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

Hacking Skills

Math & Statistics Knowledge

Substantive Expertise

Machine Learning

Data Science

Traditional Research

Danger Zone!
"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):

A predictive model combines inputs to produce output

Age
Location
Value

has lead pipes?
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**

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mimno office hours, 2-4 Gates 205</td>
<td>Presentation of new material</td>
<td>Homework due 11:59pm</td>
<td>Presentation of new material</td>
<td>Lab sessions: practice and discuss</td>
</tr>
</tbody>
</table>
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.