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- More inclusive of communities that do not publish as often or that produce more ‘grey literature,’ such as reports, and conference papers not indexed in bibliographic databases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)

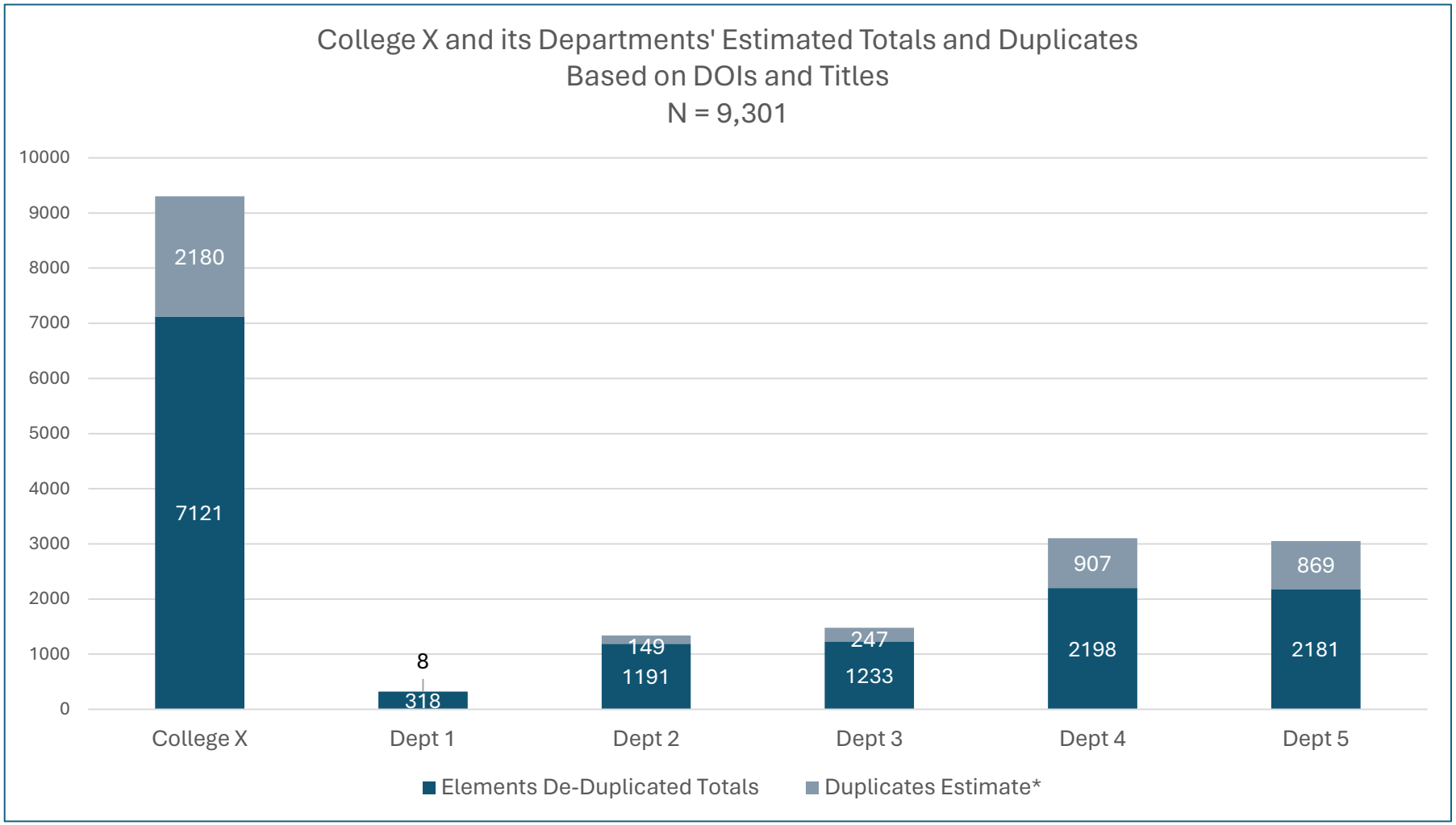
- Basic Report generated / exported for **College X**
- Filters:
  - Date from**: 01/01/2014**
  - Date to: **01/01/2024**
  - Include non-current users: **Yes**
- Select data export or report:
  - Object category: **Scholarly & creative works (linked to the selected users)**
  - Scholarly & creative 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

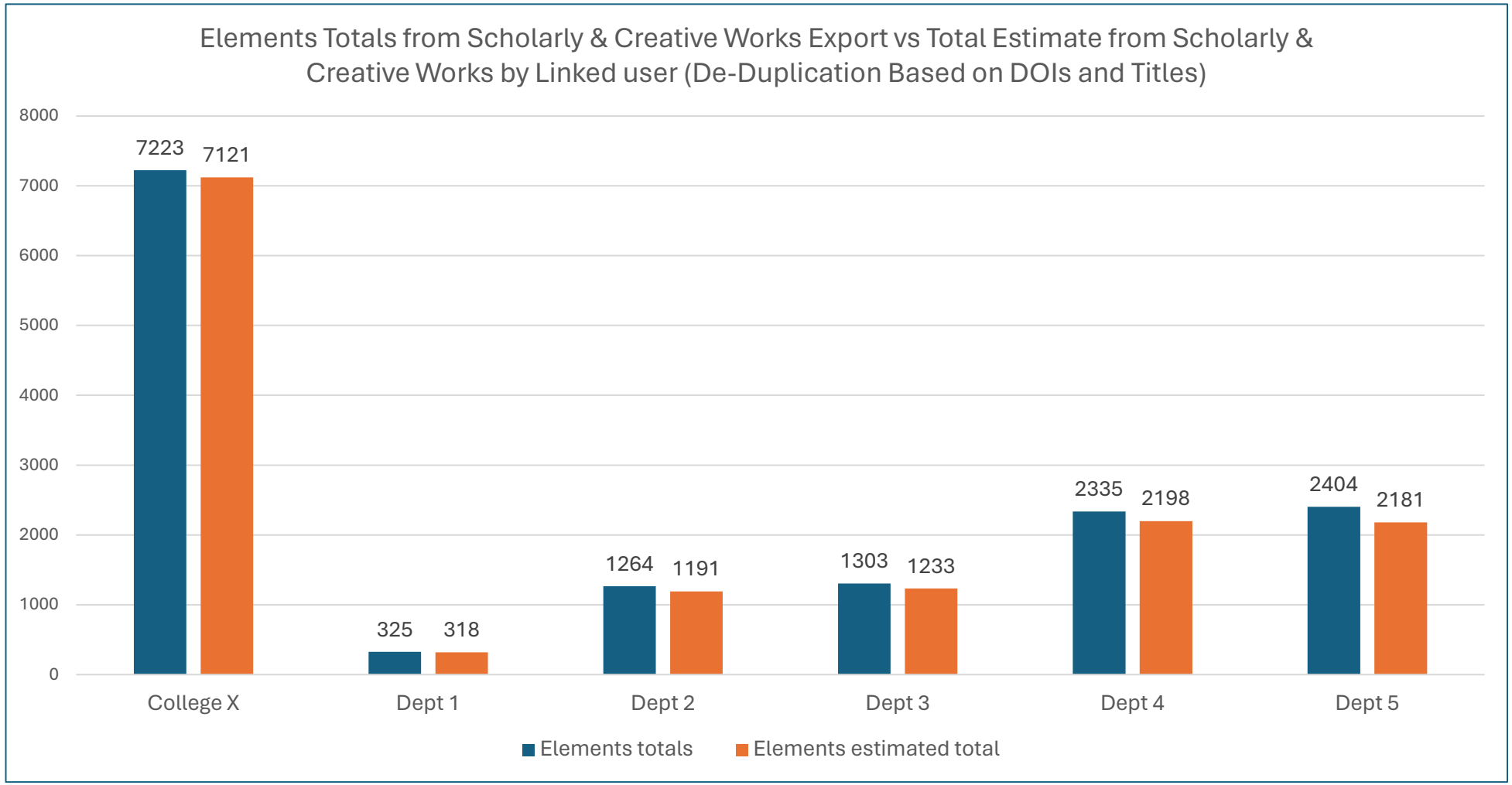
Many useful fields and values for research analytics & building dashboards (next slide)

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



UNIVERSITY LIBRARIES VIRGINIA TECH

i.e., Duplicates are overestimated, because many scholarly works have the same title but are actually separate, individual works.

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*

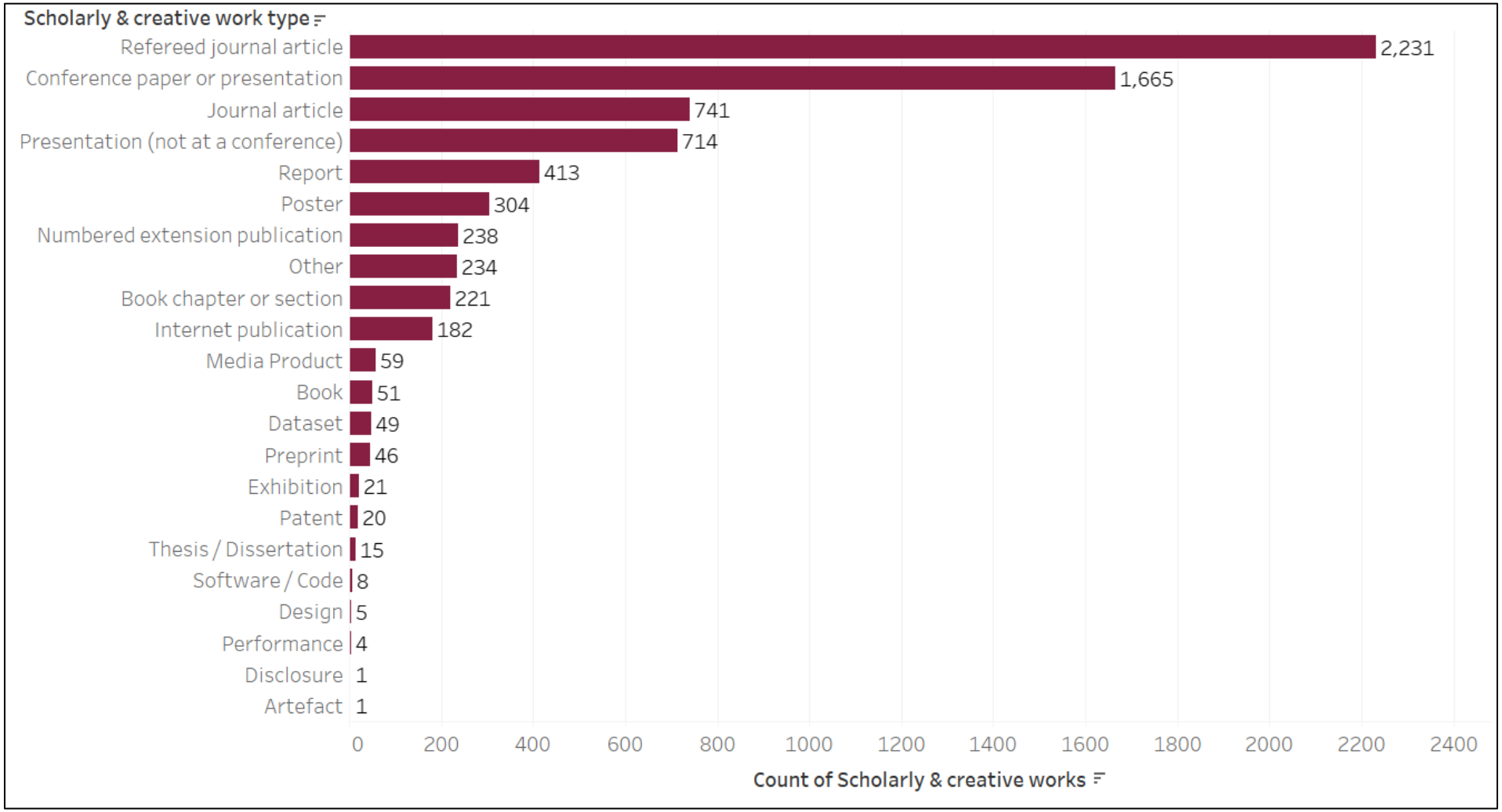
**Pulls the citation count based on the precedence we setup in the system, which has manual set as the top followed by automated sources.*

UNIVERSITY LIBRARIES VIRGINIA TECH

Dashboard Building in Tableau with Elements Data for College X

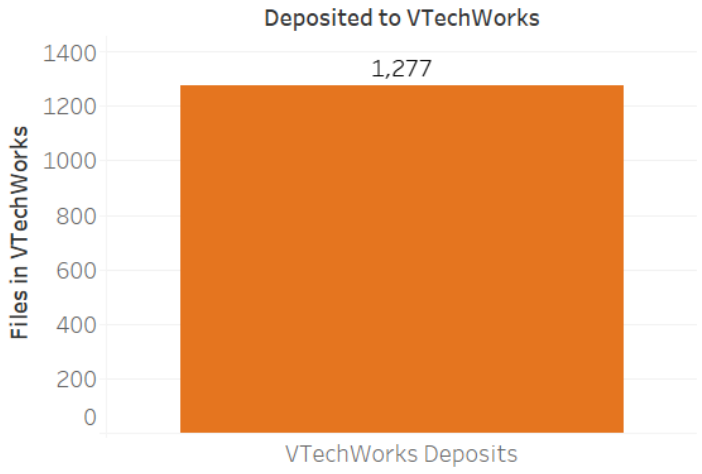
- Types of Scholarly Works (Report 2)
- VTechWorks (institutional repository) deposits (Report 1, inflated)
- Open Access records in other repositories (limited) (Report 1, inflated)
- Publication and citation counts by bibliographic data source (Report 1, inflated)
- Commissioning Bodies (Report 2)
- 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

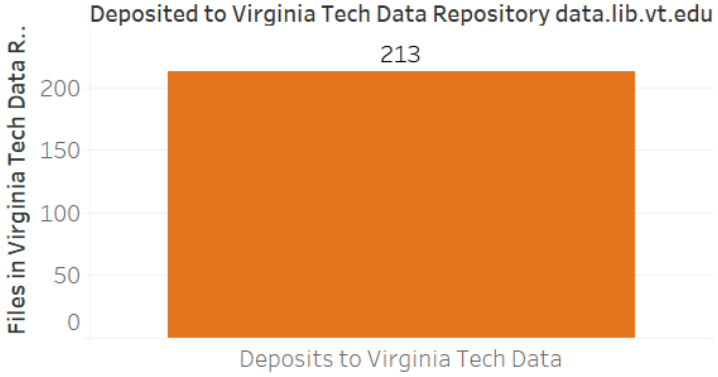


Open Access Deposits by Source

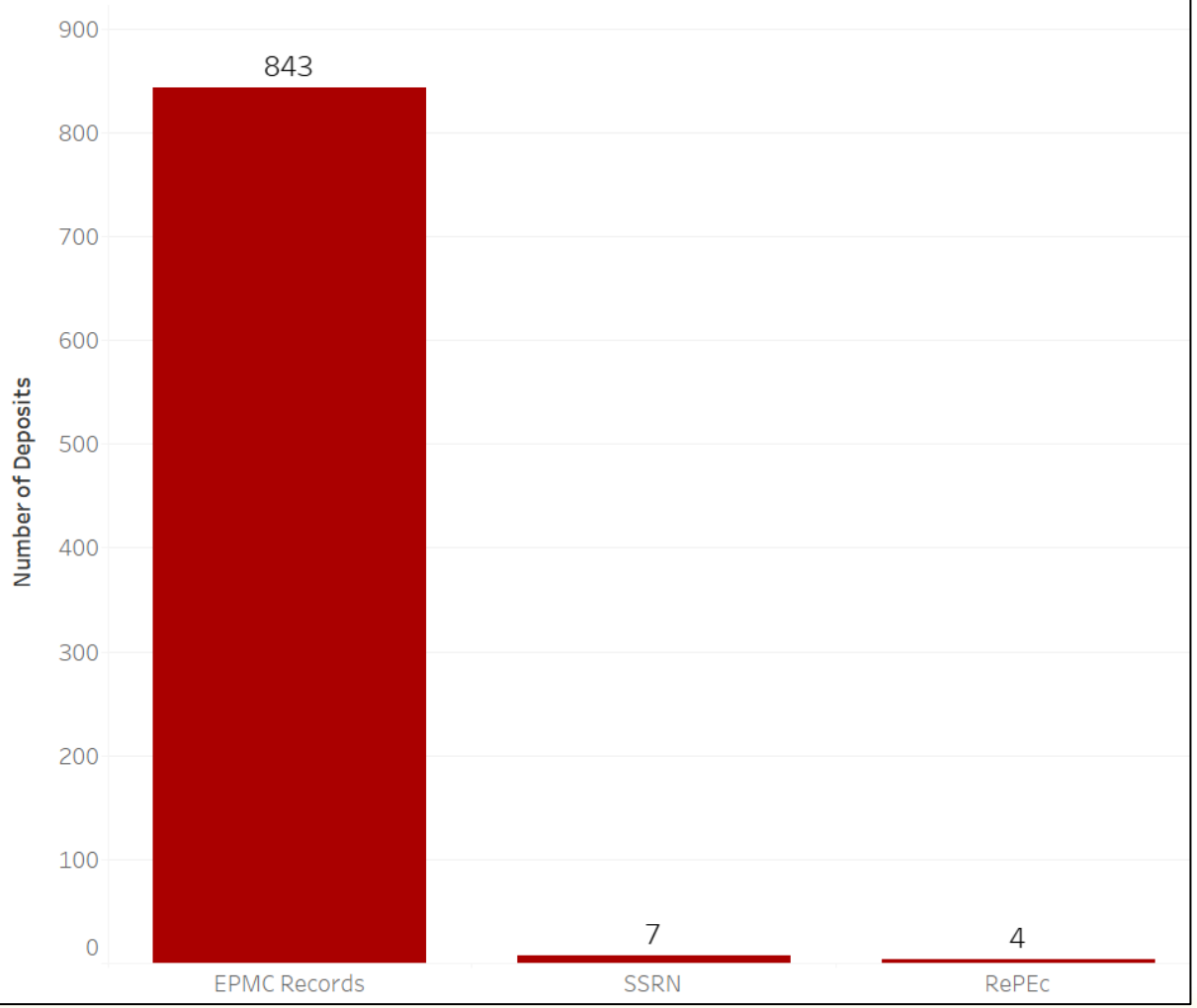
VTechWorks Deposits



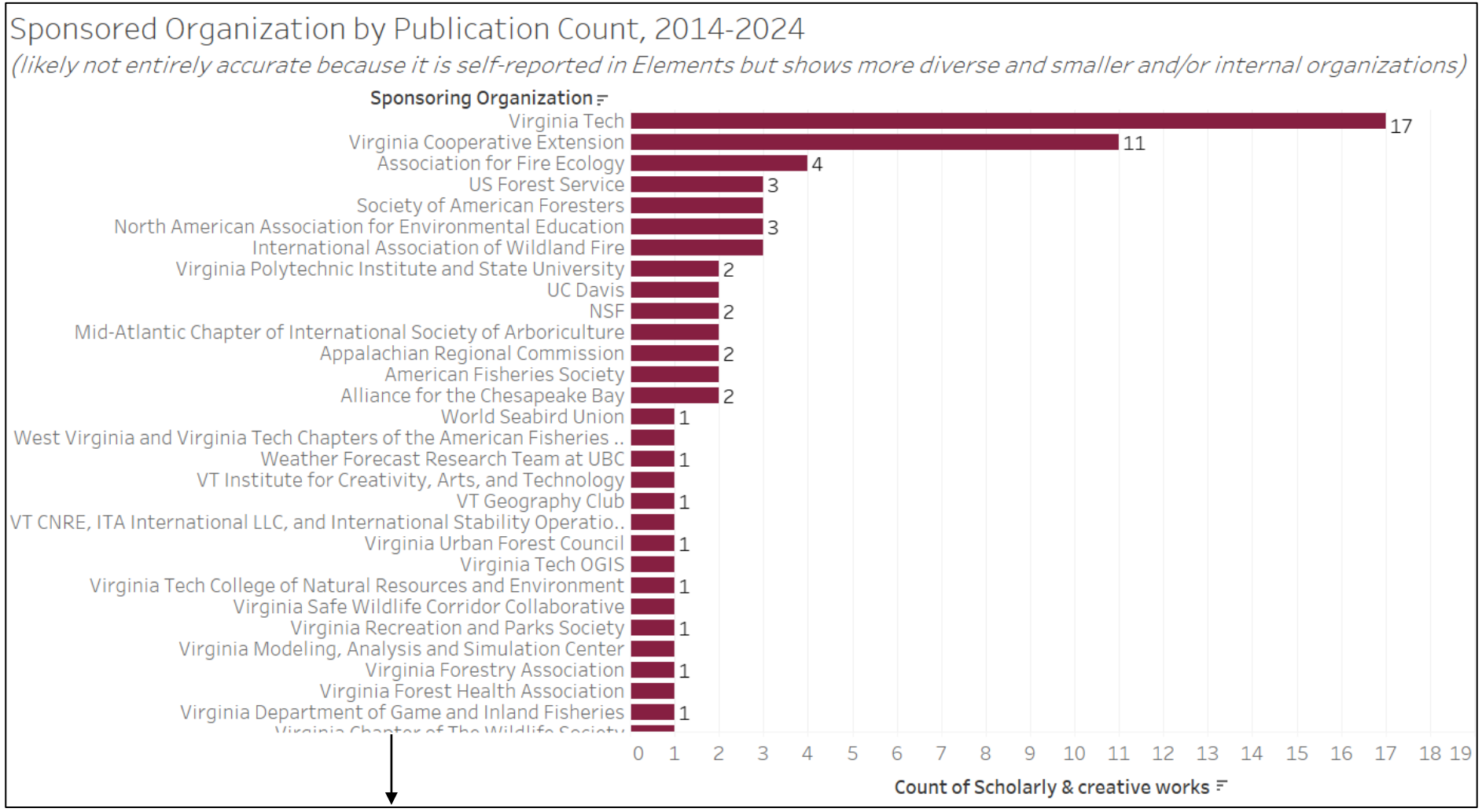
OA Records in Repositories



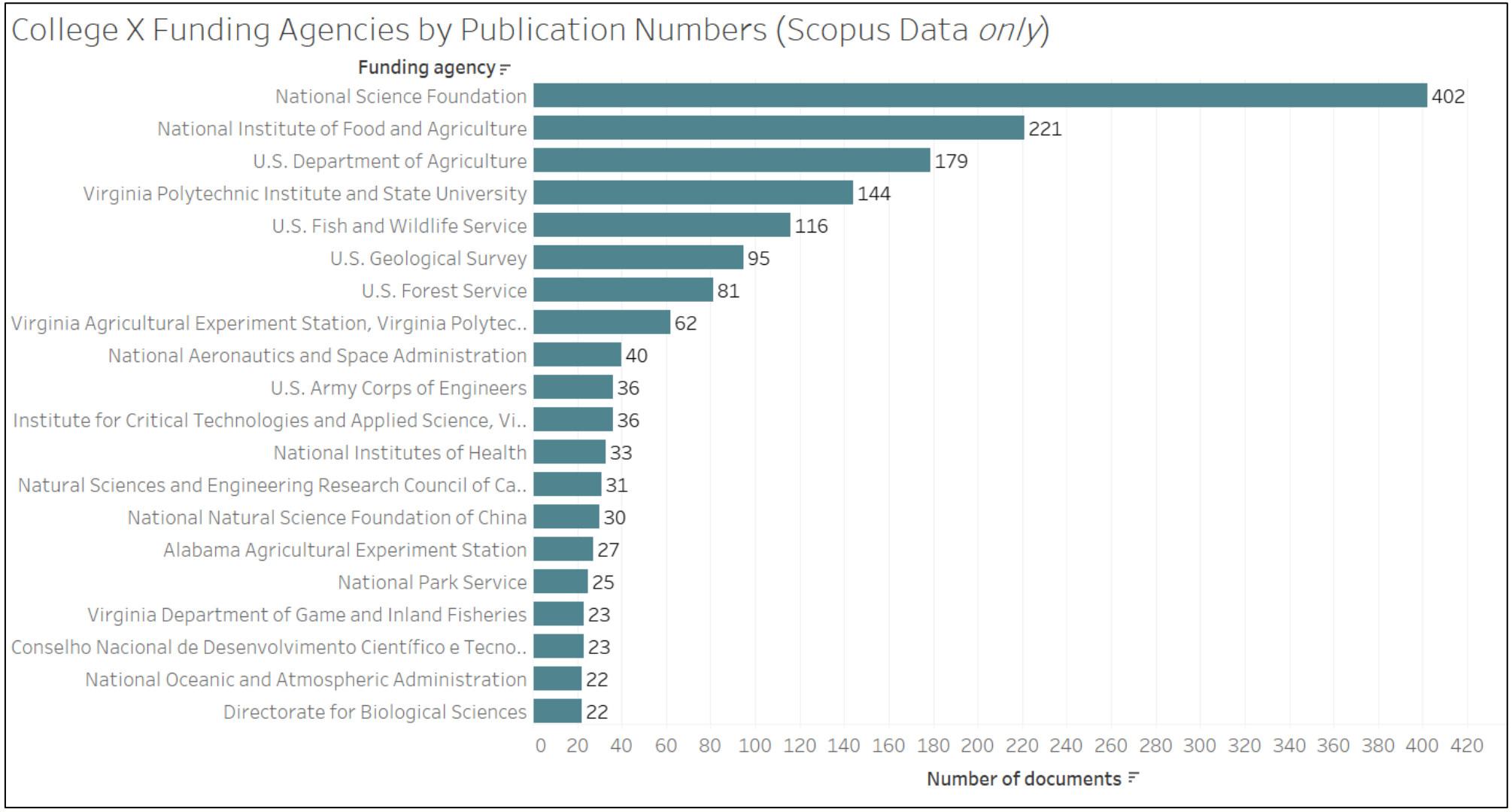
Open Access Records in Europe PubMed Central, RePEc, SSRN



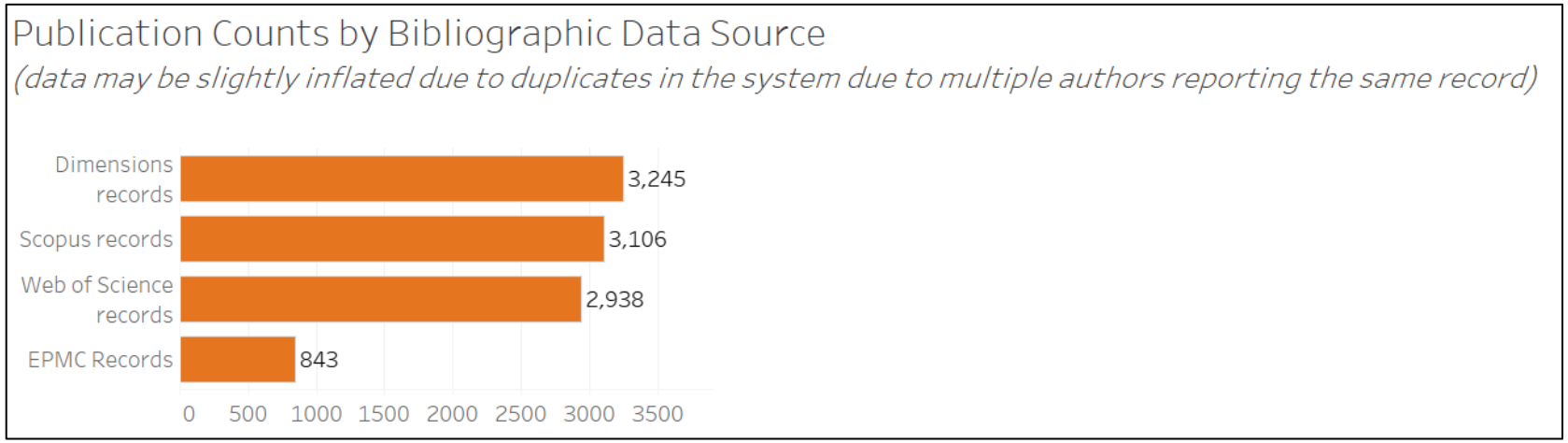
Sponsored / Funding Organization by Scholarly Work

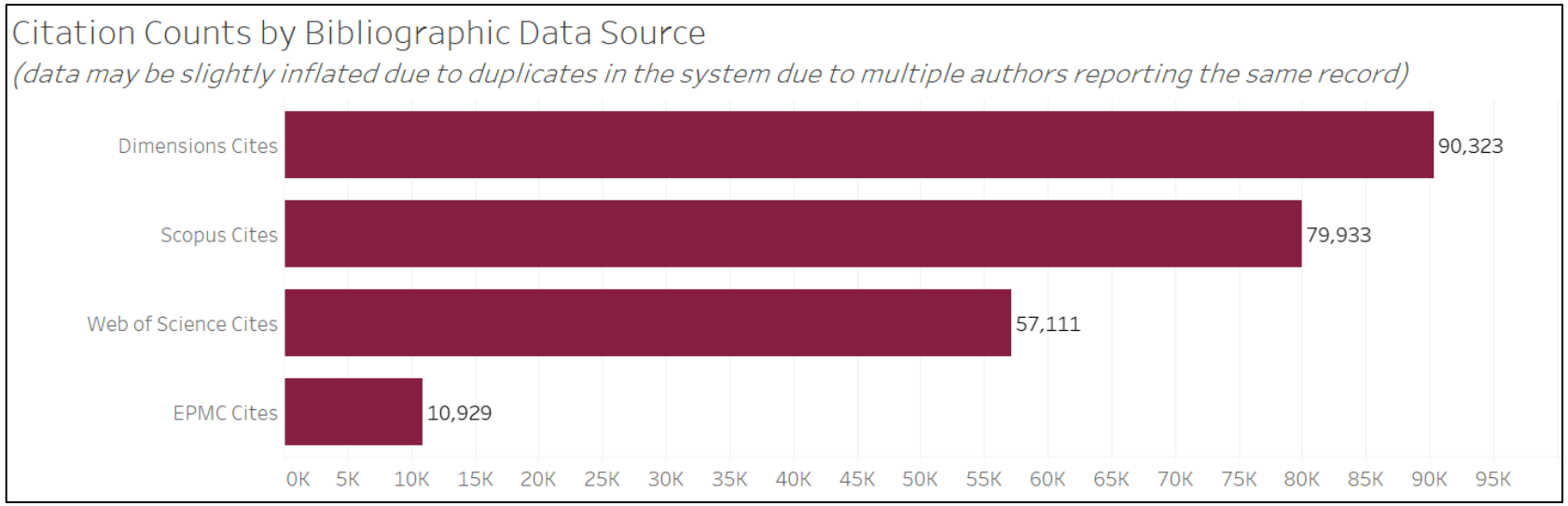


Comparing Funding Agencies to Scopus Data



Publication & Citation Counts by Bibliographic Data Source





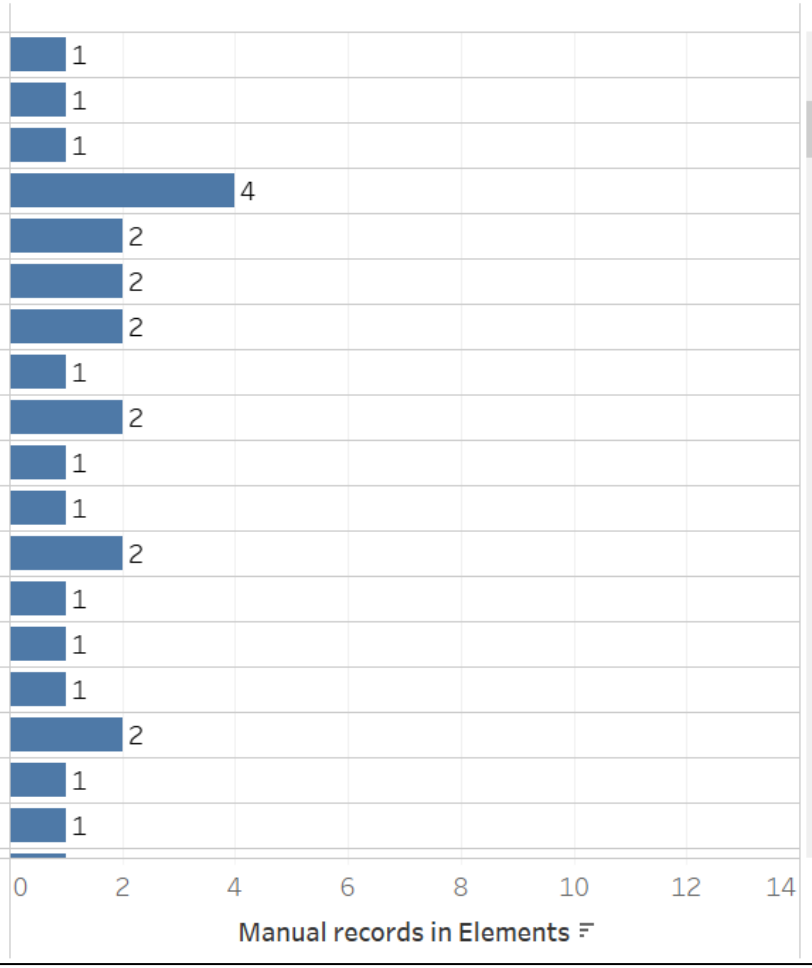
UNIVERSITY LIBRARIES
VIRGINIA TECH

Books & Reports

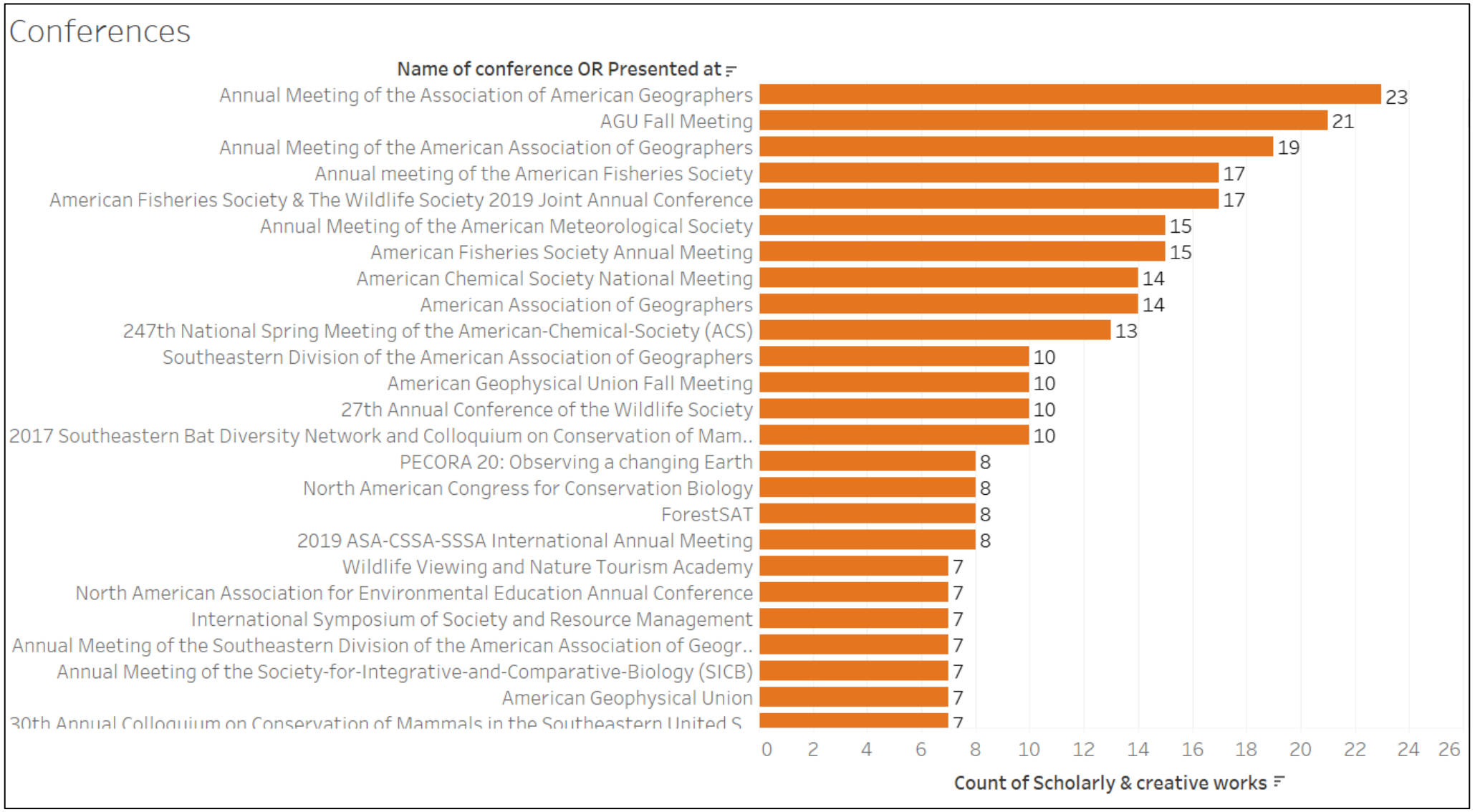
Books & Reports as reported by faculty

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

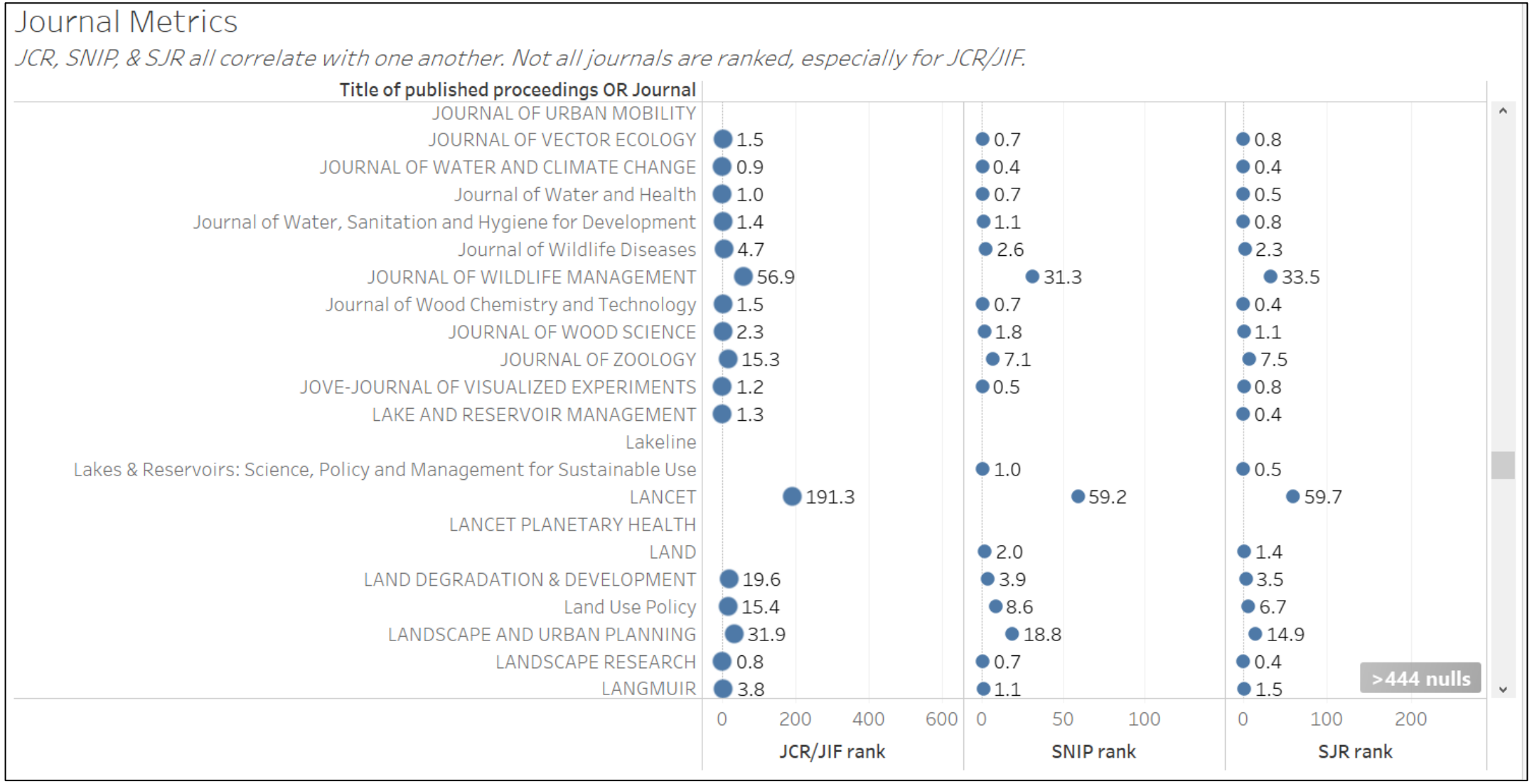
Book title OR Preprint server OR Report title	Chapter number OR Article nu..	Manual records in Elements
Analysis of engineered flooring from Vintage flooring	Null	1
Analysis of Seven Layer Engineered Flooring	Null	1
Analysis of Splits in Engineered Flooring In-service	Null	1
Annual Progress Report	NASA Award Number: NNX17AI09G	4
Annual progress report submitted to USDA CSREES	Null	2
Annual report to NOAA SEFSC	Null	2
Annual report, Great Lakes Fisheries Commission, online.	Null	2
AoB Plants	Null	1
Appalachia's Coal-Mined Landscapes: Resources and Communitie..	Null	2
Appalachian Epidemics: From Pre-Contact to COVID	Null	1
Appalachian Health: Culture, Changes, and Capacity	Null	1
Applications of Small Unmanned Aircraft Systems: Best Practices ..	4	2
April 2017 housing notes	Null	1
Aquatic Ecosystems and Environment: Concept and Method	Null	1
Aquatic protected areas as a fishery management tool. Série Área..	Null	1
Assessing the relationships between pollinator-friendly plantings..	Null	2
August 2017 housing notes	Null	1
August 2020 Housing Commentary: Section I	Null	1



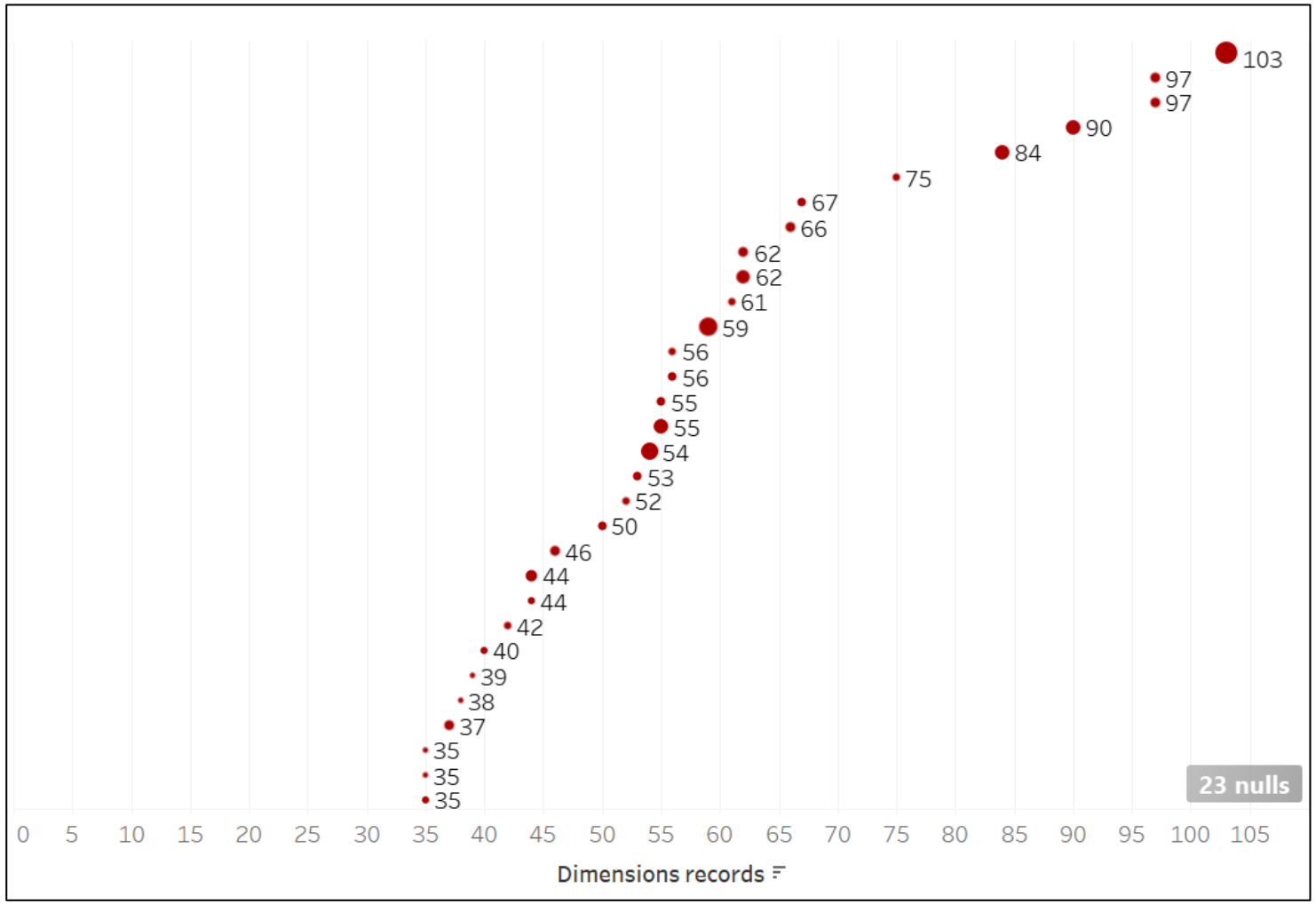
Conferences

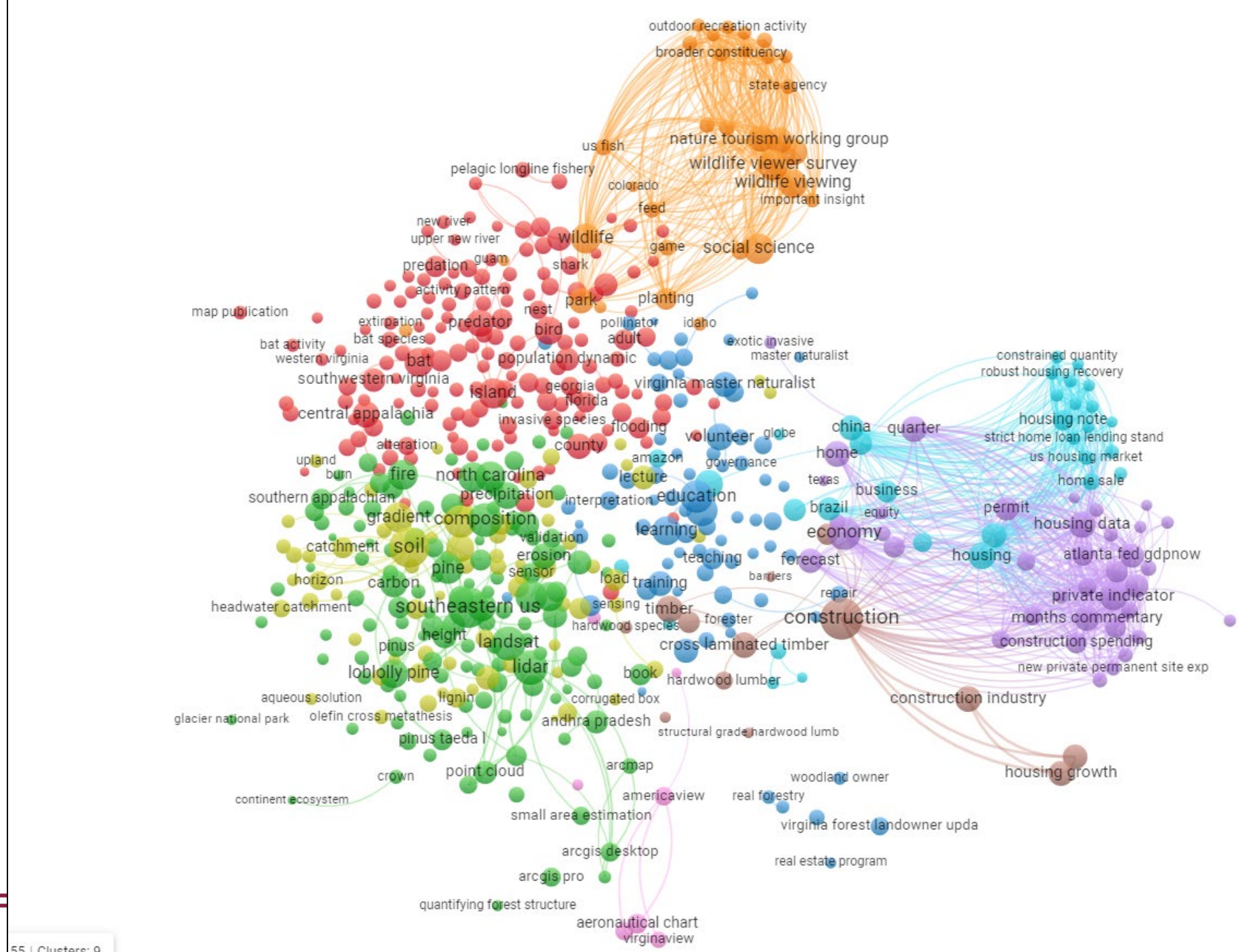


Journal Metrics

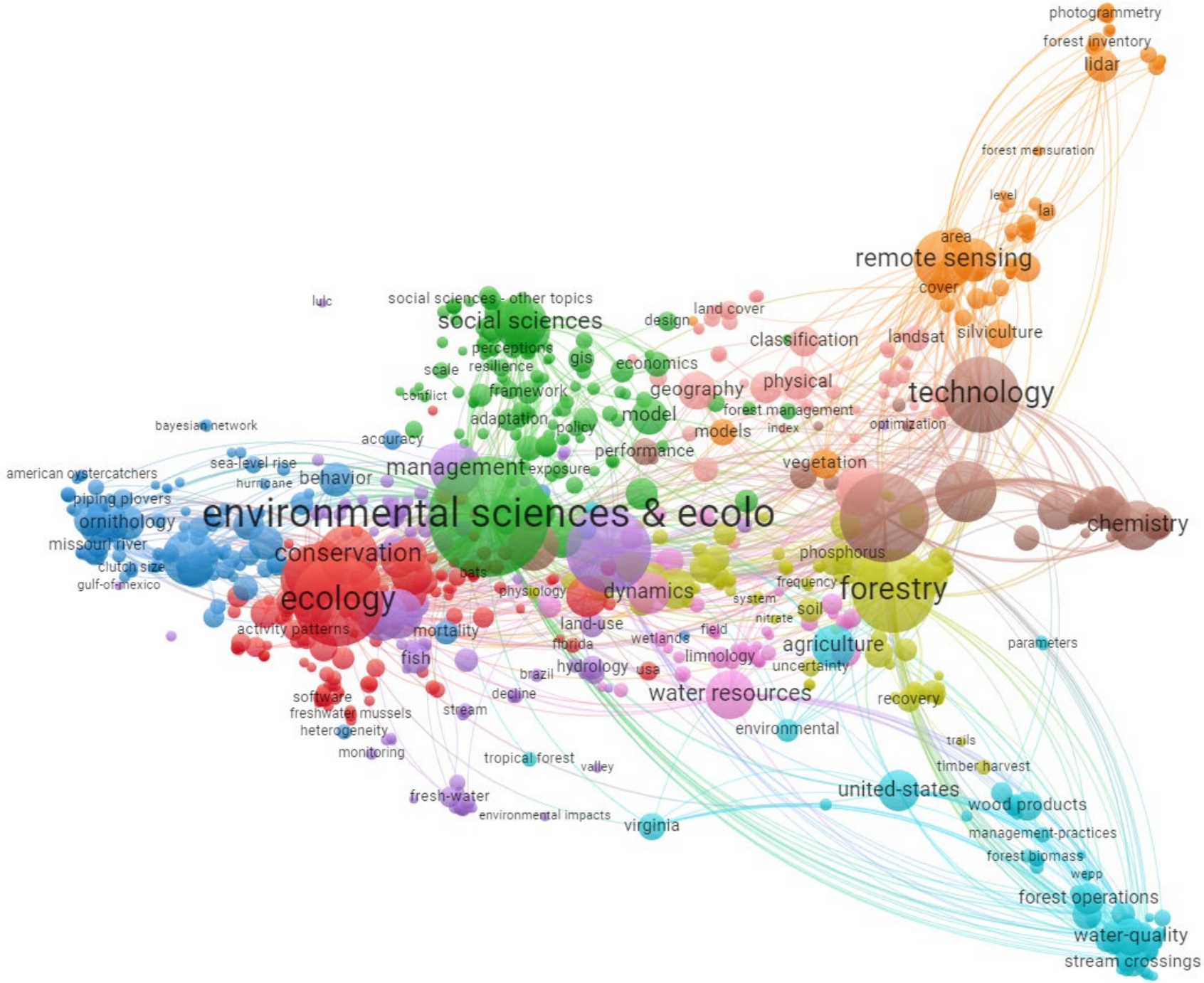


Publication & Citation Counts by Faculty Member (names not displayed)
Size of node = Number of Citations





55 | Clusters: 9


photogrammetry
forest inventory
lidar
forest mensuration
level
lai
area
remote sensing
cover
silviculture
landsat
classification
land cover
design
social sciences - other topics
lulc
social sciences
perceptions
gis
economics
scale
resilience
framework
conflict
geography
physical
technology
forest management
index
optimization
model
adaptation
policy
models
bayesian network
accuracy
performance
vegetation
sea-level rise
management
exposure
american oystercatchers
hurricane
behavior
piping plovers
environmental sciences & ecolo
ornithology
chemistry
missouri river
conservation
phosphorus
clutch size
gulf-of-mexico
bats
frequency
forestry
physiology
dynamics
system
soil
ecology
nitrate
activity patterns
mortality
land-use
wetlands
field
florida
agriculture
parameters
limnology
uncertainty
fish
hydrology
use
brazil
water resources
recovery
decline
software
freshwater mussels
stream
heterogeneity
environmental
trails
monitoring
tropical forest
valley
timber harvest
united-states
fresh-water
wood products
environmental impacts
virginia
management-practices
forest biomass
wepp
forest operations
water-quality
stream crossings

Lessons learned from building dashboards with RIM data

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

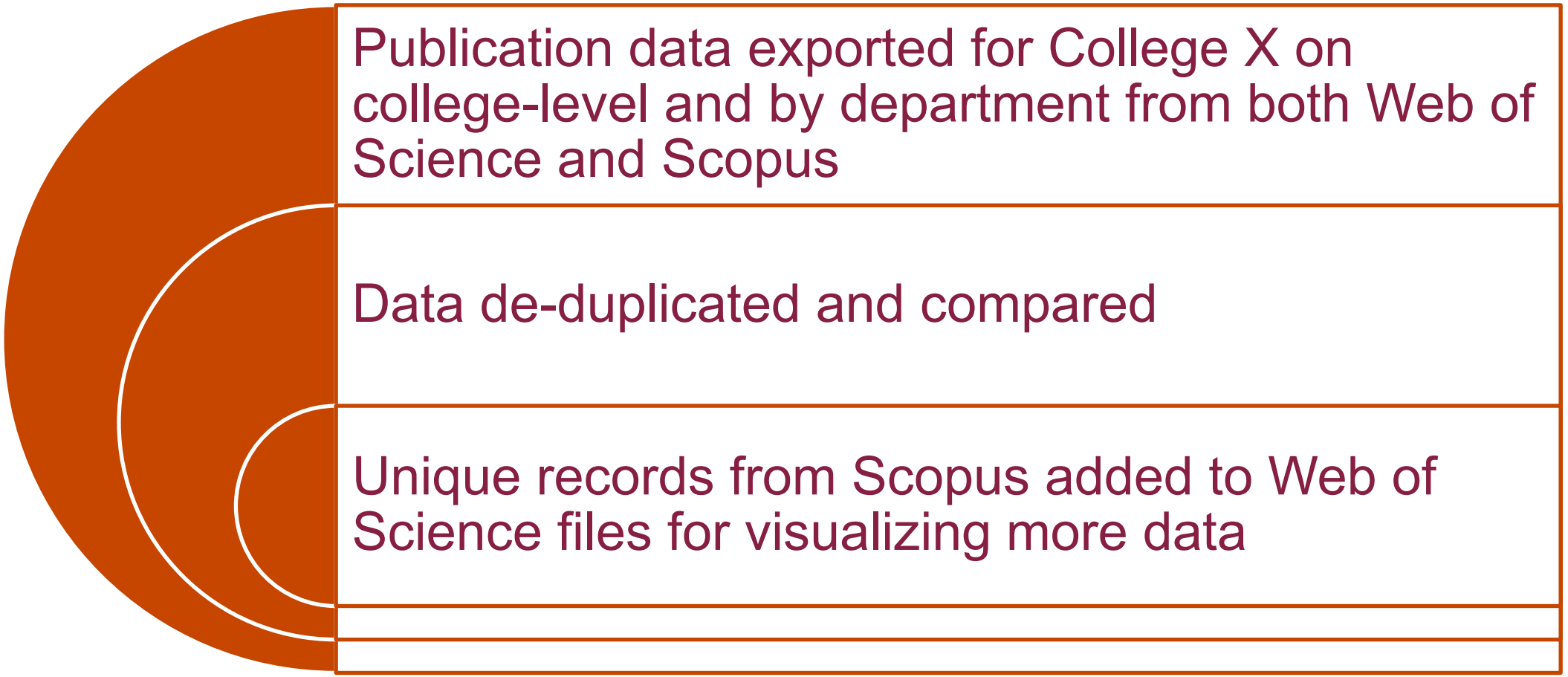
Why aggregate data from more than one bibliographic data source?

Comprehensiveness!

Demonstrate the uniqueness of each database

Create more work for yourself

Study: Export data from Web of Science & Scopus



Metadata Crosswalk

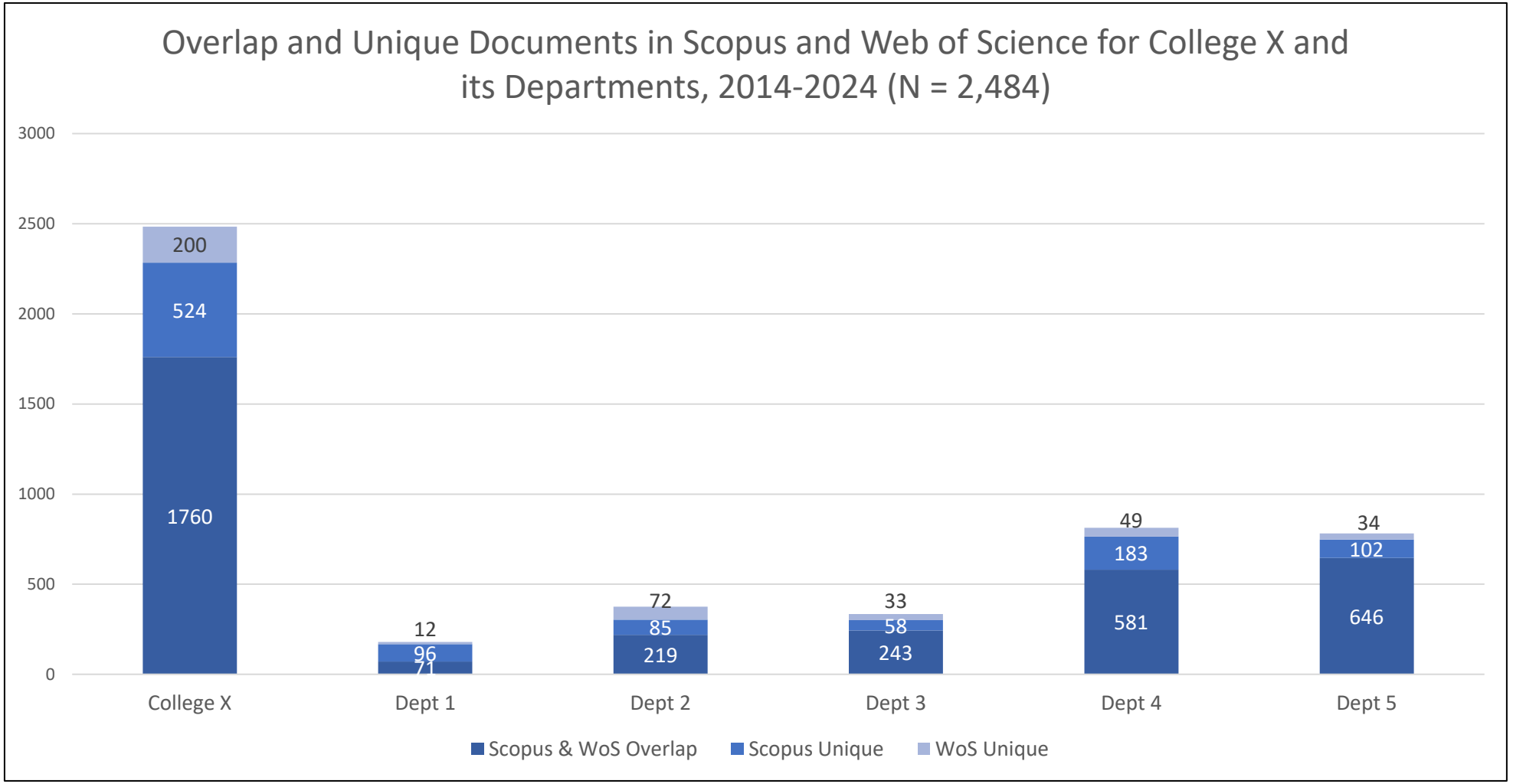
Web of Science Heading / Field	Scopus Heading / Field
AU	Authors
TI	Title
SO	Source Title
DT	Document Type*
LA	Language of Original Document
CT	Conference name
CY	Conference date
CL	Conference location
DE	Author Keywords
ID	Index Keywords
AB	Abstract
C3	Affiliation
FU	Funding details
CR	References
TC 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.

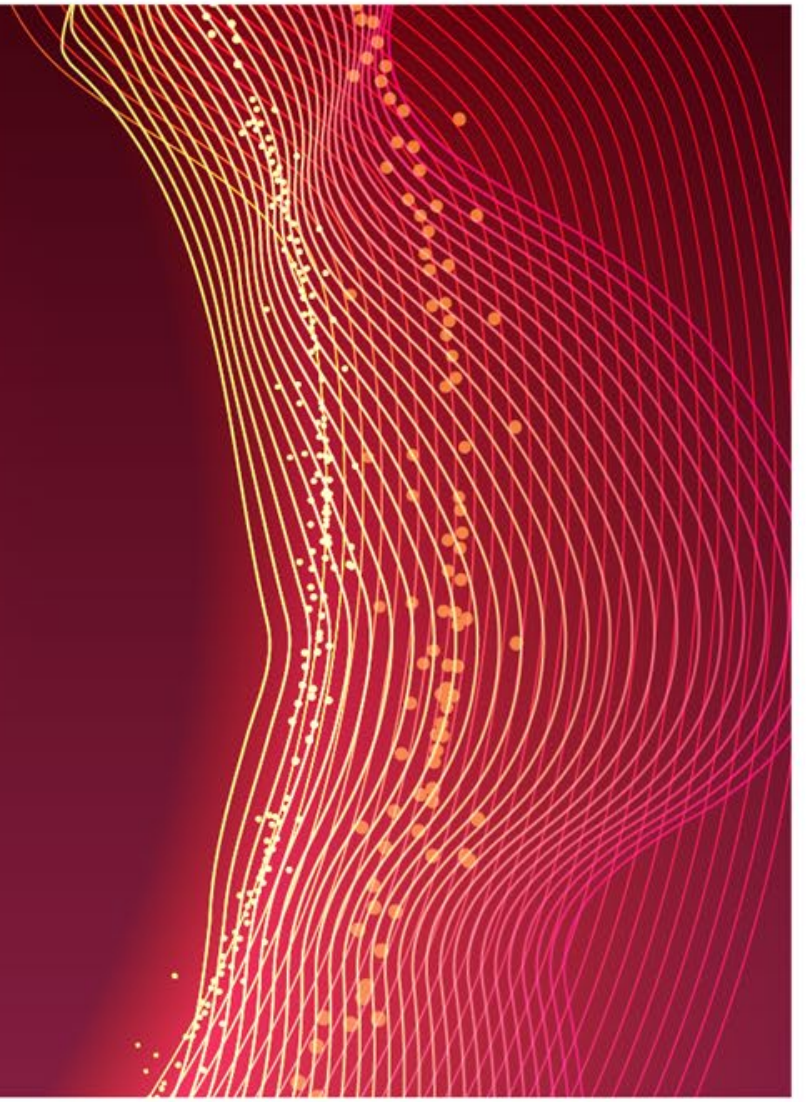
**Z9 is the total times cited across all Web of Science databases. TC is the total times cited across the Web of Science Core Collection.

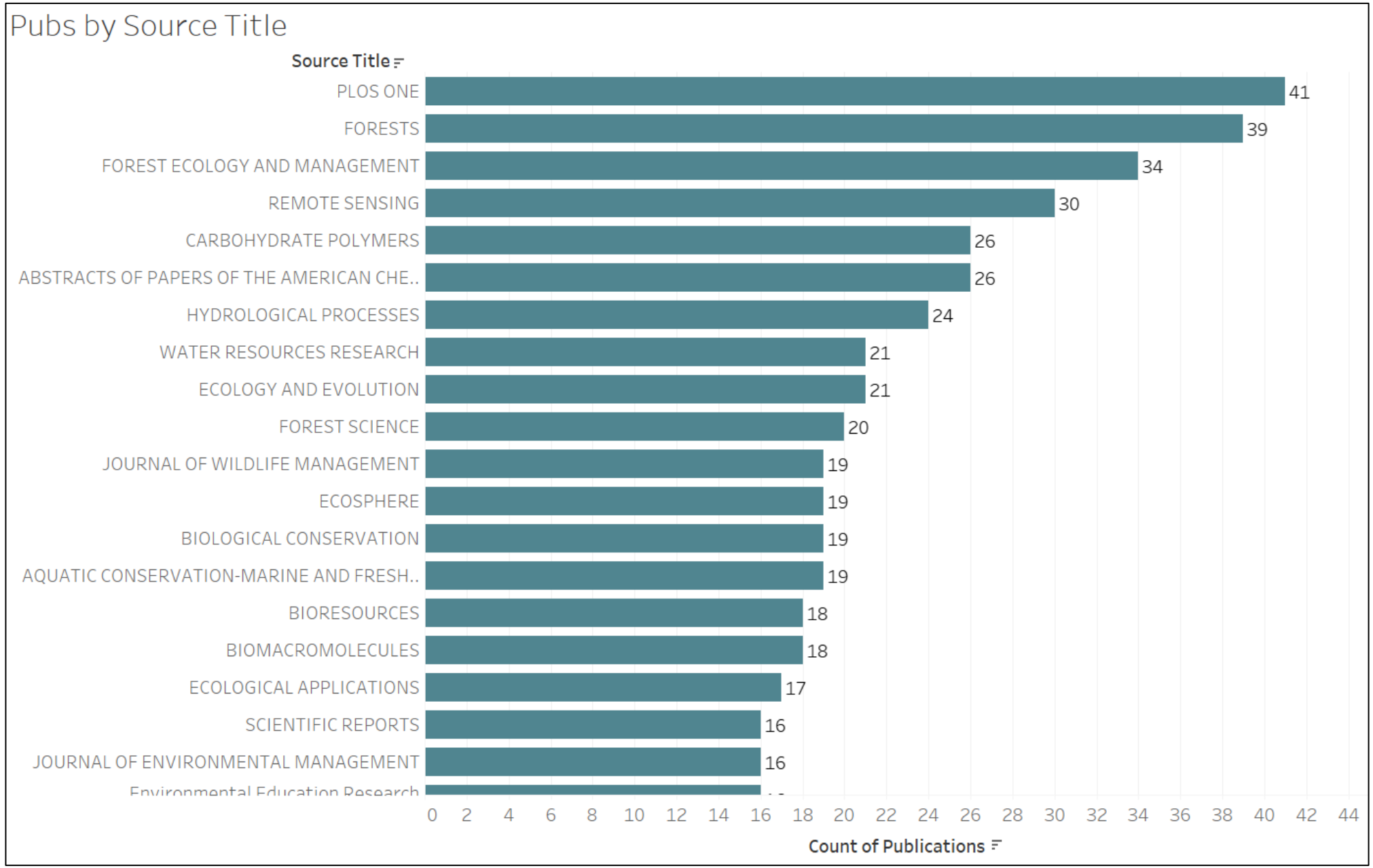
For details on all the Web of Science Core Collection field tags, see https://images.webofknowledge.com/images/help/WOS/hs_wos_fieldtags.html

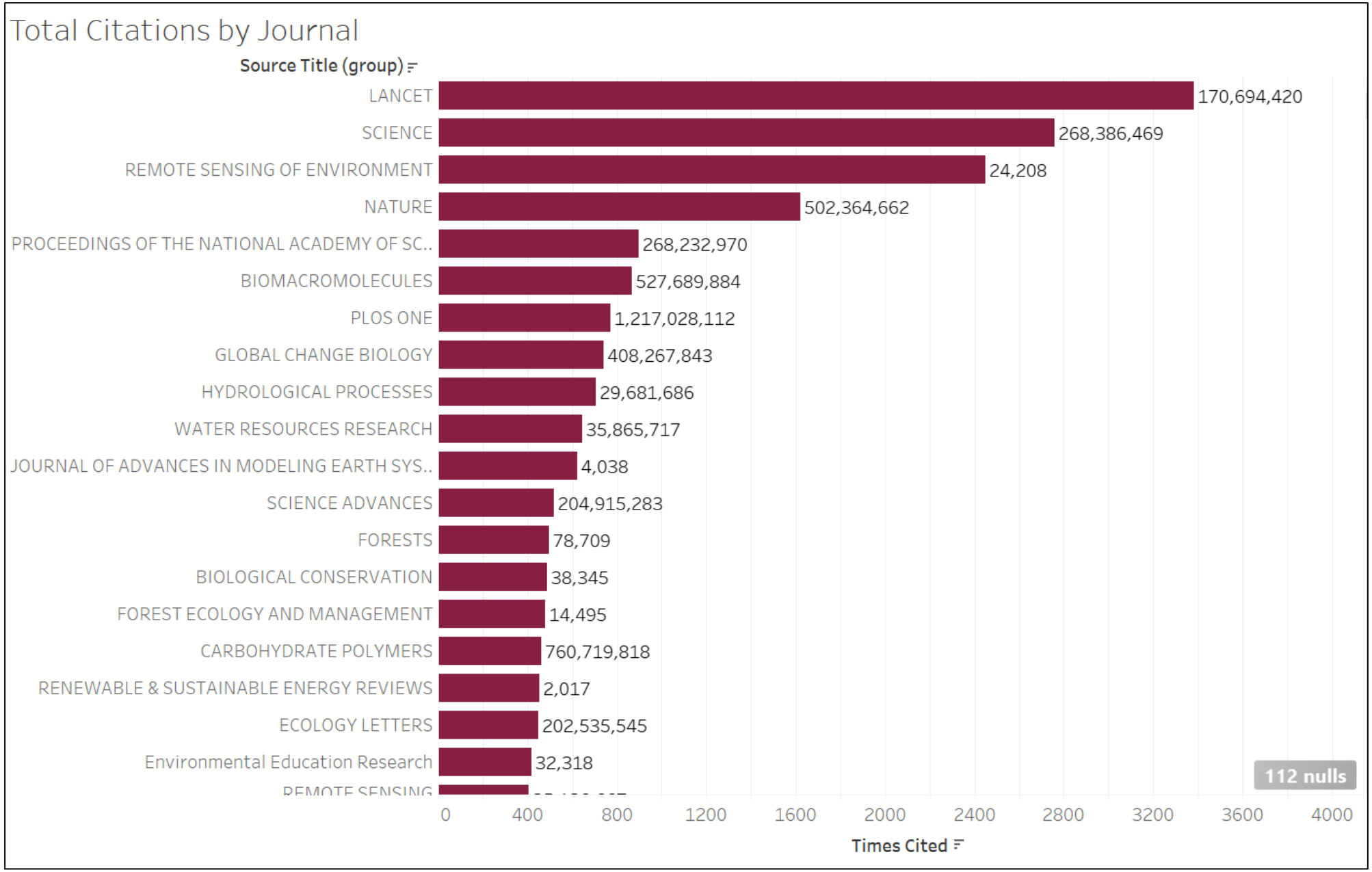
Overlap & Unique Documents in Scopus & WoS for College X, 2014-2024

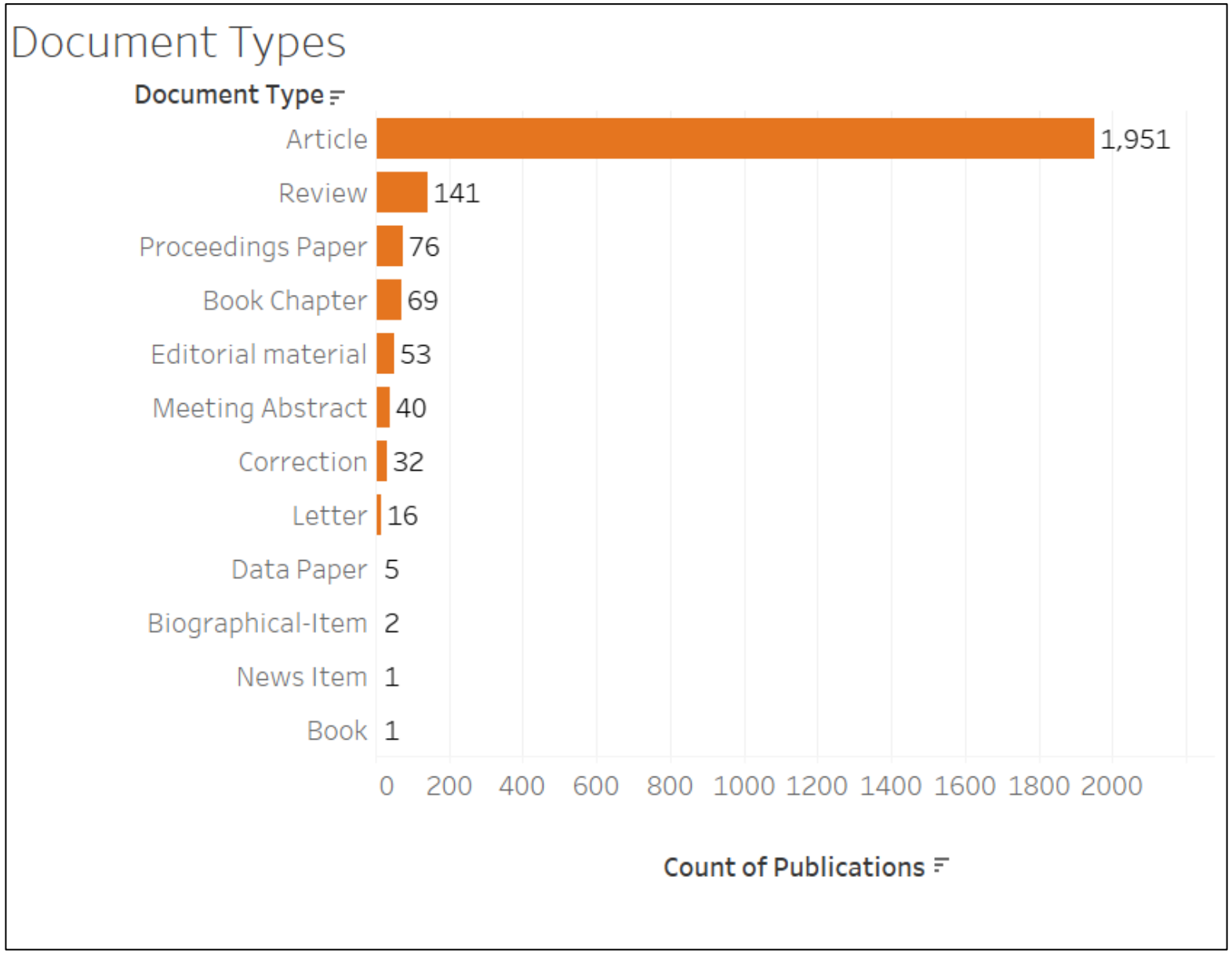


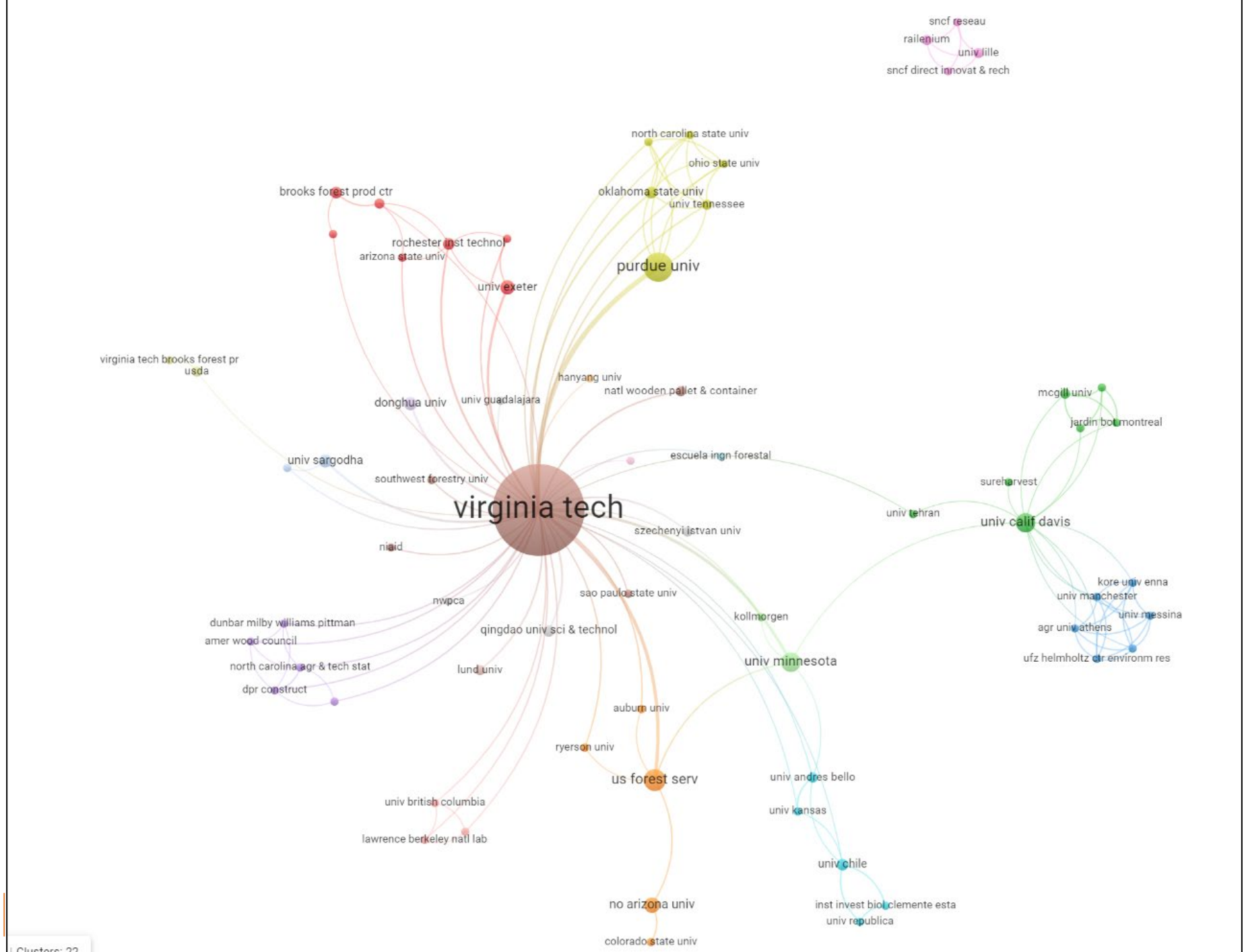
Examples of Graphs and Visualizations from the Aggregated Data



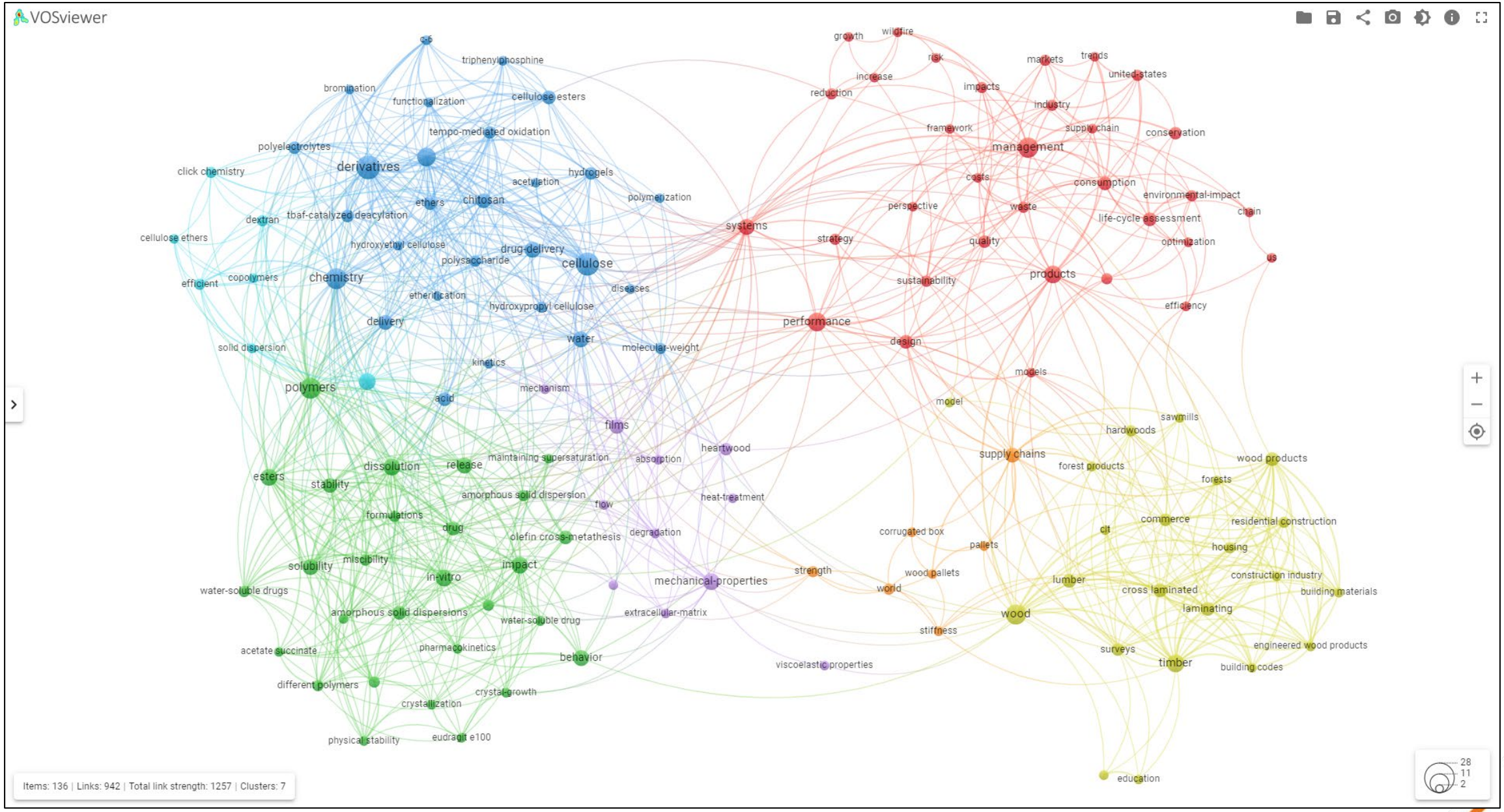

Pubs by Source Title
Source Title
PLOS ONE 41
FORESTS 39
FOREST ECOLOGY AND MANAGEMENT 34
REMOTE SENSING 30
CARBOHYDRATE POLYMERS 26
ABSTRACTS OF PAPERS OF THE AMERICAN CHE.. 26
HYDROLOGICAL PROCESSES 24
WATER RESOURCES RESEARCH 21
ECOLOGY AND EVOLUTION 21
FOREST SCIENCE 20
JOURNAL OF WILDLIFE MANAGEMENT 19
ECOSPHERE 19
BIOLOGICAL CONSERVATION 19
AQUATIC CONSERVATION-MARINE AND FRESH.. 19
BIORESOURCES 18
BIOMACROMOLECULES 18
ECOLOGICAL APPLICATIONS 17
SCIENTIFIC REPORTS 16
JOURNAL OF ENVIRONMENTAL MANAGEMENT 16
Environmental Education Research
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Count of Publications

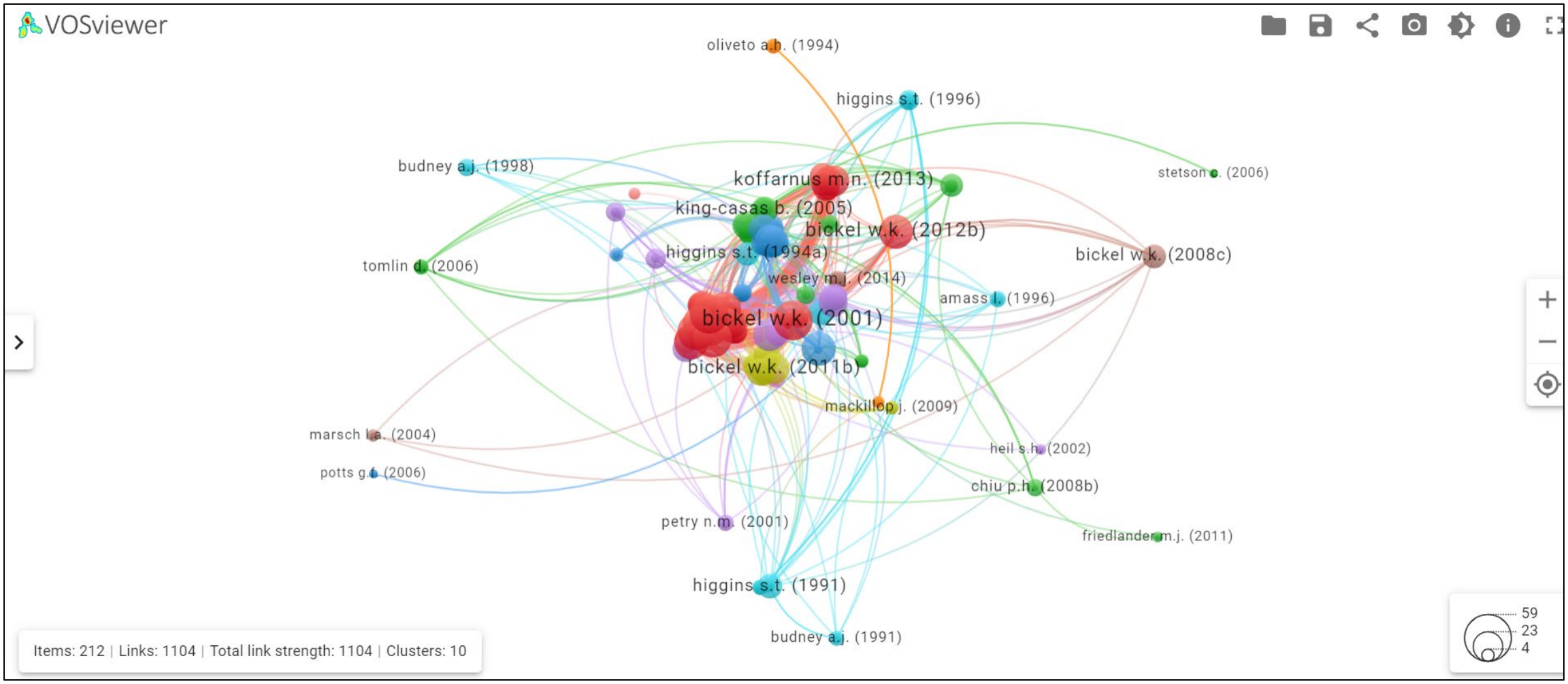

Total Citations by Journal
Source Title (group)
LANCET
SCIENCE
REMOTE SENSING OF ENVIRONMENT
NATURE
PROCEEDINGS OF THE NATIONAL ACADEMY OF SC..
BIOMACROMOLECULES
PLOS ONE
GLOBAL CHANGE BIOLOGY
HYDROLOGICAL PROCESSES
WATER RESOURCES RESEARCH
JOURNAL OF ADVANCES IN MODELING EARTH SYS..
SCIENCE ADVANCES
FORESTS
BIOLOGICAL CONSERVATION
FOREST ECOLOGY AND MANAGEMENT
CARBOHYDRATE POLYMERS
RENEWABLE & SUSTAINABLE ENERGY REVIEWS
ECOLOGY LETTERS
Environmental Education Research
170,694,420
268,386,469
24,208
502,364,662
268,232,970
527,689,884
1,217,028,112
408,267,843
29,681,686
35,865,717
4,038
204,915,283
78,709
38,345
14,495
760,719,818
2,017
202,535,545
32,318
112 nulls
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
Times Cited


Document Types
Document Type
Article
Review
Proceedings Paper
Book Chapter
Editorial material
Meeting Abstract
Correction
Letter
Data Paper
Biographical-Item
News Item
Book
1,951
141
76
69
53
40
32
16
5
2
1
1
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Count of Publications


sncf reseau
railenium
univ lille
sncf direct innovat & rech
north carolina state univ
ohio state univ
oklahoma state univ
univ tennessee
brooks forest prod ctr
rochester inst technol
arizona state univ
purdue univ
univ exeter
virginia tech brooks forest pr
usda
hanyang univ
natl wooden pallet & container
donghua univ
univ guadalajara
mcgill univ
jardin bot montreal
univ sargodha
escuela ingn forestal
southwest forestry univ
surefharvest
univ tehran
virginia tech
univ calif davis
szechenyi istvan univ
niaid
kore univ enna
univ manchester
univ messina
sao paulo state univ
nwpca
kollmorgen
dunbar milby williams pittman
qingdao univ sci & technol
agr univ athens
amer wood council
univ minnesota
ufz helmholtz ctr environm res
north carolina agr & tech stat
lund univ
dpr construct
auburn univ
ryerson univ
us forest serv
univ andres bello
univ british columbia
univ kansas
lawrence berkeley natl lab
univ chile
no arizona univ
inst invest bio clemente esta
univ republica
colorado state univ

Clusters: 22


VOSviewer
c-6
triphenylphosphine
bromination
functionalization
cellulose esters
tempo-mediated oxidation
polyelectrolytes
derivatives
click chemistry
hydrogels
acetylation
ethers
chitosan
polymerization
dextran
tbaf-catalyzed deacylation
cellulose ethers
hydroxyethyl cellulose
drug-delivery
polysaccharide
cellulose
efficient
copolymers
chemistry
diseases
etherification
hydroxypropyl cellulose
delivery
solid dispersion
water
molecular-weight
kinetics
polymers
mechanism
acid
films
heartwood
maintaining supersaturation
absorption
dissolution
release
esters
stability
amorphous solid dispersion
heat-treatment
flow
formulations
drug
olefin cross-metathesis
degradation
solubility
miscibility
impact
in-vitro
water-soluble drugs
mechanical-properties
strength
amorphous solid dispersions
water-soluble drug
extracellular-matrix
acetate succinate
pharmacokinetics
behavior
different polymers
crystal-growth
crystallization
physical stability
eudragit e100
growth
wildfire
risk
increase
reduction
markets
trends
united-states
impacts
industry
framework
supply chain
conservation
management
costs
consumption
environmental-impact
perspective
waste
life-cycle assessment
chain
systems
strategy
quality
optimization
us
products
sustainability
efficiency
performance
design
models
model
sawmills
hardwoods
supply chains
forest products
wood products
forests
commerce
cit
residential construction
corrugated box
pallets
housing
wood pallets
world
lumber
construction industry
cross laminated
building materials
laminating
wood
stiffness
surveys
engineered wood products
timber
building codes
viscoelastic properties
education
Items: 136 | Links: 942 | Total link strength: 1257 | Clusters: 7
28
11
2

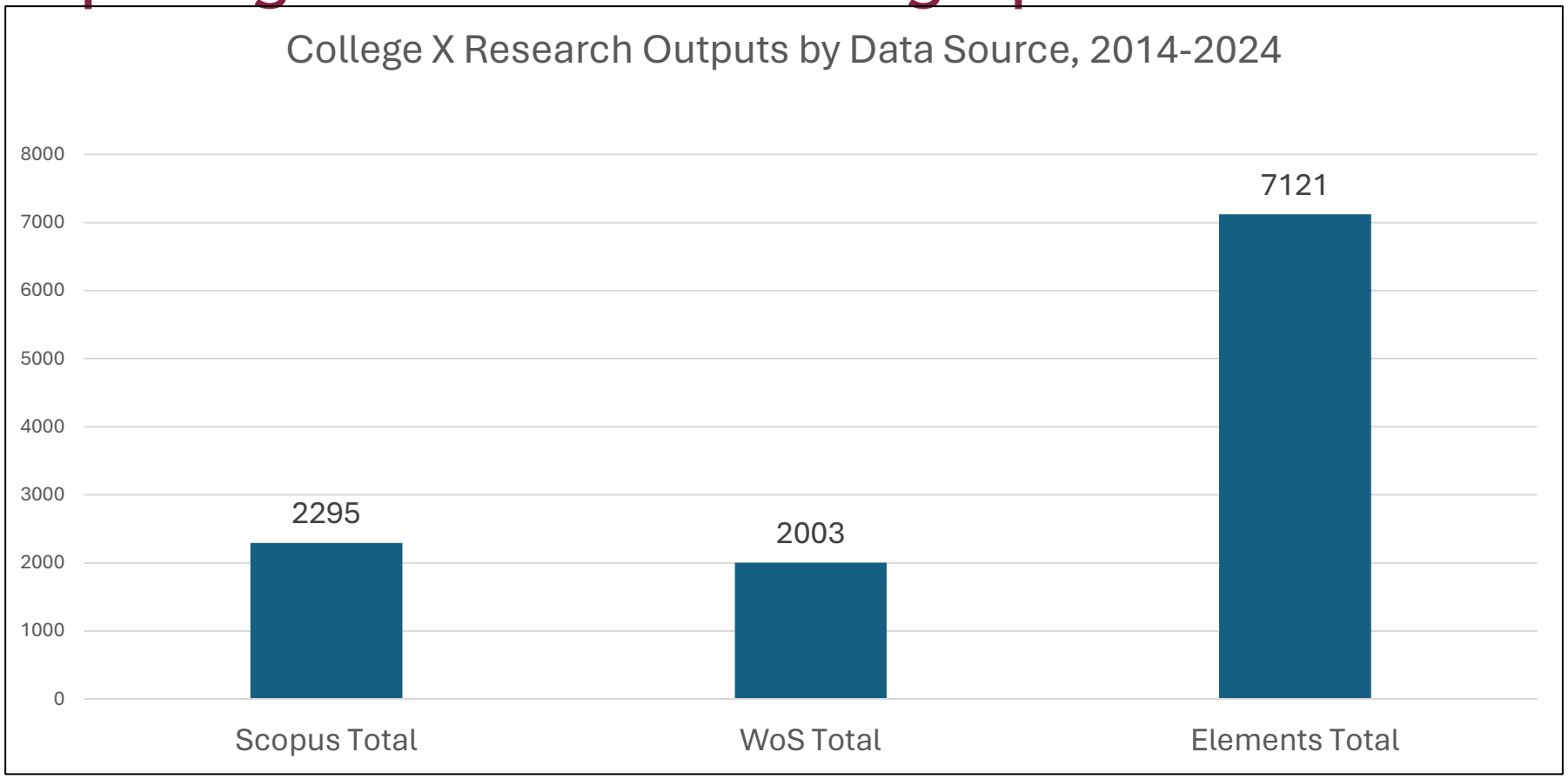

VOSviewer
oliveto a.h. (1994)
higgins s.t. (1996)
budney a.j. (1998)
koffarnus m.n. (2013)
stetson c. (2006)
king-casas b. (2005)
bickel w.k. (2012b)
higgins s.t. (1994a)
bickel w.k. (2008c)
tomlin d. (2006)
wesley m.j. (2014)
amass l. (1996)
bickel w.k. (2001)
bickel w.k. (2011b)
mackillop j. (2009)
marsch l.a. (2004)
heil s.h. (2002)
potts g.f. (2006)
chiu p.h. (2008b)
petry n.m. (2001)
friedlander m.j. (2011)
higgins s.t. (1991)
budney a.j. (1991)
Items: 212 | Links: 1104 | Total link strength: 1104 | Clusters: 10
59
23
4

Comparison of bibliographic data sources to RIM data

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

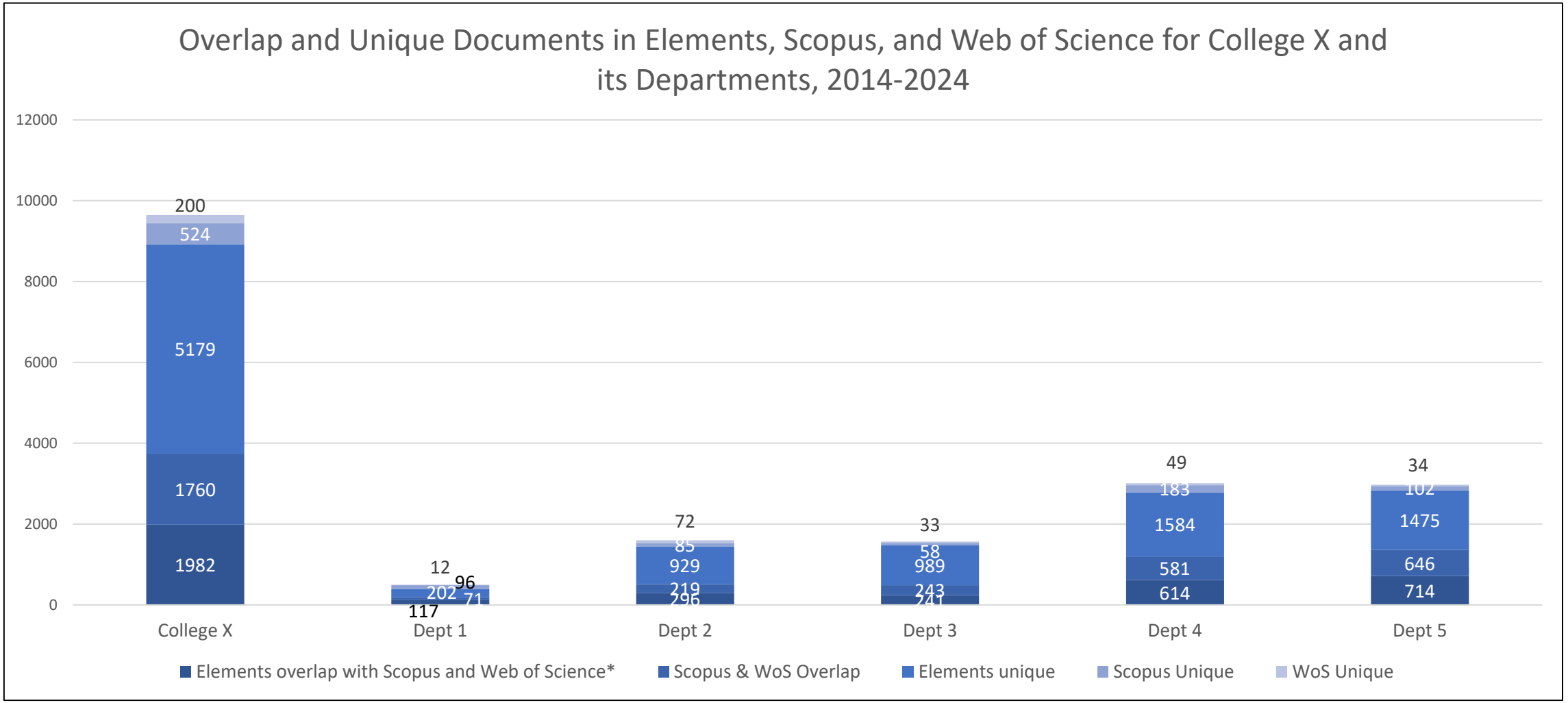
- Easier to do visualizations using VOSviewer and Bibliometrix software tools
- Difficult to crosswalk the data from RIM spreadsheets to Scopus or Web of Science spreadsheets
- Missing metadata that helps with the analyses, such references; cannot do citation analyses in VOSviewer as a result
- It may 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.

UNIVERSITY LIBRARIES
VIRGINIA TECH

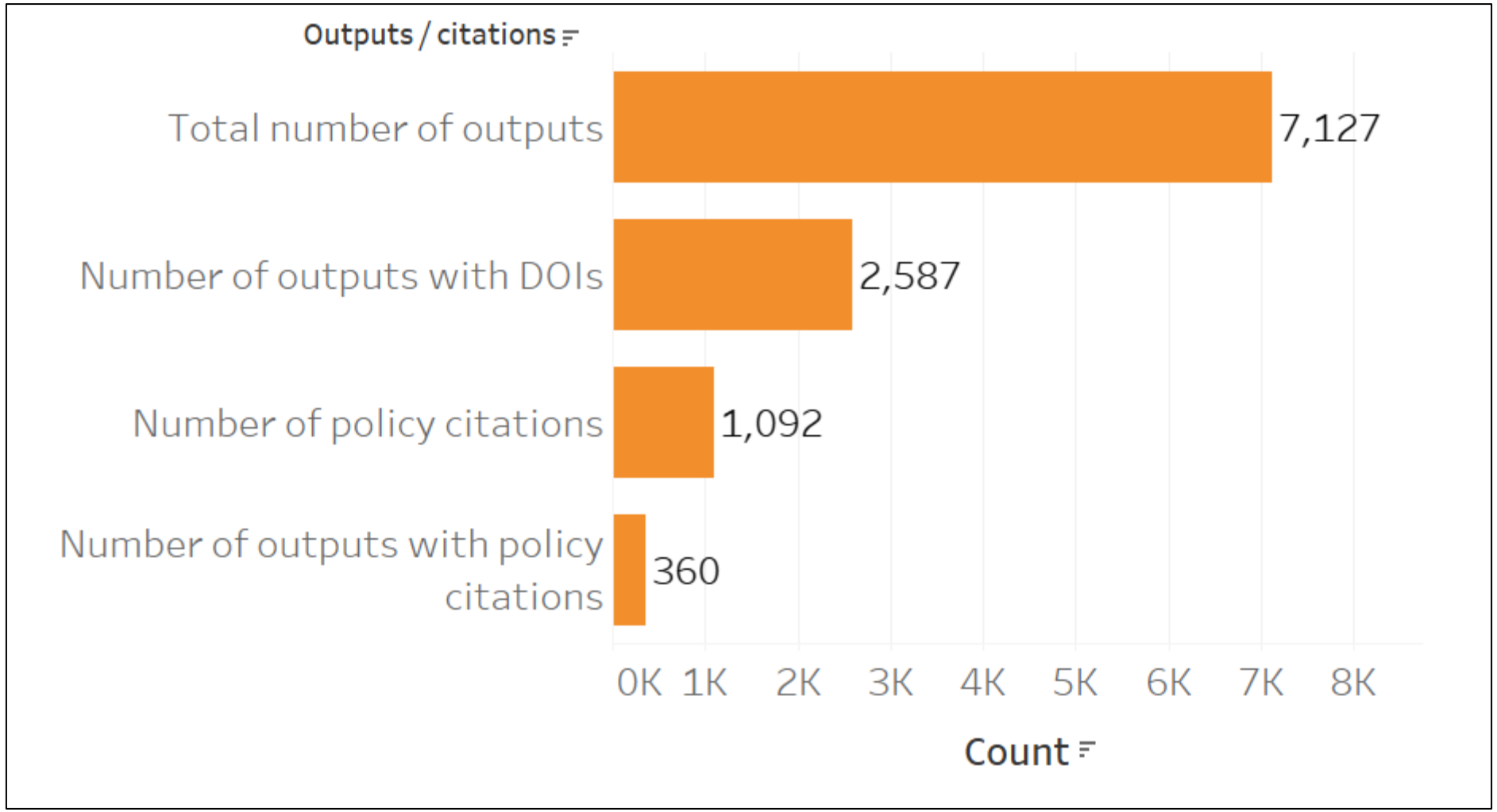
Exporting data from a policy data source & building a dashboard

UNIVERSITY LIBRARIES
VIRGINIA TECH

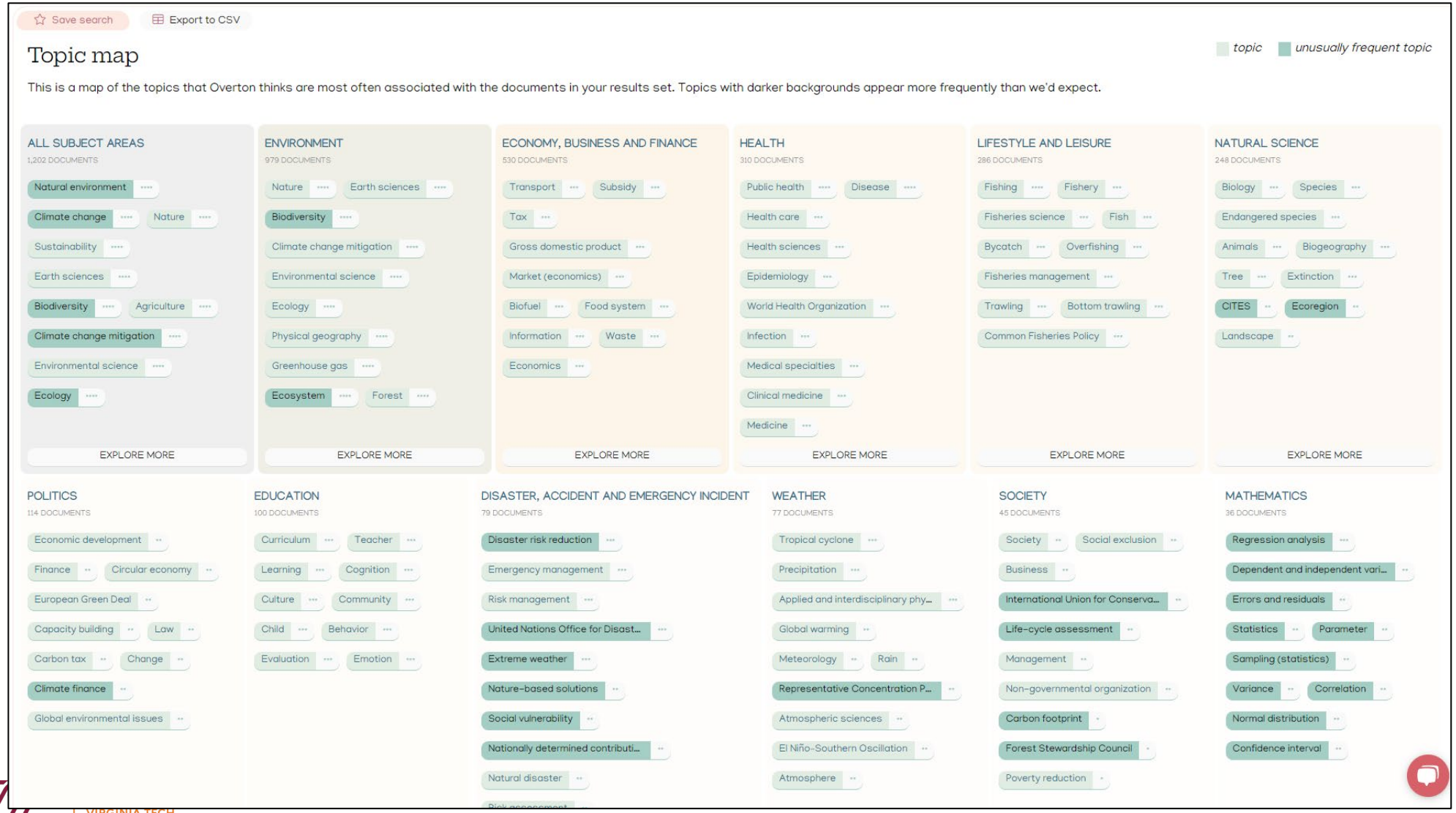
Policy Mentions to Publications

- 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

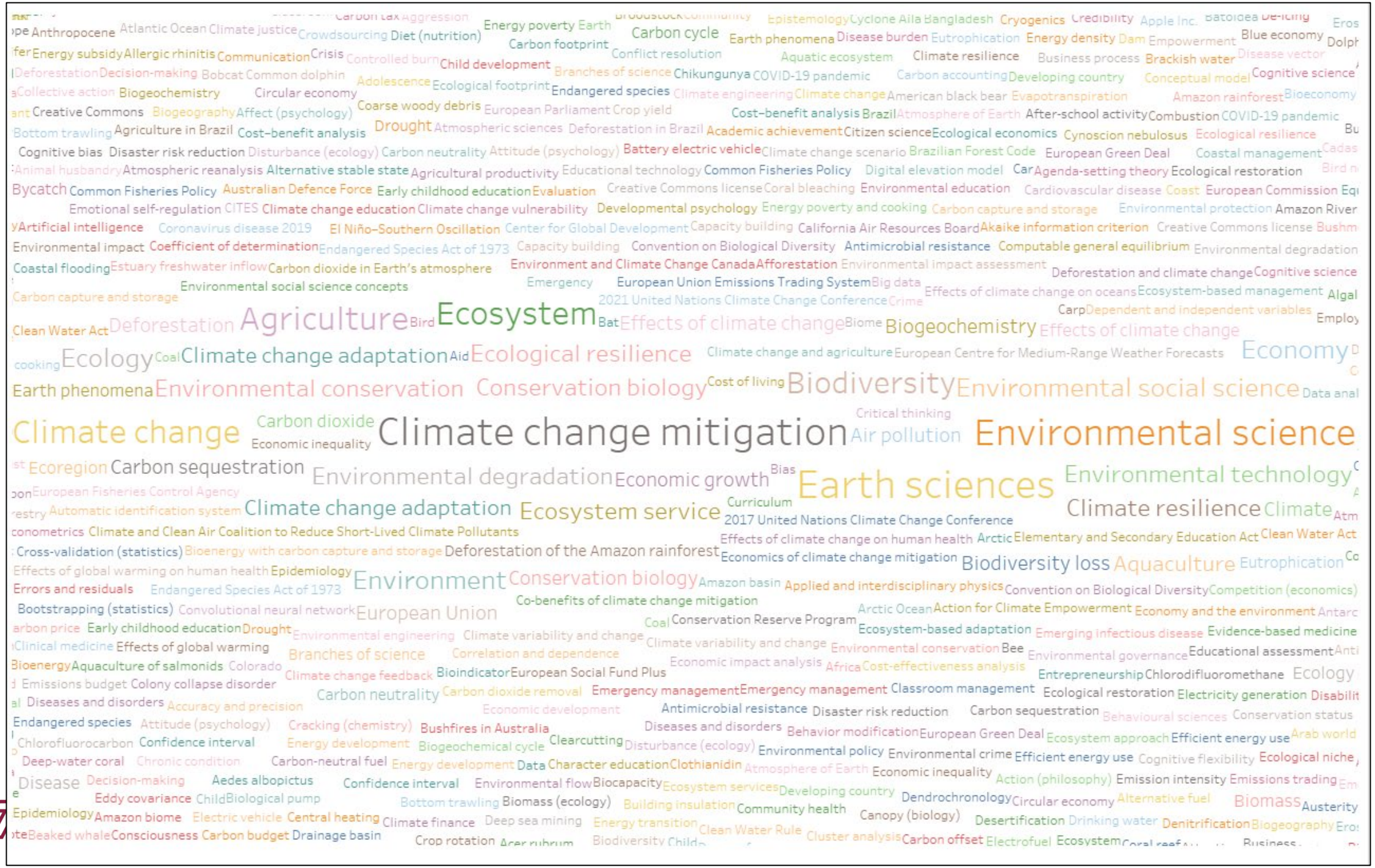
Breakdown of Data from Elements & its policy citations



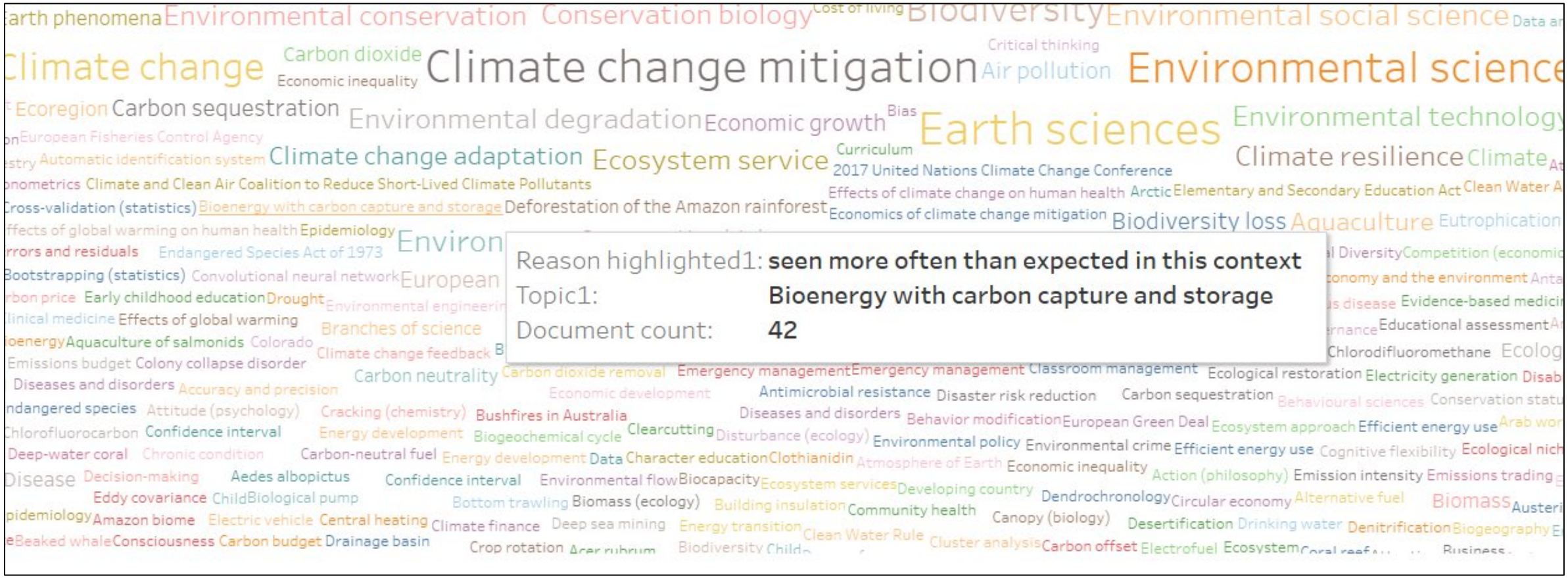
Topic Map in Overton



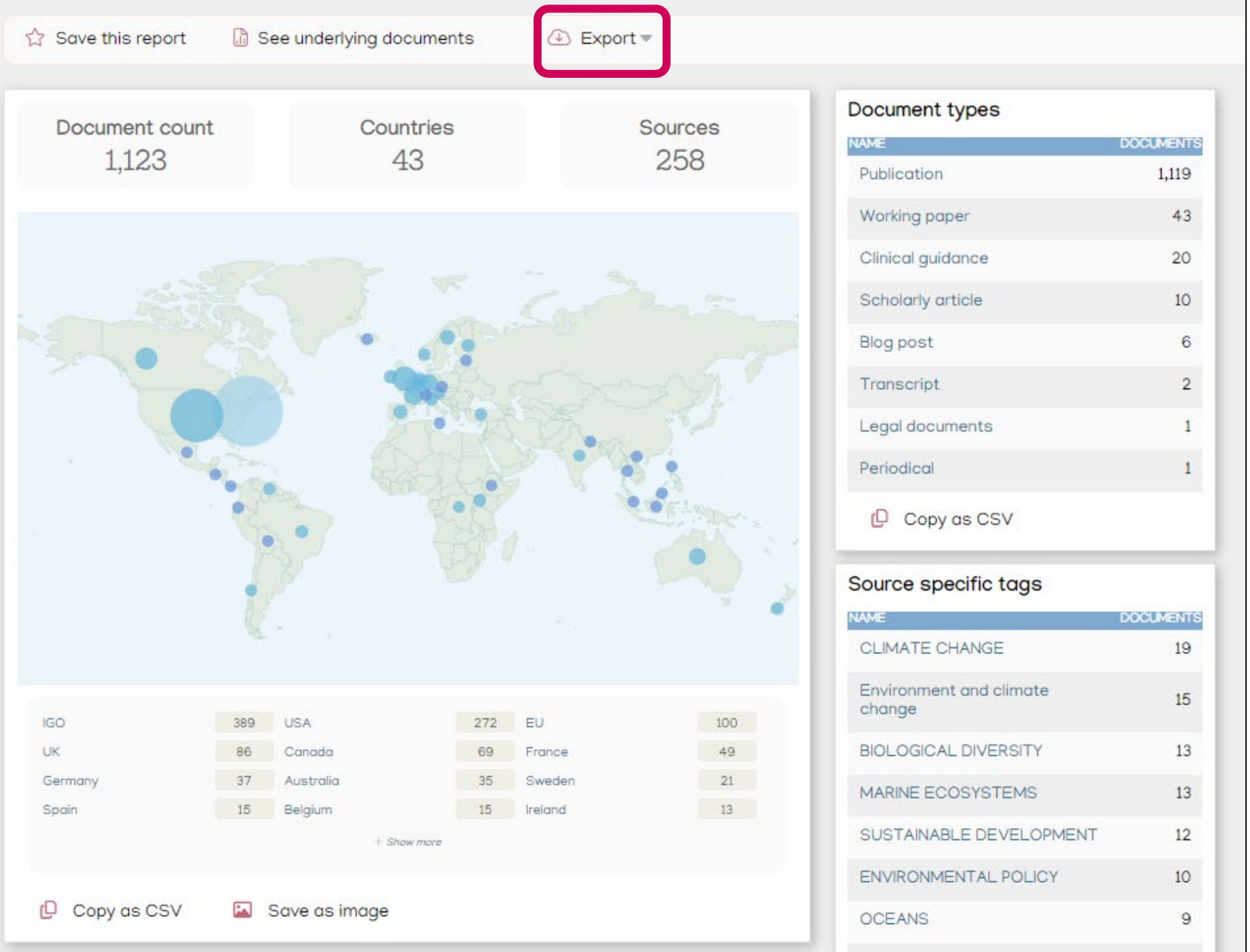
Policy Topics in Tableau



Context of Topics

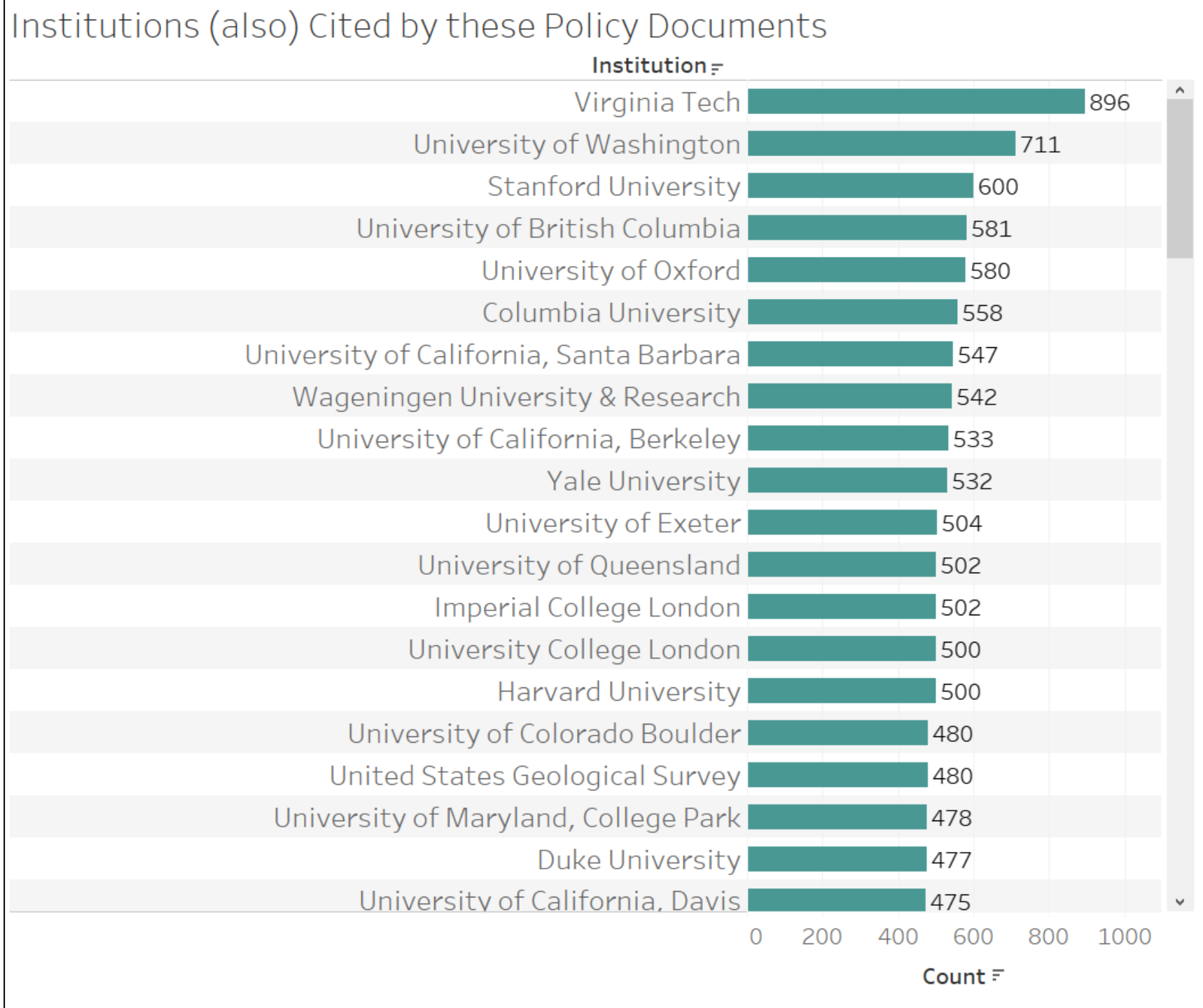


Export from the Report in Overton



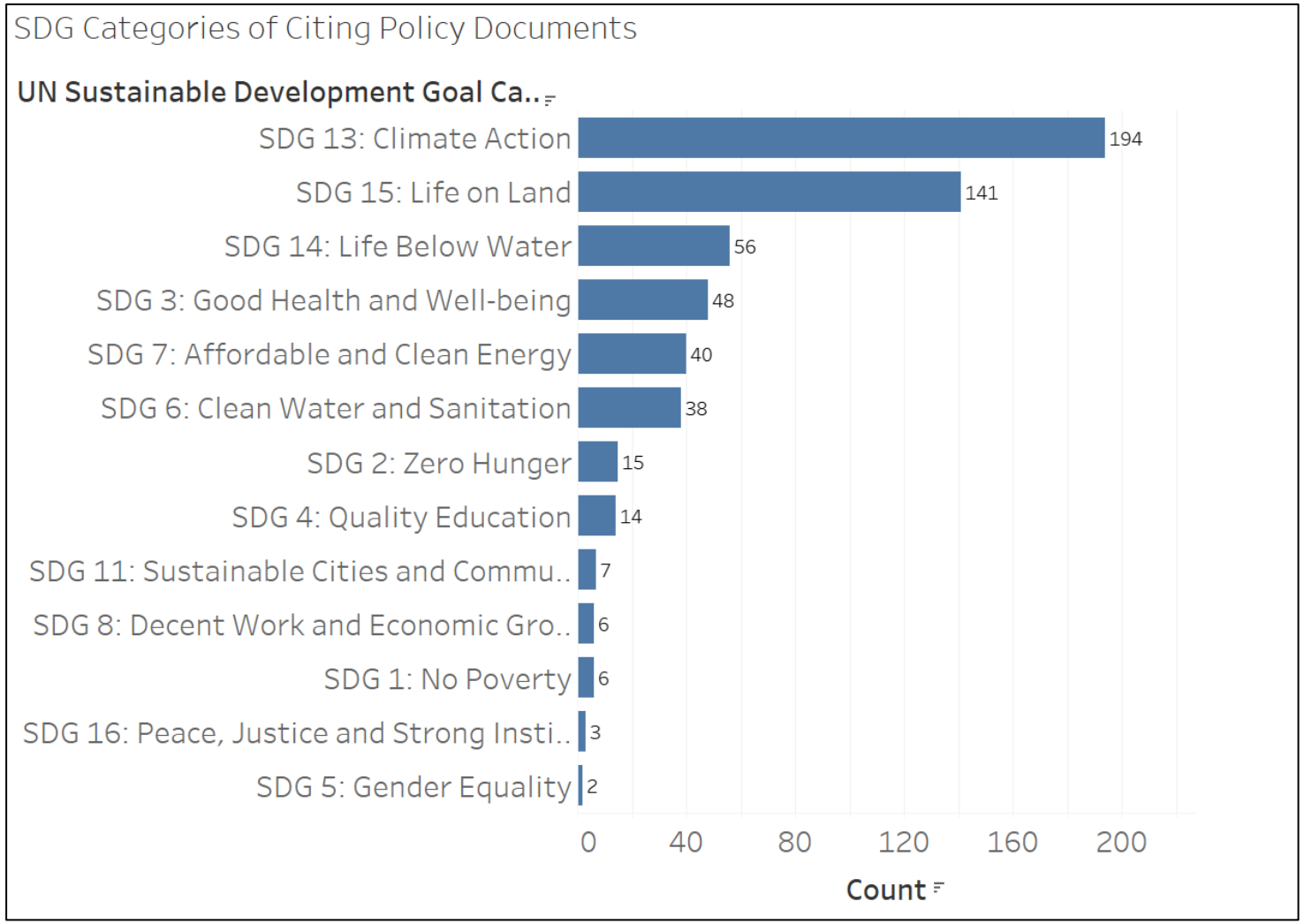
UNIVERSITY LIBRARIES
VIRGINIA TECH

Institutions also cited by these Policy Documents

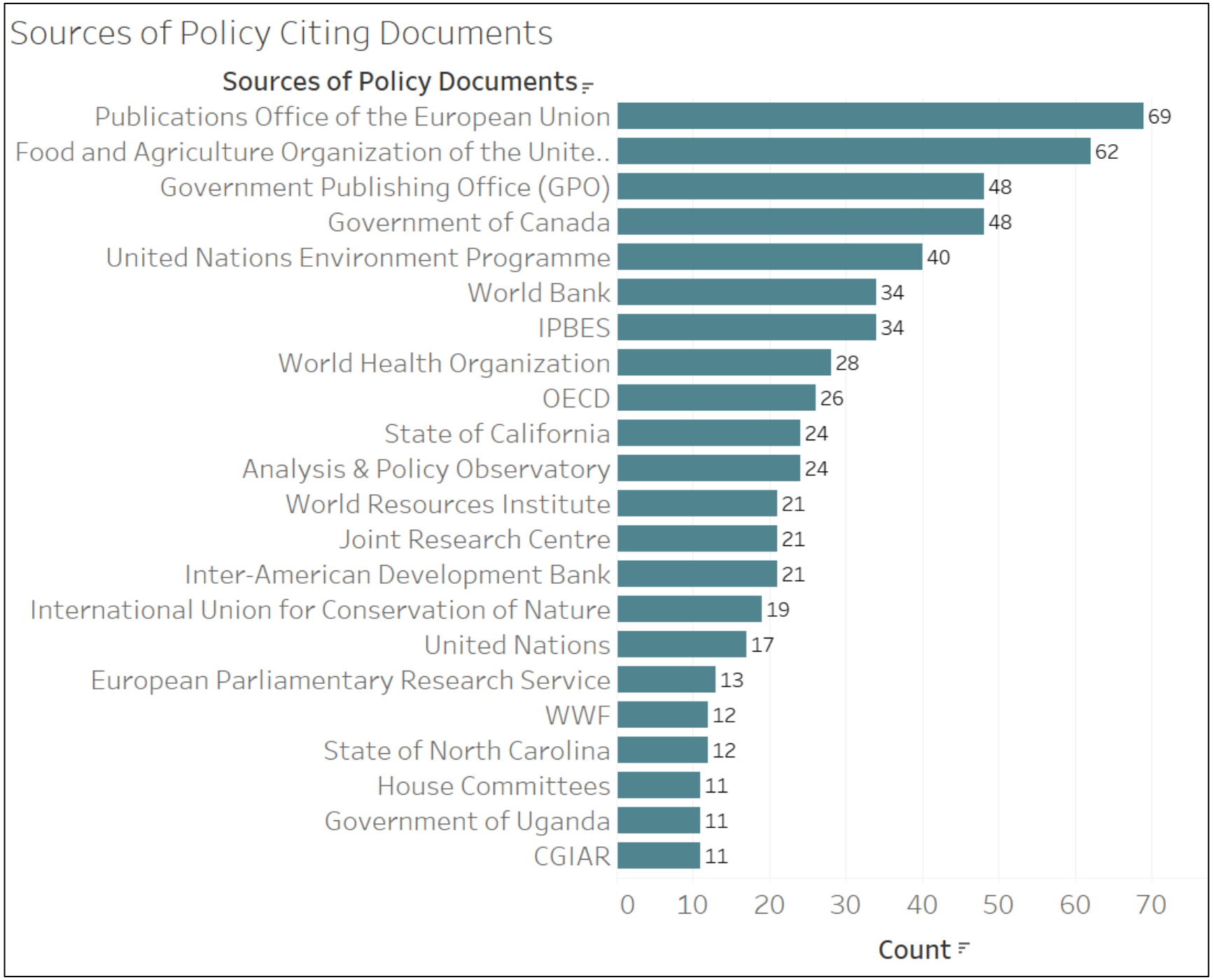


UNIVERSITY LIBRARIES
VIRGINIA TECH

SDG Categories by Citing Policy Documents

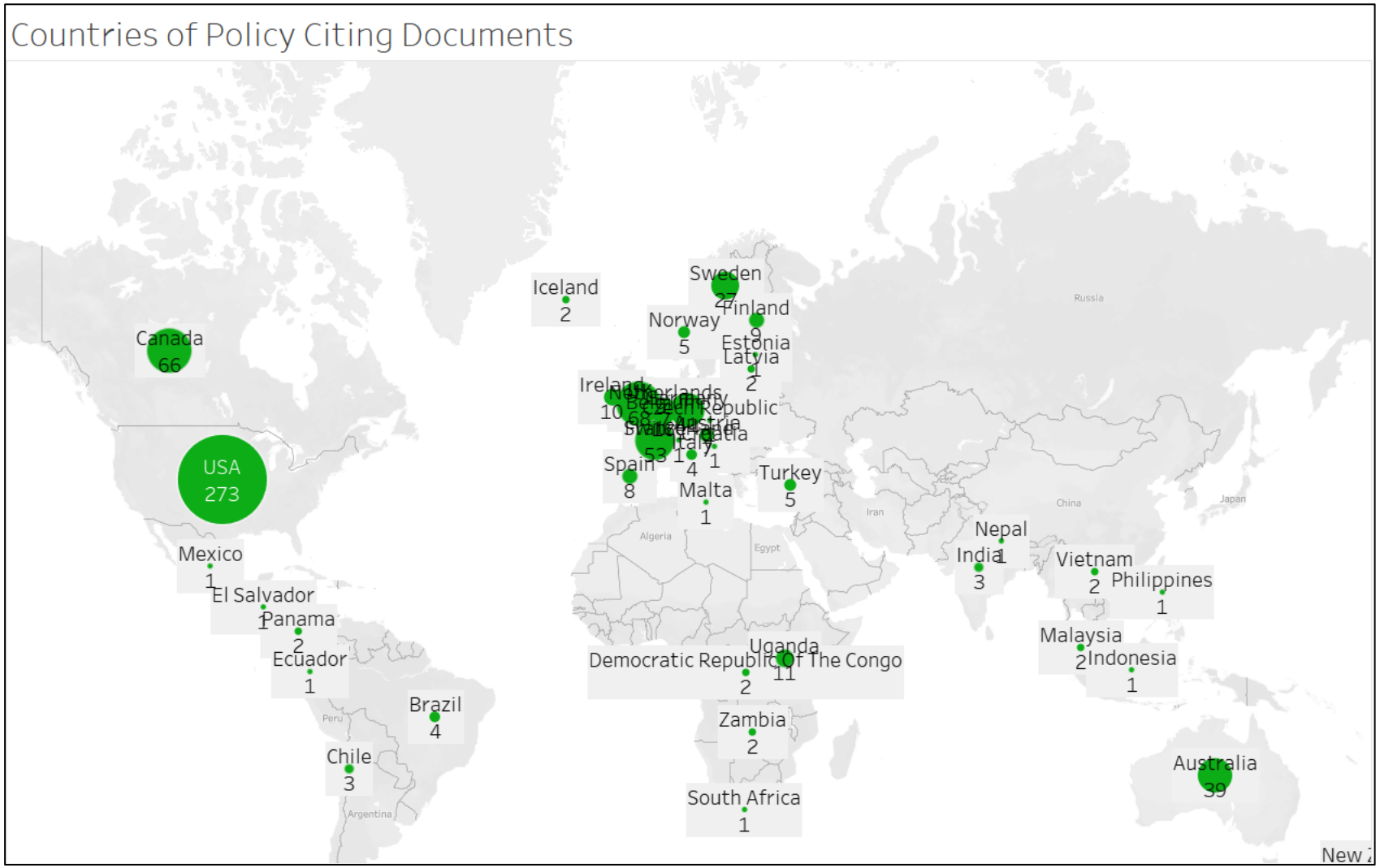


Sources of Policy Citing Documents (Organizations, NGOs, Nonprofits, etc.)



UNIVERSITY LIBRARIES
VIRGINIA TECH

Map of Countries of Policy Citing Documents



Questions?

Thank you!

Rachel Miles

Research Impact Coordinator

ramiles@vt.edu