Emory University
Department of Quantitative Theory and Methods

QTM 310 – Spring 2023
Introduction to Data Justice
M/W 10-11:15am, Math & Science Center N304

Instructors
Dr. Michal Arbilly (michal.arbilly@emory.edu)
Dr. Lauren Klein (lauren.klein@emory.edu)
Dr. Pablo Montagnes (pablo.montagnes@emory.edu)

Land Acknowledgment

Emory University is located on Muscogee (Creek) land. Emory was founded in 1836, during a period of sustained oppression, land dispossession, and forced removal of Muscogee (Creek) and Ani'yunwi'ya (Cherokee) peoples from Georgia and the Southeast. Emory owes an immense debt to the Muscogee, Ani'yunwi'ya and other original peoples, and their descendants, who have cared for and inhabited these lands.

Read the full Land Acknowledgment and History Statement developed by Emory faculty.

Instructor Office Hours

Arbilly: Thurs 10-11am, on zoom; schedule here by Wednesday night.
Klein: M/W, 11:30am-1pm; Tu, 4-5pm; schedule here.
Montagnes: Tu, 9-10am, 4-5pm; Th 10-11am in Tarbutton 311

Course Description

This course, team-taught by faculty spanning the various areas represented in QTM, will introduce students to both theoretical and methodological issues related to *data justice*. This emerging field considers how questions about data, its collection, and its use, are connected to broader social and political concerns, and how data-driven systems can be designed more equitably. Such data is expansive and expanding, and serves as the basis for automated systems that range from resume screening to voting redistricting, predictive policing to cell-phone autocomplete.

A central theme of the course is that choices (and trade-offs) are ubiquitous when bringing data to bear on technical and policy decisions. Few, if any, meaningful measures are truly "theory-free," in the sense that a measure is (perhaps implicitly) measuring *something*, and in many cases, this something is latent and not directly observable. Furthermore, even seemingly objective algorithmic systems may not be "neutral" in their effects: many of these systems rely upon data or were designed to achieve goals that reflect existing biases and inequalities embedded in the world.

Upon completing this course, students will be able to define and discuss the concepts of bias, fairness, discrimination, ethics, and justice, with respect to data science, and will gain familiarity, via case studies and labs, with how these concepts play out in data-driven inquiry.

Required Course Materials
All required readings will be posted on Canvas and/or are available online.

List of Graded Assignments
Your grade for the course will be calculated as follows:

- Class participation and in-class assignments - 10%
- Discussion questions – 5%
- First lab assignment – 5%
- Three unit tests – 25% each
- Final essay – 10%

Description of Graded Assignments
Class Participation
"Class participation" is often assumed to be a hazy concept, but it actually involves a careful assessment in five distinct areas. Here are short descriptions of each of these areas, adapted from grading criteria developed by Dr. Mark Sample of Davidson College:

- **Preparation:** Reading/reviewing any assigned material before class.
- **Presence:** Being verbally and nonverbally engaged during class.
- **Focus:** Avoiding distractions during class (both in person and online).
- **Asking questions** in class and in office hours, as well as via email when appropriate.
- **Specificity:** Referring to specific ideas from readings and prior class discussions when contributing to class discussion and/or in conversations during office hours.

In-class Assignments
During some lessons, you will work in groups on an exercise or problem, and turn in your solution on the provided worksheet for completion credit. Assignments can only be turned in by students attending the lesson and cannot be completed otherwise. To receive full credit for this grade component, you will need to complete no less than 80% of in-class assignments throughout the semester: i.e., if you complete 80% of assignments, you will receive 100% in this grade component; if you complete 70% of in-class assignments, you will receive 87.5% in this grade component, etc.

Discussion Questions
Each unit will culminate with an unscripted discussion among the three professors and the class. In order to shape the contents of these discussions, you will each be required to submit a potential discussion question via Canvas *no later than 24 hours before the class discussion* is scheduled to take place. For more detailed instructions on how to ask a good discussion question, please consult the handout posted on Canvas.

First Lab Assignment
The first lab assignment is intended to introduce you to the software, platforms, and programming language that will be used for the remainder of the course. The first lab assignment must be completed individually and submitted via Canvas.

Unit Tests
Each unit test will consist of a lab assignment followed by a series of questions that will ask you to reflect, in writing, on the lab assignment and its connection to the broader themes of the unit and course. Unit tests must be completed individually and submitted via Canvas.

Final Essay
The final essay will ask you to reflect on the course as a whole, offering your thoughts on the topics discussed in the course and your recommendations for future iterations. The final essay must be completed individually and submitted via Canvas.

Grading Process
All assignments will be graded on a 0-100 point scale. At the end of the semester, assignments will be weighted as listed above, and the weighted final grade will be converted to a letter grade according to the following cutoffs: A: 93; A-: 90; B+: 87; B: 83; B-: 80; C+: 77; C: 73; C-: 70; D+: 67; D: 60; F: any grade below 60.

Contacting your Professors
Please contact the appropriate lead professor for questions about a particular unit/day. (Information about lead professors can be found on the class-by-class schedule later in this document). We respond to email M-F 9am-5pm, and outside of those hours only if our schedules allow. Please allow 24 hours for a response, and 48 hours if your message is sent over the weekend.

For questions about grades or extensions, please also contact the lead professor for that unit test.

Policy on Late/Skipped Assignments
All assignments are mandatory. Should you submit an assignment after the due date, your grade for that assignment will decrease by 10 points for each day that it is late (e.g. 100 becomes 90). Should you fail to submit an assignment entirely, you will receive an F on that assignment. Should you need an extension, please contact us *in advance* to discuss your situation.

Attendance and Punctuality
In ordinary years, we allow three excused absences, no questions asked, with your grade beginning to be lowered with the fourth absence. However, due to the ongoing pandemic, we do not want to pressure you to come to class if you might be sick. Therefore, we will allow unlimited absences in this course.

With that said, you are responsible for finding out what was discussed in the course on any days that you miss; we do not provide copies of our lecture notes. In addition, beginning with the fourth absence, you must email us to let us know that you will be missing class for health reasons.

Finally, please be respectful to your fellow classmates and arrive on time. If you arrive more than 15 minutes late, you will be considered absent for that class.

Office of Accessibility Services
The Office of Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact OAS. It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed. Students registered with OAS who have a letter outlining their academic accommodations, are strongly encouraged to coordinate a meeting time with your professor that will be best for both to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible. Students must renew their accommodation letter every semester they attend classes. Contact the Office of Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html.

Writing Center and ESL Program
Tutors in the Emory Writing Center and the ESL Program are available to support Emory College students as they work on any type of writing assignment, at any stage of the composing process. Tutors can assist with a range of projects, from traditional papers and presentations to websites and other multimedia projects. Writing Center and ESL tutors take a similar approach as they work with students on concerns including idea development, structure, use of sources, grammar, and word choice. They do not proofread for students. Instead, they discuss strategies and resources students can use as they write, revise, and edit their own work. Students who are non-native speakers of English are welcome to visit either the Writing Center tutors or the ESL tutors. All other students in the college should see Writing Center tutors. Learn more and make an appointment by visiting the websites of the ESL Program and the Writing Center. Please review tutoring policies before your visit.

Honor Code
The Honor Code is in effect throughout the semester. By taking this course, you affirm that it is a violation of the code to cheat on exams, to plagiarize, to deviate from the teacher's instructions about collaboration on work that is submitted for grades, to give false information to a faculty member, and to undertake any other form of academic misconduct. You agree that the instructor is entitled to move you to another seat during examinations, without explanation. You also affirm that if you witness others violating the code you have a duty to report them to the honor council. Students who violate the Honor Code may be subject to a written mark on their

record, failure of the course, suspension, permanent dismissal, or a combination of these and other sanctions. The Honor Code may be reviewed at: http://catalog.college.emory.edu/academic/policies-regulations/honor-code.html.

A Closing Note on COVID

By the time this course begins, it will have been nearly three years since the pandemic began. As your professors, we are operating under the assumption that we have all experienced our share of hardships, and that we are all—in different ways and to different degrees--exhausted. Our goal is to make this course useful and interesting, and not to contribute to our collective (and ongoing) struggles. If you are having trouble related to the course in any way, *please let us know*. By the same token, we will all need to adapt as the Covid situation continues to evolve. We ask for your flexibility and understanding should the format of the course or the assignments need to change in response to new or unexpected developments.

Class-by-Class Schedule

Class schedule and readings subject to change.
Please consult Canvas for the most current class schedule.
Current versions of all course labs can also be found on GitHub.

Unit 1: Introduction and Overview

Wednesday, Jan 11 – Course overview and professor introductions

Monday, Jan 16 – MLK Day, no class

Wednesday, Jan 18 – Conceptual overview

- Readings:
  - Boyd, D., and Crawford (2012) Critical questions for big data. *Information, Communication & Society* 15:5, 662-679.
  - D'Ignazio, C., and Klein, L. F., "What Gets Counted Counts", in *Data Feminism*, MIT Press 2020.

Monday, Jan 23 – Intro to JupyterHub and Jupyter Notebook

- No assigned readings; bring laptop to class

Weds, Jan 25 – Lab Day

- No assigned readings; bring laptop to class

Unit 2: Fairness and Normative Goals– Montagnes

Monday, Jan 30 – Defining Algorithms and Goals

- Reading:
  - Sen, Amartya, 1970. "The Impossibility of a Paretian Liberal," *Journal of Political Economy*. 78 (1): 152–157.

Wednesday, Feb 1 – Ethical and Normative Frameworks Basics

- Readings:
  - Bueno de Mesquita, Ethan , "Normative Political Theory"
- **Due:**
  - Intro Lab assignment

Monday, Feb 6 – Ethical Frameworks Conflicts and Resolutions

- Readings:
  - *Fair Division: From cake-cutting to dispute resolution,* Brams & Taylor, 1996.

Wednesday, Feb 8–Algortithmic Fairness Definions

- Readings:
    - Mitchell et al., "Algorithmic Fairness: Choices, Assumptions, and Definitions," *Annual Review of Statistics and its Application* (2021).
    - Corbett-Davies et al., "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning," *arXiv preprint*, (2018).
    - Chouldechova, "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments," *arXiv preprint*, (2017).
    - Angwin et al., "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks," *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Monday, Feb 13–Algortithmic Fairness Conflicts and Impossibilities

Wednesday, Feb 15–Lab/Case Study: Using Machine Learning to Target Tax Audits

- Readings:
    - Battaglini et al, "Refining Public Policies with Machine Learning: The Case of Tax Auditing"
- **Due**
    - Unit Test #2

Monday, Feb 20 – Discussion

- **Due:**
    - Submit discussion question via Canvas by 10am on Tues, Feb 19th

Unit 3: Justice – Klein

Wednesday, Feb 22 – Conceptual framework

- Readings:
    - Catherine D'Ignazio and Lauren Klein, "The Power Chapter"
    - Charles Mills, "'Ideal Theory' as Ideology," *Hypatia* 20.3 (2005): 165-184 (Canvas).
    - Optional: Ben Green, ""Data science as political action: Grounding data science in a politics of justice"

Monday, Feb 27 – Predictive Policing

- Readings:
    - Cathy O'Neil, "Civilian Casualties," from *Weapons of Math Destruction* (Crown, 2016) (Canvas)
    - Optional: D'Ignazio and Klein, "Collect, Analyze, Imagine, Teach," from *Data Feminism*

- Optional: Chris Gilliard and Hugh Culik, "Digital Redlining, Access, and Privacy" and WaPo profile of Chris Gilliard and his work

Wednesday, March 1 – Facial Recognition

- Readings:
  - Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research 81*: 1–15 (2018) (Canvas)
  - Simone Browne, "Notes on Surveillance Studies," from *Dark Matters: On the Surveillance of Blackness* (Duke UP, 2015) (Canvas)
  - Optional: Blaise Aguera y Arcas et al., "Physiognomy's New Clothes," *Medium* (2017)

SPRING BREAK – NO CLASS

Monday, March 13 – Lab Day

- No assigned readings; bring laptop to class

Wednesday, March 15 – Imagining Otherwise

- Readings:
  - Ruha Benjamin, "Retooling Solidarity, Reimagining Justice," from *Race after Technology: Abolitionist Tools for the New Jim Code* (Pollity, 2019) (Canvas)
  - Algorithmic Justice League (explore website)
  - Karen Gregory, "Worker Data Science Can Teach Us How to Fix the Gig Economy," *WIRED*, December 7, 2021.

Monday, March 20 – Current Challenges (Large Language Models and ChatGPT)

- Readings:
  - Kevin Roose, "The Brilliance and Weirdness of ChatGPT," *The New York Times*, December 5, 2022.
  - Emily Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (March 2021): 610-623 (Canvas).
- **Due:**
  - Unit Test #3

Wednesday, March 22 – Discussion

- **Due:**
  - Submit discussion question via Canvas by 10am on Tues, March 21st

Unit 4: Diversity and Equity – Arbilly

Monday, March 27 – Genetic Diversity

- Readings:
    - Dawkins, The Selfish Gene, Chapters 2 and 3. 40th anniversary edition. Oxford University Press (2016).

Wednesday, March 29 – Heritability and the Nature-Nurture debate

- Readings:
    - Zimmer, "Nine foot high complete", from *She Has Her Mother's Laugh*, Dutton 2018.
    - Zimmer, "Ed and Fred", from *She Has Her Mother's Laugh*, Dutton 2018.

Monday, April 3 – Hereditarianism: Old and New

- Reading:
    - Feldman, M. W. (2014). Echoes of the Past: Hereditarianism and A Troublesome Inheritance. *PLoS Genetics* 10(12): e1004817.

Wednesday, April 5 – Forensic Genetics

- Reading:
    - Jobling, M.A. (2022). Forensic genetics through the lens of Lewontin: population structure, ancestry and race. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377: 20200422.

Monday, April 10 – Gendered Brains

- Reading:
    - Joel, D., et al. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences of the United States of America* 112(50): 15468–15473.

Wednesday, April 12 – Lab Day

- No assigned readings; bring laptop to class.

Monday, April 17 – Discussion

- **Due:**
    - Submit discussion question via Canvas by 10am on Tues, April 16th

Course Wrap-Up

Wednesday, April 19 – Semester Review and Course Evaluation

- No assigned readings; bring laptop to class
- **Due:**
    - Unit Test #4

Monday, April 24 – Final Discussion

- **Due:**
    - **Final essay!**
    - Extra credit: Submit discussion question via Canvas by 10am on Tues, April 20th

NB: There is no final exam for this course.