~~Privacy~~ Issues?

- Helen Nissenbaum, Cornell Tech
- Paul Ohm, Georgetown Law
- Daniel Barth-Jones, Columbia Public Health
- Omer Tene, International Association of Privacy Scholars
- danah boyd, Data & Society and Microsoft Research
- Daniel Goroff, Alfred P. Sloan Foundation
  *** Opinions expressed are his own. ***

Participation Issues!

- The applications of Decennial Data are truly impressive. Conducted with great care about data quality and utility.

- Many threats: sampling error; coding error; human error; imputation; swapping; suppression; clamping; noise, etc.

- All these might hardly matter without the truthful, representative, and safe participation in the Decennial by hundreds of millions of our fellow residents.

- They will reasonably ask, “Should I participate in the Census and, if so, should I answer truthfully?”

Should I Participate?

- If many refuse or dissemble, that worsens other threats. Especially since this is unlikely to be randomly distributed.

- As always, some want to skew the response distribution. People like us do not feel very vulnerable. Others may, either because of disinformation or for very good reasons. E.g., elderly alone, undocumented, publicly housed, etc.

- What can we truthfully tell them about participating? How can we help fellow residents and fellow Census users understand how data releases can actually protect **both** utility and privacy by trading some of one for the other?

Formal Privacy

Lemma: Suppose an analyst uses a query mechanism *M* that satisfies ε-differential privacy to study a dataset *z*.

The analyst does not know whether *z* = *x* or *x'*, where these denote two neighboring datasets, ie, ones that differ in at most one line.

As a Bayesian, the analyst does have a prior belief about whether *z* = *x* or *x'* that is expressed as an odds ratio *Pr* (*z*=*x*) / *Pr* (*z*=*x'*).

After receiving the query answer, *M*(*z*), the analysts' posterior odds can only differ from the prior odds by a factor that is between *exp*(- ε) and *exp*(ε).

Privacy Guarantees

Corollary: For small values of ε, the analyst's prior and posterior odds about my participation in the data collection can differ by no more than (100 x ε) percent.

Example: Consider an analyst who has no idea to begin with about whether z=x, a dataset that contains my personal information, or z=x', a dataset that does not. This corresponds to even odds of 50:50 or, in other words, an odds ratio of one. The answer provided by a query mechanism that satisfies (0.1)-differential privacy could change those odds by about 10% to 52.5:47.5. For epsilon = 3, the odds could change to about 20:1.

Other Considerations

- Your truthful participation matters. Differential Privacy just makes it very hard to tell that it was you who participated.

- **Systematic** undercounting for any reason can be deeply troubling. Many would more readily attribute it to lack of participation rather than to methodological concerns.

- **Statistical bias** is not due to DP, but to cosmetic post-processing. Researchers should have access without that.

- The Bureau actually will add random numbers to the counts it reports, but these are likely to be small compared with usual utility threats like non-participation and other errors. The alternative is to offer **no** formal privacy guarantees.

Privacy and Participation

- Helen Nissenbaum, Cornell Tech
- Paul Ohm, Georgetown Law
- Daniel Barth-Jones, Columbia Public Health
- Omer Tene, International Association of Privacy Scholars
- danah boyd, Data & Society and Microsoft Research
- Daniel Goroff, Alfred P. Sloan Foundation
  *** Opinions expressed are his own. ***

Panel Questions

- Helen: Context matters. What privacy considerations should help people decide about responding to a survey?

- Paul: What legal protections apply? What is the worse that could happen to me or to my data?

- Daniel: Many feel that health data can find cures or be used against us. How is Census participation different?

- Omer: How much of a threat is re-identification anyway?

- danah: What worries you most?