Welcome to MIDS!

Nick Eubank

What is Data Science?

1. What (in theory) do we think Data Science should be?

What is Data Science?

1. What (in theory) do we think Data Science should be?
2. What (empirically) is Data Science?

What (in theory) should Data Science be?

What (in theory) should Data Science be?

Study of how to solve problems

What (in theory) should Data Science be?

Study of how to solve problems by answering questions

What (in theory) should Data Science be?

Study of how to solve problems by answering questions using quantitative data.

What (in theory) should Data Science be?

Study of how to solve problems by answering questions using quantitative data.

- Problem-first approach.

What (in theory) should Data Science be?

Study of how to solve problems by answering questions using quantitative data.

- Problem-first approach.
  - Data Science is an *applied* discipline!

What (in theory) should Data Science be?

Study of how to solve problems by answering questions using quantitative data.

- Problem-first approach.
  - Data Science is an *applied* discipline!
- The tool you use should be dictated by the question you seek to answer

What (empirically) is Data Science?

How did Data Science become a thing?

Over the past several decades:

1. Availability of data ↑
2. Computational power ↑

How did Data Science become a thing?

Over the past several decades:

1. Availability of data ↑
2. Computational power ↑

⇒ Huge proliferation and increase in sophistication of computational methods

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.

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
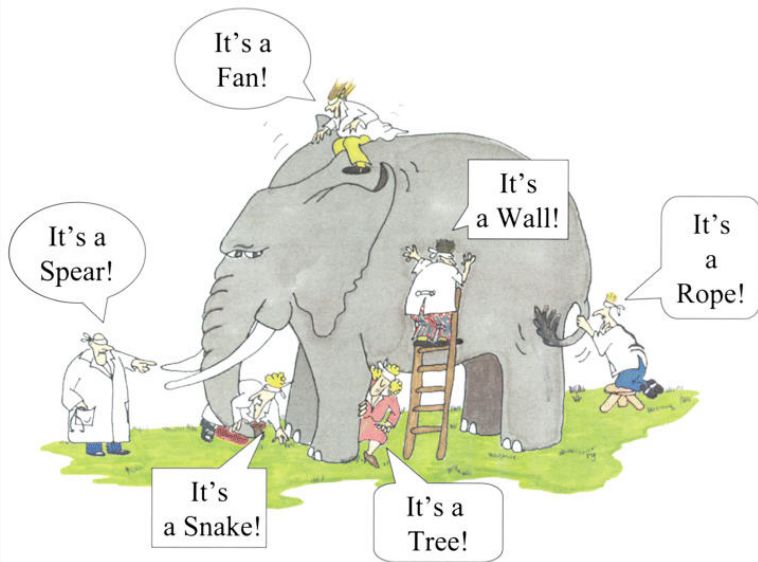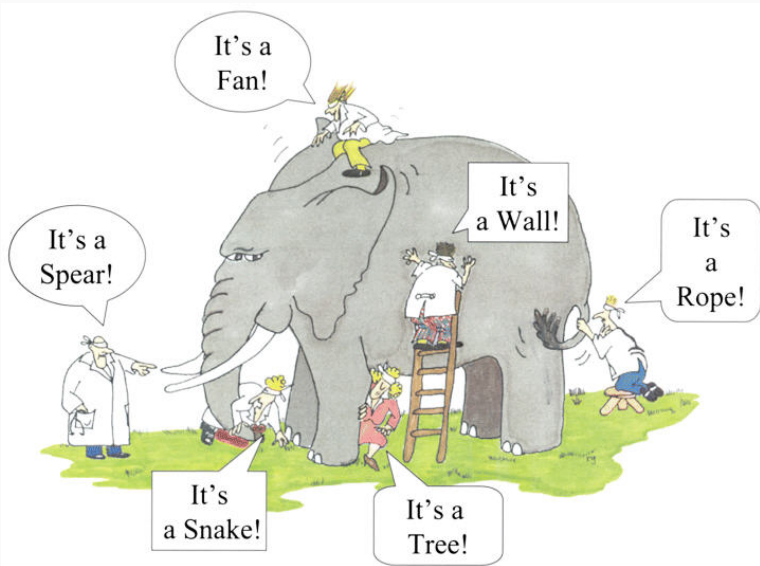
$\Rightarrow$ This is where we are *now*.

What is (empirically) Data Science?

An effort to unify the development of quantitative methods

An effort to unify the development of quantitative methods
$\rightarrow$ Recognize the elephant

Why does this matter to you?

- Many of you will have a conception of what is (and is not) data science.

Why does this matter to you?

- Many of you will have a conception of what is (and is not) data science.
  - Don't hold to those beliefs too strongly.

Why does this matter to you?

- Many of you will have a conception of what is (and is not) data science.
    - Don't hold to those beliefs too strongly.
- MIDS aims to ensure *all* our students have a solid, wide foundational understanding of the discipline.

Why does this matter to you?

- Many of you will have a conception of what is (and is not) data science.
    - Don't hold to those beliefs too strongly.
- MIDS aims to ensure *all* our students have a solid, wide foundational understanding of the discipline.
- My people who talk about data science were trained in a single silo, and may not fully recognize what's happening beyond.

Why does this matter to you?

- Many of you will have a conception of what is (and is not) data science.
  - Don't hold to those beliefs too strongly.
- MIDS aims to ensure *all* our students have a solid, wide foundational understanding of the discipline.
- My people who talk about data science were trained in a single silo, and may not fully recognize what's happening beyond.
  - That's why we have faculty from many fields.

Software Engineering DS   **Data Analysis DS**

Software Engineering DS

- Recommendation engines
- Financial trading
  algorithms
- Self-driving cars

Data Analysis DS

Areas of Data Science

Software Engineering DS

- Recommendation engines
- Financial trading algorithms
- Self-driving cars

Data Analysis DS

- Impact of policy change
- Effectiveness of health interventions
- Plan political campaigns

Areas of Data Science

Software Engineering DS

- Recommendation engines
- Financial trading algorithms
- Self-driving cars

Data Analysis DS

- Impact of policy change
- Effectiveness of health interventions
- Plan political campaigns

Nearly all data scientists will use some of both sets of skills.

Areas of Data Science

Software Engineering DS

- Recommendation engines
- Financial trading algorithms
- Self-driving cars

Data Analysis DS

- Impact of policy change
- Effectiveness of health interventions
- Plan political campaigns

Nearly all data scientists will use some of both sets of skills.
Within MIDS, you will do lots of both!

Definitely emphasizes the Software Engineering side of Data Science.

Bootcamp

Definitely emphasizes the Software Engineering side of Data Science.

- Drew and Genevieve are *extremely* talented software engineers, and will be providing a rigorous foundation for your future programming endeavors.

Bootcamp

Definitely emphasizes the Software Engineering side of Data Science.

- Drew and Genevieve are *extremely* talented software engineers, and will be providing a rigorous foundation for your future programming endeavors.
- Even if you have programmed before, please be open to what they teach.
    - LOTS of industry experience feeds into their recommendations.
    - Great opportunity to break some bad habits.

Bootcamp & LLMs

- It is a *really* bad idea to use LLMs at this stage in your education.

Bootcamp & LLMs

- It is a *really* bad idea to use LLMs at this stage in your education.
- Why? In short:

Bootcamp & LLMs

- It is a *really* bad idea to use LLMs at this stage in your education.
- Why? In short:
  - For real data science applications, LLMs are valuable but error-prone tools and require careful supervision.

Bootcamp & LLMs

- It is a *really* bad idea to use LLMs at this stage in your education.
- Why? In short:
  - For real data science applications, LLMs are valuable but error-prone tools and require careful supervision.
    - Essentially, they are like undergraduate RAs, good at drafting code, but not fully trustworthy.

Bootcamp & LLMs

- It is a *really* bad idea to use LLMs at this stage in your education.
- Why? In short:
    - For real data science applications, LLMs are valuable but error-prone tools and require careful supervision.
        - Essentially, they are like undergraduate RAs, good at drafting code, but not fully trustworthy.
    - But for beginner programming exercises, they are deceptively powerful, and so using them is like learning multiplication using a calculator — it deprives you of the opportunity to develop the intuitive "number sense" required to do advanced work.

Bootcamp & LLMs

- It is a *really* bad idea to use LLMs at this stage in your education.
- Why? In short:
  - For real data science applications, LLMs are valuable but error-prone tools and require careful supervision.
    - Essentially, they are like undergraduate RAs, good at drafting code, but not fully trustworthy.
  - But for beginner programming exercises, they are deceptively powerful, and so using them is like learning multiplication using a calculator — it deprives you of the opportunity to develop the intuitive "number sense" required to do advanced work.

If you use LLMs now, you will never develop the skills needed to supervise them effectively for real projects! So please don't cheat yourself.

You Belong Here

- We admit a very diverse set of students.

You Belong Here

- We admit a very diverse set of students.
    - By the time you graduate, all of you will have come to appreciate the different strengths and weaknesses of your classmates.

You Belong Here

- We admit a very diverse set of students.
  - By the time you graduate, all of you will have come to appreciate the different strengths and weaknesses of your classmates.
- But the MIDS curriculum front-loads foundational technical training.

You Belong Here

- We admit a very diverse set of students.
    - By the time you graduate, all of you will have come to appreciate the different strengths and weaknesses of your classmates.
- But the MIDS curriculum front-loads foundational technical training.
    - Tends to leave those with less technical backgrounds feeling out of place.

I promise, you are right where you are supposed to be.