Census Tree

Nathan Schill
schilln@byu.edu

What is the Census Tree?

- Link people across each pair of US censuses

- Machine learning with XGBoost in Python

- Use FamilySearch "true" data for training

- Comprehensive data across time for study!
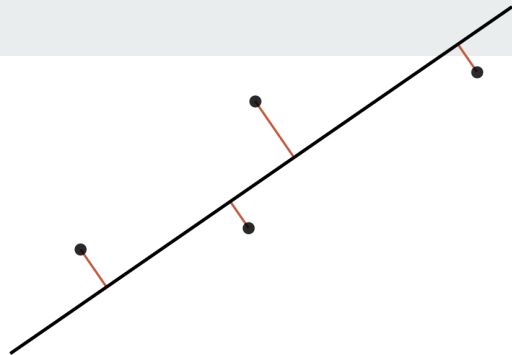
What features?   Total: ~60

- Birth year difference
- Residence distance
- Last and first name similarity
- Relationship to head match
- Parents' birthplaces
- Change from single to married

- Occupation match
- . . .

Blocking?

- Birthplace
- Gender
- Race . . .

What's next?

- Train a model just on women

- Experiment with hyperparameters and features

- Bridge 50-year gaps between censuses

- Share!