SAGE ocean

Software and tools for social science

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne | Toronto

Goal of this webinar

- Share with you our findings
- Get feedback and questions

Quick overview:

- Why we looked at 400+ tools
- Summary findings
- Future of tools
- Where next

2016
A SAGE White Paper
Who Is Doing Computational Social Science?
Trends in Big Data Research
Katie Metzler
Publisher for SAGE Research Methods, SAGE Publishing
David A. Kim
Stanford University, Department of Emergency Medicine
Nick Allum
Professor of Sociology and Research Methodology, University of Essex
Angella Denman
University of Essex
www.sagepublishing.com
SAGE Publishing

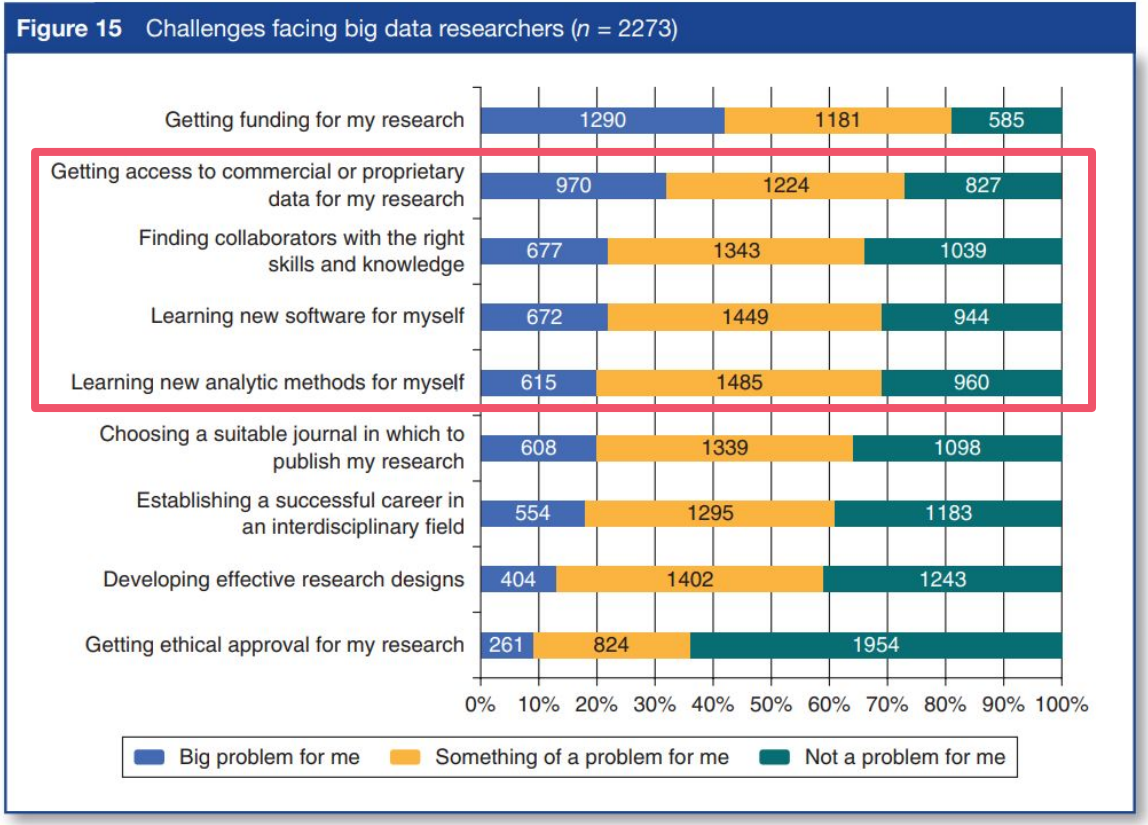
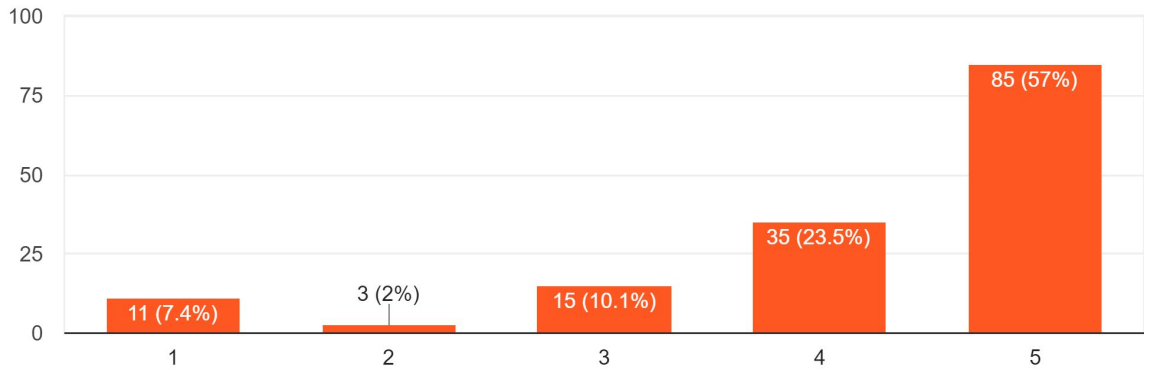


Figure 15 Challenges facing big data researchers (n = 2273)
Getting funding for my research 1290 1181 585
Getting access to commercial or proprietary data for my research 970 1224 827
Finding collaborators with the right skills and knowledge 677 1343 1039
Learning new software for myself 672 1449 944
Learning new analytic methods for myself 615 1485 960
Choosing a suitable journal in which to publish my research 608 1339 1098
Establishing a successful career in an interdisciplinary field 554 1295 1183
Developing effective research designs 404 1402 1243
Getting ethical approval for my research 261 824 1954
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
Big problem for me
Something of a problem for me
Not a problem for me

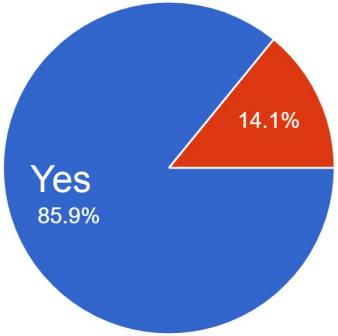
Software in social science

How important is research software to your work?

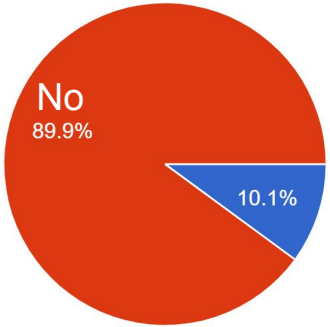
149 responses



Do you use research software?



Do you develop your own research software?

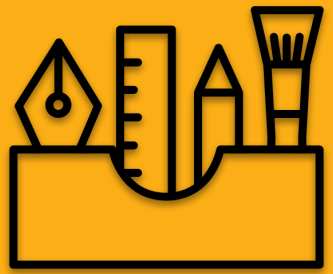


149 responses, August 2019, thank you SSI

SAGE ocean

Enabling social scientists to work with BIG DATA & TECHNOLOGY

- Find and promote new tech and methods
- Concept grants: DDNA, Quanteda, miniVAN, Text Wash
- Investments: Tag.Works



How?

- Constantly reviewed alternatives to existing tools
- Conference websites and archives
- Prominent labs
- Other tool directories and lists
- Academic papers referencing or describing the tools

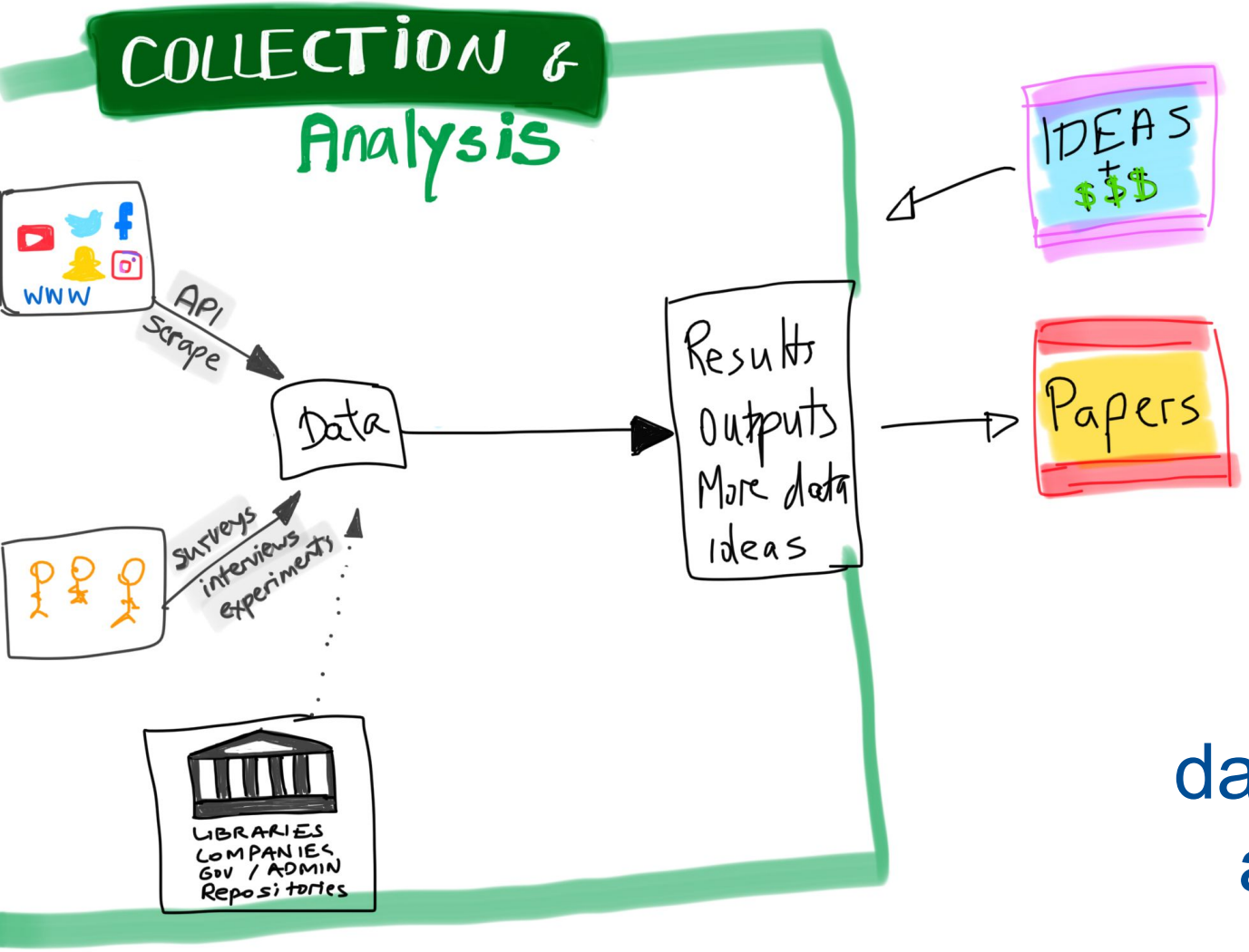
- Wikipedia
- Crunchbase

Tools, footnote

Will refer to a tool, technology, software service or package interchangeably.

Our list of 418 includes:

cloud and locally installed apps, software applications, packages and code, licensed and open source, essentially any piece of code, whether neatly packaged within an interface or delivered raw on GitHub and used by academic researchers

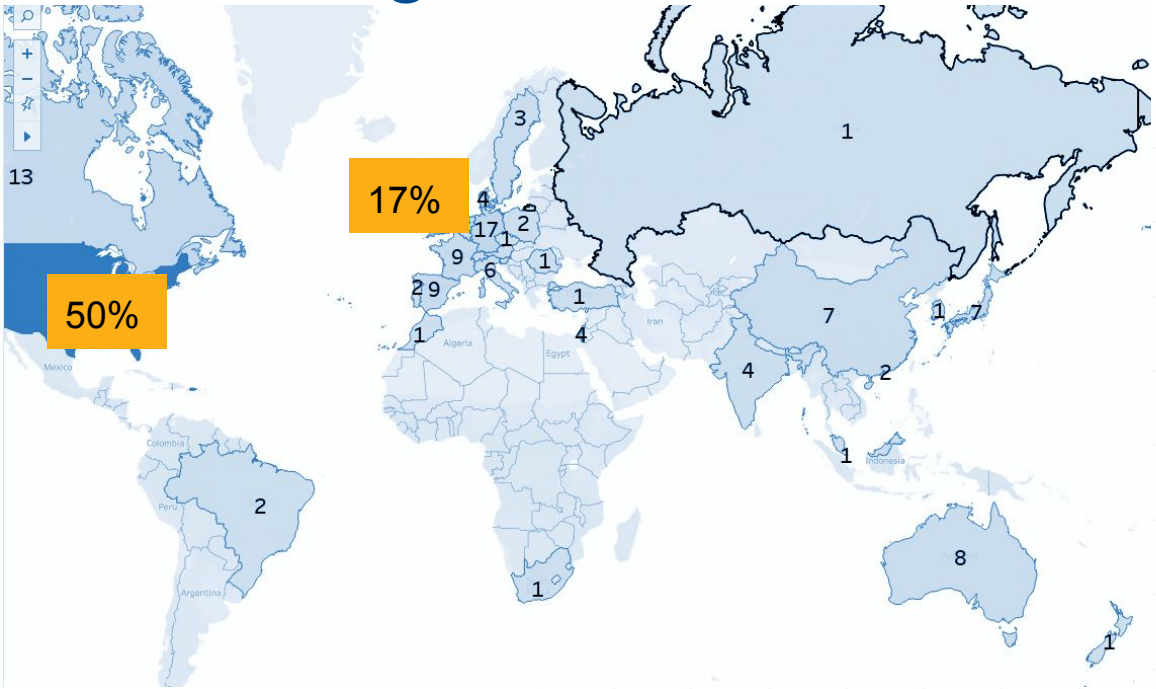


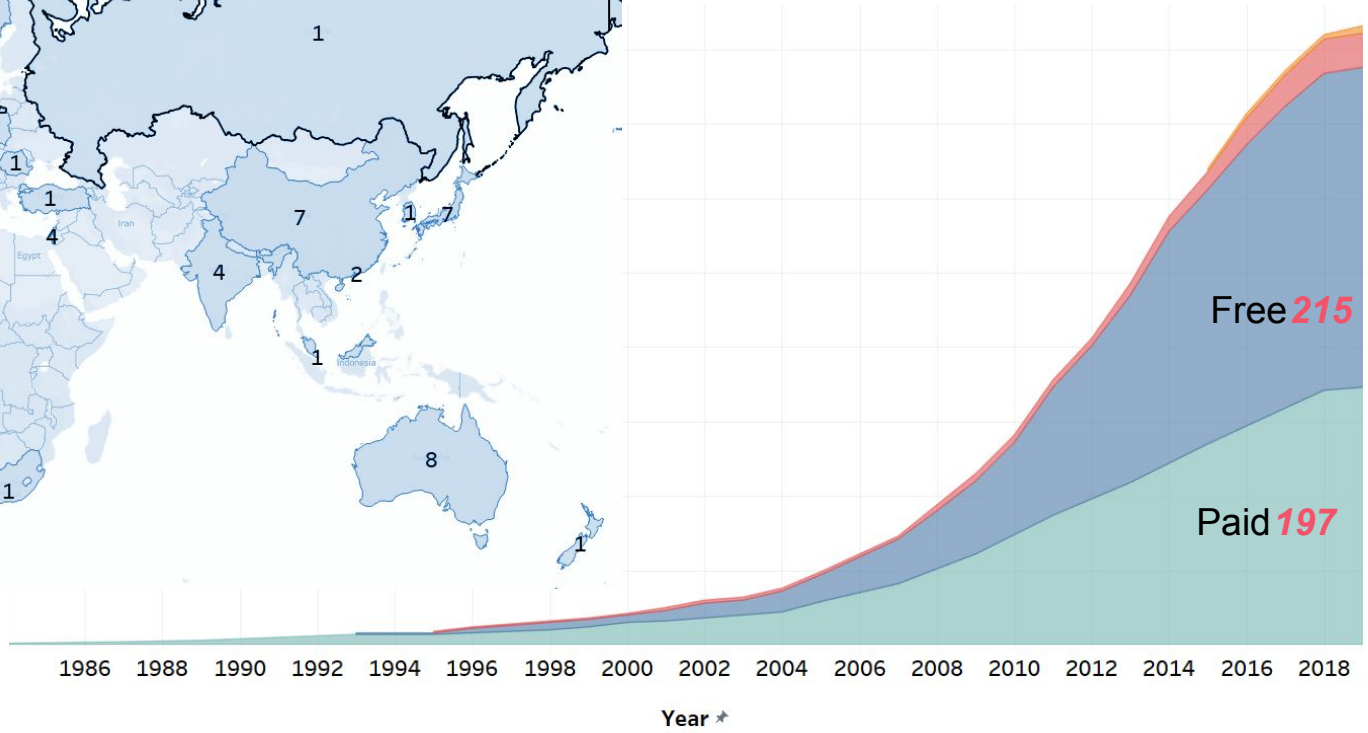
Tools focusing on data collection and analysis

Summary findings

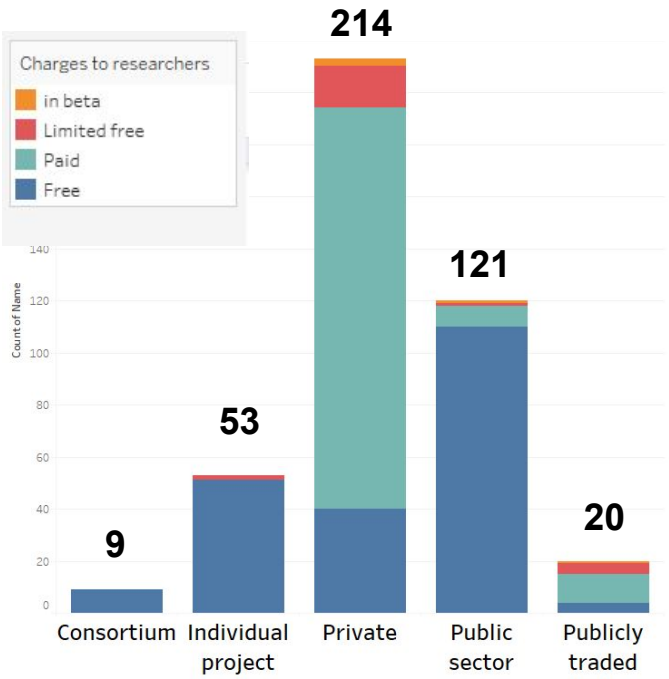
SAGE Publishing

Growing number of tools





Who’s behind the tools?



socialmedia RESEARCH FOUNDATION

Pelagios network

NUMFOCUS OPEN CODE = BETTER SCIENCE

Software Sustainability Institute

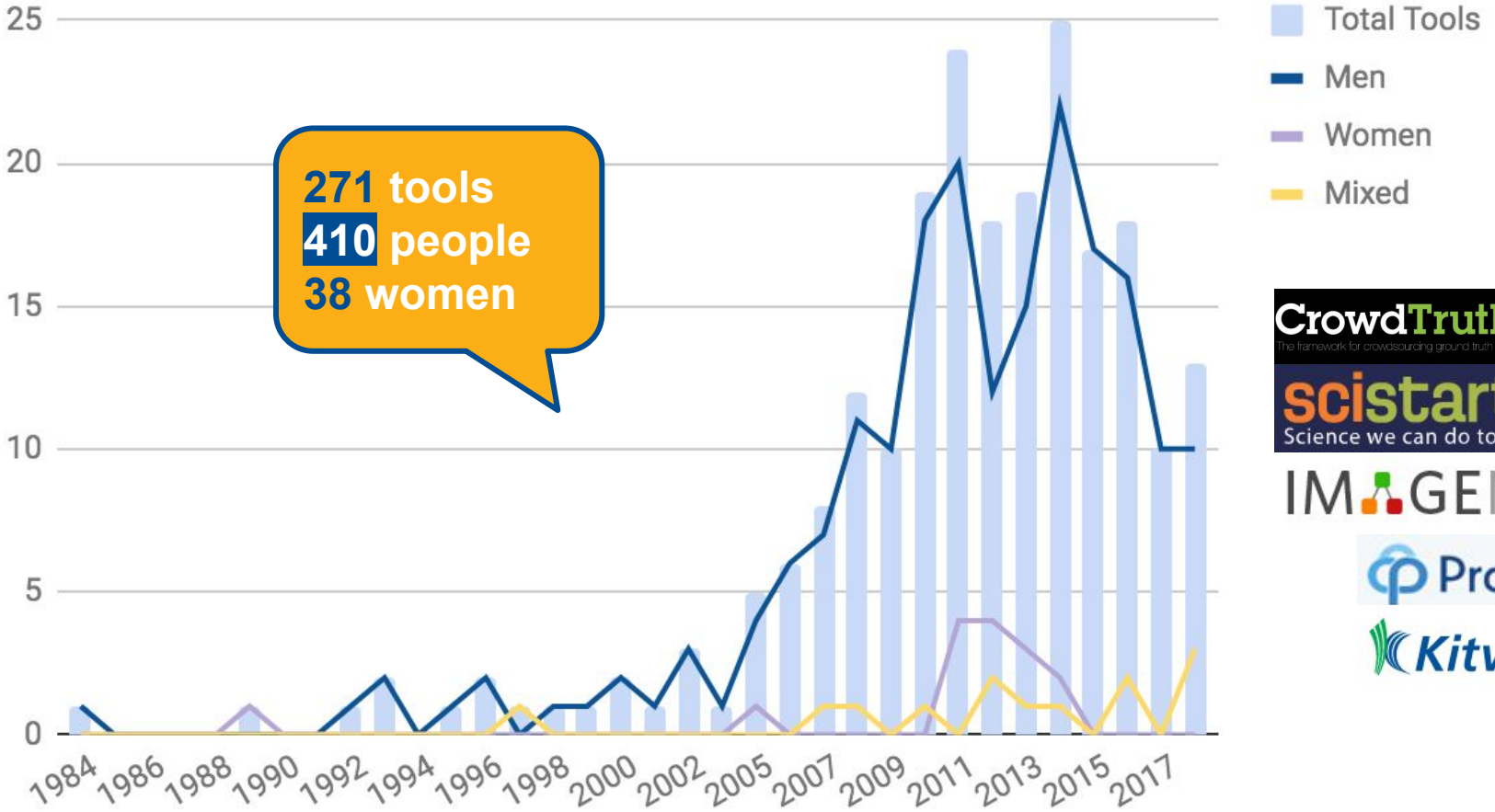
Social Mining & Big Data Ecosystem SoBigData RESEARCH INFRASTRUCTURE

cessda DARIAH-EU

COS CENTER FOR OPEN SCIENCE

nectar

Tool Leadership Gender, 1984-2019



CrowdTruth
The framework for crowdsourcing ground truth data
scistarter®
Science we can do together.
IMAGENET
Prolific
Kitware

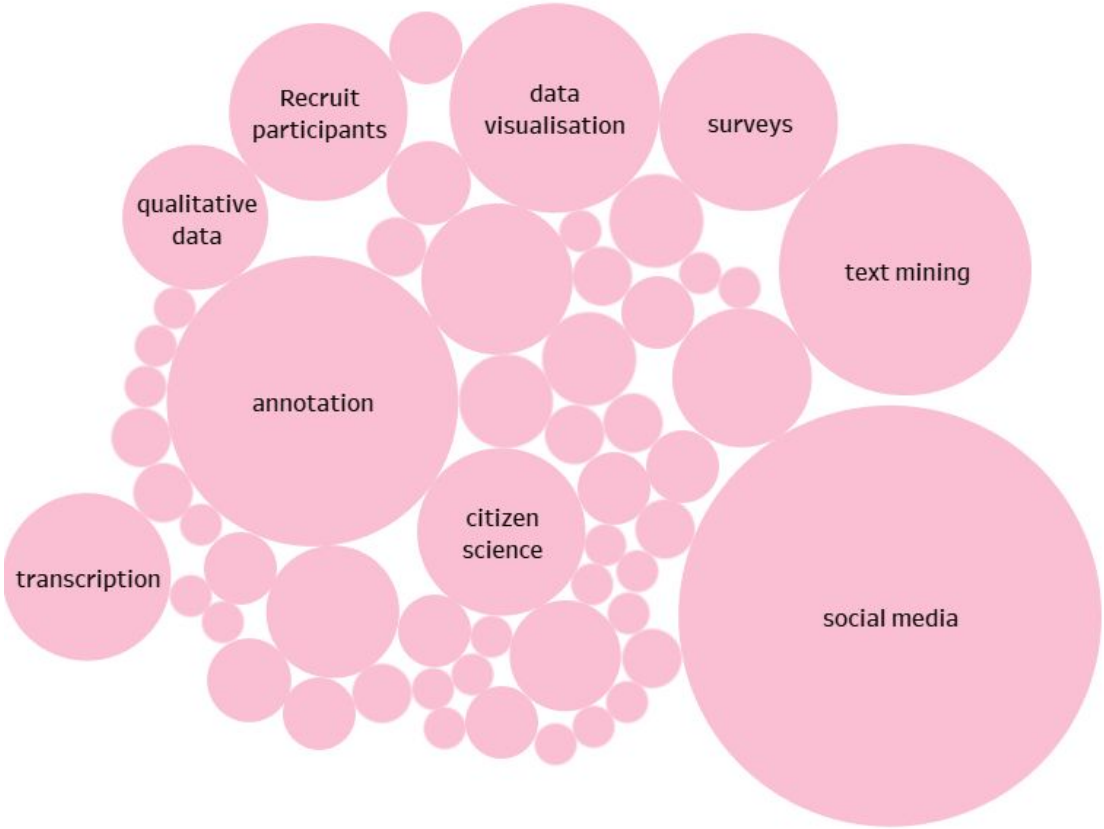
Insights: successful tools

Open Source/Free

- Team or individual dedicated fully to raising funds up to $6M
- When applicable beyond academia backed by a growing and large community

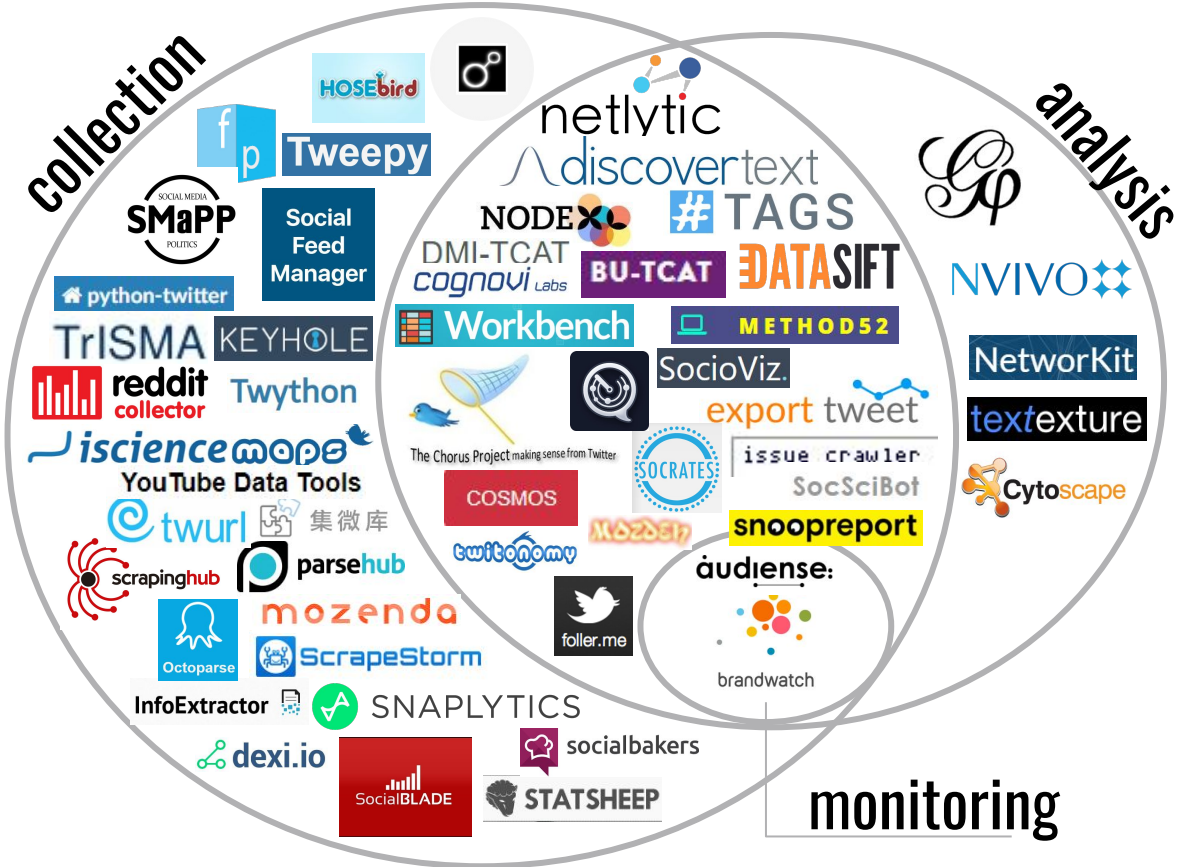
Commercial

- Applicability beyond academia
- VC-backing up to $400M
- Big brand clients
- Founding team with business background

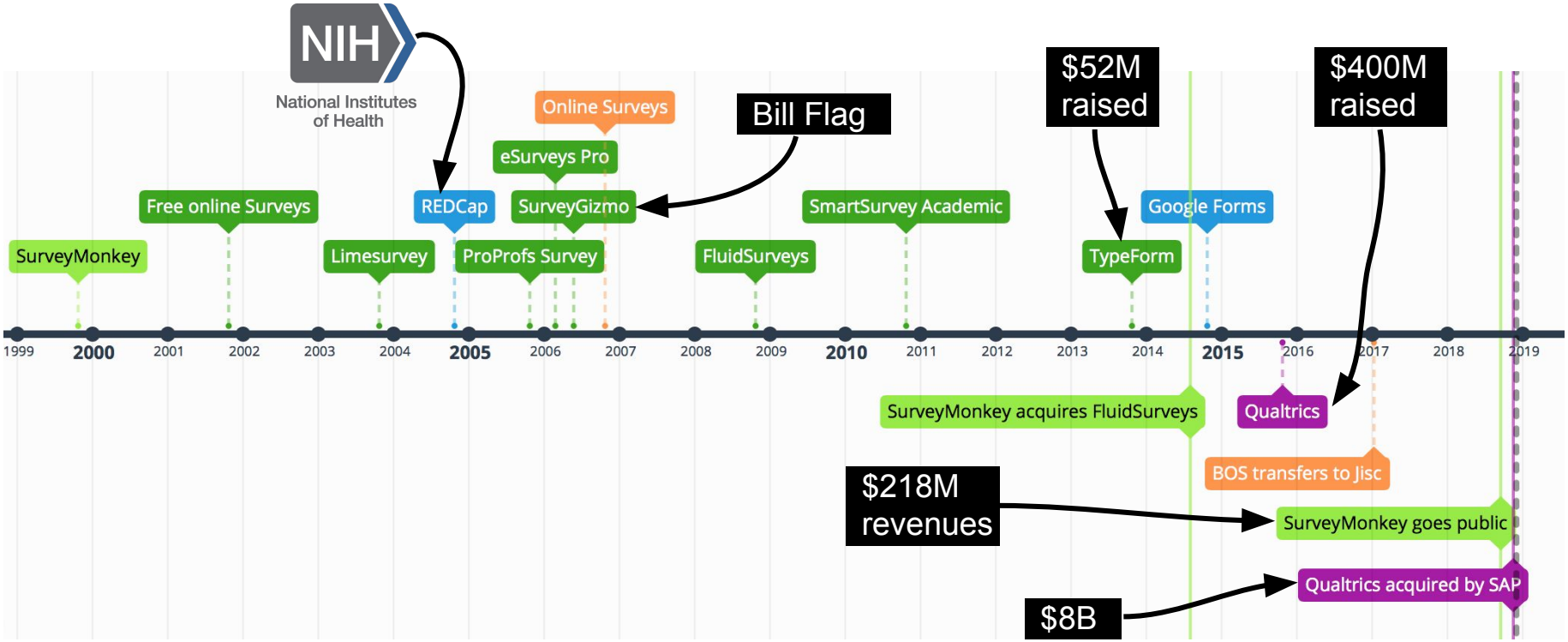

Recruit participants
data visualisation
surveys
qualitative data
text mining
annotation
citizen science
transcription
social media

Social media analysis tools

- A lot of tools, esp for twitter
- Somewhat hard to automate text analysis
- Very little on image analysis



Survey tools



Beyond surveys:

- Recruit participants
- Design experiments

Recruit and/or manage participants

classic surveys

qualtrics

Prolific

REDCap
Research Electronic Data Capture

SONA SYSTEMS

amazon mechanical turk

VOLUNTEER SCIENCE

DAEMO

www.gorilla.sc

INQUISIT by millisecond

E-Prime 3
The world leader in behavioral research software

design experiments

bos

Google FORMS

Typeform

SurveyMonkey

surveygizmo

SmartSurvey

LimeSurvey

EXPERIMENT FACTORY

tatool

PEBL

DIY recruitment

Annotating text

Large scale and for ML

Basic labelling

amazon mechanical turk

Amazon SageMaker

SEQUENCE

tagtog

Light TAG

DAEMO

Recogito

doccano

PRP VBP RB Web Anno

TagWorks

discovertext

GATE

A.nnotate

Anafora

dedoose Great Research Made Easy

NVIVO

atlas.ti

CATMA COMPUTER AIDED TEXTUAL MARKUP AND ANALYSIS

brat

Complex annotation, i.e. hierarchical with code books

Small scale

Automating text analysis

70+

Maximum performance for large corpora: High

Low

AYLIEN rapidminer

MALLET

other python and R packages/libraries

VOYANT see through your text

atlas.ti

GATE

brat

dedoose Great Research Made Easy

NVIVO

orange

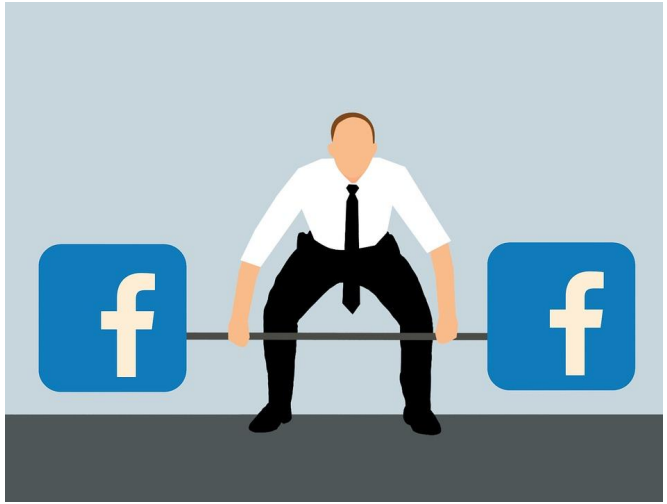
quanteda Quantitative Analysis of Textual Data

LEXIMANCER

Skill level needed: High → Low

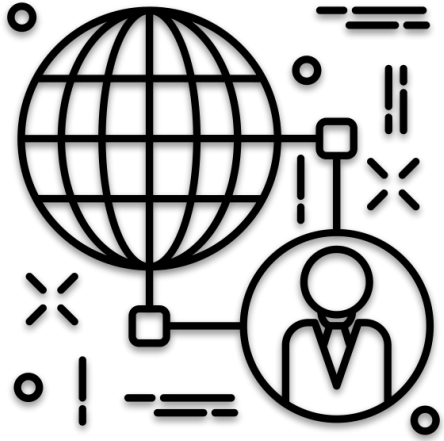
Challenges for researchers

- How do I find the right tool?
- Is it good enough?
- How to cite it?
- When I develop, how do I get credit?
- Sustainability & open source
- How to build when data is with big tech



400+ tools...

- 50% from US
- 50% private/commercial
- 38 women involved in 271 tools
- 9 tools supported by consortia
- $6M in grants
- $400M in VC funding
- Survey and social media - saturated market
- Automating text analysis
- Sharing and citation



The future of tools

- What are we going to be obsessed with?
- Where will the data about us sit?
- Where will the most promising researchers be?

More about tools...

1. SAGE Ocean [Blogs on tools](#):
   - Matchmaking
   - Transcription
   - Social media, Weibo, Linkedin
   - Running experiments
   - Search data
   - Survey on who uses/develops software

2. Emerging tech spotlight in Big Data Newsletter
3. [SAGE Ocean tools directory](#)
4. White paper

The SAGE Ocean big data newsletter

YOUR MONTHLY BIG DATA & SOCIAL RESEARCH ROUNDUP

JUNE 2019

Featuring details of the next SAGE Ocean Speaker Series: Digital support for more creative science journalism with Neil Maiden.

The ecosystem of technologies for social science research

A SAGE White Paper

View on GitHub

To download the whitepaper as a pdf, please follow this link.

The underlying **data** is available as csv in this repository's data folder

Both the whitepaper and the underlying data are open under CC-BY and CC0 license respectively. Please feel free to use and reuse, and contribute. Watch this space for an updated version of the white paper and the data.

https://bit.ly/35RaAzG

What next

- Webinar: automating text analysis
- GitHub page next iteration: 500 tools
- Wikidata

- Concept grants: £2k & £15k, Feb ‘20
- Ideas, suggestions, more data, developing software - get in touch!

Thank you

Poll