DISCOVERY

#1: Welcome to Data Science Discovery

January 14, 2019

Karle Flanagan and Wade Fagen-Ulmschneider

Welcome to STAT/CS/IS 107: "Data Science Discovery"

Course Website: http://courses.las.illinois.edu/spring2019/stat107/
...or just Google Search: `stat107 uiuc`

Description: Data Science Discovery is the intersection of statistics, computation, and real-world relevance. As a project-driven course, students perform hands-on-analysis of real-world datasets to analyze and discover the impact of the data. Throughout each experience, students reflect on the social issues surrounding data analysis such as privacy and design.

Instructors:

Karle Flanagan <kflan@>
Instructor of Statistics, College of Liberal Arts and Science

Wade Fagen-Ulmschneider <waf@>
Teaching Assistant Professor of Computer Science, College of Engineering

---

Coursework and Grading

A total of 1,000 points are available in Data Science Discovery, along with many opportunities to earn extra credit (capped at +107 points). The points are broken down in the following way:

- **40 points**: Participation (42 × 1 point, up to 42 pts)
  - +1 /lecture, points beyond 40 are extra credit!
- **105 points**: Lab Sections (10 × 15 points, up to 130 pts)
  - Points over 105 are extra credit!
- **150 points**: Homework Assignments (3 × 50 points)
- **300 points**: Projects (75, 100, and 125 points)
- **405 points**: Midterm and Final (135 and 270 points)
  - Midterm Exam (March 4, 2019, computer-based; 140 points)
  - Final Exam (Final Exam Week, computer-based; 280 points)
- **Extra Credit:**
  - Lots of opportunities for +X points
  - *Extra credit is capped at +107 points from all sources!*

Final Course Grades

Your final course grade is determined by the number of points you earned during the semester:

Points	Grade	Points	Grade	Points	Grade
[1070, ∞)	A+	[930, 1070)	A	[900, 930)	A-
[870, 900)	B+	[830, 870)	B	[800, 830)	B-
[770, 800)	C+	[730, 770)	C	[700, 730)	C-
[670, 700)	D+	[630, 670)	D	[600, 630)	D-
		(600, 0]	F		

We never curve individual exam or assignment scores; instead, if necessary, we may lower the points required for each grade cutoff to be lower than the stated cutoff. In no case will we raise the cutoff.

Lab Sections: As part of registering for Data Science Discovery, you have also registered for a lab section. **Labs start this week** and are held by your contact TA – your lab TA is the first person who you should contact if you're having any trouble with the course!

Contacting Us: We want to do everything we can to make sure you have a great time learning data science. There are a lot of ways to reach out if you get stuck:

1. **Course Piazza Site**: Online Q&A discussion forum to ask and answers from your peers and course staff. Join it!
   URL: http://piazza.com/illinois/spring2019/stat107

2. **Course Staff**: There's a whole team of people here to help you succeed and do amazing things:
   - Teaching Assistant (TA): Wenjing Yin,
   - Course Assistants (CAs): Ishani Desai, Tina Abraham, and Kelly Mack

3. **Office Hours**: Course staff will hold office hours each week to help you if you're stuck! Times and details to be announced.
   Location: 23 Illini Hall (basement)
   Times: TBA

Does the death penalty have a deterrent effect?
Is chocolate good for you?
What causes breast cancer?

All of these questions attempt to assign a cause to an effect. A careful examination of data can help shed light on questions like these. In this section, you will learn some of the fundamental concepts involved in designing experiments.

Experimental Design
Statistical studies or experiments are done to see if a _____ has an effect. This effect is called the _____.

Examples:

Experiment vs Observational Study
Consider a study based on the point of view of the researcher (you!) designing the study:

- If the researcher decides who gets the treatment, the investigation is called an _____.

- If the researcher just studies the effects of the treatment but plays no part in choosing who gets the treatment then the investigation is called an _____.

If the experiment compares the effects of getting the treatment to not getting the treatment, it's called a **controlled experiment**. In a controlled experiment the researcher decides who gets the treatment (treatment group) and then compares their responses to the responses of those who didn't get the treatment (control group).

The fundamental question is whether the treatment has an effect on the outcome. Any relation between the treatment and the outcome is called an **association (or correlation)**. If the treatment causes the outcome to occur, then the association is **causal**.

Your goal when designing experiments is to ensure that the treatment and control groups are AS ALIKE AS POSSIBLE.

Visualizing Experimental Design:

Data Science Discovery – Things To Be Doing:

1. Join the Data Science Discovery Piazza
2. Visit our course website!
3. Attend your lab section (lab sections start **tomorrow!**)
4. Attend lecture (every Monday, Wednesday, and Friday)
5. HW1 is released next week