Impacts of Differential Privacy on data about American Indian and Alaska Native People and Tribal Lands

Carolyn A. Liebler, Ph.D.
Associate Professor
Department of Sociology &
Minnesota Population Center
University of Minnesota

Data on tribal lands & AIAN people is hard to get! Census data is often the only data source.

Core users include

- Tribal leaders for tribal governance, local governments, state and federal governments, researchers, judges, lawmakers, policy makers, advocacy groups

Some uses of these data

- Funding allocations, policy development, policy evaluation, planning
- Fertility rate denominators, other vital rate denominators
- Water rights and other litigation and judicial decisions
- Enforcement of equal employment, housing, and other anti-discrimination laws

Impacts of Differential Privacy on data about American Indian and Alaska Native People and Tribal Lands

AIAN people

People reported as American Indian and/or Alaska Native in the Census race question, including people who marked other races.

Note that this is self-described race; tribal membership is not required.

% difference = $100*(Pop_{DP}-Pop_{SF})/Pop_{SF}$

Percent difference between the Differentially Private population size (Pop_{DP}) and the population size in the 2010 SF data (Pop_{SF})

A positive % difference indicates that the DP number is larger than the original number.

Example:
3,000 people in the original data: $Pop_{SF} = 3,000$
3,300 people in the DP data: $Pop_{DP} = 3,300$
% difference = 100 * (3,300-3,000)/3,000 = 300/3000*100
= the numbers are 10% different and the DP version is larger

Figure 1

Extremes within states of % diff btwn DP and non-DP total AIAN population in counties. **Finalized and unchangeable DP settings.**

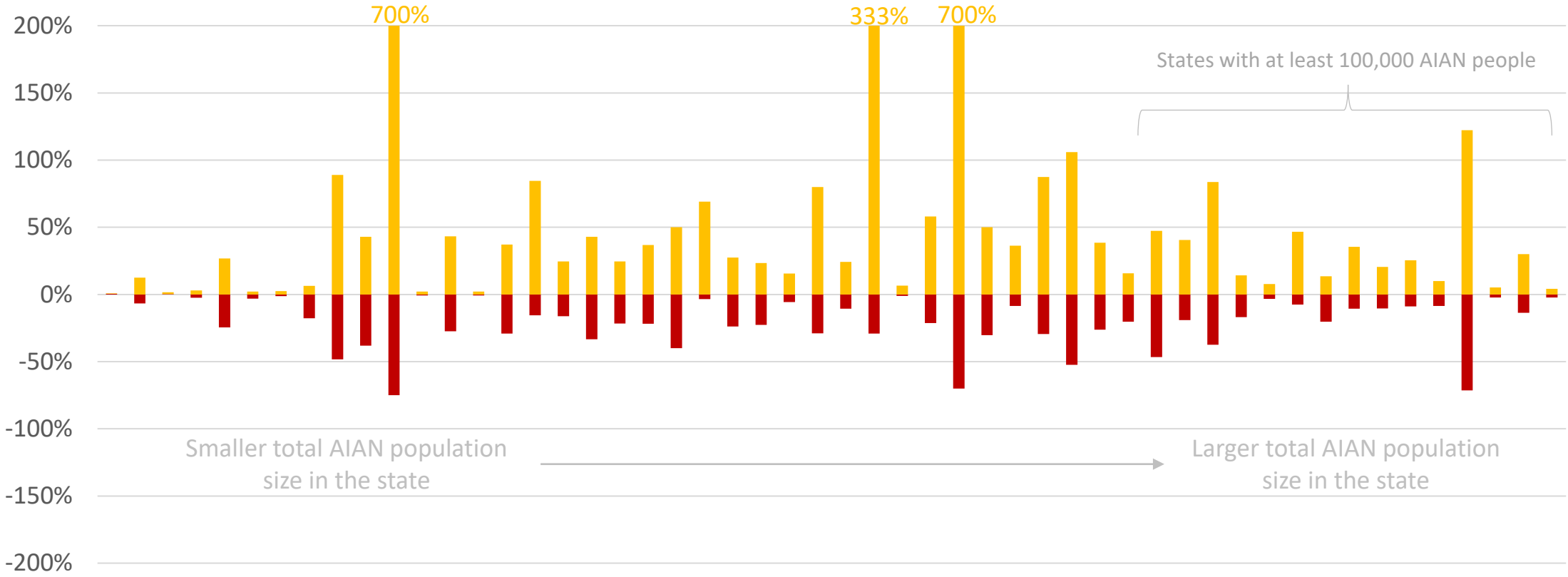
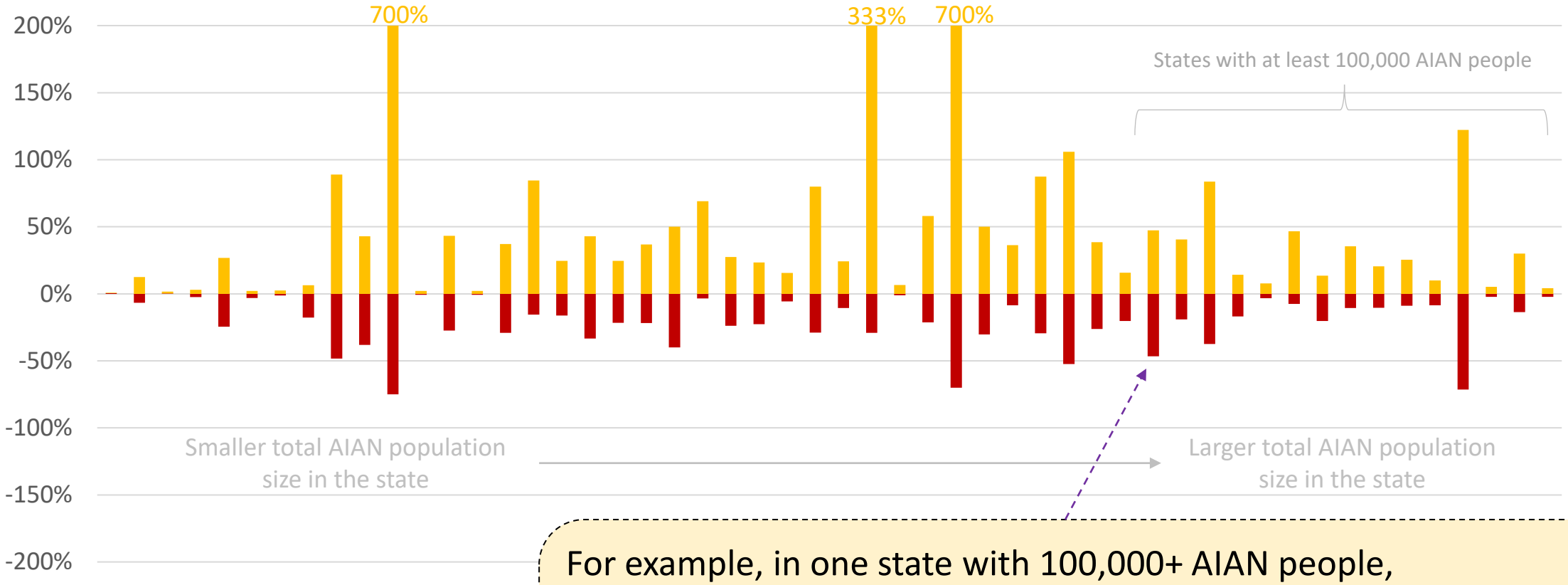


Figure 1

Extremes within states of % diff btwn DP and non-DP total AIAN population in counties. **Finalized and unchangeable DP settings.**



For example, in one state with 100,000+ AIAN people,

- the DP data in one county shows the AIAN population listed as 47% lower than is true.
- In a different county in the same state, the DP data shows the AIAN population listed as 47% higher than is true.

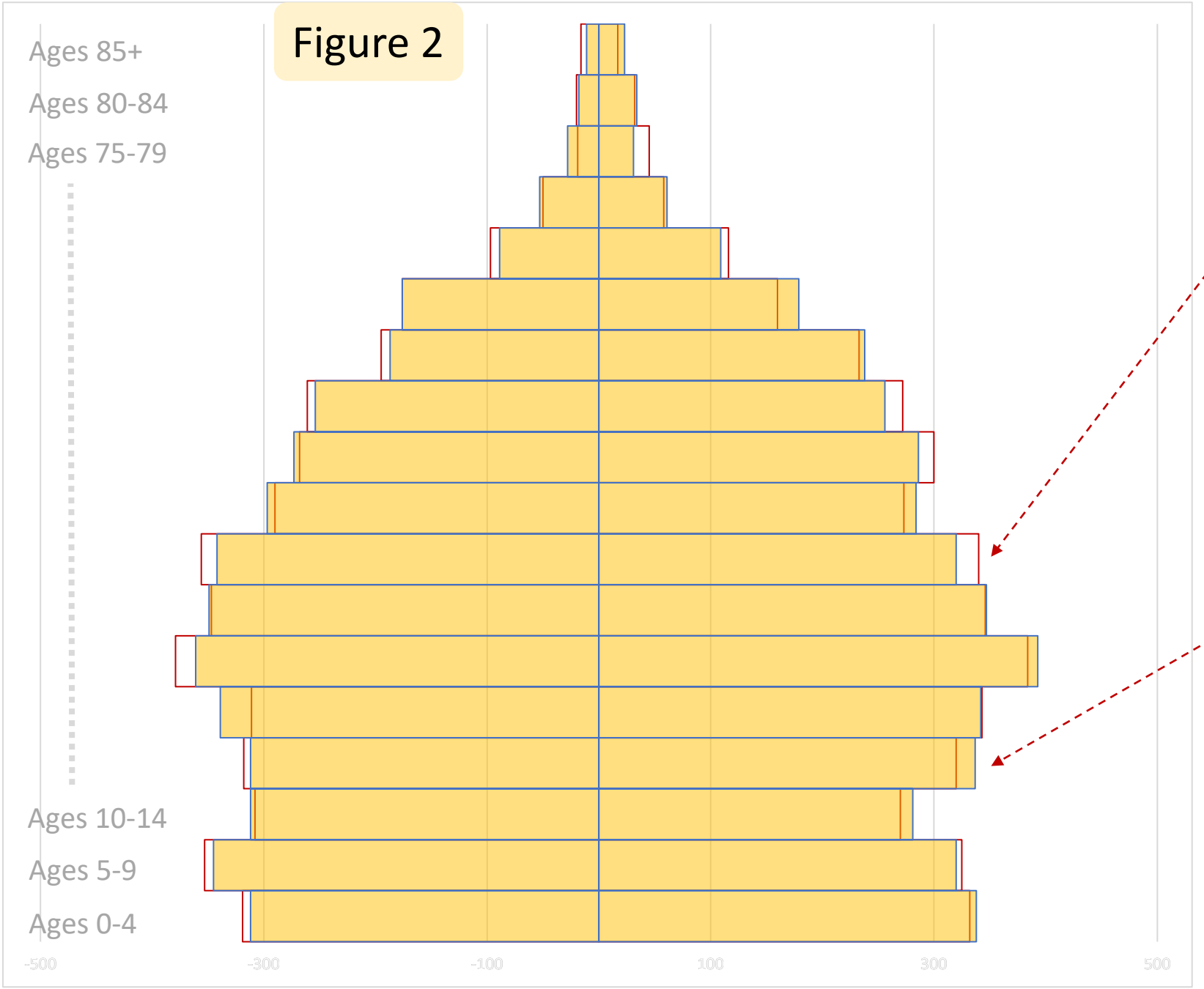
Single-race AIAN populations in counties

220 counties have at least 2,000 single-race AIAN,
within states with at least 100,000 AIAN AOIC

Need accurate age and sex structure for policies and planning

Need accurate age by sex by race for fertility rate denominators

Figure 2



In this county, there were 320 single-race AIAN women ages 35-39 but DP said there were 340

% diff = 6%

In this county, there were 337 single-race AIAN women ages 15-19 but DP said there were 320

% diff = -5%

A county with at least 2,000 single-race AIAN people, within a state with at least 100,000 AIAN people

Table 1 **Birth rates to single-race AIAN women in a county in 2010, with non-DP and with DP**

	Births to AIAN mothers	Pop. of AIAN women		Birth rate per 1000 AIAN	
		SF	DP	SF	DP
Age 10-17	3				
Age 18-19	14	155	139	90.3	100.7
Age 20-29	39	735	727	53.1	53.6
Age 30+	31				

SF = Actual data in 2010

DP = Data with differential privacy settings as of 3/16/2022

Source for birth data: https://www.cohealthdata.dphe.state.co.us/chd/Resources/vs/2010/Denver.pdf

Table 1 **Birth rates to single-race AIAN women in a county in 2010, with non-DP and with DP**

	Births to AIAN mothers	Pop. of AIAN women		Birth rate per 1000 AIAN	
		SF	DP	SF	DP
Age 10-17	3				
Age 18-19	14	155	139	90.3	100.7
Age 20-29	39	735	727	53.1	53.6
Age 30+	31				

SF = Actual data in 2010

DP = Data with differential privacy settings as of 3/16/2022

Percent difference between SF=735 and DP=727 is 100*[(727-735)/735] = 100*[(-8/735] or 1.09% . No problem.

Source for birth data: https://www.cohealthdata.dphe.state.co.us/chd/Resources/vs/2010/Denver.pdf

Table 1 **Birth rates to single-race AIAN women in a county in 2010, with non-DP and with DP**

	Births to AIAN mothers	Pop. of AIAN women		Birth rate per 1000 AIAN	
		SF	DP	SF	DP
Age 10-17	3				
Age 18-19	14	155	139	90.3	100.7
Age 20-29	39	735	727	53.1	53.6
Age 30+	31				

SF = Actual data in 2010

DP = Data with differential privacy settings as of 3/16/2022

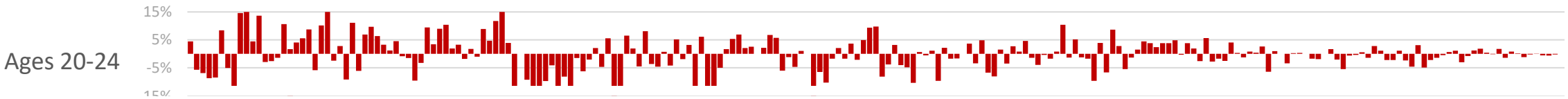
Percent difference between SF=155 and DP=139 is 100*[(139-155)/155] = 100*[(-16/155] or 10.3%

Percent difference between SF=735 and DP=727 is 100*[(727-735)/735] = 100*[(-8/735] or 1.09% . No problem.

This difference would probably cause alarm!

Source for birth data: https://www.cohealthdata.dphe.state.co.us/chd/Resources/vs/2010/Denver.pdf

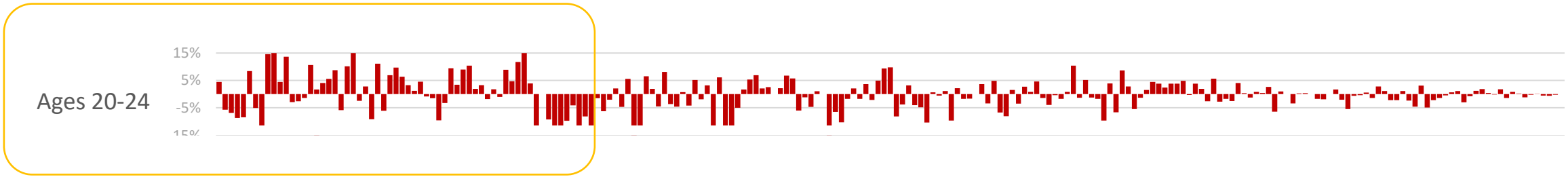
Figure 3 Percent difference in non-DP and DP versions of the number of single-race AIAN women of fertility ages in 220 counties with 2,000+ AIAN people



Smaller total AIAN population size → Larger total AIAN population size

Note: Percent differences range from -65% to 52% in these counties.

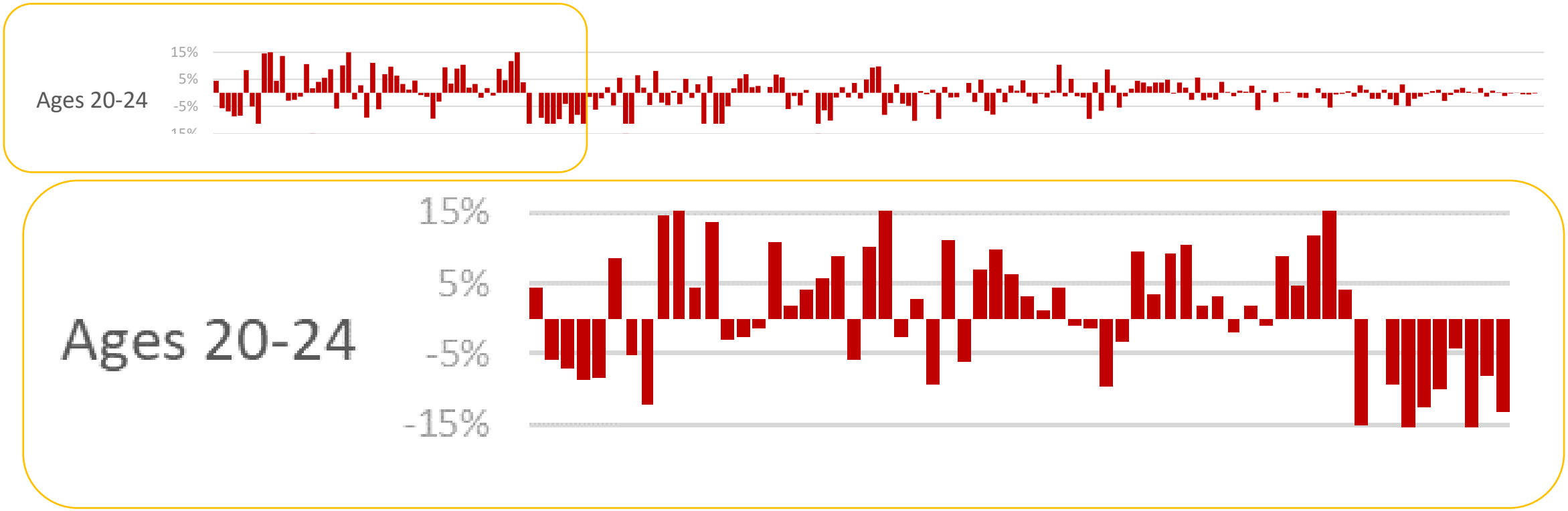
Figure 3 Percent difference in non-DP and DP versions of the number of single-race AIAN women of fertility ages in 220 counties with 2,000+ AIAN people



Smaller total AIAN population size → Larger total AIAN population size

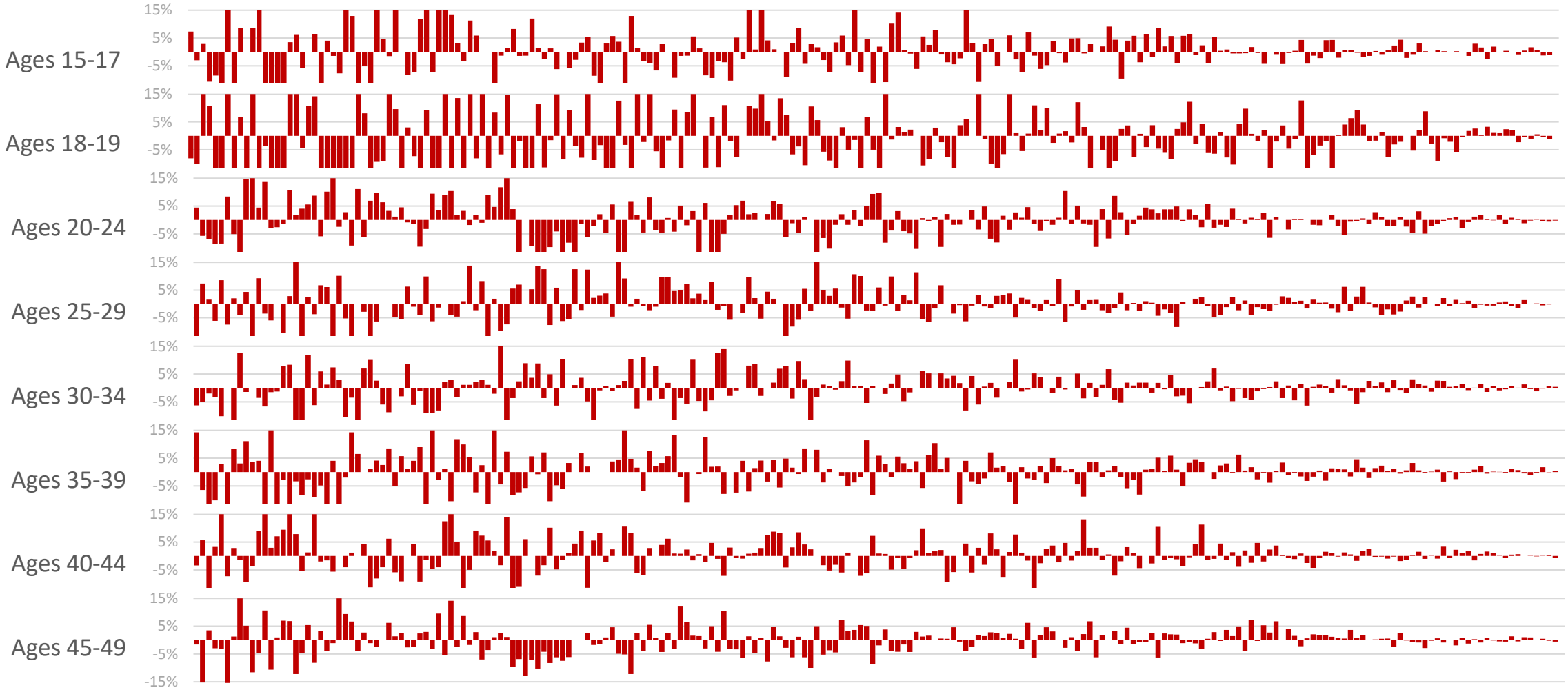
Note: Percent differences range from -65% to 52% in these counties.

Figure 3 Percent difference in non-DP and DP versions of the number of single-race AIAN women of fertility ages in 220 counties with 2,000+ AIAN people



Note: Percent differences range from -65% to 52% in these counties.

Figure 3 Percent difference in non-DP and DP versions of the number of single-race AIAN women of fertility ages in 220 counties with 2,000+ AIAN people



Note: Percent differences range from -65% to 52% in these counties.

Impacts of Differential Privacy on data about American Indian and Alaska Native People and Tribal Lands

Need accurate population totals for every reason!

(Later) will need accurate tribe responses for programs

Housing information on own/rent, overcrowding: for policies and planning

Tribal lands

Reservations and other land areas defined and used by tribal governments.

History is complex so there are many legal statuses of lands.

- People of any race(s) in these areas
- Borders cross state and county boundaries

Tribal governance requires accurate, geographically-detailed information at the tribal tract level.

Tribal census tracts are commonly used for governance. They nest within tribal lands and have a minimum population of 1,200, a maximum population of 8,000, and an optimum population of 4,000.*

*Source: https://www.census.gov/newsroom/blogs/random-samplings/2012/07/decoding-state-county-census-tracts-versus-tribal-census-tracts.html

I use the term "tribal lands" to describe Census Bureau geographies labeled "AIANHH" (not including those in Hawaii). See https://www.ncai.org/policy-research-center/research-data/prc-publications/Overview_of_2020_AIAN_Redistricting_Data_FINAL_8_13_2021.pdf

For the 2020 Census Tribal Tract Reference map, see: https://www.census.gov/geographies/reference-maps/2020/geo/2020-census-tribal-tract.html

Source: https://mapsmith.net/wp-content/uploads/2017/06/reservations_poster_final-5k.jpg

Table 2 **Percent difference between non-DP and DP versions of the 2010 total population size on AIAN tribal lands**

(whole or disaggregated lands). **Finalized and unchangable DP settings.**

Panel A:

Entire tribal land, not disaggregated (p_aianhh)

Population size range of tribal lands	# of tribal lands	Extremes of % difference in total pop size		
		Min	Mean	Max
10 to 99	115	-23%	-1%	14%
100 to 499	183	-6%	0%	7%
500 to 999	83	-9%	0%	2%
1,000 to 2,999	55	-5%	0%	2%
3,000 to 4,999	38	-3%	0%	0%
5,000 to 9,999	36	-3%	0%	0%
10,000 or more	52	0%	0%	0%

DP = Differentially Private

Numbers larger than 5% or smaller than -5% are highlighted in red.

Tribal land population totals have been released. The current DP settings have reported tribal lands as substantially larger or smaller than the true enumeration.

← For example, there were 83 whole tribal lands with 500 to 999 population. With differential privacy, their total populations were reported as up to 2% larger than enumerated or up to 9% smaller than enumerated.

Table 2 **Percent difference between non-DP and DP versions of the 2010 total population size on AIAN tribal lands**

(whole or disaggregated lands). **Finalized and unchangable DP settings.**

Panel A:

Entire tribal land, not disaggregated (p_aianhh)

Population size range of tribal lands	# of tribal lands	Extremes of % difference in total pop size		
		Min	Mean	Max
10 to 99	115	-23%	-1%	14%
100 to 499	183	-6%	0%	7%
500 to 999	83	-9%	0%	2%
1,000 to 2,999	55	-5%	0%	2%
3,000 to 4,999	38	-3%	0%	0%
5,000 to 9,999	36	-3%	0%	0%
10,000 or more	52	0%	0%	0%

Panel B:

Disaggregated tribal lands (p_aianhh_144)

Population size range of tribal lands	# of tribal lands	Extremes of % difference in total pop size		
		Min	Mean	Max
10 to 99	244	-58%	0%	27%
100 to 499	336	-14%	0%	27%
500 to 999	153	-9%	0%	4%
1,000 to 2,999	555	-12%	0%	4%
3,000 to 4,999	468	-2%	0%	1%
5,000 to 9,999	253	-1%	0%	0%
10,000 or more	4	0%	0%	0%

DP = Differentially Private

Numbers larger than 5% or smaller than -5% are highlighted in red.

Table 3 **Percent difference between non-DP and DP versions of the population size in 2010 of American Indian and Alaska Natives (AIAN)** (alone or in combination with other races) **on AIAN tribal lands** (whole or disaggregated lands). **DP settings are not finalized.**

Panel A:
Entire tribal land, not disaggregated (p_aianhh)

Population size range of tribal lands	# of tribal lands	Extremes of % difference in AIAN pop size		
		Min	Mean	Max
10 to 99	115	-100%	2%	250%
100 to 499	182	-17%	0%	56%
500 to 999	83	-22%	0%	13%
1,000 to 2,999	55	-57%	-1%	31%
3,000 to 4,999	38	-3%	1%	10%
5,000 to 9,999	36	-17%	-1%	3%
10,000 or more	52	-5%	0%	12%

Panel B:
Disaggregated tribal lands (p_aianhh_144)

Population size range of tribal lands	# of tribal areas	Extremes of % difference in AIAN pop size		
		Min	Mean	Max
10 to 99	221	-100%	11%	900%
100 to 499	328	-100%	12%	1000%
500 to 999	153	-73%	-2%	155%
1,000 to 2,999	555	-45%	1%	136%
3,000 to 4,999	468	-34%	0%	47%
5,000 to 9,999	253	-18%	0%	15%
10,000 or more	4	-5%	-1%	0%

DP = Differentially Private

Numbers larger than 5% or smaller than -5% are highlighted in red.

Data: nhgis_ppdd_20220316_p_aianhh and nhgis_ppdd_20220316_p_aianhh_144

Housing characteristics on tribal lands

- very good results (~2% diff) except if less than 100 population for:
  - Percent of all households who own with loan vs. own free and clear vs. rent
  - Reason for vacancy
  - AIAN householder: owns vs. rents

- Too much noise, even in bigger areas ☹
  - AIAN householder is a renter: Household size
  - Householder is a renter: Presence of children under 18

Data: nhgis_ppdd_20220316_p_aianhh_144

Final thoughts

- Data on tribal lands & AIAN people is hard to get! Census data is often the *only* data source.

Final thoughts

- Data on tribal lands & AIAN people is hard to get! Census data is often the *only* data source.

- Additional DP on the new detail of data has a compounded effect
  - Always keep in mind that tribal land population totals have already been released for 2020. Differential privacy has made them look much larger or smaller than the true enumeration.

Final thoughts

- Data on tribal lands & AIAN people is hard to get! Census data is often the *only* data source.

- Additional DP on the new detail of data has a compounded effect
  - Always keep in mind that tribal land population totals have already been released for 2020. Differential privacy has made them look much larger or smaller than the true enumeration.

- Especially need accurate population totals used for denominators
  - 5-year age group by sex by race for mortality and fertility rates.

Data for this research: 2010 Data Demonstration Products (vintage 3/16/2022), downloaded from IPUMS NHGIS: https://www.nhgis.org/privacy-protected-2010-census-demonstration-data#v20220316 .

IPUMS NHGIS is cited as: Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS. 2021. http://doi.org/10.18128/D050.V16.0

I am grateful to Randall Akee and Yvette Roubideaux for conversations supporting this presentation.