36 - 200: Reasoning with Data
Course Policies and Syllabus, Summer 2018

Instructor: Purvasha Chakravarti (FMS 302)
Email: pchakrav@stat.cmu.edu *(Please allow 24 hours for email responses.)*

Teaching Assistant: Yue Li (FMS)
Email: yuel5@andrew.cmu.edu *(Please allow 24 hours for email responses.)*

Office Hours: TBD

Lectures: Mondays, Wednesdays, Fridays 10:30AM - 11:50AM, Wean Hall 5302

Labs: Tuesdays, Thursdays 10:30AM - 11:50AM, Baker Hall 140E

Prerequisites: None   **Units:** 9

Course Website: http://www.cmu.edu/canvas/

Textbook: None required, recommended readings will be provided.

Recommended References: Statistics: The Art and Science of Learning from Data by Agresti & Franklin; Introduction to the Practice of Statistics by Moore & McCabe

COURSE DESCRIPTION

In recent years, the growth of statistics has made itself felt in almost every phase of human activity. Statistics no longer consists merely of the collection of data and their presentation in charts and tables (*descriptive statistics*) - it is now considered to encompass the science of basing inferences on observed data and the entire problem of making decisions in the face of uncertainty (*inferential statistics*). This covers considerable ground since uncertainties are met when we flip a coin, when a dietitian experiments with food additives, when an actuary determines life expectancy, when a quality control engineer accepts or rejects manufactured products, when a teacher compares the abilities of students, when an economist forecasts market trends, when a newspaper predicts an election, and so forth.

This course is an introduction to learning how to make statistical decisions and "reason with data. The approach will emphasize thinking through an empirical problem from beginning to end and using statistical tools to look for evidence for/against an explicit argument/hypothesis. Types of data will include continuous and categorical variables, images, text, networks, and repeated measures over time. Applications will largely drawn from interdisciplinary case studies spanning the humanities, social sciences, and related fields. Methodological topics will include basic exploratory data analysis, elementary probability, hypothesis tests, and empirical research methods. There is no calculus or programming requirement. There will be two weekly computer labs for additional hands-on practice using an interactive software platform that allows student-driven inquiry.

This course is the credit-equivalent to 36-201 and will be honored appropriately as a prerequisite for downstream Statistics courses. As such, this course is not currently open to students who have received credit for 36-201, 36/70-207, 36-220, 36-247, or any 300- or 400-level Statistics course.

OBJECTIVES

1. Learn the empirical research process including data collection and design methods.

2. Develop and use methods for summarizing and evaluating numerical data.

3. Learn and apply the basic concepts of probability and hypothesis tests.

4. Develop skills in the applications of statistical methods to problems in the sciences and the social sciences, including interpretation and communication of results.

HOMEWORK

There will be two homework assignments per week, except weeks which have projects due. One will be assigned no later (and perhaps earlier) than 2:00pm on Tuesday and due at the beginning of class on Thursday. Another will be assigned no later (and perhaps earlier) than 2:00pm on Thursday and due at the beginning of class on the following Tuesday. The assignments may differ in length, roughly commensurate with the time given to complete them, and may include pages of the textbook to read along with a problem set. *You will be able to find the assignment on the course's Canvas website.*

Depending on the week, assignments will be submitted either in class or on Canvas. This deadline is strict to ensure efficient grading and because the instructor may choose to discuss the homework problems due that day during the class period. Late homework will not be accepted for a grade. Illness, family problems, travel problems, last minute computer problems, etc. are not an acceptable excuse for handing in homework late. Rather, one homework grade will be dropped. Since there are no excuses for turning in homework late, save the dropped homework for emergencies.

It is your responsibility to check your homework answers against the solutions. You may find better ways to do what you got correct, and you may find more extensive discussion and solutions for what you got wrong. Do not rely only on the comments on your assignment. Solutions will be made available on the Canvas site shortly after the assignment deadline.

- **Format.** Write your full name, your Andrew ID, the course number, and the homework number at the top of each page. **STAPLE** your entire assignment together with a staple (not with a paper clip, folding the pages together, etc.). Otherwise, we will take off points. It is your responsibility to make your homework legible - if the grader cannot read your work, you will receive a zero on the assignment.

- **Show all your work.** You will not receive credit for simply writing down a numerical answer even if the calculations seem simple enough to do in your head. How you arrived at your answer is more important. If the HW requires output or graphs, just circling the answers or attaching graphs with no labels or explanation is not acceptable. Answers should be written up appropriately. Any required graphs should be as close to the corresponding problem as possible.

- Please start homework assignments early so that you can get help if you need it. Each homework assignment will be worth 100 points.

- You are encouraged to discuss homework problems with your fellow students, however the work you submit must be your own. Acknowledge any help received on your assignments. Copied work will receive no credit.

LABS

On lab days, that is, on Tuesdays and Thursdays there will be 30 minutes of lecture followed by the lab. Discussion labs will involve hands-on practice with the concepts learned in previous classes. Lab assignments will consist roughly of exploring case study data sets, applying data science and statistical tools to answer questions about real scenarios, and providing short responses. They will be graded primarily for completeness.

PROJECTS

There will be two projects during the course. For the projects, students will use case studies and real research examples to demonstrate and apply concepts from class. The projects will include creating reports with summary information. More details will follow.

EXAMS

There will be an in-class midterm exam and a cumulative final examination. The exams will be closed-book and closed-notes, except that you may use one 8.5" by 11" sheet of paper (both sides) with whatever formulas, facts, or explanations you find helpful. Computers are not permitted, but calculators may be used. All exams are required and there will be no make-up exams. Cellphones must be turned off and put away during exams. Cheating/copying on exams results in a zero for the exam and a letter to your dean. Do your own work.

· A student who misses an examination because of a medical reason must provide *documented* evidence of medical incapacitation. Other reasons for missing an examination must be discussed with the instructor as soon as possible *before* the day of the examination. Each case will be considered on an individual basis.

COURSE AND EXAM SCHEDULE

Homework	Throughout	20%
Labs	Throughout	10%
Midterm	Monday, June 11 (in class)	15%
Projects	Friday, June 1 and Friday, June 22	30%
Final Exam	Friday, June 29 (in class)	25%
Total		100%

Course grade. Your overall course score will be determined as a weighted average of each element as noted above. A letter grade will be assigned based on:

A: 90 - 100   **B:** 80 - 89   **C:** 70 - 79   **D:** 60 - 69   **F:** Below 60

Grades may be curved at the instructors discretion.

Computing: For this class, we will be using a new software platform being designed by collaborators in the Department of Statistics and Heinz College. The platform will allow students to interact with data and case studies without requiring them to learn a programming language.

From the Eberly Center: For this class, a PhD student Philip Burckhardt, is conducting research on the impacts of various teaching methods and materials on student learning. You will not be asked to do anything above and beyond the normal learning activities and assignments that are part of this

course. You are free not to participate in this research, and your participation will have no influence on your grade for this course or your academic career at CMU. Please note that students choosing not to participate in the research will not be excused from required course activities. Participants will not receive any compensation. The data collected as part of this research will include student grades. All analyses of data from participants coursework will be conducted after the course is over and final grades are submitted. The Eberly Center may provide support on this research project regarding data analysis and interpretation. To minimize the risk of breach of confidentiality, the Eberly Center will never have access to data from this course containing your personal identifiers. All data will be analyzed in de-identified form and presented in the aggregate, without any personal identifiers. Please contact Philip Burckhardt at pburckhardt@cmu.edu, if you have questions or concerns about your participation.

Communication: Be advised that sending an email to the instructor or a teaching assistant does not create a responsibility or obligation to respond to it. Do not send complicated questions or requests to us via email. Replies will not be given for email questions or problems requiring lengthy (more than a couple of sentences) or complicated responses. These types of communications should be done in person. Assignments and class information will be posted on Canvas. **BE SURE TO TURN ON YOUR CANVAS NOTIFICATIONS.** Help is available at `www.cmu.edu/canvas`.

Please note that the instructor will only respond to email from an address taking the form [**andrewID**]**@andrew.cmu.edu.** To increase the chances of a response, start the subject line with "36-200:" so that it is clear that the email pertains to this class.

Office hours. The purpose of office hours is to provide you with an opportunity for additional conversation, guidance or help. Please feel free to come to our office hours at the time designated above. If you are not able to attend Purvasha's office hour, she would be happy to meet with you at a different time; *please email her to set up an appointment.*

Grades and regrades. Course grades will appear on the Canvas website. Each student is responsible for verifying his or her recorded scores on an ongoing basis.
Although we strive for consistency and accuracy in grading, we understand that grading errors can occur.

- We will gladly correct all errors in tabulation or overlooked material.

- All regrading requests must be accompanied by a written statement carefully highlighting and explaining the items that were misgraded. Note that regrading requests can end in a positive, negative, or no change in points.

- Regrade requests should be submitted to the instructor within one week of when the assignment or exam is returned.

Academic Integrity. The ethical guidelines and cheating and plagiarism policies are defined in the *Student Handbook* at `http://www.cmu.edu/policies/documents/Cheating.html`. You will be held accountable for violations of these guidelines and policies that come to our attention. We encourage you to be helpful to your classmates and to work together, the work you turn in must be your own. Any student who turns in work for credit that is identical, or similar beyond coincidence, to that of another student may face appropriate disciplinary action at the department, college, or universty level. *Cheating and/or plagiarism will not be tolerated.*

The purpose of homework assignments is to help you learn the material. It is okay to discuss an assignment with other students *but the written solutions to homework problems must be your own*

and not copied from someone else. If you receive significant help on an assignment from a classmate, please acknowledge that help explicitly in writing on your assignment; you will not be penalized for doing so, provided the written solutions are clearly your own. If you are uncertain as to whether the help you received was significant, err on the side of more acknowledgment. If you are uncertain about what degree of collaboration is acceptable, please discuss the situation with the instructor.

Attendance. Attendance is expected and you will be asked to perform in-class work. Only the combination of class lectures, homework sets, projects, labs and your own reading will give you full exposure to the material and prepare you adequately for examinations. Participation is encouraged; use of cellular phones and laptops (except for taking notes or topics pertaining directly to computing) during the instruction period is strongly discouraged.

Study skills and time management. Planning and time management are essential skills for academic success. It is important that you make every effort to keep up with the lectures, readings and assignments. Manage your time effectively and bear in mind that you are ultimately responsible for your own performance in this course. Plan to start each assignment early so that you finish on time. This strategy will allow you to have time for questions or react to any problems. We do not want to hear excuses for failure to complete your work on time or at a low quality. Your procrastination or failure to manage your time effectively will not be considered as an extenuating circumstance for poor performance.

Disability services. If you have a disability and require accommodations, please contact Catherine Getchell, Director of Disability Resources, 412-268-6121. If you have an accommodations letter from the Disability Resources office, we encourage you to discuss your accommodations and needs with the instructor as early in the semester as possible. We will work with you to ensure that accommodations are provided as appropriate.

Take care of yourself: Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

CaPS: 412-268-2922

Re:solve Crisis Network: 888-796-8226

If the situation is life threatening, call the police:

On campus: CMU Police: 412-268-2323

Off campus: 911

36 - 200: Reasoning with Data
Course Program

Lecture	Weekday	Date	Content	Due
1	Monday	May 21	"Big Picture"; Variable Types	
2	Tuesday	May 22	1-var EDA (categorical)	Lab 1
3	Wednesday	May 23	1-var EDA (quantitative)	
4	Thursday	May 24	2-var EDA	Lab 2; HW 1
5	Friday	May 25	2-var EDA	
6	Monday	May 28	Holiday - No class	
7	Tuesday	May 29	Relationships between Variables	Lab 3; HW 2
8	Wednesday	May 30	Bias; Lurking Variables	
9	Thursday	May 31	Experimental Design	Lab 4
10	Friday	June 1	Experimental Design	Project 1
11	Monday	June 4	Elementary Probability	
12	Tuesday	June 5	Elementary Probability	Lab 5; HW 3
13	Wednesday	June 6	Discrete Distributions; Binomial	
14	Thursday	June 7	Density Curves	Lab 6; HW 4
15	Friday	June 8	Normal Distribution/Review	
16	Monday	June 11	**Midterm Exam**	
17	Tuesday	June 12	Standardization	Lab 7
18	Wednesday	June 13	Sampling Distribution; CLT	
19	Thursday	June 14	Sampling Distribution; CLT	Lab 8; HW 5
20	Friday	June 15	Confidence Intervals	
21	Monday	June 18	Significance Tests	
22	Tuesday	June 19	t-distribution	Lab 9; HW 6
23	Wednesday	June 20	Inference for two means/proportions	
24	Thursday	June 21	Inference for two means/proportions	Lab 10
25	Friday	June 22	Inference for Linear Regression	Project 2
26	Monday	June 25	One-Way ANOVA	
27	Tuesday	June 26	Relationship between Categorical Variables	Lab 11; HW 7
28	Wednesday	June 27	Chi-Square Test; Special Topics	
29	Thursday	June 28	Review	
30	Friday	June 29	**Final Exam**	

Note that some of the timing and topics here are tentative.