ICPSR

SHARING DATA TO ADVANCE SCIENCE

A Follow-up Study on Data Management and Data Sharing Training in Graduate Education in the Social Sciences

Ashley Doonan and Evan Cosby

ICPSR INSTITUTE FOR SOCIAL RESEARCH
UNIVERSITY OF MICHIGAN

The Case for Graduate Training

- Need for researchers to be transparent
  - Sharing data is the best way to accomplish this (going beyond writing papers)
- Good data management is essential for good science
  - Replicability, secondary analysis, reducing cost and effort, etc.
- On the job training is insufficient – learning these skills needs to happen in graduate research training
- The role of data science professionals

Previously…

2016-2017 Survey of Graduate Programs

- Programs were confident in research skills (67%), 42% believed graduates have good data management skills

- 20% had an ethics course; only one had a data management course

- Nearly all programs required research methods course
  - Most programs reported data management included in methods
  - Less than half covered data sharing within the methods course

- Programs interested in repository programming
  - 62.5% maybe interested - 20.8% definite interest

(Ebersole & Holt, 2017)

Project Progression

1. ~~Initial Program Survey~~
2. Syllabus Analysis ←
3. Expanded Program Survey
4. Graduate Student Survey
5. Cross Discipline Comparisons

Research Questions

- Objectively, what content are instructors covering in methodology courses?
  - What information do instructors present to graduate students when they are learning how to conduct research?

- Is material regarding data management or data sharing included in the overview of coursework?
  - Is it significant enough in the course to include?
  - Would it be obvious to someone outside of the program that students received instruction on this material?

Methods

1. Creating the Sampling Frame
2. Gathering Syllabi
3. Conducting Text Analysis

Methods - Sampling

- Seven fields included
  - Previous inclusion criteria: Social Science, research oriented, major governing body with an ethical code
  - Matching the previous six inclusions, with the addition of Economics
- A randomly selected list of U.S. programs was created using gradschools.com
  - Only programs with in-person courses were included
  - Mixed inclusion of both Master's and Doctorate programs

Sample

- 140 graduate programs included in syllabus search
  - Random selection included 20 programs per field
  - Private and public universities
  - Schools of all sizes
  - All Carnegie classification levels included, as well as unclassified
  - Within all 50 states

Methods - Collection

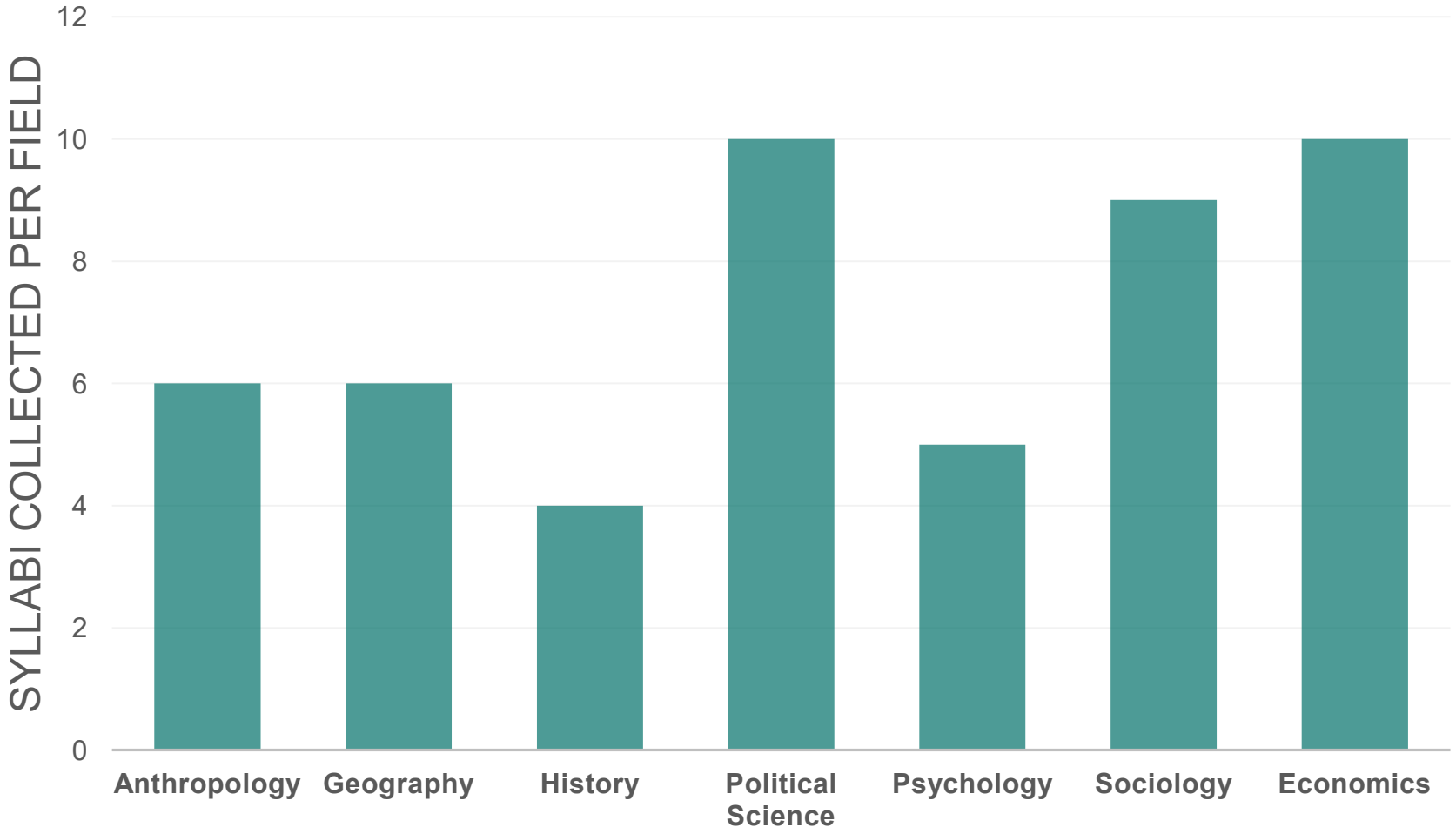
- Required coursework was identified for programs using public webpages

- Syllabi were sought for required courses in the following areas:
  - Methodology, discipline specific statistics, discipline ethics, ethics in research, or similar

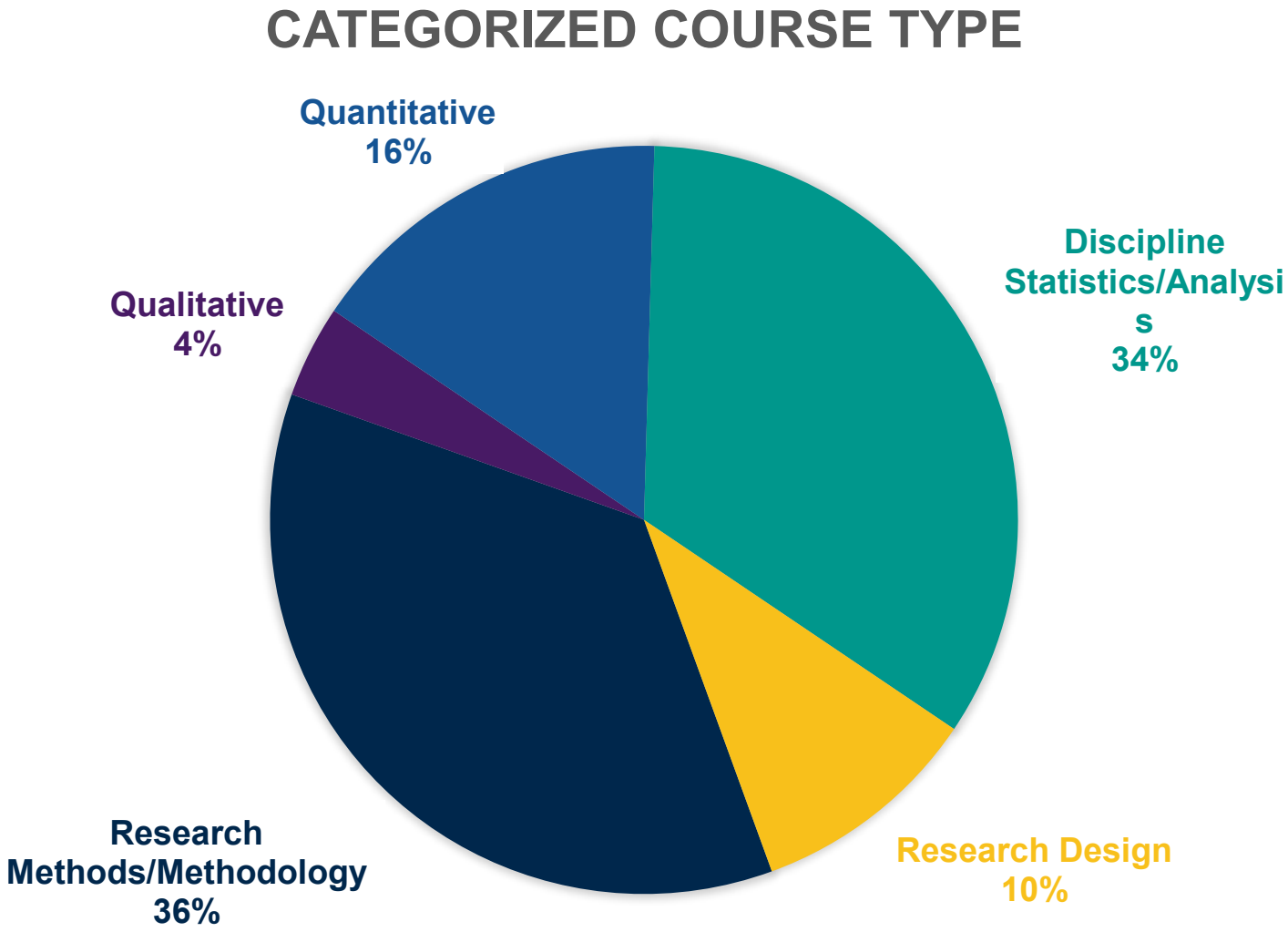
- Syllabi were found in databases such as OER Commons, or on the programs' websites.

- Syllabi were available for 50 programs in total.

Methods – Available Syllabi



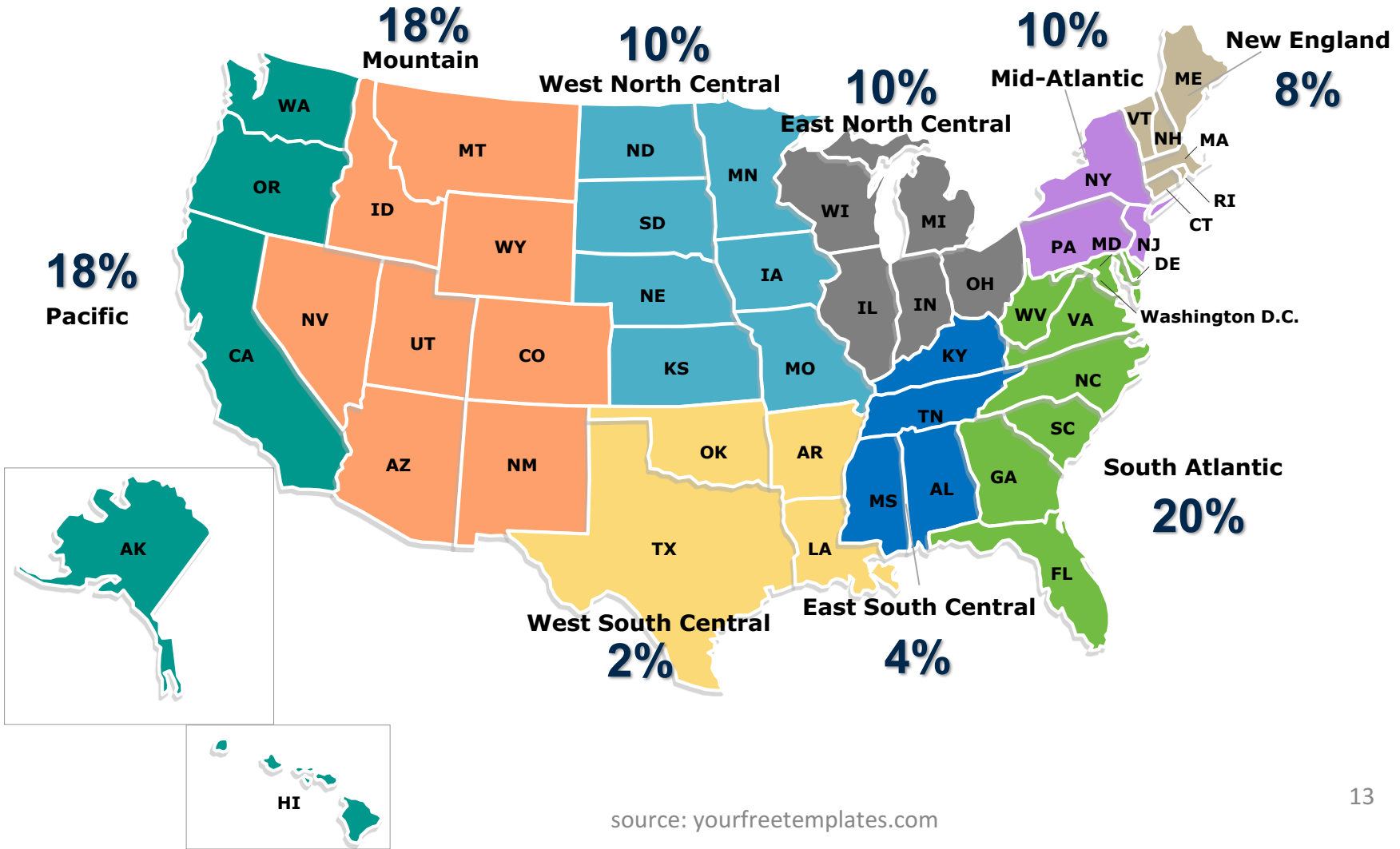
Syllabi Source Characteristics



Syllabi Source Characteristics

- Course Requirements:
  - Completion of a research project (66%)
  - Collection of data (30%)
- Program Type
  - Masters (28%)
  - Doctoral (72%)

Syllabi Source Characteristics



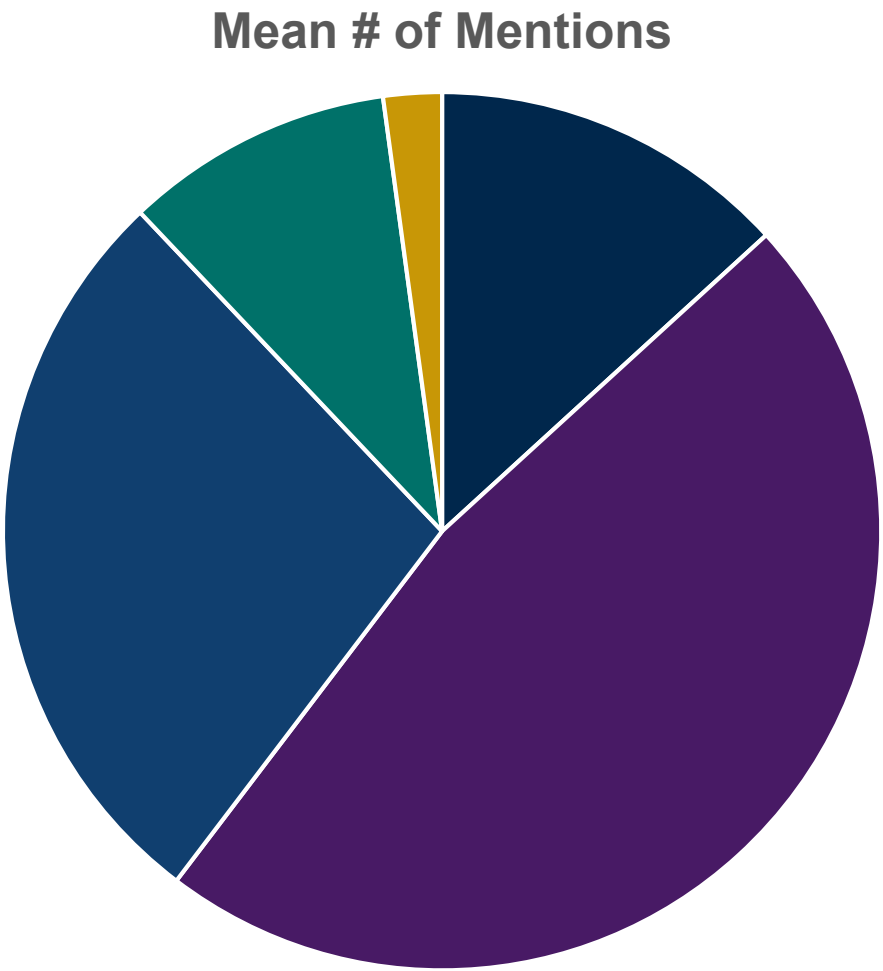
source: yourfreetemplates.com

Methods – Text Analysis

- Search function and visual scan
  - Highlighting applicable words or phrases
  - Full document search
  - Headers and titles were excluded

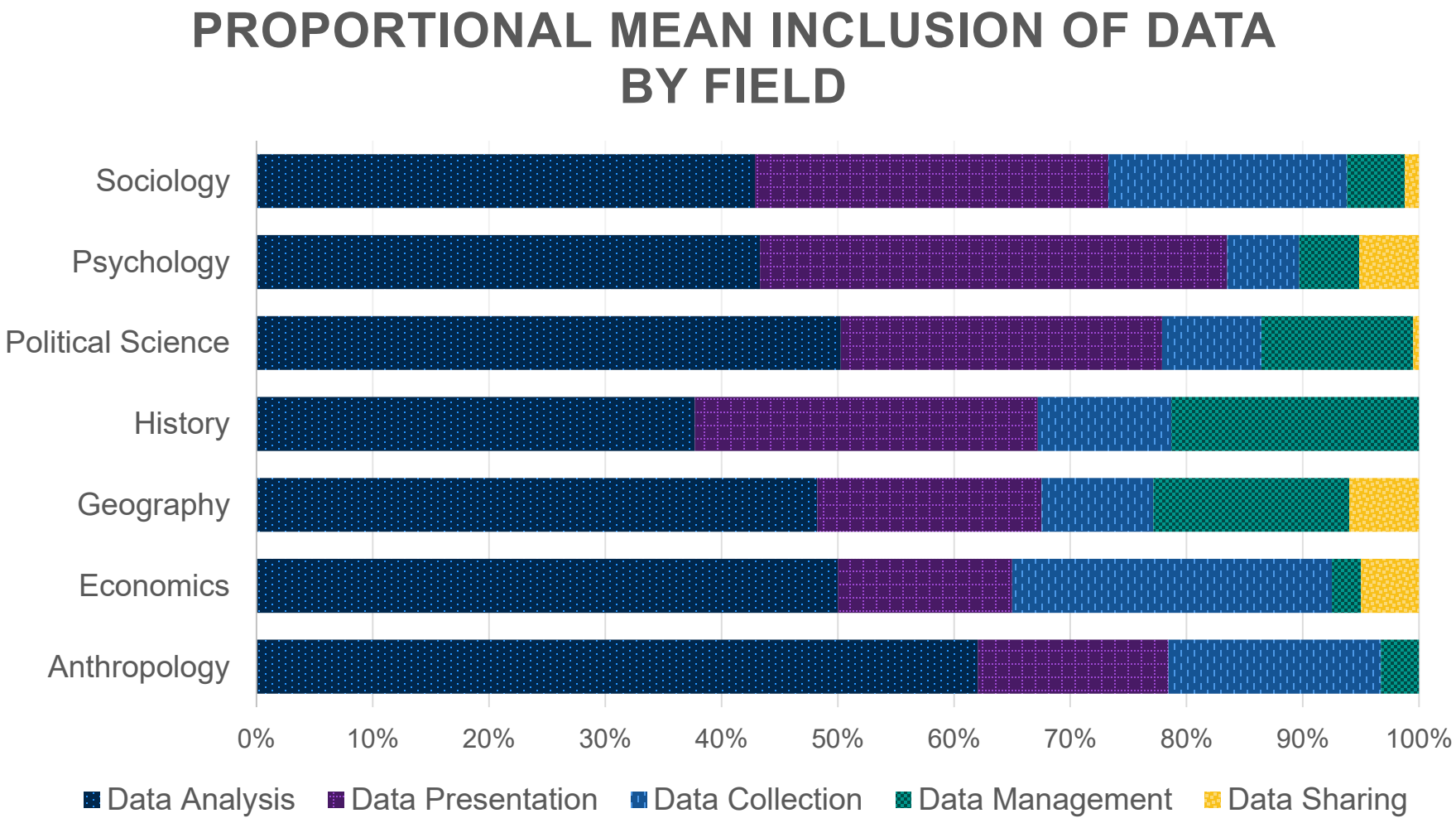
- Allowing for permutations or implied mentions:
  - Synonyms or similar meanings were used
  - "Survey data collection, interpretation, and analysis" was 3 total mentions

Results



	Mean	SD
Data Total	11.56	9.06
Collection	1.84	6.84
Analysis	6.56	6.84
Interpretation	3.84	3.58
Management	1.38	2.47
Sharing	.30	.84

Results


PROPORTIONAL MEAN INCLUSION OF DATA BY FIELD
Sociology
Psychology
Political Science
History
Geography
Economics
Anthropology
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Data Analysis
Data Presentation
Data Collection
Data Management
Data Sharing

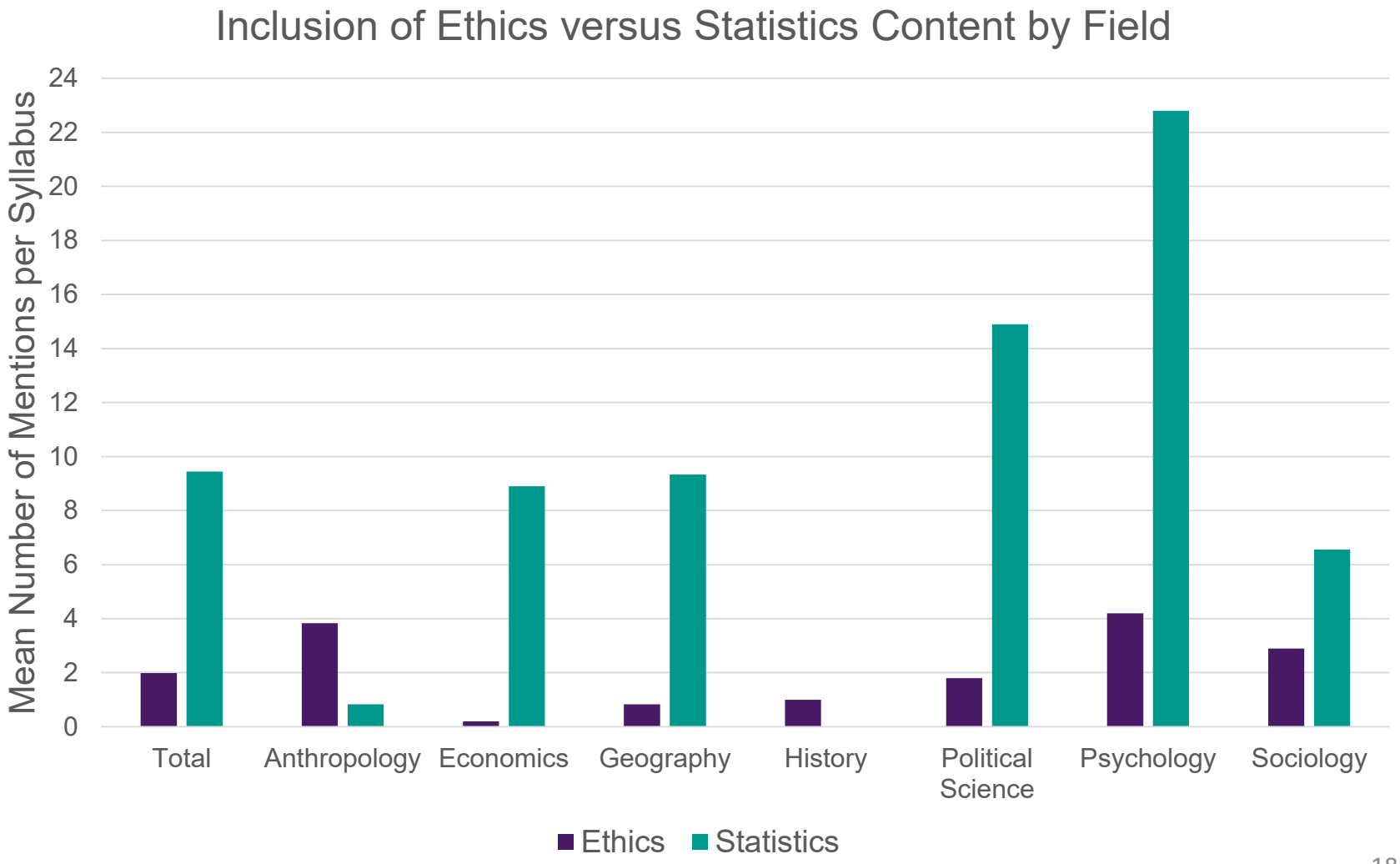
Results

- Indications of Using or Interacting with an Archive, Database, or Library

	Total Mentions	Mean	SD	% Not Mentioned
Archive Mentions	22	.44	1.64	90%
Database Mentions	43	.92	1.99	72%
Secondary Analysis	61	1.22	2.41	68%
Data Archival	34	.68	1.72	82%

- Any reference to an archive, library, or database (for any reason) in 22% of syllabi

Results


Inclusion of Ethics versus Statistics Content by Field
Mean Number of Mentions per Syllabus
24
22
20
18
16
14
12
10
8
6
4
2
0
Total
Anthropology
Economics
Geography
History
Political Science
Psychology
Sociology
Ethics
Statistics

Conclusions

- Previous survey suggested data management content would be represented in methods courses
  - Data management references appeared in limited amounts
- Training in data sharing or data repository use appears to be limited
  - Minimal references to archives or databases, even as resources
- Syllabi discuss statistics and performing analyses heavily, but how to care for data and comply with ethical requirements is often overlooked

Conclusions

- Is material regarding data management or data sharing included in the overview of coursework?

Rarely, if at all

  - Is it significant enough in the course to include?

  Even when high level of collection detail is included, what to do with data after is not mentioned

  - Would it be obvious to someone outside of the program that students received instruction on this material?

  In the majority of cases, no

Possible Actions

- Supplemental coursework
  - Self-directed online content
  - Webinars or video series
  - Workshops
  - Use of example datasets with syntax
  - "Self-curation" in multiple stats packages
- Certification in data management skills

Discussion

- Providing language for professors to use in syllabi?
  - How to practice good data management or where to seek out help
  - Ethical importance of data sharing, how to share data, and how to find data
- Room for consultations or data management user support services?

Future Questions and Project Steps

1. ~~Initial Program Survey~~
2. ~~Syllabus Analysis~~
3. Expanded Program Survey
4. Graduate Student Survey
5. Cross Discipline Comparisons