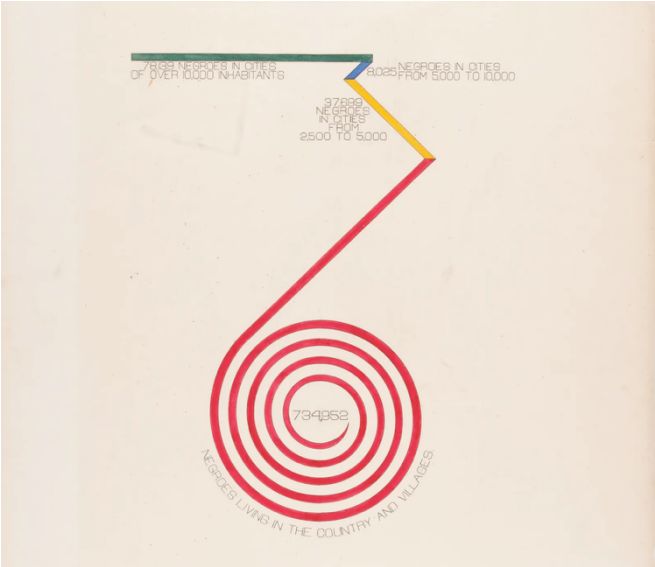
INST 878Y: Race & Data



Dr. Daniel Greene dgreene1@umd.edu

Class time: Tuesdays, 2-4:45 PM

Class location: JMZ 0208

Office hours: Mon/Thu by appointment
https://calendly.com/dgreene1/office-hours

Course site: ELMS

Race and racism are products of, inputs to, and organizing structures for the big data stores that power contemporary business and governance. There is a long history here. Actuarial data gained its shape during the transatlantic slave trade, and the Progressive Era saw a boom in measures of human performance and wellbeing deeply entangled with eugenics. Black scientists like WEB Du Bois explained these data structures as they emerged, and proposed alternative methods of collection, analysis, and visualization with an eye towards freedom. But the contemporary moment is different in its stakes and scope. This course will help doctoral students rise the to the challenge, exploring race and racism as input, output, and organizing principle of data structures.

As an interdisciplinary course in information studies, this class pushes students from computational fields to engage with history, sociology, and law, while students from the social sciences will engage with computer science, statistics, and the biosciences. Readings will mix foundational texts and cutting-edge research, theoretical debate and applied studies. The semester will begin by introducing students to scholarly debates on the definition and measurement of race and racism; then class will explore the historical development of big data stores in government and business, and the racial politics therein; following these foundational weeks, class will explore race and data in the theory and practice of domains such as work, health, education, and law. Throughout, we will attend to the intersections between race and racism and other social fields such as disability, sexuality, gender, and nationality.

Assignments encourage students to take ownership of sections of the course, to engage race and data through both prose and computation, and to produce scholarship that fits their course of study. Weekly blog posts will assess understanding of core themes. Weekly leaders will be responsible for providing intellectual-historical background and political context for readings. Class sessions will include both traditional seminar discussions of readings and hands-on experimentation with different analytical and visualization techniques. A ‘dataset biography’ will have students describe the racial politics of a specific large, publicly available dataset. The final project will ask students to produce either an annotated

bibliography that draws on topics in race and data to support their doctoral research, or a novel data visualization that explains some of the principles discussed in class.

Learning Outcomes

After completing this course, students will be able to…

- Situate novel developments in data collection, analysis, and circulation in historical and organizational context
- Explain and apply core concepts from the sociology of race and related fields
- Explain and apply core concepts from critical data studies and related fields
- Propose, design, and evaluate visualizations related to human subjects data
- Propose new research related to the collection of data on race and racism

Assignments

Participation -- 300 points [weekly]

Like most graduate seminars, this class does not work without regular, considered, and collegial participation from everyone involved. The instructor will manage discussion and provide context when necessary but by and large will not lecture. Please bring copies of all readings to class every day (print or digital), along with your notes on them. Students must complete all assigned readings prior to the start of class to effectively participate in class. Effective participation can look like:

- Asking open-ended questions to seek clarity or prompt discussion

- Quoting or paraphrasing a reading to make a point

- Collaborating on small-group activities

- Providing your analysis of an issue or reading, or responding to a colleague's

- Giving context colleagues may need on a particular point of discussion

It is not necessary to master every point of every reading to effectively participate. Indeed, it is sometimes most helpful to arrive to class with questions in mind. To receive full credit, students need only 1) come to class ready to share a question or discussion point (we will often start class by writing these on the board); 2) engage with their peers through one or more of the methods above. Only 12 of our 14 class sessions are graded for participation, meaning you have two "freebie" days. These can be used in lieu of providing an explanation for an excused absence, or just as a moment to recharge during (another) crisis year. There is no need to contact the instructor to request permission to use a freebie day.

Blogs -- 300 points [weekly]

Beginning in week 2, students will be required to write a weekly blog post of approximately 500 words that engages with some aspect of the assigned readings. The instructor will include prompts--generally around implications for category definition or data collection, analysis, and circulation--but students should write posts that are useful to them in processing these ideas. To encourage this process of reflection, posts are only graded for participation. Only 10 of a possible 13 blog posts are graded, meaning students can choose which to complete. All blog posts are due at midnight Monday, the night before class.

Dataset Biography -- 100 points [proposal due Oct. 7, report due Nov. 11]

Whether students' goals are to design better measurements of race and racism or better understand our experiences with them, the starting point is the same: Understanding what relevant datasets are already out there, how they work, and what work they're doing. Students will choose a dataset relevant to class and their own doctoral studies, with suggestions provided, and write a brief report detailing its creation, usage, structure, and reception. Students with quantitative training are encouraged to more directly measure internal biases, while students with qualitative training are encouraged to build an archive of sources that help describe its life story.

Show-and-Tell -- 100 points [assignments by Sept. 9, due dates vary]

Students will be assigned different weeks in class for which they must lead the class in an informal show-and-tell session, with an accompanying written brief. Students may choose to either review a recent book relevant to—a terrific first publication opportunity—or a new technology or dataset being used in that week's domain—a potential first draft of the dataset biography. We will create shared documents to provide inspiration for both options.

Final Project -- 200 Points [due Dec. 16]

The final project provides students with the opportunity to expand on course content in a way that serves their own professional goals. Students may choose to either write a literature review on a particular aspect of race and data (e.g., colorism in Brazil, facial recognition in policing) or, building off the week on visualization, create a visualization of a novel dataset relevant to the course.

Course Schedule

Note: Readings on the open web are hyperlinked and readings from books have PDFs provided. But journal articles must be found by students through lib.umd.edu, in order to give them experience navigating our databases and to produce the data necessary to justify these institutional subscriptions.

1) September 2: Introductions

2) September 9: Race

- Bonilla-Silva, E., 2022. "Color-blind racism in pandemic times." *Sociology of Race and Ethnicity*, *8*(3), pp.343-354.
- Fields, B.J. and Fields, K.E., 2022. "Chapter 1: A Tour of Racecraft." In *Racecraft: The soul of inequality in American life*. Verso Books.
- Mills (1997) "Overview" in The Racial Contract. Cornell University Press.
- Omi, M. and Winant, H., 1994. "Racial Formation. In *Racial Formation in the United States From the 1960s to the 1990s*, pp 53-76. Routledge.

3) September 16: Data

- Bouk, D., 2017. "The history and political economy of personal data over the last two centuries in three acts." *Osiris*, *32*(1), pp.85-106.
- Porter, T. (1995). "Cultures of objectivity" and "How social numbers are made valid." In *Trust in Numbers.* Princeton University Press.
- Rosenberg, D., 2013. "Data before the Fact." In *Raw Data is an Oxymoron*, pp. 15-40.
- Viljoen, S., 2021. A relational theory of data governance. *The Yale Law Journal*, pp.573-654.
- Yates, J. (1989). "Managerial methods and the functions of internal communication" in *Control Through Communication: The Rise of System in American Management,* pp. 1-20.

4) September 23: Research

- Zuberi, T. and Bonilla-Silva, E. eds., 2008. *White logic, white methods: Racism and methodology*. Rowman & Littlefield.
    - Zuberi & Bonilla-Silva "Introduction"
    - Gallagher "The 'end of racism' as the new doxa"
    - Holland "Causation and race"
    - Horton & Sykes "Critical demography and the measurement of racism"

5) September 30: Categories

- Bowker, G. and Star, S.L., 1999. "Some tricks of the trade in analyzing classification" and "The case of race classification and reclassification under apartheid." In *Sorting things out: Classification and its consequences*. MIT Press.
- McKay, Dwanna L. 2021. "Real Indians: Policing or Protecting Authentic Indigenous Identity?" *Sociology of Race and Ethnicity* 7(1):12-25.
- Panofsky, A. and Donovan, J., 2019. Genetic ancestry testing among white nationalists: From identity repair to citizen science. *Social studies of science*, *49*(5), pp.653-681.
- Telles, E. and Paschel, T., 2014. Who is black, white, or mixed race? How skin color, status, and nation shape racial classification in Latin America. *American Journal of Sociology*, *120*(3), pp.864-907.

6) October 7: Population

- Ruggles, S. and Magnuson, D.L., 2020. Census technology, politics, and institutional change, 1790–2020. *Journal of American History*, *107*(1), pp.19-51.
- López, I.C., 2020. Alien data: Immigration and regimes of connectivity in the United States. *Critical Ethnic Studies*, *6*(2).
- Ghosh, Arunabh (2020.). "Counting China." *Aeon,* July 23.
- Sahoo, S., 2023. Biometric data's colonial imaginaries continue in Aadhaar's minimal data. *BJHS Themes*, *8*, pp.205-220.
- Zimmerman, A., 1999. Anti-Semitism as skill: Rudolf Virchow's Schulstatistik and the racial composition of Germany. *Central European History*, *32*(4), pp.409-429

7) October 14: Fall break, no class, no blog

8) October 21: Crime

- Brayne, S., 2017. Big data surveillance: The case of policing. *American sociological review*, *82*(5), pp.977-1008.
- Muhammad, KG. 2010. Saving the nation: The racial data revolution and the negro problem. In *The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America.* Harvard University Press.
- Jefferson, B.J., 2018. Policing, data, and power-geometry: Intersections of crime analytics and race during urban restructuring. *Urban geography*, *39*(8), pp.1247-1264.
- Lavigne et al, 2017. White Collar Crime Risk Zones. *The New Inquiry.*
- Larregue, J. and Rollins, O., 2019. Biosocial criminology and the mismeasure of race. *Ethnic and Racial Studies*, *42*(12), pp.1990-2007.

9) October 28: Health

- Tsai, J., 2021. COVID-19 is not a story of race, but a record of racism—our scholarship should reflect that reality. *The American Journal of Bioethics*, *21*(2), pp.43-47.
- Possin, K.L., Tsoy, E. and Windon, C.C., 2021. Perils of race-based norms in cognitive testing: The case of former NFL players. *JAMA neurology*, *78*(4), pp.377-378.
- Creary, M.S., 2018. Biocultural citizenship and embodying exceptionalism: Biopolitics for sickle cell disease in Brazil. *Social Science & Medicine*, *199*, pp.123-131.
- **Roberts, D. 2011. "Color-coded pills" in *Fata invention: How science, politics, and big business re-create race in the 21st century.* The New Press.**
- TallBear, Kim. "Genomic articulations of indigeneity." *Social Studies of Science* 43, no. 4 (2013): 509-533.

10) November 4: Work

- Amrute, Sareeta. "Bored techies being casually racist: Race as algorithm." *Science, Technology, & Human Values* 45, no. 5 (2020): 903-933.
- Rosenthal, C.C., 2013. From memory to mastery: accounting for control in America, 1750–1880. *Enterprise & Society*, *14*(4), pp.732-748.

- Van Doorn, N., 2017. Platform labor: on the gendered and racialized exploitation of low-income service work in the 'on-demand' economy. *Information, communication & society*, *20*(6), pp.898-914.
- Greene, D., Beard, N., Clegg, T. and Weight, E., 2023. The visible body and the invisible organization: Information asymmetry and college athletics data. *Big Data & Society*, *10*(1), p.20539517231179197.
- Ray, V., 2019. A theory of racialized organizations. *American sociological review*, *84*(1), pp.26-53.

11) November 11: Education

- Chávez-Moreno, L.C., 2021. The problem with Latinx as a racial construct vis-à-vis language and bilingualism: Toward recognizing multiple colonialisms in the racialization of Latinidad. In *Handbook of Latinos and education* (pp. 164-180). Routledge.
- Crooks, Roderic. Productive myopia: Racialized organizations and edtech. *Big Data & Society* 8, no. 2 (2021): 20539517211050499.
- Nettles, M.T., 2019. History of testing in the United States: Higher education. *The ANNALS of the American Academy of Political and Social Science*, *683*(1), pp.38-55.
- Irby, D.J., 2018. Mo'data, mo'problems: Making sense of racial discipline disparities in a large diversifying suburban high school. *Educational Administration Quarterly*, *54*(5), pp.693-722.
- Park, J.J., Kim, B.H., Wong, N., Zheng, J., Breen, S., Lo, P., Baker, D.J., Rosinger, K., Nguyen, M.H. and Poon, O.A., 2025. Inequality beyond standardized tests: Trends in extracurricular activity reporting in college applications across race and class. *American Educational Research Journal*, *62*(2), pp.336-377.

12) November 18: Visualization

- Battle-Baptiste, W. and Rusert, B. eds., 2018. "Introduction" and Plates" in *WEB Du Bois's data portraits: Visualizing black America*. Chronicle Books.
    - Access via UMD Libraries' subscription to Proquest eBook Central so as to see the Plates in full color
- Loukissas, Y.A. 2019 "Visualizing the Social." In *The DigitalSTS Handbook*, edited by Vertesi, Janet, David Ribes, Carl DiSalvo, Laura Forlano, Steve Jackson, Yanni A. Loukissas, Daniela Rosner, Hanna Rose Shell. Princeton: Princeton Press.
- Housing Justice League. 2017. "Beltlining: Gentrification, broken promises, and hope on Atlanta's southside."
    - Supporting visualization: Loukissas, Y.A, "A BeltLine for all?"
- Gray, V et al. "Data colonialism in Canada's chemical valley: Aamjiwnaang First Nation and the failure of the Pollution Notification System." Yellowhead Institute.
        - Map produced as part of the same project
- Onuha, M. 2020. "When proof is not enough." FiveThirtyEight, July 1.

13) November 25: Design [no class, but still blog]

- Benjamin, R., 2019. “Default discrimination” and “Retooling solidarity, reimagining justice” in *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.
- Brown, T.H., Homan, P. and Ray, V., 2025. Advancing the scientific study of structural racism: concepts, measures, and methods. *Annual review of sociology*, *51*.
- Laufer, B., Raghavan, M. and Barocas, S., 2025. Fundamental Limits in the Search for Less Discriminatory Algorithms—and How to Avoid Them. In *NeurIPS 2024 Workshop on Regulatable ML*.

14) December 2: Prediction

- Friedler, S.A., Scheidegger, C. and Venkatasubramanian, S., 2021. The (im)possibility of fairness: Different value systems require different mechanisms for fair decision making. *Communications of the ACM*, *64*(4), pp.136-143.
- Chasalow, K. and Levy, K., 2021, March. Representativeness in statistics, politics, and machine learning. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 77-89).
- Harcourt, B.E., 2015. Risk as a proxy for race: The dangers of risk assessment. *Federal sentencing reporter*, *27*(4), pp.237-243.
- Stop LAPD Spying (2023). *Automating Banishment.*
- Wang, A., Kapoor, S., Barocas, S. and Narayanan, A., 2024. Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. *ACM Journal on Responsible Computing*, *1*(1), pp.1-45.

15) December 9: Resistance

- Garvey, E.G., 2013. “facts and FACTS”: Abolitionists’ Database Innovations. *Raw Data is an Oxymoron,* pp. 89-102.
- Khovanskaya, V. and Sengers, P., 2019, June. Data rhetoric and uneasy alliances: Data advocacy in US labor history. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (pp. 1391-1403).
- Yanchapaxi, María Fernanda, and M. Murphy. “Indigenous Environmental Data Justice: Confronting Colonial Data and Activating Indigenous Sovereignty.” *Science, Technology, & Human Values* (2025)
- Harmon, Ellie, and M. Silberman. (2019). "Rating working conditions on digital labor platforms." Computer Supported Cooperative Work 28(5): 911-960
- Hirschman, D. and Bosk, E.A., 2020. Standardizing biases: Selection devices and the quantification of race. *Sociology of Race and Ethnicity*, *6*(3), pp.348-364.