

iSchools Transform: Data Driven Trailblazers

6th National Data Service Workshop

October 20, 2016 Pittsburgh Supercomputing Center (PSC) and Carnegie Mellon University's (CMU) Software Engineering Institute, Pittsburgh, PA

Jane Greenberg, Drexel University
Virginia Ortiz-Repiso and Javier Calzada,
Universidad Carlos III de Madrid, Spain

Outline

1. Ask a few questions
2. Tell you about me (*briefly*)
3. Research
 - iSchools' data related curriculum
4. Conclusion: Why I'm excited about NDS?
5. Open discussion



©Prawny * illustrationsOf.com/216139

Questions

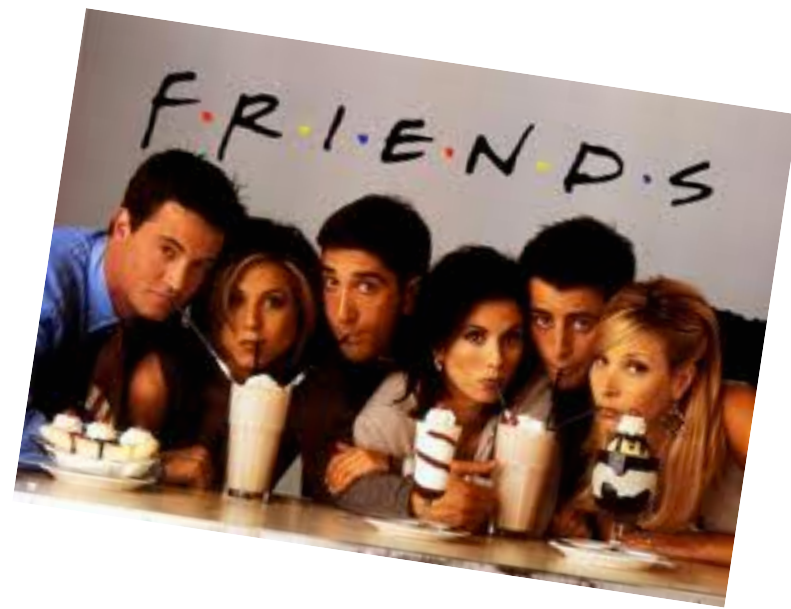
- Librarians, archivists/curators?
- Computer/information scientists?
- Disciplinary scientist - e.g., biologist, geologist?
- Data scientists?

.....

Concerned with data infrastructure design, functionality, and sustainability



Outline



1. Ask a few questions
2. Tell you about me (*briefly*)

3. Research

- iSchools' data related curricula
4. Conclusion: Why I'm excited about NDS?
 5. Open discussion



Not an
advertisement



- Founded in 2005
- International consortium of Schools - advancing the information field
 - Information science - unifying discipline
 - Library and archival science, informatics, human and social computing, business intelligence, and computational and data driven activities.
 - Share a fundamental interest in the relationships among *information, people and technology*
 - Data competent workforce needs

Jane Greenberg, Drexel University
*Virginia Ortiz-Repiso and Javier Calzada,
Universidad Carlos III de Madrid, Spain

Research goals

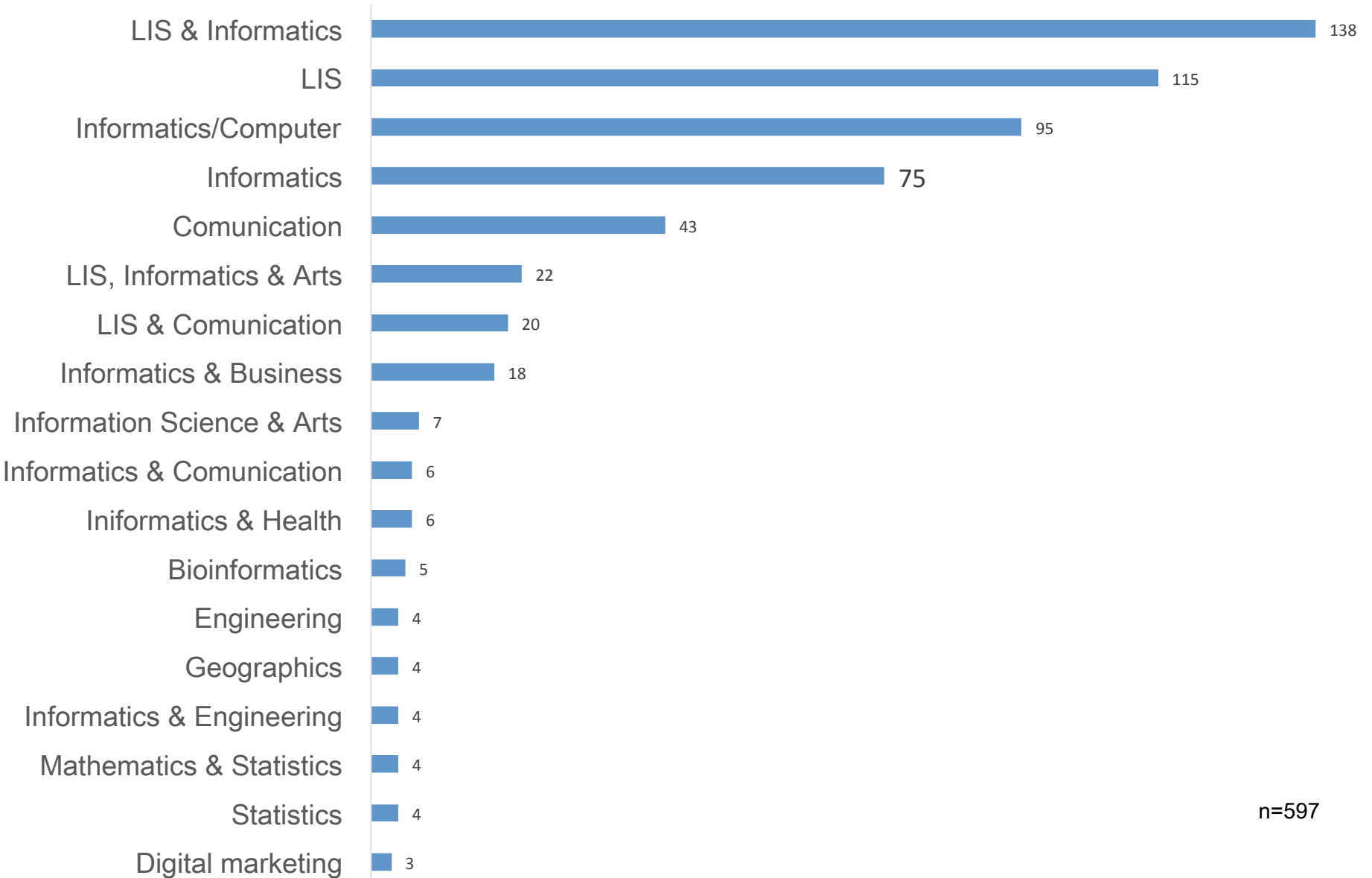
How are iSchools responding - need for for a data competent workforce?

- What is the extent of data related iSchool curricula activities?
- What data driven emphases and foci are found in iSchools?

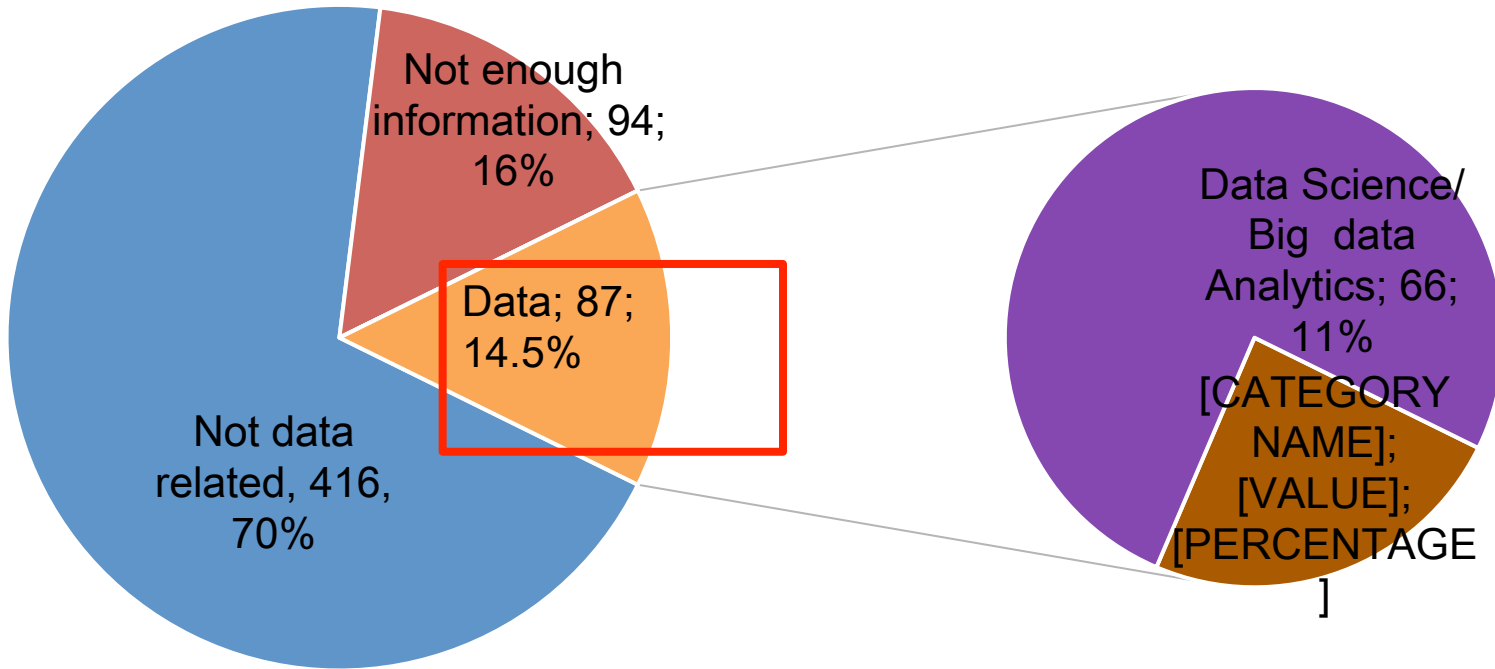
Approach

- Cross-institutional survey of iSchools, including a cluster analysis of courses offerings
- 65 iSchools (Jan. 2016)
 - Classified degree programs by country, type of degree, discipline, and concentration
 - Normalization
 - Rubric of [data science, big data analytics, and digital curation](#)

Most relevant disciplines involved in iSchools' degree programs

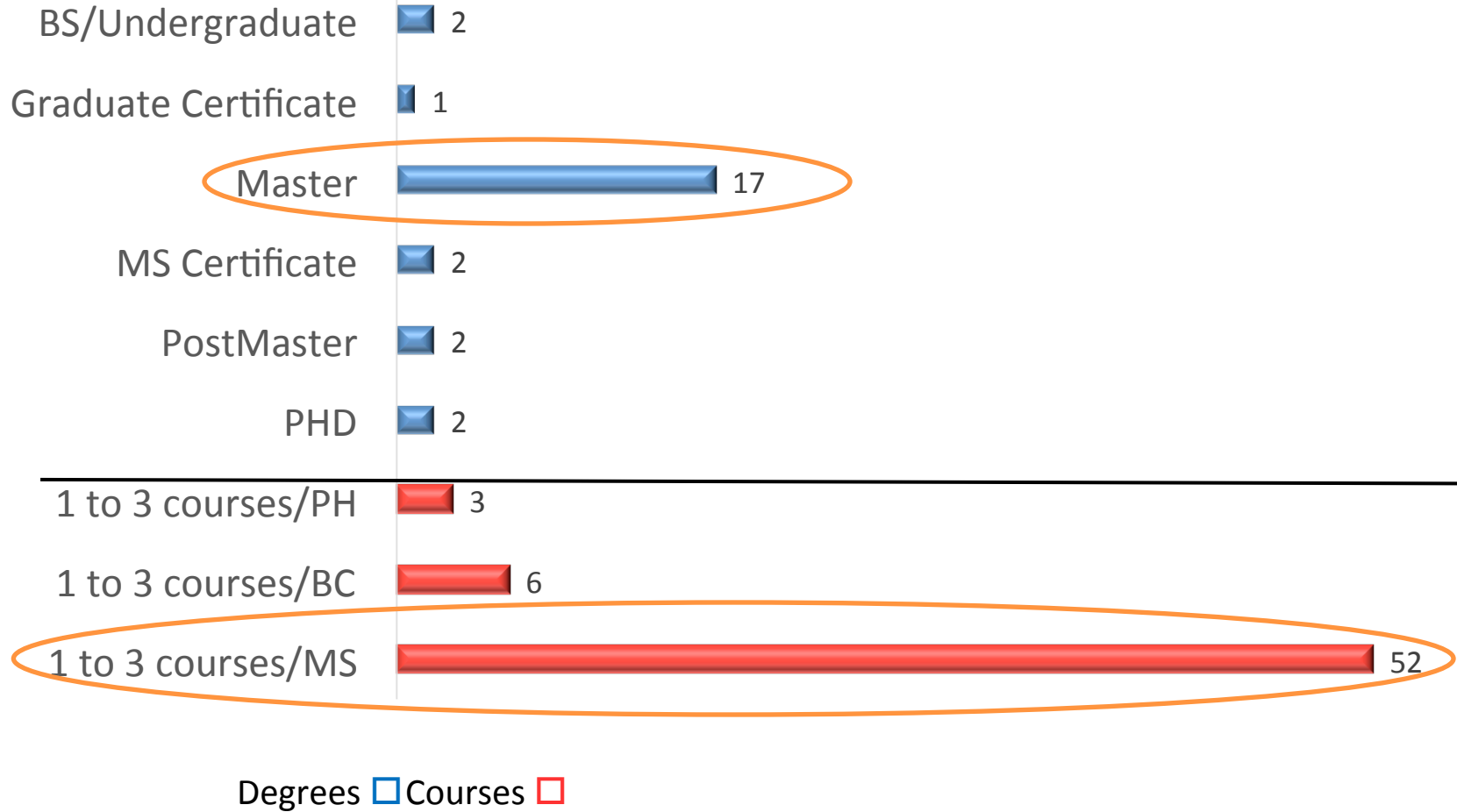


iSchools degree programs and data education

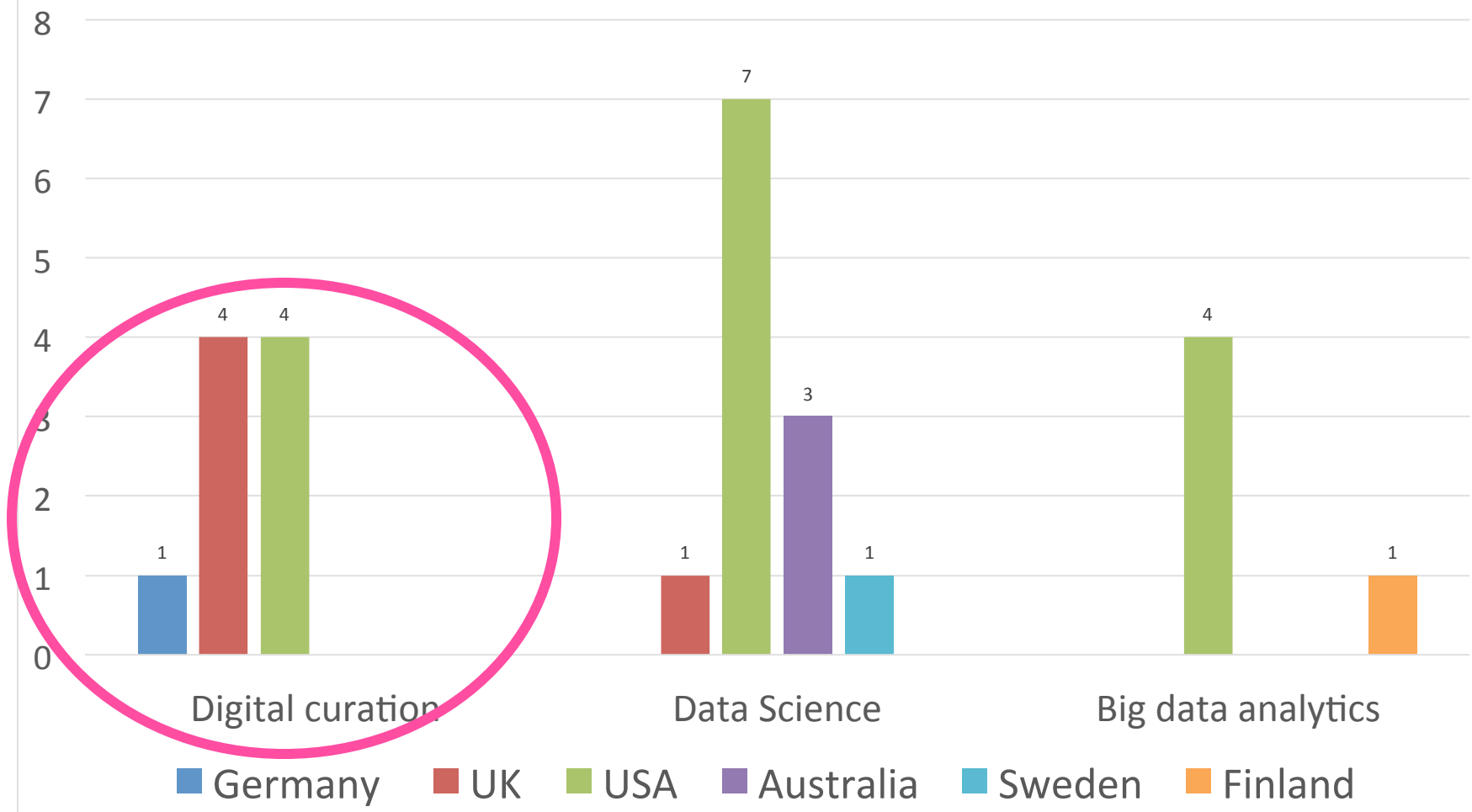


n=597

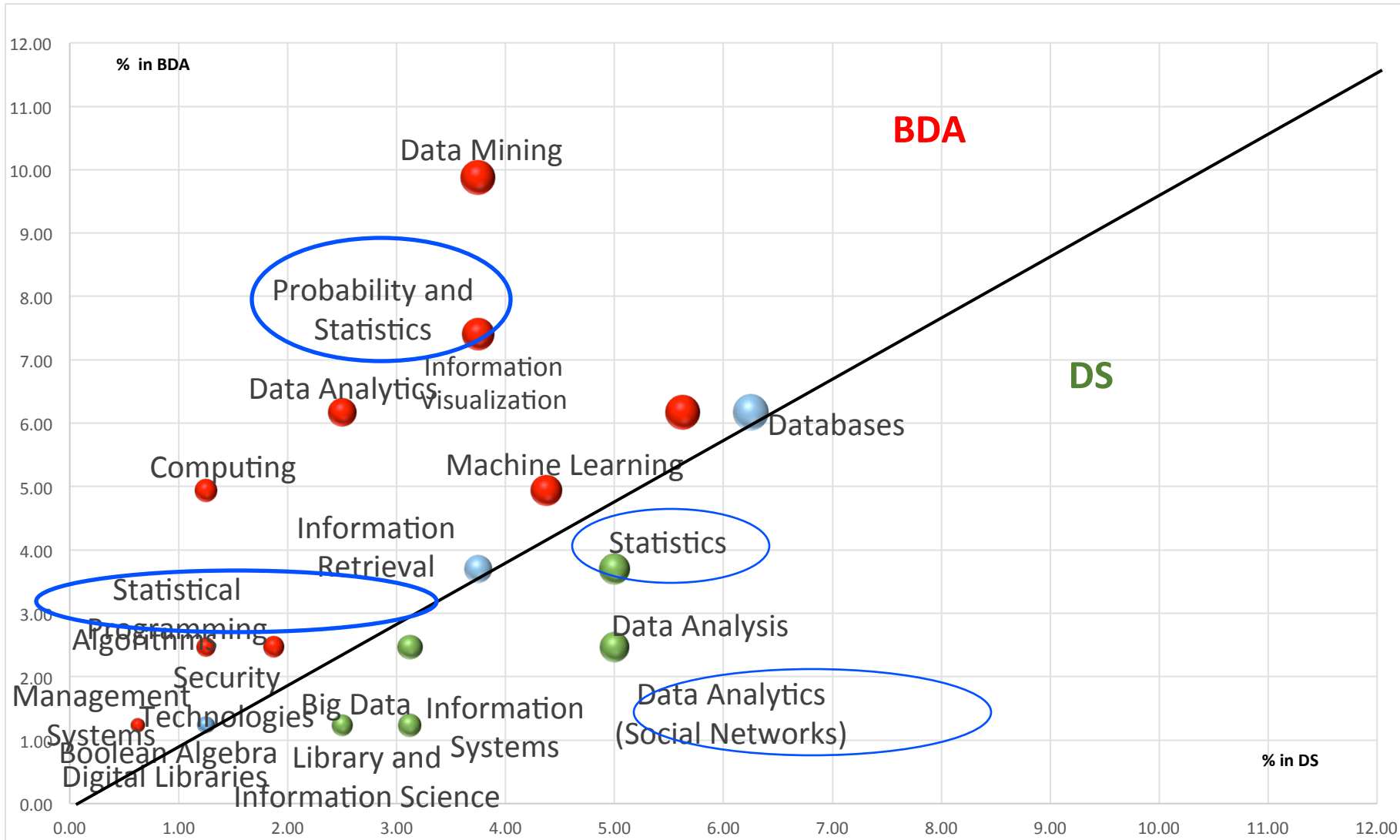
87 degrees



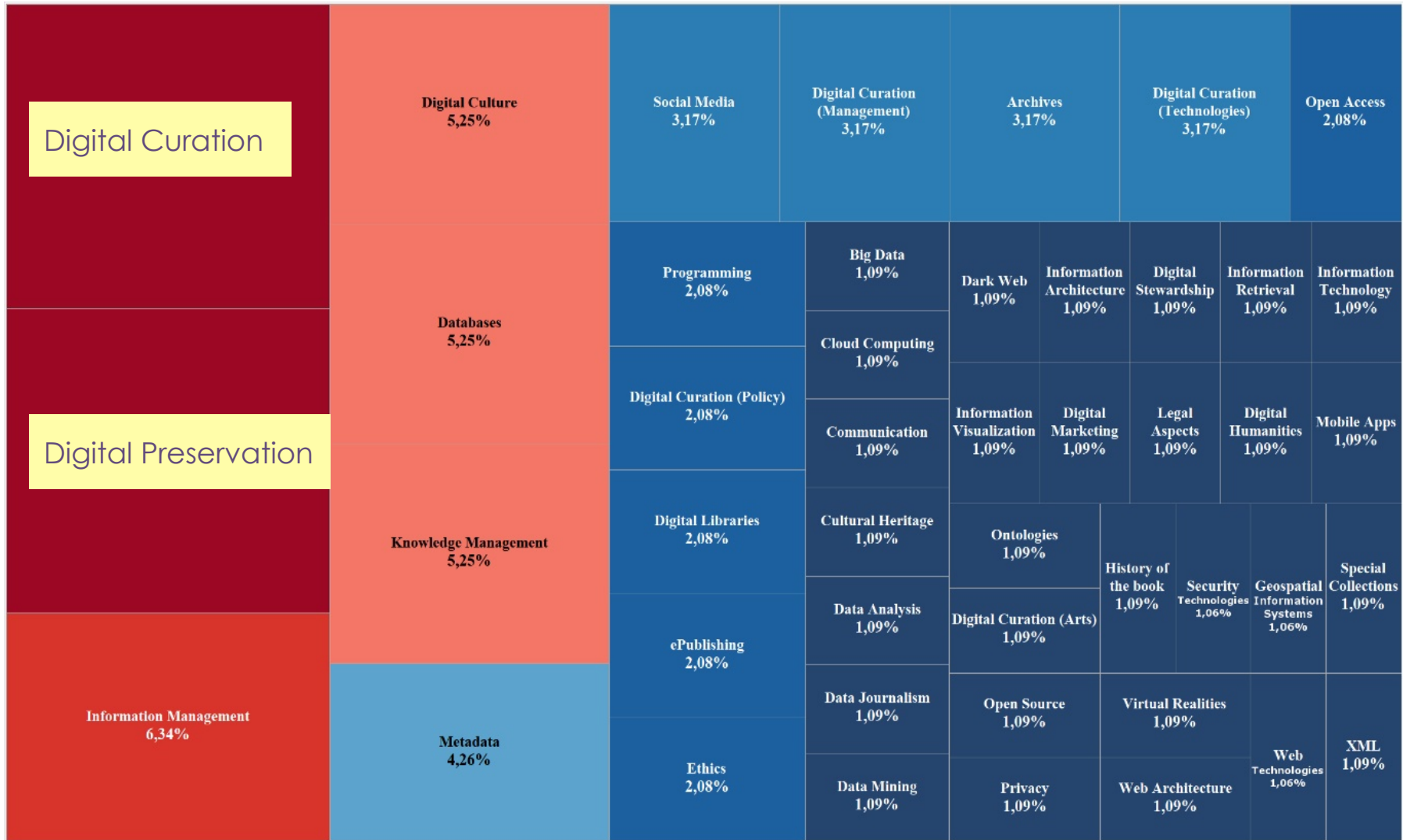
26 data degrees by country



What is being taught ?



Digital Curation, found in 21 degree programs, of the 87 that have data-related curricula



A few more results

- Data-driven education in iSchools can be classified data science (DS), big data analytics (BDA), and digital curation.
- Most data science and big data analytics degrees are in US iSchools
- Digital curation fairly equal in Europe + U.S.
- Data degrees fairly recent, starting 2013-2016
- Lack of clear distinction between DS and BDA, but DS has larger spectrum of diversity

A few more results

- Digital curation degrees smaller percentage of iSchools compared to DS and DA.
- ✚ New job titles - “data research scientists”, “data services librarian”, and “research data and digital curation officer”, suggests iSchools give more attention digital curation.
- ✚ Limited interdisciplinary evidence;
- Limitations: web, inconsistent info., 77 iSchools, etc., but...

1. Ask a few questions
2. Tell you about me (*briefly*)
3. Research
 - iSchools' data related curricula



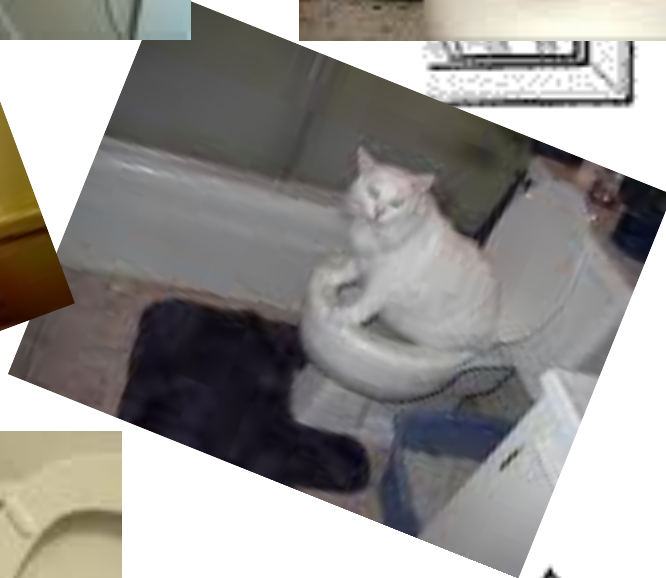
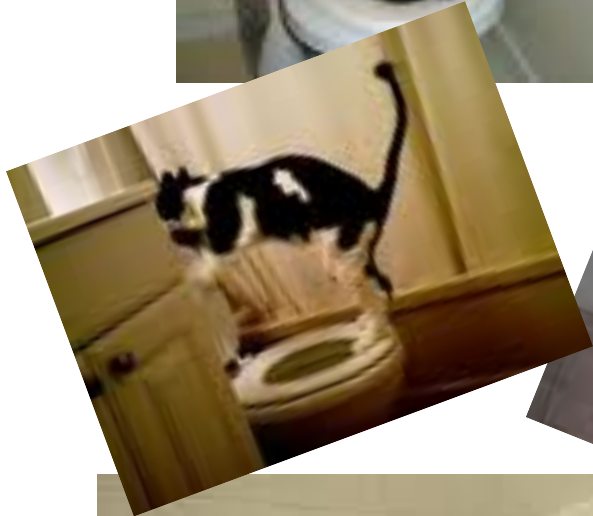
4. Conclusion: Why I'm excited about NDS?

- Hard problems, continuum of expertise
- Library science part of the larger information and data ecosystem
 - Cross fertilization (curation, DS, etc.)
 - 2013+ trailblazers
 - Stone age – to Bronze age

ALL RIGHTS RESERVED
<http://www.cartoonbank.com>



"Never, ever, think outside the box."



RESERVED
onbank.com



"Never, ever, think outside the box."

Optional slide...



1



2



3



4



Librarian



MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants




PROGRAMS & DATABASES

- ☆ Computer science
- ☆ Scripting languages
- ☆ Statistical computing
- ☆ Databases: SQL
- ☆ Relational algebra
- ☆ Parallel database processing
- ☆ MapReduce computing
- ☆ Hadoop and HBase
- ☆ Custom reducers
- ☆ Experience with

Data Scientists

DATA SCIENTIST

- Math
- Statistics
- Programming
- Database
- Domain Knowledge
- Soft Skills
- Communication
- Visualization




Historian/
Data Scientists