

Data Ethics: Understanding Big Data, Algorithmic Bias, and Research Ethics

Cara Marta Messina and Jeff Sternberg
CRIM 3600 Research Methods
Megan Denver
Fall 2019



Northeastern University
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point
during the presentation!*

Discussion

When you are doing research—whether you are researching for a class, your career, or personal interest—what are some ethical guidelines you follow?



Workshop Agenda

- Objectives
- Introduce 'Big Data' Concepts
- Activity: Animal or Plant?
- Algorithmic bias and the criminal justice system
- Research ethics

Slides, handouts, and data available at <http://bit.ly/33xzyUr>



Workshop Goals

- Understand the ways in which technologies reflect cultural, social, and political biases.
- Explore the basic process for machine learning algorithms
- Understand the ways data is being used in society as well as how algorithms impact and shape our daily lives and the criminal justice system
- Explore the ways in which these questions and methods are influencing how humanists and social scientists do research and practice their craft



What is 'Big Data'?

Big data has been called the 'new oil' by some.

Shoshana Zuboff argues that we now live in an era of 'surveillance capitalism.'

The four components of big data are: **volume**, **variety**, **velocity** and **veracity**



Why should we care?

- Big data is characterized by its **scale**
- Big data **sources** include: digitized records, social media/internet activity, or sensors from the physical environment.
- Big data is often **privately owned**
 - Example: an insurance company purchasing social media activity from facebook in order to make specific insurance sales decisions.



Ethical Implications

- Cambridge Analytica Controversy
- Big data also raises questions of autonomy, anonymity, privacy, discrimination, and bias.
- Questions to consider:
 - How are we being represented online?
 - How is our data being used?
 - Who is using it and for what purposes?
 - How might it be used in the future?



DIY Cybersecurity and Tightening your Privacy

Want to make your life more private? Follow this “DIY Guide to Feminist Cybersecurity”

<https://hackblossom.org/cybersecurity/>



Algorithms

Big data relies on the collection of high amounts of information and **algorithms** to parse through, categorize, and “read” that information.

Algorithms are a set of procedures to be followed by certain technologies (computers, cell phones, etc). Algorithms typically rely on data and a set of instructions to “read” that data in some way.



**So what do algorithms have to
do with the criminal justice
system?**



Northeastern University
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point
during the presentation!*

Risk Assessment: Algorithmic Bias

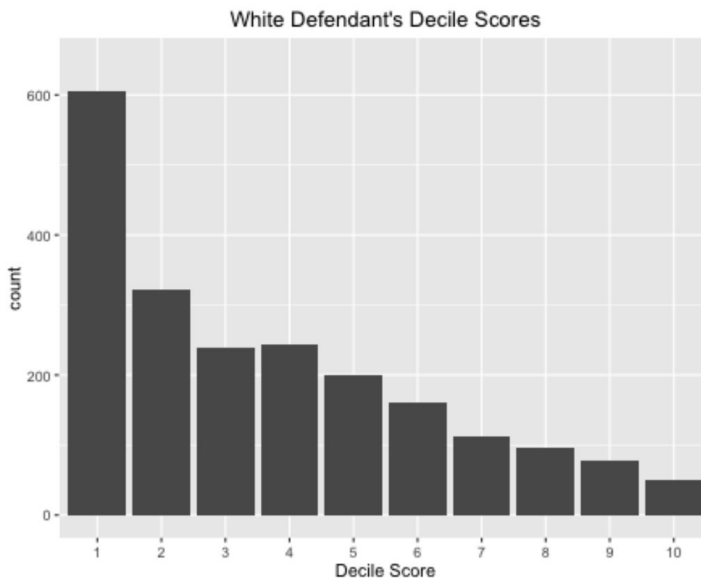
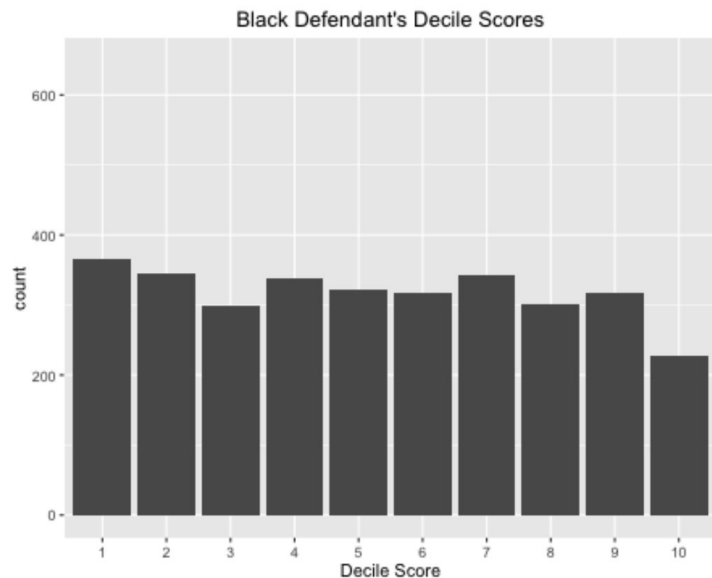
Risk assessment: used to determine the likelihood that someone will reoffend, not appear for trial, etc..

What happens when machine learning algorithms are used to help determine risk assessment?



COMPAS Algorithm & ProPublica's Analysis

The COMPAS recidivism algorithm does not “see” race. Yet...



<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>



Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Algorithmic Bias



Northeastern University
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point
during the presentation!*

Class Discussion

Based on the readings (the ProPublica article and the FAT/ML site) and your own knowledge:

- What is your opinion using risk assessment algorithms? In what ways are they beneficial and/or harmful?
- Based on the ProPublica analysis, what are some recommendations you might have for those in the judicial system making decisions?
- What are the best and worst practices for decision making?



So what can we do?



Northeastern University
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point
during the presentation!*

Questions Researchers Must Ask

- What **information** is being collected and from where? To whom does this data **belong**?
- How is it being **collected**? Do **participants** know that it is collected, how it will be collected, and how will it be used?
- **How** will the data be analyzed? What **biases** and **ideologies** may be implicit in this analysis?
- Who will this research impact? Who will it **benefit**? Who will it potentially **harm**?



Thank you!

If you have any questions, contact us at:

Cara Marta Messina

Digital Teaching Integration

Assistant Director

messina.c@husky.neu.edu

Jeff Sternberg

Digital Teaching Integration

NULab Research Fellow

sternberg.je@husky.neu.edu

Slides, handouts, and data available at <http://bit.ly/33xzyUr>

Office Hours: **Tuesdays from 1–3PM in 401 Nightingale Hall**



Northeastern University

NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point
during the presentation!*