



Losing sleep in a data-driven dream

Jevin West

Information School, University of Washington

Data Science Summit, Suncadia Resort, Cle Elem, WA

Oct. 5, 2015





TRUST ME BRO, I'M A
DATA SCIENTIST

Karl Ove Hovind, Ph.D. 2013

TRUST ME BRO, I'M A
DATA SCIENTIST

Karl Ove Hovind, Ph.D. 2013

mes

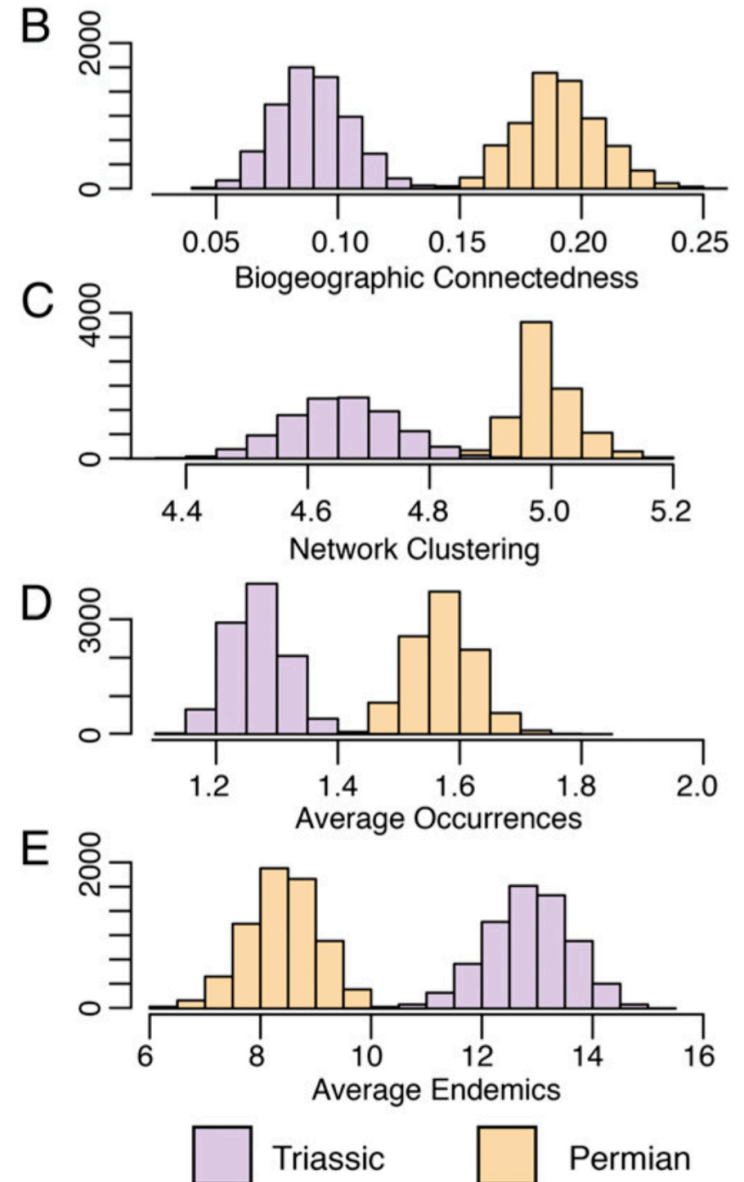
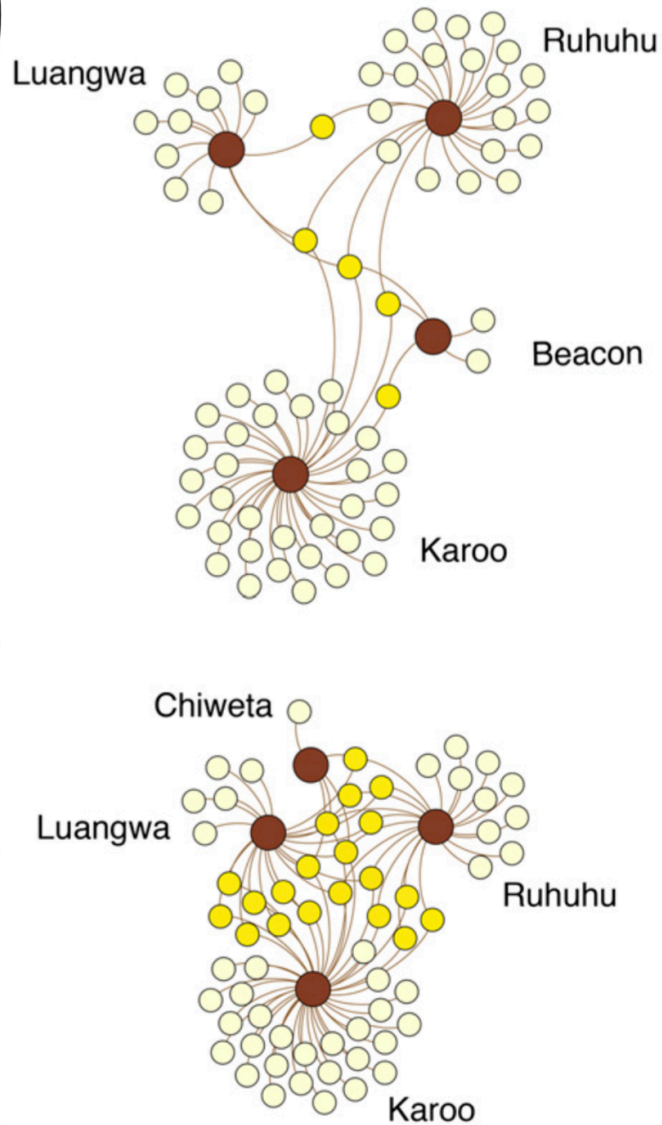
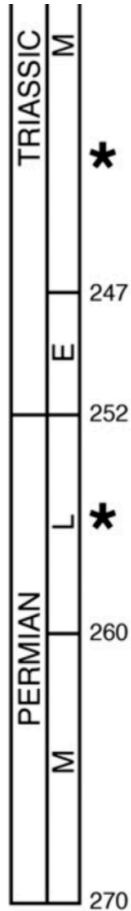
ting site

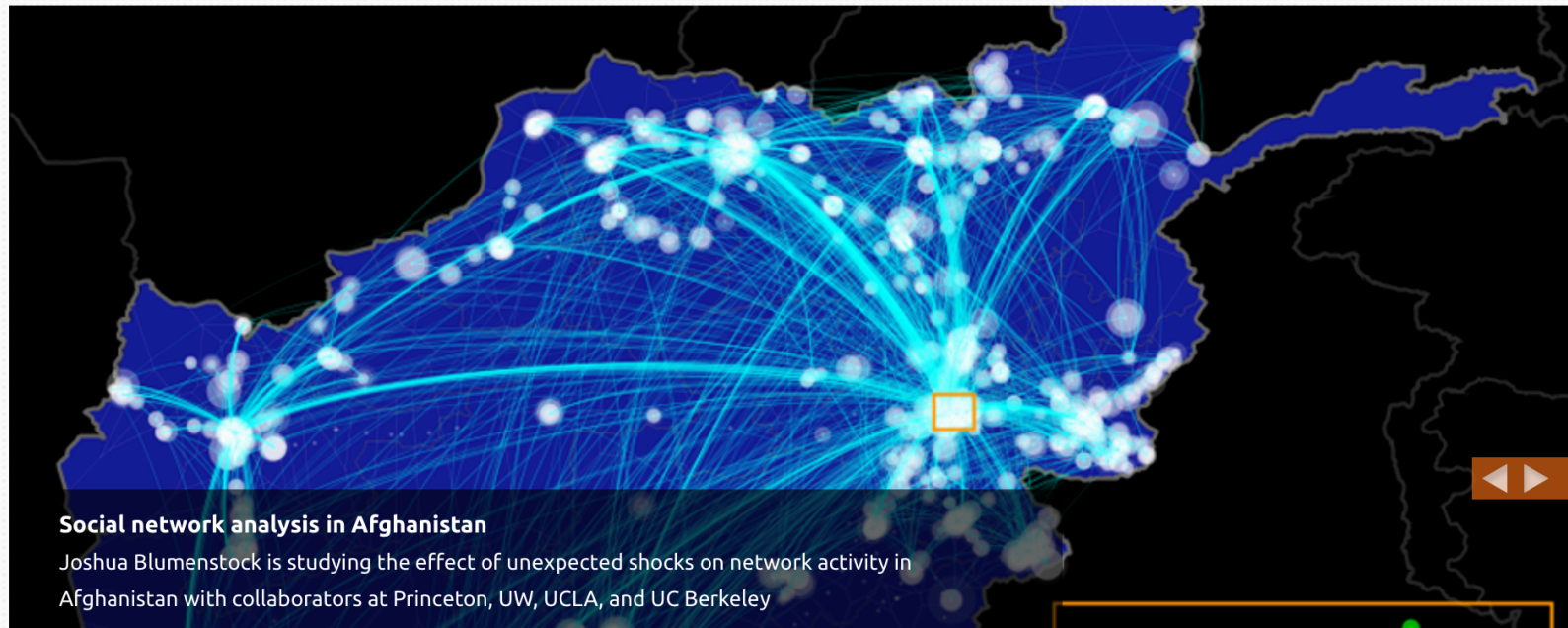
users,

h threads

online?

Data Science and Mass Extinction





Social network analysis in Afghanistan

Joshua Blumenstock is studying the effect of unexpected shocks on network activity in Afghanistan with collaborators at Princeton, UW, UCLA, and UC Berkeley

Research Focus Areas



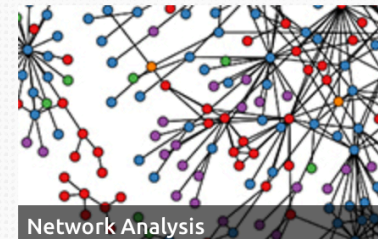
Business Analytics



Social Behavior



Poverty and social change



Network Analysis

News and Updates

28

Blumenstock at Population Association of America

What we do

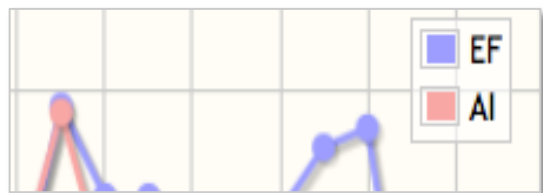
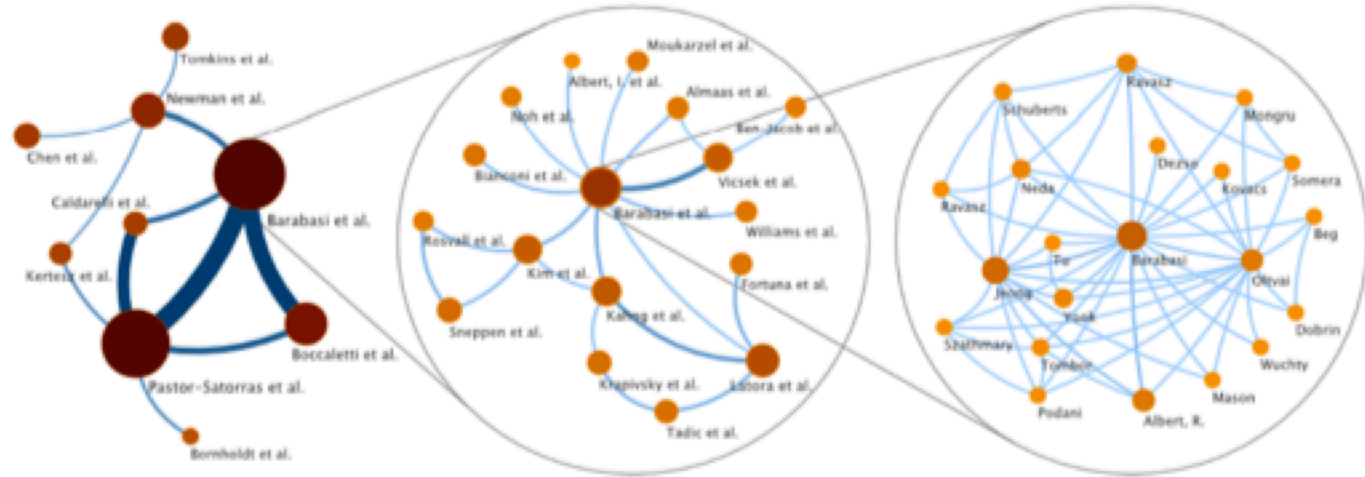
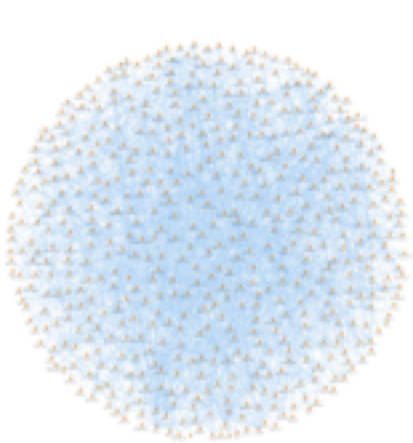
The DataLab is the nexus for research on Data Science and Analytics at the UW iSchool. We study **large-scale, heterogeneous human data** in an

Jevin West

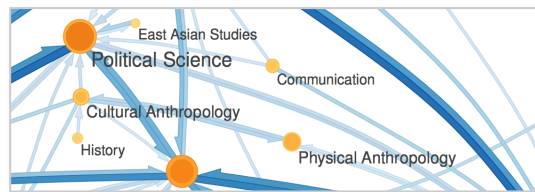
Assistant Professor | iSchool



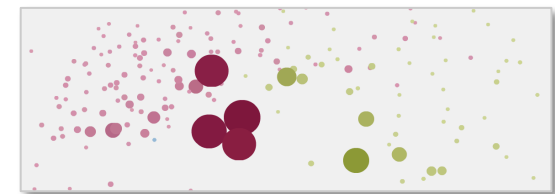
eigenFACTOR.org
RANKING AND MAPPING SCIENTIFIC KNOWLEDGE



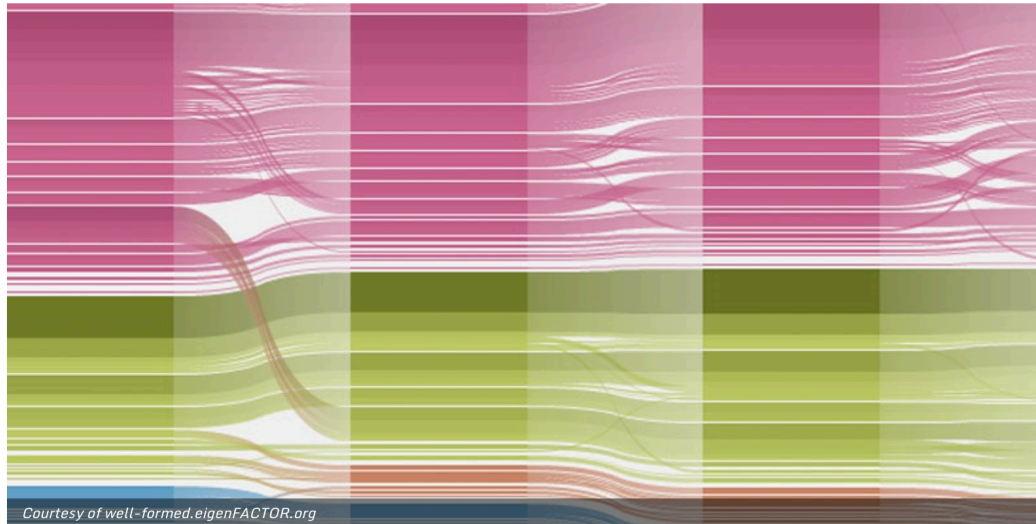
Ranking



Mapping



Navigating



DATA-DRIVEN DISCOVERY

Data Science
Environments



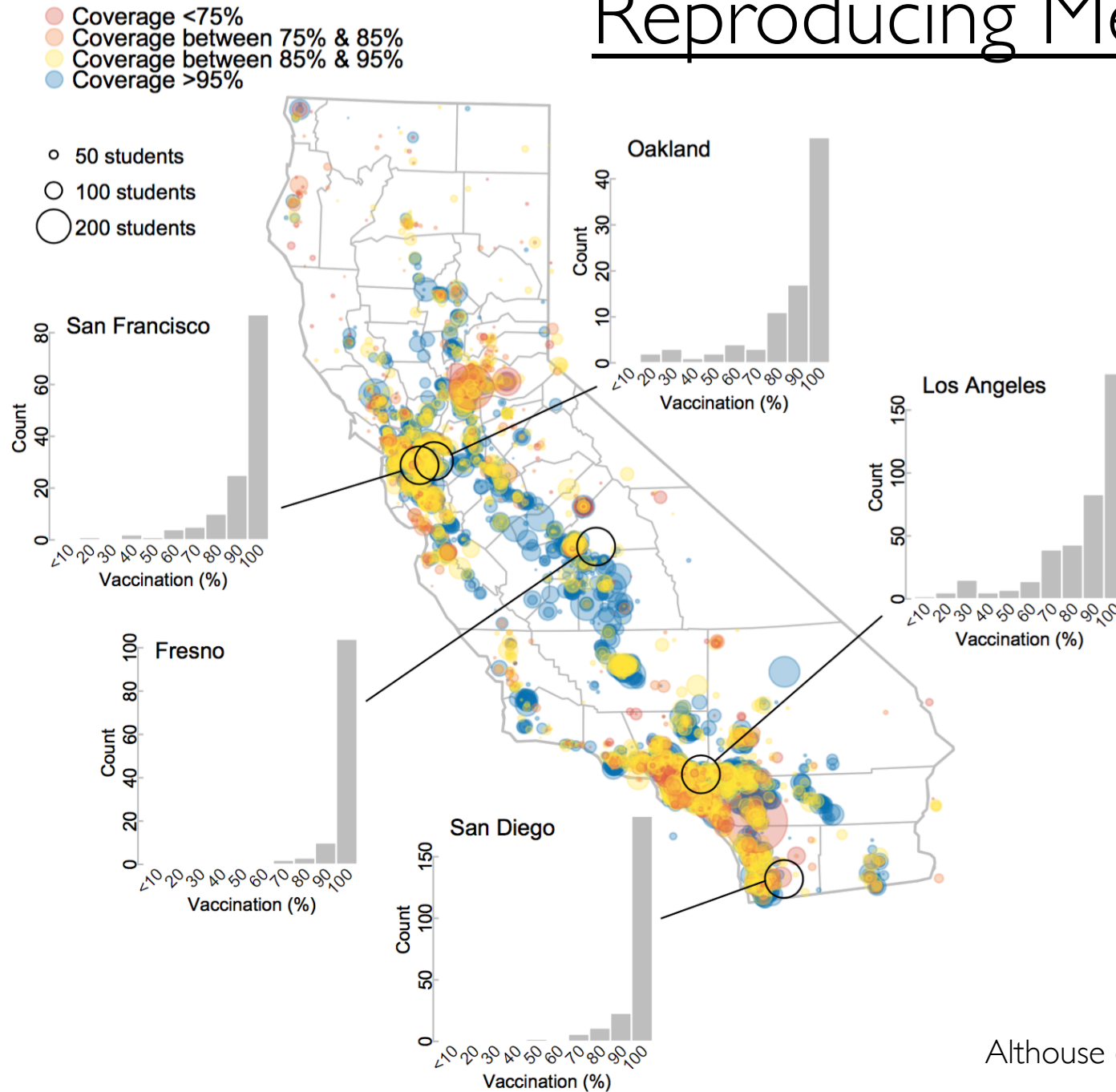
ALFRED P. SLOAN FOUNDATION

Losing Sleep



Reproducibility

Reproducing Measles





Tenure



DOI:10.1145/2753507

Moshe Y. Vardi

Incentivizing Quality and Impact in Computing Research

Over the past few years, the computing-research community has been conducting a public conversation on its publication culture. Much of that conversation has taken place

in the pages of *Communications*. (See <http://cra.org/scholarlypub/>.) The underlying issue is that while computing research has been widely successful in developing fundamental results and insights, having a deep impact on life and society, and influencing almost all scholarly fields, its publication culture has developed certain anomalies that are not conducive to the future success of the field. A major anomaly is the reliance of the fields on conferences as the chief vehicle for scholarly publications.

While the discussion of the computing-research publication culture has led

be a game changer. By advising research organizations to focus on quality and impact, the memo aims at changing the incentive system and, consequently, at changing behavior.

The key observation underlying the memo is that we have slid down the slippery path of using quantity as a proxy for quality. When I completed my doctorate (a long time ago) I was able to list four publications on my CV. Today, it is not uncommon to see fresh Ph.D.'s with 20 and even 30 publications. In the 1980s, serving on a single program committee per year was a respectable sign of profes-

careful scholarship. Indeed, academic folklore has invented the term LPU, for "least publishable unit," suggesting that optimizing one's bibliography for quantity rather than for quality has become common practice.

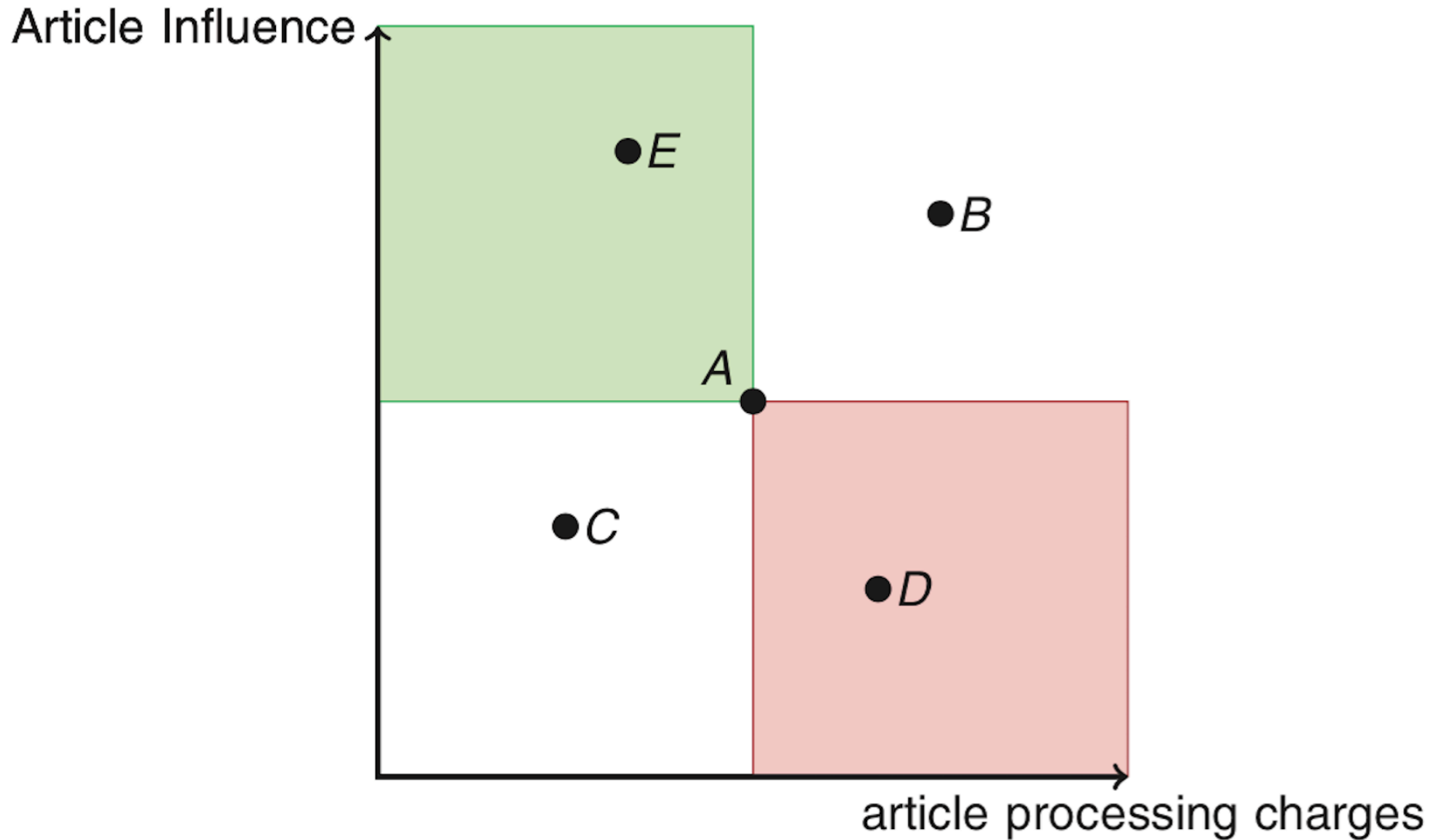
To cut the Gordian knot of mutually reinforcing norms and expectations, the memo advises hiring units to focus on quality and impact and pay little attention to numbers. For junior researchers, hiring decisions should be based not on their number of publications, but on the quality of their top one or two publications. For tenure candidates, decisions should be based on the quality and impact of their top three-to-five publications.

Focusing on quality rather than quantity should apply to other areas as well. We should not be impressed by large research grants, but ask what the actual



Open Access

Cost Effectiveness





Data Scientists

[illegible]



Data Ethics

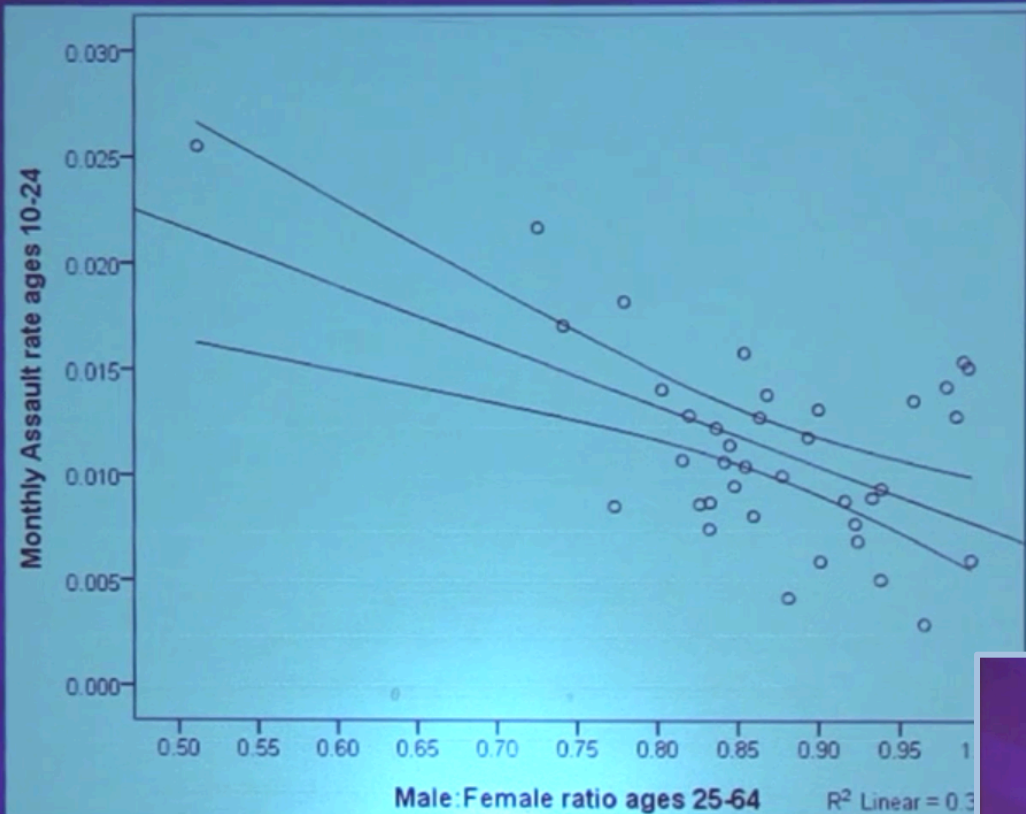




Calling Bullshit

Sex Ratios and Crime

Results

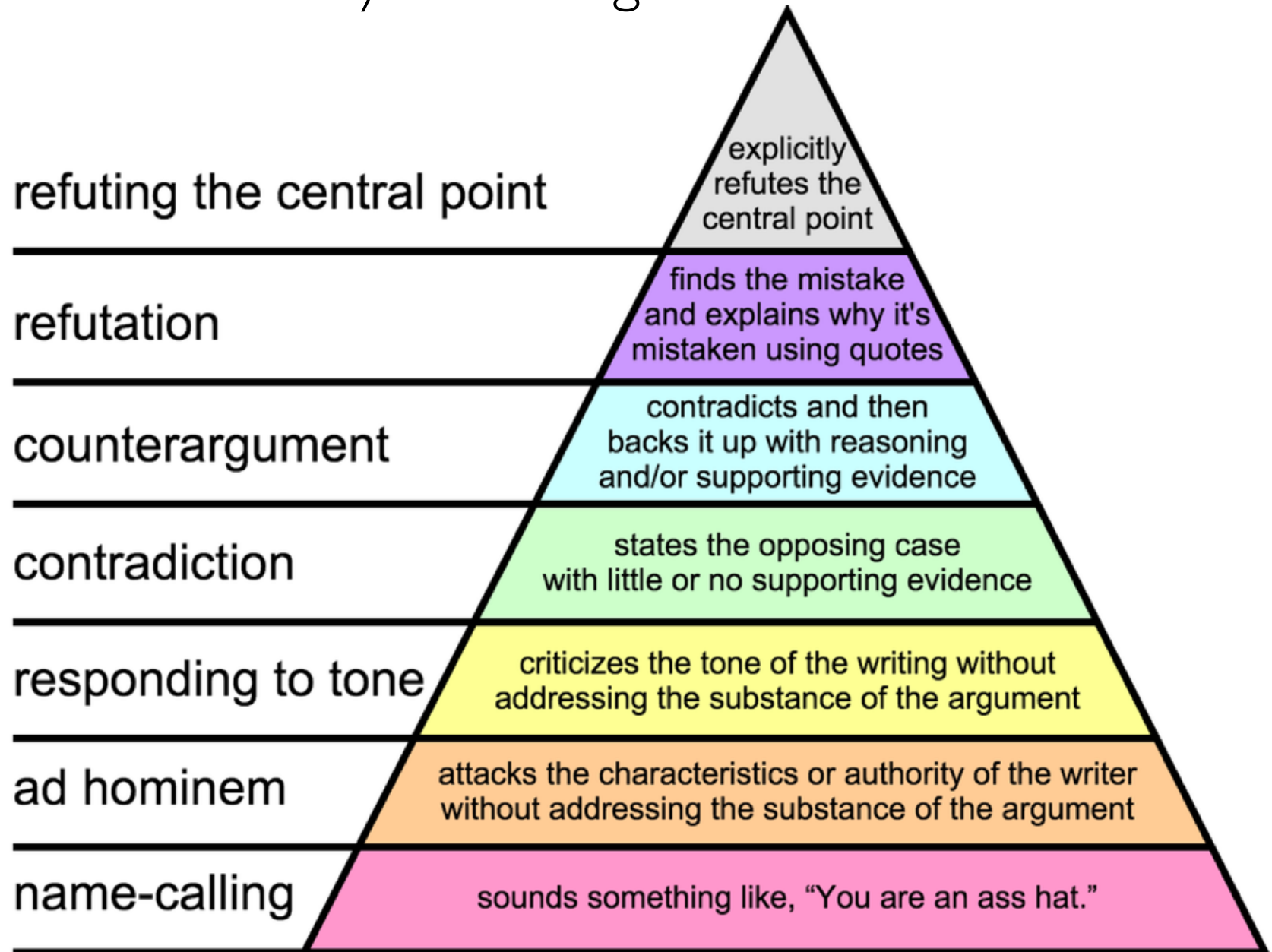


Results



* $p < .05$; ** $p < .001$

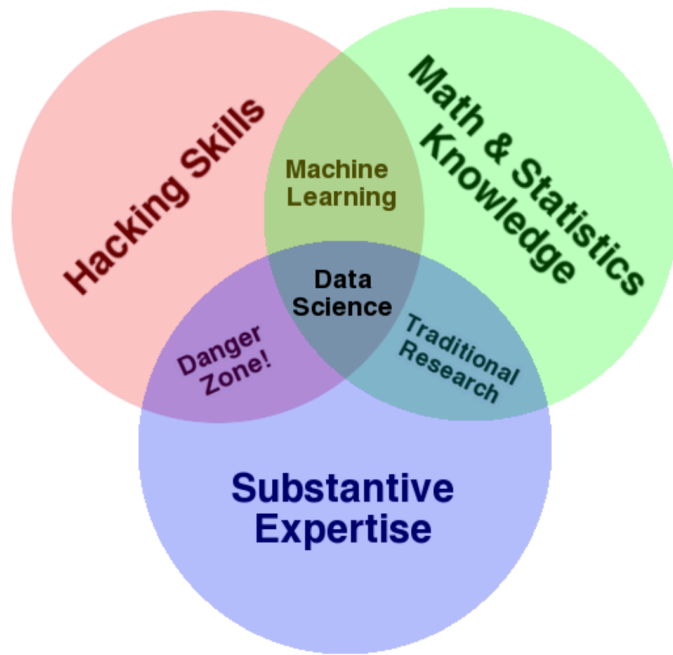
Bloome's Taxonomy of Calling Bullshit



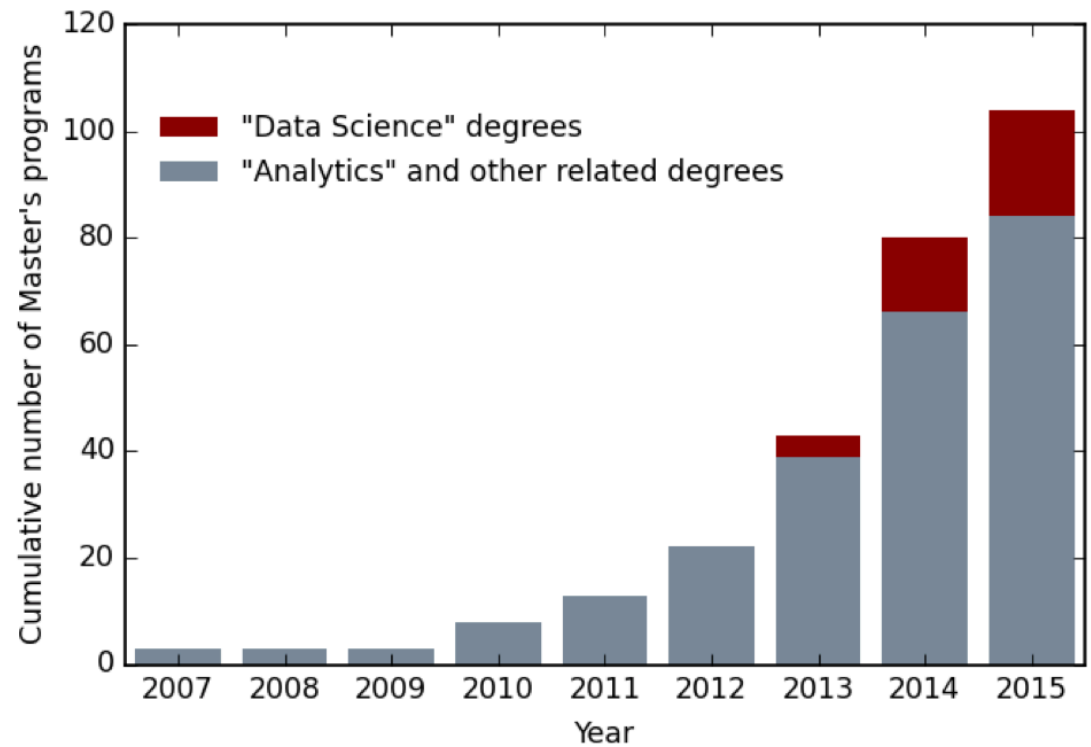


Education

Transcriptable Option in Data Science



Drew Conway, 2009



Institute for Advanced Analytics, NC State



Full Stack

- Unknown algorithm
- Unknown corpus
- Non-customizable
- Non-extensible
- No community development

My updates: recommended based on My Citations [Learn more](#)

Faster unfolding of communities: speeding up the Louvain algorithm

VA Traag - arXiv preprint arXiv:1503.01322, 2015

Networks of Communities and Communities of Networks in Online

Government




Hennrich, R.M. Panayir - Electronic Journal of e-Government, 2014

[See all updates](#)




When you know what you are looking for,
Scholar can usually find it. When you don't,
Scholar is useless. We need tools for *navigation*.




Recommendations




Results for:




   Ecological Theory Suggests That **Antimicrobial** Cycling Will Not Reduce Antimicrobial Resistance In Hospitals - 2003




Expert

   The Relationship Between The Volume Of **Antimicrobial** Consumption In Human Communities And The Frequency Of Re




   Evaluating Treatment Protocols To Prevent **Antibiotic** Resistance - 1996




   The Epidemiology Of **Antibiotic** Resistance In Hospitals: Paradoxes And Prescriptions - 1999




   The Transmission Dynamics Of **Antibiotic**-Resistant Bacteria: The Relationship Between Resistance In Commensal Orga




   Persistent Colonization And The Spread Of **Antibiotic** Resistance In Nosocomial Pathogens: Resistance Is A Regional Pr




Classic

   The Crisis In **Antibiotic** Resistance - 1991

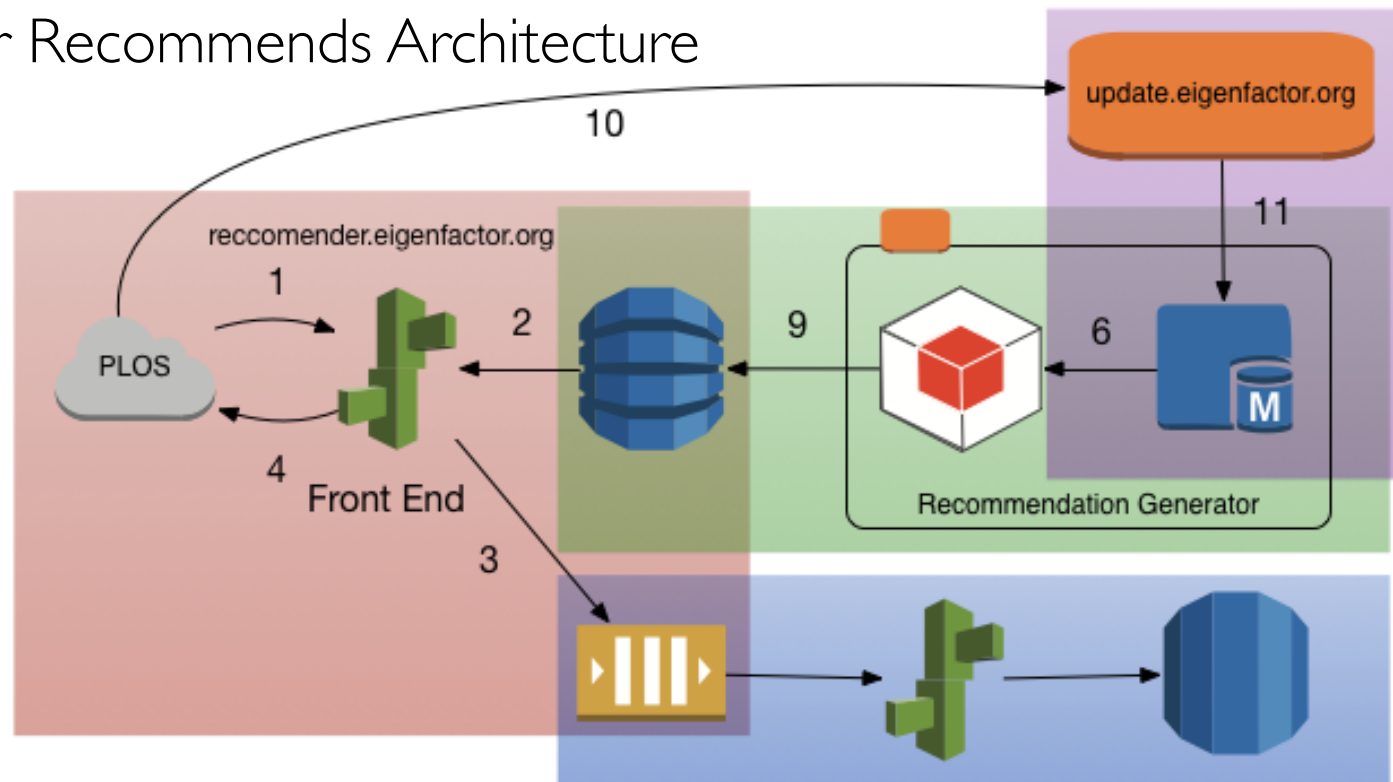
   Epidemiology Of Drug Resistance: Implications For A Post-**Antimicrobial** Era - 1991

   **Drug-Resistant** Salmonella In The United States: An Epidemiologic Perspective - 1985

   The Relationship Between The Volume Of **Antimicrobial** Consumption In Human Communities And The Frequency Of Re

   Evaluating Treatment Protocols To Prevent **Antibiotic** Resistance - 1996

Eigenfactor Recommends Architecture



Recommendation Request

1. Request for recommendation for paper
2. Front-end looks up DOI on a DynamoDB
3. Front-end logs recommendations
4. Front-end returns recommendations

Feedback

1. PLOS sends feedback to the front end
3. The front end logs the feedback in SQS

Analytics

TBD

Recommendation Generation

5. Cron job starts recommender
6. Application reads citation network from DB
7. Application writes recommendations to CSV
8. CSV file is backed up to S3 for offline analysis
9. Transformer takes CSV file and pushes to DynamoDB

Citation DB Updates

10. PLOS calls a private API, providing new DOIs and citations in those documents
11. Application pushes changes to SQL DB



Full Stack Software Engineer

Database



Services & Servers



Backend



WordPress



Frontend



ANGULARJS



jQuery



JS



Monthly Spend



Welcome to the AWS Account Billing console. Your current monthly balance appears below. The accompanying graph shows the proportion of costs spent for each service you use.

Current month-to-date balance for August 2014

\$50,436.95



\$213.99

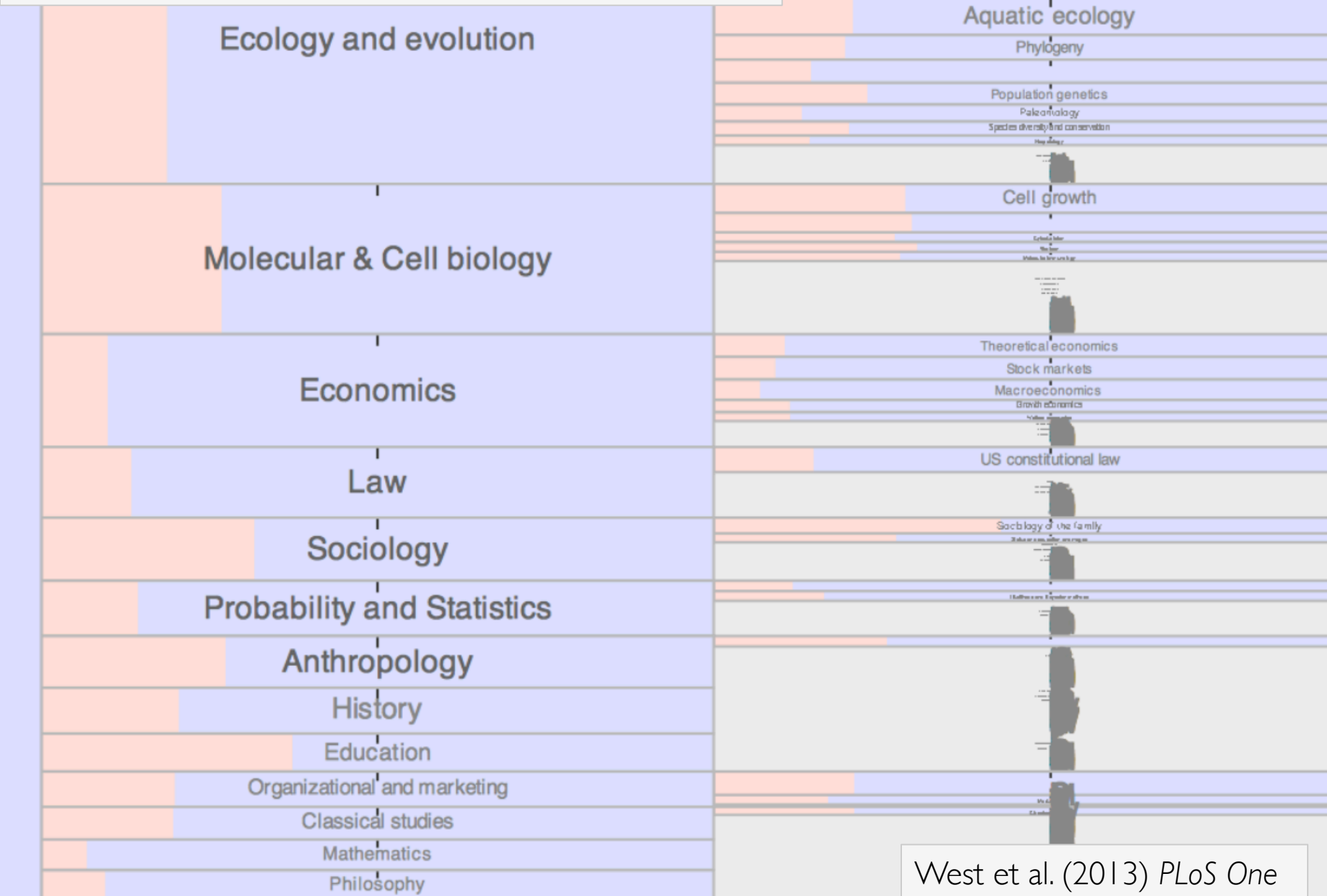
Previous month bill

Image source: <http://www.digibrady.com>



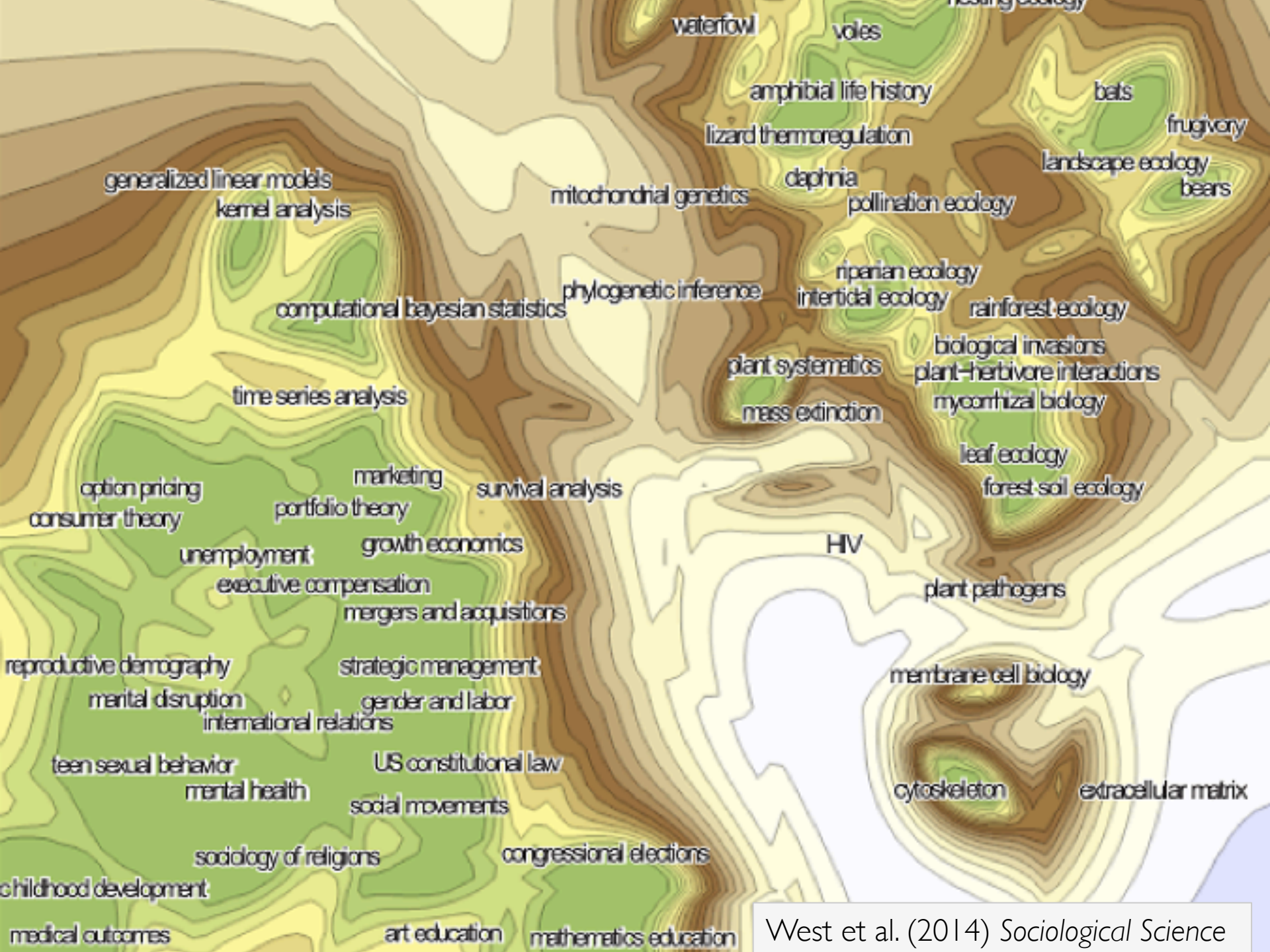
Diversification

Gender Composition in Science





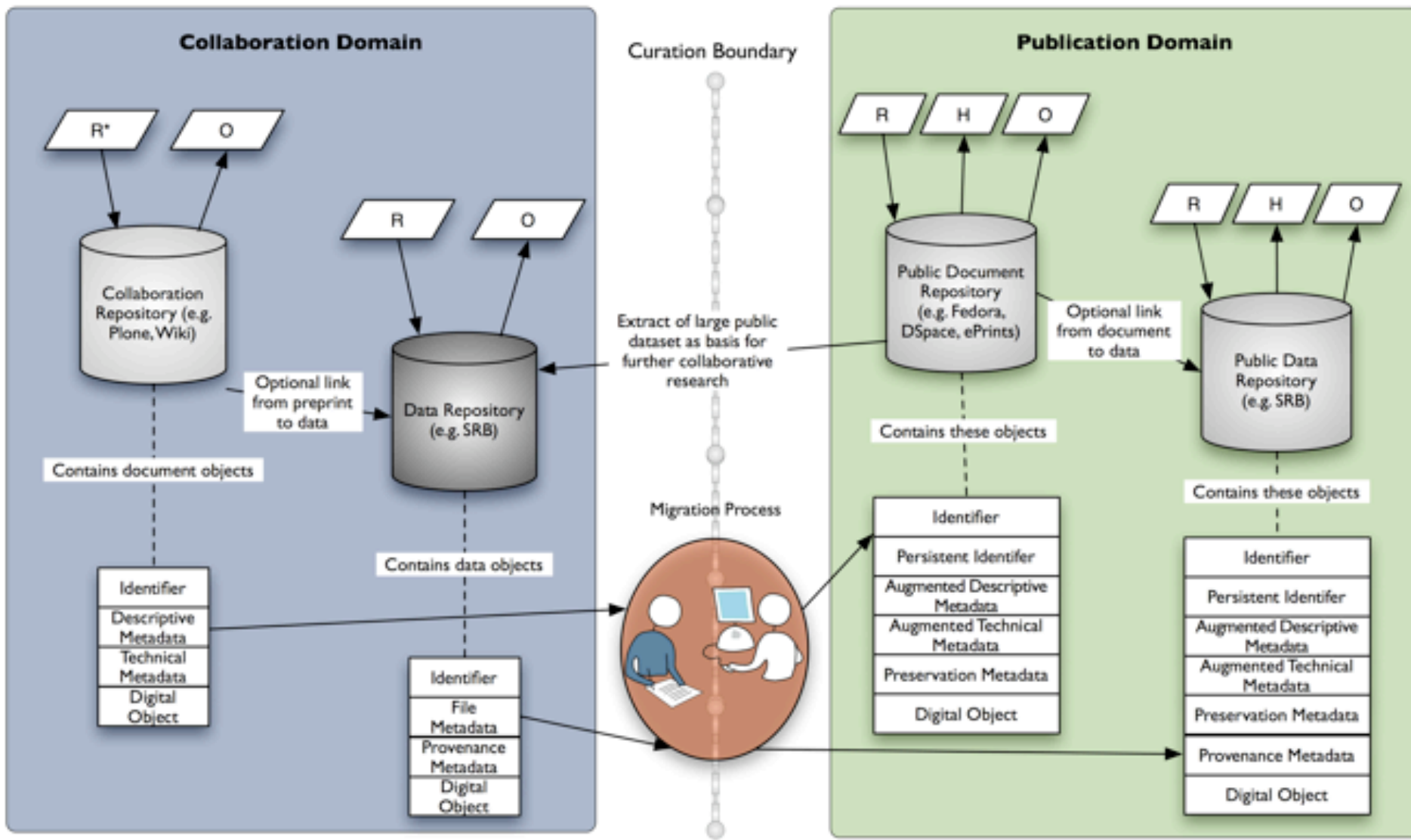
Translation





Data Curation

Collaboration, Publication, and the Curation Boundary



* R = Register, H = Harvest, O = Obtain

Version 1.3, <http://andrew.treloar.net/>, 15/09/07



Tech Transfer



Record of Innovation

Record of Innovation (ROI) Form

This ROI form is used for disclosing innovations to UW CoMotion including mechanical devices, materials, software, digital media and copyrighted works.

Please note the following steps to the ROI submission process:

Step 1: Complete and submit your ROI information online using the form below. If needed, you may save your work using the **Save** feature at the bottom of this form and submit your ROI when completed.

Step 2: After submitting your information online you will be prompted to print a copy of the form to collect the necessary signatures from your contributors. Forward to UW CoMotion (Attn: ROI Coordinator) via campus mail (Box 354990).

You will receive a confirmation email within 24 hours of receipt of this electronic ROI. Please reply to this email and attach any additional information such as manuscripts, grant applications or any other materials that help describe the innovation if available. Within two weeks the technology manager assigned to your ROI will contact you.

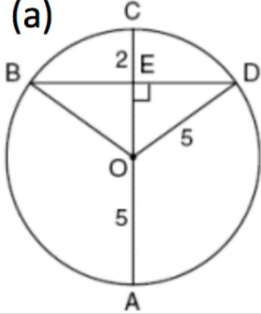
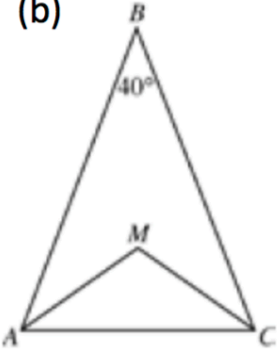
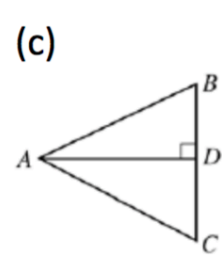
If you submitted an ROI and you do not hear from us within the above timeframe please contact our office at (206) 543-3970 to speak to the ROI Coordinator. We look forward to working with you.

For definitions of the fields, [click here](#) or click the [\[?\]](#) beside the field name.



Robots

Geometry Problem Solver

Questions	Interpretations
<p>(a)</p>  <p>In the diagram at the left, circle O has a radius of 5, and $CE = 2$. Diameter AC is perpendicular to chord BD. What is the length of BD?</p>	<p> <i>Equals(RadiusOf(O), 5)</i> <i>IsCircle(O)</i> <i>Equals(LengthOf(CE), 2)</i> <i>IsDiameter(AC)</i> <i>IsChord(BD)</i> <i>Perpendicular(AC), BD)</i> <i>Equals(what, Length(BD))</i> </p> <p>correct</p> <p>a) 12 b) 10 c) 8 d) 6 e) 4</p>
<p>(b)</p>  <p>In isosceles triangle ABC at the left, lines AM and CM are the angle bisectors of angles BAC and BCA. What is the measure of angle AMC?</p>	<p> <i>IsIsoscelesTriangle(ABC)</i> <i>BisectsAngle(AM, BAC)</i> <i>IsLine(AM)</i> <i>CC(AM, CM)</i> <i>CC(BAC, BCA)</i> <i>IsAngle(BAC)</i> <i>IsAngle(AMC)</i> <i>Equals(what, MeasureOf(AMC))</i> </p> <p>correct</p> <p>a) 110 b) 115 c) 120 d) 125 e) 130</p>
<p>(c)</p>  <p>In the figure at left, The bisector of angle BAC is perpendicular to BC at point D. If $AB = 6$ and $BD = 3$, what is the measure of angle BAC?</p>	<p> <i>IsAngle(BAC)</i> <i>BisectsAngle(line, BAC)</i> <i>Perpendicular(line, BC)</i> <i>Equals(LengthOf(AB), 6)</i> <i>Equals(LengthOf(BD), 3)</i> <i>IsAngle(BAC)</i> <i>Equals(what, MeasureOf(BAC))</i> </p> <p>correct</p> <p>a) 15 b) 30 c) 45 d) 60 e) 75</p>





Translation



Reproducibility



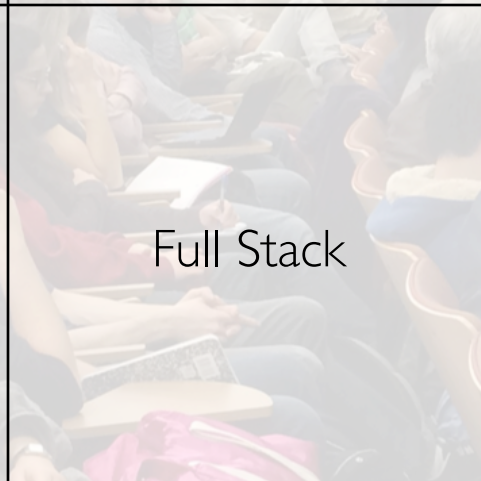
Data Scientists



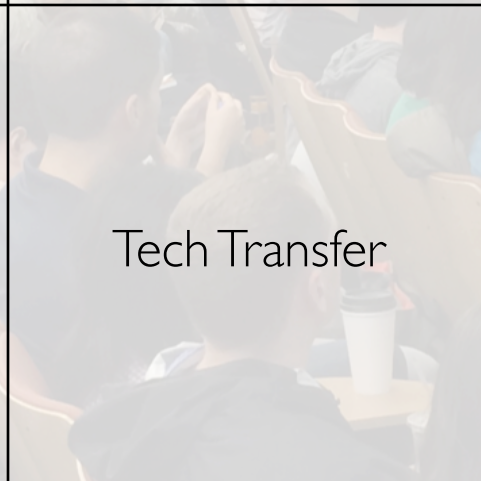
Data Ethics



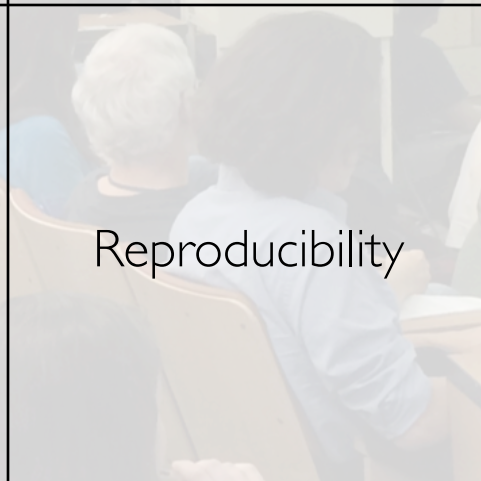
Calling Bullshit



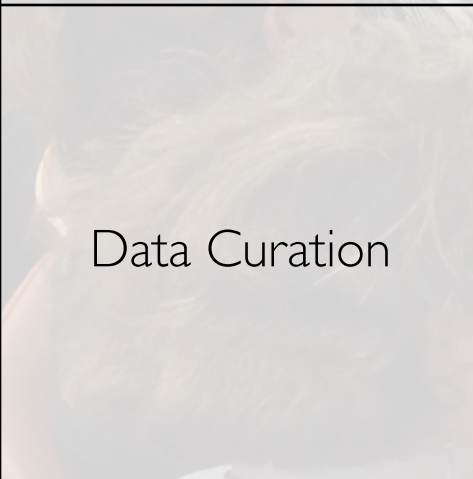
Full Stack



Tech Transfer



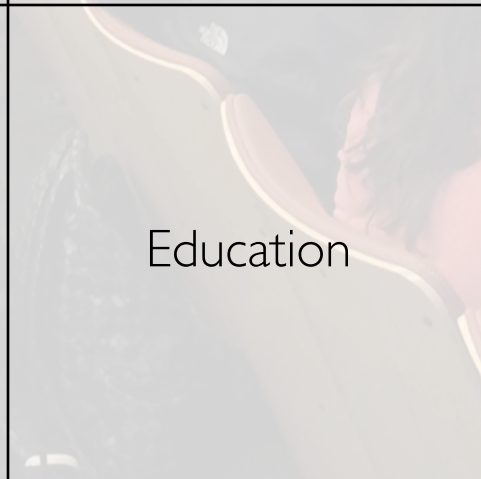
Reproducibility



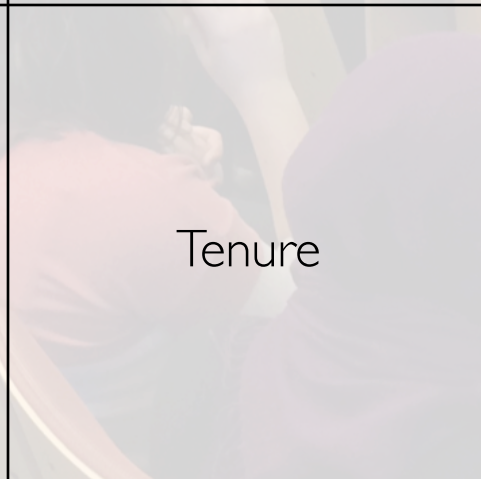
Data Curation




Diversification



Education



Tenure

A surreal landscape featuring a long, straight path that leads towards a large, dark cave opening. The sky above the cave is a vibrant mix of blue, purple, and pink, suggesting a sunset or sunrise. The path is flanked by dark, rocky terrain. The overall mood is dreamlike and mysterious.

Lately, I've been losing sleep.
Dreaming of all the things that we could be.

Translation	Reproducibility	Data Scientists	Data Ethics
Calling Bullshit	Full Stack	Open Access	Reproducibility
Data Curation	Diversification	Education	Tenure