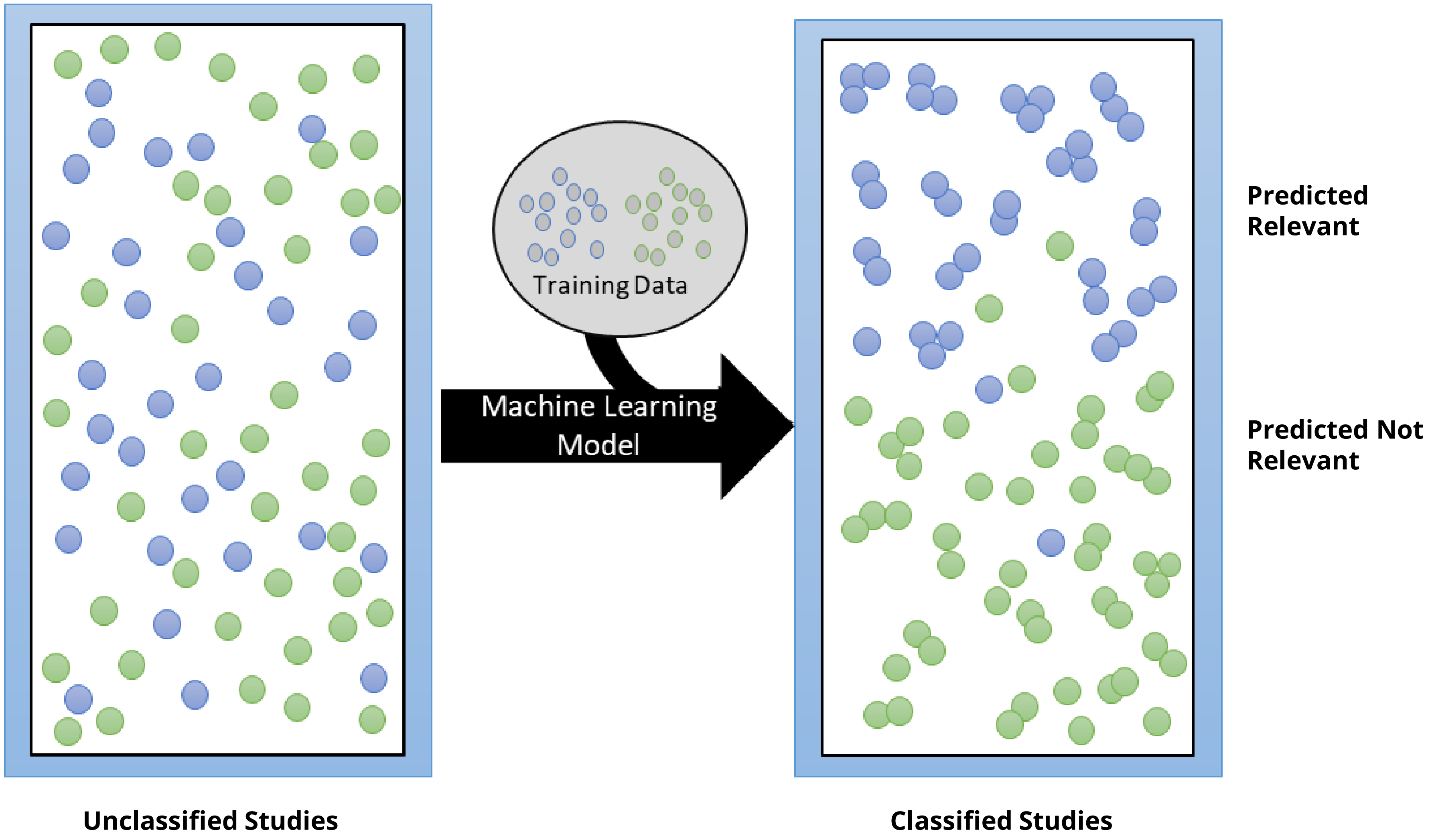
Cluster based on text similarities in titles and abstracts.

Review high priority clusters based on percent of seed studies.

Machine Learning At a Glance

- Fully supervised method
  - Potential algorithms: Naïve-Bayes, Support Vector Machines (SVM)
- Can be active or passive training
- Benefits
  - Regarded as being more precise method (i.e., fewer false positives)
- Challenges
  - Training dataset required
  - Active ML predictions may be unreliable

Visualizing Machine Learning



Techniques Side-by-Side

Reduced False Positives

Machine Learning	Supervised learning; training dataset required; predictive metrics.
Supervised Clustering	Semi-supervised learning; seeds required; predictive metrics.
Clustering	Unsupervised learning; no input required; no predictive metrics.
Keyword Search Strategy	Traditional approach; difficult to balance precision and recall.

UNC Researchers Using Automation

Project	Change in Precision[1]	Studies Not Screened Manually	Time Saved (Hours)[2]
1. ExHa	+ 8.0%	4,697	78-157
2. DisHF			
3. CoEHS			

[1] Precision is calculated as # relevant studies identified/total reviewed. Change in precision refers to calculation of precision in prioritized studies (after machine learning) compared to randomly selected studies. A positive change indicates that precision increased following machine learning.

[2] Time saved calculated as a range of 30 seconds (low end) to 1 minute (high end) x number of studies x 2 reviewers.

Impact Measurement & Visualization

Adam Dodd

Data/Applications Analyst
Co-Lead Impact Measurement and Visualization (IMV) Team

UNC | UNIVERSITY LIBRARIES
Health Sciences Library

IMV @ HSL

IMV partners with the UNC-Chapel Hill community on impact measurement and visualization projects in support of research, education, and clinical care.

- Identify patterns, trends, and gaps in unit-supported research, education or clinical care activities
- Measure and assess research impact by discipline area
- Discover the scope and pattern of research collaborations (e.g. author, institution, and country collaboration)
- Communicate research impact to audiences such as funders or promotion and tenure committees

Analysis Process

- Identify project scope, goals, and research questions
- Construct search strategy in consultation with domain experts
- Retrieve and clean data (most time-consuming step)
- Perform analyses; produce visualizations
- Review initial analyses with team members for quality assurance; identify additional questions
- Revise analyses / visualizations as needed

Data Types and Sources

Data Types

- Publications
- Grants
- Patents

Data Source Examples

- Scopus, Web of Science (citation tracking databases)
- PubMed, CINAHL, CT.gov, etc. (more specific databases)
- NIH Federal Reporter, UNC RAMSeS

Project Example:
Nurse Research Capacity within the Caribbean

Research collaborators:

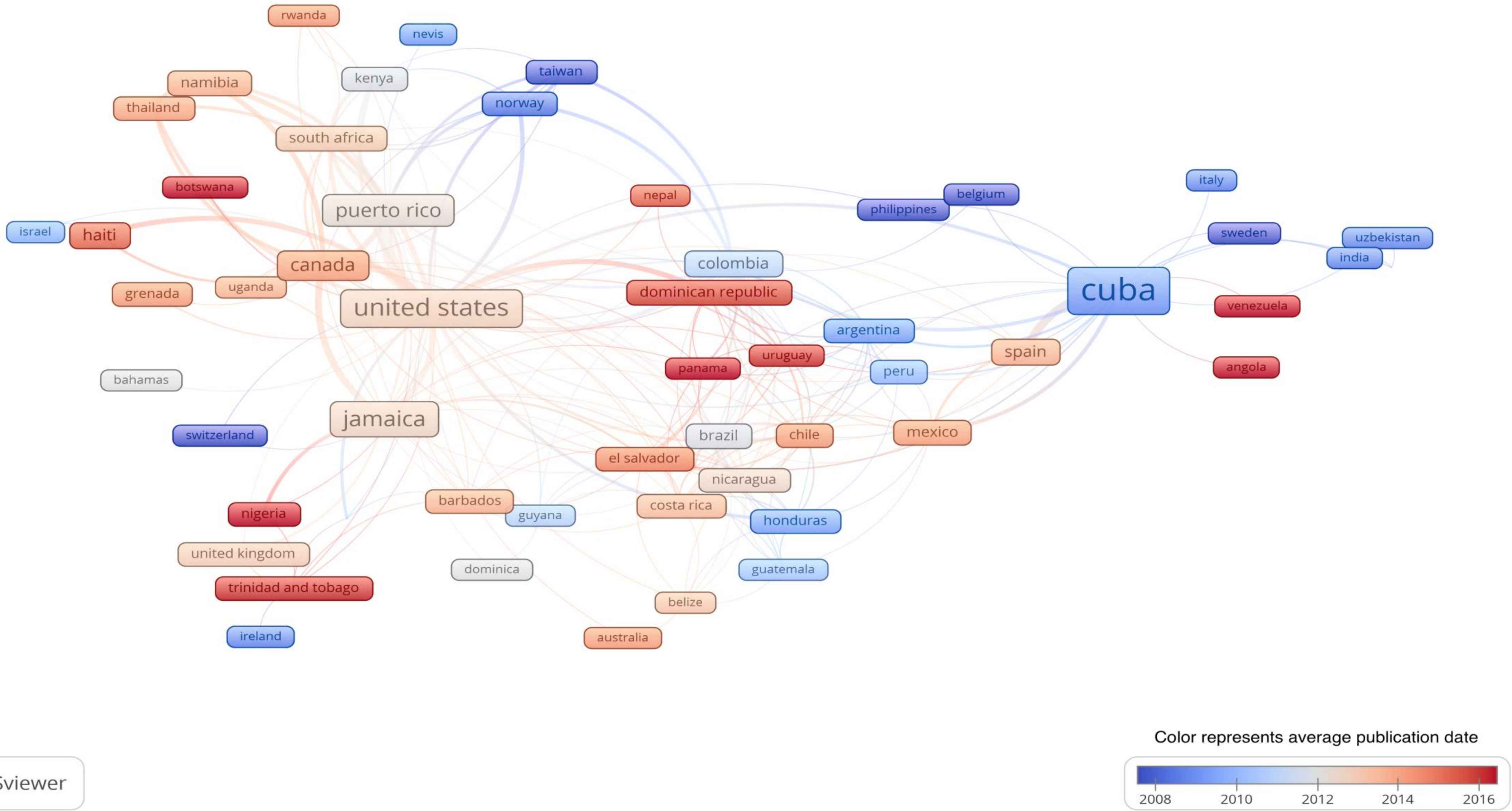
- UNC School of Nursing
- University of the West Indies School of Nursing
- UNC Health Sciences Library

Research Questions

Within the Caribbean region:

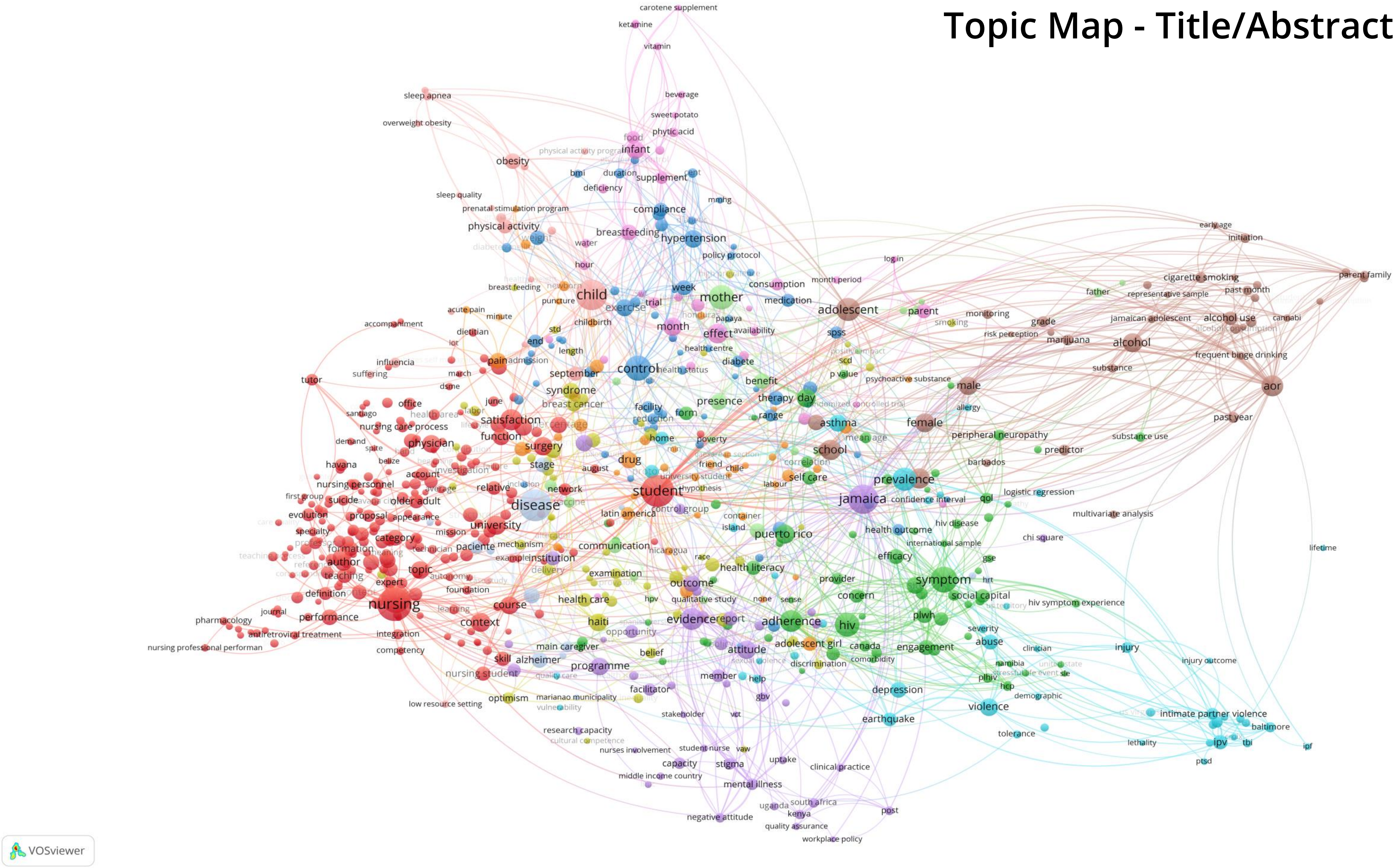
- Which organizations support nurse research?
- What is the level of research output, from the region and by institution?
- What are the areas of research focus among nurse researchers?
- How has the research focus changed over time?

Overview of Country Collaboration



VOSviewer

Topic Map - Title/Abstract Terms

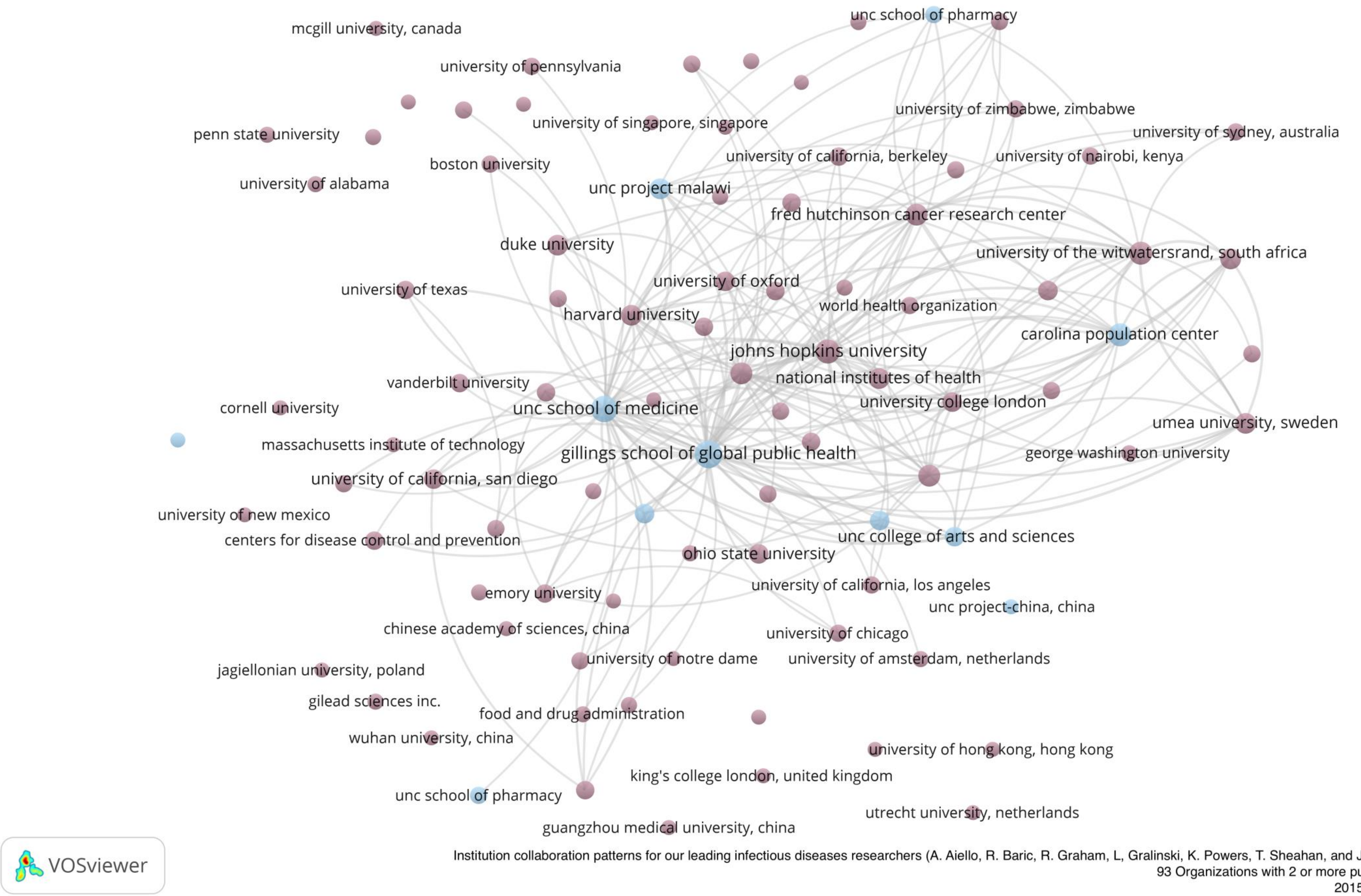


Project Example:
UNC Gillings School of Global Public Health - Infectious Diseases

Research collaborators:

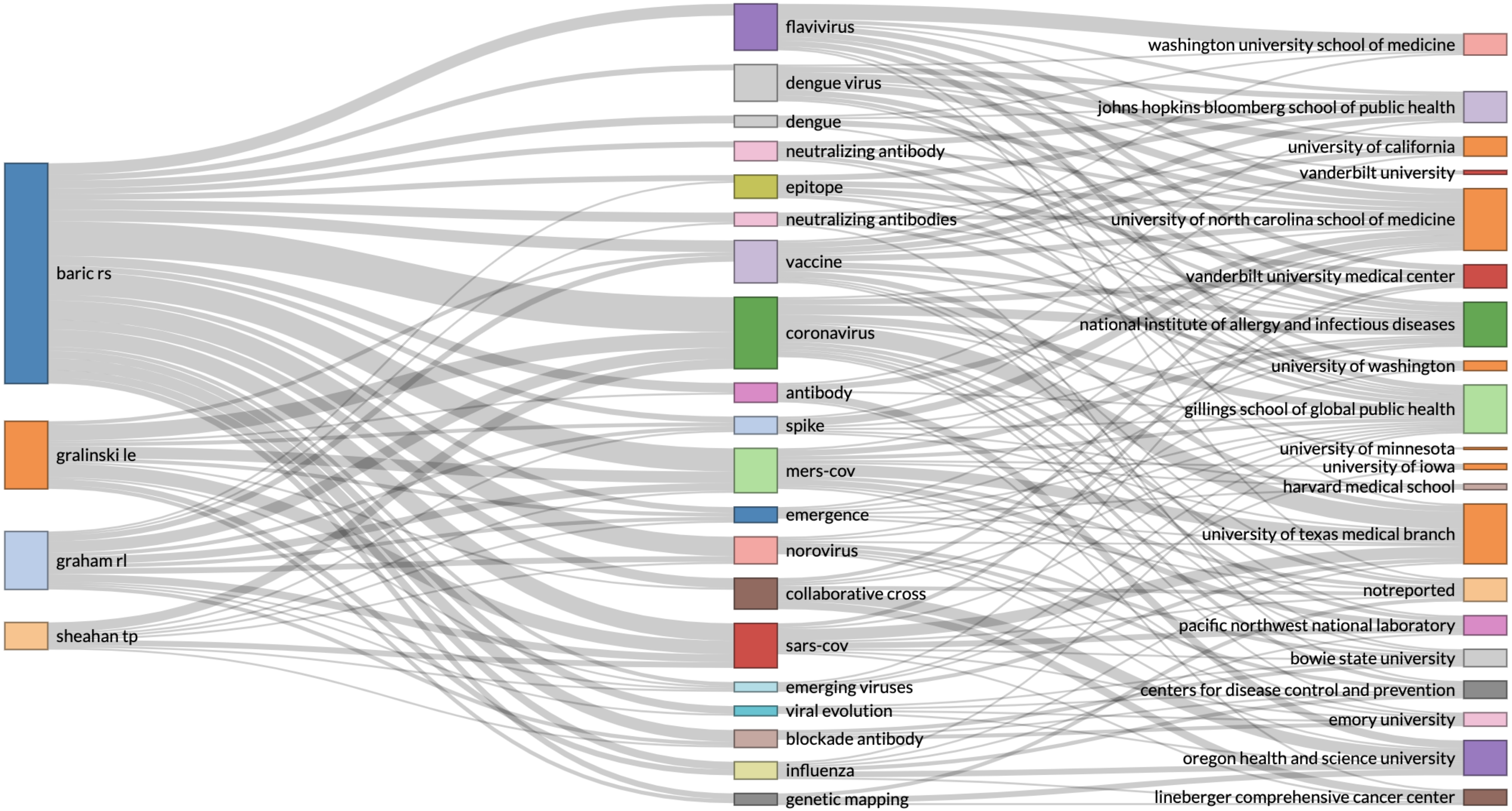
- UNC SPH Research Administration
- UNC Health Sciences Library

UNC Research Collaboration (Infectious Diseases)



Institution collaboration patterns for our leading infectious diseases researchers (A. Aiello, R. Baric, R. Graham, L, Gralinski, K. Powers, T. Sheahan, and J. Stewart)
93 Organizations with 2 or more publications
2015 - present

Three-fields Plot (Author / Keyword / Organization)



Questions?
Comments?

Go To: www.menti.com
Enter Code: 3582 1864

What were the most useful parts of today's presentation?

Mentimeter