

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

---

Cara Marta Messina and Garrett Morrow  
SOCL 2321 Research Methods  
Ineke Marshall  
Fall 2019



Northeastern University  
*NULab for Texts, Maps, and Networks*

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

# Discussion: China's Social Credit System

- What is China's Social Credit system? How does it work?
- In what ways might America have similar or different technological infrastructures?



# Workshop Agenda

- Objectives
- Introduce 'Big Data' Concepts
- Discuss data, privacy, and data categorization
- Activity: Adopt or Not?
- Discuss ethical implications of big data and more generally digital research

Slides, handouts, and data available at <https://bit.ly/2H48Yc1>



# Workshop Goals

- Understand the ways data is being used in society as well as how algorithms impact and shape our daily lives
- Understand the ways in which technology reflects cultural, social, and political biases.
- Explore the ways in which privacy and security are being reshaped and redefined through the use of big data, algorithms, and policy
- 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.



# Facebook Preferences

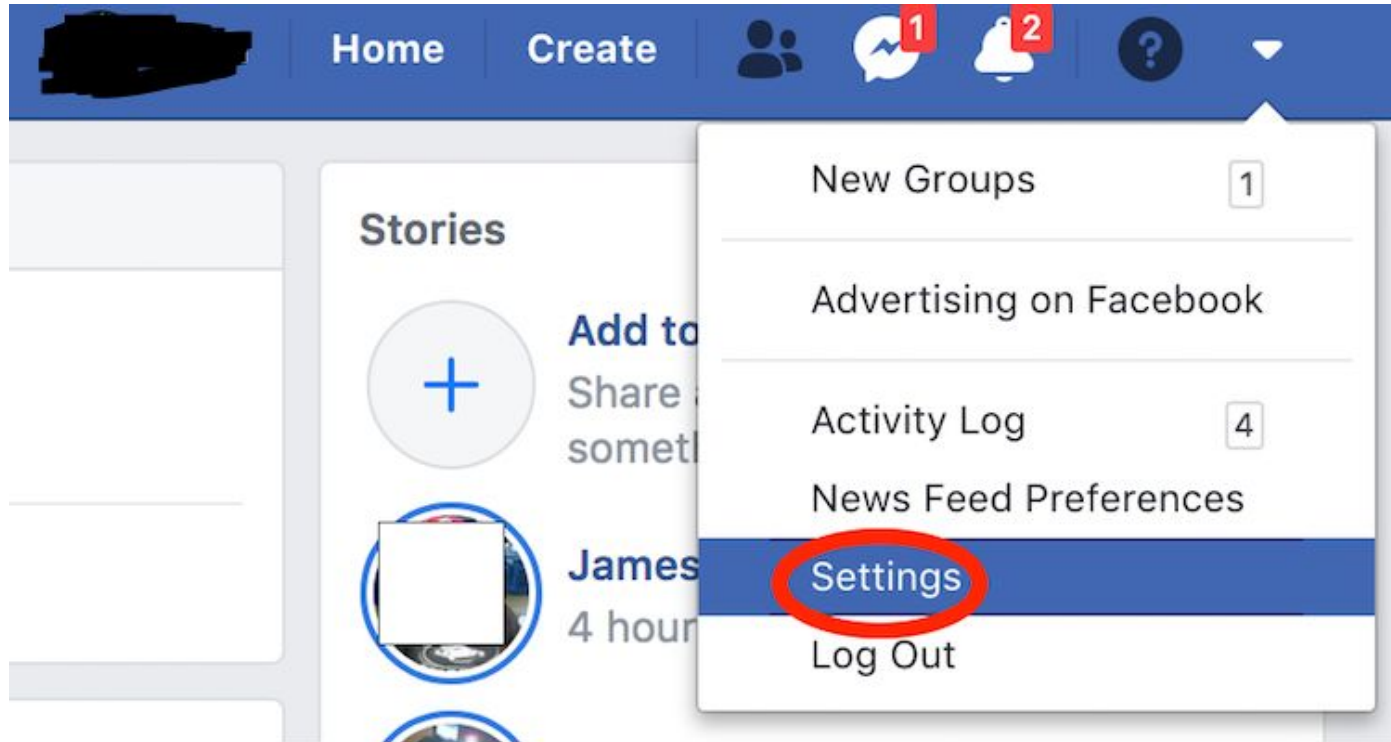
Facebook collects, stores, and sells information about you so you get better targeted ads and your newsfeed is tailored to your categories.

Other social media sites that do this:

- Instagram (owned by Facebook)
- Google
- YouTube (owned by Google)
- Twitter

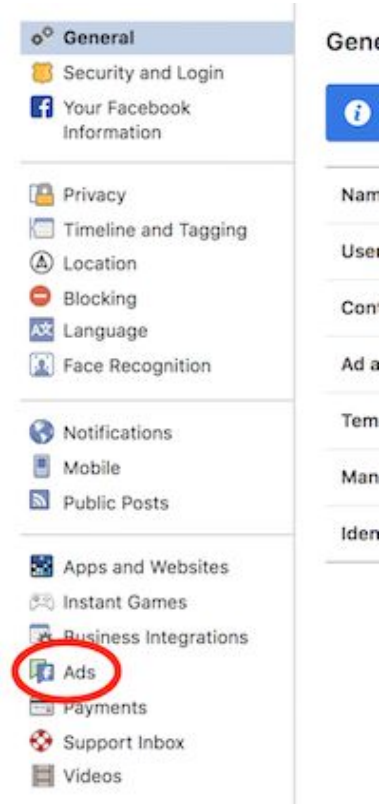


# Settings > Ads > Your information > Categories

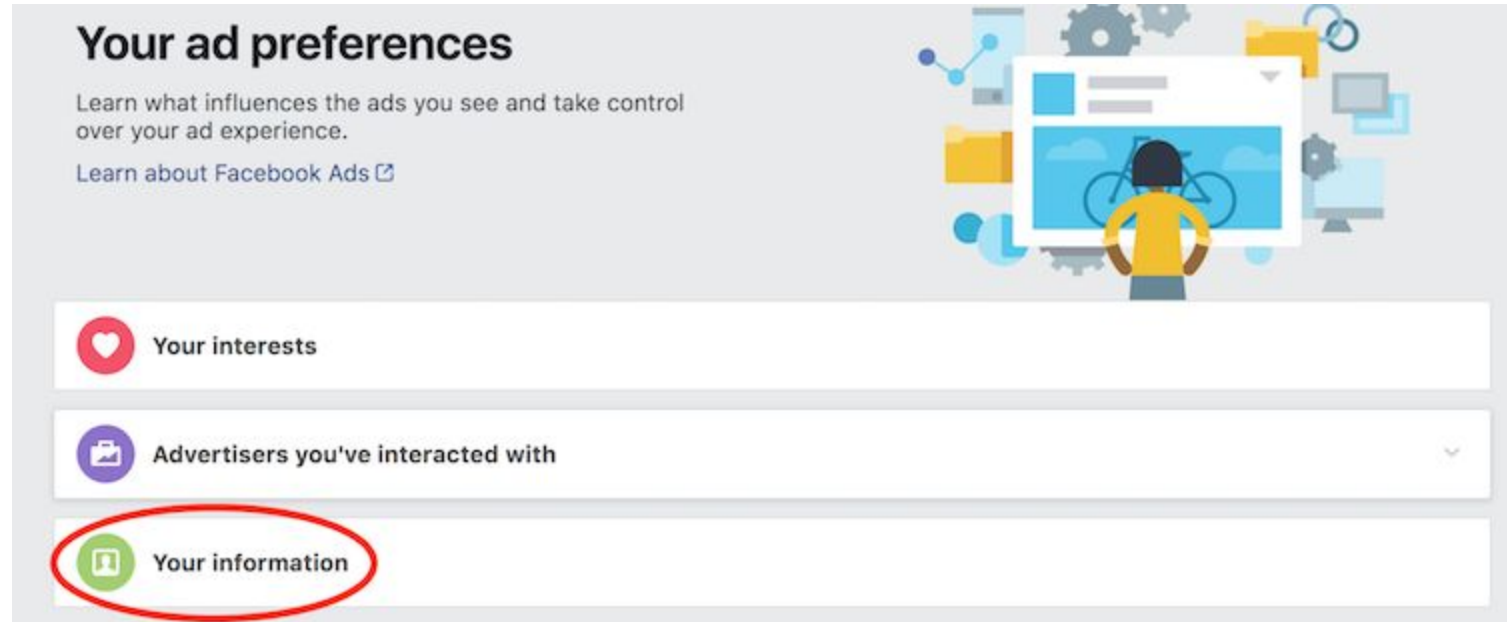




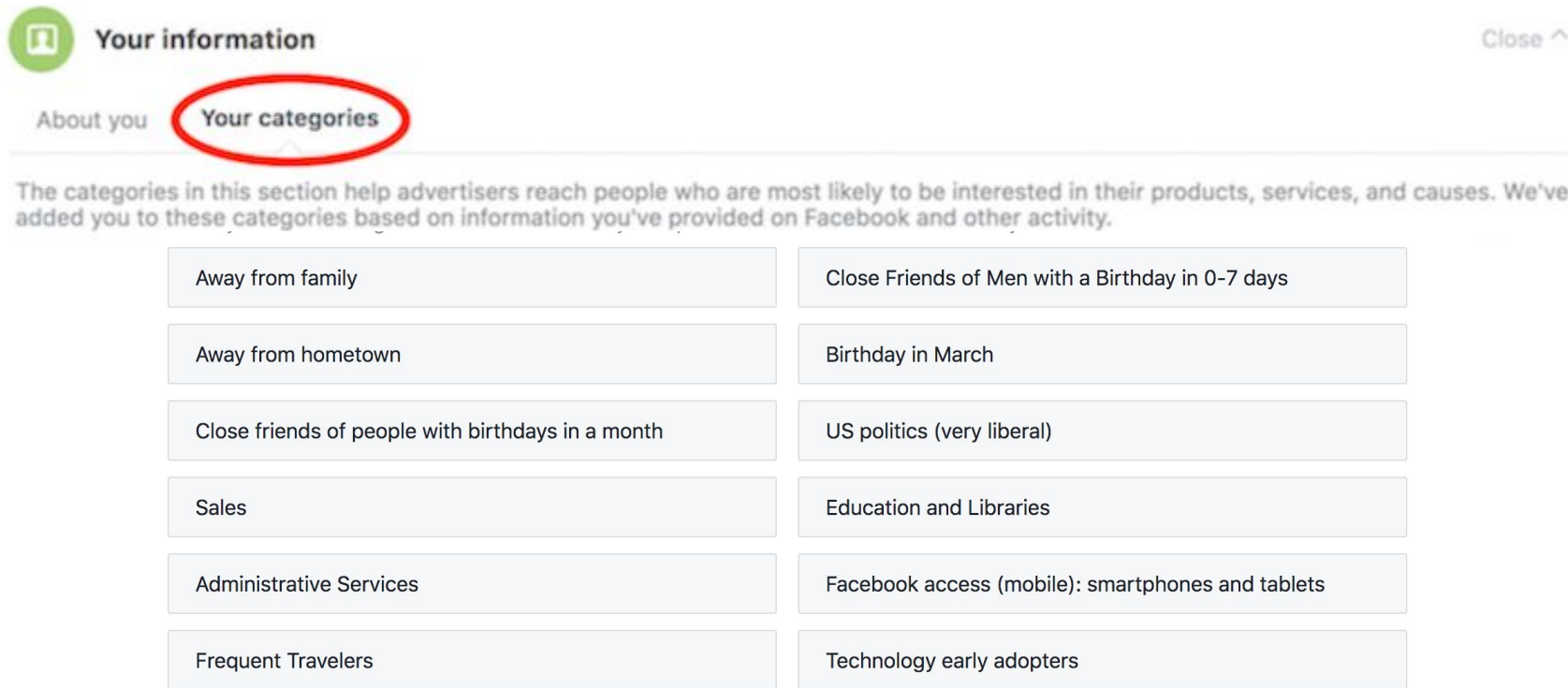
# Settings > Ads > Your information > Categories



# Settings > Ads > Your information > Categories



# Settings > Ads > Your information > Categories



The screenshot shows the Facebook 'Your categories' settings page. At the top, there's a green circular icon with a white 'i' and the text 'Your information'. Below it, two tabs are visible: 'About you' and 'Your categories', with the latter being circled in red. A 'Close ^' link is in the top right. A descriptive paragraph states: 'The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.' Below this, there are two columns of category tiles. The left column includes: 'Away from family', 'Away from hometown', 'Close friends of people with birthdays in a month', 'Sales', 'Administrative Services', and 'Frequent Travelers'. The right column includes: 'Close Friends of Men with a Birthday in 0-7 days', 'Birthday in March', 'US politics (very liberal)', 'Education and Libraries', 'Facebook access (mobile): smartphones and tablets', and 'Technology early adopters'.

**Your information** Close ^

About you **Your categories**

The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.

Away from family	Close Friends of Men with a Birthday in 0-7 days
Away from hometown	Birthday in March
Close friends of people with birthdays in a month	US politics (very liberal)
Sales	Education and Libraries
Administrative Services	Facebook access (mobile): smartphones and tablets
Frequent Travelers	Technology early adopters



# 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>



# Downloading Your Data

Facebook: Settings > Your Facebook Information > Download your Information

Google:

<https://support.google.com/accounts/answer/3024190?hl=en>.

Instagram: Settings > Security > Access Data/Download Data



# 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 and Bias



Northeastern University  
*NULab for Texts, Maps, and Networks*

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



# Activity: Adopt or Not?

**Small Group:** Find a partner or two! You all work for an adoption agency and have to decide if someone can adopt a dog. On your handouts, please read the four previous adoption applications and decide if the new adoption applicant can adopt or not.

**Do you think this new applicant can adopt a dog? Why or why not?**



# Class Discussion: Adopt or Not?

- Do you ACCEPT or REJECT their application? Why?
- What questions from the application did you weigh more? Why?
- What might be some implicit biases in this application and in your choices?



# Adopt or Not? Algorithm

Algorithms to “read” through applications exist.

- Where might you see these algorithms being used to make decisions?
- What biases may be ingrained in those applications?
- How might the impact choices?



# Want to learn more about accountability and best practices when creating algorithms?

Visit <https://www.fatml.org/>, or Fairness, Accountability, and  
Transparency in Machine Learning



# So what do 'big data' & algorithms have to do with research?



**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

**Garrett Morrow**

Digital Teaching Integration

Research Fellow

morrow.g@husky.neu.edu

Slides, handouts, and data available at <https://bit.ly/2H48Yc1>

DTI office hours: **Tuesdays, 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!*