

# Characteristics of big data (02-03)

Matthew J. Salganik  
Department of Sociology  
Princeton University

Soc 596: Computational Social Science  
Fall 2016



Big data → repurposed data

- ▶ good things
- ▶ bad things

Good things

- ▶ big

## Good things

- ▶ big
- ▶ always-on

## Good things

- ▶ big
- ▶ always-on
- ▶ non-reactive

Bad things:

- ▶ incomplete

Bad things:

- ▶ incomplete
- ▶ inaccessible

Bad things:

- ▶ incomplete
- ▶ inaccessible
- ▶ non-representative



Bad things:

- ▶ incomplete
- ▶ inaccessible
- ▶ non-representative
- ▶ drifting

Bad things:

- ▶ incomplete
- ▶ inaccessible
- ▶ non-representative
- ▶ drifting
- ▶ algorithmically confounded

## Bad things:

- ▶ incomplete
- ▶ inaccessible
- ▶ non-representative
- ▶ drifting
- ▶ algorithmically confounded
- ▶ dirty

## Bad things:

- ▶ incomplete
- ▶ inaccessible
- ▶ non-representative
- ▶ drifting
- ▶ algorithmically confounded
- ▶ dirty
- ▶ sensitive

Bad things:

- ▶ incomplete
- ▶ inaccessible
- ▶ non-representative
- ▶ drifting
- ▶ algorithmically confounded
- ▶ dirty
- ▶ sensitive

For each bad thing, what is a counter-strategy?