AST 200: Intro to Data Science (with Astronomical Applications)
Spring 2019

Instructor: Dr. Kate Follette
Office: NSC A016
Lab: NSC A120
Office Phone: (413) 542-5938
Office Hours: Fridays 2-3pm in NSC A126 or by appointment

Course Teaching Assistants:
Mark Schoen (mschoen18@amherst.edu)
Clare Leonard (cleonard19@amherst.edu)
TA Help Sessions: Wednesdays 2-4pm in NSC A126 or by appointment

Class Meetings: 1-1:50pm MWF, 1-4pm Th
Location: NSC A126

Course Description
The purpose of this course is to introduce computational, statistical and data visualization techniques that will allow students to excel in further coursework in astronomy and other STEM majors. Students will be introduced to how to use the Python programming language to analyze and manipulate data; how to create, interpret, and present visualizations of those data; and how to apply statistical analysis techniques to data. We will sharpen these skills through the lens of astronomical data collection and analysis, though the skills themselves are applicable in many other fields.

Recommended (Pre or Co)Requisites: ASTR 111 or 112 and COSC 111

Required Texts*
The Visual Display of Quantitative Information by Tufte
Practical Statistics for Astronomers by Wall & Jenkins
OpenIntro Statistics by Diez, Barr & Cetinkaya-Rundel
**If purchase of these texts are outside of your financial means for whatever reason, please let me know right away*

Supplemental Texts
Available on hold in the library and in the Follette Research Lab (NSC A120)
Beautiful Evidence by Tufte
Envisioning Information by Tufte
Visual Explanations by Tufte
Demon Haunted World: Science as a Candle in the Dark by Sagan
An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements by Taylor

Goals and Expectations

Course Learning Goals

By the end of this course, students will be able to:

- Write efficient, well-commented python programs for analysis and visualization of (astronomical) data
- Interpret, compare, critique, construct, and refine a variety of types of (astronomical) data visualizations according to ethical and aesthetic graphical design principles
- Apply descriptive statistics, hypothesis testing, regression, and probabilistic statistical analyses to (astronomical) data and use the results to support data-driven arguments
- Spot poorly designed and misleading data visualizations, improperly applied statistics, and unsubstantiated arguments regarding (astronomical) data

What I Expect from You

1. Much of the learning in this course will happen in real time during class. I expect that you'll come to ***every*** class on time and prepared to actively engage with the material and your classmates (except in the case of a serious illness or emergency) and with assigned reading and prelabs completed.
2. That you respect both your instructor and your fellow students by listening when others are speaking.
3. That you work collaboratively with your classmates and aim to contribute equivalently (if not equally) to group discussions and assignments. At the same time, I expect you to follow the expectations laid out for individual work and abide by the college's honor code at all times.
4. That you participate actively and conscientiously. If you come to this class with a lot of prior knowledge about the subject, please take a step back and try not to overparticipate in class discussions. If you are less comfortable with the course material or format, challenge yourself to actively participate. ***Active participation may take many forms***. For example:
    - Asking a clarifying question of me or a classmate
    - Explaining a concept to a classmate
    - Speaking up in a large or small group discussion
    - Answering a question that I ask the class

What You Can Expect from Me

1. I am here to help you learn. I will do my best to facilitate learning by providing opportunities for you to engage with the course material authentically, being open to and available for questions, and eliciting feedback on course content and structure regularly[1].

[1] Throughout the course, I will be seeking your feedback through periodic anonymous surveys. A welcoming classroom environment in which all students feel supported in their learning is very important to me. Recognizing that you each bring insights and experiences to the classroom that are different from my own, *I invite your feedback regarding course atmosphere at any point during the course* and in whatever form you feel most comfortable (e-mail, in person conversation, anonymous note). We all occasionally say things without thinking, and it is very important to me that if I or a classmate ever use an example/analogy or make a comment that feels isolating or derogatory to you, that you feel comfortable telling me so that I can make it right.

2. That I will clearly communicate my expectations for acceptable coursework, and will grade fairly and anonymously
3. That I will be present, prepared and engaged at every class and office hour session (except in the case of serious illness, emergency, or unavoidable travel).

Course Outline and Calendar
See attached

Flipped Class Structure
This class will be operating with a "flipped" structure. Rather than being introduced to the concepts in class and left to struggle with them on your own on problem sets, you will be asked to learn new ideas outside of class by reading from various texts and we will spend the majority of our in-class time working through problems and examples together. You should think of class sessions as an opportunity for you to interrogate, clarify, and explore course material that you were introduced to in the reading. We will spend only a brief time reviewing and clarifying the material from the reading in class and will focus mainly on working through problems and examples. A typical class might look like:

10-20min for review, clarification, examples and questions from the reading(s)
20min group problem session
10min for groups to present solutions

Thursday labs**,** like the lectures, are designed so that you encounter the toughest computational problems during the lab period where you can get immediate help and feedback and are introduced to the relevant computational techniques during the prelab.

Friday discussion sections will be reserved for discussions of data visualization concepts, which we will be learning about in parallel with the other course material. Most of the readings will be from the required course text *The Visual Display of Quantitative Information*, with select readings from Edward Tufte's other texts.

Grading
The flipped structure of this course means that the onus is on you to keep on top of the material, but there will be some accountability (worth a total of 35% of your grade) built in to help keep you on track. Namely:

1) ***Periodic Reading Quizzes (5%)*** will be administered at the start of a M/W class approximately biweekly. Generally they will involve between 1 and 3 problems from the course text or will be closely related to problems from the text.
2) ***Periodic Coding Quizzes (5%)*** will be administered at the start of lab approximately biweekly. These are quick closed note quizzes that will involve your writing a small piece of code. They are very low stakes, and will be graded leniently. Please look at them as an opportunity to test your mastery of the material so that you know what you

need to work on rather than as something that will substantially impact your grade. Generally they will involve a short exercise that looks very similar to something that you were asked to do in the prelab or the lab the previous week.

3) ***Annotation of Course Readings in Perusall (10%)*** will be required before each M/W class. At a minimum, you must:
    a. Make at least two substantive comments on the assigned reading.
    b. Pose at least one question about the reading
    c. Answer or comment on one of your peers' questions
    d. Attempt one new problem from the end of the chapter and post your solution in the text at the spot where it is most relevant (rather than at the end of the chapter). Each member of the class should attempt a different problem.
    e. Comment on one of your peers' problem solutions

    Grading of these annotations is done automatically by Perusall based on the substance of what you write, but I will review and override them as necessary.
4) ***Prelabs (10%)*** will be due at the start of each Thursday lab. Prelabs will introduce you to the basic computational syntax needed to complete the lab for a given week and are designed to be completed in ~1-2hr. Coding is not something that you can learn to do just by reading, so the prelabs will generally also include a couple of simpler exercises to allow you to begin applying the ideas on your own. It is these exercises that will be graded.
5) ***Pre-Discussion Reading Assignments (5%).*** Except for the first two discussions, these assignments will be designed by your classmates and will likely take a number of different forms. They might involve annotating a reading in Perusall, finding an example of something discussed in the reading to bring to class, redesigning a plot, etc. Their purpose is to prepare you to apply and expand on the ideas from the reading during the in-class discussion.

The remaining 65% of your course grade is comprised of:

1) ***Discussion Leading (10%).*** You will be asked to lead one Friday discussion during the semester together with 1-3 groupmates. I will lead the first two discussions and will demonstrate several potential formats. You and your groupmates will decide the format for your assigned discussion and will lead the class on your assigned day. You will also design a brief assignment to accompany the reading and designed to prepare your classmates for the discussion you've planned. You and your groupmates will meet with me one week ahead of your assigned discussion. You should have already thoroughly read the relevant text(s), developed a draft of the reading assignment, and developed a proposed discussion format before meeting with me.
2) ***Participation (10%)****.* Engaged, respectful communication with your peers is <u>crucial</u> to the success of this course. In particular, explaining your thinking and listening to others' explanations in turn is extremely important. If you arrive to class on time, are well-prepared, listen, are respectful, and participate, you will generally receive 5 participation points for each class you attend. Due to their negative impact on the

learning environment, the following things will negatively impact your daily participation grade:

- Being late to class or leaving early[2] - 1 point
- Being very late to class or leaving very early (>5 min) – 2 points
- Failing to abide by the device usage policy (see below)– 1 point per incident
- Disrespecting a fellow student – 1-5 points, depending on the severity
- Failing to participate actively (remember active participation can take many forms, see above) in group discussions, pair activities, etc. – 1-5 points, depending on the severity
- Coming to class unprepared and unable to contribute to group activities – 1-5 points, depending on the severity

If you lost participation points, there will be a comment on your Moodle grade for that day explaining the reason.

3) ***Labs (15%).*** Lab activities are an extremely important part of this class and are designed to give you the skills you need to apply statistical and computational techniques to real data. Over the course of the semester, you will build up to doing an analysis that you design on a dataset of your choosing (see final project, below). Labs are designed to be finished in class, however everyone works at a different pace. You are free to continue working on the lab after the lab period, and labs are formally due at the start of the following Monday class. Please arrange to meet with your lab partners and complete the lab together if and when you are finishing up outside of class.
4) ***Projects 1 and 2 (5% each, 10% total*).** You will complete two smaller scale projects to give you some practice at applying your developing computational and statistical skills to a problem, designing data visualizations, and presenting the results.
5) ***Final Project(20%).*** Your efforts in this class will culminate in a data investigation where you choose the dataset, question to investigate, analysis techniques, and visualizations yourself. To keep you on track, there will be a number of intermediate deadlines interspersed throughout the second half of the semester and the TAs and I will work closely with you on designing your investigation.

Deadlines

I am a big believer in frequent small assignments as the best way to master course material. As such, there are frequent deadlines for this course. Please make use of the course calendar to keep track of what is due when. If and when the deadlines are modified, I'll post a new version of this calendar as a Moodle announcement. You will soon fall into a pattern for the course. The regular weekly deadlines are:

- Reading annotations before Monday and Wednesday classes
- Reading and reading assignment before Friday classes
- Prelabs before Thursday classes
- Labs at the end of the lab period on Thursday or before class on Mondays

Guidelines, point breakdowns, and rubrics for the three course projects will be distributed later

[2] Except in the case of a documented flexible attendance need

in the semester with deadlines clearly indicated.

Submitting Assignments
All assignments for this course should be submitted via Moodle. Before submitting any assignment that includes a Jupyter notebook, please go through the following exercise:

1) In the menu at the top of the notebook, select Kernel → Restart and Run All.
2) Make sure that all of the cells in your notebook have executed properly and produce the desired output. **We cannot grade code that doesn't compile.**
3) Once you've checked that the cells execute, select Kernel → Restart and Clear All.
4) Combine your notebook and any other files required to run your code (images, standalone .py files, data) into a **single zip repository**.
5) **Rename this .zip file** with the last names of all group members and the assignment name. For example, if Mark, Clare and I were working in a group on Lab #2, we would name the file FolletteLeonardSchoen_Lab2.zip before submitting. If I were submitting my prelab #6 I would name it Follette_Prelab6.zip.

Late Work Policy
Prelabs and Reading Annotations/Assignments are essential to the success of our in-class work and cannot be handed in late (though some will be dropped, see below). Late labs and projects will be subject to a 10% per day penalty up to 10 days after the official due date.

Illness, Injury, Etc. and Dropped Assignments
There will likely be times during the semester when you are under the weather, overwhelmed by other coursework, or something else is going on in your life that prevents you from handing in an assignment or attending a class. This is inevitable, and understandable. To accommodate such situations, your lowest-scoring lab, prelab, reading annotation, reading quiz, and coding quiz will be dropped before computing your grade. You may miss up to two M/W class sessions and one discussion without a participation point penalty. You may also miss one lab without penalty, however if this happens you must complete the lab individually by the start of the lab period the following week. If you need help with it, please make sure to attend office hours.

Office Hours and Help Sessions (NSC A126)
Attendance at Office Hours is by far the best ways to get your questions about assignments, grades, and course content answered. Students consistently say in course evaluations that they wish they'd gone more often, so please take advantage of these help sessions as often as you can. You can come to discuss specific problems, but also just to check your understanding of a tricky topic, or to say hello. Mark will hold a TA help session on Wednesday afternoons 2-4pm. I will hold my office hours on Fridays from 2-3pm, and both of us and Clare are also available by appointment. My schedule is often quite tight, so ***please request appointments with me at least one full day in advance*** so that I can make room to meet with you. Five College students are welcome to attend office hours in person or virtually. I have an iPad on a rolling stand that will allow you to participate robotically.

Computer Requirements

You are welcome to install python on your personal computer according to the instructions posted on Moodle if you wish to use it for your coursework, however you are not required to do so. The computers in the classroom (NSC A126) are available for you to use at any point when the room is not reserved. If you are a Five College student and do **not** plan to work on a personal computer, please let me know ASAP so that I can work with your astronomy department to get you access to an appropriate machine on your campus.

A Note on the Textbook

Data science is inherently multidisciplinary and we will be drawing ideas from a number of areas – statistics, computer science, astronomy, graphical design, etc. There are several supplemental texts on reserve in the library, and I am happy to point you to others if you encounter a concept that you want to explore in more depth or that needs further clarification.

Technology Usage Policy

Data suggests the following:

1) That students who use cell phones during class receive grades that are up to a full letter grade lower than their peers who do not use them in class[3].
2) Students who use technology in class score lower on assessments that measure their comprehension of the material, and *their peers who are within view of their screens score even lower yet*[4].

For this reason, cell phones and computers are not allowed during class except when required for an assignment or activity. If you have a documented reason for needing to use technology in class, I am happy to make accommodations. All other technology use will have an impact on your participation grade. Please plan to stow your cell phones, computers, and tablets in your backpacks or pockets during class rather than leaving them on the table as a temptation.

Where to Go To Find What and When

On the Course GitHub:

- Lab and prelab assignments
- Project rubrics and instructions

On the Course Moodle Site:

- Your grades, as they become available
- Problem solutions
- Announcements for the course (***please make sure your Moodle settings have you receiving course announcements by e-mail***)
- Submission links for assignments

Accommodations for Disabilities

If you have any kind of disability, whether apparent or non-apparent, learning, emotional, physical, or cognitive, and you need some accommodations or alternatives to lectures,

---

[3] Duncan, Hoekstra & Wilcox. (2012). "Digital Devices, Distraction, and Student Performance: Does In-Class Cell Phone Use Reduce Learning?". *Astronomy Education Review*. 11. 10.3847/AER2012011.

[4] Sana, Weston & Cepeda (2013). "Laptop multitasking hinders classroom learning for both users and nearby peers", *Computers & Education*, Volume 62, p. 24-31.

assignments, or exams, please feel free to contact me to discuss reasonable accommodations for your access needs.

I rely on Accessibility Services for assistance in developing accommodation strategies. If you have not previously contacted Accessibility Services, I encourage you to do so:
email - a ccessibility@amherst.edu or 413-542-2337.

I look forward to working with you this semester in Astronomy 200!

Course Calendar (subject to change)

Dates	Mon	Tue	Wed	Thu	Fri	Sat-Sun
1/28-2/3	Lecture #1		Lecture #2	Lab #1	Discussion #1	
2/4-2/10	Lecture #3		Lecture #4	Lab #2	Discussion #2	
2/11-2/17	Lecture #5		Lecture #6	Lab #3	Discussion #3	
2/18-2/24	Lecture #7		Lecture #8	Project 1 Work Time	Discussion #4	
2/25-3/3	Lecture #9		Lecture #10	Lab #4	Project 1 Presentations	
3/4-3/10	Lecture #11		*Slush Class*	Lab #5	*Slush Class*	*Spring Break*
3/11-3/17	*Spring Break*	*Spring Break*	*Spring Break*	*Spring Break*	*Spring Break*	*Spring Break*
3/18-3/24	Lecture #12		Lecture #13	Lab #6	Discussion #5	
3/25-3/31	Lecture #14		Lecture #15	Lab #7	Discussion #6	
4/1-4/7	Lecture #16		Lecture #17	Project 2 Work Time	Discussion #7	
4/8-4/14	Lecture #18		Lecture #19	Lab #8	Project 2 Presentations	
4/15-4/21	Lecture #20		Lecture #21	Lab #9	Discussion #8	
4/22-4/28	Lecture #22		Lecture #23	Lab #10	Discussion #9	
4/29-5/5	*Slush Class*		*Slush Class*	Final Project Work Time	*Discussion #10*	
5/6-5/12	*Makeup Day*	*Makeup Day*	*Makeup Day*	*Reading Period*	*Reading Period*	*Reading Period*
5/13-5/19	*Final Exam Week*	*Final Exam Week*	*Final Exam Week*	*Final Exam Week*	*Final Exam Week*	

Course Outline (subject to change)

Class	Date	Topic	Assignments Due	Reading Due
Lec 1	Mon 1/28	What is data? What is data science? How is it used in Astronomy?	N/A	Syllabus
Lec 2	Wed 1/30	Types of variables Populations and Samples	Perusall annotations	Diez p. 1-26
Lab 1	Thu 1/31	Variables and Arrays	Prelab #1 – Intro to Jupyter Notebooks	
Disc 1	Fri 2/1	Graphical Excellence	Reading Assignment #1	Tufte Ch 1
Lec 3	Mon 2/4	Descriptive Statistics	Lab #1, Perusall annotations	Diez p. 26-43
Lec 4	Wed 2/5	Types of Data Visualizations	Perusall annotations	Diez p. 43-54
Lab 2	Thu 2/6	Functions and Control Flow	Prelab #2 – Intro to Functions	
Disc 2	Fri 2/7	Graphical Integrity	Reading Assignment #2	Tufte Ch 2
Lec 5	Mon 2/11	Linear Regression – correlation and fitting	Lab #2, Perusall annotations	Diez p. 331-344
Lec 6	Wed 2/13	Linear Regression – outliers and error handling	Perusall annotations	Diez p. 345-355
Lab 3	Thu 2/14	Plotting with Matplotlib	Prelab #3 – Intro to Plotting	
Disc 3	Fri 2/15	Graphical Deception	Reading Assignment #3	Handouts: Tufte “Corruption in Evidence Presentations”+ Sagan “The Fine Art of Baloney Detection”
Lec 7	Mon 2/18	Statistical Distributions – The Normal Distribution	Lab #3, Perusall annotations	Diez p. 127-140
Lec 8	Wed 2/20	Statistical Distributions – Other	Perusall annotations	Diez p. 141-157
	Thu 2/21	Open Work Time for Project #1		
Disc 4	Fri 2/22	Use of Graphics to Make an Argument – The Good and the Bad	Reading Assignment #4	Handout: Tufte “Displays of Evidence for Making Decisions”
Lec 9	Mon 2/25	Sampling, Estimation and Confidence Intervals	Perusall annotations	Diez p. 168-180

Class	Date	Topic	Assignments Due	Reading Due
Lec 10	Wed 2/27	Hypothesis Testing	Perusall annotations	Diez p. 180-194
Lab 4	Thu 2/28	Advanced Data Structures	Prelab #4 – Intro to Advanced Data Structures	
	Fri 3/1	Project #1 Presentations	Project #1 Presentations	
Lec 11	Mon 3/4	The Central Limit Theorem	Lab #4, Perusall annotations	Diez p. 194-202
	Wed 3/6	*Slush Class*	Project #1 Writeup	TBD
Lab 5	Thu 3/7	Tables	Prelab #5 – Intro to Pandas	
	Fri 3/8	*Slush Class*	TBD	TBD
3/9-3/17 Spring Break				
Lec 12	Mon 3/18	Intro to Probability in Astronomy	Lab #5, Perusall annotations	Wall & Jenkins p. 1-19
Lec 13	Wed 3/20	Bayes' Theorem	Perusall annotations	W&J p. 20-33
Lab 6	Thu 3/21	Graphical Exploration of the IPAC Database	Prelab #6 – Intro to IPAC Database	
Disc 5	Fri 3/22	Data-Ink and Chartjunk	Reading Assignment #5	Tufte Ch 4 + 5 (jigsaw)
Lec 14	Mon 3/25	Statistical Distributions in Astronomy	Lab #6, Final Project Proposal, Perusall annotations	W&J p. 33-46
Lec 15	Wed 3/27	Random Numbers and Monte Carlo Simulations	Perusall annotations	W&J p. 46-52
Lab 7	Thu 3/28	Monte Carlo Simulation	Prelab #7 – Random Numbers	
Disc 6	Fri 3/29	Graphical Aesthetics and Design Principles	Reading Assignment #6	Handout: Tufte "Principles of Analytical Design" + Tufte Ch 9
Lec 16	Mon 4/1	Types of Error	Lab #7, Perusall annotations	W&J p. 55-62
Lec 17	Wed 4/3	Error Propagation	Perusall annotations	W&J p. 62-69
	Thu 4/4	Open Work Time for Project #2	Project #2 Presentations	
Disc 7	Fri 4/5	Graphical Design, Cont'd	Reading Assignment #7	Tufte Ch 6 + 7 (jigsaw)
Lec 18	Mon 4/8	Correlation Metrics in Astronomy	Perusall annotations	W&J p. 71-84
Lec 19	Wed 4/10	Hypothesis Testing in Astronomy	Perusall annotations, Project #2 Writeup	W&J p. 92-96
Lab 8	Thu 4/11	Correlation Testing with IPAC Database	Prelab #8 – Functions for correlation testing	

Class	Date	Topic	Assignments Due	Reading Due
	Fri 4/12	Project #2 Presentations		
Lec 20	Mon 4/15	Parametric Hypothesis Tests	Lab #8, Final Project Outline, Perusall annotations	W&J p. 96-106
Lec 21	Wed 4/17	Non-Parametric Hypothesis Tests		W&J p. 97-123
Lab 9	Thu 4/18	Hypothesis Testing with IPAC Database	Prelab #9 – Functions for hypothesis testing	
Disc 8	Fri 4/19	Multidimensional Data	Reading Assignment #8	Handout: Tufte "Escaping Flatland"
Lec 22	Mon 4/22	Maximum Likelihood Modeling	Lab #9, Perusall annotations	W&J 126-133
Lec 23	Wed 4/24	Least Squares and Minimum Chi-Squared Modeling	Perusall annotations	W&J 134-142
Lab 10	Thu 4/25	Model Fitting with IPAC Database	Prelab #10 – Model Fitting Intro	
Disc 9	Fri 4/26	Data Narratives	Reading Assignment #9	Handout: Tufte "Narratives of Space and Time"
	Mon 4/29	*Slush Class*	Lab #10, TBD	TBD
	Wed 5/1	*Slush Class*	TBD	TBD
	Thu 5/2	Open Work time for Final Project		
Disc 10	Fri 5/3	TBD	TBD	TBD
Finals Week		Final Project Presentations	Final Project Writeup and Presentation	