DATA SCIENCE DISCOVERY
University of Illinois Urbana-Champaign

Syllabus for STAT/CS/IS 107

Data Science DISCOVERY is offered at the University of Illinois as STAT/CS/IS 107. The content of this website (https://discovery.cs.illinois.edu/) is used as the pre-lecture materials for the lecture where we do interactive activities, practice problems, and nerd out with data science.

The syllabus here provides the structure of the course, schedule, grades, course policies, and the overview of all assignments, labs, and exams. Information about your teaching assistants (TA), office hours, lab sections, discussion forums, and other resources available to students enrolled in STAT 107 is available on the Canvas page:

- Fall 2025 Canvas Page for STAT/CS/IS 107

Course Overview

Course Description: Data Science DISCOVERY or simply "DISCOVERY" is UIUC's largest introductory data science course. This course focuses on the intersection of inferential thinking using statistics, computation using industry standard tools like Python and git, and real-world relevance. As a data driven course, students perform hands-on analysis of real world datasets to analyze and discover the impact of data science. Throughout each experience, students reflect on the social issues surrounding data analysis and practice communicating their results.

Lecture Section: DISCOVERY has synchronous in-person lectures that meet every Monday, Wednesday, and Friday! You must be enrolled in one lecture section. Lecture sections

are listed as the "Lecture" sections under STAT 107 in Course Explorer.

Lab Section: DISCOVERY has synchronous in-person lab sections that meet once each week. Lab sections are led by graduate student TAs and focus on small group programming exercises and discussion. Sections are "Bring Your Own Device" and participation in your lab section makes up a third of your lab score each week. The labs are 80 minutes long and capped at 30 students each. You must attend the lab section you are officially registered for to earn credit. Lab sections are listed as the "Laboratory-Discussion" sections under STAT 107 in Course Explorer.

Prerequisites: None
General Education Credit at UIUC: Quantitative Reasoning I (QR1)
Credit Hours: 4 credit hours (3 lecture hours, 1 lab hour)
Proficiency Exam: A proficiency exam is offered at the beginning of each semester. Details about the proficiency exam can be found at
https://discovery.cs.illinois.edu/proficiency-exam/

Course Contact

DISCOVERY is co-taught by two professors from two different departments:

- Dr. Karle Flanagan, Department of Statistics, (kflan@illinois.edu)
- Dr. Wade Fagen-Ulmschneider, Siebel School of Computing and Data Science, (waf@illinois.edu)
- *When contacting us, please include* **BOTH OF US** *in one email!*

Lab TA: Your primary point of contact with the course staff will be your Lab Teaching Assistant (TA) based on the lab you

are registered for in DISCOVERY. The TA contact information by lab section for Fall 2025 is found on Canvas.

Course Discussion Forum: Discord is our class discussion forum. All of the course staff (instructors, TAs, and CAs) will be on Discord to answer any questions you have related to the course. Access to the Discord will be given to you via Canvas.

Even as an online and more casual space, Discord is an extension of the classroom and we expect you to be kind, respect each other, and uphold the student code. Please do not DM or "@" any course staff, and do not share screenshots of solution code. Instead, Discord is a community to help people learn how to master the material themselves. The Discord "Join Link" for Fall 2025 is found on Canvas.

Course Content

In DISCOVERY, there are 3 websites we use regularly:

1. **The DISCOVERY Website** (https://discovery.cs.illinois.edu/): This website contains many useful resources for students, including data science guides. We will reference it frequently throughout the semester.
2. **Canvas** (https://canvas.illinois.edu/): We use Canvas as a grade book in DISCOVERY! Grades are updated in Canvas each week on Mondays.
3. **Mastery Platform** (https://mastery.cs.illinois.edu/): All homework and CBTF exams will be done using Illinois Mastery Platform.

Required Course Materials

Laptop Computer: You need a laptop running Windows, Mac OS, or Linux. Android Tablets, Chromebooks, and iPads are

not supported. You will need to be able to install both Python and git to complete the labs (instructions provided).

Course Notebook: DISCOVERY has a fill in the blank notebook that you can purchase from the Illini Union Bookstore (IUB) for under $40. This includes all of the course notes for the semester! You will use this notebook during each lecture. Filling out the entire notebook will earn you extra credit at the end of the semester.

Grades

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

- **Labs**: 140 points (14 × 10 points)
- **Homework**: 150 points (33 total, 30 × 5 points and lowest 3 dropped)
- **Projects**: 90 points
- **Participation**: 60 points (maximum of 36/40 points from lectures and 24/28 points from labs, 2pts /lab)
- **Midterm Exam 1**: 140 points (Computer-based, in the CBTF)
- **Midterm Exam 2**: 140 points (Computer-based, in the CBTF)
- **Comprehensive Final Exam**: 280 points (Computer-based, in the CBTF)
  = 1,000 total points

Grade Letter Cutoffs

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

Points Earned	Minimum Grade	Points Earned	Minimum Grade	Points Earned	Minimum Grade
Exceptional	A+	[930, 1000+)	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		

It is the course policy that we will never "curve" any individual assignment. If needed, we may lower these cutoffs but will never raise the cutoffs. This means earning 930 points will always guarantee an "A" in the course. We have not lowered any cutoffs in any recent semester.

Extra Credit (EC)

There are multiple opportunities for extra credit in DISCOVERY (usually called "+1 EC points"). Points for extra credit work will be assigned after grade cutoffs are determined, so they are a true bonus to your score. The total amount of extra credit you can earn varies each semester. We will announce extra credit opportunities throughout the semester! Some of them will include:

- **Surveys**: These will be done using Google Forms with your @illinois.edu email address.
- **MicroProjects**: Each week, there will be a MicroProject available for you to complete for +1 EC. These are

designed for you to do "real data science in under an hour".

- **Notebook**: During the last week of class, if your entire notebook is filled out from lecture this semester and you've answered at least 70% of the Clicker questions showing you went to lecture, you will earn several EC points. This is usually the largest single opportunity for EC in the course.
  ...and more!

Course Assignments

Labs: Labs will be completed each week in your lab section. Labs are due Mondays at 11:59pm. There will be a lab every week, even during exam weeks.

Homeworks (HWs): In general, there will be one short mastery-based homework assignment for each lecture. Homeworks will be due at 11:59pm on Monday, Wednesday, or Friday (homeworks will be due the following lecture after it's assigned; ex: if it's assigned on Monday, it will be due Wednesday at 11:59pm).

Projects: There will be 3 projects throughout the semester, which are longer and more involved than the weekly labs. You will have at least one week (and usually more) to complete the project. The projects are an opportunity to apply everything you've learned in DISCOVERY and explore your individual interests and passions.

Exams: There will be 2 midterm exams and a cumulative final. Exams will be taken in the Computer Based Testing Facility (CBTF). See https://cbtf.engr.illinois.edu/ for details.

Course Calendar

(Note: For specific due dates for Fall 2025, make sure to check the Fall 2025 calendar on Canvas.)

Lecture	Topics	Assignment

Module 1: Basics of Data Science with Python

Lecture	Topics	Assignment
#1	What is Data Science? and Types of Data	Homework #1: Party Words Survey
#2	Experimental Design and Blocking and Python for Data Science: Introduction to DataFrames	Homework #2: Mastery Platform (m1-02)
#3	Row Selection with DataFrames	Homework #3: Mastery Platform (m1-03)
#4	Observational Studies, Confounders, and Stratification	Homework #4: Mastery Platform (m1-04)
#5	Simpson's Paradox	Homework #5: Mastery Platform (m1-05)
#6	DataFrames with Conditionals	Homework #6: Mastery Platform (m1-06/07)
#7	Software Version Control with git and Exploratory Data Analysis (EDA) Overview	Homework #7: Mastery Platform (Module 1 Review)

Module 2: Exploratory Data Analysis

Lecture	Topics	Assignment
#8	Descriptive Statistics	Homework #8: Mastery Platform (m2-02)

Lecture	Topics	Assignment
#9	Grouping Data in Python	Homework #9: Mastery Platform (m2-03)
#10	Histograms and Quartiles and Box Plots	Homework #10: Mastery Platform (m2-04/05)
#11	Basic Data Visualization in Python	Homework #11: Practice Exam #1

Module 3: Simulation and Distributions

Lecture	Topics	Assignment
#12	Overview of Simulation and Random Numbers in Python	Homework #12: Mastery Platform (m3-01/02)
#13	For-Loops in Python and Simple Simulations in Python	No Homework (Midterm 1 Upcoming!)
#14	No Lecture: Midterm #1 (Covers Module #1 and Module #2)	—
#15	Sample Space amd Conditionals in Python	Homework #13: Mastery Platform (m3-03/04)
#16	Day #1 of Functions in Python and Project #1 Released	Homework #14: Mastery Platform (m3-05)
#17	Day #2 of Functions in Python	Homework #15: Mastery Platform (m3-06)

Lecture	Topics	Assignment
#18	Normal Distribution	Homework #16: Mastery Platform (m3-07)
#19	Law of Large Numbers	Homework #17: Mastery Platform (m3-08)

Module 4: Prediction and Probability

Lecture	Topics	Assignment
#20	Probability Introduction and The Monty Hall Problem	Homework #18: Mastery Platform (Module 3 Review)
#21	Multi-event Probability: Multiplication Rule	Homework #19: Mastery Platform (m4-03)
#22	Multi-event Probability: Addition Rule	Homework #20: Mastery Platform (m4-04)
#23	Conditional Probability	Homework #21: Mastery Platform (m4-05)
#24	Bayes' Theorem	Homework #22: Mastery Platform (m4-06)
#25	Review/Complex Examples: Conditional Probability and Bayes' Theorem	Homework #23: Midterm #2 Review

Module 5: Polling, Confidence Intervals, and the Normal Distribution

Lecture	Topics	Assignment
#26	Random Variables	No Homework (Project #1 Due Soon!)
#27	Bernoulli and Binomial Random Variables	No Homework (Project #1 Due!)
#28	Beyond DISCOVERY: STAT 207, CS 277, CS 307 and more!	No Homework (Midterm #2 Upcoming!)
#29	No Lecture: Midterm #2 (Covers Module #3 and #4)	—
#30	Python Functions for Random Distributions	Homework #24: Mastery Platform (m5-02/03)
#31	Central Limit Theorem	Homework #25: Mastery Platform (m5-04)
#32	Polling and Sampling	Homework #26: Mastery Platform (m5-05)
#33	Confidence Intervals	Homework #27: Mastery Platform (m5-06)
#34	Hypothesis Testing	Homework #28: Mastery Platform (m5-07)

Module 6: Towards Machine Learning

Lecture	Topics	Assignment
#35	Overview of Machine Learning	Homework #29: Mastery Platform (Module 5 Review)
#36	Correlation	Homework #30: Mastery Platform (m6-02)
#37	Linear Regression	Homework #31: Mastery Platform (m6-03)
#38	Machine Learning Models in Python with sk-learn	No Homework (Project #2 Due Soon!)
#39	Clustering	No Homework (Project #2 Due Today!)
#40	Towards Machine Learning in Python	Homework #32: Mastery Platform (m6-04/05)
#41	Last Day of Lecture: Topics in Data Science	Homework #33: Final Exam Review
—	Comprehensive Final Exam	—

Office Hours and Getting Help

Open office hours are fantastic for getting help with understanding course concepts assignments, debugging your code, and more! All open office hours will have multiple TAs and/or CAs available to help you out! We have open office hours **every single weekday** and Prof. Wade and Prof. Karle

will be available at specific times each week at office hours. We will also have online office hours available each week.

[See the "Office Hours" section in the Canvas syllabus for Fall 2025 office hour times and locations.](#)

Course Policies

Attendance: Attendance is strongly recommended for lecture but is not required. You must answer at least 70% of the Clicker questions to get extra credit for the notebook at the end of the semester so keep that in mind! Attendance in your lab section makes up one third of the points of your total lab grade. If you cannot come to your lab due to illness, you can email your lab TA before your lab starts to discuss an alternative.

Late Work: No late submissions are accepted. However, we do drop your 3 lowest homework assignments at the end of the semester. So missing 3 homework assignments won't hurt your grade. Late projects and late labs will not be accepted.

Rounding Grades/Curves: Due to the large amount of extra credit offered in DISCOVERY, we do not round grades or have curves on exams.

DRES Accommodations: We are happy to offer accommodations for disabilities verified through DRES (http://www.disability.illinois.edu/). Please email us a copy of your DRES letter during Week 1. Since all of our exams will be in the CBTF, you can request extra time through this link: https://cbtf.illinois.edu/students/dres. If you have any other questions or need any other accommodations, don't hesitate to reach out! :)

Learning Collaboratively: Data Science is a collaborative science. Do not try to tackle this course alone. We strongly

encourage you to discuss all of your course activities (with the exception of exams) with your friends and classmates! You will learn more by talking through the problems, teaching others, and sharing ideas.

Continue to read on "Academic Integrity" to understand the difference between collaboration and giving an answer away.

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, the loss of all extra credit in STAT 107, and that the final course grade is lowered by one whole letter grade (100 points). A second incident, or any 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.

On This Page

4. Required Course Materials

5. Grades

    i. Grade Letter Cutoffs

    ii. Extra Credit (EC)

6. Course Assignments

    i. Course Calendar

7. Office Hours and Getting Help

8. Course Policies

Data Science Discovery is an open-source data science resource created by The University of Illinois with support from The Discovery Partners Institute, the College of Liberal Arts and Sciences, and The Grainger College of Engineering. The aim is to support basic data science literacy to all through clear, understandable lessons, real-world examples, and support.