

PIT-UN 2022 - 04/2023

**Integrating Community
Knowledge and Core Resources
into a Data Science for Social
Impact Program at CWRU**



SASS 471: Introduction to Data Analytics for Social Impact, 3c, Spring Year 1

DSCI 451: Exploratory Data Science, 3c, Fall Year 2

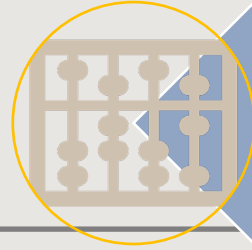
DSCI 453 Modeling and Prediction, 3c, Spring Year 2

SASS 4YY DSSI Research Project, 3c, Spring Year 2

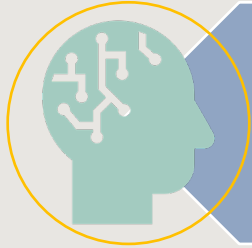
Offered jointly by Jack, Joseph and Morton Mandel School of Applied Social Sciences and Case School of Engineering

- **Help shape how technology can be used to advance social justice with a Certificate in Data Science for Social Impact from Case Western Reserve University.**
- Learn to use data-driven solutions to improve social welfare, while critically assessing the human impact of these digital technologies from an ethical perspective.
- With our specialized certificate, you'll be prepared for careers in consulting, philanthropy, journalism, local government, nonprofit management, urban planning and development, and more.

Product: Online platform with resources for DSSI project **with the goal of making it publicly available.**

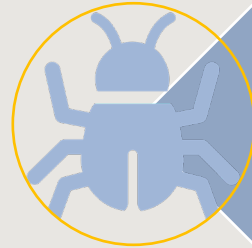


Using Administrative/Public Data



Guide to integrating community knowledge into data science projects.

Help students critically examine/challenge assumptions



Identifying/addressing bias in data analysis – C & P modeling.

Embedded within the Data Verification process applied in DSSI classes.



Ethics and fairness considerations

12 Community Collaborators

Community Partners:

Alice Jackson, Founding Director of ThinkingCapp.

Claire Gauntner, Senior Program Officer, Enterprise.

Chris Knestrick, NorthEast Ohio Coalition for the Homeless

Catherine Tkachyk, Chief Innovation & Performance Officer,
Cuyahoga County

Aysha Wilburn, Director of 2Gen Strategy – The Centers for Families
and Children

Student Collaborators:

Nicole Coury – MSASS and WSOM

Kelly Orr - Economics

Callie Ogland-Hand - Swetland Center
Graduate Assistant

Faculty/Staff Partners:

Laura Bruckman, Engineering

Shanina Knighton, Nursing

Francisca G.-C. Richter, MSASS

Brian Gran, Sociology

Claudia Coulton & **Joe Andre**, Poverty Center

Darcy Freedman & Rachael Sommer, Swetland Center

Andy Podgursky, Computer Science

Where Bias and Discrimination May Arise in Predictive Modeling

From Kleinberg et al. (2019).

Decision making goal / screening objective?

Who should be hired?

Choice of label or outcome variable Y (s5.1.1):

M_y : measured (y) versus desired Y

B_y : biases reflected in Y , y

Choice of predictors to collect, construct X (s5.1.2):

M_x : measured (x) versus desired X

B_x : biases reflected in X , x

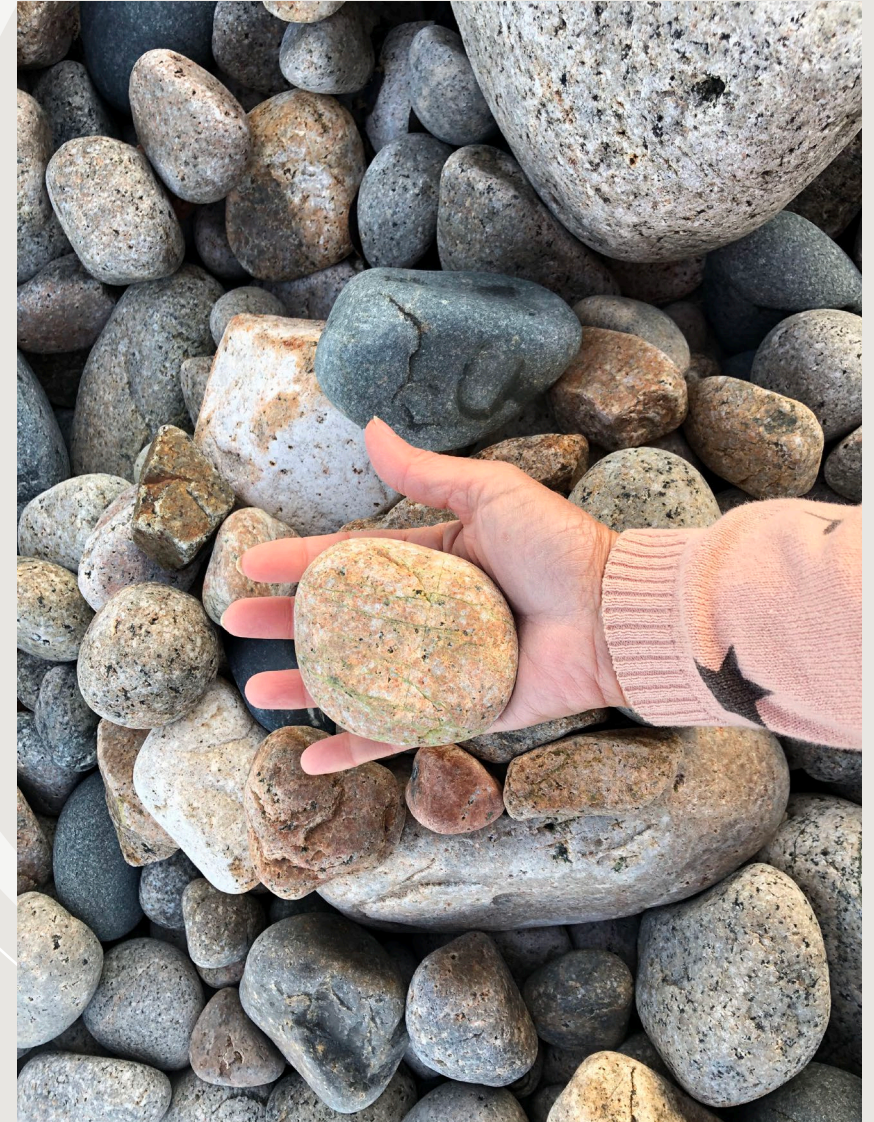
Choice of data or training sample (s5.1.3):

Selective Labels: for whom are labels Y missing?

Role of sensitive attribute A

How does data document A and how does representation vary across A ?

Identifying/addressing bias in data analysis



Causal Modeling to clarify assumptions and biases

Colliders explain bias when using administrative data.

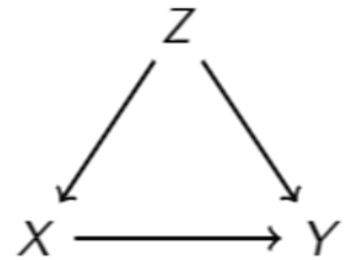
Zhao, Q. & Hastie, T. 2021. **Causal Interpretations of Black-Box Models**, Journal of Business & Economic Statistics, 39:1, 272-281, DOI: 10.1080/07350015.2019.1624293

Lakkaraju, H., Kleinberg, J., Leskovec, J. Ludwig, J., & Mullainathan, S. 2017. **The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables**. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining 23rd.



A Toolkit for causal modeling

- Confounding Bias
- Fork ($X \leftarrow Z \rightarrow Y$) induces spurious correlation between X and Y
- **Z** confuses the true causal relationship ($X \rightarrow Y$).
 - Z: season; X: ice cream sales; Y: petty theft
 - Z: age; X: going to the bar; Y: developing Covid-19 symptoms
- Solution: controlling or adjusting for Z
- Problem: Over-controlling without understanding causal hypotheses implied



Directed Acyclic Graph or DAG: Edges are directed & no feedback loops

Integrating community knowledge into data analysis – learning by doing

- Homelessness (measured in HMIS and ACS data)
- Food insecurity (measured in SNAP and ACS data)
- Held meetings with all community partners and faculty leads
- Completed process to hire our community partner Ms. Alice Jackson
- Established collaboration with students
- Attended Informed Consent training to develop IRB protocol
- Identified themes that seem relevant to stakeholders

- **Churn – SNAP program exit & reentry w/in a 4-month period.**

Integrating community knowledge into data analysis – Fall 2022 collaborative meetings

Understanding the Data: Administrative processes in this system collect data on XYZ.

[Show elements from data]. Based on your experience,

- Is the intake of this data necessary to provide the appropriate services?
- What is missing and what may be over emphasized?
- Where do you think the information will be unreliable?

Relevance of Research Questions: from literature

- Are these relevant questions; others?

Graphical Modeling: focusing on main research question, using graphical tools.

- Based on your experience, what are the main factors that influence Y?

While we are interested in understanding Y for all stakeholders, we only see data for people who interacted with the system. Who do you think we would miss and why?





Your feedback is essential to the project!