CORE-UA 111: From Data to Discovery
Syllabus

Instructor	Mutiara Sondjaja	**Lecture**	Tu Th 2:00-3:15PM
Email	sondjaja@nyu.edu	**Classroom**	Silver 401
Office	WWH 725	**Course Page**	via NYU Classess

Teaching Asisstant	Elizabeth Zhao (rz1280@nyu.edu)
Lab Sections	Section 2: F 9:30-10:45AM, 194 Mercer, Room 304
	Section 3: F 11:00-12:15PM, 194 Mercer, Room 304

Goals of the Course

We live in an era of almost too much information. Today's technology enables us to collect massive amounts of data, including images of distant planets, the ups and downs of the economy, and the patterns of our tweets and online behavior. As educated citizens living in a world of data and technology, we need to be discerning about how journalists, politicians, and scientists draw conclusions from data, and to be mindful about how data are collected and used.

This is a hands-on course that aims to equip students with skills in data, quantitative, and computational thinking.

1. **Working with Data.** Students can comfortably use R and Jupyter Notebook to download openly-available data, explore data, and visualize data.
2. **Thinking Quantitatively.** Students understand how to interpret results from sampled data, understand what a mathematical model is, can make prediction using basic models, and can interpret results of these predictions. Students will also develop the habit to ask questions about the validity of claims and conclusions drawn from data before accepting them.
3. **Thinking Computationally.** Students understand and can fluently use basic programming elements, including: types of data, functions, conditional statements, and loops.

Throughout the course, we will also examine issues such as **data privacy and ethics**.

Prerequisites: The course has no prerequisites beyond basic arithmetic and algebra.

Expectations: As your instructor, my aim is to help you gain firm understanding of the above topics. You can expect that course material and information are presented clearly, and that I am available to answer your questions: in class, during the scheduled drop-in office hours, and via Piazza within 2 business days.

I am rooting for you success! but your learning cannot happen without your active involvement. You are expected to actively participate during in-class problem-solving, contribute to a positive learning environment, spend time and thoughtful effort in all assignments, and maintain the highest level of academic integrity.

Textbooks and Required Technology

1. **CoCalc**: You are required to create a free account on `CoCalc.com` using your @nyu.edu email address. All labs, homework assignments, and lecture demonstrations will be available through CoCalc.
2. **NYU Classes** and **Piazza**: Announcements and other course material will be posted on NYU Classes. We will use Piazza for questions and answers.
3. **Reference Textbooks:**
   - *Open Intro Statistics with Randomization and Simulation*, by David M. Diez, Christopher D. Barr, and Mine Çetinkaya-Rundel.
     PDF copy (free) or print copy ($8.49): `https://www.openintro.org`.
   - *R For Data Science*, by Hadley Wickham.
     Online copy (free): `http://r4ds.had.co.nz/`

Assessments and Grading

Labs (12%)

- Labs are computer-based work you will do in the weekly recitations. If you like, you can choose to bring your laptop computer to recitation. Each lab worksheet is **due at the end of your assigned Friday lab session** on CoCalc. No make-ups or late submissions are allowed. Two lowest lab grades are dropped.

Weekly Homework (12%)

- Each homework assignment is computational, a writing assignment, or project-related. Homeworks are **due at 9AM on Fridays** through CoCalc. See further instructions on NYU Classes. **No emailed or hardcopy homework** will be accepted.
- **Late homework policy**: To maintain fairness, homework submitted within 24 hours of the official deadline will be accepted with penalty. Homeworks submitted more than 24 hours after the deadline are not accepted and will receive 0 points. Technical issues is not a valid excuse for additional homework extension. Two lowest homework scores will be dropped.

Clicker Participation (2%)

- A "clicker app" will be used to encourage active learning. Further instructions will be provided on NYU Classes. You must bring your clicker device to each lecture.
- A clicker participation grade of 75% or better corresponds to all 2% of clicker credit. **Manual attendance will not be taken due to the size of the class.** The generous grading of clicker participation take into account the possibility of students': (1) having to miss class due to other obligations or minor illness, (2) forgetting their clicker on a non-regular basis, (3) non-persistent technical problem with the clicker app/equipment.

Projects (24%)

There are two projects in this course. This is your opportunity to use your newly acquired data skills on problems that are more involved than the weekly homework. Further detail will be posted on NYU Classes.

- **Project 1** (8%): You will work in groups of 3 with a dataset that we provide for you. See calendar.
- **Project 2** (16%): You will work in groups of 3 with a dataset of your choice. See calendar.

Exams (50%)

Midterm 1	October 2 (in class)	Exploring and Visualizing Data
Midterm 2	November 8 (in class)	Statistical Inference
Final Exam	December 18, 2PM-3:50PM, Location TBA	All Topics

The lowest exam score will be dropped; the remaining two exams are weighed equally: each exam counts towards 25% of your overall course grade.

Academic Integrity

We value hard work and integrity, and do not tolerate academic dishonesty. You are expected to uphold academic integrity as specified by the university and the College of Arts and Sciences. Remember that we are here to learn.

Group work, assignments, and academic integrity: You are encouraged to form study groups and to exchange ideas with classmates when wrestling with homework and projects. However, you must write up your weekly homework and lab individually. A good rule of thumb: make sure that you can reproduce or explain your submitted work individually without looking at what you have written up. And finally, give credit where it's due: indicate with whom you have held discussions as you worked on your assignments.

Course Policies

There will be no accommodation for missed homework, labs, and exams, except in the cases of illness, observance of religious holidays, official/university-approved travels (e.g., athletic meets), or exceptional, extenuating circumstances. In the case of observance of religious holidays or university-approved travels, you must make arrangements to make up missed work **at least one week in advance**. In the case of illness, you must present a letter from a physician/health care provider as soon as possible. Students with disabilities or requiring special accomodations must make individual arrangements with Moses Center.

Advice for making most of this course

- **Attend lectures and labs.** It is very hard to catch up if you regularly miss classes. Your instructor and TAs are there to help you learn; use this opportunity well.
- **Get your hands dirty in class.** Actively participate when we solve problems in class. Passively listening to lectures and taking notes are generally not sufficient to really internalize new, challenging ideas.
- **Spend time** on each homework, lab, and projects. Expect to spend 4-8 hours each week on assignments and study outside of class. This is your opportunity to wrestle with and to internalize new ideas introduced in class. When working on assignments, strive to really understand the ideas behind the methods.
- **Don't hesitate to get help, and to do so early**:
  - **Attend instructor and TA's office hours**. Do not wait to seek help if you have any questions about the ideas introduced in class. Office hours schedule will be posted in the **NYU Classes** page for our section. No appointment is needed to attend these regular office hours.
  - **Form study groups**, but it's critical that you write up your own homework individually. Homework and assignments that are copied from or written by someone else are not hard to spot. Academic dishonesty is not tolerated.
  - **Piazza**: Use the course Piazza page to post questions and to respond to classmates' questions. When you do, make sure to be courteous and respectful. For homework-related questions, full solutions to homework/worksheet problems should not be requested or provided.

Tentative Calendar

	Week	Date	Topics	Work Due
Exploring and Visualizing Data	1	09/04	1. What is this class about?	
		09/06	2. Introduction to computing using R	
		09/07	*No Lab*	
	2	09/11	3. Data frames and types of data	
		09/13	4. Working with data frames	
		09/14	*Lab 1: Using R and Jupyter Notebook to explore data*	HW 1, Lab 1
	3	09/18	5. Data visualization	
		09/20	6. Data visualization	
		09/21	*Lab 2: Exploring and visualizing data*	HW 2, Lab 2
	4	09/25	7. Application: Analyzing text data	
		09/27	8. Review/Catch-Up/Guest Speaker	
		09/28	*Lab 3: Analyzing text data*	HW 3, Lab 3
Statistical Inference	5	10/02	MIDTERM EXAM 1	
		10/04	9. Understanding randomness and probability	
		10/05	*Lab 4: Project 1 Warm-up (team assignment; dataset intro)*	HW 4, Lab 4
	6	10/09	No Class (Legislative Day—Monday schedule)	
		10/11	10. Random sampling, point estimates, and sampling distribution	
		10/12	*Lab 5: Simulating random events*	HW 5, Lab 5
	7	10/16	11. Random sampling, point estimates, and sampling distribution	
		10/18	12. Averages, standard deviation, and the normal distribution	
		10/19	*Lab 6: Understanding the central limit theorem via simulations*	Project 1, HW 6, Lab 6
	8	10/23	13. Confidence intervals	
		10/25	14. Confidence intervals with bootstrap	
		10/26	*Lab 7: Bootstrap and confidence intervals*	HW 7, Lab 7
	9	10/30	15. Testing statistical hypotheses	
		11/01	16. Testing statistical hypotheses	
		11/02	*Lab 8: Making a hypothesis and performing hypothesis test*	HW 8, Lab 8
	10	11/06	17. Review/Catch-Up/Guest Speaker	
		11/08	MIDTERM EXAM 2	
		11/09	*Lab 9: Project 2 Warm-up (team assignment; dataset intro)*	HW 9, Lab 9
Models and Predictions	11	11/13	18. Scatterplots and correlation	
		11/15	19. Linear regression and prediction	
		11/16	*Lab 10: Scatterplots and linear regression*	HW 10, Lab 10
	12	11/20	20. Introduction to graphs and networks	
		11/22	No Class (Thanksgiving Break)	
		11/23	No Class (Thanksgiving Break)	
	13	11/27	21. Clustering: Identifying clusters on graphs	
		11/29	22. Clustering: k-Nearest Neighbors	
		11/30	*Lab 11: Modeling and prediction*	HW 11, Lab 11
	14	12/04	23. Application: Analyzing social networks	
		12/06	24. Catch-Up/Guest Speaker	
		12/07	*Lab 12: Applying k-Nearest Neighbors clustering*	HW 12, Lab 12
	15	12/11	25. Project presentations	
		12/13	26. Model assessment and validation	
		12/14	*Lab: Finish working on Project 2 write-up*	Project 2
		12/18	FINAL EXAM: All Topics (2PM-3:50PM, Location: TBA)	