Big data
(02-02)

Matthew J. Salganik
Department of Sociology
Princeton University

Soc 596: Computational Social Science
Fall 2016

- Observing behavior
- Asking questions
- Running experiments
- Mass collaboration

What is big data?

Ambiguity makes a big tent

Big data:

- Volume
- Velocity
- Variety

Definition originates from business (Gartner)

Big data $\rightarrow$ repurposed data

Big data → repurposed data

- Business administrative data
- Government administrative data

Big data $\rightarrow$ repurposed data

Big data $\rightarrow$ repurposed data

- good things
- bad things

Problems with the chapter:

- ▶ not all observed behavior is repurposed data, but most is
- ▶ this is kind of residual category (not surveys, not experiments, not mass collaboration)

General advice:

- Imagine the ideal data. How does your data compare?

General advice:

- Imagine the ideal data. How does your data compare?
- Understand the data generating process (sometimes require reverse engineering goals of data creators)

General advice:

- Imagine the ideal data. How does your data compare?
- Understand the data generating process (sometimes require reverse engineering goals of data creators)
- Remember big data is not all of computational social science

Readymades

Custommades

Big data $\rightarrow$ repurposed data

Big data $\rightarrow$ repurposed data

- ▶ good things
- ▶ bad things