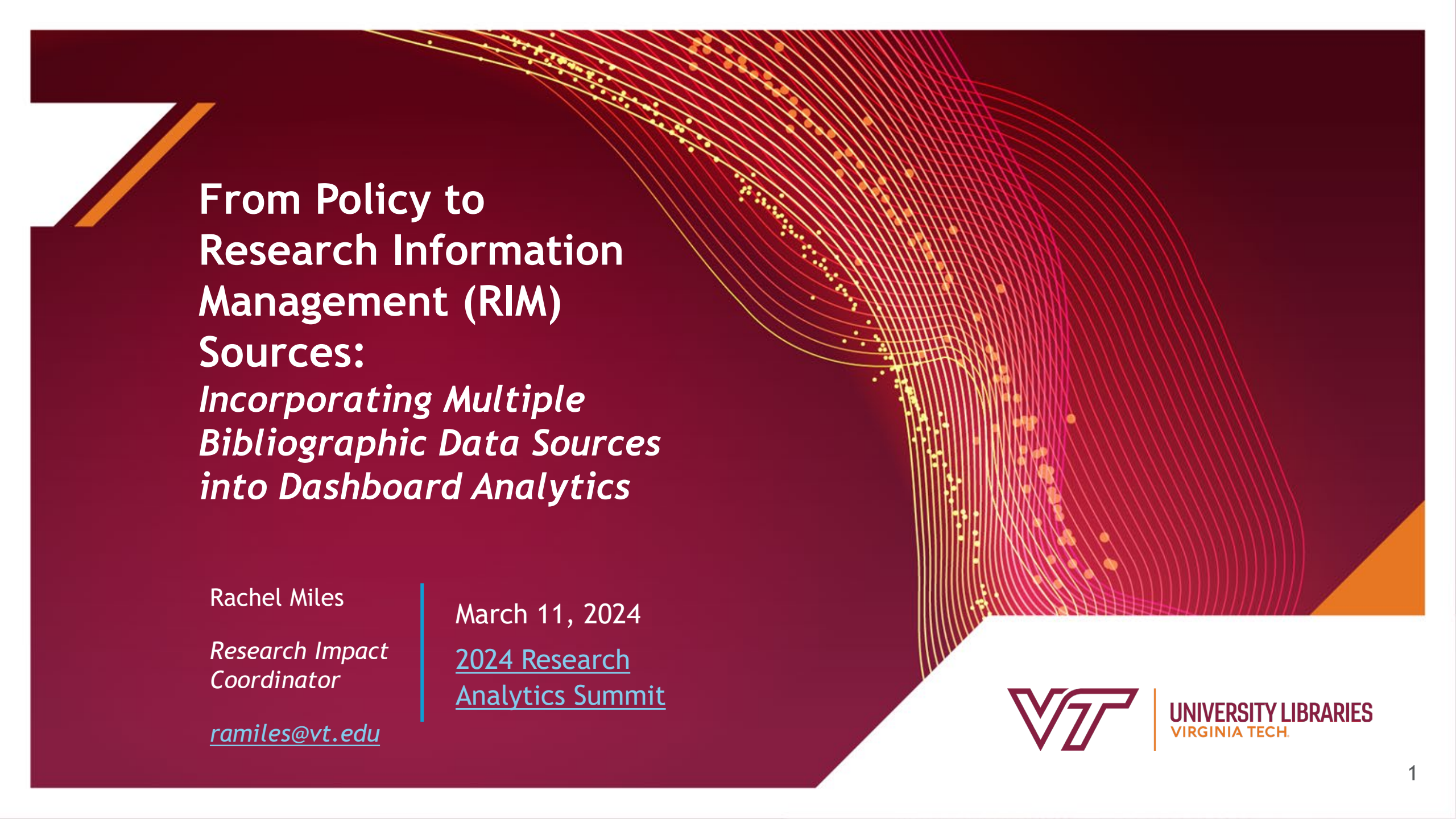




From Policy to Research Information Management (RIM) Sources: *Incorporating Multiple Bibliographic Data Sources into Dashboard Analytics*

Rachel Miles

Research Impact
Coordinator

ramiles@vt.edu

March 11, 2024

[2024 Research
Analytics Summit](#)



UNIVERSITY LIBRARIES
VIRGINIA TECH.

Outline

- Research Information Management Systems: Benefits and Disadvantages
- Exporting RIM data and building dashboards
- Lessons learned from building dashboards with RIM data
- Aggregating bibliographic data from multiple sources
- Comparison of bibliographic data sources to RIM data
- Exporting data from a policy data source & building a dashboard





Research Information Management (RIM) Systems: Advantages & Disadvantages



UNIVERSITY LIBRARIES
VIRGINIA TECH

Advantages of using RIM Systems for Research Analytics

- If a university, college, or department is required to use it, the data can be quite comprehensive
- Includes data from a number of automated sources, databases, optional records from BibTeX or RIS files, as well as manually entered records
- More diverse and customized document types can be included, such as reports, compositions, exhibitions, media products, posters, performances, software, extension publications, etc.
- More inclusive of the arts and humanities
- More inclusive of communities that do not publish as often or that produce more 'grey literature,' such as reports, and conference papers not indexed in bibliographic databases



Disadvantages of using RIM systems for Research Analytics

- Incomplete and low-quality data on manually entered records
- Reliance on faculty to input their data and maintain their scholarly record accurately and completely*
- Limitations on the types of data and fields exported from RIM systems for analytics purposes
- Duplicates in the system as a result of manual entry from multiple users as well as automated importing for multiple users**
- Difficulties parsing data for analytics in bibliometric software tools, such as Bibliometrix and VOSviewer

*For the college that is analyzed in today's example, we did profile curation services, so we know that the data is fairly complete and accurate.

**Can be eradicated, but valuable fields and data are lost in the process





Exporting RIM data and Building Dashboards



UNIVERSITY LIBRARIES
VIRGINIA TECH

Some background

- Symplectic Elements used at Virginia Tech for 10+ years
- Elements Implementation Team = partnership between the Provost's Office and University Libraries
- Elements only recently required by *all* colleges at the university
 - All colleges implement this requirement differently



Exporting RIM Data from Elements (Symplectic Elements, from Digital Science)

- Basic Report generated / exported for **College X**
- Filters:
 - Date from: **01/01/2014**
 - Date to: **01/01/2024**
 - Include non-current users: **Yes**
- Select data export or report:
 - Object category: **Scholarly & creative works (linked to the selected users)**
 - Scholarly & creative work type: **All**
 - Return (first report): **Scholarly & creative works by linked user**
 - Returns all works entered by each user; duplicates exist
 - Return (second report): **Scholarly & creative works**
 - Returns all scholarly & creative works, no duplicates (supposedly)



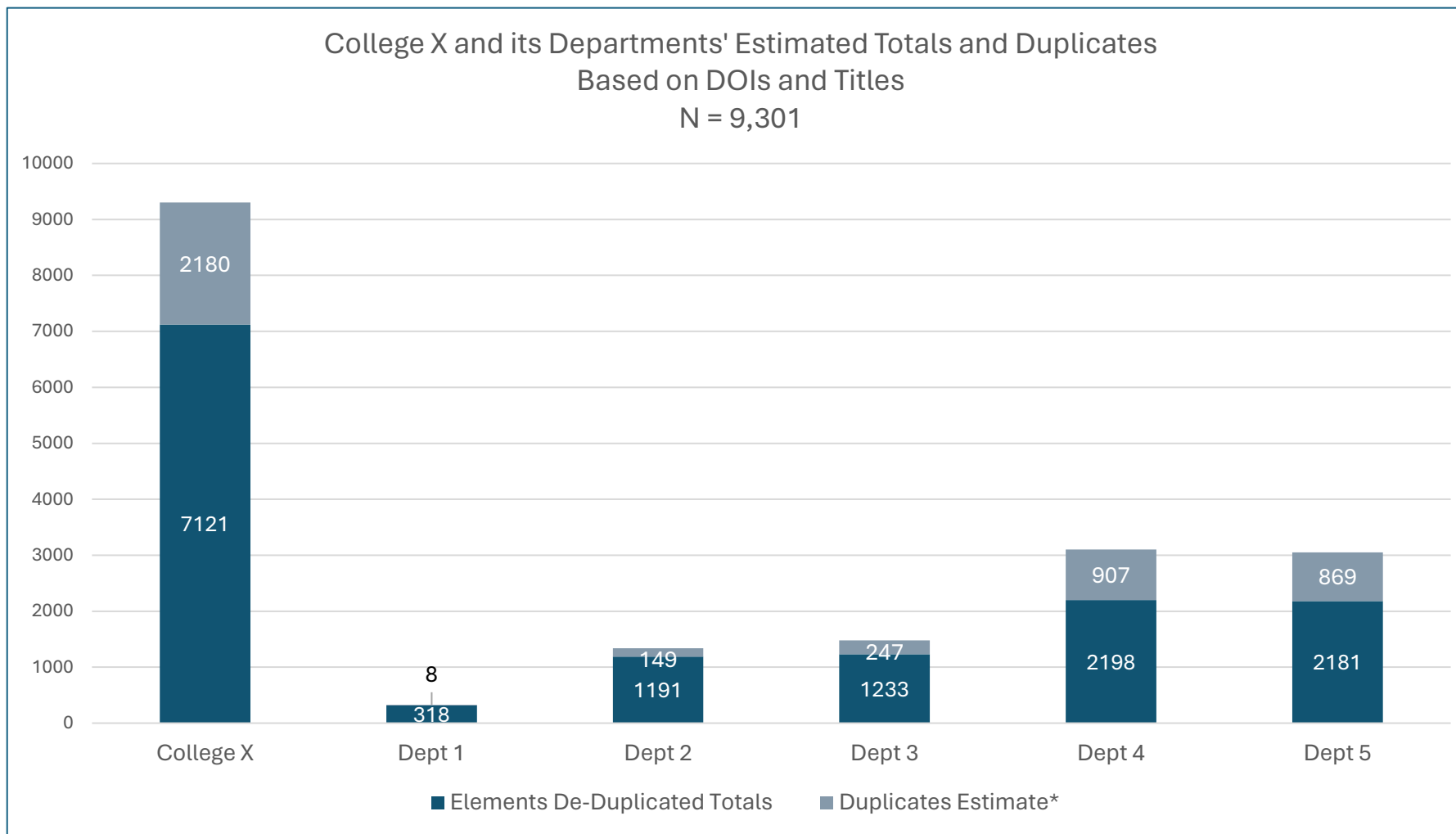
First report: Scholarly & creative works by linked user

Many useful fields and values for research analytics & building dashboards (next slide)

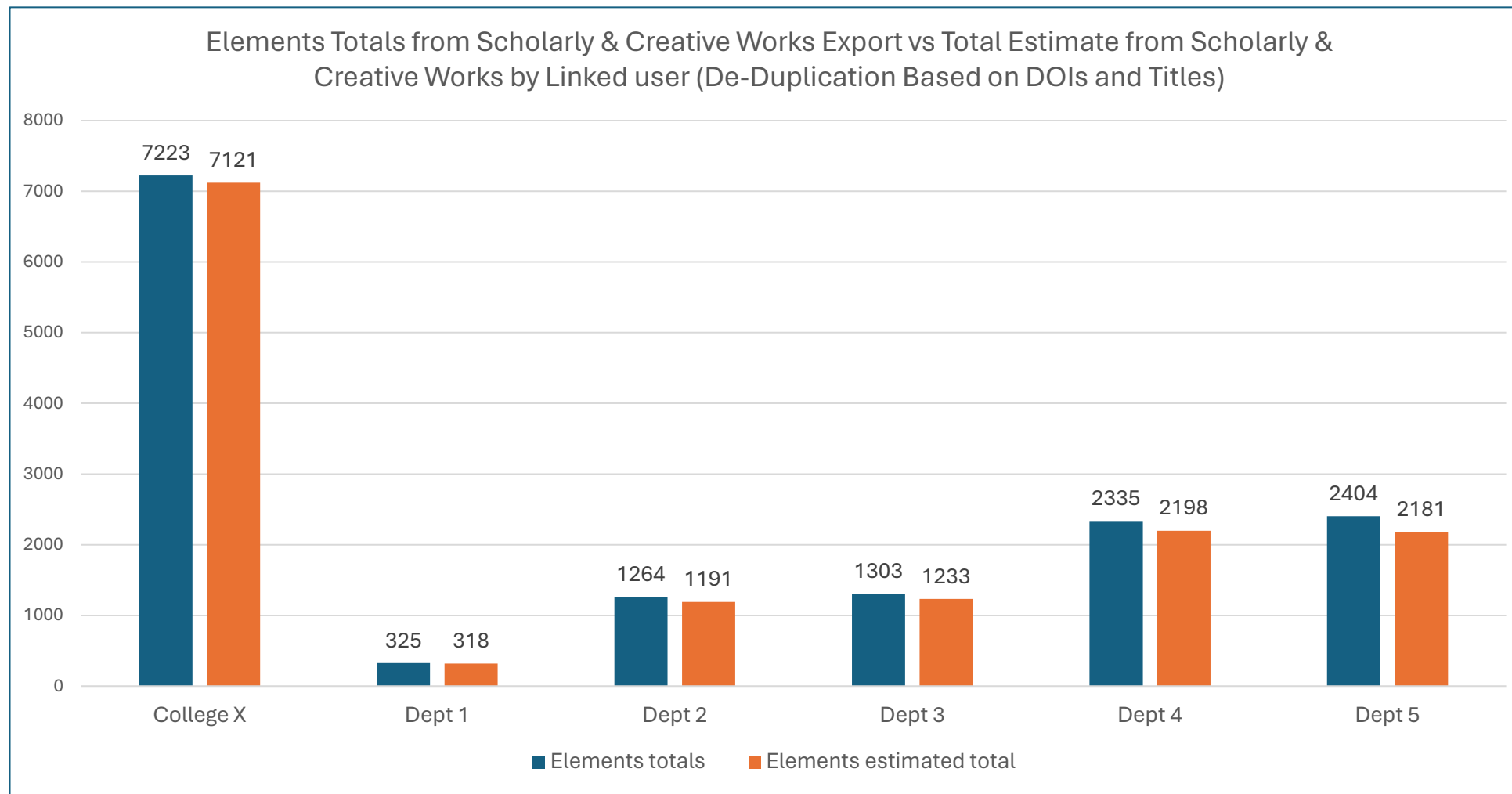
Unfortunately, because it is based on the linked user, there are thousands of duplicates for one college over a ten-year period



Estimate of Duplicates in Elements for College X, 2014-2024



Elements Totals Based on Report 2: Scholarly & creative works (not by linked user) vs. Elements Estimated Totals from Report 1



Comparison of Data & Fields Exported

Report 1:
Scholarly &
creative works by
linked user

- 152 fields

Report 2:
Scholarly &
creative works

- 106 fields



Report 1: Scholarly & creative works by linked user:

Unique fields

Name
Username
Email
User's Proprietary ID
Primary group descriptor
Primary group
Is current staff
Visible
Favourite
Crossref record exists?
DBLP record exists?
Digital Commons record exists?
Dimensions record exists?
Dimensions for Universities record exists?
Elements record exists?
EPrints record exists?
Europe PubMed Central record exists?
figshare.com record exists?
Google Books record exists?
Hyrax record exists?
ISI Proceedings record exists?
Local Source 1 record exists?
Local Source 2 record exists?

Manual record exists?
MLA record exists?
ORCID record exists?
PubMed record exists?
R3 record exists?
RePEc record exists?
SciVal Experts record exists?
Scopus record exists?
SSRN record exists?
Virginia Tech Data Repository data.lib.vt.edu record exists?
VTechWorks record exists?
Web of Science record exists?
Web of Science (Lite) record exists?
Not externally funded
Times cited (Web of Science)
Times cited (Scopus)
Times cited (Dimensions)
Times cited (Europe PubMed Central)
Relative citation ratio (Dimensions)
Field citation ratio (Dimensions)

'Times
Cited' in
Report 2*

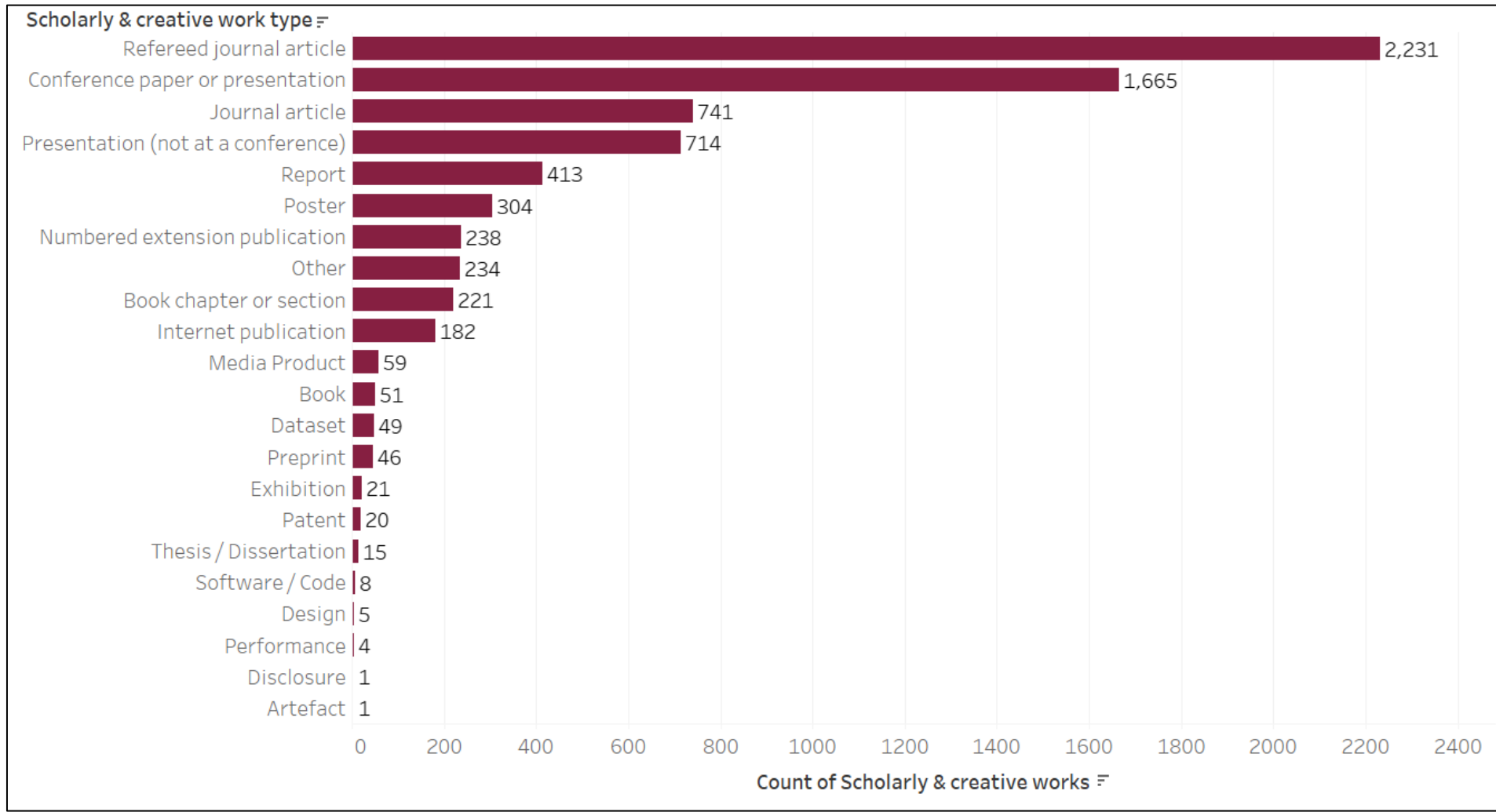
**Pulls the citation count based on the precedence we setup in the system, which has manual set as the top followed by automated sources.*

Dashboard Building in Tableau with Elements Data for College X

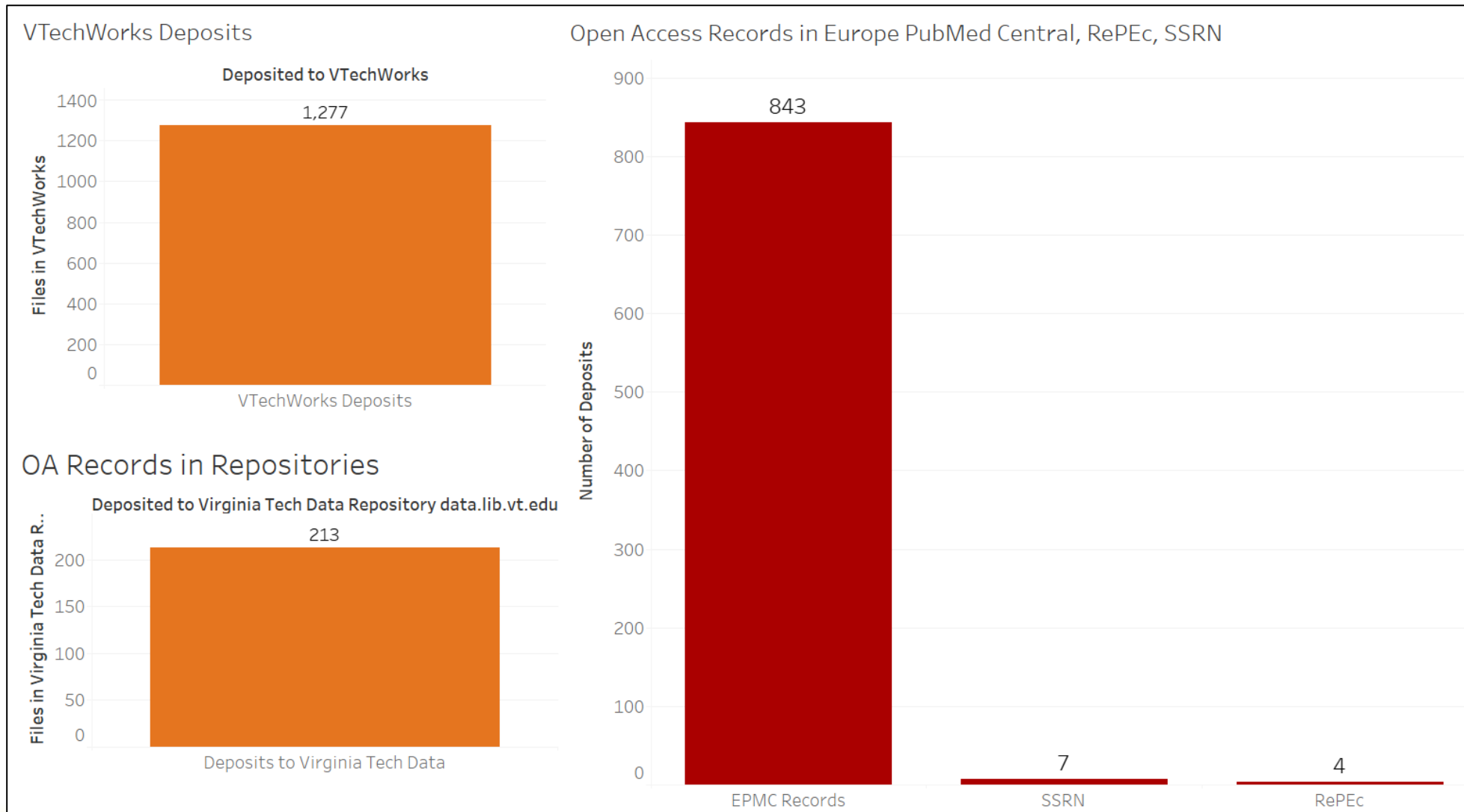
- Types of Scholarly Works (Report 2)
- VTechWorks (institutional repository) deposits (Report 1, inflated)
- Open Access records in other repositories (limited) (Report 1, inflated)
- Publication and citation counts by bibliographic data source (Report 1, inflated)
- Commissioning Bodies (Report 2)
- Fields of Research (FoR)
- Conferences (Report 2)
- Books & Reports (Report 1, inflated)
- Journal Metrics (Report 1, simple metrics)
- Term Co-Occurrence of Titles & Abstracts VOSviewer Network (Report 1)
- Fields of Research VOSviewer Network (Report 1)



Scholarly & Creative Work Types for College X, 2014-2024



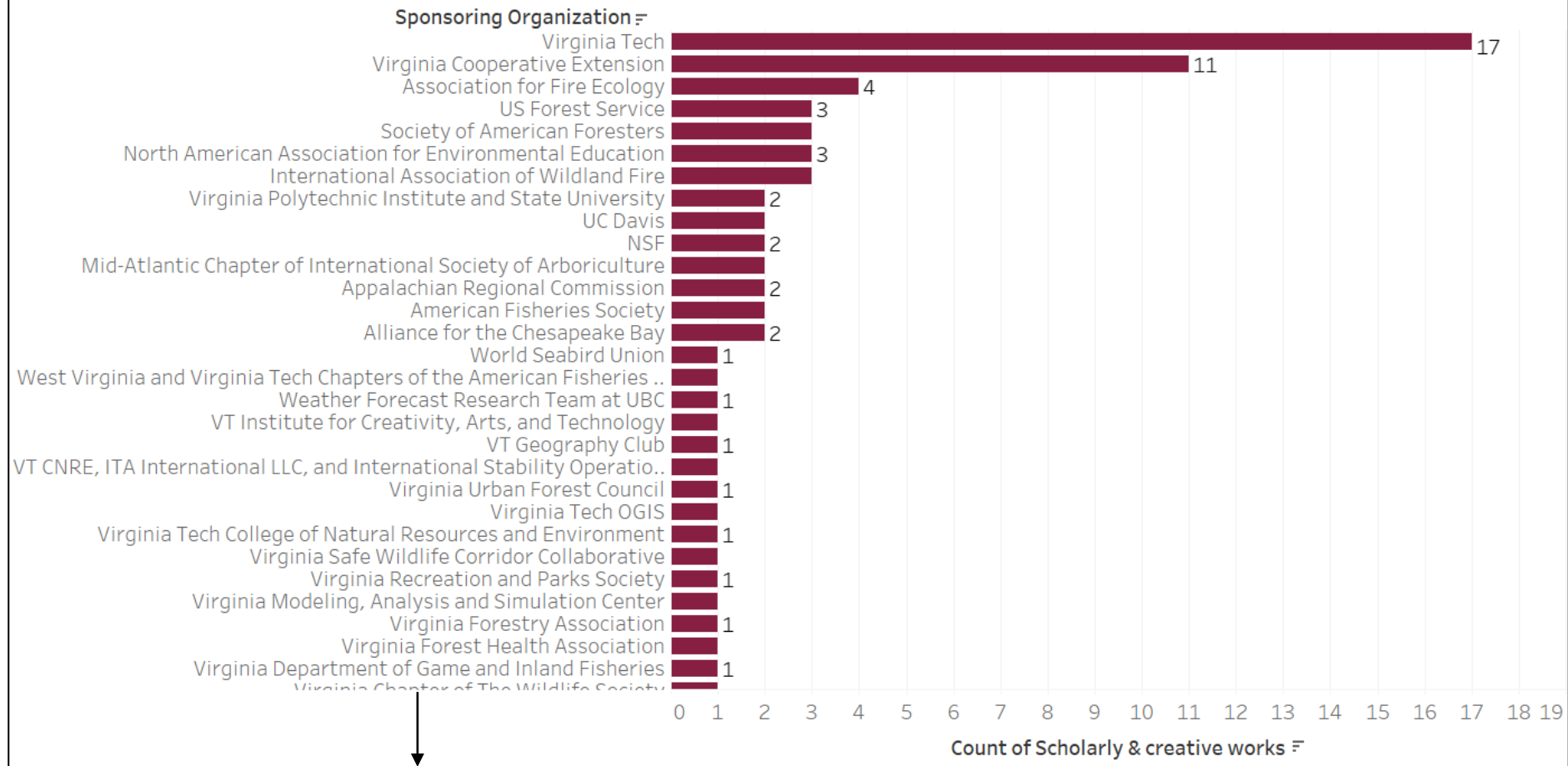
Open Access Deposits by Source



Sponsored / Funding Organization by Scholarly Work

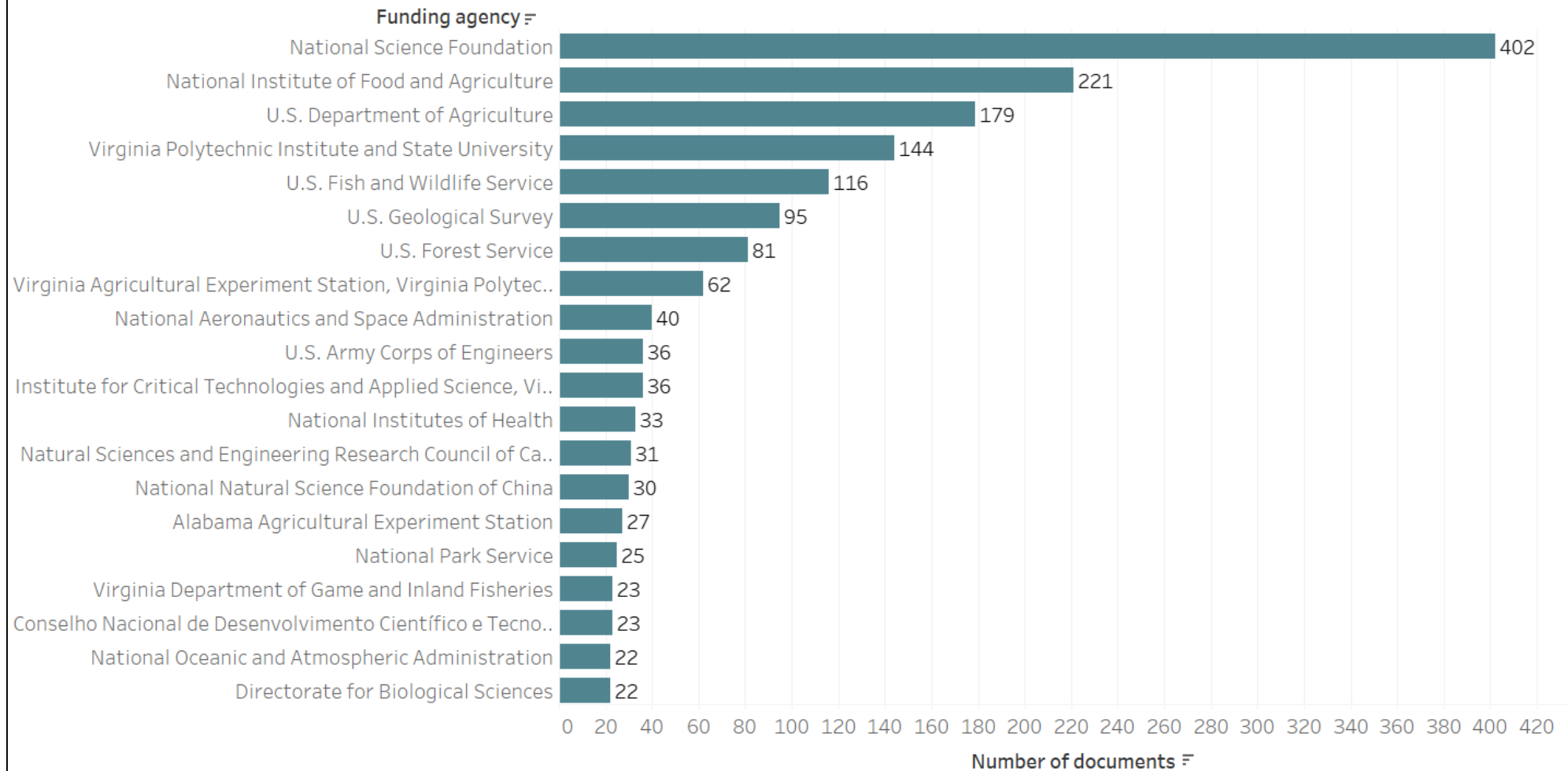
Sponsored Organization by Publication Count, 2014-2024

(likely not entirely accurate because it is self-reported in Elements but shows more diverse and smaller and/or internal organizations)



Comparing Funding Agencies to Scopus Data

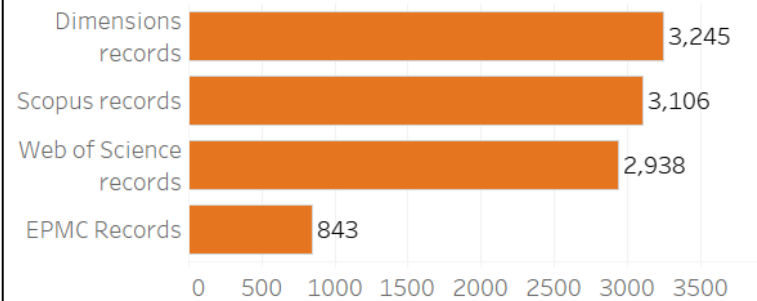
College X Funding Agencies by Publication Numbers (Scopus Data *only*)



Publication & Citation Counts by Bibliographic Data Source

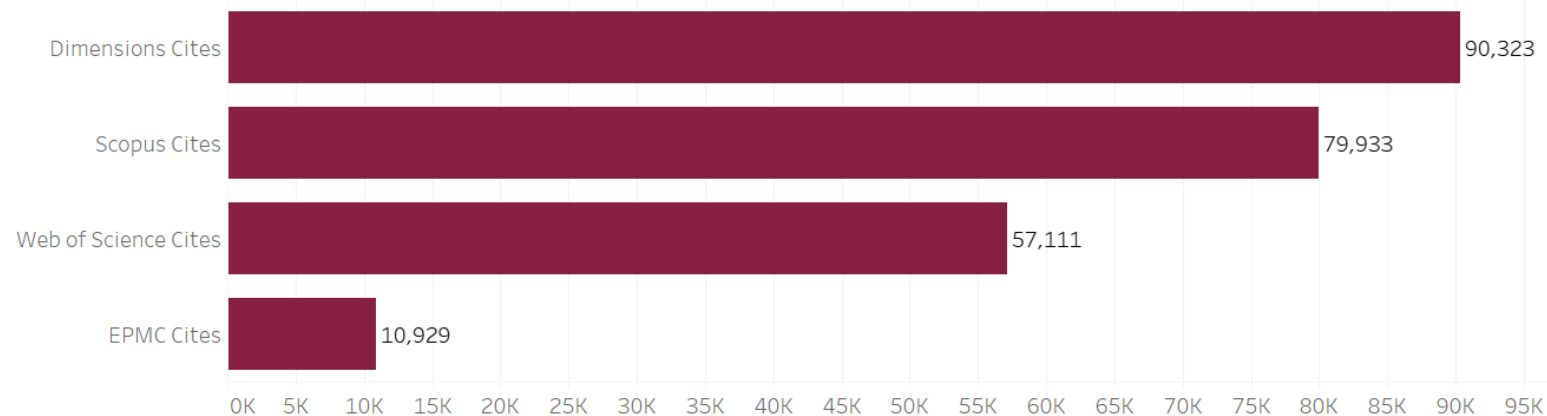
Publication Counts by Bibliographic Data Source

(data may be slightly inflated due to duplicates in the system due to multiple authors reporting the same record)



Citation Counts by Bibliographic Data Source

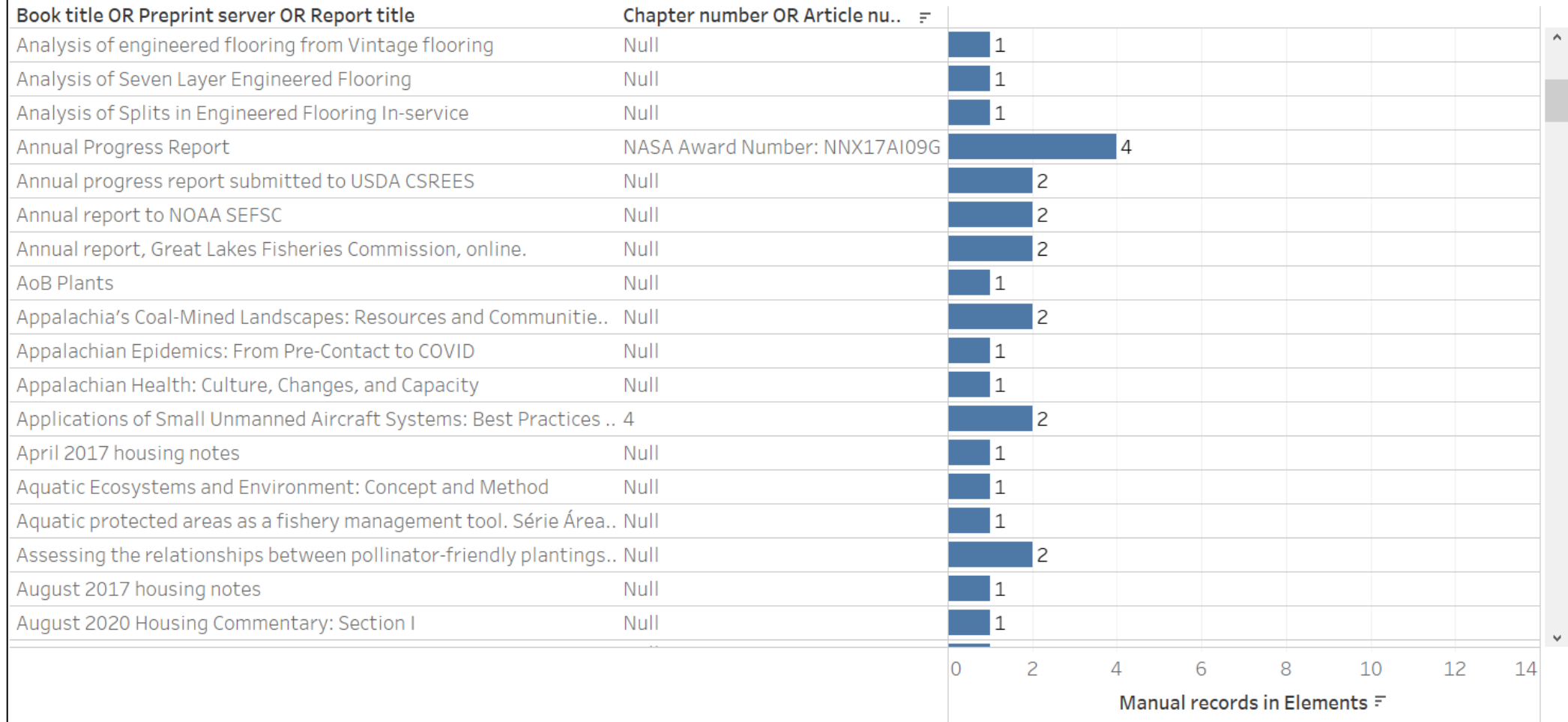
(data may be slightly inflated due to duplicates in the system due to multiple authors reporting the same record)



Books & Reports

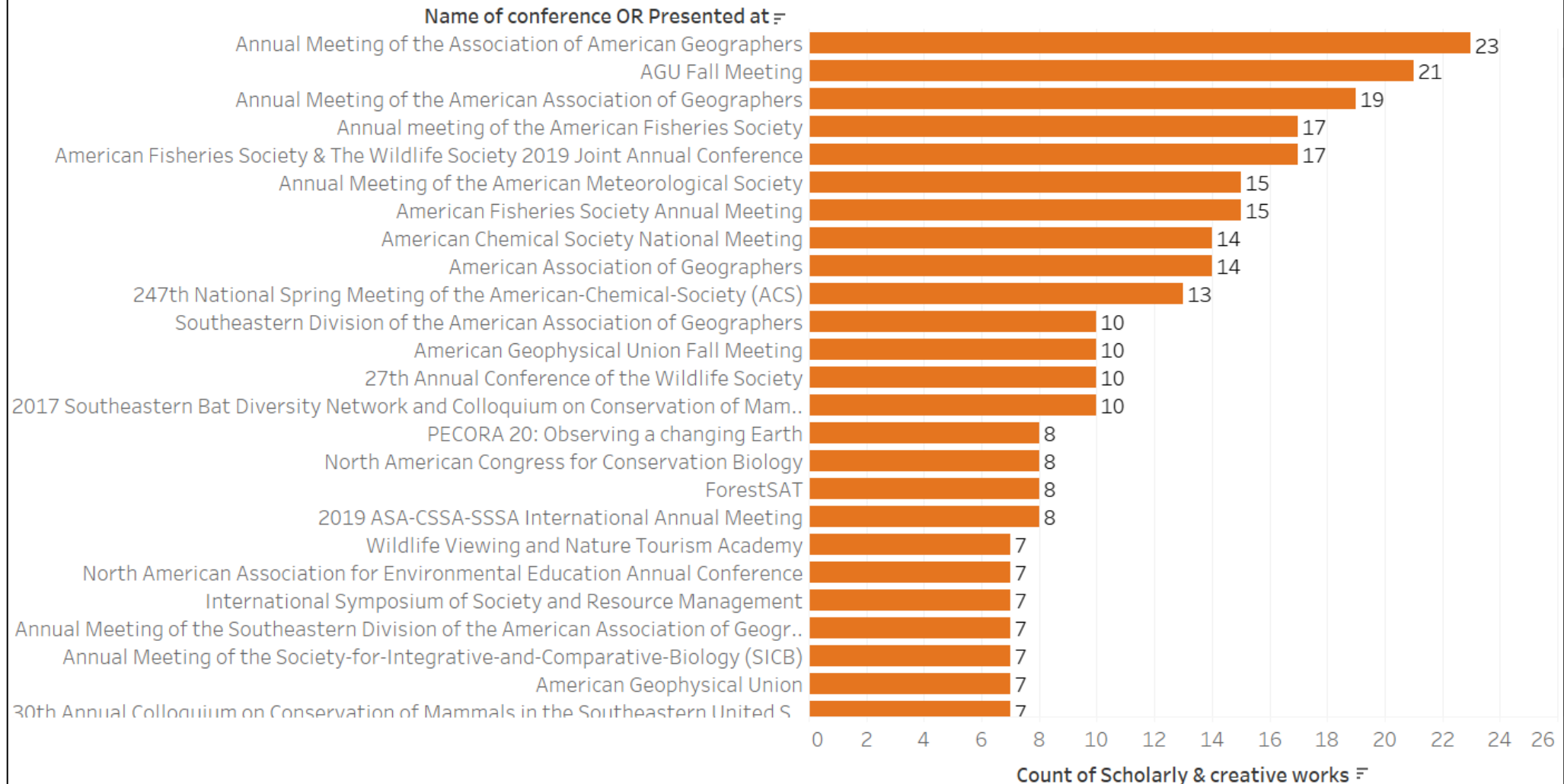
Books & Reports as reported by faculty

(data may be slightly inflated due to duplicates in the system due to multiple authors reporting the same record)



Conferences

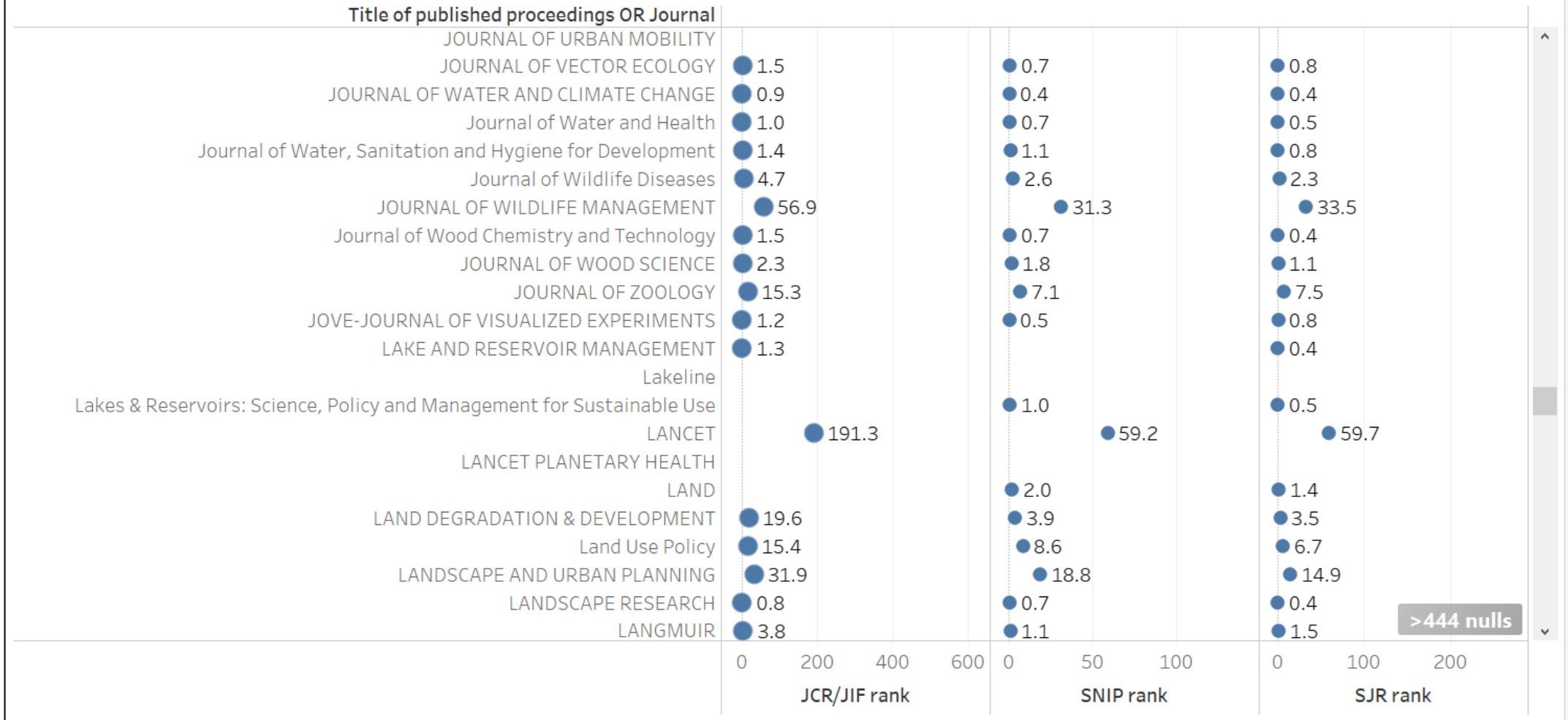
Conferences



Journal Metrics

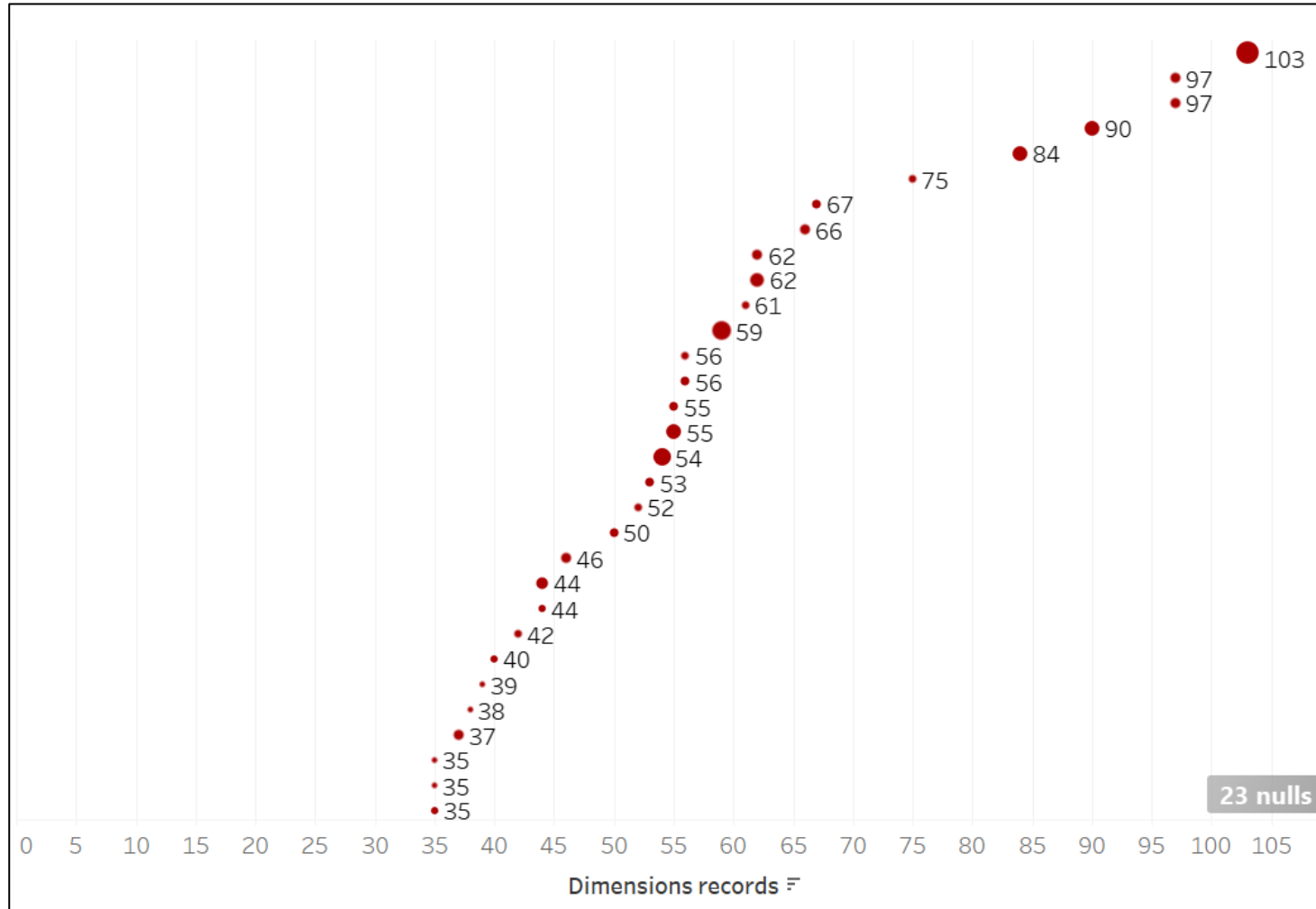
Journal Metrics

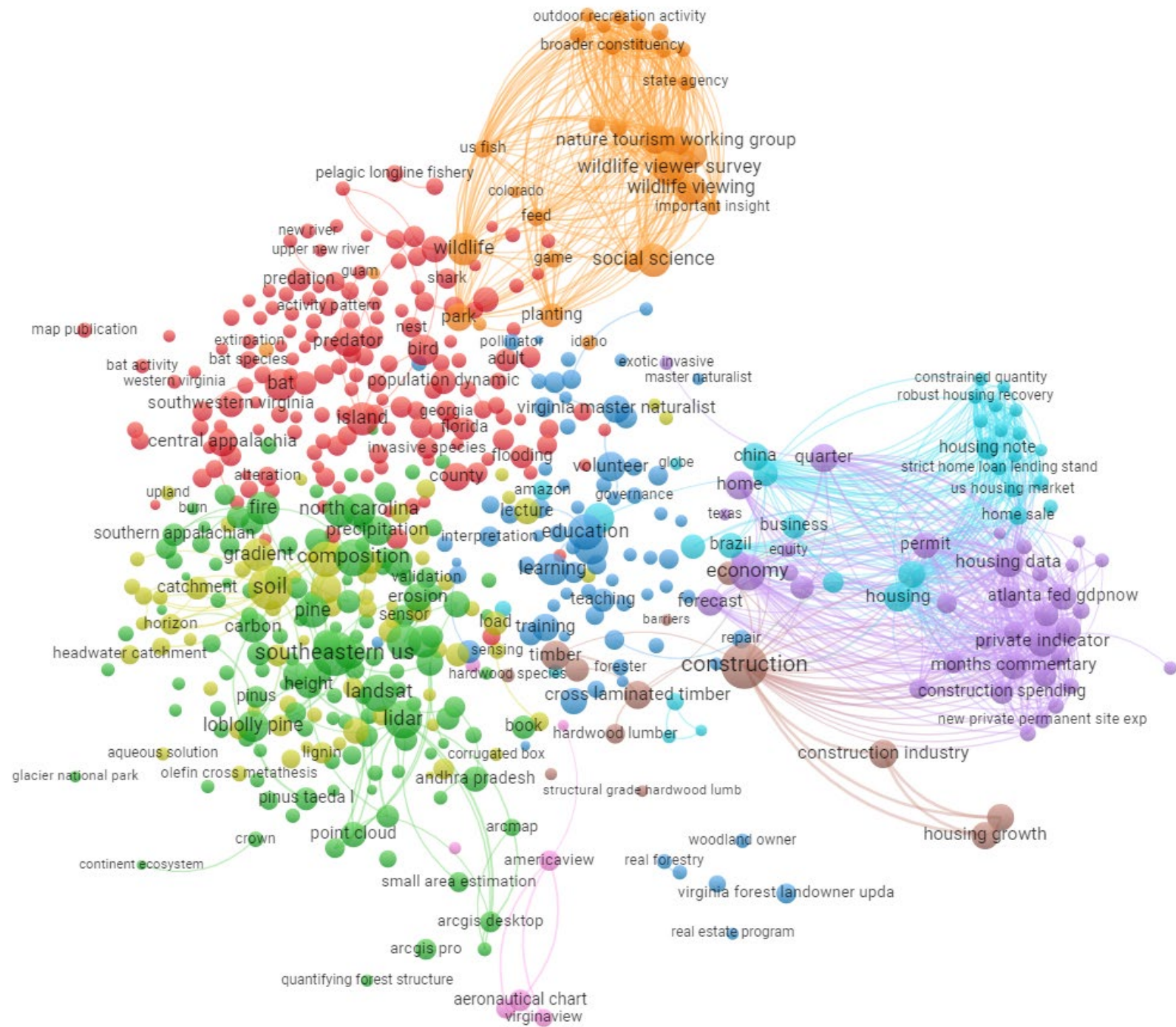
JCR, SNIP, & SJR all correlate with one another. Not all journals are ranked, especially for JCR/JIF.

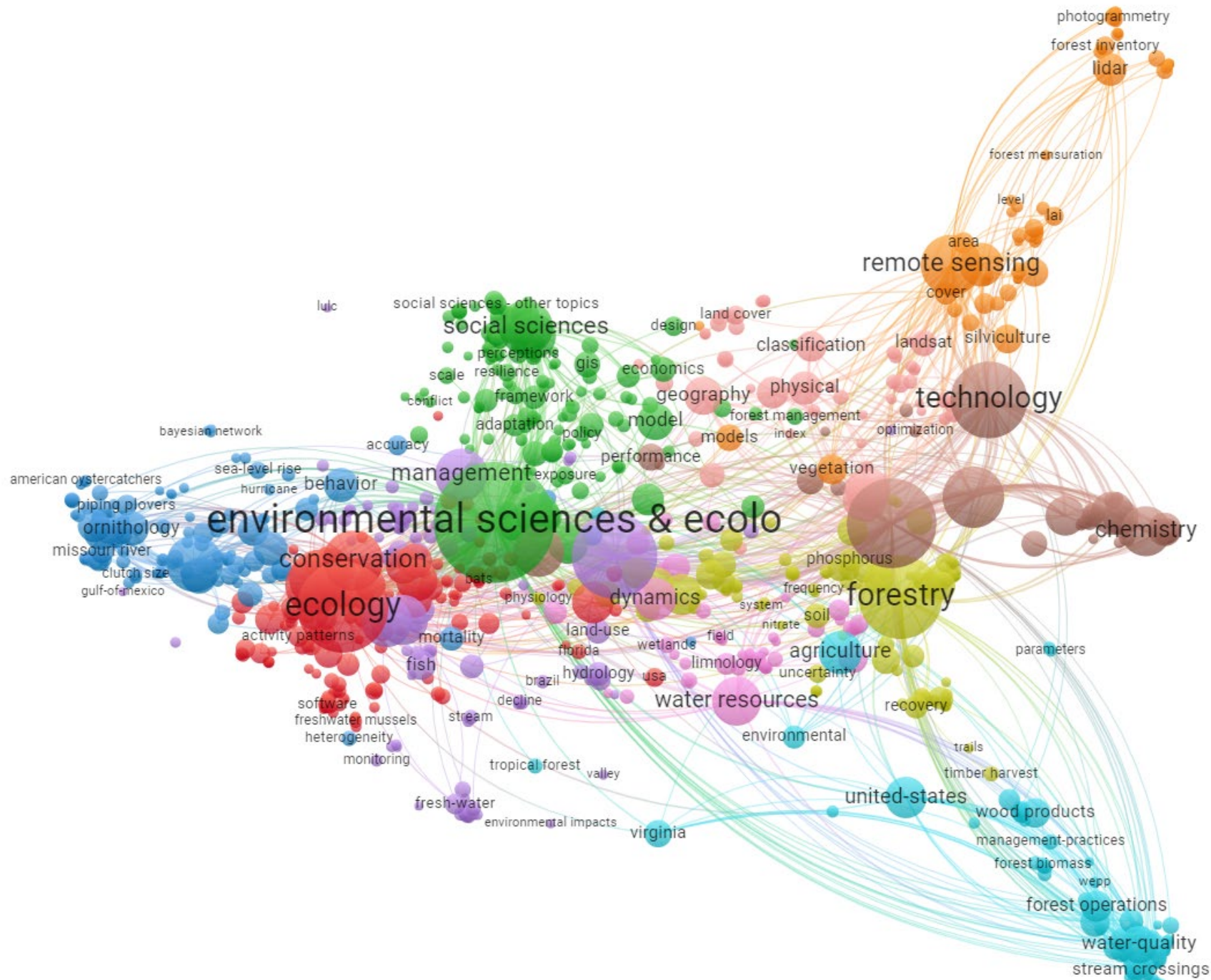


Publication & Citation Counts by Faculty Member (names not displayed)

Size of node = Number of Citations









Lessons learned from building dashboards with RIM data



UNIVERSITY LIBRARIES
VIRGINIA TECH

Lessons Learned: Using RIM Data for Research Analytics

- Expect some low-quality data, especially from manually entered records
- In our use case, manual records are set to the highest precedence for faculty activity reporting purposes, which means that automated records with high quality metadata that have manual records attached may now have low quality metadata
- Duplicates are inevitable and may (will) muddy the data, depending on how the data is exported and used.
- In the case of Elements, the superior data export means more duplicates
 - I did not de-duplicate Report 1 data due to time constraints; instead, some analytics were appropriate without de-duplicating, such as the publication and citation counts by faculty member / author.
 - De-duplicating may be inappropriate in the case of titles without DOIs, in which many of those duplicate titles are unique works with the same titles.
 - Some analytics were not very affected by the duplicates, such as the term co-occurrence, fields of research visualizations, and journal metrics.
 - Where it was appropriate, Report 2 was used to report more accurate numbers.



Lessons Learned: Using Elements Data to Build Tableau Dashboards

- Overwhelming amount of data and potential uses of the data
 - However, limitations on the data (as discussed).
- Not as easy as ‘clicking a button’ in a research analytics tool and getting a result (of course)
- Benefits the college administration: they get a better understanding of what their faculty are doing beyond peer reviewed publications (plus publications not indexed in databases)
- Does not link up to other points of data; e.g., references are not part of the data export from Elements, which is a limitation for doing citation analyses.
 - So far, most of the ‘analytics’ are basic counts of different types of scholarly works, but it has potential to be linked with citation impact, depending on the inquiry or request.
 - Caveat: smaller data sets should be used cautiously with citation data in terms of making conclusions about ‘citation impact,’ such as the ‘Field Citation Ratio (FCR)’ which can be exported from Elements, especially when using multiple sources of publication and citation data in one data set.
 - Have not yet attempted these analytics, since more care needs to be taken before doing so.
- This has been a huge learning experience, and I can only imagine it will improve from here!



Aggregating bibliographic data from multiple sources



UNIVERSITY LIBRARIES
VIRGINIA TECH

Why aggregate data from more than one bibliographic data source?

Comprehensiveness!

Demonstrate the uniqueness of each database

Create more work for yourself

Study: Export data from Web of Science & Scopus



Publication data exported for College X on college-level and by department from both Web of Science and Scopus

Data de-duplicated and compared

Unique records from Scopus added to Web of Science files for visualizing more data

Metadata Crosswalk

Web of Science Heading / Field	Scopus Heading / Field
AU	Authors
TI	Title
SO	Source Title
DT	Document Type*
LA	Language of Original Document
CT	Conference name
CY	Conference date
CL	Conference location
DE	Author Keywords
ID	Index Keywords
AB	Abstract
C3	Affiliation
FU	Funding details
CR	References
TC or Z9**	Cited by
PU	Publisher
SN	ISSN
PY	Year
VL	Volume
IS	Issue
DI	DOI

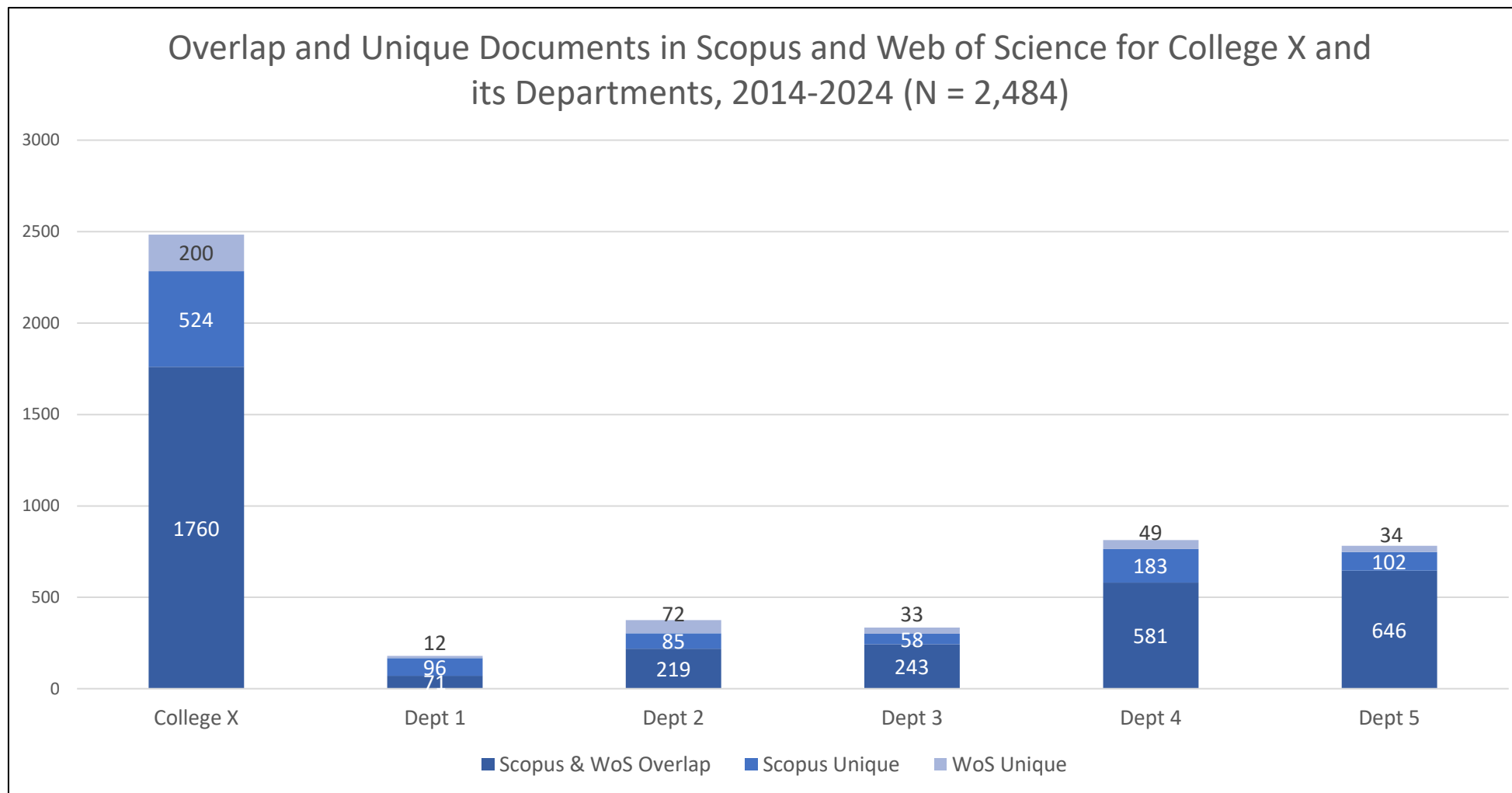
*Certain document types have to be changed, e.g., “Conference paper” is used in Scopus whereas “Proceedings Paper” is used in Web of Science.

**Z9 is the total times cited across all Web of Science databases. TC is the total times cited across the Web of Science Core Collection.

For details on all the Web of Science Core Collection field tags, see

https://images.webofknowledge.com/images/help/WOS/hs_wos_fieldtags.html

Overlap & Unique Documents in Scopus & WoS for College X, 2014-2024

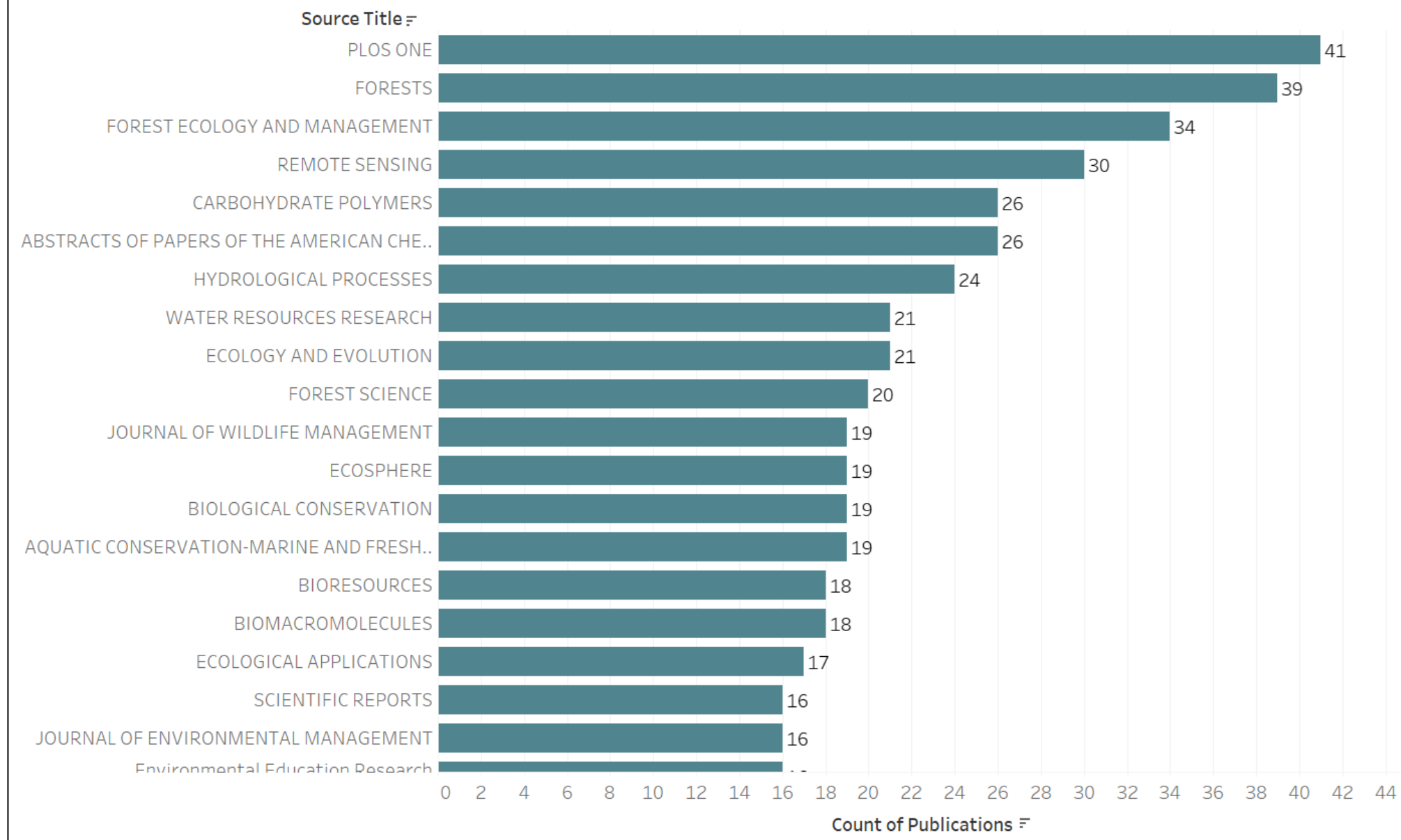




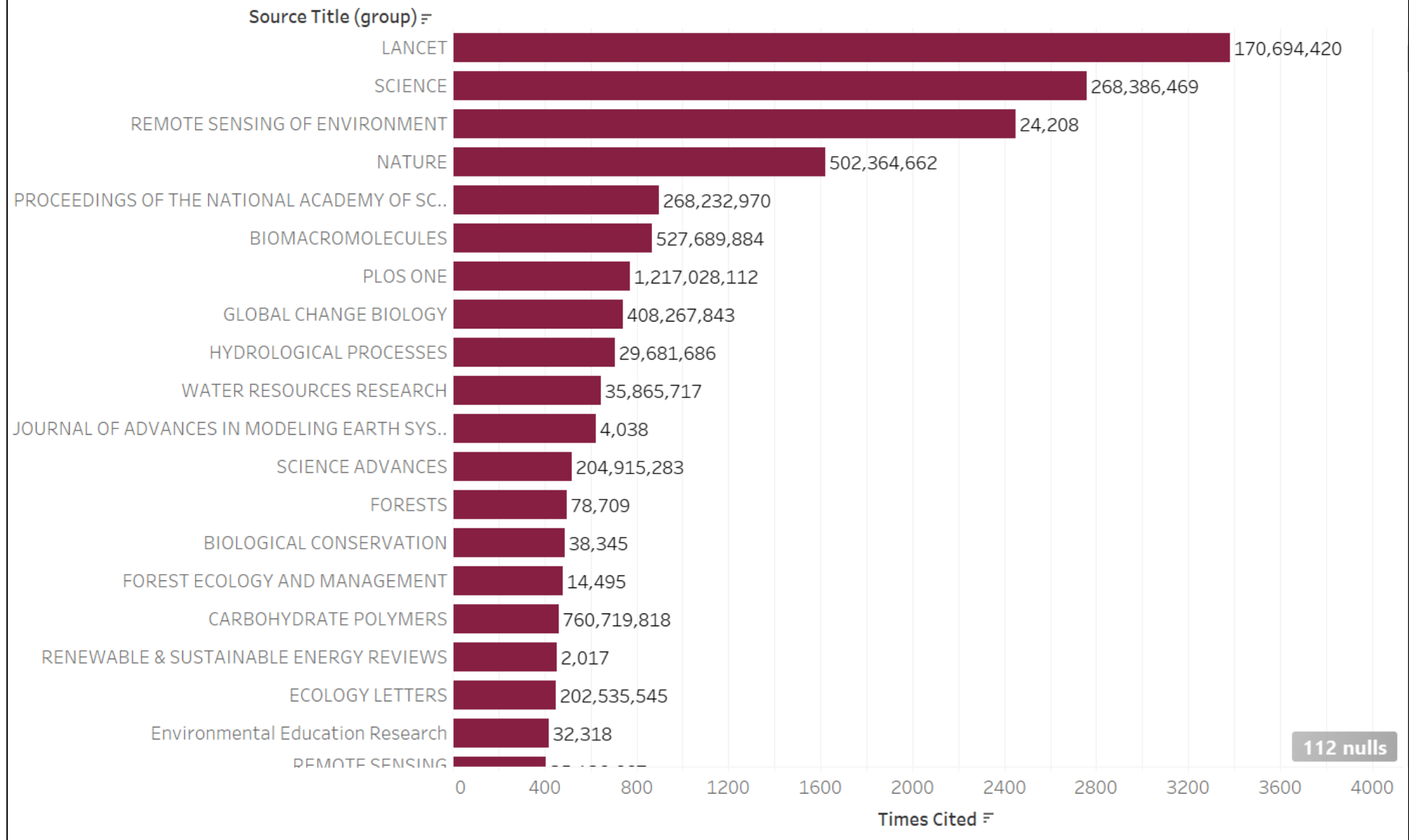
Examples of Graphs and Visualizations from the Aggregated Data



Pubs by Source Title

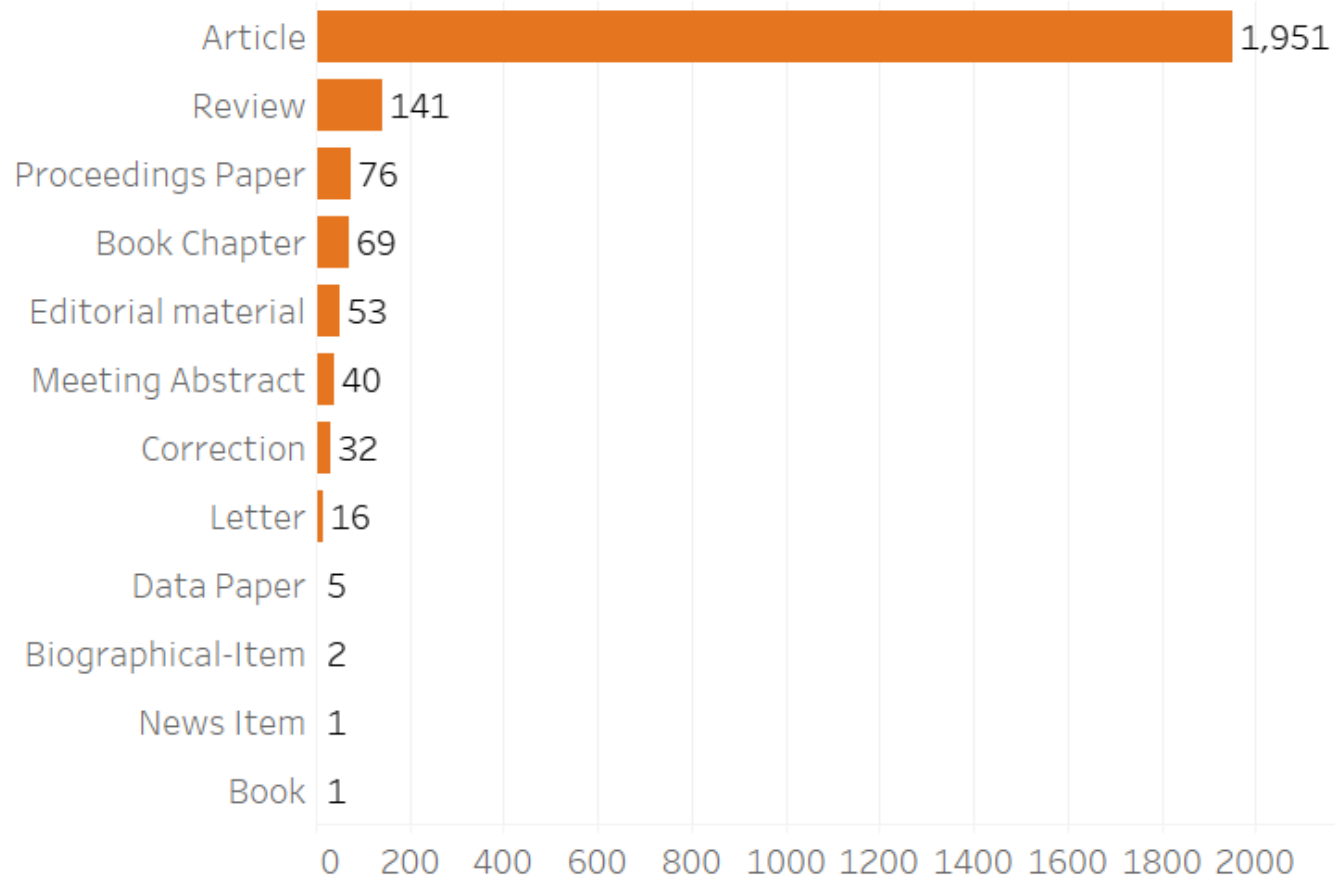


Total Citations by Journal

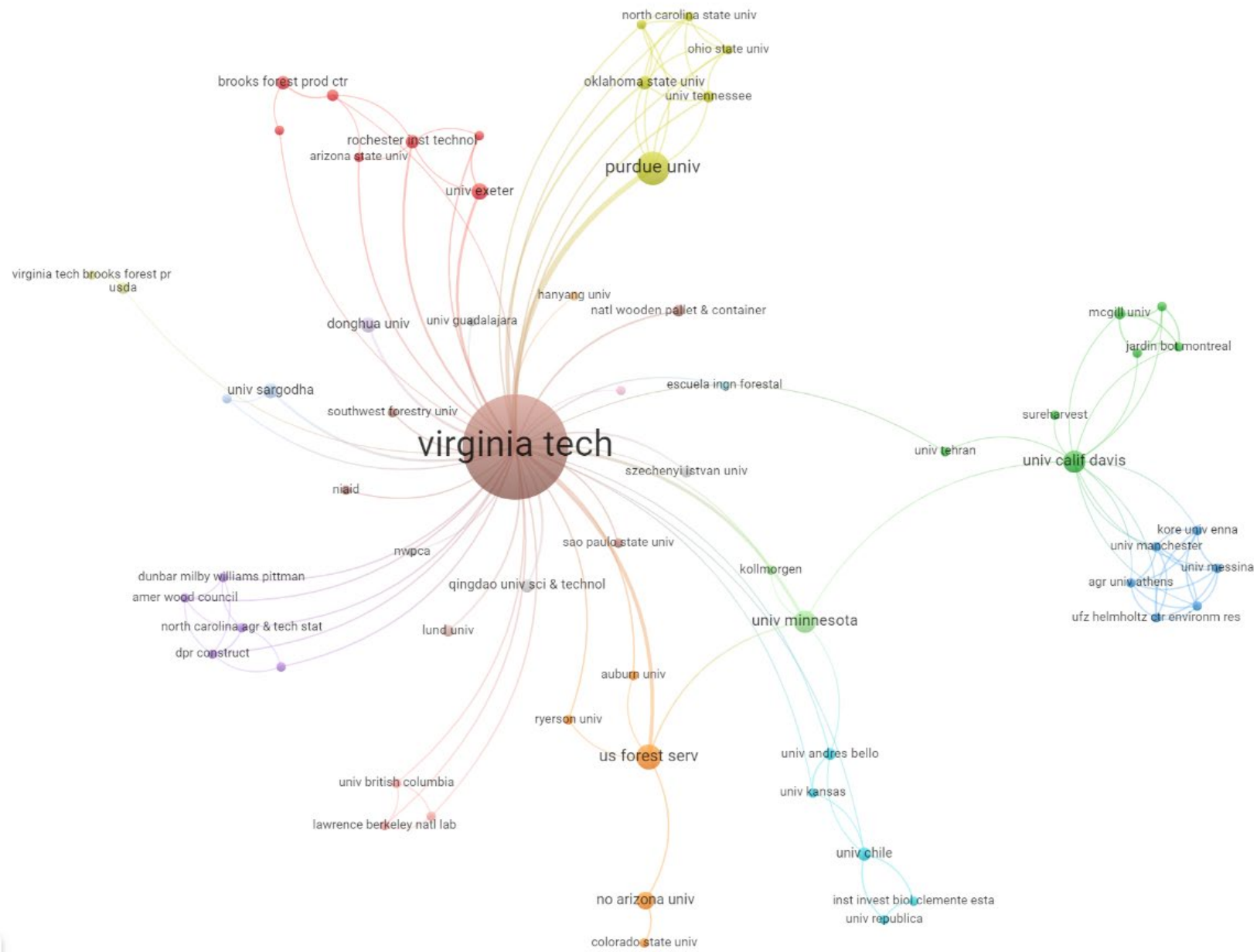


Document Types

Document Type =



Count of Publications =

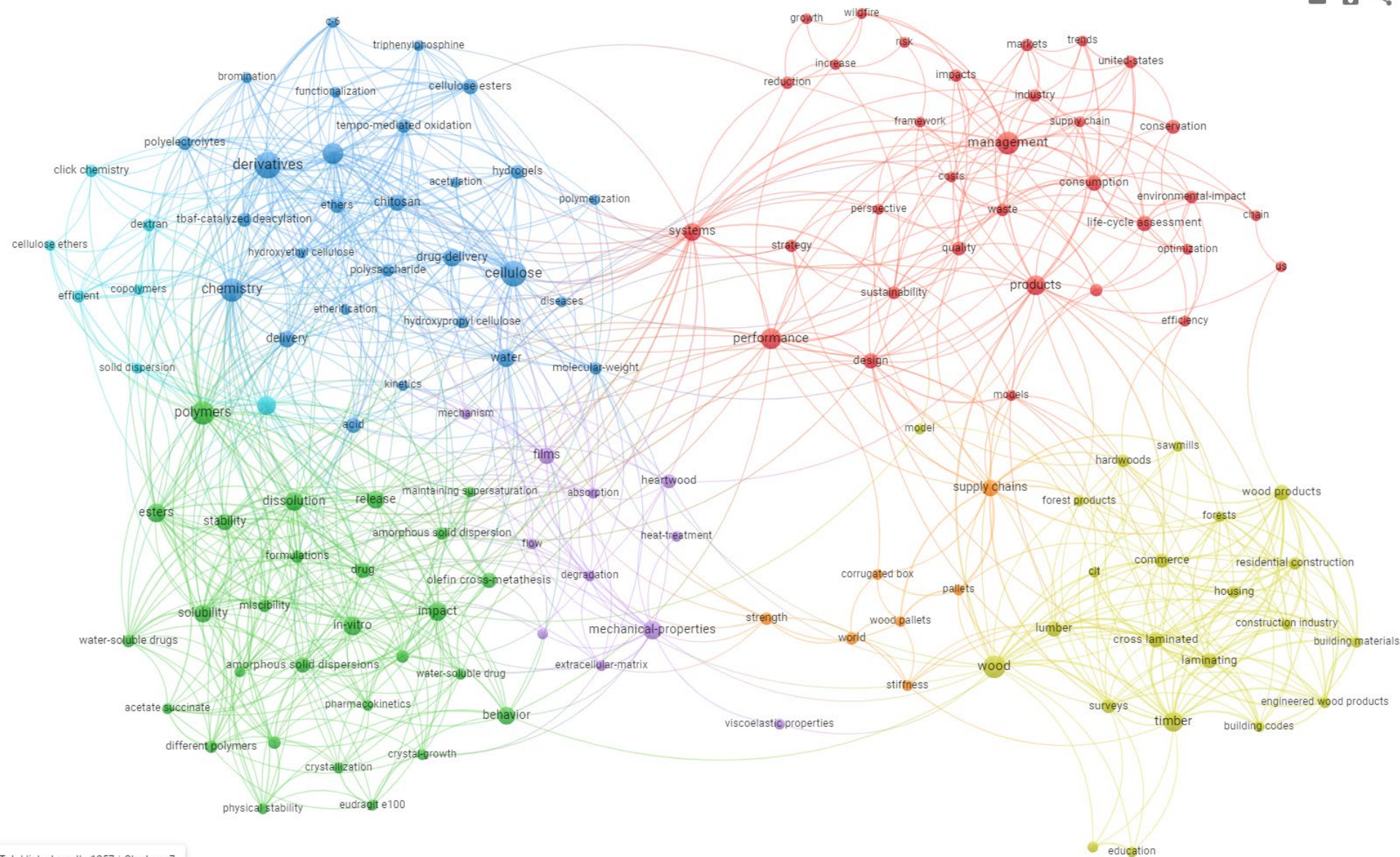


snf reseau
railegium
univ lille
snf direct innovat & rech

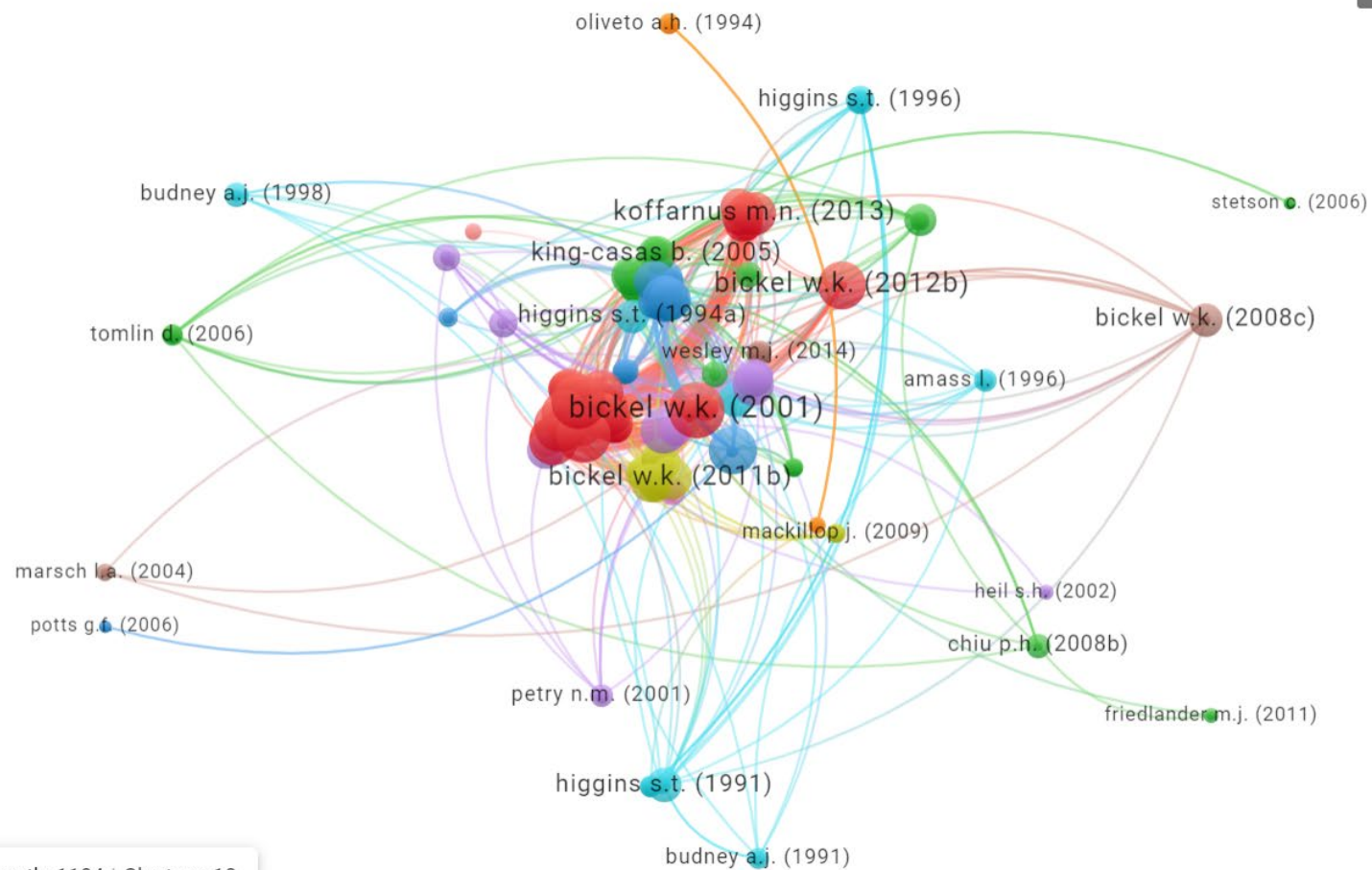


Clusters: 22

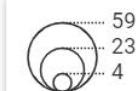




Items: 136 | Links: 942 | Total link strength: 1257 | Clusters: 7



Items: 212 | Links: 1104 | Total link strength: 1104 | Clusters: 10





Comparison of bibliographic data sources to RIM data



UNIVERSITY LIBRARIES
VIRGINIA TECH

Why rely on bibliographic data at all when you have RIM data?

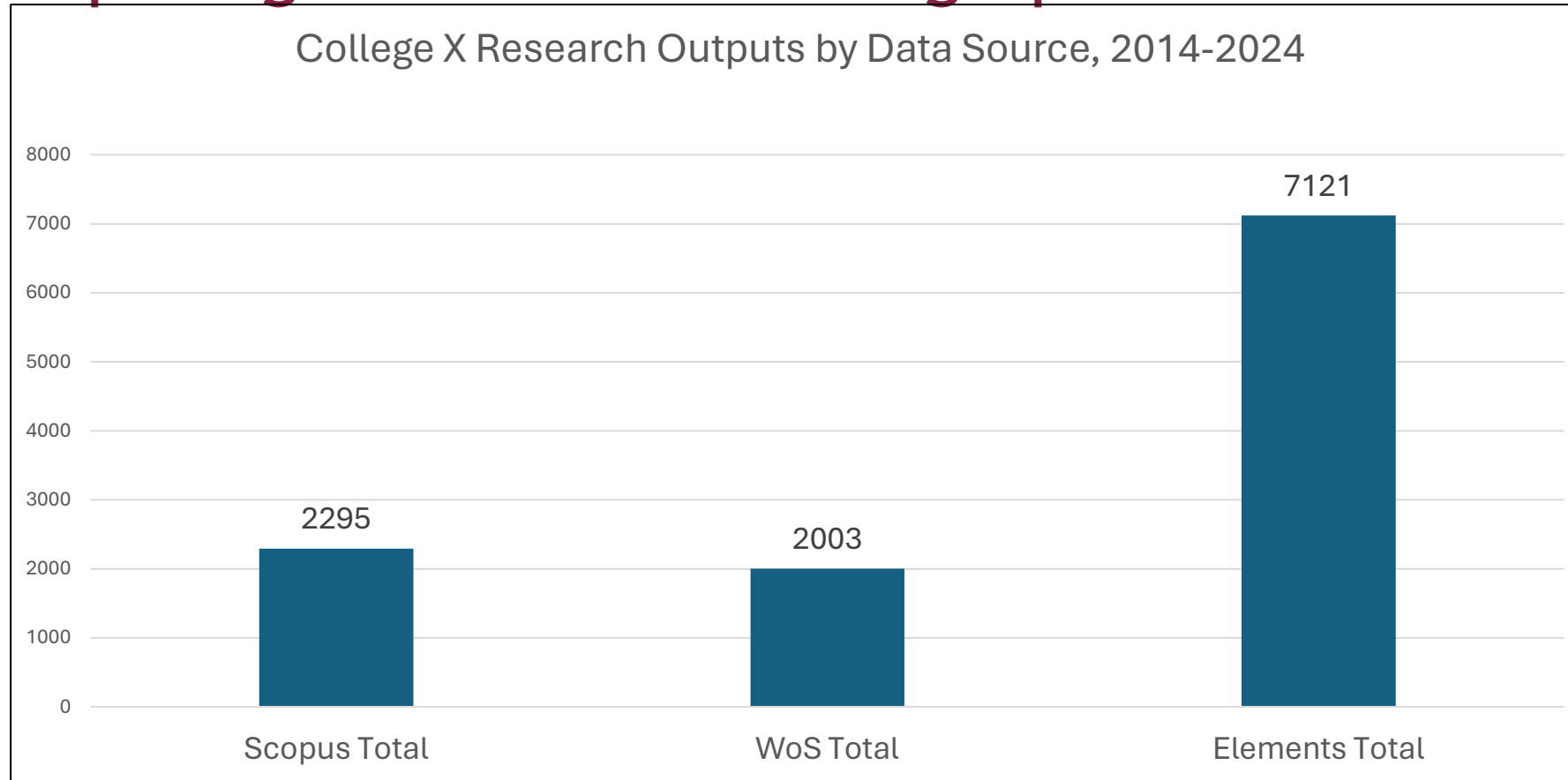
Easier to do visualizations using VOSviewer and Bibliometrix software tools

Difficult to crosswalk the data from RIM spreadsheets to Scopus or Web of Science spreadsheets

Missing metadata that helps with the analyses, such references; cannot do citation analyses in VOSviewer as a result

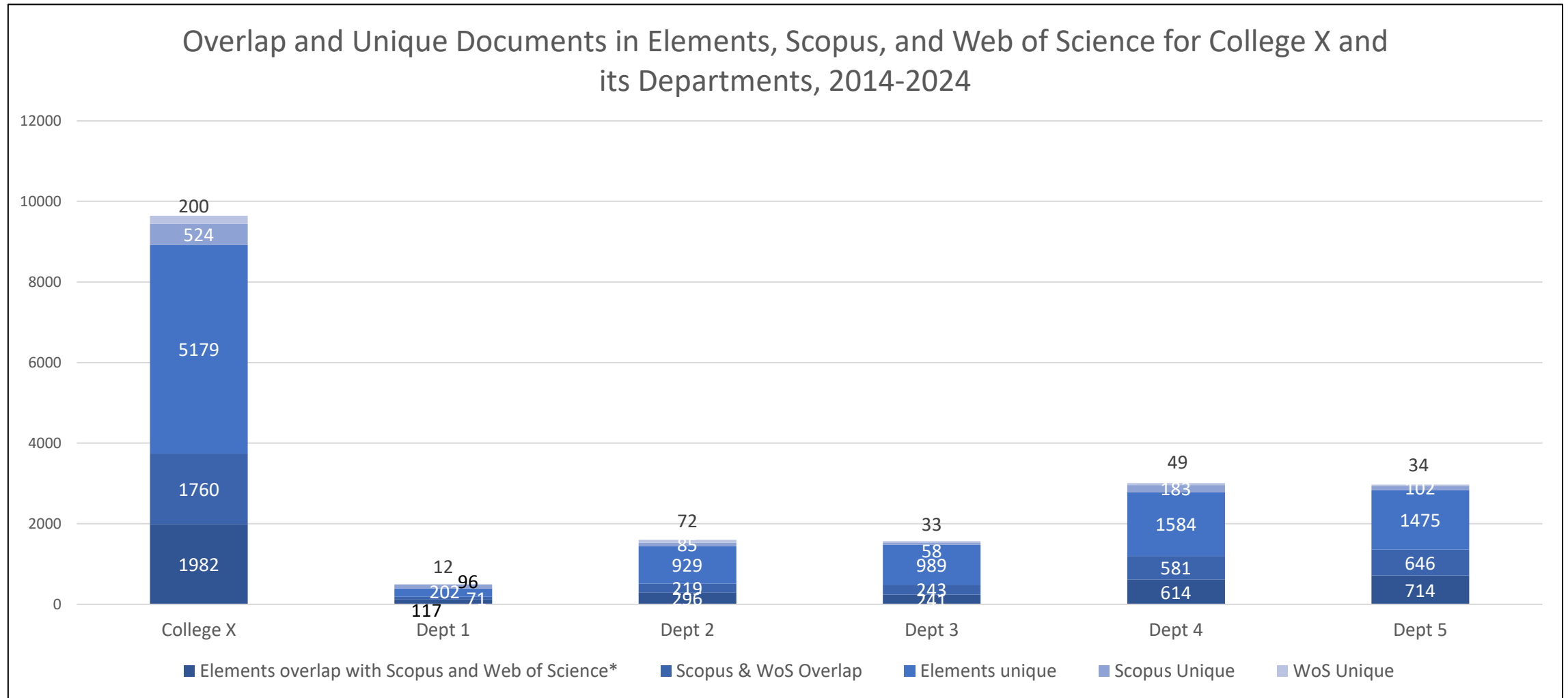
It may not be appropriate to visualize and analyze diverse document types together in certain contexts, because not all document types are 'citable' and the visualizations may give the wrong impression.

Comparing RIM Data to Bibliographic Data



Many documents in Elements / RIM system have been entered manually and include posters, lectures, conference presentations, presentations not given at conferences, and so on; these may or may not be deemed crucial scholarly works, so it is likely that this number is inflated if viewing it from this perspective.

Overlap & Unique Documents in Elements, Scopus, & WoS



*Most of the Elements overlap (1982) with Web of Science and Scopus is likely significantly overlapped with the Scopus and WoS overlap (1760); therefore, this stacked bar chart is not entirely representative of the *total* numbers.



Exporting data from a policy data source & building a dashboard



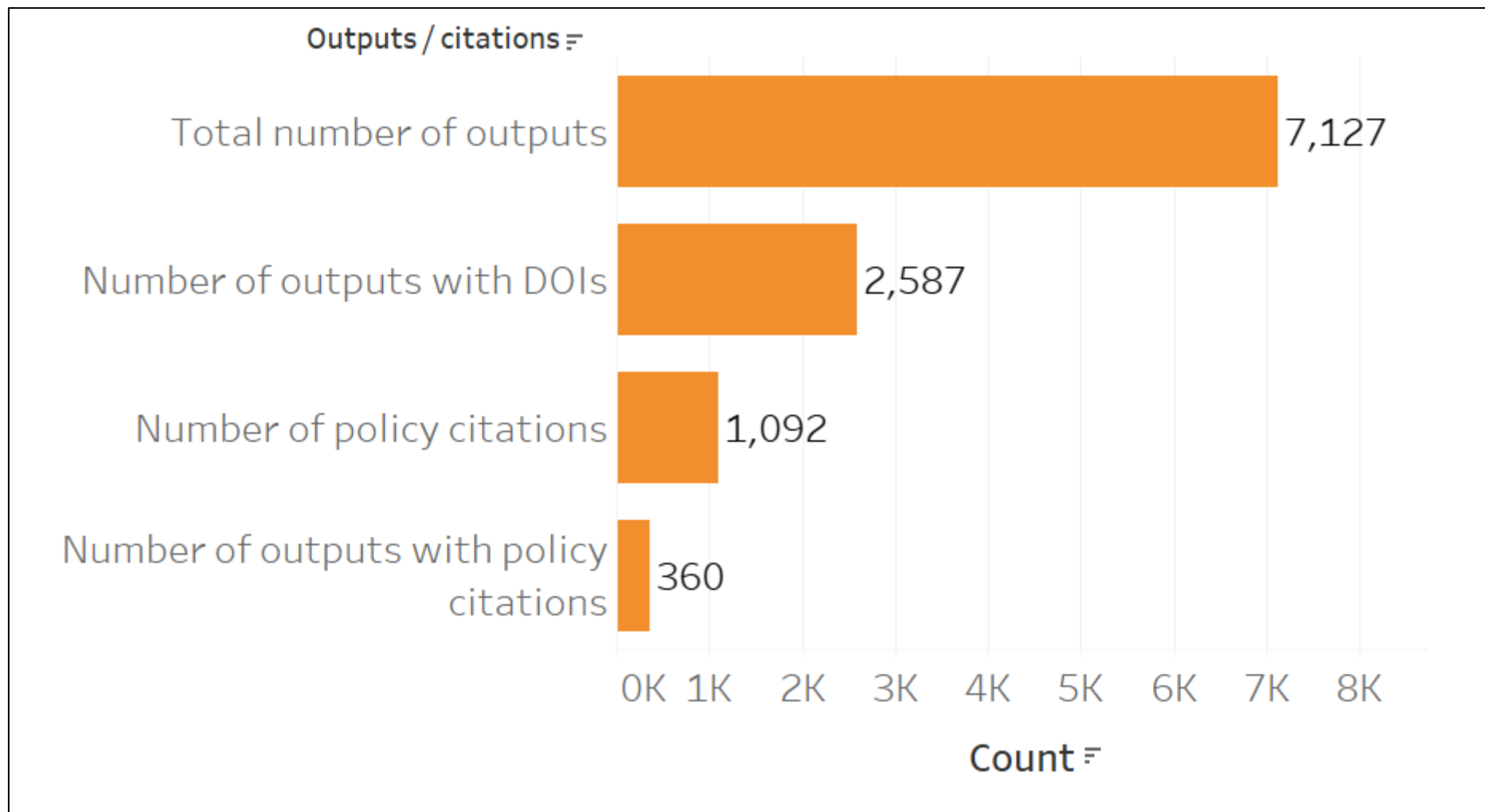
UNIVERSITY LIBRARIES
VIRGINIA TECH

Policy Mentions to Publications

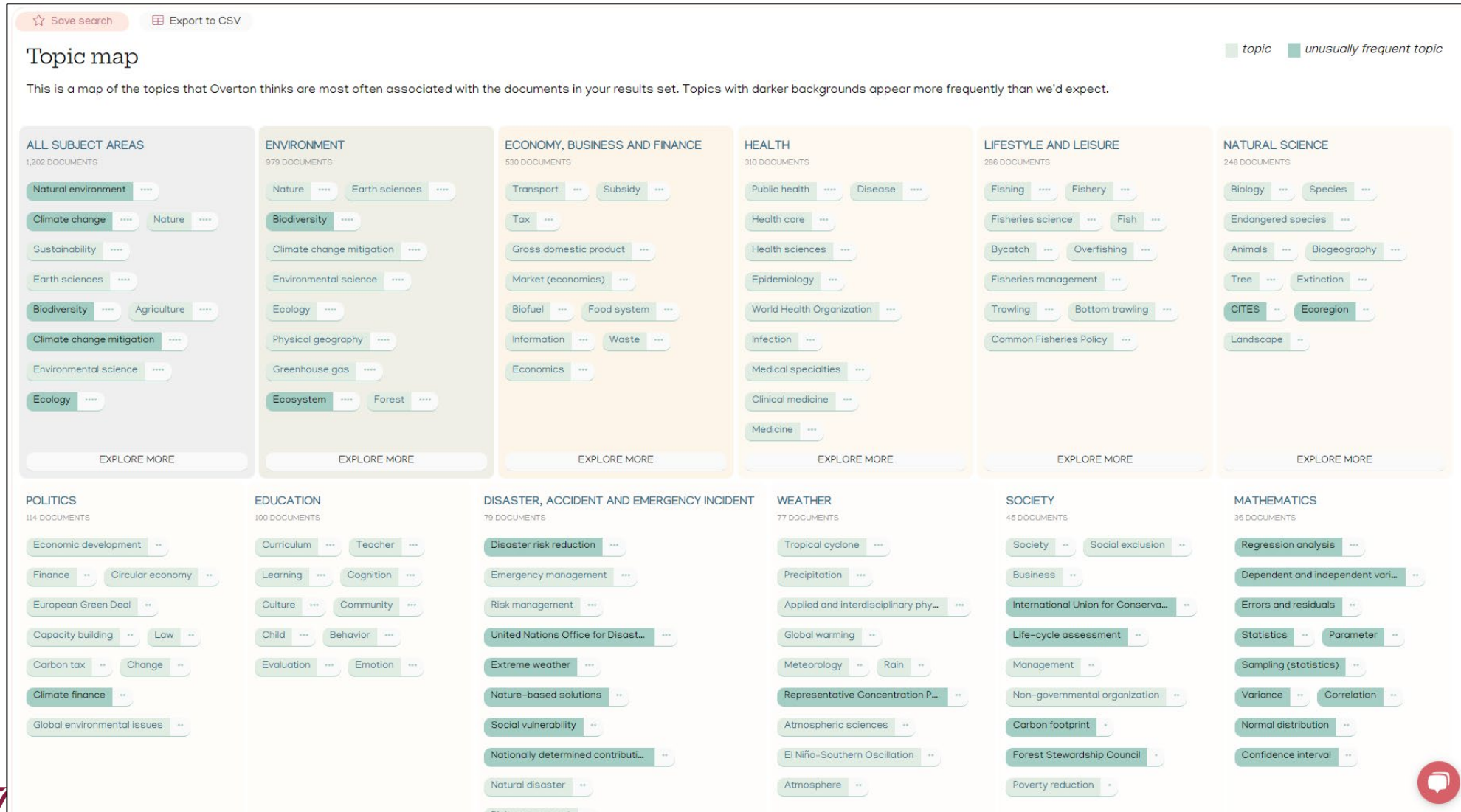
- Policy data source: Overton
- Used list of DOIs from Elements to analyze policy mentions in Overton first
- Several export options used in Overton for reporting
- Imported those spreadsheets into Tableau
- Analyzed the data based on:
 - Policy Topics
 - Policy Citing Country
 - Policy Sources
 - Sustainable Development Goals
 - Institutions also cited by the same policy documents



Breakdown of Data from Elements & its policy citations



Topic Map in Overton



Context of Topics



Export from the Report in Overton

★ Save this report

📄 See underlying documents

📄 Export ▼

Document count
1,123

Countries
43

Sources
258



IGO	389	USA	272	EU	100
UK	86	Canada	69	France	49
Germany	37	Australia	35	Sweden	21
Spain	15	Belgium	15	Ireland	13

⬆ Show more

📄 Copy as CSV

🖼 Save as image

Document types

NAME	DOCUMENTS
Publication	1,119
Working paper	43
Clinical guidance	20
Scholarly article	10
Blog post	6
Transcript	2
Legal documents	1
Periodical	1

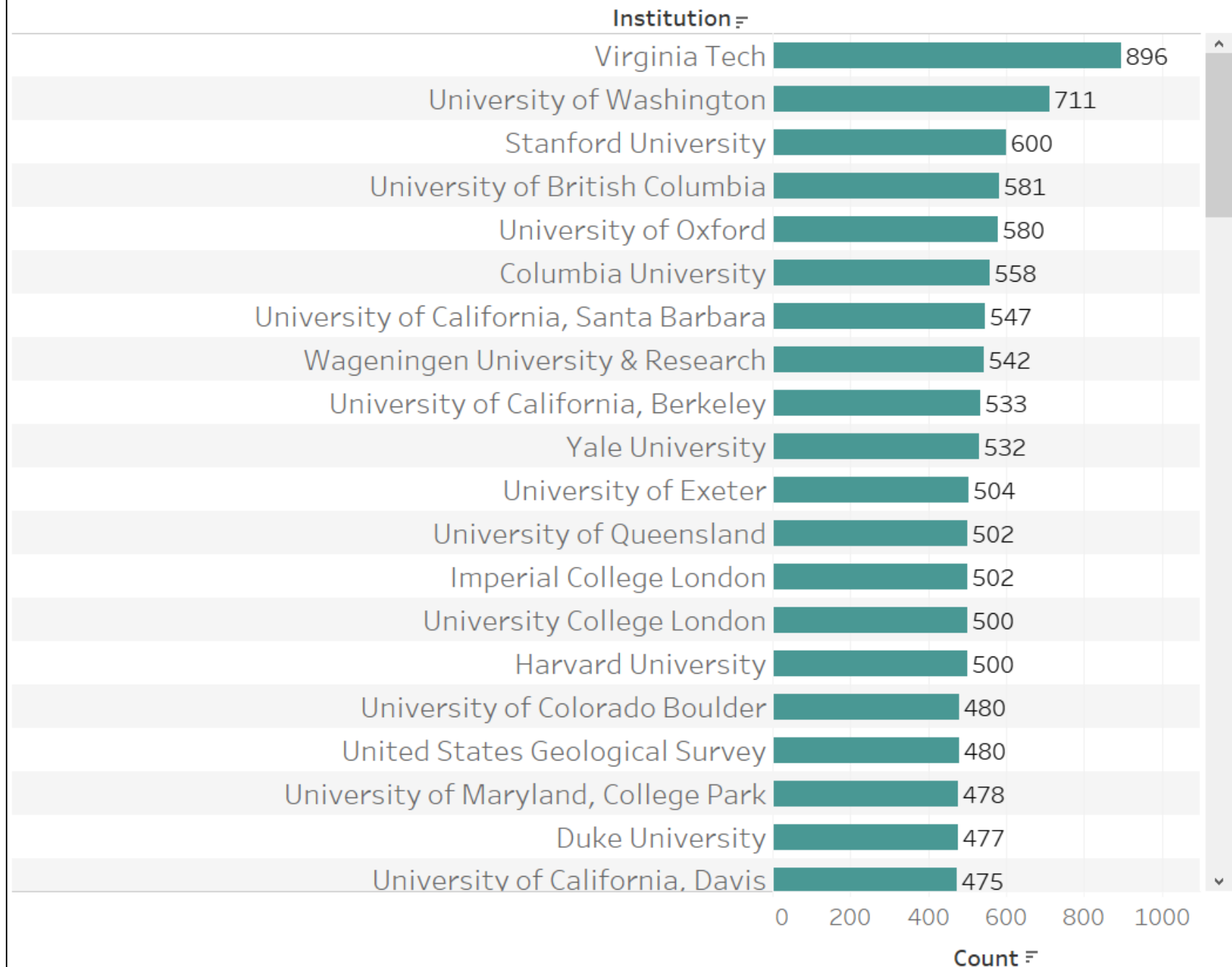
📄 Copy as CSV

Source specific tags

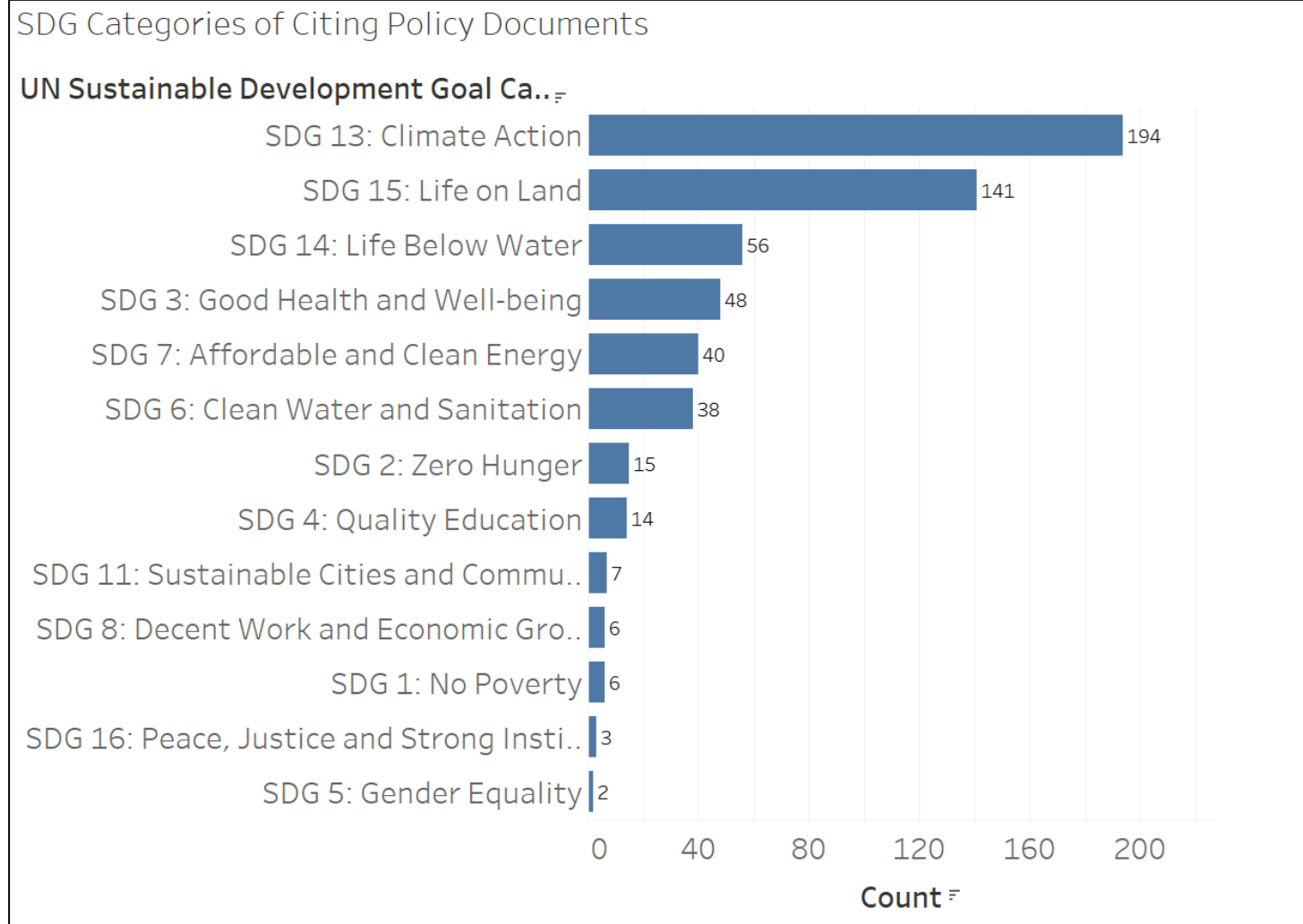
NAME	DOCUMENTS
CLIMATE CHANGE	19
Environment and climate change	15
BIOLOGICAL DIVERSITY	13
MARINE ECOSYSTEMS	13
SUSTAINABLE DEVELOPMENT	12
ENVIRONMENTAL POLICY	10
OCEANS	9

Institutions also cited by these Policy Documents

Institutions (also) Cited by these Policy Documents

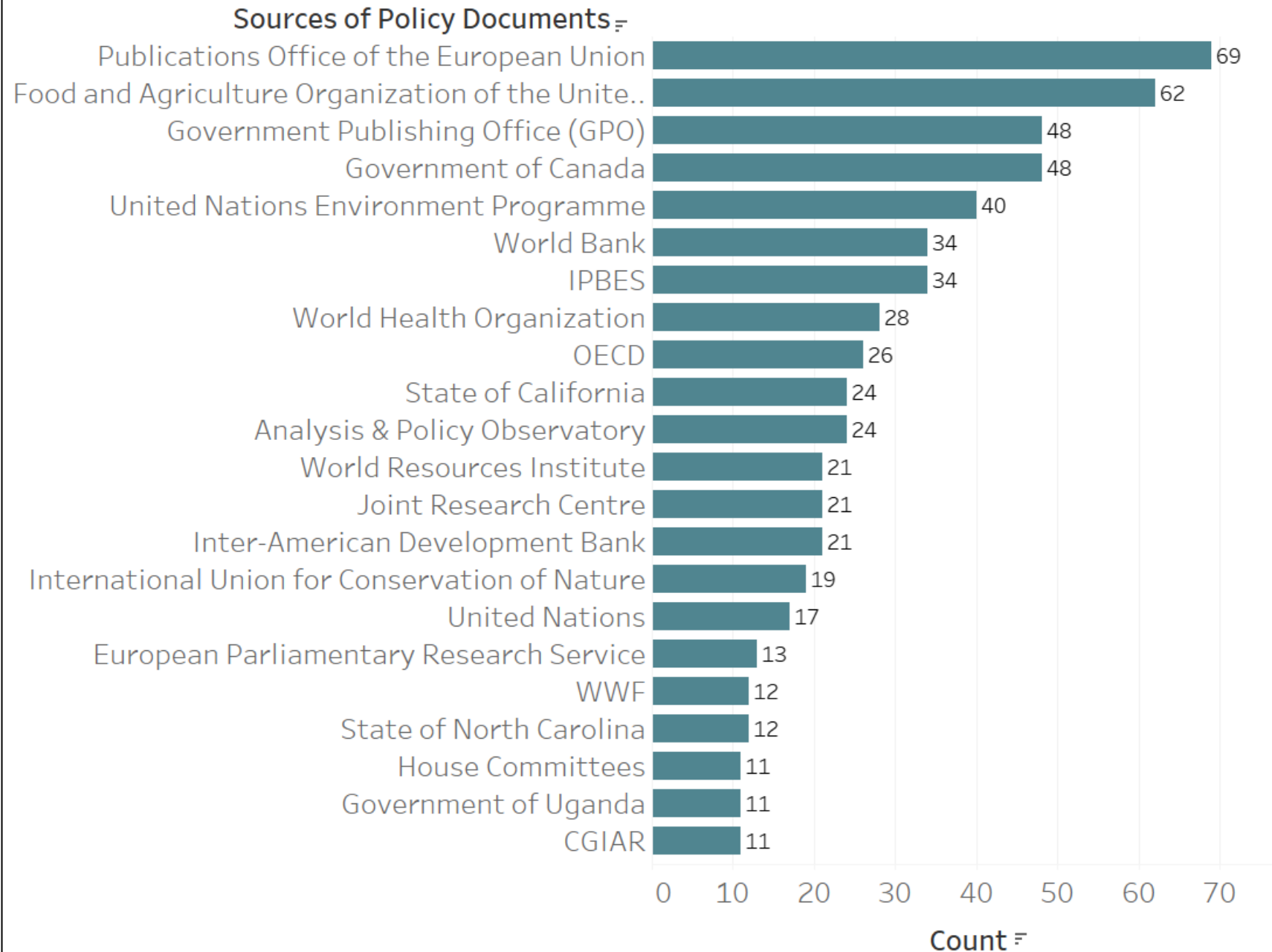


SDG Categories by Citing Policy Documents

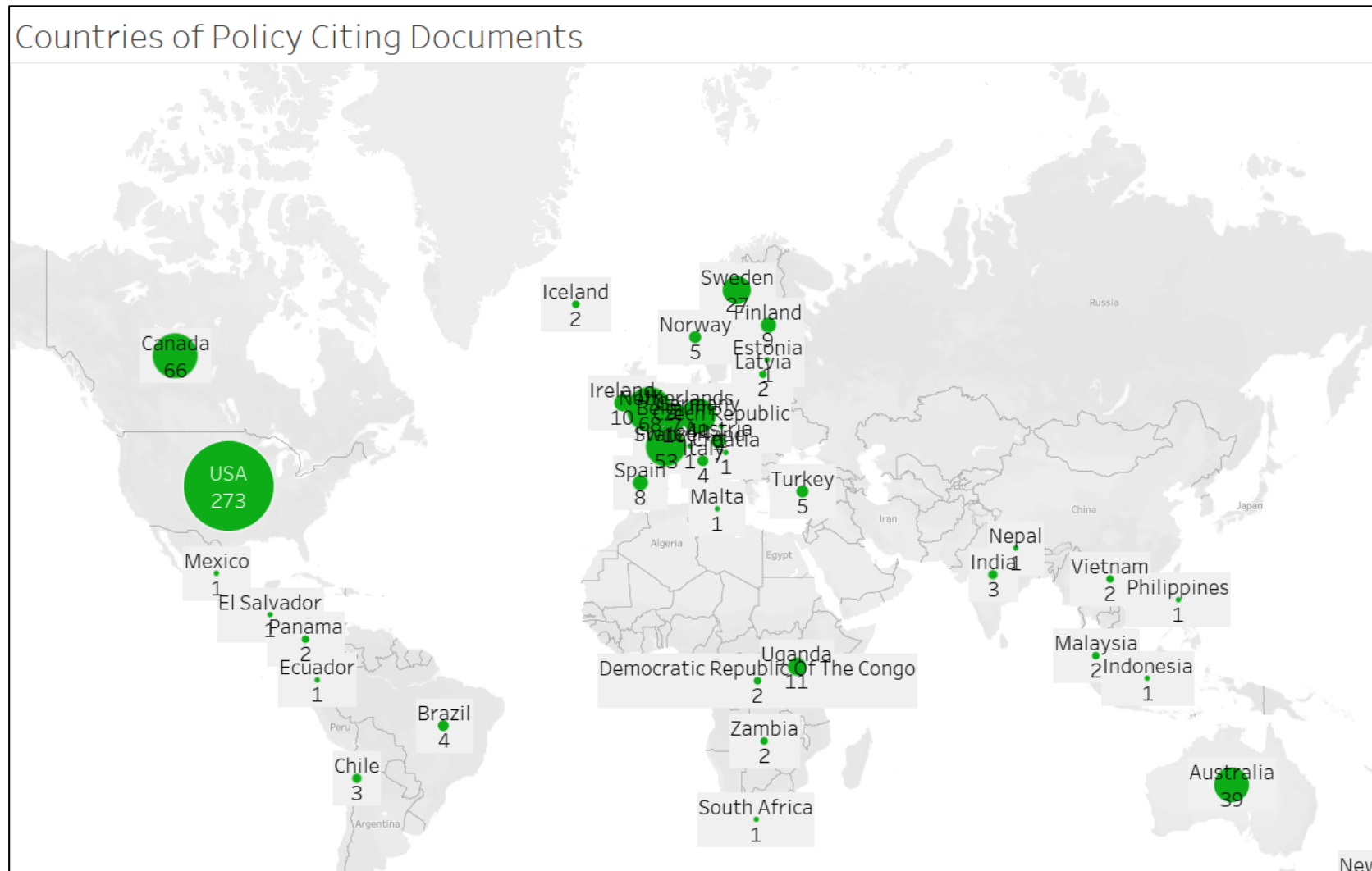


Sources of Policy Citing Documents (Organizations, NGOs, Nonprofits, etc.)

Sources of Policy Citing Documents



Map of Countries of Policy Citing Documents





Questions?

Thank you!

Rachel Miles

Research Impact Coordinator

ramiles@vt.edu