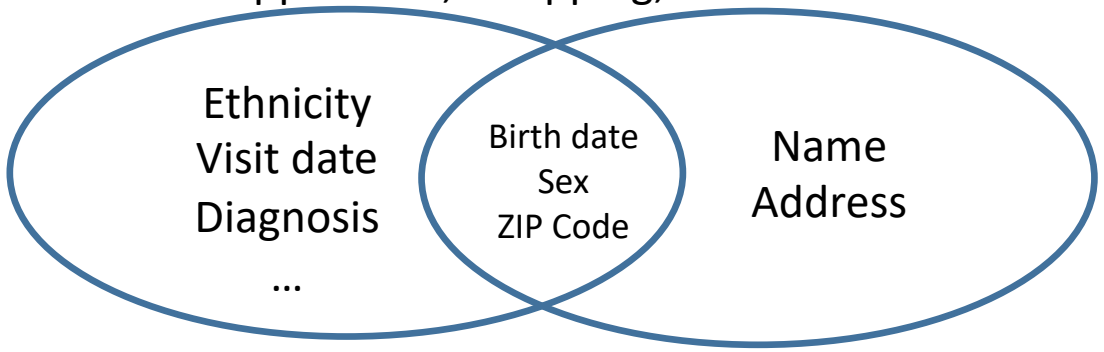
There is privacy risk in de-identified data — what can we do about it?

Sweeney, 2000

Field Suppression, Swapping, Generalization



Medical Data **Voter List**

HIPAA "Safe Harbor" standard.
k-anonymity, *ℓ*-diversity, *t*-closeness
Differential Privacy

Today:

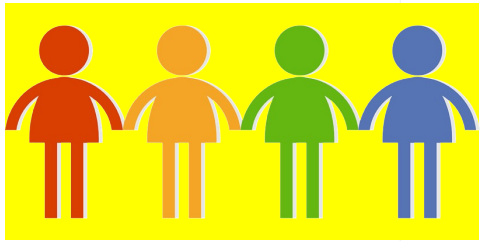
How do we balance Utility vs. Risk?

How does risk change over time?

New data types to de-identify.

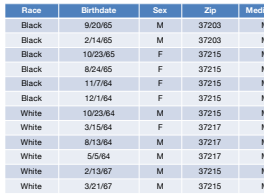
NYC Taxi drivers Re-identified

Bikeshare riders re-identified?

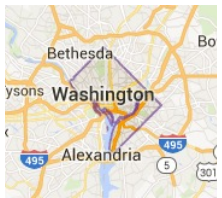


Community harms without re-identification

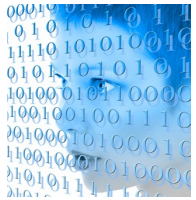
Options moving forward:
- Better de-identification
- Synthetic data
- Data use agreements
- Privacy remediation



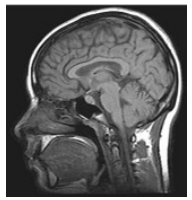
Tabular



Geospatial



Social Media



Medical

DNA

Bodycams/Surveillance

NIST FY2016 GOALS:
- Develop guidance
- Gov De-ID Stakeholder meeting
- Pilot de-identification evaluation

Simson L. Garfinkel, Ph.D., Senior Advisor, Information Access Division, simson.garfinkel@nist.gov

NIST National Institute of Standards and Technology U.S. Department of Commerce