

# **Ethics, Privacy and Social Justice in Data Science**

## **Data Science for Social Justice**

Week 7

Ksenia Polson, Ph.D.

# Recap Last Week

- Ethics in social media research
- Seek informed consent in all types of online research?
- Importance of anonymity and avoiding harm
- Applications of social media in medical research
- Mind maps

# Weekly Outline

Apple and Google's Covid-19 contact tracing

- Privacy concerns?

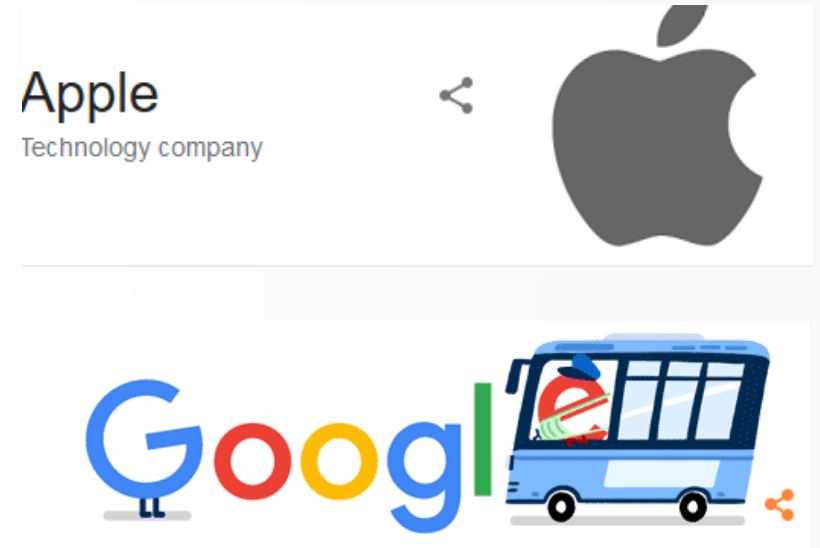
Social Justice Changemakers

- Jake Porway's 5 principles of data science for social good
- Public-private partnerships

# Covid-19 Contact Tracing

- New features on mobile phones to use Bluetooth radios to track proximity between phones
- If someone receives a positive diagnosis- they can report it through the app
- Issues with tracking people, ad-targeting, location tracking, identity of Covid-19 patients and false positives

Is contact tracing ethical if it compromises privacy?



# What is Social Justice?

American Public Health Association defines social justice as the view that everyone deserves equal rights and opportunities... “ [link](#)

David Miller (1976)

- 3 principles of social justice:

“To each according his **rights**; to each according to his **deserts**; and to each according to his **needs**.”

John Rawls (1971)

- “the ways rights and duties are assigned the basic **social institutions**, (such as governing agencies, courts, schools, and hospitals) and the principles that guide **institutional decisions** about how wealth, rights, and resources are **distributed**.”

Which principle if the hardest to achieve using data science?

**Urban Institute's  
Alex Engler on  
future of social  
science research**



“Data are changing, and we can use these data for social good, but only if we are willing to adapt to new tools and emerging methods.”



[New York City](#) identifies illegal building conversions that can pose a fire hazard.



[Chicago](#) preemptively searches for homes with young children and lead paint to prevent lead poisoning



A call for social scientists to open city databases on which the future of research is dependent.

**Dr. Phillip Solomon Use  
of Big Data for Social  
Justice**

What will big data tell us  
about our democracy?

- Disparities?

Use of Big data in law  
enforcement

- CompStat for Justice (C4J) is an approach to map crime in regions of a given city and measure patterns of crime in those regions.
- Implementation underway in San Francisco, CA and Norfolk, VA

# University of Washington's Collaborative Projects with a social good dimension



<https://escience.washington.edu/using-data-science/data-science-for-social-good/>



# Social Justice Changemakers

Megan Price- Human Rights Data Analysis Group

- Human rights projects in Guatemala, Syria, and Columbia
- Analysis of documents from the National Police Archive.

Human Rights Data Analysis Group  
Statisticians for human rights

**HRDAG** 30 years and counting

Jake Porway – DataKind

- Connects data scientists and social-change organizations
- Short to multi-month projects
- Demand on both sides

**DataKind**®

# Datakind's Social Justice Partnerships

## Crisis Text Line (CTL)

- Volunteer staff
- Repeat Callers are a problem- change in policies and tools to serve a greater number of people
- Power of storytelling

**Text HOME  
to 741741**

**for free, 24/7 crisis  
support in the US.**

# Social Change Maker- Charity “Service Provider”

Donations channeled directly to the poor in Kenya and Uganda

Model based on scientific inquiry

Transparent and redundant system of multiple teams and a verification process

**GiveDirectly**

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

# Jake Porway's 5 Principles for data science and social good

1. what constitutes data, what it can be used for, and why it's valuable.

- Move to real time, granular data

2. collaboration between the problem holders, the data holders, and the skills holders.

- What is possible to predict from the data?

# Jake Porway's 5 Data Science Principles (continued)



**3. encourage people to speak openly and honestly. Engage in curiosity.**

Relate how the findings help an organization function and improve its operations



**4. To tackle sector-wide challenges, we need a range of voices involved.**

Call for “collective impact”



**5. Data is not truth, and tech is not an answer in-and-of-itself. Keep the end of the project in mind.**

How will a project benefit the people being served?

# Private Public Partnerships

Software screened reviews posted on Yelp for potential cases of foodborne illnesses

- 468 reviews were flagged
- 3% of reviews had been reported to the Health Department

Dr. Balter of DOMHH, “the team from Columbia helped us focus on the small number of restaurant reviews that might indicate real problems.”

Goal to build infrastructure and machine learning tools

Application to other health departments



# Social Good Books

## Book



### AI for Social Good

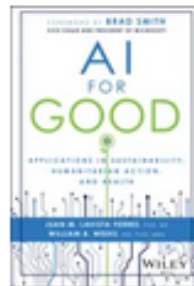
By [Rahul Dodhia](#)

[Wiley](#) • March 2024

Write the [first review](#)  336 pages  6h 19m

---

## Book



### AI for Good

By [Juan M. Lavista Ferres](#), [William B. Weeks](#) and [Brad Smith](#)

[Wiley](#) • April 2024

Write the [first review](#)  432 pages  9h 44m

# Assignment: Bolivia Case Study

Rathburn, M. K., & Baum, K. J. (2009). Originally published 7/22/09 by Division of Natural Sciences, College of General Studies, Boston University.

1. Read through the entire ["Water Bolivia" case study](#), also located in the Course Resources folder.

*Note:* You do not need to answer all the questions in the case study.

2. Answer the following questions:

**Scenario 1:** You are a data scientist working *on behalf of the social good of the Cochabamba people*. After reading the Water Bolivia case study, what is your definition of the problem as a data scientist? As an individual, not as a data scientist, do you agree with that definition? What ethical dilemmas could you be facing?

**Scenario 2:** You are a data scientist for the Company Bechtel. Bechtel is interested in increasing their profits, so their data scientist is in direct conflict with the Scenario 1 data scientist. What definition of the problem might Bechtel want you to have as a data scientist for the Cochabamba water system? What data collection, analysis, and reporting ethical dilemmas could you be facing?



# Assignment (continued)

3. Select one of the following roles:

- A data scientist for the Cochabamba water system.
- A data scientist for the company Bechtel.
- A data scientist working on behalf of the people in the community.

4. Using your chosen role, identify your desired outcome.

5. Using your chosen role, review questions 2 and 3 on page 5 of the case study.

Provide responses to these questions as a data scientist and explain how you could represent and misrepresent the data in response to these questions.

6. Using your chosen role, describe what data is needed.

7. Describe ethical issues you might face relating to each of the answers you provided for your responses to questions 4 through 6 above.

8. Format your writing as an APA-formatted paper, and submit your assignment to the Week 7 dropbox.

# Bolivia Case Study Rubric

Name	Points
Definition and analysis of the problem	2
Conflicting definition of the problem	2
Ethical dilemmas (and ethical dilemmas as a data scientist) for the Cochabamba water system	2
Desired outcome	1
Description of how to represent the data ethically, unethically, and data needed for the project	3
Total Points	10

# **Ethics Grand Rounds Presentations**

Groups 3 & 4 present!