

GESIS Fall Seminar in Computational Social Science 2022

Introduction to Computational Social Science with Python



Day 1-1: What is CSS?



Who we are

- Dr. Milena Tsvetkova
- Dr. Patrick Gildersleve
- And you?
 - Name and where you come from
 - Institution and discipline
 - Prior experience with programming, large data, computational methods
 - What you hope to take away from this course



Course outline

1. Computation for Social Science
2. Writing legible, modular, and optimized code
3. Obtaining data
4. Analysing Rectangular Data
5. Visualising data and analysing non-rectangular data



Why do social scientists need computation?





Why do social scientists need computation?

- Collect data from new sources
 - Crawling websites and using APIs
 - Online/mobile surveys and experiments
 - Computational models and simulations
- Manage, analyze, and visualize new forms of data
 - Large data
 - Non-rectangular data (e.g. networks, text, video, audio)
- Work in new ways
 - Be autonomous and work independently
 - Collaborate with natural scientists and engineers
 - Generate and share reproducible workflows



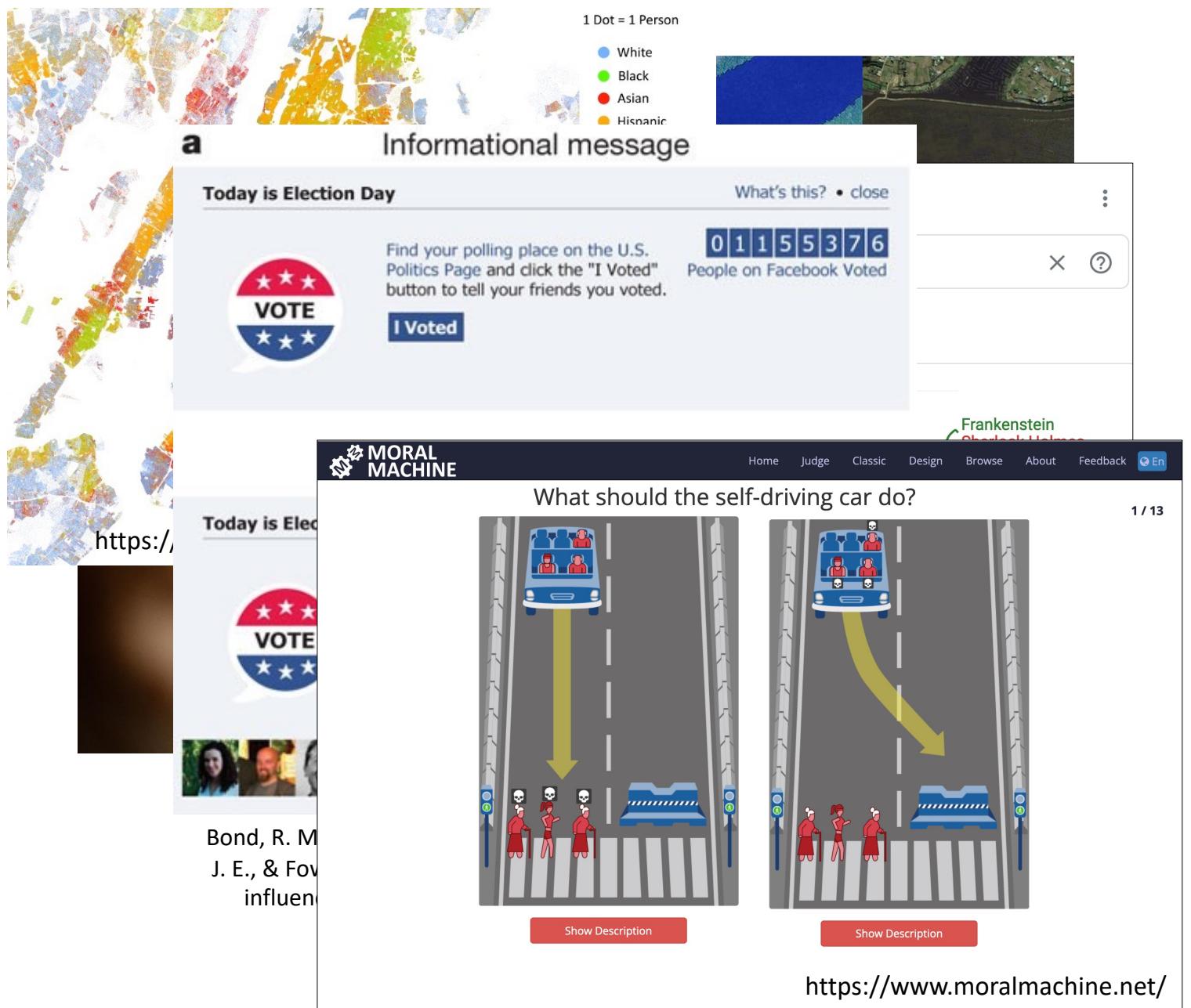
Computational Social Science (CSS)

- An academic sub-discipline that uses computational approaches to study social phenomena
- Populated by social scientists, computer scientists, and physicists
- Notable institutes/venues
 - [CSS team at GESIS Cologne](#)
 - [International Conference on Computational Social Science](#)
 - [Summer Institutes in Computational Social Science](#)
 - [International Society for Computational Social Science](#)



Data in CSS

- Big data
 - High volume
 - High velocity
 - High variability/complexity
- Sources
 - Digital trace data
 - Satellite imagery
 - Digitized historical archives
 - Census/register data
 - A/B testing
 - Crowdsourced participants





Methods in CSS

- Data mining – discovering patterns
- Machine learning – making predictions
- Causal inference – identifying causal relationships
- Modeling and simulation – proposing general principles

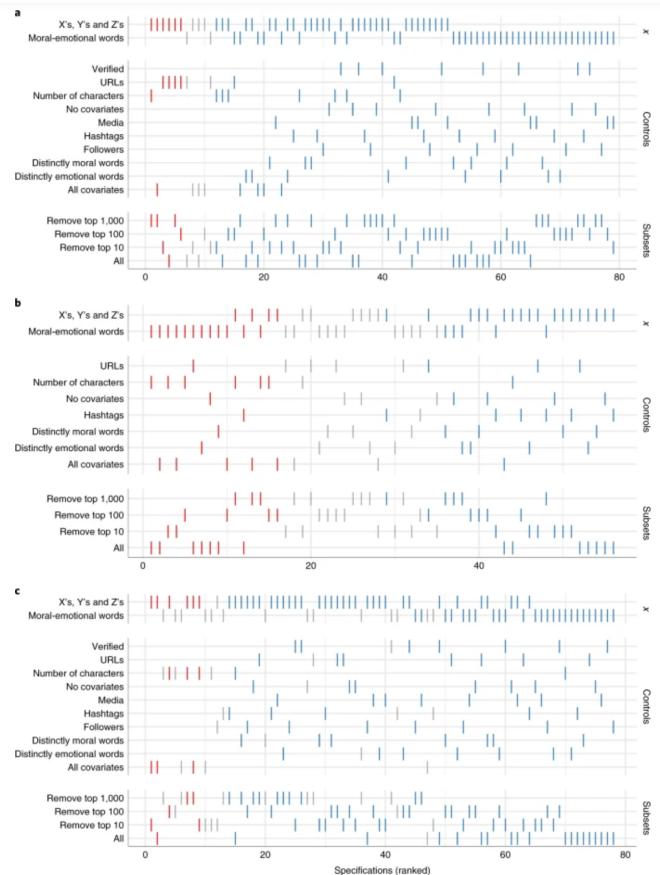


Questions in CSS

- Social influence
- Contagion
- Poverty and inequality
- Misinformation
- Political polarization

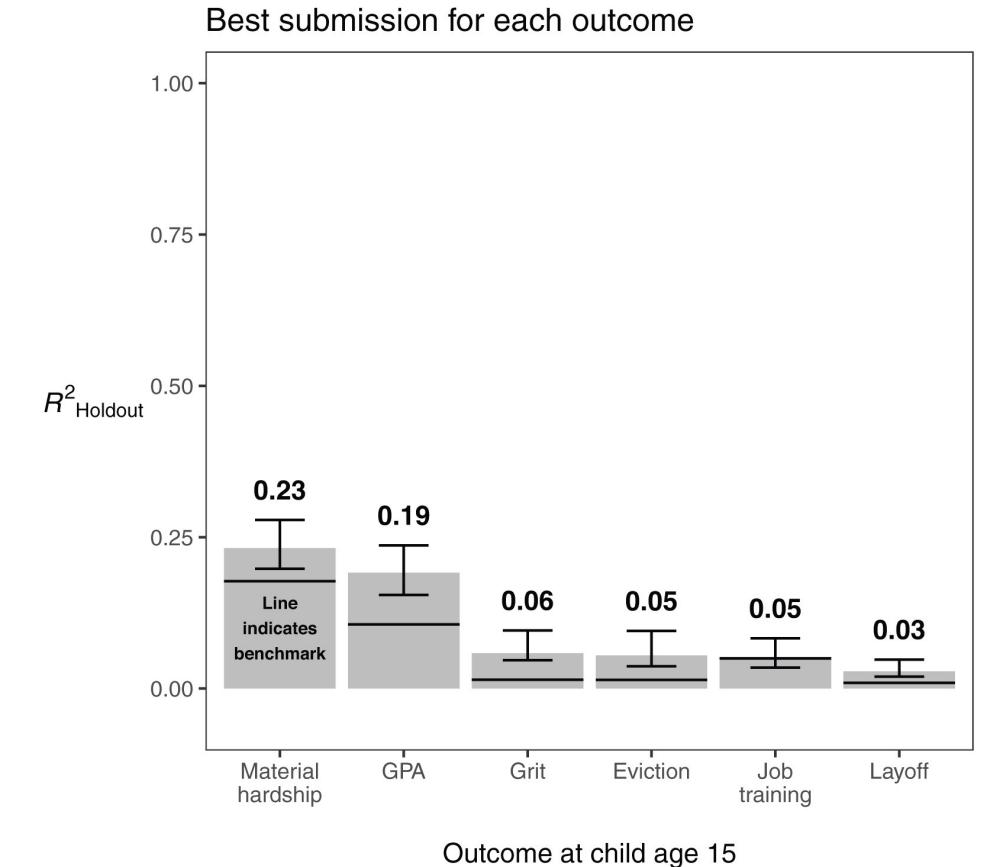


Problems in CSS: Reproducibility



a, COVID-19 corpus. **b**, #MeToo corpus. **c**, #MuellerReport corpus. Each possible model specification (x axis) is represented by three vertically aligned points corresponding to the outliers removed and covariates and independent variable included in the negative binomial regression equation (y axis). Red indicates a significant ($P < 0.05$) negative regression coefficient, grey indicates a non-significant coefficient and blue indicates a significant positive coefficient.

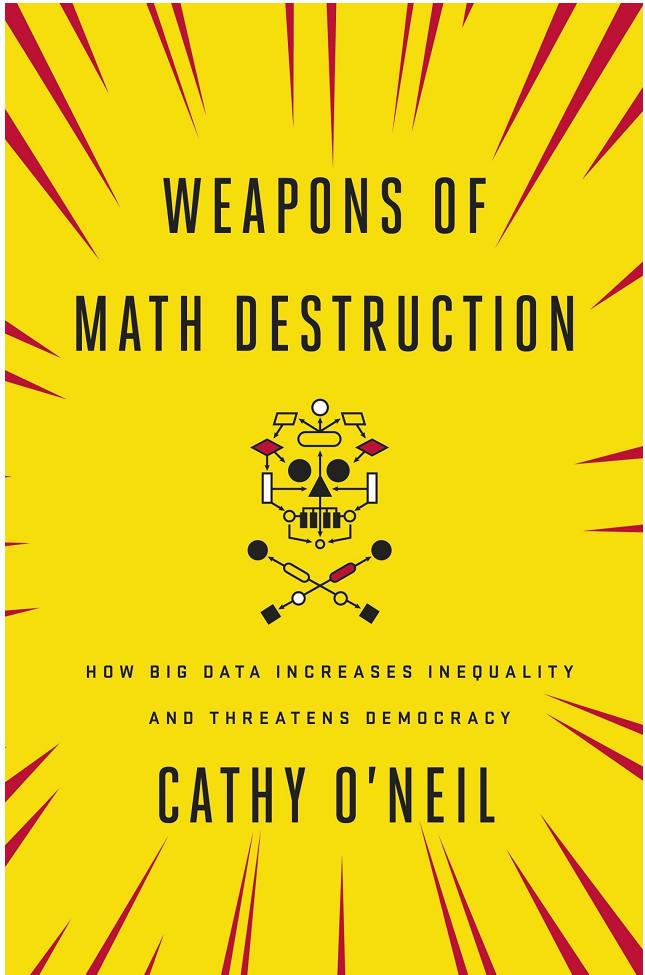
Burton, J. W., Cruz, N., & Hahn, U. (2021). Reconsidering evidence of moral contagion in online social networks. *Nature Human Behaviour*, 5(12), 1629-1635.



Salganik, M. J., et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398-8403.



Problems in CSS: Accountability



O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.

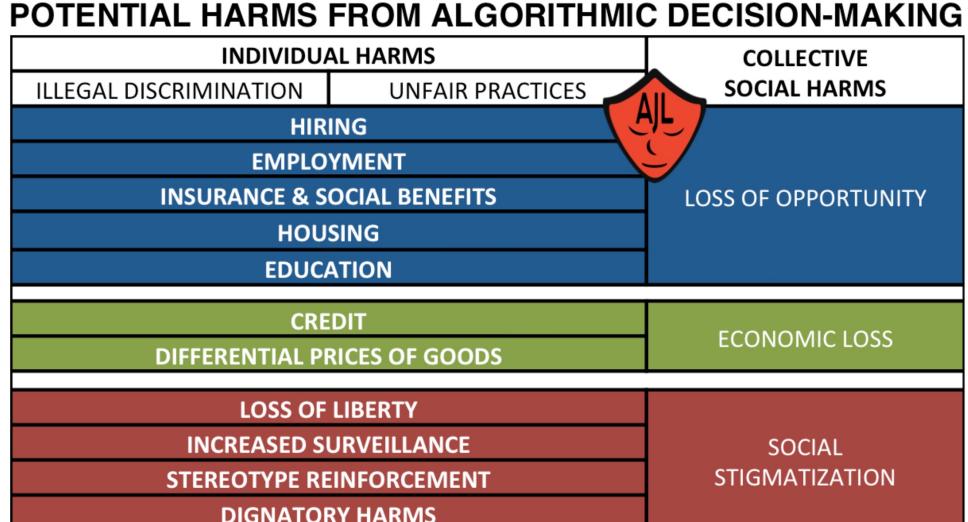


Chart Contents Courtesy of Megan Smith, Former CTO of the United States



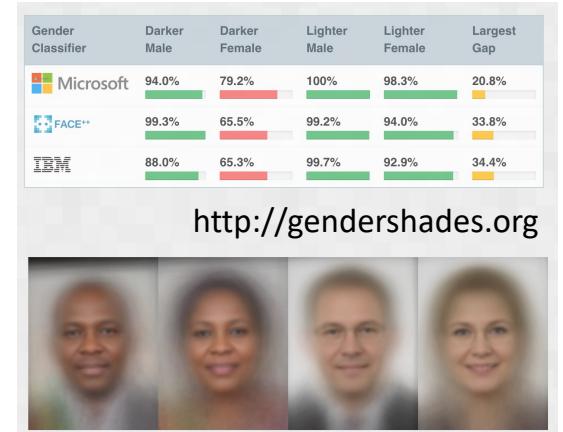
Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77-91). PMLR.

Rosalia
@BonKamona

I saw a tweet saying "Google unprofessional hairstyles for work". I did. Then I checked the 'professional' ones 😊😊😊

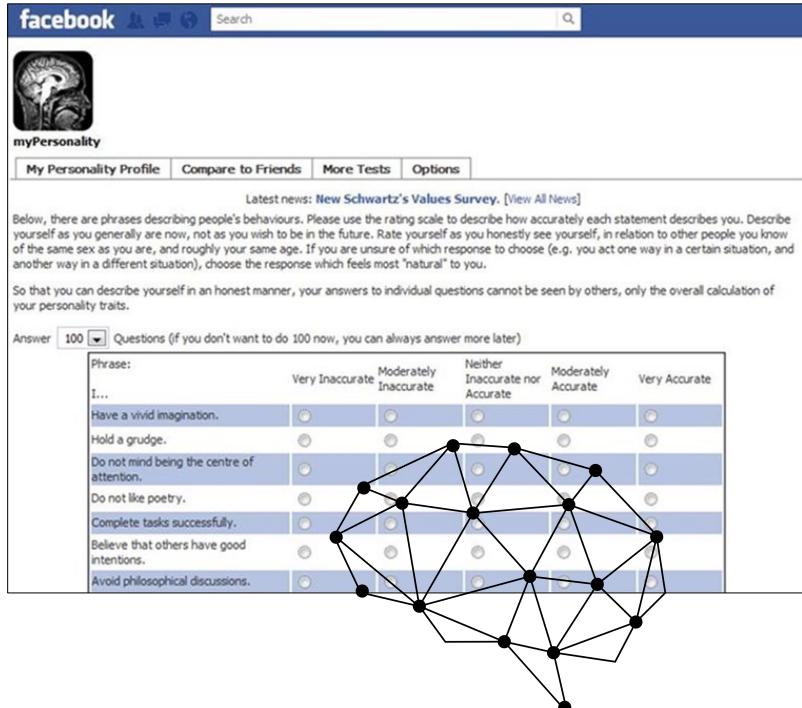
RETWEETS 141 LIKES 41

2:04 PM - 5 Apr 2016



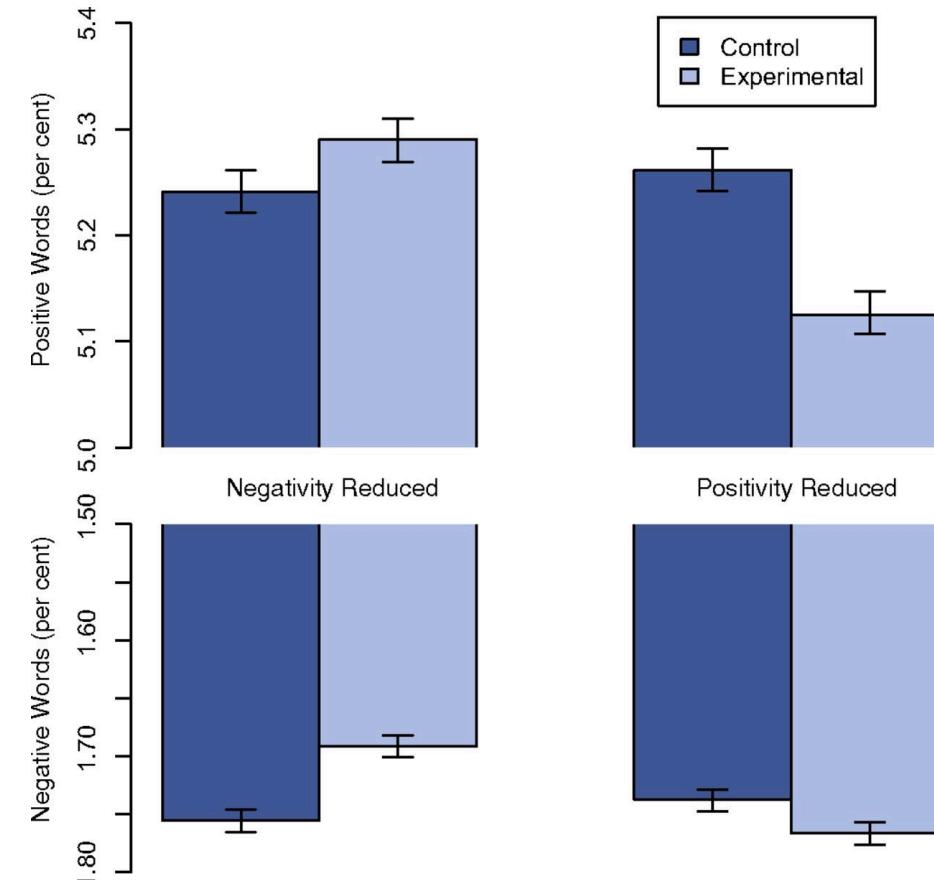


Problems in CSS: Research Ethics



Cambridge Analytica

Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*, 17, 22.



Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.