2-24-2021

Big Data: Ethics, Resources, and Potential Collaboration

Matthew Zook
University of Kentucky, zook@uky.edu

Big Data: Ethics, Resources, and Potential Collaboration

VPR Lunch & Learn
University of Kentucky
February 24, 2021

Dr. Matthew Zook
Geography, University of Kentucky

PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

EDITORIAL

Ten simple rules for responsible big data research

Matthew Zook, Solon Barocas, danah boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson, Frank Pasquale

Published: March 30, 2017 • https://doi.org/10.1371/journal.pcbi.1005399

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399

Search Results

Results: 1 – 20 of 168 for All ethics within Big Data & Society ×

Articles(168)

Download selected citations Refine Search Sort: Relevance

Select all

***AI ethics should not remain toothless!* A call to bring back the teeth of ethics**

Anaïs Resseguier, Rowena Rodrigues

Big Data & Society, vol. 7, 2, First Published July 22, 2020.

Abstract

› Preview

https://journals.sagepub.com/home/bds

Ten Simple Rules for Responsible Big Data Research

1. Acknowledge that data are people and can do harm.

1. Acknowledge that data are people and can do harm

- Unanticipated insight from data
  - EXIF records from photos with location coordinates or determining heart rates of people from YouTube videos;
- Datasets about population-wide effects that impact groups
  - Social network maps shaping credit-access
  - Recidivism metrics that shape parole decisions in a racially disparate manner
  - Zip code categorization resulting in less access to Amazon Prime for African-Americans in US cities
- "Public" datasets are easily adapted for highly invasive research by incorporating other data
  - Such as Hague et al.'s (2016) identifying the artist Banksy.

Ten Simple Rules for Responsible Big Data Research

1. Acknowledge that data are people and can do harm.
2. Recognize that privacy is more than a binary value.

2. Recognize that privacy is more than a binary value.

- Privacy is contextual (Nissenbaum, 2010) and situational (Marwick & boyd, 2014)
- Just because something has been shared publicly does not mean any subsequent use is unproblematic. (single observation vs. ALL)
- Pushing past social norms, particularly in novel situations created by new technologies, is perceived by individuals as "creepy" (Tene and Polonetsky, 2013)
  - True even when no violation of ToS or privacy laws.
- Privacy also goes beyond single individuals and extends to groups.
  - Particularly important for communities who have been historically discriminatory against, e.g., the practice of redlining in the US

Ten Simple Rules for Responsible Big Data Research

1. Acknowledge that data are people and can do harm.
2. Recognize that privacy is more than a binary value.
3. Guard against the re-identification of your data.

3. Guard against the re-identification of your data

- Long history of "anonymized" data that can be de-anonymized
  - See Barbaro et al, 2006; Cox, 2016; Panduragnan, 2014.
- The identificatory power of birthdate, gender, and zip code is well known (Sweeney, 2002).
- Hard to know vulnerable points *a priori* and "harmless data" may very well prove to be a significant vector of identification (e.g., battery usage).
  - Metadata associated with digital activity, location, battery usage
  - Unlabeled network graphs - such as location and movement, DNA profiles, call records and even high-resolution satellite images of the earth - can be used to re-identify people (Kloumann and Kleinberg, 2014).
  - Google's reverse image search can connect previously separate personal activities – such as dating and professional profiles – in unanticipated ways (see also Acquisti, Gross and Stutzman, 2014).

Ten Simple Rules for Responsible Big Data Research

1. Acknowledge that data are people and can do harm.
2. Recognize that privacy is more than a binary value.
3. Guard against the re-identification of your data.
4. Practice ethical data sharing.
5. Consider the strengths and limitations of your data; big does not automatically mean better.

Ten Simple Rules for Responsible Big Data Research

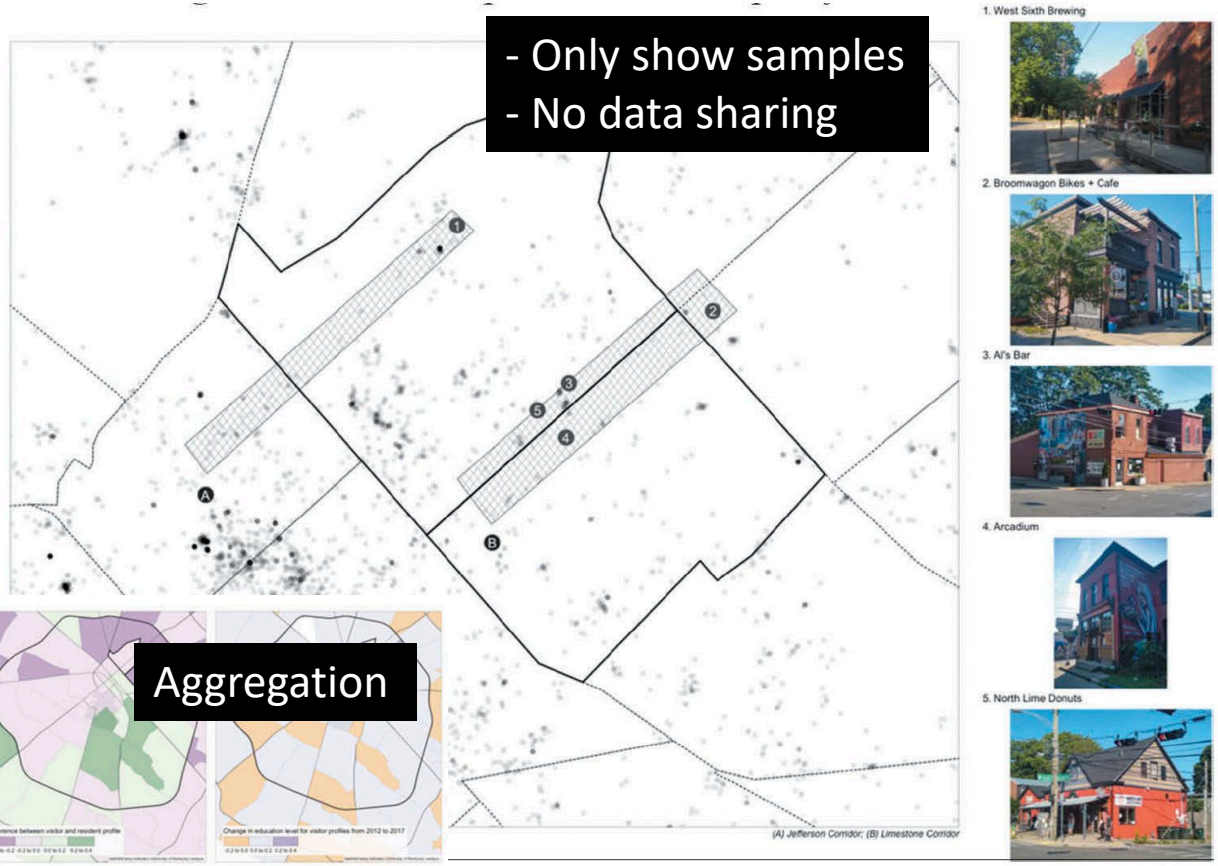
6. Debate the tough, ethical choices.
7. Develop a code of conduct for your organization, research community, or industry.
8. Design your data and systems for auditability.
9. Engage with the broader consequences of data and analysis practices.
10. Know when to break these rules.

My Own Big Data Challenges

- Geotagged Twitter data (18+ billion obs.)
  - Locational data is sensitive (home, work, leisure locations)
  - Abide by Twitter's ToS and have alerted them of the project (but hard to get responses)
  - Data is secured at collaborator's university
- Approach has largely focused on "ethics"
  - We're often the first to work with these data.

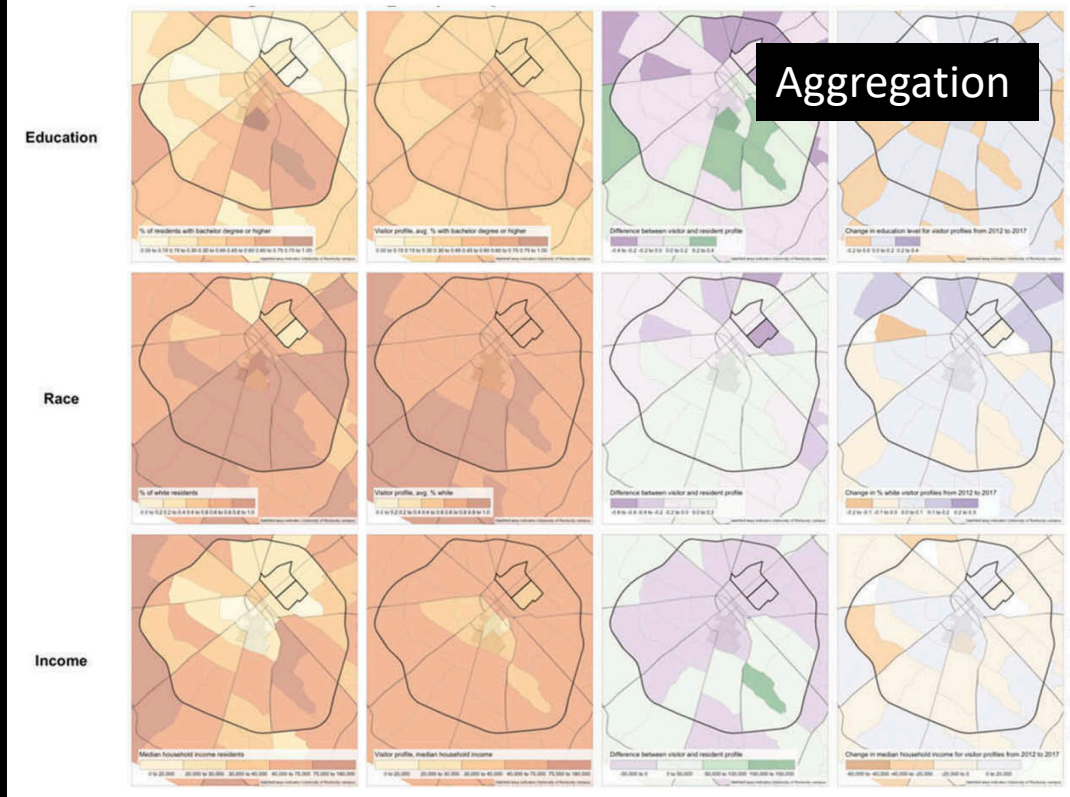
Using Twitter data to gentrification in Lexington via mobility and relational connections between neighborhood

- Only show samples
- No data sharing



efferson Avenue Corridors (delineated with dashed lines). Thicker
nd 3 used in our analysis to represent the Northside neighbor-
tweets in the dataset.

Aggregation



Poorthuis, A., Shelton, T. and M. Zook. 2021. Mapping gentrification 'in real time (and space)': using social media data to understand changing urban relationalities. *Urban Geography*. Forthcoming.

How do we balance this?