

Reconstruction of age distributions from differentially private (DP) data

SHARYGIN Ethan (PSU),
DYRTING Sigurd (CDU),
FLAXMAN Abraham (UW)

INTRO

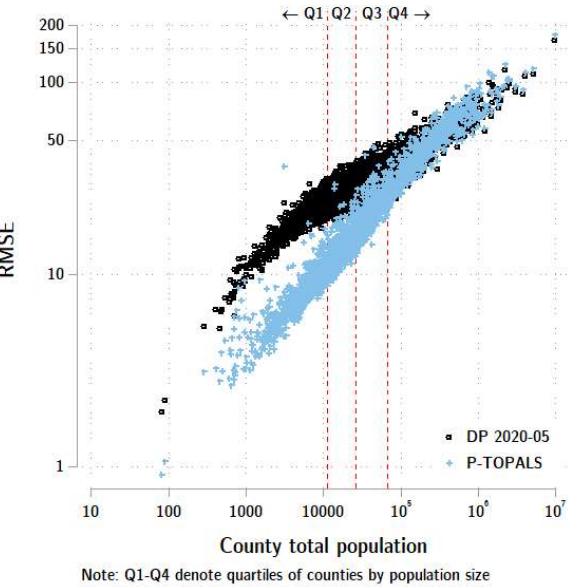
- US Census Bureau will apply differential privacy to protect confidentiality of responses to the 2020 Census.
- Can we leverage demographic methods to improve the accuracy of the age structured data -- without compromising the privacy guarantee?

METHODS

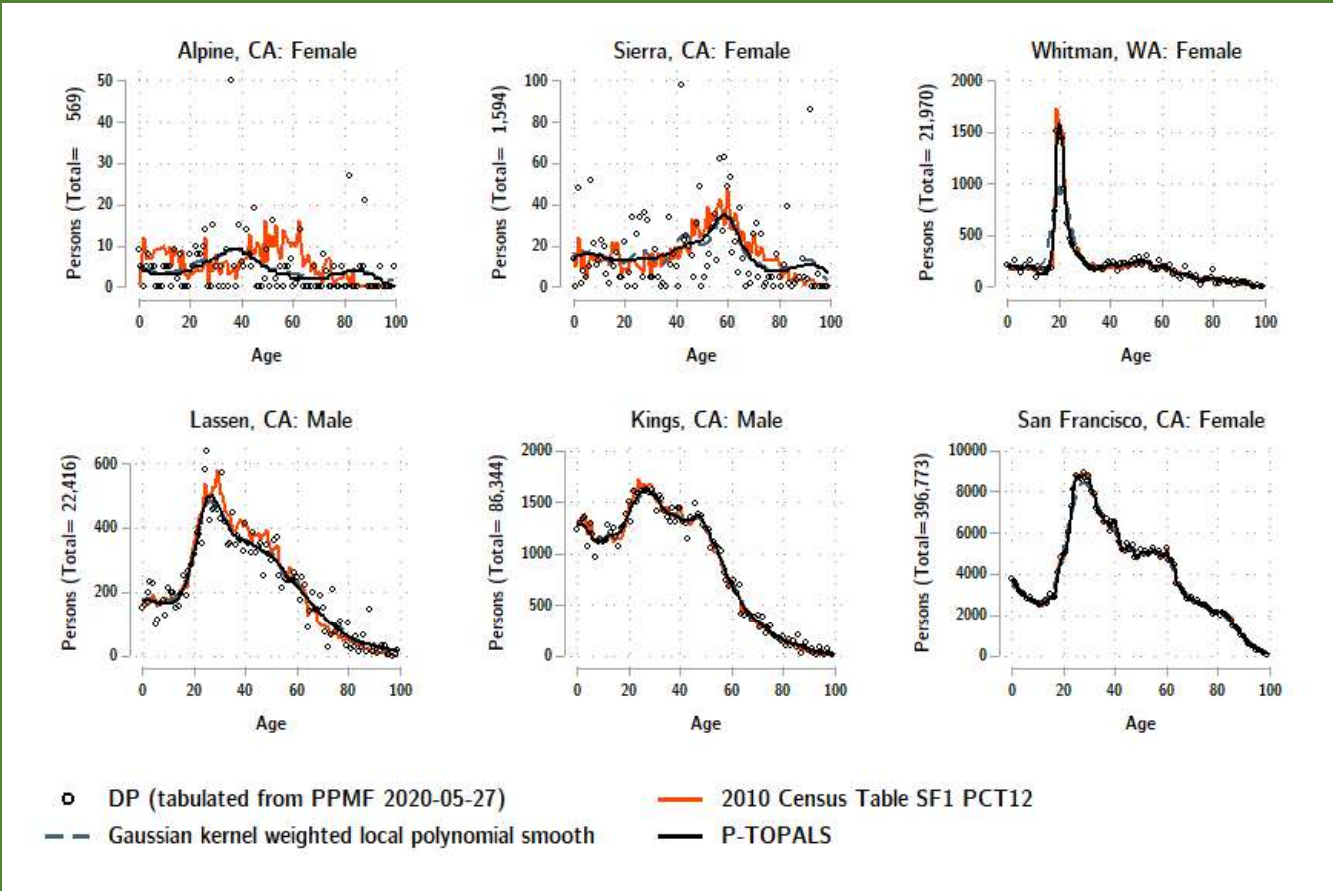
- Downloaded the 2020-05 PPMF release from the U.S. Census Bureau and tabulated the data into specific planned county-level table products.
- Applied P-TOPALS (Dyrting 2020), a novel smoothing method using penalized splines.
- Compared the RMSE of the DP data versus the PT results + Accuracy of derived measures (life expectancy, TFR, etc).

RESULTS

- RMSE declines for small populations, meaning more counties attain a coefficient of variation within usable standards



Demographic age smoothing techniques can reduce the observed differences between a differentially private population dataset and population counts by age as enumerated.



SO WHAT

Measures that use age detail from the census (there are **many**) become more accurate ...

		Total Fertility Rate (CA births 2009-11)		
P7 table (alone)	Race	SF	DP	PT (U)
B	Black	1.61	3.15	2.12
C	AIAN	1.02	2.07	1.23
D	Asian	1.89	3.50	2.67
E	NHPI	2.05	5.16	2.93
H	Hispanic	2.37	2.71	2.40
I	White NH	1.82	1.94	1.80
-	Total	2.03	2.07	2.01

Fertility rates for example: average error in TFR declines from 65% to 25%.

sharygin@pdx.edu
@ejsharygin

