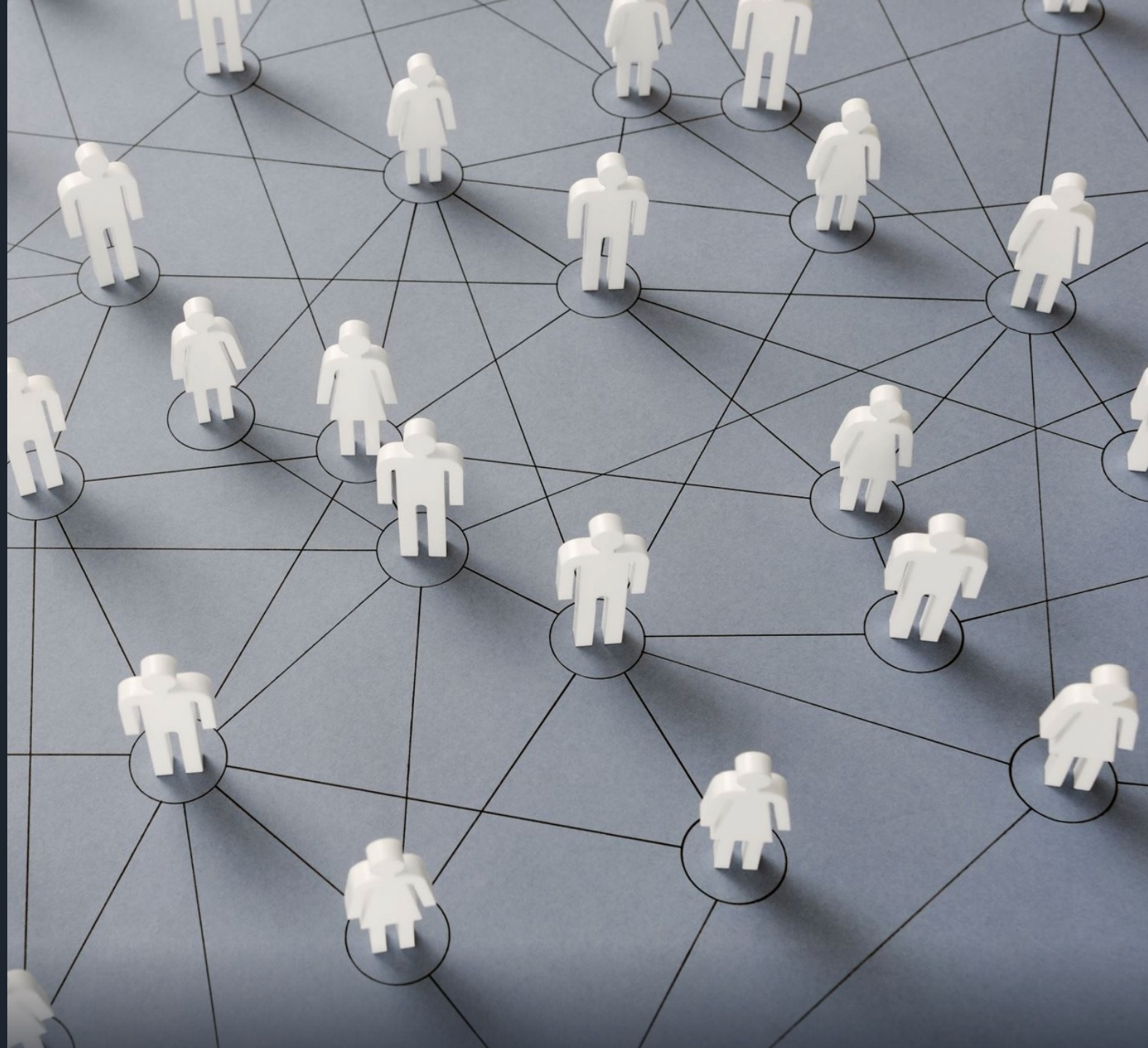
Bias in conflict data

Evan Muzzall, PhD
Software and Services for Data Science, Stanford University
https://ssds.stanford.edu/
2021 December 8

Pop Fam 8676: Epidemiological Methods to Measure Human Rights Abuses, Fall 2021, Columbia University



Today’s talk

- What is data science?
- Bias in research, how research products are stored as media, and considering if research is a good “fit”
- Impacts of bias can have far-reaching consequences
- Basic forms of bias in conflict data, research, and interpretation

Computational and Inferential Thinking

Search this book...

1. What is Data Science?

Data Science is about drawing useful conclusions from large and diverse data sets through exploration, prediction, and inference. Exploration involves identifying patterns in information. Prediction involves using information we know to make informed guesses about values we wish we knew. Inference involves quantifying our degree of certainty: will the patterns that we found in our data also appear in new observations? How accurate are our predictions? Our primary tools for exploration are visualizations and descriptive statistics, for prediction are machine learning and optimization, and for inference are statistical tests and models.

Statistics is a central component of data science because statistics studies how to make robust conclusions based on incomplete information. Computing is a central component because programming allows us to apply analysis techniques to the large and diverse data sets that arise in real-world applications: not just numbers, but text, images, videos, and sensor readings. Data science is all of these things, but it is more than the sum of its parts because of the applications. Through understanding a particular domain, data scientists learn to ask appropriate questions about their data and correctly interpret the answers provided by our inferential and computational tools.

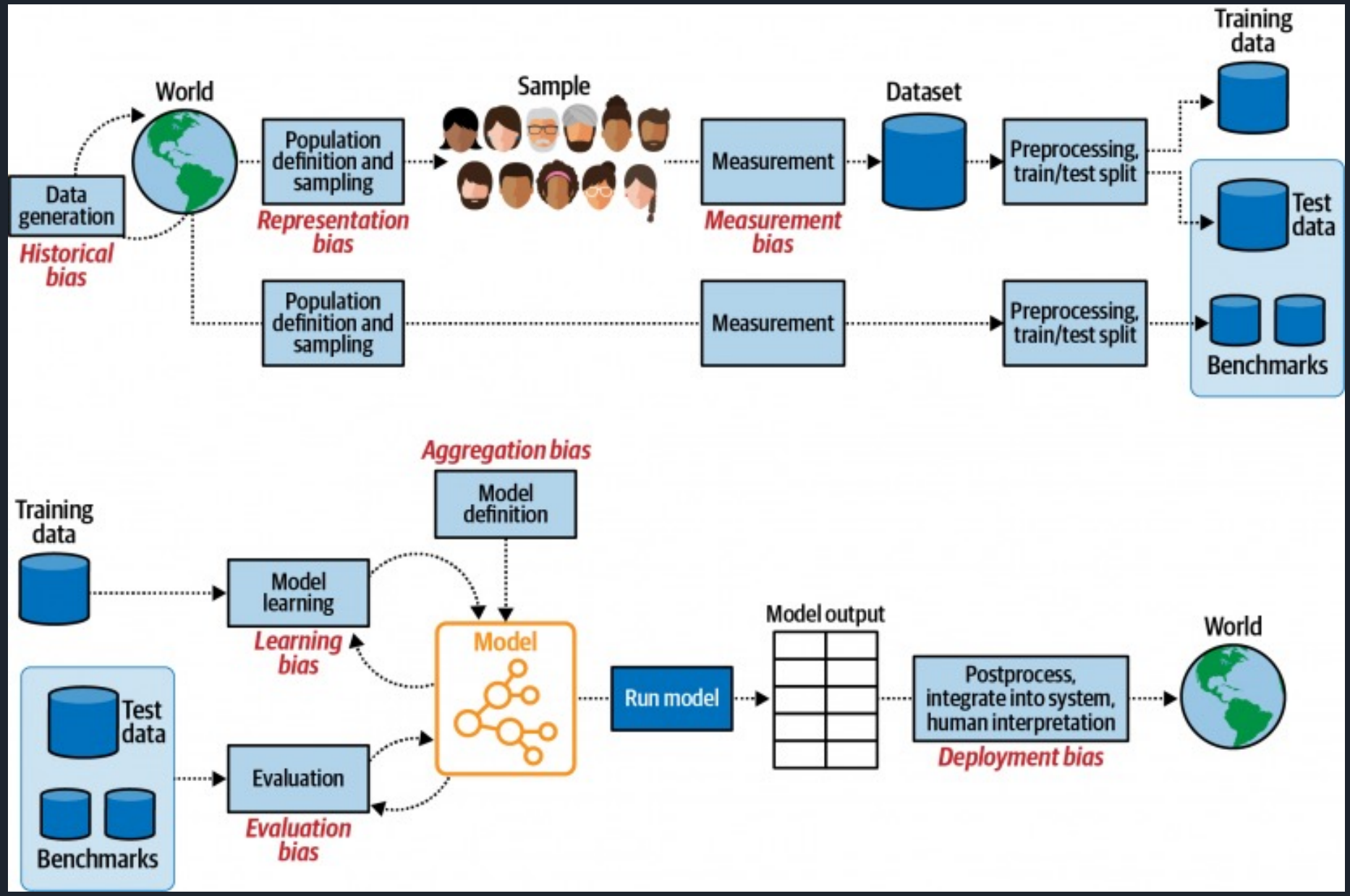
previous
Computational and Inferential Thinking: The Foundations of Data Science

next
1.1. Chapter 1: Introduction

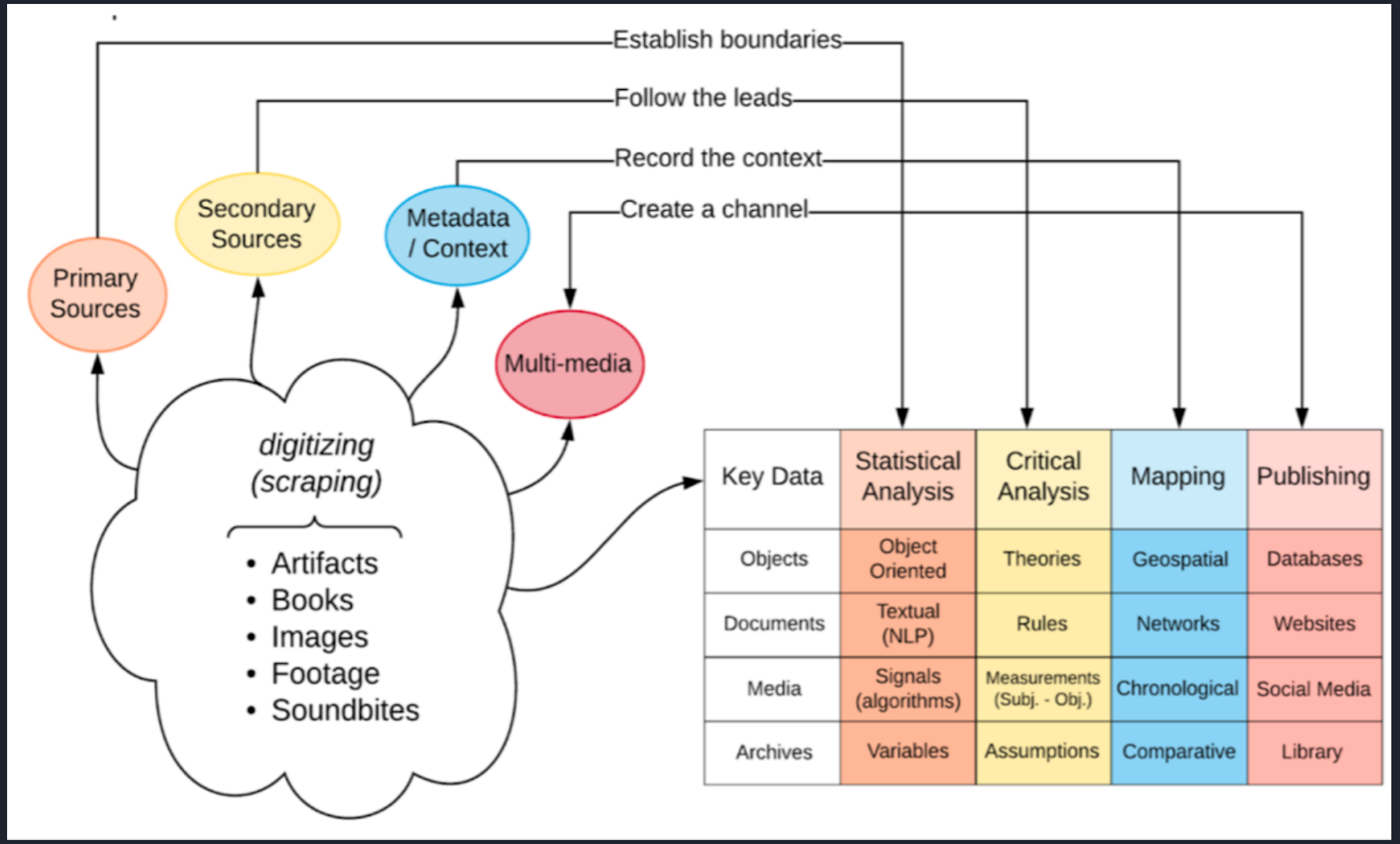
By Ani Adhikari and John DeNero and David Wagner

https://inferentialthinking.com – Chapter 1

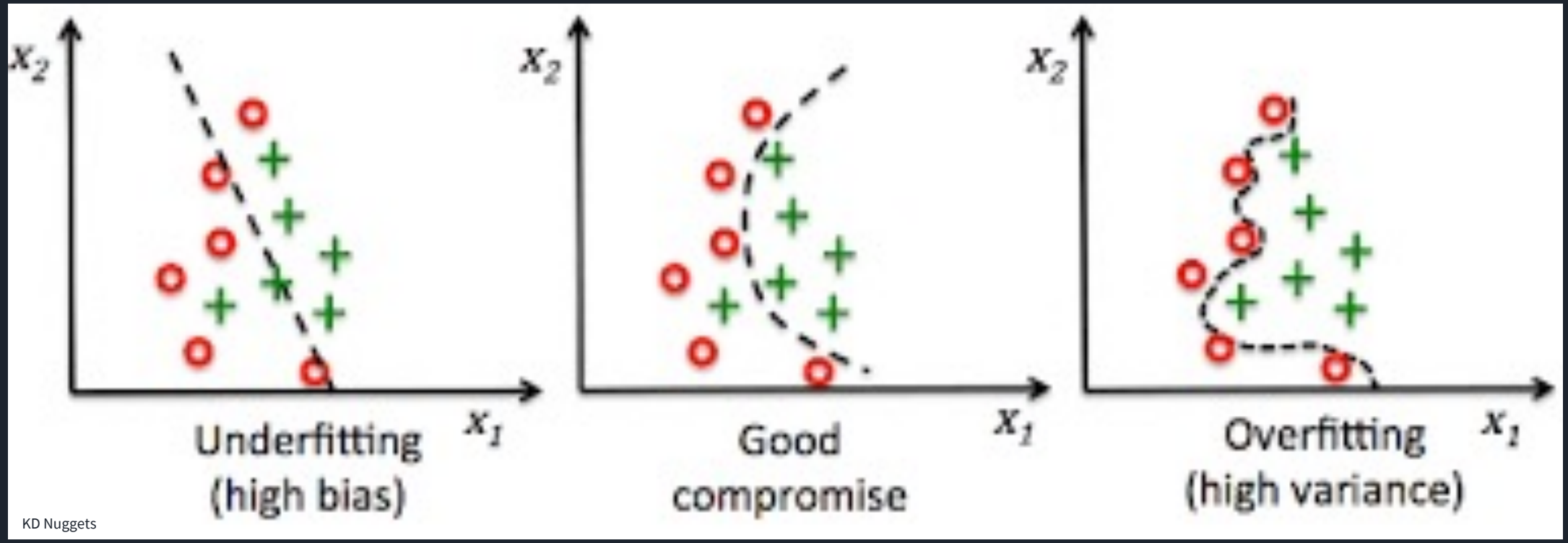
Bias in research



Research as media

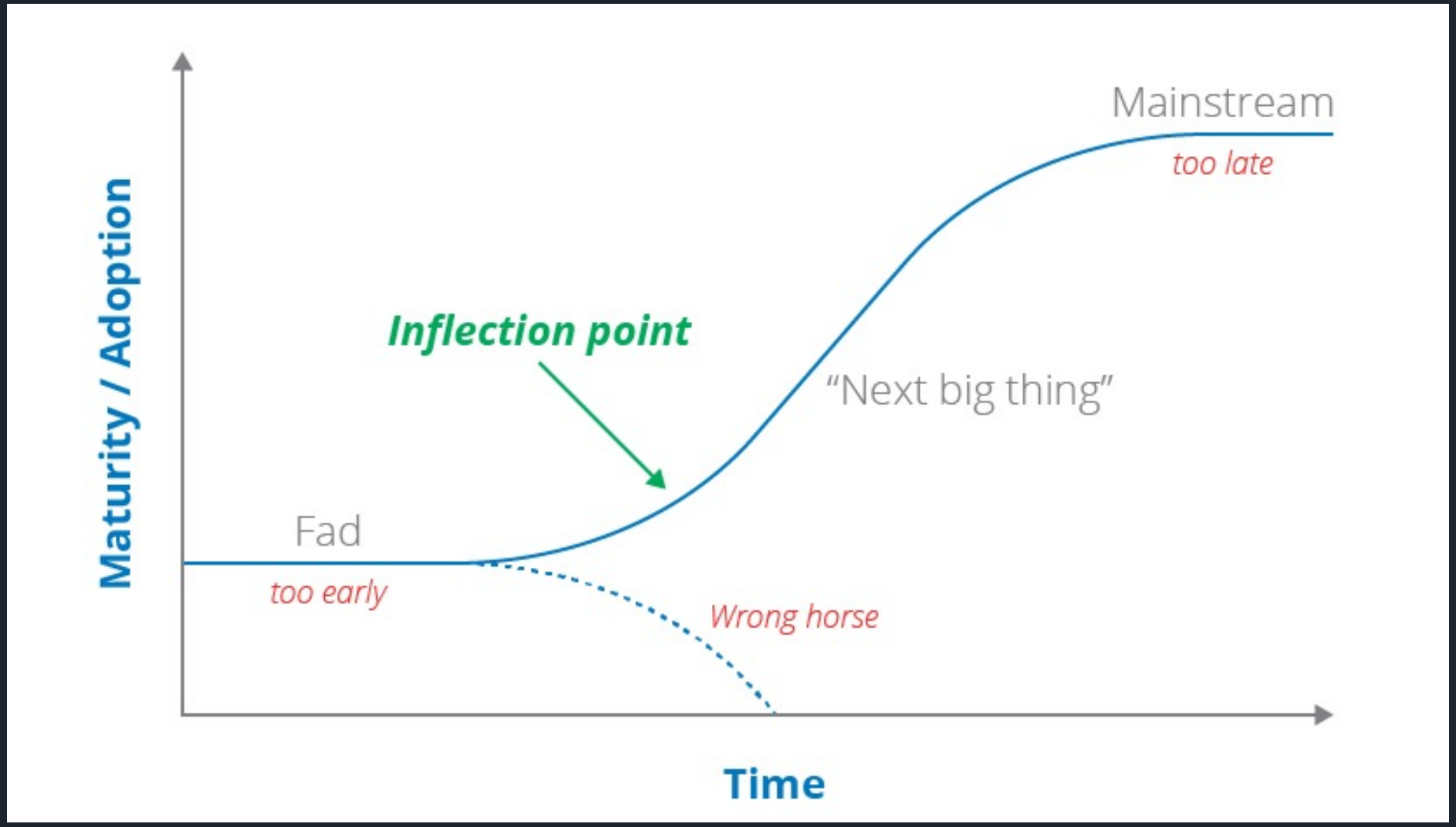


Research “fit”



KD Nuggets

Impacts of bias can be far-reaching



Armed conflict

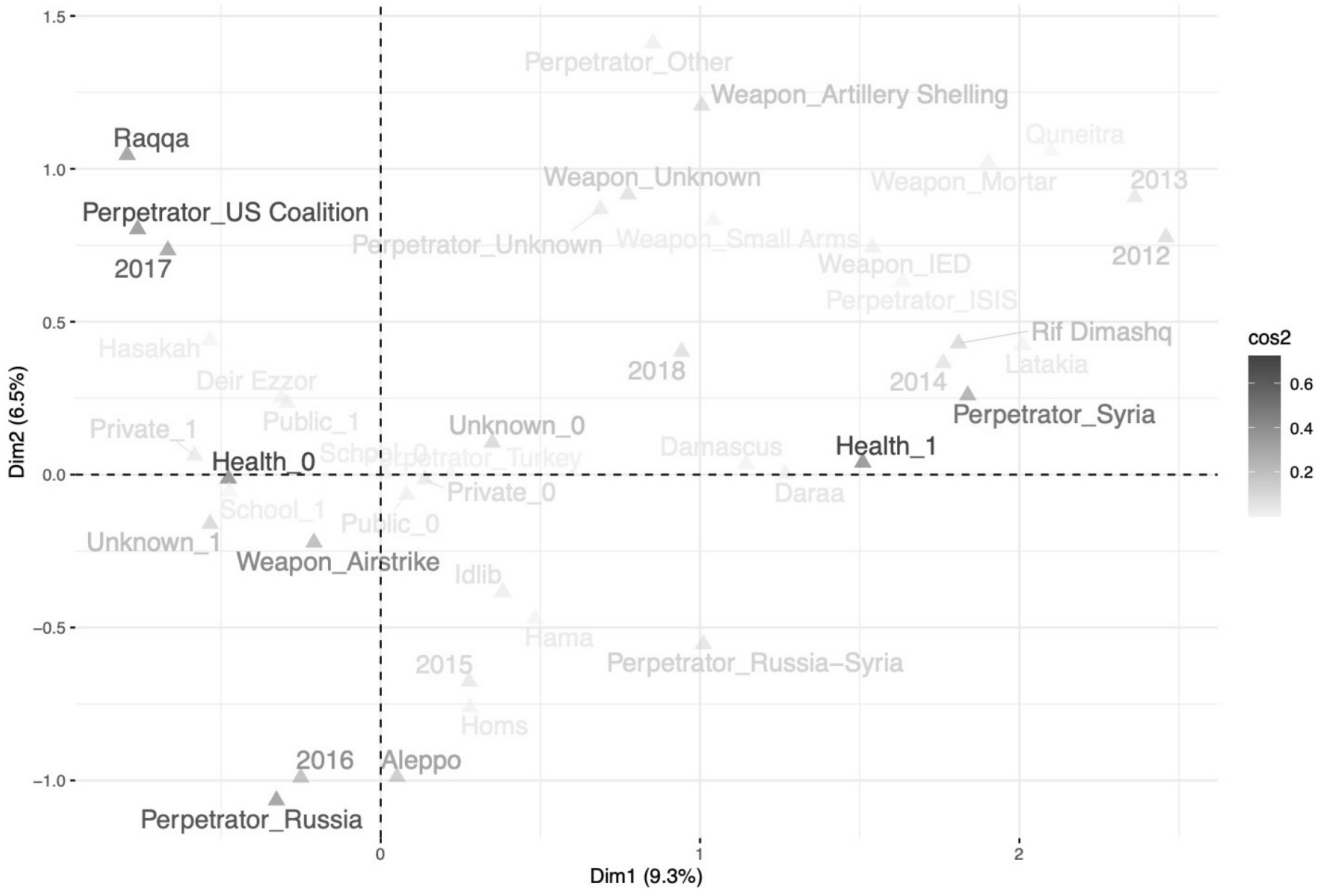


Figure 3 Multiple correspondence analysis squared cosine distance (cos2) plot. Very high confidence of point placement is labelled with darker text, high confidence placement with grey and moderate confidence with light grey text. Note disparate locations of whether healthcare was attacked (Health 1) or not (Health 0) along the x-axis.

Bias in conflict data

- Perspectives: individual, local, regional, and global histories and contexts
- Interests and alliances between governments, economies, and warring parties
- Researcher categorizations of people and things in space and time
- Logistical challenges for data collection and curation; what data are missing versus unrecorded/not able to be recorded?
- Statistical misrepresentation; biased interpretation of the quantification of conflict
- Dissemination of inaccurate information by academics, governments, NGOs, human rights groups, etc.