

CS 109A, STAT 121A, AC 209A, CSCI E-109A

Introduction to Data Science

Syllabus - Fall 2019

Welcome to CS109a/STAT121a/AC209a, also offered by the DCE as CSCI E-109a, Introduction to Data Science. This course is the first half of a one-year introduction to data science. The course focuses on the analysis of messy, real life data to perform predictions using statistical and machine learning methods.

The material of the course is divided into 3 modules. Each module will integrate the five key facets of an investigation using data:

1. data collection - data wrangling, cleaning, and sampling to get a suitable data set;
2. data management - accessing data quickly and reliably;
3. exploratory data analysis – generating hypotheses and building intuition;
4. prediction or statistical learning; and
5. communication – summarizing results through visualization, stories, and interpretable summaries.

Only one of CS 109a, AC 209a, or Stat 121a can be taken for credit. Students who have previously taken CS 109, AC 209, or Stat 121 cannot take CS 109a, AC 209a, or Stat 121a for credit.

Course Logistics

Prerequisites

You are expected to have programming experience at the level of CS 50 or above, and statistics knowledge at the level of Stat 100 or above (Stat 110 recommended). HWO is designed to test your knowledge on the prerequisites. Successful completion of this assignment will show that this course is suitable for you. HWO will not be graded but you are required to submit.

Course Components

Lectures

The class consists of two weekly lectures. Lectures are held **Mon and Wed 1:30pm - 2:45 pm** in **Northwest Building (NW), Lecture Hall B-103 (tentative)**.

Attendance to lectures is mandatory for FAS students. We will have in class quizzes to assess your understanding of the material and to help us identify gaps.

Labs

Labs are designed as hands-on in-class activities. The instructor will go over practice problems similar to the homework problems and review difficult material. Attendance to labs is optional but **strongly encouraged**.

Two lab sessions with identical content are held **TBD (tentatively, Thur 4:30-6:00 pm and Fri 10:30-11:45 am in Pierce 301)**. You should plan to attend one of the two.

Sections

Lectures and labs are supplemented by 1 hour sections led by teaching fellows. There are two types of sections:

- a) **Standard Sections**: which will be a mix of review of material and practice problems similar to the HW
- b) **Advanced Sections** which will cover advanced topics like the mathematical underpinnings of the methods seen in lecture and lab and extensions of those methods. The material covered in the Advanced Sections is **required** for all AC 209A students.

Video Recordings

All learning instances—lectures, labs, and sections—will be recorded. DCE students can stream these events in real time with a TF monitoring the accompanying chat room. Recordings will then be made available to all students within 24 hours via Canvas.

Instructor Office Hours

Pavlos: TBD, IACS lobby

Kevin: TBD, IACS lobby

Chris: TBD, IACS lobby

Assignments

There will be an initial self-assessment homework called HWO and 8 more graded homework assignments. Some of them will be due in a week and some of them in two weeks. You have the option to work and submit in pairs for all the assignments **except** two which you will do individually. You will be working in Jupyter Notebooks which you can run in your own environment or in the SEAS JupyterHub cloud (*details on this to come*).

Quizzes

Quizzes will be taken at the end of class and the material will be based on what was discussed in lecture. DCE students will have 72 hours to complete the quizzes.

40% of the quizzes will be dropped from your grade.

Final Project

There will be a final group project (2-4 students) due during Exams period. See Calendar for specific dates.

Recommended Textbook

An Introduction to Statistical Learning by James, Witten, Hastie, Tibshirani.

The book is available here:

Free electronic version: <http://www-bcf.usc.edu/~gareth/ISL/> ([Links to an external site](#)).

HOLLIS:

<http://link.springer.com.ezp-prod1.hul.harvard.edu/book/10.1007%2F978-1-4614-7138-7>

Amazon:

<https://www.amazon.com/Introduction-Statistical-Learning-Applications-Statistics/dp/1461471370> ([Links to an external site](#)).

Course Policies

Getting Help

For questions about homework, course content, package installation, JupyterHub, and after you have tried to troubleshoot yourselves, the process to get help is:

1. Post the question in **Piazza** and hopefully your peers will answer. Note that in Piazza questions are visible to everyone. The TFs monitor the posts.
2. Go to **Office Hours**, this is the best way to get help.
3. For private matters send an email to the **Helpline**: cs109a2019@gmail.com.

The Helpline is monitored by TFs.

4. For personal and confidential matters send an email to the **instructors**.

Grading Guidelines and Regrading Policy

Homework will be graded based on:

1. How correct your code is (the Notebook cells should run, we are not troubleshooting code)
2. How you have interpreted the results -- we want text not just code. It should be a report.
3. How well you present the results.

The scale is 1-5.

For more details, check out [The CS109A Grade](#)

Questions on Graded Homework and Regrading Policy

We take great care in making sure all homework assignments are graded properly.

However, if you feel that your assignment was not fairly graded you may:

Contact the grader by emailing the helpline with subject line "Regrade HW1:

Grader=johnsmith" within **48 hours of the grade release**.

If still unhappy with the initial response, then submit a reason via email to the Helpline with subject line "Regrade HW1: Second request" within 2 days of receiving the initial response. Note: once regrading is done, you may receive a grade that is higher or lower than the initial grade.

Collaboration Policy

We encourage you to talk and discuss the assignments with your fellow students (and on Piazza), but you are not allowed to look at any other student's assignment or code outside of your pair. Discussion is encouraged; copying is not allowed. Please refer to the Academic Honesty section in The CS109A Grade link above.

Late Day Policy

Homework is due on Tuesdays. There are no late days.

Communication from Staff to Students

Class announcements will be through **Canvas**. All homework and quizzes will be posted and submitted through Canvas, as well as all feedback forms.

NOTE: make sure you adjust your account settings so you can receive emails from Canvas.

Submitting an assignment

Instructions for turning in assignments will be posted when the semester starts.

Course Grade

Your final score for the course will be computed using the following weights:

Homework 50%

Quizzes 10%

Project 40%

Total 100%

Software

We will be using Jupyter Notebooks, Python 3, and various python modules. You can access the notebook viewer either on your own machine by installing the [Anaconda platform \(Links to an external site\)](#) which includes Jupyter/IPython as well all packages that will be required for the course, or by using the SEAS JupyterHub from Canvas. Details in class.

Accommodations for students with disabilities

Students needing academic adjustments or accommodations because of a documented disability must present their Faculty Letter from the [Accessible Education Office](#) (AEO) and speak with Kevin by the end of the third week of the term: Friday, September 15. Failure to do so may result in us being unable to respond in a timely manner. All discussions will remain confidential.

Academic Honesty

Ethical behavior is an important trait of a Data Scientist, from ethically handling data to attribution of code and work of others. Thus, in CS109 we give a strong emphasis to Academic Honesty.

As a student your best guidelines are to be reasonable and fair. We encourage teamwork for problem sets and allow submission in pairs for most problem sets, but you should not split the homework and you should work on all the problems together.

For more detailed expectations, please refer to the Academic Honesty section in The CS109A Grade link above.