

Data Acumen in Action

USPTO Data Science Ideation

Social and Decision Analytics Division
Biocomplexity Institute

21 June 2022



Welcome and Introductions

Welcome

- Scott Beliveau
- Matthew Baker

Introductions

- Your name
- Your technical background
- Where do you work in the organization
- What do you do

USPTO Club for Open Data Enthusiasts (CODE)

New today

1. Chikaodili Anyikire
2. Maria Ayad
3. Christine Behncke
4. Branka Bowman
5. Claudine Deslauriers
6. Alexandria Emgushov
7. Elman (Al) Krinker
8. James Kucab
9. Chris Marokov
10. Adam Marcetich
11. Paul McCord
12. James Nosal
13. David Orange
14. Michael Robinson
15. Minnah Seoh
16. Brian Slawski
17. Gerard Torres
18. John Wasaff

USPTO attendees at April

Ideation Session and are here today – Thx!

Matthew Baker
Jesse Frumkin
Soren Harward
Austin Hicks

USPTO-Satsyil-Grant Thornton

Scott Beliveau, USPTO Analytics
Kiran Gunda, Satsyil
Don Cencki, Satsyil
Duc Duong, Grant Thornton

Social and Decision Analytics Team

Stephanie S. Shipp, Deputy Director and Professor (economist)

Kathryn Linehan, Research Scientist (mathematician)

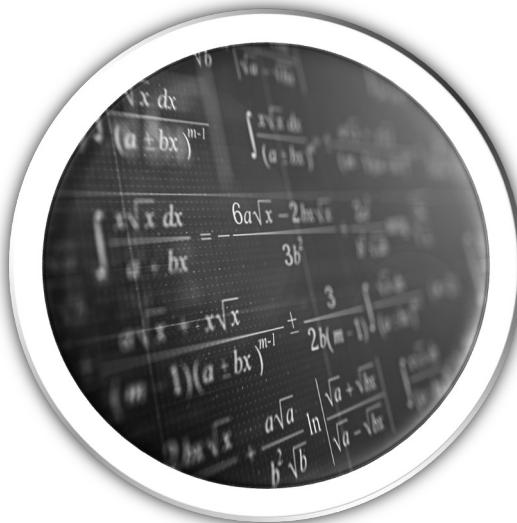
Neil Kattampallil, Research Scientist (systems engineer)

Leonel Swie, Postdoctoral Research Associate (economist)

Micah Iserman, Postdoctoral Research Associate (experimental psychologist)

Biocomplexity Institute

Team science exploring the behavior of massively interacting living systems
to develop practical solutions to real-world problems.



Mathematical
Biocomplexity



Network Systems Science
and Advanced Computing



Social and Decision
Analytics



UNIVERSITY OF VIRGINIA

BIOCOMPLEXITY INSTITUTE

UVA Objectives in Satsyil Proposal

- Leverage UVA's technical and data science expertise to provide **ideation workshops on analytics** that include methods to integrate statistical learning and machine learning
- **Conduct research** on science and technology trends through a **disciplined approach** of (1) problem identification, (2) contextual point of view, (3) framed research methods, and (4) review of traditional literature, leveraging academic research, and grey literature to glean best practices and insight.
- **Translate research findings into information briefings** that help USPTO by providing the context to define, evaluate, and interpret findings in support of technical direction and priority.
- Present the analysis of findings using **visualizations and interpretations** to convey the context, purpose, and implications of the research and findings.

Ideas from April 19, 2022 Ideation Session

There were many interesting ideas discussed or entered in the zoom chat, both internally and externally focused.

Here are a few of them:

- Patent Emerging Trends Analysis – speed, size, acceleration, momentum
- Patent Application Tool to check for potential infringements.
- Improve current patent classification process using Natural Language Processing/machine learning
- Build Network graph of patent citations to better understand similarities and connections between patents

Agenda

Data Acumen - What is it? Why should you care? How do you gain it?

Art of the Possible – data-driven learning case study example

Discussion #1 What keeps you up at night?

- What challenges do you face in your work that you think data and data science methods could help address?

Discussion #2 – Emerging patents/trademarks trends topic - proposed pilot project – *we would like to divide you into two breakout rooms*

- What aspects of this topic interest you?
- How would you use this information?
- How would you scope for a pilot?

Synthesize and next steps

What is data science?

Data science is learning from data in the context of
problems.

No problems, no data science

Guiding process for Data Science Ideation

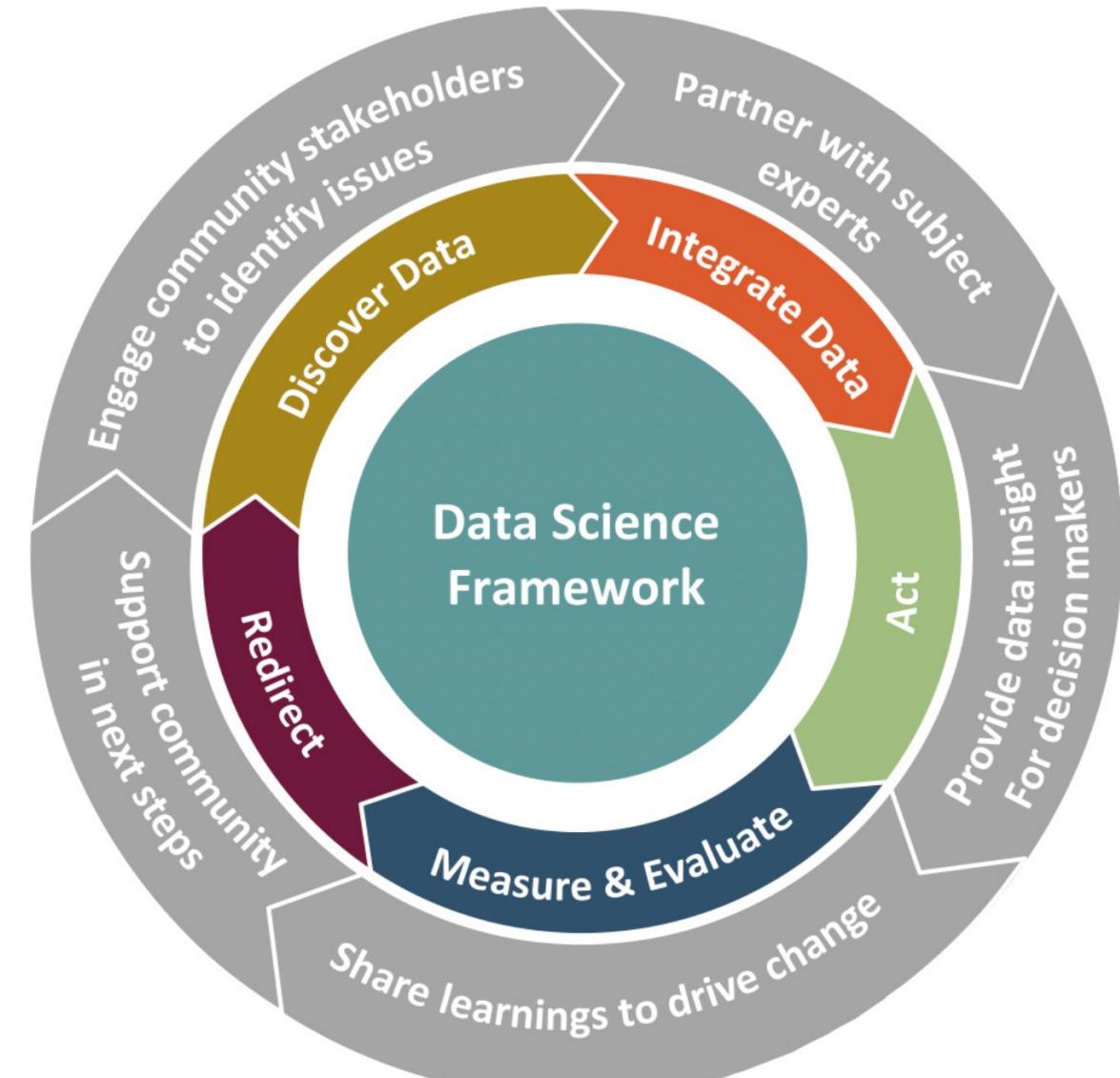
Learning Through Data Driven Discovery

Outer wheel: continuous interaction and communication across stakeholders

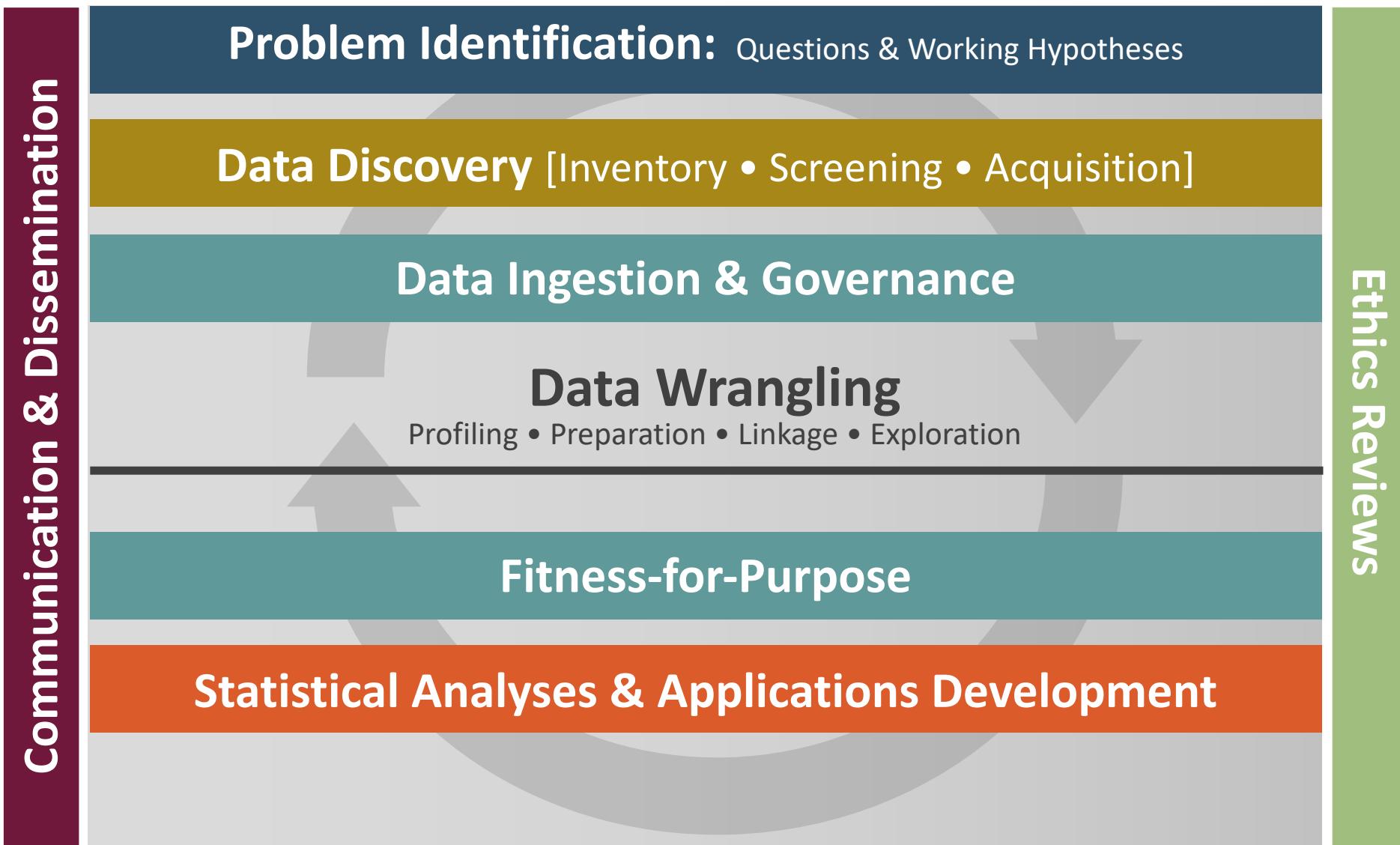
Middle wheel: data-driven learning process

Frontier between the outer and middle wheels: active collaboration between all partners

Inner circle: rigorous research framework to guide the data science



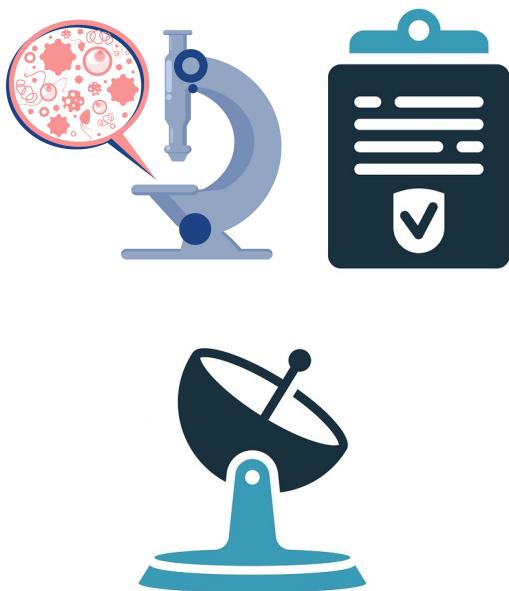
Data Science Framework



Harnessing All Data

Massive data repurposing

Designed
Data



Administrative
Data



Opportunity
Data



Procedural
Data



Data acumen

Data acumen is the ability to make good judgments about the use of data to support problem solutions.

How do you trust and use the results for data-informed decisions?

Levels of data acumen



Data scientists



Subject matter experts



Consumers

Which are you?

ART-OF-THE-POSSIBLE

Case Study – What R&D trends are detected using administrative data?

R&D Trends

Problem: To assess the feasibility of detecting R&D trends using administrative data for strategic planning

Questions:

- What specific research areas does this public funding support?
- Can we conduct case studies for strategic planning, e.g., pandemics, and artificial intelligence?
- What are R&D trends over time by funding agencies?
- Can we leverage existing data in lieu of new data collection & reporting?

Types of Analyses: Information Retrieval, Topic Modeling



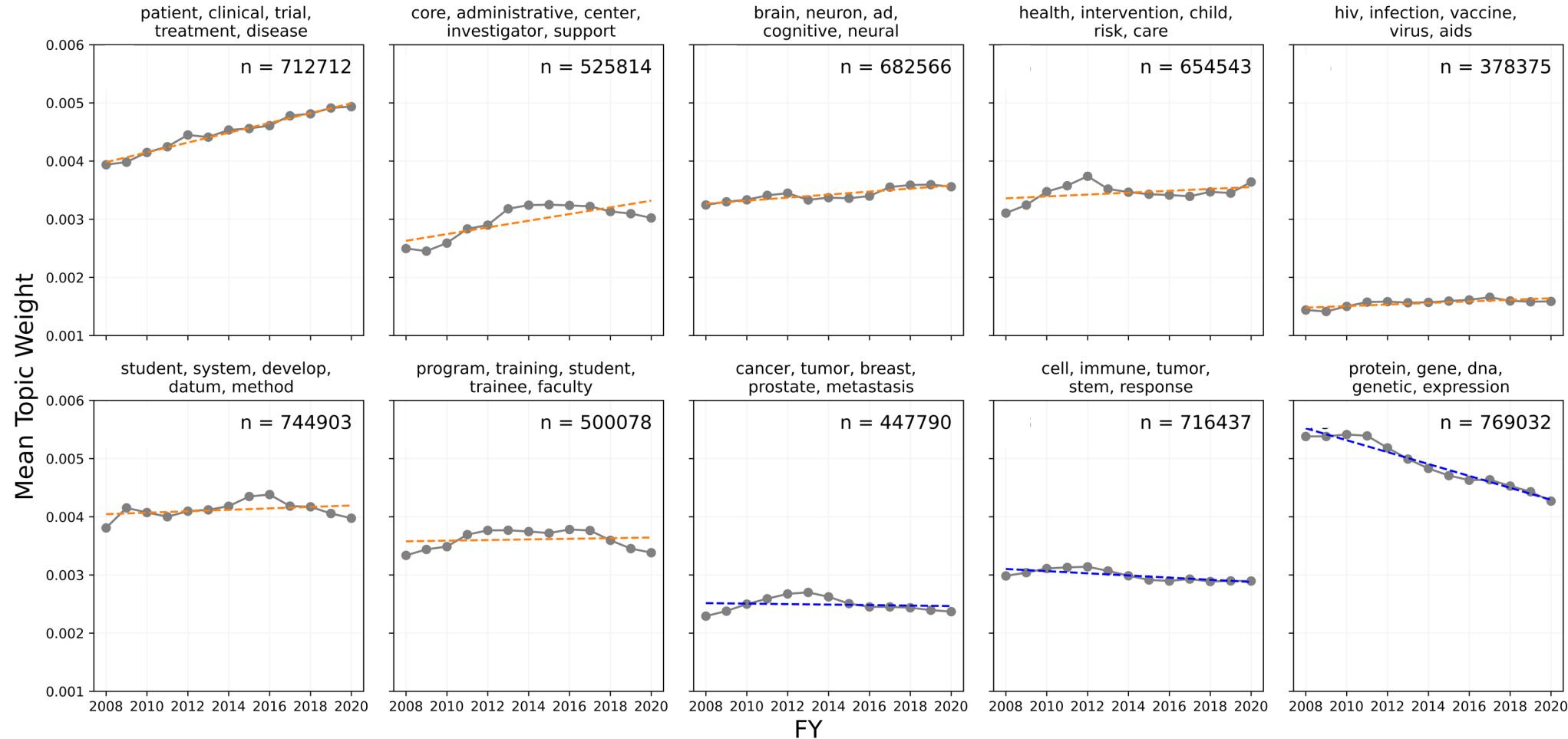
BIOCOMPLEXITY INSTITUTE



Topic Trends

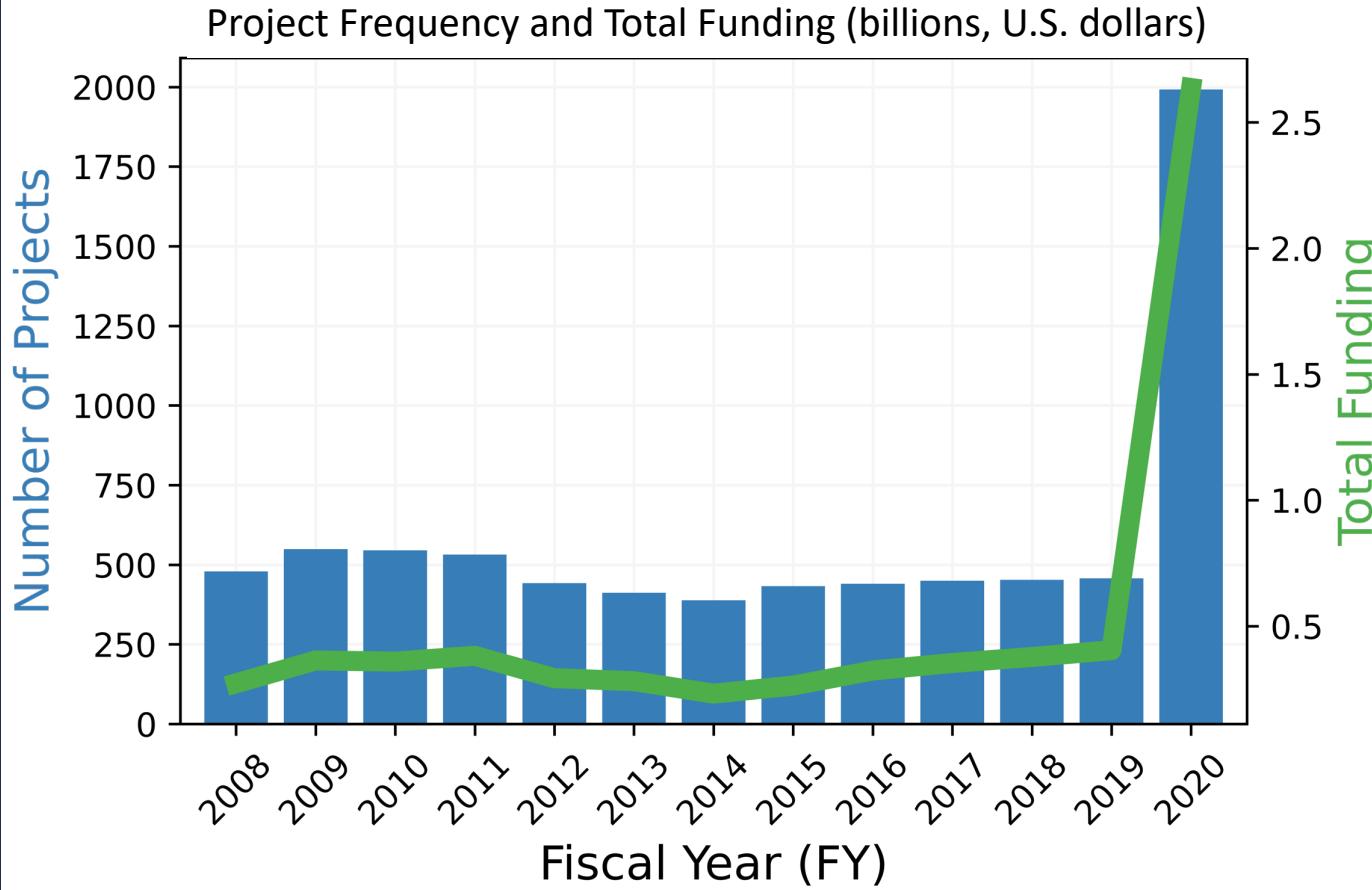
In addition to finding topics in a text corpus, we can also track topic trends
Trend line is found using regression on fiscal year and mean topic weight,
2008-2020

Federal RePORTER: Topic Trends Over Time (10 Topics)

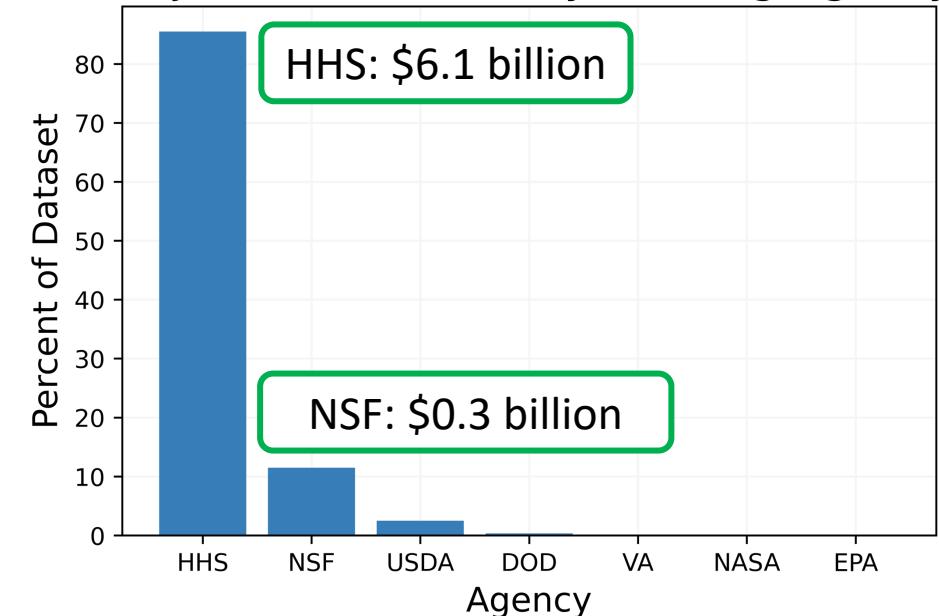


Pandemics R&D Trends

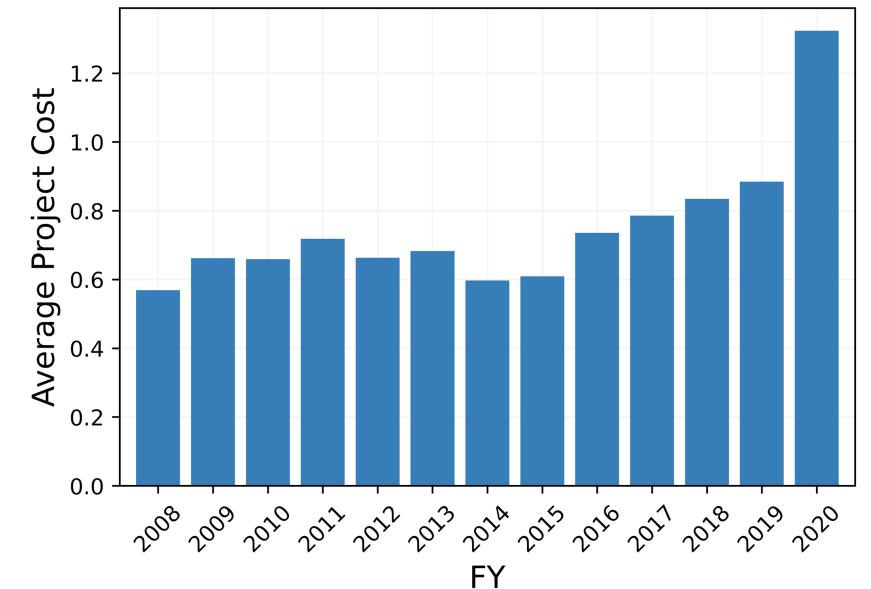
Pandemics Corpus: **7,571 abstracts**



Project Distribution by Funding Agency

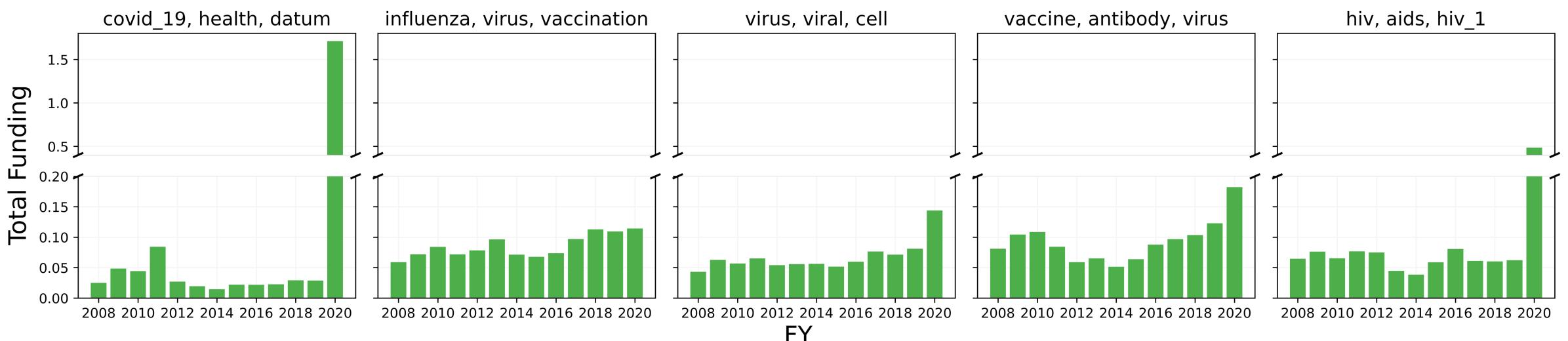
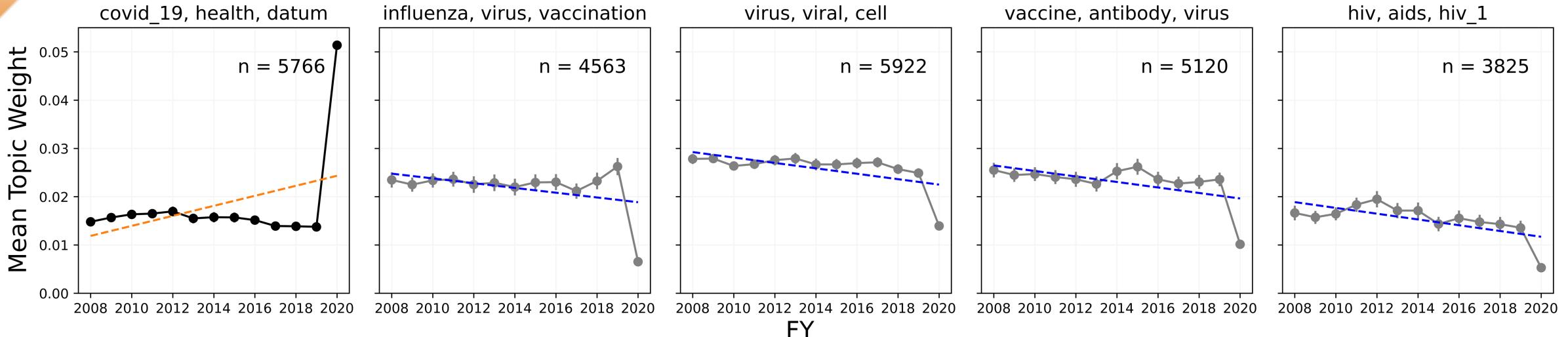


Average Project Cost (millions, U.S. dollars)



Pandemics R&D Trends

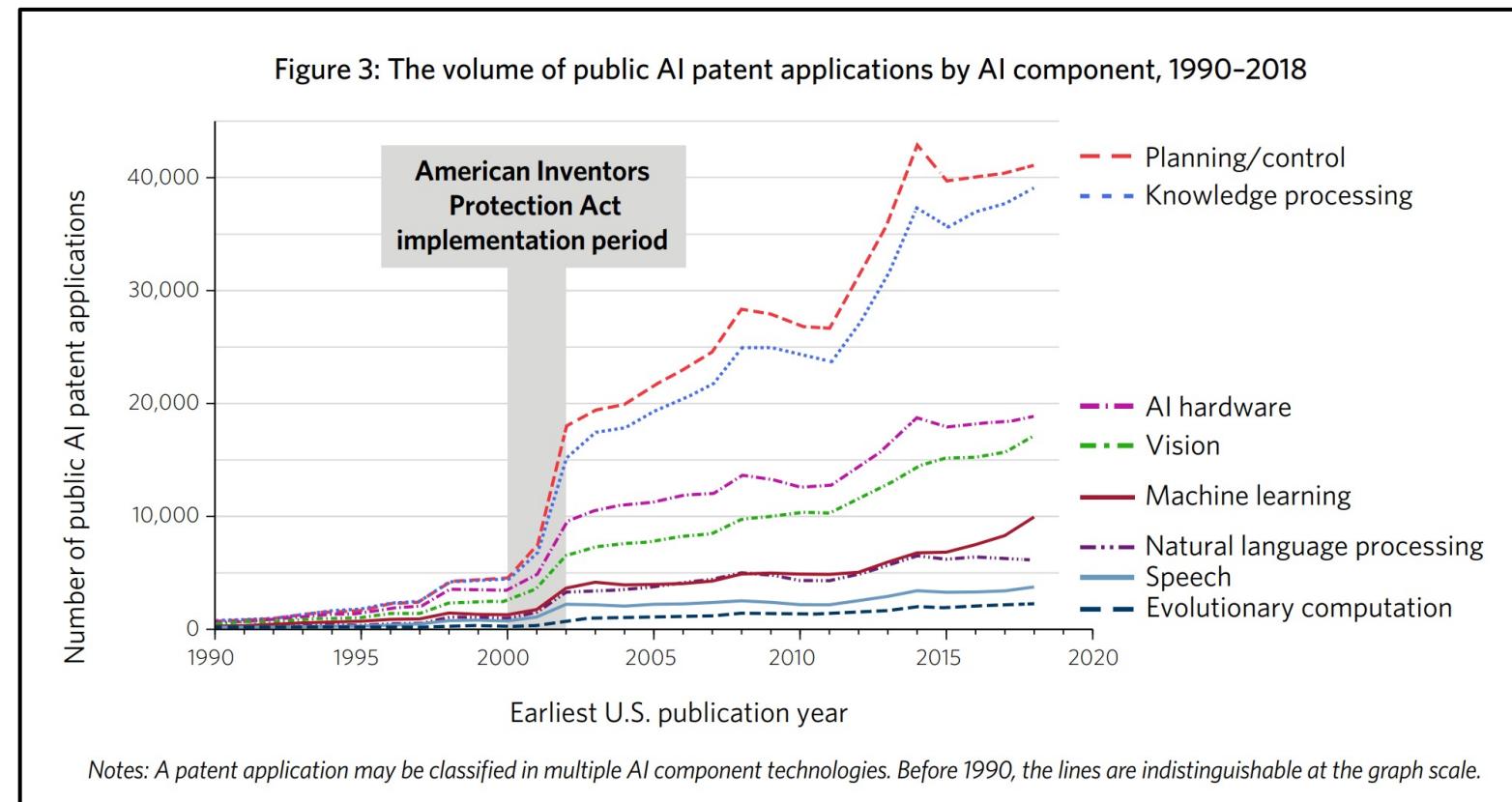
Topic Trends and Funding (billions, U.S. dollars)



Propose Patent Trends Pilot

Leveraging USPTO and UVA work, can we address questions, e.g.,

- What are characteristics of patents by field?
- What are emerging trends in patent topics?
- What are interesting case studies for a deep dive?



Graphic Source: Office of the Chief Economist. (2020). *Inventing AI: Tracing the diffusion of artificial intelligence with U.S. patents* (IP Data Highlights No. 5). U.S. Patent and Trademark Office. <https://www.uspto.gov/sites/default/files/documents/OCE-DH-AI.pdf>

Discussion #1 – Identifying Your Challenges

Goal – learn about USPTO needs and data science opportunities

- What challenges do you face in your work that you think data and data science methods could help address?

Discussion

Discussion #2 – Your Input

Pilot Project to Examine Patent/Trademark Trends

Goal – Define the pilot project such that engages data science and provides new insights

Divide into two breakout rooms

**What are Patent Emerging Trends Analysis,
e.g., speed, size, acceleration, momentum?**

- What aspects of this topic interest you?
- How would you use this information?
- What sources of data might be useful?
- How would you scope for a pilot?

USPTO Data Learning Team

A USPTO team to work with the UVA team to provide input, context, and feedback throughout the pilot (involves ~ 4 hours per month for 3 months).

