STAT 200, Section DS = STAT 207 Data Science Exploration

- **Lectures**: 1090 Lincoln, MWF 1:00-1:50pm
- **Open Lab**: Monday/Tuesday 5:00-7:00pm, English Building, Room 69
- **Instructor**: [Douglas Simpson](), Department of Statistics
- **Office Hours**: 118 Illini Hall, Wed 2:00-3:30, Thu 1:30-2:30
- **Teaching Assistant**: Yuxuan Liu, Department of Statistics
- **Course Website**: http://go.illinois.edu/stat207
- **Course Piazza**: http://piazza.com/illinois/spring2020/stat207

Syllabus

Building on the foundation of STAT 107, Data Science Discovery, we use Python, Jupyter notebooks, and GitHub to explore data science techniques, statistical concepts, and data analytic workflows, combined with the statistical analysis of STAT 200. As we explore data science we:

- Develop understanding of probability models for noisy data and how these translate into uncertainty analysis and statistical inference
- Understand how modeling assumptions and sampling frames affect our conclusions
- Become adept with multiple regression modeling, basic machine learning, and inference
- Become proficient in Python coding for data management, analytics, visualization
- Understand and use GitHub repositories, the industry standard for submitting code and reports

Prerequisites: None. Students who completed STAT 100 or 107 previously are welcome. Equally welcome are students in quantitative fields taking STAT 200 as their first course in statistical analysis. The early programming will comprise an accelerated introduction for those who did not take STAT 107 prior and a quick review for those who did. You will receive course credit and Quant I or Quant II credit for STAT 200 while participating in the launch of Data Science Exploration!

Course Topics

See: Schedule

Course Section

The primary contact hours for this course are comprised of two major components:

- **Lecture Section**: Meets three times a week (M/W/F), 50 minutes each lecture, lead by University of Illinois faculty.

- **Weekly Open Labs**:

    - Monday and Tuesday evenings 5:00-7pm, 69 English
    - Go here for personalized and small group help
    - These are ideal settings to complete your lab assignments

Course Materials

- **Required Calculator**: Any non-programmable calculator (no phones, graphing calculators, etc.)

- **Laptop Computer**: You need a laptop running Windows, OS X, or Linux. 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).

- **Lecture notes**: These will be posted on the course website.

- **Online Books**: To read more about the topics in this course.

    - J. VanderPlas (2016) *Python Data Science Handbook*, https://jakevdp.github.io/PythonDataScienceHandbook/
    - Diez, Barr, and Cetinkaya-Rundel, (2015), *OpenIntro Statistics* https://www.openintro.org/download.php?file=os3&redirect=/stat/textbook/os3.php

Course Assignments and Grades (updated April 28)

Course grades are computed based on your percentage out of 700 points for the course. The components are as follows:

- Homework Labs - 250 pts
- Midterm Exams - 200 pts total
- Final Exam - 200 pts

Total: 650 pts

Bonus points:

- Any homework/lab points above 250 (total possible for 10 out of 12) will be added to bonus points, up to 50 bonus points.

- Project - Up to 50 bonus pts

Final Course Grade

Course points will be translated into a course grade at the end of the semester. The grade thresholds will be based on your percentage score out of 650:

Grade	Min Pct	Min Pts	Grade	Min Pct	Min Pts	Grade	Min Pct	Min Pts
A+	97	630.5	A	93	604.5	A-	90	585
B+	87	565.5	B	83	539.5	B-	80	520
C+	77	500.5	C	73	474.5	C-	70	455
D+	67	435.5	D	63	409.5	D-	60	390

Participation

Number one: attend class and go to open lab! Being present will help you keep up with what is going on, gain hands on experience in learning activities, and benefit from interacting with other students and instructional staff.

Please ask questions whenever anything is confusing. If you find errors in the notes, please report them to the instructor. The instructor will be very happy that you detected them so they can be corrected!

Late Submissions

No late submissions are accepted. However, it is only necessary to complete 10 of 12 labs to achieve full credit.

Learning Collaboratively

We encourage you to discuss all of your course activities (with the exception of exams) with your friends and classmates! You will learn more though 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 207 and all cases will be brought to the University, your college, and your department. You should understand how academic integrity applies specifically to STAT 207: 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 (70 points). A second incident, or cheating on an exam, results in an automatic F in the course.

Academic integrity includes protecting your work. If you 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.