Data 102, Spring 2024
Homework 6

Due: (partial) First 2 posts by **3:30 PM** Tuesday, December 3, 2024
(complete) All 4 posts by **5:00 PM** Friday, December 6, 2024

Instructions

Homework 6 will have a different format than the first 5 homeworks. Homework 6 is a reading assignment instead of a problem set. You will be asked to **complete two assigned readings** that pair to the lectures on November 26^{th} and December 3^{rd}, then to **complete a discussion activity** in your assigned groups. Group assignments will be released through email. The readings have been chosen to demonstrate the kind of analysis and reflection that you'll need for the concluding sections of your final project.

Please see the instructions below for the required reading and for the discussion activity. As for all other assignments, late submissions will count towards your slip days; it is your responsibility to ensure you have enough time to submit your work. As this assignment asks for less work than the previous five, it will be worth 50% the weight of the previous problem sets.

Reading

Required reading:

- Breznau, Nate, et al. *Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty.* Proceedings of the National Academy of Sciences 119.44 (2022): e2203150119. https://doi.org/10.1073/pnas.2203150119
- Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* St. Martin's Press, 2018. Chapter 4. The Allegheny Algorithm. https://drive.google.com/file/d/1KqWQQnbCLDB9R1b6T1wP9myb90KFt6ow/view?usp=sharing

Discussion Activity

You will be graded based on the completion of a discussion activity on Ed.

You will be assigned into a specific randomly chosen megathread, and you must participate under that megathread. Please watch for an announcement **in your email** containing your randomly chosen group assignment. Groups are numbered from 1 to 10. Your corresponding Ed megathread will be given a matching number.

Your HW6 group is completely unrelated to your project group.

In your assigned megathread, you should make four posts: each one should either ask a question, highlight a quote, raise a thoughtful comment, or reply to another student, using

the format guidelines below to mark which one it is. Following these instructions is important since the posts will be tallied automatically. If you do not include the [Category] text at the start of a post, it will not be counted.

- **Question:** Make a new thread under the megapost with a question about the reading. Provide any context for others to follow up. *You must include the text [Question] at the beginning of your post.*
- **Quote:** Make a new thread under the megapost that highlights a quote, and your reaction to it. *You must include the text [Quote] at the beginning of your post.*
- **Comment:** Make a new thread under the megapost with a comment about the reading. *You must include the text [Comment] at the beginning of your post.*
- **Reply:** Respond constructively to another student's thread. *You must include the text [Reply] at the beginning of your post.*

You should also include which reading assignment you're referring to (or both if your question/quote/comment/reply pertains to both of them), to help other students put your comments into context. For example, you might post:

> [Comment] Breznau: This is an example comment based on the reading. It mentions ideas and concepts from the reading, and then makes an interesting comment about the reading drawing on my perspectives and analysis of the situation. It demonstrates thought and reflection, and can be read and responded to by other students.

To which another student might respond:

> [Reply] Breznau: This might also be because of some of the other factors about the researchers, such as X, Y, and Z. In fact, on page 5... *(at least 50 words here)*

Policies:

- **Deadlines:** You must make 4 posts in total. At least 2 must be replies; the other two can be questions, quotes, comments, or replies. You must make at least 2 posts by 3:30 pm (i.e., the start of class) on Tuesday, December 3rd. The rest should be posted by 5:00 pm on Friday, December 5th.
- **Sincerity:** Please post sincerely. Flippant or automatically generated responses may be flagged and will not count toward completion.
- **Culture:** Be respectful to your classmates. Ed is a formal academic space, and as such, we expect you to maintain the same standards of behavior on Ed as you would in class.

  Some of the topics raised by this reading are difficult and relate to profound personal, familial, and social challenges. Please recognize that your peers' opinions and experience may differ from your own. Consider your peers and act with integrity.
- **Length:** All posts must exceed 50 words. 50 words is about three sentences. Note that for [Quote] posts, the 50 word minimum does not include the quote.
- **Location:** You must post in the Ed megathread that was sent to you by email and that matches your group number. Misplaced Ed posts will not earn credit.

Optional Follow-up Sources:

If you're interested in learning more after finishing the assigned reading, here are some interesting follow-up papers:

On reliability, following Breznau:

- **On the importance of stability in statistical models. How sensitive are our conclusions to reasonable perturbations in the data *or model* used?**

  Yu, Bin. *Stability.* (2013): 1484-1500. https://projecteuclid.org/journals/bernoulli/volume-19/issue-4/Stability/10.3150/13-BEJSP14.full

- **On best practices in data science to promote reliable research.**

  Yu, Bin. *Veridical data science.* Proceedings of the 13th international conference on web search and data mining. 2020. https://dl.acm.org/doi/abs/10.1145/3336191.3372191

- **On the variability in results produced by different analysts, even given the same data. How much do the results of a study depend on who performs it?**

  Gould, Elliot, et al. *Same data, different analysts: Variation in effect sizes due to analytical decisions in ecology and evolutionary biology.* EcoEvoRxiv (2024).https://egouldo.github.io/ManyAnalysts/

On the social impact of automated tools, following Eubanks:

- **On the interaction between human and machine decision making. How do automated tools interact with human expertise in real decision making settings?**

  Cheng, Hao-Fei, et al. *How child welfare workers reduce racial disparities in algorithmic decisions.* Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022. https://dl.acm.org/doi/abs/10.1145/3491102.3501831

- **On the opinions of individuals impacted by, or working with, automated risk prediction tools in child welfare. How do the individuals affected by a tool regard its use?**

  Stapleton, Logan, et al. *Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders.* Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. 2022. https://dl.acm.org/doi/abs/10.1145/3531146.3533177