INFO 2950: Intro to Data Science

Prof. David Mimno

Can I add this class?

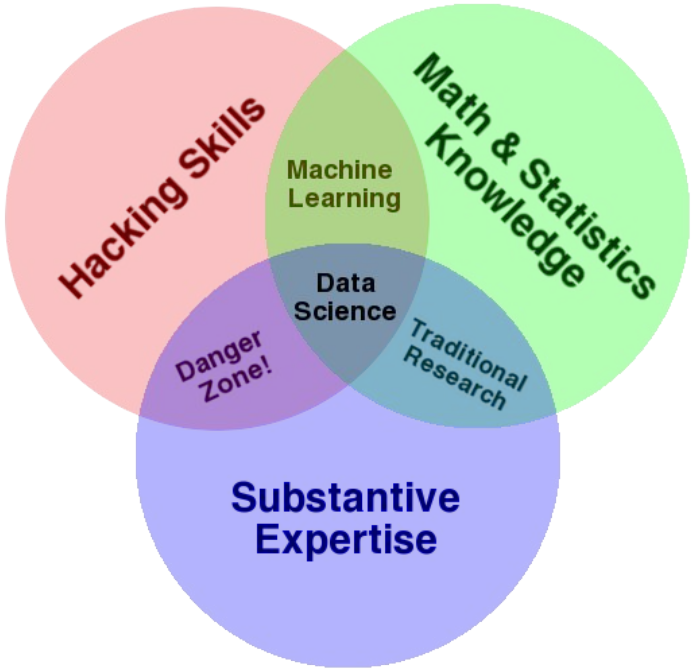
Terry Horgan (tmh233) is handling the waiting list. We expect all majors and minors to be able to enroll.

Where to find things

- Course website: http://mimno.infosci.cornell.edu/info2950
- Question answering: Campuswire.com
- Assignments: CMS (enrollment will sync in the next 24 hrs)

No book, no clickers

Drew Conway's Venn diagram



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

"Data" is an uncountable mass noun

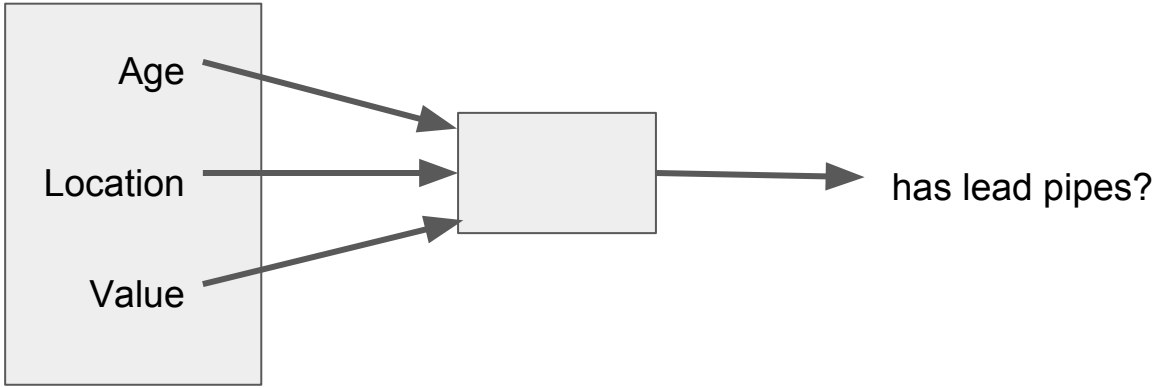
come at me, bro

Case study: Finding lead pipes in Flint, MI

Article from The Atlantic (by Alexis Madrigal): https://www.theatlantic.com/technology/archive/2019/01/how-machine-learning-found-flints-lead-pipes/578692/

Technical description: https://arxiv.org/pdf/1806.10692.pdf

A predictive model combines inputs to produce output



Data science pattern

1. Map real-world entities to a computational *representation*
2. Perform mathematical *operations* on those representations
3. Interpret *results* of those operations

Data science pattern

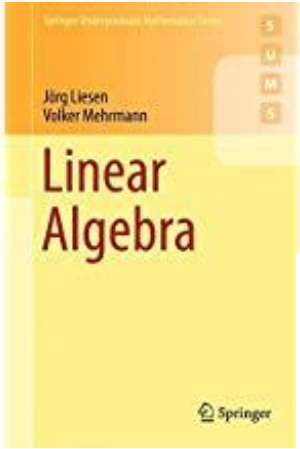
1. Map real-world entities to a computational *representation*
2. Perform mathematical *operations* on those representations
3. Interpret *results* of those operations
4. [go to step 1]

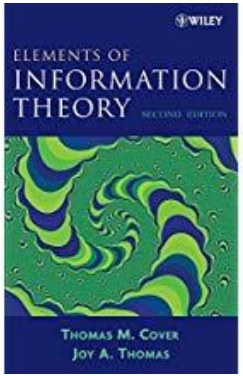
Math questions

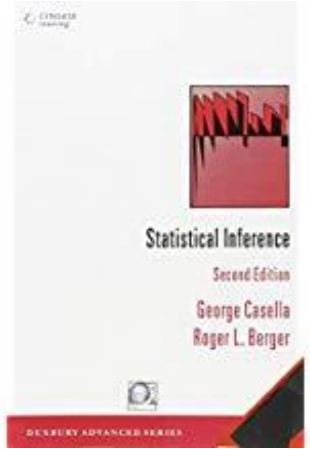
What representations are good for supporting mathematical operations?

How can we create accurate mathematica models of real-world events?

How can we convince ourselves and others that this isn't just randomness?







The math is the **easy part**

- Is the data reliable and complete?
- Are we answering the right question?
- How can we balance between what is useful and what is easily available?
- Will anyone believe that we have the right answer? Should they?

Wikipedia "Town hall meeting"

Live experiment! Find a study group

Live experiment! Find a study group

poop 💩

https://goo.gl/forms/cOflyFHdl2cUZFKI3

Where to find things

- Course website: http://mimno.infosci.cornell.edu/info2950
- Question answering: Campuswire.com
- Assignments: CMS (may not be ready yet)

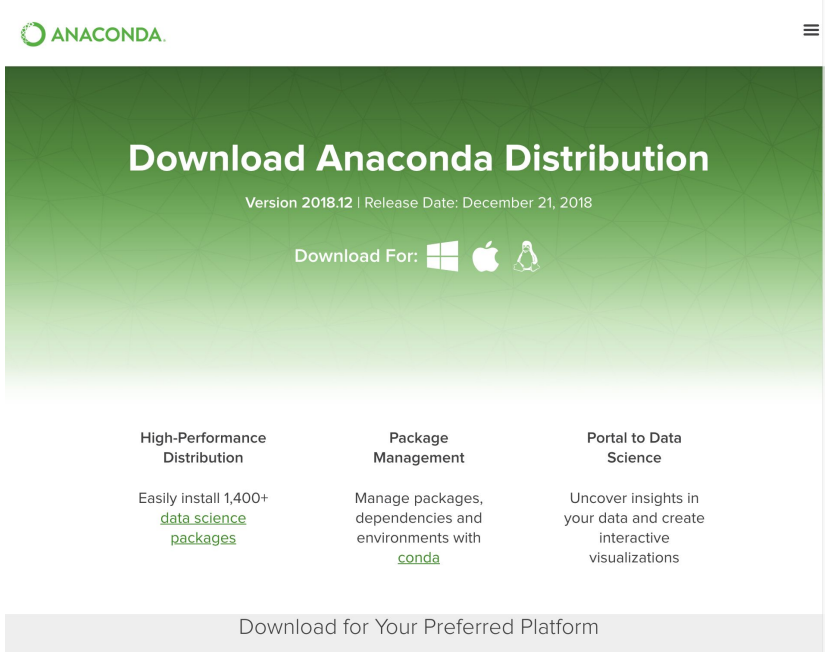
Weekly pattern

Monday	Tuesday	Wednesday	Thursday	Friday
Mimno office hours, 2-4 Gates 205	Presentation of new material	Homework due 11:59pm	Presentation of new material	Lab sessions: practice and discuss

For Friday: Install Python 3

- Anaconda is the easiest, most reliable installation: https://anaconda.com/download
- NO PYTHON 2.
  - To check: type `print "hello"` with no (parentheses). You should get an error.

We will work in notebooks, scripts, and the command line (>>>)



Can I add this class?

Terry Horgan (tmh233) is handling the waiting list. We expect all majors and minors to be able to enroll.