Unifying Data Science

Nick Eubank

1. Provide unified conceptual framework for relating data science tools
2. Learn causal inference
3. Get practice developing data science project start-to-finish

How did Data Science become a thing?

How did Data Science become a thing?

- Academic research is organized into silos:

How did Data Science become a thing?

- Academic research is organized into silos:
  - Computer Science
  - Statistics
  - Economics
  - Political science
  - Engineering

How did Data Science become a thing?

- Academic research is organized into silos:
  - Computer Science
  - Statistics
  - Economics
  - Political science
  - Engineering

$\Rightarrow$ Development of new tools occurred *within* each silo.

Where are we today?

Very little cross-pollination across silos

Where are we today?

Very little cross-pollination across silos

- Lots of duplication of development.

Where are we today?

Very little cross-pollination across silos

- Lots of duplication of development.
- Every silo has its own vocabulary.

Where are we today?

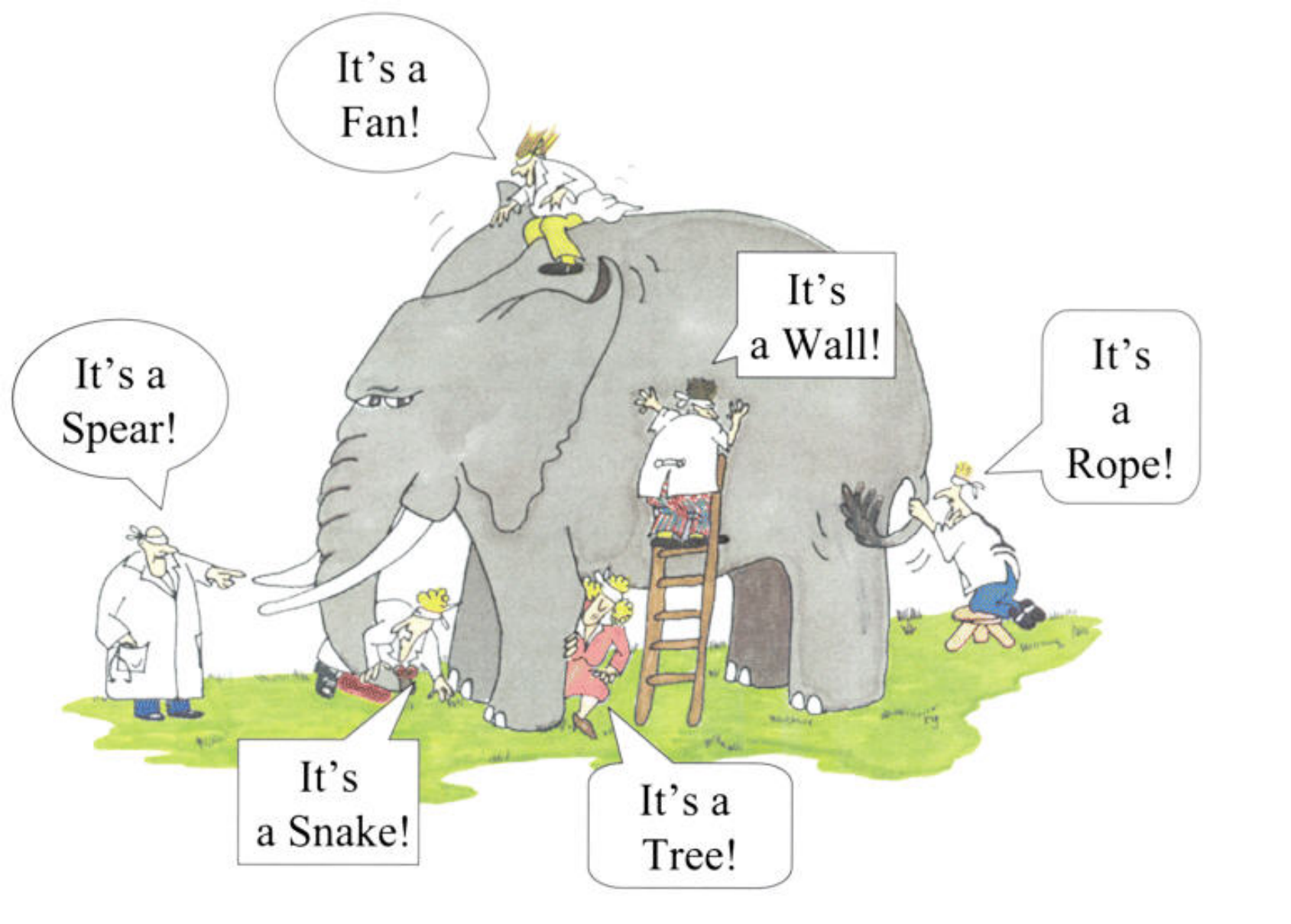
Very little cross-pollination across silos

- Lots of duplication of development.
- Every silo has its own vocabulary.
- Each silo has focused on the aspects most relevant to their applications. e.g.:

Where are we today?

Very little cross-pollination across silos

- Lots of duplication of development.
- Every silo has its own vocabulary.
- Each silo has focused on the aspects most relevant to their applications. e.g.:
  - CS likes to classify things and make predictions, don't care how model works
  - Social scientists like to make causal statements, don't care about predictive power



⇒ This is where we are *now*.

What is (empirically) Data Science?

What is (empirically) Data Science?

An effort to unify the development of quantitative methods

What is (empirically) Data Science?

An effort to unify the development of quantitative methods

$\rightarrow$ Recognize the elephant

This Class

Discipline of learning how best to answer questions using quantitative data.

This Class

1. Introduce a taxonomy of questions
   Descriptive, causal, predictive

This Class

1. Introduce a taxonomy of questions

   Descriptive, causal, predictive

2. For each class of questions, we will discuss the relative strengths and weaknesses of different empirical approaches

This Class

1. Introduce a taxonomy of questions

   Descriptive, causal, predictive
2. For each class of questions, we will discuss the relative strengths and weaknesses of different empirical approaches
   - Unsupervised machine learning
   - Supervised machine learning
   - Range of causal inference techniques

     e.g. experiments, matching, regression, differences-in-differences
   - Other approaches to descriptive analysis

This Class

The tool you use should be dictated by the question you seek to answer

This Class

1. Provide unified conceptual framework for relating data science tools
2. Discuss descriptive questions
3. Learn causal inference
   Deep dive – $\sim$ half the semester
4. Discuss prediction

This Class

1. Provide unified conceptual framework for relating data science tools
   Practice generating questions
2. Discuss descriptive questions
3. Learn causal inference
   Deep dive – $\sim$ half the semester
4. Discuss prediction

This Class

1. Provide unified conceptual framework for relating data science tools
   Practice generating questions
2. Discuss descriptive questions
   Relatively brief
3. Learn causal inference
   Deep dive – $\sim$ half the semester
4. Discuss prediction

This Class

1. Provide unified conceptual framework for relating data science tools
   Practice generating questions
2. Discuss descriptive questions
   Relatively brief
3. Learn causal inference
   Deep dive – $\sim$ half the semester
4. Discuss prediction
   Relative merits of supervised machine learning v. causal methods

Data Science Project

Over semester, you will also develop a data science project from start-to-finish

- Teams of 2-3, grouped by interest and experience
- On topic of your own choosing

Data Science Project

Over semester, you will also develop a data science project from start-to-finish

- Teams of 2-3, grouped by interest and experience
- On topic of your own choosing

$\rightarrow$ Nice portfolio piece

Data Science Project

Over semester, you will also develop a data science project from start-to-finish

- Teams of 2-3, grouped by interest and experience
- On topic of your own choosing

$\rightarrow$ Nice portfolio piece
$\rightarrow$ MIDS first-years: Capstone with training wheels

Who Are We?

I am a social scientist

Who Are We?

I am a social scientist

- PhD in Political Economy, Masters in Economics, BA in Economics and Political Science
- Research on international development, social networks, election administration, gerrymandering

Who Are We?

I am a social scientist

- PhD in Political Economy, Masters in Economics, BA in Economics and Political Science
- Research on international development, social networks, election administration, gerrymandering

Zeren Li (TA)

Who Are We?

I am a social scientist

- PhD in Political Economy, Masters in Economics, BA in Economics and Political Science
- Research on international development, social networks, election administration, gerrymandering

Zeren Li (TA)

- PhD Candidate in Political Science
- Studies Chinese politics
- Strong background in causal inference and machine learning