

Big data characteristics (02-03)

Matthew J. Salganik
Department of Sociology
Princeton University

Soc 596: Computational Social Science



Big data → repurposed data

- ▶ good things
- ▶ bad things

Good things

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

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?