

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

Jeffrey Sternberg
ENGL 1111 :First Year Writing
Emily Avery Miller
Spring 2020



Northeastern University
NULab for Texts, Maps, and Networks

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

Mini Get-to-know-you Activity

What are some recent advertisements you have seen pop up on your Snapchat, Instagram, Facebook, Youtube, and other online spaces?

- Which advertisements interested you?
- What are some surprising advertisements?
- Why do you think you received those advertisements?



Workshop Agenda

- Objectives and Goals
- Introduce 'Big Data' Concepts
- Algorithmic bias and policy implications
- Algorithm activity
- Research ethics

Slides, handouts, and data available at

<http://bit.ly/diti-spring2020-averymiller>



Workshop Objectives

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



What is 'Big Data'?

Big data has been called the “new oil” by some, including Andrew Yang.

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**



Big Data: What is it and why should we care?

- Big data **sources** include: digitized records, social media/internet activity, and 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 insurance sales decisions.
- Big data can often reproduce results that may harm certain communities.



Google's File on You is 10 Times Bigger Than Facebook's — Here's How to View It

Google, Amazon, Apple, and Microsoft are all central players in “surveillance capitalism” and prey on our data.



Example: If you have **location services** turned on for Google (like if you use Google maps), Google can track your every move. Go to:

<https://www.google.com/maps/timeline>



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?
 - If I use big data sources in my research, what ethical issues must I think about? Is my big data source representative?



DIY Cybersecurity and Tightening your Privacy

Want to make your life more private? Follow this guide:

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



Criminal Justice Example:

COMPAS Risk Assessment Algorithm



Northeastern University
NULab for Texts, Maps, and Networks

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

Risk Assessment: Algorithmic Bias

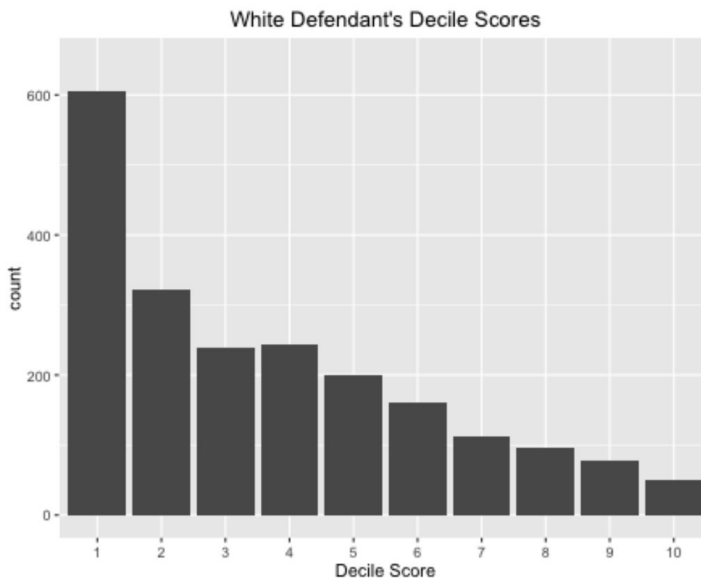
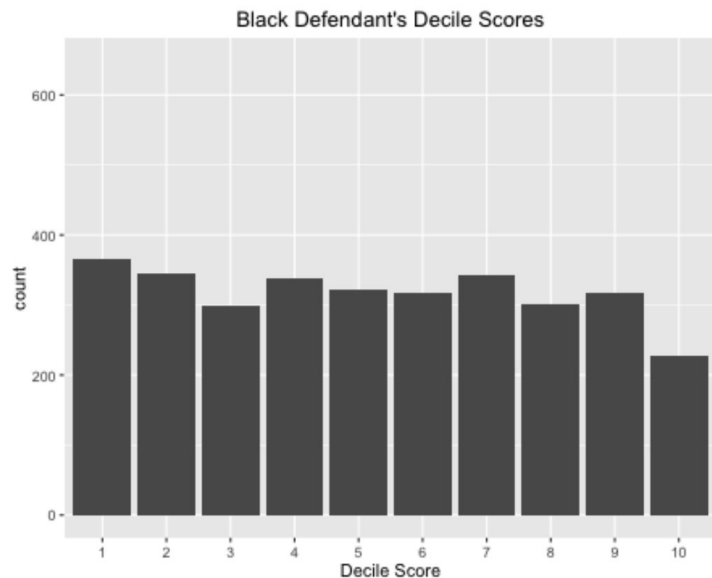
Risk assessment: used to determine the likelihood that someone will reoffend, not appear for trial, etc..

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!

COMPAS Algorithm & ProPublica's Analysis

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

Prediction Fails Differently for Black Defendants

| | WHITE | AFRICAN AMERICAN |
|---|-------|------------------|
| Labeled Higher Risk, But Didn't Re-Offend | 23.5% | 44.9% |
| Labeled Lower Risk, Yet Did Re-Offend | 47.7% | 28.0% |

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

<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!

Algorithms and Bias



Northeastern University
NULab for Texts, Maps, and Networks

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

Class Discussion

Open Google's search engine and type in "Greatest authors of all time."

- What are some of the results? What do you notice about these results?
- How many authors on this list have you read? Do you agree with the list?
- Where do you think these results came from?
- What do these results suggest to you in terms of defining "greatest" and "authors"?



Class Discussion

Now try these results:

- Greatest women authors
- Greatest Black women authors
- Greatest Black authors
- Greatest white authors

“Black” leads to substantial results, while “white” does not. Why do you think this might be?



Technology is Not Neutral

Do white men really make better authors than white women, Black women, and Black men? According to the literary canon and Google, yes, they do! But should we just accept this?

Information systems like Google as well as data collection, data analysis, and algorithms are **not neutral**. They can reinforce and make explicit systemic, political, and cultural biases. They are affected by input data, the way that data is presented, how the data is interpreted by machines, and more. This means we also have the ability to challenge these biases, norms, and forms of discrimination.



OPTION: Watch this video



Safiya Noble's "Challenging the Algorithms of Oppression"



Northeastern University
NULab for Texts, Maps, and Networks

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

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**?



Using Big Data in Your Own Research: Opportunities and Vulnerabilities

Opportunities

- Massive & Passive
- “Nowcasting”
- Data on social systems themselves
- Natural and Field Experiments
- Making big data small

Vulnerabilities

- Generalizability
- Too many big data
- Artifacts, reactivity, and drift
- Ideal user assumption

Source: Lazer and Radford, 2017



Zotero for Research: Insights into Managing Sources

Jeffrey Sternberg
ENGL 1111 :First Year Writing
Emily Avery Miller
Spring 2020



Managing Sources Class Discussion

Break up into small groups and talk amongst yourselves for five minutes; choose one person to report back to the larger group. Think about these questions:

- How do you manage your sources when you are doing research?
- Where do you look for sources?
- How do you save your sources?
- How do you remind yourself what each source is about and how you plan to use it?



Workshop Agenda

- Objectives
- Why are citation management tools useful?
- Introduction to Zotero
- Important Zotero vocabulary
- Tutorial
- In-class hands on activity with Zotero

Slides & handout available at

<http://bit.ly/diti-spring2020-averymiller>



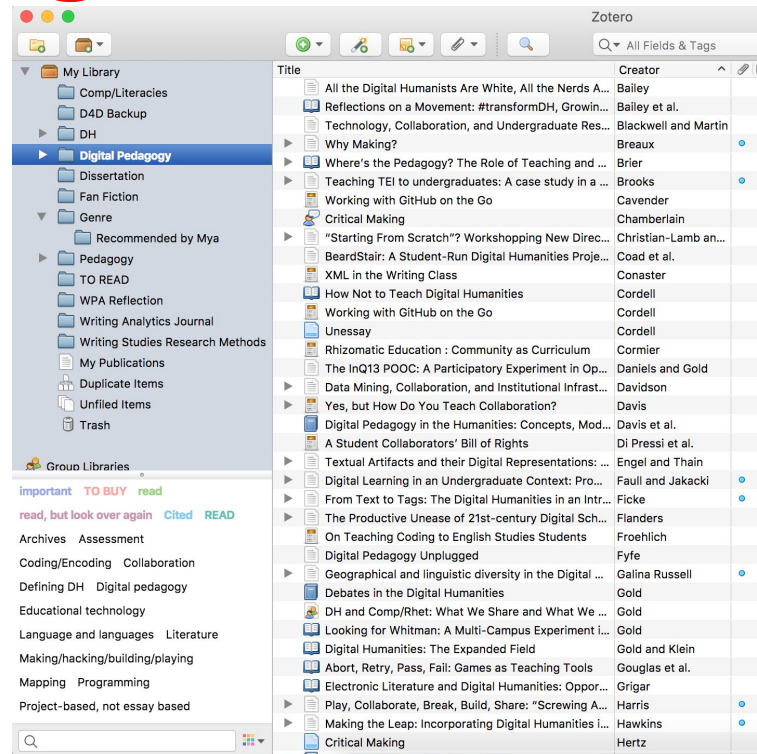
Workshop Objectives

- Understand how to use a citation management software to organize research materials
- Understand Zotero's functions and features
- Define several important functions of Zotero and citation management, including libraries, ISBN, tags, and more.
- Input bibliographic information both manually and automatically into Zotero
- Transfer a library into a Works Cited page



Why Use Citation Management Tools?

Citation management tools provide researchers with the ability to organize their primary and secondary sources, easily create citations and works cited pages for their research, and organize/maintain notes.



What is Zotero?

“Zotero is a free, open-source research tool that helps you collect, organize, and analyze research and share it in a variety of ways” (“About”, Zotero).

- Store citation information on a cloud
- Finds bibliographic information for a source
- Provides flexible methods for organizing your bibliographies (libraries, tags, notes, etc)
- Stores PDFs (even annotated PDFs!) or links to the source
- Provides outputs of your citation information
- Connects to GoogleDrive, Chrome, Microsoft Word, and other

applications



Northeastern University

NULab for Texts, Maps, and Networks

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

Before we use Zotero:

Zotero is most useful when you are collecting sources. So, let's talk about finding proper sources for Advanced Writing

What are **databases**, **journals**, and **peer-reviewed articles**?

- Northeastern Library Database: <https://library.northeastern.edu/>
- Advanced Writing Subject Guides: <https://subjectguides.lib.neu.edu/awd>
- Useful Databases: JSTOR, PolicyFile, ProQuest Political Science, and Academic Search Complete

Choose a database and search for an article related to your research.



Install Zotero

Create an Account

<https://www.zotero.org/>

****IMPORTANT:** Your zotero libraries and info are stored in your Zotero *account* (not on your computer)



Northeastern University
NULab for Texts, Maps, and Networks

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

Anatomy of Zotero

Toolbar

The screenshot shows the Zotero application window. On the left is a sidebar with 'My Library' and a list of group libraries. Below this is a section for 'important TO BUY read' and 'read, but look over again Cited READ' with various keywords. The main pane displays a list of sources with columns for Title, Creator, and other metadata. The right pane shows the 'Info' tab for a selected source, displaying details like Item Type, Title, Editor, Series, and Date Added. Red arrows point from external text labels to specific parts of the interface: 'Your individual and group libraries' points to the sidebar; 'Your chosen keywords to tag each source' points to the keyword list; 'Information about an individual source' points to the right pane; 'Your list of sources in a particular library' points to the main list; and 'at any point in time!' points to the bottom right corner.

Your individual and group libraries

Your chosen keywords to tag each source

Information about an individual source

Your list of sources in a particular library

at any point in time!

Zotero Vocabulary

Knowing the vocabulary of a particular tool makes looking up documentation significantly easier.

- **Library:** the folders that store your citation lists
- **My library:** your individual libraries that can only be edited on your account
- **Group libraries:** collaborative libraries
- **Info:** the bibliographic info about a particular source
- **Notes:** your own notes taken about a source
- **Tags:** keywords chosen to organize and arrange sources



Zotero's Bibliographic Metadata

Search ISBN 978-0801497643

The screenshot shows the Zotero application interface. On the left, a list of books is displayed. The book 'Myths of Empire: Domestic Politics and International Ambition' by Jack Snyder is selected. On the right, the 'Info' tab is active, showing detailed bibliographic information for the selected book. A red box highlights the 'Info', 'Notes', 'Tags', and 'Related' tabs at the top of the right pane.

| Title | Creator | Date Added |
|---|----------------------|------------------|
| > Why Iran should get the bomb: Nuclear balancing would mean stability | Waltz | 2/8/2018, 4:... |
| > Deterrence and Coercive Diplomacy: The Contributions of Alexander Geor... | Levy | 2/8/2018, 4:... |
| > Gendering Power Transition Theory | Sjoberg | 2/8/2018, 4:... |
| > Democratization and war | Mansfield and Sny... | 2/8/2018, 4:... |
| > Rationalist Explanations for War | Fearon | 2/8/2018, 4:... |
| > Conceptual models and the Cuban missile crisis | Allison | 2/8/2018, 4:... |
| > A view from the top: International politics, norms and the worldwide grow... | Reimann | 2/8/2018, 4:... |
| > Political Community and the North Atlantic Area | Deutsch | 2/8/2018, 4:... |
| > Refocusing the IMF | Feldstein | 2/8/2018, 4:... |
| > The utopian nightmare | Easterly | 2/8/2018, 4:... |
| > Emerging multipolarity: why should we care? | Posen | 2/8/2018, 4:... |
| > The new history of World War I and what it means for international relatio... | Lieber | 2/8/2018, 4:... |
| > One world, rival theories | Snyder | 2/8/2018, 4:... |
| > The Melian Dialogue | Thucydides | 2/8/2018, 4:... |
| > International regimes | Krasner | 2/8/2018, 3:... |
| > After hegemony: cooperation and discord in the world political economy | Keohane | 2/8/2018, 3:... |
| > Theories and empirical studies of international institutions | Martin and Simmo... | 2/8/2018, 3:... |
| > Taking preferences seriously: A liberal theory of international politics | Moravcsik | 2/8/2018, 3:... |
| > Cooperation under the Security Dilemma | Jervis | 2/8/2018, 3:... |
| > The politics, power, and pathologies of international organizations | Barnett and Finne... | 2/8/2018, 3:... |
| > Soft Power: The Means To Success In World Politics | Nye | 1/25/2018, 4:... |
| > Myths of Empire: Domestic Politics and International Ambition | Snyder | 1/25/2018, 4:... |
| > State Power and the Structure of International Trade | Krasner | 1/25/2018, 4:... |
| > War and Change in World Politics | Gilpin | 1/25/2018, 4:... |
| > The origins of alliances | Walt | 1/17/2018, 1:... |
| > Theory of international politics | Waltz | 1/17/2018, 1:... |
| > Power Analysis and World Politics: New Trends versus Old Tendencies | Baldwin | 1/13/2018, 3:... |
| > Neoclassical Realism and Theories of Foreign Policy | Rose | 1/13/2018, 3:... |
| > Man, the state, and war : a theoretical analysis | Waltz | 1/13/2018, 3:... |
| > A Critique of Morgenthau's Principles of Political Realism | Tickner | 1/13/2018, 3:... |
| > Politics among nations : the struggle for power and peace | Morgenthau | 1/13/2018, 1:... |
| > Why Do States Build Nuclear Weapons?: Three Models in Search of a Bomb | Sagan | 11/28/2017, ... |
| > Stigmatizing the Bomb: Origins of the Nuclear Taboo | Tannenwald | 11/27/2017, ... |

Info Notes Tags Related

Item Type: Book

Title: Myths of Empire: Domestic Politics and International Ambition

Author: Snyder, Jack

(...) Abstract: Overextension is the common pitfall of e...

Series

Series Number

Volume

of Volumes

Edition: 50579th edition

Place: Ithaca

Publisher: Cornell University Press

Date: March 28, 1993

of Pages: 344

Language: English

ISBN: 978-0-8014-9764-3

Short Title: Myths of Empire

URL

Accessed

Archive

Loc. in Archive

Library Catalog: Amazon

Call Number

Rights

Extra

Date Added: 1/25/2018, 4:05:45 PM

Modified: 1/25/2018, 4:07:18 PM

Info: bibliographic info

Notes: your notes (also sometimes notes are included when you enter ISBN numbers)

Tags: keywords (also sometimes keywords are included when you use ISBN numbers)

Related: choosing other sources that may be related



Inputting Information Manually

1. Select the “New Item” button and find the type of source the article is
2. Input as much of the bibliographic information as you can
3. Use the “Notes” to keep track of important information (we will talk about this more in a second)
4. Use the “tags” to create keywords. Do not overload yourself with keywords. They should be chosen with purpose to help find similar content and organize your sources, not overwhelm you



Demonstration: Inputting Manually

In your Zotero, create a new folder titled “Class Tutorial” (you can delete this folder later. Open up the readings sent over via email (Adcock & Collier, 2001; Lazer & Radford, 2017). We will:

- Figure out the correct citation information
- Determine the type of source
- Fill out the information from there
- Attach the PDF by dragging and dropping the file onto the proper source



Drag and Drop PDFs for Biblio Data

If you already have a PDF of an article downloaded, you can **drag** and **drop** the PDF to Zotero and it will immediately input all the information.

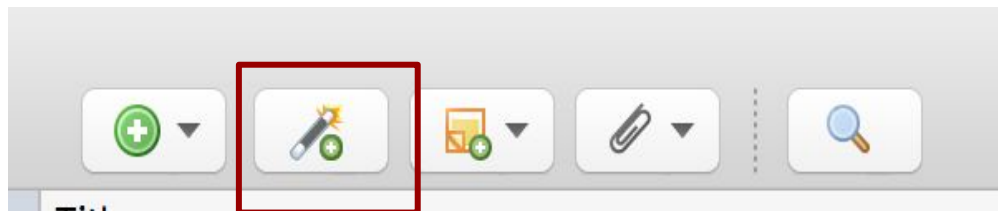
The PDF, however, needs to have metadata attached to it; not every PDF will, but most you get from library databases should.



Adding Items by Identifiers

Look for the DOI number (unique identifier for scholarly articles, ex: 10.2307/358774) or ISBN number (unique identifier for books, ex: 978-0801497643)

Click the “Add Items by Identifier(s)” and paste in the DOI or ISBN number



Exporting your Bibliography

1. As a bibliography saved directly into your clipboard
 - a. Right-click your library
 - b. Click “Create bibliography from collection”
 - c. Choose your proper citation format
 - d. Paste the bibliography in your works cited page
2. As a CSV file (which can be opened on Excel) if you’re ever interested
3. Export it to other Citation Management (EndNotes, for example)



Zotero Web Browser Extension

If you use Chrome or Firefox, there is a **Zotero extension** that you can download onto your browser and connect to your Zotero account. When you're researching articles, you can easily save your bibliographic information by clicking on the "Zotero" button on your toolbar (just make sure it's being saved to the correct library!). It will also save the HTML link or the PDF.

Always double check that the citation information is correct in your Zotero library (some websites do not provide proper metadata)



Zotero on Google Docs and Word

Your Zotero can now be connected to your Google Drive account or Microsoft Word. While you are writing a paper, you can use the Zotero plugin to create your bibliography for you!

****Always double-check that all the bibliographic information is correct and follows the proper citation format****



Your Turn!

Break up in a group and in your “Class Tutorial” folder, use Zotero’s features and/or manually input information for:

- Any *New York Times* article
- The two readings provided via email (Adcock & Collier, 2001; Lazer & Radford, 2017)
- Singleton, Royce A., and Bruce C. Straits. *Approaches to Social Research*. Oxford University Press, 2017.

Slides & handout available at <http://bit.ly/diti-spring2020-averymiller>



Class Discussion

- First of all, any general questions?
- What did you find easy or challenging about using Zotero?
- How might you use Zotero in the future outside of this class?



Thank you!

If you have any questions, contact us at:

Jeff Sternberg

DITI Research Fellow

sternberg.je@husky.neu.edu

Slides & handout available at: <http://bit.ly/diti-spring2020-averymiller>

Sign up for office hours at: <https://calendly.com/sternberg-je/15min>



Northeastern University
NULab for Texts, Maps, and Networks

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