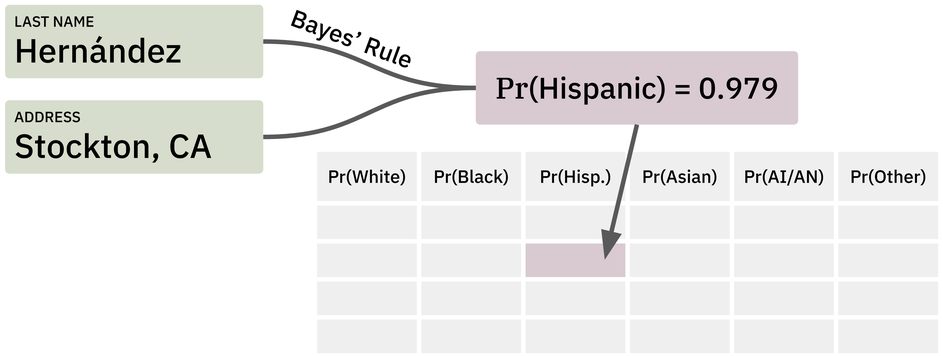
We can accurately predict race from proxies like name and address.

But how do we use these predictions to estimate racial *disparities*?



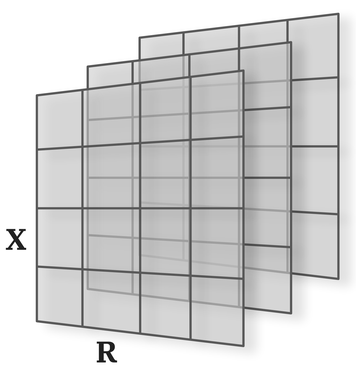
1 Bayesian Improved Surname Geocoding (BISG) generates **probabilistic predictions** of individual race using Census data on locations, last names, and race.

How can we use these BISG predictions to estimate a disparity like $\mathbf{E[X \mid White] - E[X \mid Black]}$?

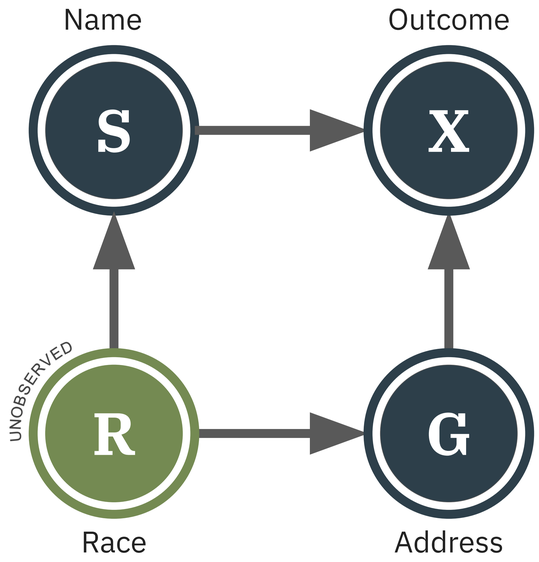
Fiscella, K. and Fremont, A. M. (2006). "Use of geocoding and surname analysis to estimate race and ethnicity."

Elliott, M. N., Fremont, A., Morrison, P. A., Pantoja, P., and Lurie, N. (2008). "A new method for estimating race/ethnicity and associated disparities where administrative records lack self-reported race/ethnicity."

4 We develop a simple **Bayesian model** to produce disparity estimates from BISG probabilities, using the identifying assumption.



- **Scalable**: Marginalize out unknown race vector for improved sampling, fit using stochastic variational inference (SVI)
- **Flexible**: can add covariates, combine random effects, or produce subgroup estimates
- **Theoretical results** on the magnitude and direction of the effect of non-sampling error on model estimates

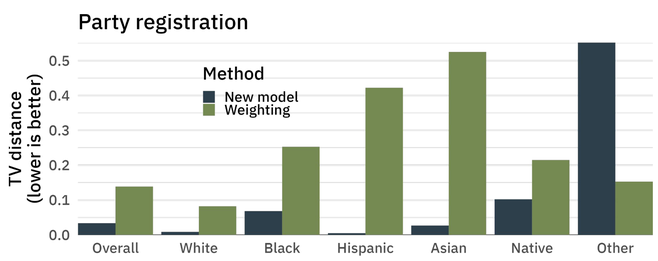


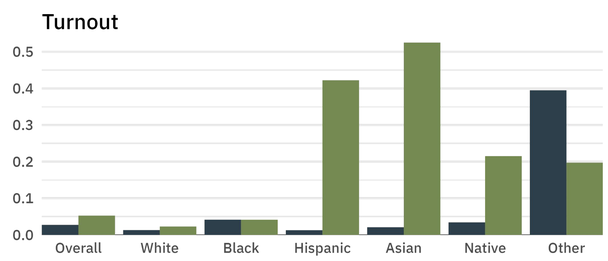
2 Researchers often **weight** the outcome $\mathbf{X}$ by the BISG probabilities.

This approach is **biased** unless the effect of race is fully mediated by name and address: $\boldsymbol{X \perp\!\!\!\perp R \mid S, G}$.

Chen, J., Kallus, N., Mao, X., Svacha, G., and Udell, M. (2019). "Fairness under unawareness: Assessing disparity when protected class is unobserved."

5 The new method produces better estimates of racial differences in party registration and turnout in North Carolina **validation data**.

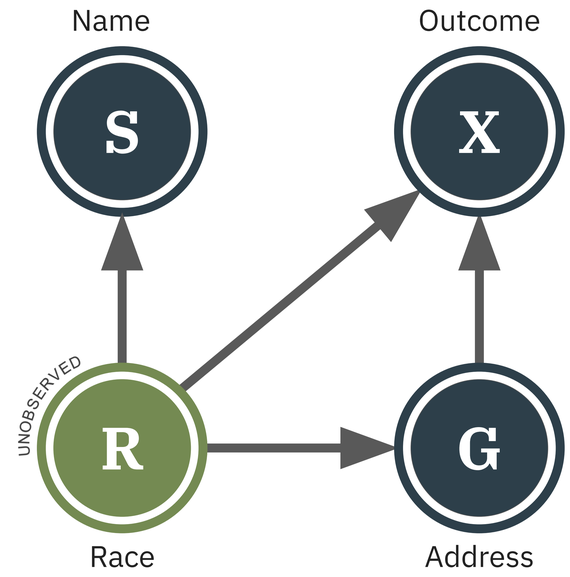




Estimation of Racial Disparities When Race is Not Observed

Cory McCartan,[1] Robin Fisher,[2] Daniel E. Ho,[3] Jacob Goldin,[2,3] AND Kosuke Imai [1,4]

1. Department of Statistics, Harvard University
2. Office of Tax Analysis, U.S. Department of the Treasury
3. Law School, Stanford University
4. Department of Government, Harvard University



3 We provide an **alternative identifying assumption:**

$$\boldsymbol{X \perp\!\!\!\perp S \mid R, G}$$

This assumption allows us to accurately estimate racial disparities using name as an independent high-dimensional proxy for race.