

## Syllabus :: Data Science Discovery

### Course Instructors

**Karle Flanagan (kflan)**, Instructor, Department of Statistics

Office: 110 Illini Hall

**Wade Fagen-Ulmschneider (waf)**, Teaching Asst. Professor, Department of Computer Science

Office: 2215 Siebel Center for Computer Science

### Course Description

Data Science Discovery is the intersection of inferential thinking, computational thinking, and real-world relevance. This course empowers a student to analyze data arising from real-world events to understand the phenomenon. Throughout the semester, students will have hands-on-analysis of real-world datasets. In each experience, students reflect on the social issues surrounding data analysis such as privacy and design.

Four (4) credit hours. Prerequisites: None.

### Topics Covered

Data Science	Inferential Thinking	Computational Thinking
<ul style="list-style-type: none"><li>- Introduction to Data Science</li><li>- Visualizations</li><li>- A/B Testing</li><li>- Testing Hypotheses</li><li>- Confidence intervals</li><li>- Case Studies</li></ul>	<ul style="list-style-type: none"><li>- Causality and Experiments</li><li>- Randomness</li><li>- Sampling and Empirical Distribution</li><li>- Comparing two Samples</li><li>- Estimation</li><li>- Why the Mean Matters</li><li>- Prediction</li><li>- Inference for Regression</li><li>- Classification</li><li>- Updating Predictions</li></ul>	<ul style="list-style-type: none"><li>- Programming in Python</li><li>- Data Types</li><li>- Sequences</li><li>- Tables</li><li>- Functions</li><li>- Visualizations</li></ul>

### Course Sections

**Lecture:** Lecture three times a week (M/W/F), 50 minutes each lecture.

**Discussion Section:** Weekly conceptual and problem-solving discussion sections lead by course staff. Not computer-based. *(50 minutes /week)*

**Lab Sections:** Weekly computer-based programming sections lead by course staff. Some sections will be BYOD (“Bring Your Own Device”) and others will be in computer labs for students who do not have or choose not to use their own device. *(50 minutes /week)*

### Course Materials

1. **Required Text (FREE):** “The Foundations of Data Science” by Ani Adhikari and John DeNero
2. **Required Calculator:** Any non-programmable calculator is accepted (no phones, graphing calculators, etc.)

## Course Grades

Course grades are given in points, totaling 1,000 points throughout the semester. The breakdown of points is as follows:

Category	Total Points /1,000 points
Participation	10 points
Labs	100 points (10 x 10 points)
Homework	150 points (3 x 50 points)
Projects	300 points (3 x 100 points)
Midterm Exam	140 points
Final Exam	300 points

Course points will be translated into a course grade at the end of the semester:

A+: [970, 1000+)	A: [930, 970)	A-: [900, 930)
B+: [870, 900)	B: [830, 870)	B-: [800, 830)
C+: [770, 800)	C: [730, 770)	C-: [700, 730)
D+: [670, 700)	D: [630, 670)	D-: [600, 630)
	F: [0, 600)	

## Late Submission

No late submissions are accepted. However, your lowest homework assignment and your lowest two (2) lab sections will be dropped.

## Learning Collaboratively

Data Science is a collaborative science. We strongly encourage you to discuss all of your course activities, with the obvious exception of exams, with your friends and classmates. You will learn more through talking through the problems, teaching others, and sharing ideas. Do not try to tackle this course alone.

## Academic Integrity

Collaboration is about working **together**. Collaboration is **not** giving the direct answer to a friend or sharing the source code to an assignment. Collaboration requires you to make a serious attempt at every assignment and discuss your ideas and doubts with others so everyone gets more out of the discussion. Your answers **must** be your own words and your code must be typed (not copied/pasted) by you.

Academic dishonesty is taken very seriously in STAT 107 and all cases will be brought to the University, your college, and your department. You should understand how academic integrity applies specifically to STAT 107: the sanctions for cheating on an assignment includes a loss of all points for the assignment **and** that the final course grade is lowered by one whole letter grade (100 points). A second incident, or cheating on an exam, results in an automatic F in the course.

Academic integrity includes protecting your work. If your work ends up submitted by someone else, we have considered this a violation of academic integrity just as though you submitted someone else's work.

## Course Outline

Lecture	DoW	Topic	Reading	Discussion	Assignments
1	1-*	Introduction	1.1, 1.2, 1.3	Introduction	Lab 1
2	1-*	Cause and Effect	2		
3	X-*	Tables	3	Data Handling	Lab 2
4	2-W	Data Types	4, 5		
5	2-F	Building Tables	6.1, 6.2	Data Privacy	Lab 3
6	3-M	Charts; Census	7, 7.1, 6.3, 6.4		
7	3-W	Census	6.3, 6.4	Functions and Groups	Lab 4
8	3-F	Histograms	7.2, 7.3		
9	4-M	Functions	8, 8.1	Project 1	Proj. 1 Checkpoint
10	4-W	Groups	8.2, 8.3		
11	4-F	Joins	8.4	Result Validation	Lab 5
12	5-M	Table Examples	8.5		
13	5-W	Iteration	9, 9.1, 9.2, 9.3	Causality	Lab 6
14	5-F	Chance	9.4, 9.5		
15	6-M	Sampling	10, 10.1, 10.2	Project 2	Proj. 2 Checkpoint
16	6-W	Models	10.3		
17	6-F	Comparing Distributions	11.1, 11.2	Midterm Review	Proj. Due
18	7-M	Decisions and Uncertainty	11.3		
19	7-W	A/B Testing	12.1, 12.2	Midterm Reflection	Lab 7
20	7-F	Causality	12.3		
21	8-M	Examples	12.2	Project 3	Proj. 3 Checkpoint
24	8-W	Confidence intervals	13, 13.1, 13.2		
24	8-F	Interpreting Confidence	13.3, 13.4	Reporting Statistical Error	Lab 8
25	9-M	<b>Midterm Exam</b>			
26	9-W	Center and Spread	14, 14.1, 14.2	Impact of Classifiers	Lab 9
27	9-F	The Normal Distribution	14.3, 14.4		
28	10-M	Sample Means	14.5	Project 3	Proj. Due
29	10-W	Designing Experiments	14.6		
30	10-F	Correlation	15, 15.1	Open Discussion on Cases	Lab 10
31	11-M	Linear Regression	15.2		
32	11-W	Least Squares	15.3, 15.4	Project 3	Proj. 3 Checkpoint
33	11-F	Residuals	15.5, 15.6		
34	12-M	Regression Inference	16	Open Discussion on Cases	Lab 10
35	12-W	Classification	17.1, 17.2, 17.3		
36	12-F	Classifiers	17.4	Project 3	Proj. Due
37	13-M	Decisions	17		
38	13-W	Privacy		Open Discussion on Cases	Lab 10
39	13-F	Case Studies	Supplement		
40	14-M	Case Studies	Supplement	Project 3	Proj. Due
41	14-W	Case Studies	Supplement		
42	14-F	Review			
		<b>Final Exam</b>			

### Disability Accommodations

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak Street, Champaign, call 333-4603 (V/TTY), or email a message to [disability@illinois.edu](mailto:disability@illinois.edu).

### Course Logistics

The main lectures (M/W/F) will be delivered by the subject-expert for the specific area of focus. For topics primarily from statistics, lectures will be delivered by Karle Flanagan. For topics primarily from computer science, lectures will be delivered by Wade Fagen-Ulmschneider.

Each week, you will attend a two-hour lab/discussion section that is led by a Teaching Assistant (TA). Your lab TA will be **your TA** and is the best first-point-of-contact with the course. One or more Course Aides (CA) will also be there to help out in labs!

### Office Hours

Open Office Hours will be held weekly by each member of the course staff and will be announced at the beginning of the semester.

### Final Exam

The final exam for STAT 107 will be given at the University-scheduled time based on your lecture section. The time and location will be announced in class.

### Pilot Section (Spring 2019 ONLY)

The course a pilot course for STAT 107. The pilot will be offered as STAT 430 and requires enrollment in a STAT 199 lab section. Your grade in STAT 430 will depend on both your performance in STAT 430 and your work completed as part of the associated STAT 199 lab section.

Instructor approval is required for STAT 430. Approval will only be given for enrollments that enroll in BOTH the lecture (STAT 430) and the lab section (STAT 199).