Differential Privacy and Census Data: 2020 Census Demographics and Housing Characteristics File

Jonathan Buttle
California Department of Finance
Demographic Research Unit

What We Will Cover

- Why Differential Privacy?
- What is Differential Privacy?
- The components of Differential Privacy
- The Differential Privacy Mechanism
- Census Bureau and Differential Privacy
- 2020 Demographics and Housing Characteristics File and Privacy
- 2020 DHC Demonstration Products
- Submitting Feedback to the Census Bureau

Why Differential Privacy?

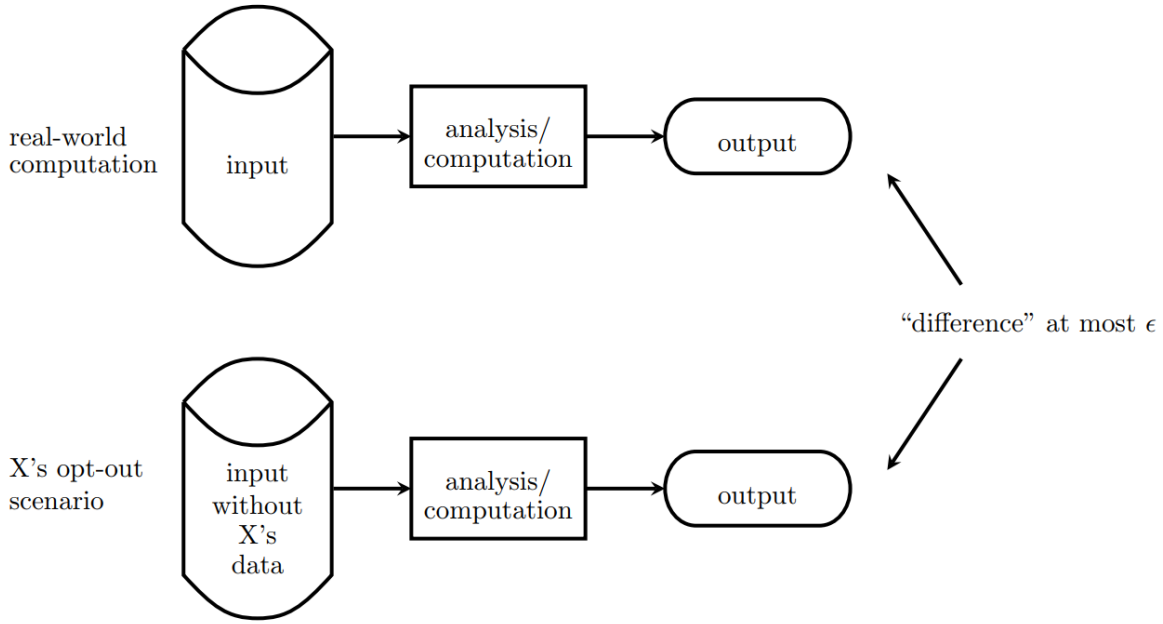
- Title 13 specifies that “the Census Bureau shall not make any publication whereby the data furnished by any particular establishment or individual ... can be identified” (Title 13 U.S.C. § 9(a)(2), Public Law 87-813);
- Title 5 further prohibits “any representation of information that permits the identity of the respondent to whom the information applies to be reasonably inferred by either direct or indirect means” (Title 5 U.S.C. §502 (4), Public Law 107–347);

What is Differential Privacy?

- Differential Privacy (DP) is a mathematical technique that allows for the formal quantification of the risk of data disclosure;
- Formally, DP is a property of algorithms for answering queries;
- As a result, DP allows for mathematically quantifying the risk of identifying a specific element in a dataset;
- Specifically, differentially private algorithms provide formal bounds as to how many queries can be made before the probability of learning specific information about a database increases beyond acceptable levels.

What is Differential Privacy?

- Imagine two databases – one that contains your information and one that doesn't. DP stipulates that the probability that the result generated by a statistical query from either database will be (nearly) identical.
- In other words, an observer cannot determine which is the true database by observing the output.
- DP 'quantifies' the effect your information will have on a query



The Components of Differential Privacy

- The privacy loss budget. The privacy loss budget is typically represented by epsilon (ε).
- When $\varepsilon = 0$, the resulting data would be random and essentially useless (perfect privacy).
- When $\varepsilon = \infty$, the resulting data would allow for full identification of survey participants (perfect accuracy).
- Values of epsilon between 0 and ∞ represent a trade off between privacy and accuracy.

The Privacy Budget

- An alternative interpretation of epsilon is that of a “privacy budget”.
- If only a single query on the data is expected to be performed, that query might use up the entirety of the budget;
- However, performing a series of queries on the data requires allocation of the budget over all the queries;
- There are two methods of allocating the privacy budget – sequential and parallel.

The Privacy-Accuracy Tradeoff

This graph illustrates the privacy-accuracy trade off for a privacy mechanism with epsilon values between 1 and 6.

Census Bureau and DP

- Differential Privacy Implementation for the 2020 Census.
  - Employs top-down methodology.
  - Creates a histogram of demographic attributes (total population, voting age, race/ethnicity, group quarters type, and combinations of attributes).
  - Assigns them iteratively to various geographies (Nation, State, County, Place, Tract, Block Group, Block, etc.).
  - Applies ‘noise’ to the attributes by adding results from random number generator (discrete Gaussian) to the attribute counts.
  - Post-processes the resulting noisy data subject to ‘invariants’ – total population at the state and national level, and total housing unit and group quarters counts at the block level.

The DP Mechanism – Noise

- The DP mechanism works by injecting statistically calibrated ”noise” into the data.
- The amount of noise injected is determined by three parameters:
  - Epsilon – the privacy loss budget;
  - Sensitivity – the amount that one or more individuals (or records) can influence the output of the mechanism; and
  - Delta – for ‘nearly’ pure DP, the probability of a catastrophic data breach.
- Statistical “noise” is typically derived from two distributions:
  - The geometric distribution (the discrete variant of the Laplace distribution)
    - ✓ Returns pure DP (used in previous 2020 Census DP testing); and
  - The discrete Gaussian distribution (the discrete variant of the Normal distribution).
    - ✓ Returns ‘nearly’ pure DP (current noise engine).

Post-Processing

- One important characteristic of DP is that once a dataset has been privatized through a DP algorithm, additional processing on the privatized dataset maintains the differential privacy;
- Therefore, additional data processing can address issues such as:
  - Counts less than zero;
  - Ensuring the sum of counts for lower geographies are equal to counts for higher geographies (for example, the sum of the counts for all counties in a state equal the total count for the state).

2010 Census DAS-DHC Changes

- The DHC implementation includes updates to the geographic hierarchy –
- Added new geographic category - “Population Estimates Primitive Geographies”:
  - Most granular geographic unit used by the Census Bureau’s Population Estimates Program to derive tables for every geography with population estimates;
  - Because the primitive geographies do not always align with census tracts, the hierarchy also incorporates tracts subsets and tract groups;
  - Tract subsets are the intersection of Population Estimates Geographies with tabulation tracts;
  - Tract subset groups are the union of multiple tracts subsets within the same primitive estimates geography;
  - These geographies do not impact how the data tabulated within the DHC;

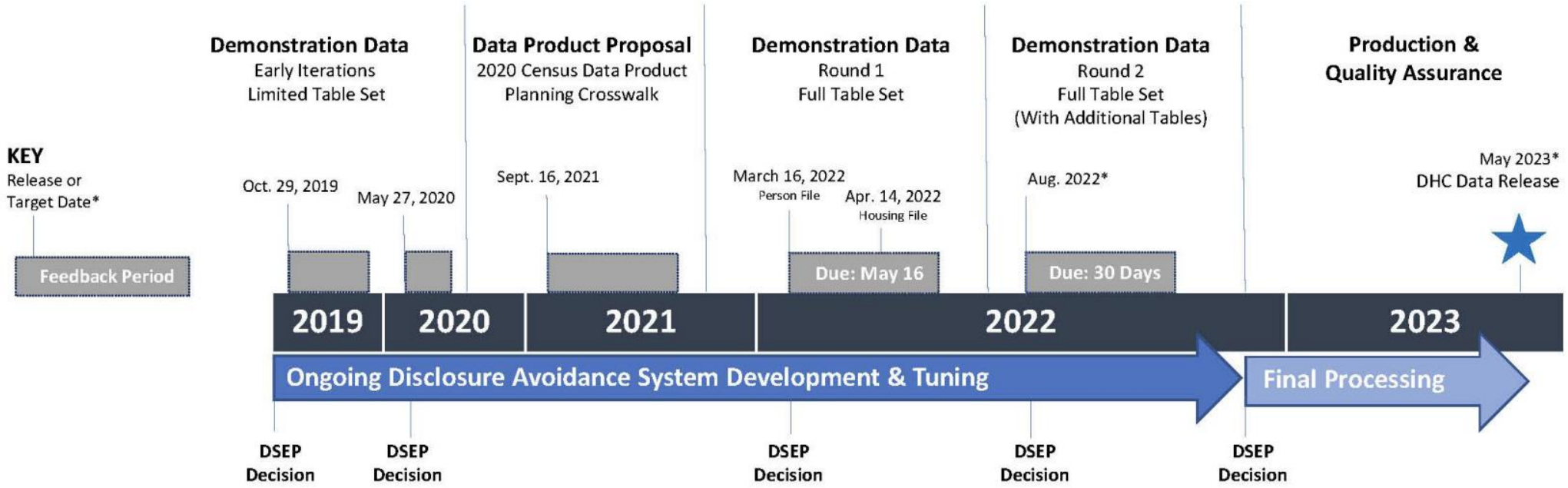
2010 Census DAS-DHC Changes

- The second release of the 2010 DHC-DAS demonstration incorporates changes to the DAS from previous versions;
- Mechanism changes:
  - Increased the global epsilon for units tables from 3.87 to 6.14;
  - Revised distribution of privacy-loss budgets to favor national and state tables at the expense of block and block group tables (county and tract tables relatively unchanged);
- Table changes:
  - A total of 249 tables are proposed for inclusion in the 2020 DHC release;
  - Reduced the geography to census tract level from state or county for 16 tables and 63 iterated tables or tables repeated by race and ethnicity;
  - Added sex by single year of age table repeated by race and ethnicity at the tract level;

2020 Census Product Release Schedule

- Released:
  - Apportionment File – April 26, 2021
  - Redistricting File – August 12, 2021 (FTP)/September 16, 2021 (data.census.gov)
- Planned Future Releases:
  - Demographic Profile/DHC – May 2023
  - Detailed DHC-A (Total population and sex by age by detailed race/ethnicity) – August 2023
  - Detailed DHC-B (Household and tenure by detailed race/ethnicity) – TBD
  - Supplemental DHC (S-DHC – People in households) – TBD
- Future Efforts (include PUMS File and Special Tabulations) – TBD

2020 Census Product Development Schedule



As of August, 31, 2022. Source: US Census Bureau

The DP Mechanism - Parameters

- The August 2022 DP mechanism release incorporates the following parameter distributions:
  - ➢ Person Tables:
    - Global ρ (rho): 3.65 (which returns global ε $(epsilon)$: 21.97);
    - Global δ $(delta)$: 10^{-10}
  - ➢ Units (Housing) Tables:
    - Global ρ (rho): 6.14 (which returns global ε $(epsilon)$: 29.92);
    - Global δ $(delta)$: 10^{-10}

The DP Mechanism - Parameters

- The percent distributions for ρ (rho) by geometry are:

Geography - Persons	*rho* Allocation by Geographic Level
US	2.0%
State	27.4%
County	8.5%
Population Estimates Primitive Geography†	13.1%
Tract Subset Group‡	13.1%
Tract Subset‡	23.8%
Optimized Block Group◊	11.8%
Block	0.3%

Geography - Units	*rho* Allocation by Geographic Level
US	7.87%
State	29.43%
County	11.79%
Population Estimates Primitive Geography†	11.79%
Tract Subset Group‡	11.79%
Tract Subset‡	20.13%
Optimized Block Group◊	6.94%
Block	0.26%

†Population Estimates Primitive Geographies are the most granular geographic unit used by the Census Bureau's Population Estimates Program. These geographic units are the most granular geographic areas that are required in order to derive tables for every geography for which official population estimates are produced.

‡ Tract Subsets are defined as the intersection of Population Estimates Primitive Geographies with census tabulation tracts. Tract Subset Groups are defined as the union of multiple tract subsets that are all within the same Population Estimates primitive geography.

◊ Optimized Block Groups are defined as sequentially grouped blocks within the same Tract Subset in the order of the geoid until either there are no more blocks within the Tract Subset left or there are sqrt(number_of_blocks_in_tract_subset) + 13 blocks in the block group.

The DP Mechanism - Parameters

- The percent distributions for ρ (rho) by geographical level are:

Query - Persons	Per Query *rho* Allocation by Geographic Level							
	US	State	County	Population Estimates Primitive Geography†	Tract Subset Group‡	Tract Subset‡	Optimized Block Group◊	Block
AGE_18_64_116 * RELGQ_4_GROUPS	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
AGE_18_64_116 * SEX	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
AGE_26_GROUPS * SEX	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
HISPANIC * SEX	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
SEX * RELGQ_4_GROUPS	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
GQ_CONSTR_GROUPS * AGE_10_GROUPS	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
POPSEHSDTARGETSRELSHIP	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
HISPANIC * SEX * AGE_29_GROUPS * RELSHIP_AND_EIGHT_LEVEL_GQ * CENRACE	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
RELGQ * AGE_29_GROUPS * HISPANIC *	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%
DETAILED	0.20%	2.74%	0.85%	1.31%	1.31%	2.38%	1.18%	0.03%

The DP Mechanism - Parameters

- The percent distributions for ρ (rho) by geographical level are:

Query - Units	Per Query *rho* Allocation by Geographic Level							
	US	State	County	Population Estimates Primitive Geography†	Tract Subset Group‡	Tract Subset‡	Optimized Block Group◊	Block
DETAILED	1.71%	7.10%	1.81%	1.58%	1.58%	5.03%	1.73%	0.07%
SEX * HISP * HHTENSHORT_3LEV * RACE * FAMILY_NONFAMILY_SIZE	0.00%	0.00%	0.00%	0.00%	0.00%	5.03%	1.73%	0.07%
SEX * HISP * HHTENSHORT_3LEV * RACE * HHAGE * FAMILY_NONFAMILY_SIZE	0.00%	0.00%	0.00%	0.00%	0.00%	5.03%	1.73%	0.07%
TENVACGQ	0.42%	5.81%	1.81%	1.58%	1.58%	5.03%	1.73%	0.07%
MULTIG * HISP * HHTENSHORT_2LEV	1.29%	1.29%	1.29%	1.58%	1.58%	0.00%	0.00%	0.00%
HISP*HHTENSHORT_2LEV	0.35%	0.35%	0.35%	0.35%	0.35%	0.00%	0.00%	0.00%
PARTNER_TYPE_OWN_CHILD_STATUS * SEX * HHTENSHOT_2LEV	1.29%	1.29%	1.29%	1.58%	1.58%	0.00%	0.00%	0.00%
COUPLED_HH_TYPE * HISP * HHTENSHORT_2LEV	1.29%	1.29%	1.29%	1.58%	1.58%	0.00%	0.00%	0.00%
SEX * HISP * HHTENSHORT_3LEV * RACE * DETAILEDCOUPLETYPEMULTGENDETOWN CHILDSIZE	0.42%	5.81%	1.81%	1.58%	1.58%	0.00%	0.00%	0.00%
CHILDSIZE	0.42%	5.81%	1.81%	1.58%	1.58%	0.00%	0.00%	0.00%
HHTENSHORT_3LEV * HHAGE * DETAILEDCOUPLETYPEMULTGENDETOWN CHILDSIZE	0.35%	0.35%	0.35%	0.35%	0.35%	0.00%	0.00%	0.00%
HISP * HHTENSHORT_2LEV * RACE	0.35%	0.35%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

The DP Mechanism - Parameters

- The σ^2 distributions for persons tables by geographical level are:

Query - Persons	Per Query *sigma*^{2} Allocation by Geographic Level							
	US	State	County	Population Estimates Primitive Geography†	Tract Subset Group‡	Tract Subset‡	Optimized Block Group◊	Block
AGE_18_64_116 * RELGQ_4_GROUPS	274	20	64	42	42	23	46	1,826
AGE_18_64_116 * SEX	274	20	64	42	42	23	46	1,826
AGE_26_GROUPS * SEX	274	20	64	42	42	23	46	1,826
HISPANIC * SEX	274	20	64	42	42	23	46	1,826
SEX * RELGQ_4_GROUPS	274	20	64	42	42	23	46	1,826
GQ_CONSTR_GROUPS * AGE_10_GROUPS	274	20	64	42	42	23	46	1,826
POPSEHSDTARGETSRELSHIP	274	20	64	42	42	23	46	1,826
HISPANIC * SEX * AGE_29_GROUPS * RELSHIP_AND_EIGHT_LEVEL_GQ * CENRACE	274	20	64	42	42	23	46	1,826
RELGQ * AGE_29_GROUPS * HISPANIC * CENRACE * SEX	274	20	64	42	42	23	46	1,826
DETAILED	274	20	64	42	42	23	46	1,826

Source: author's calculations.

The DP Mechanism - Parameters

- The σ^2 distributions for units tables by geographical level are:

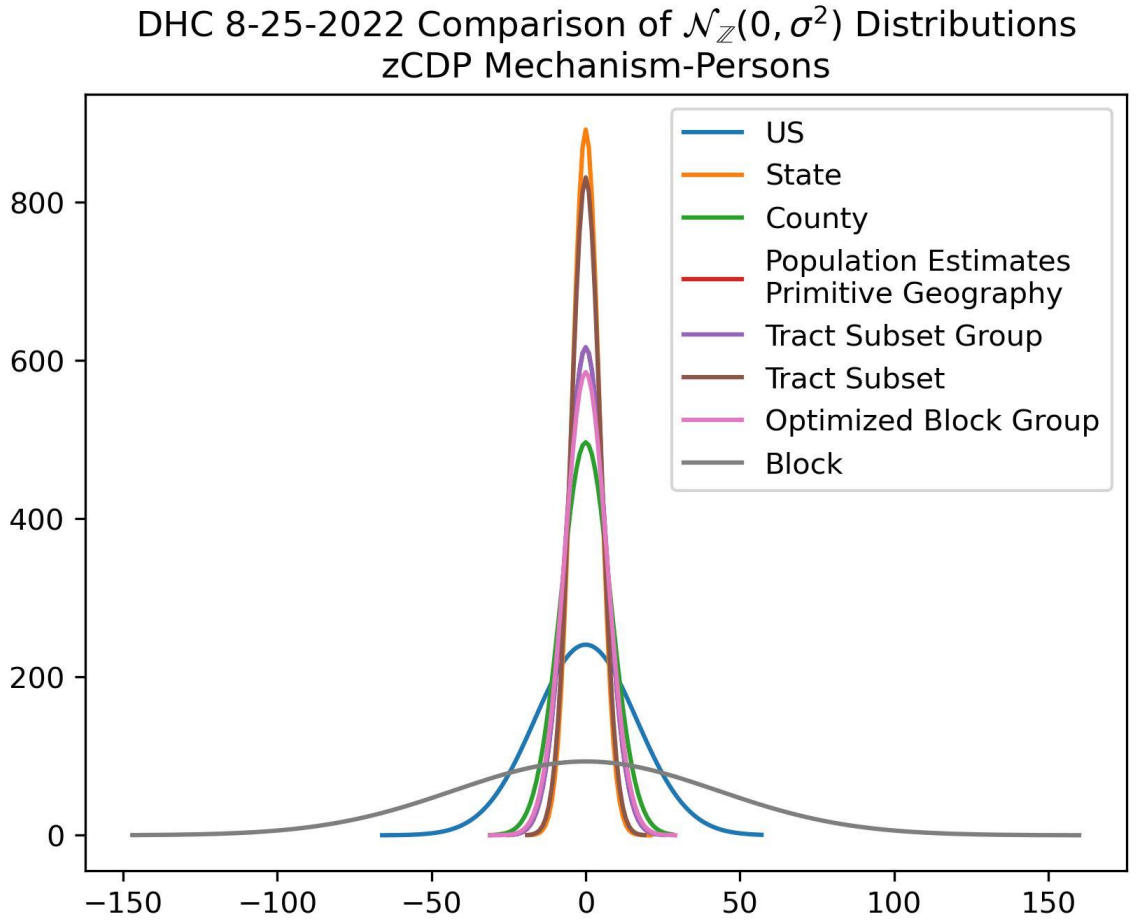
Query - Units	Per Query *sigma*^{2} Allocation by Geographic Level							
	US	State	County	Population Estimates Primitive Geography†	Tract Subset Group‡	Tract Subset‡	Optimized Block Group◊	Block
DETAILED	19	5	18	21	21	6	19	465
SEX * HISP * HHTENSHORT_3LEV * RACE * FAMILY_NONFAMILY_SIZE	NA	NA	NA	NA	NA	6	19	465
SEX * HISP * HHTENSHORT_3LEV * RACE * HHAGE * FAMILY_NONFAMILY_SIZE	NA	NA	NA	NA	NA	6	19	465
TENVACGQ	78	6	18	21	21	6	19	465
MULTIG * HISP * HHTENSHORT_2LEV	25	25	25	21	21	NA	NA	NA
HISP*HHTENSHORT_2LEV	93	93	93	93	93	NA	NA	NA
PARTNER_TYPE_OWN_CHILD_STATUS * SEX * HHTENSHOT_2LEV	25	25	25	21	21	NA	NA	NA
COUPLED_HH_TYPE * HISP * HHTENSHORT_2LEV	25	25	25	21	21	NA	NA	NA
SEX * HISP * HHTENSHORT_3LEV * RACE * DETAILEDCOUPLETYPEMULTGENDETOWN CHILDSIZE	78	6	18	21	21	NA	NA	NA
HHAGE * DETAILEDCOUPLETYPEMULTGENDETOWN CHILDSIZE	78	6	18	21	21	NA	NA	NA
HHTENSHORT_3LEV * HHAGE * DETAILEDCOUPLETYPEMULTGENDETOWN CHILDSIZE	93	93	93	93	93	NA	NA	NA
HISP * HHTENSHORT_2LEV * RACE	93	93	NA	NA	NA	NA	NA	NA

Source: author's calculations.

The DP Mechanism - Parameters

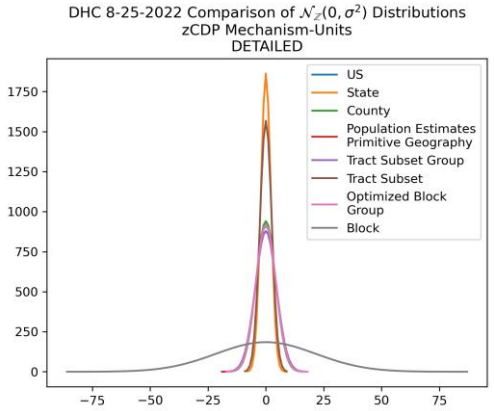
- The σ^2 distributions for persons tables by geographical level (graph);

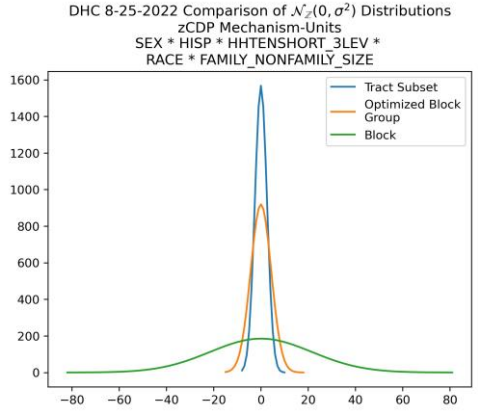
- Because the σ^2 are the same for each query and differ only by geography, only one graph is needed to display the distributions by query.

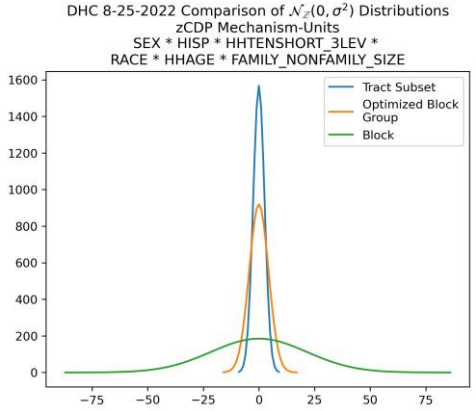


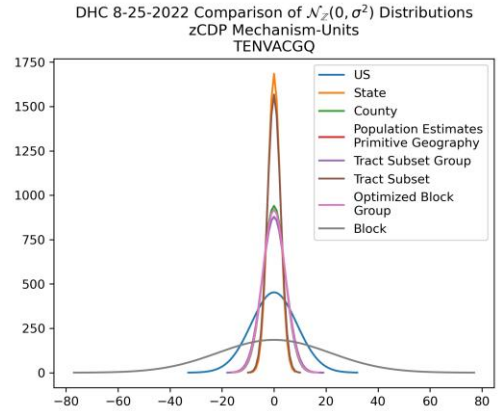
The DP Mechanism - Parameters

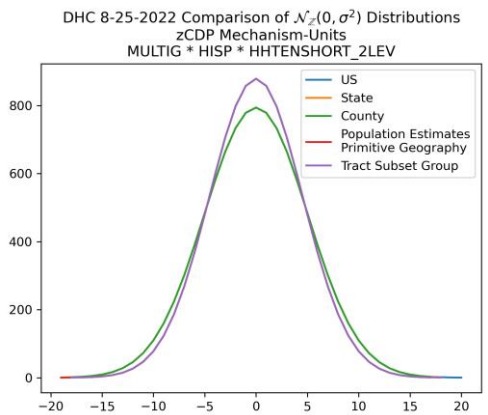
- The σ^2 distributions for units tables by geographical level (graphs);

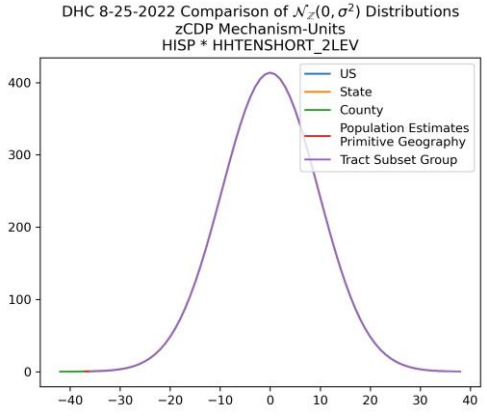


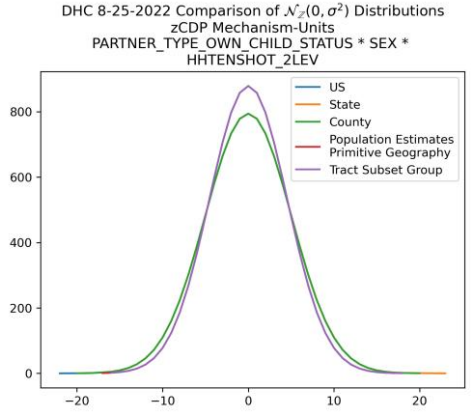












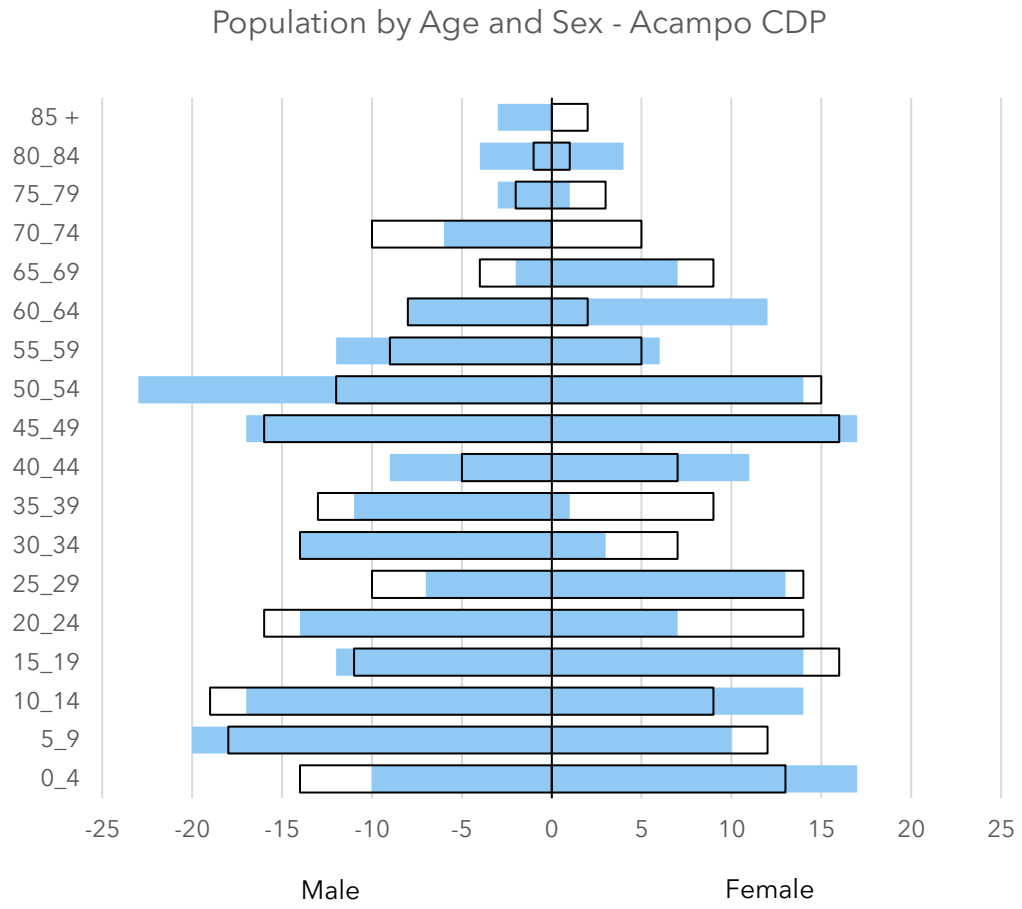
A Tale of 3 Population Pyramids – Small Population

This pyramid compares the population distribution derived from the 2010 SF1 published data with data derived from the 2010 DHC-DAS for Acampo CDP.

2010 Census with DP

Published 2010 Census Data

2010 SF1 Population: 341



	Absolute Error
Age 85 +	1
Age 80 to 84	6
Age 75 to 79	1
Age 70 to 74	9
Age 65 to 69	4
Age 60 to 64	10
Age 55 to 59	4
Age 50 to 54	10
Age 45 to 49	2
Age 40 to 44	8
Age 35 to 39	10
Age 30 to 34	4
Age 25 to 29	4
Age 20 to 24	9
Age 15 to 19	1
Age 10 to 14	3
Age 5 to 9	0
Age 0 to 4	0
Mean Absolute Error	4.8

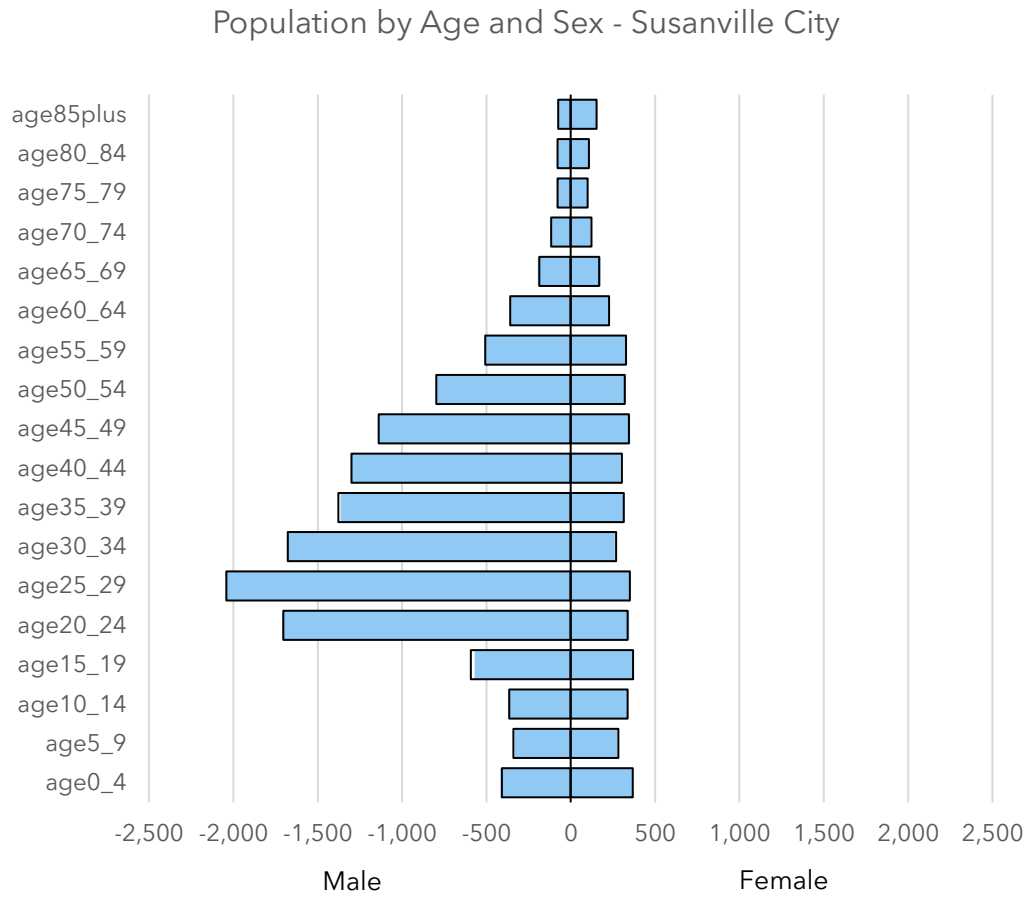
A Tale of 3 Population Pyramids – Mid Population

This pyramid compares the population distribution derived from the 2010 SF1 published data with data derived from the 2010 DHC-DAS for Susanville city.

2010 Census with DP

Published 2010 Census Data

2010 SF1 Population: 17,947



	Absolute Error
Age 85 +	3
Age 80 to 84	2
Age 75 to 79	4
Age 70 to 74	7
Age 65 to 69	18
Age 60 to 64	3
Age 55 to 59	0
Age 50 to 54	7
Age 45 to 49	7
Age 40 to 44	6
Age 35 to 39	18
Age 30 to 34	4
Age 25 to 29	2
Age 20 to 24	13
Age 15 to 19	26
Age 10 to 14	3
Age 5 to 9	13
Age 0 to 4	6
Mean Absolute Error	7.9

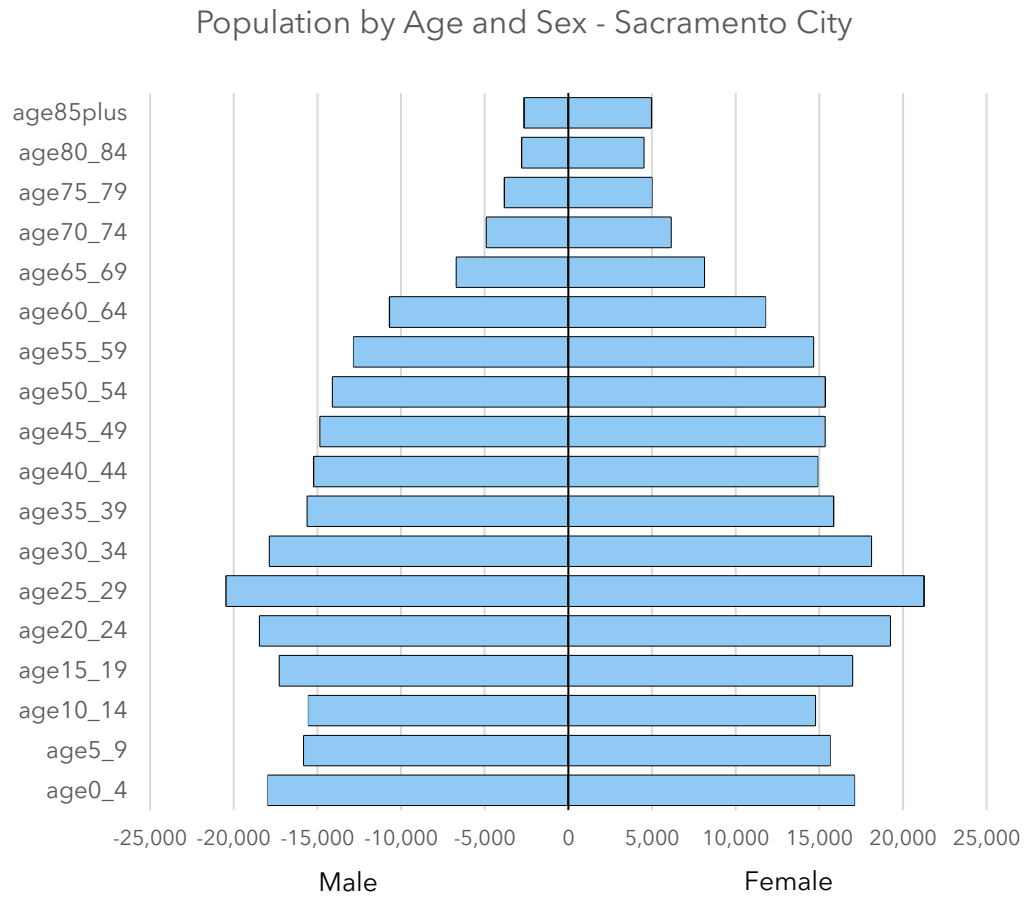
A Tale of 3 Population Pyramids – Large Population

This pyramid compares the population distribution derived from the 2010 SF1 published data with data derived from the 2010 DHC-DAS for Sacramento city.

2010 Census with DP

Published 2010 Census Data

2010 SF1 Population: 466,488



	Absolute Error
Age 85 +	57
Age 80 to 84	18
Age 75 to 79	6
Age 70 to 74	11
Age 65 to 69	1
Age 60 to 64	9
Age 55 to 59	11
Age 50 to 54	1
Age 45 to 49	18
Age 40 to 44	15
Age 35 to 39	7
Age 30 to 34	1
Age 25 to 29	51
Age 20 to 24	24
Age 15 to 19	19
Age 10 to 14	32
Age 5 to 9	25
Age 0 to 4	57
Mean Absolute Error	20.2

Demonstration Products – Metrics Tables

- Starting in March 2020, Census began releasing updated metrics designed around use cases and stakeholder feedback;
- The purpose is to allow users/stakeholders to see improvements from changes to the DAS mechanism;
- The metrics include measures of accuracy, bias, and outliers;
- For users needing to explore tabulations not included in metrics tables, the Census Bureau has provided a complete dataset with all tables and geographies.

Demonstration Products – Metrics Tables - Accuracy

- Measures of accuracy.
  - Accuracy is measured by comparing the post-disclosure protected tabulations to the original, publicly available tabulations from the 2010 Census and the internal pre-disclosure avoidance microdata from the 2010 Census.

- Accuracy measures include –
  - Mean Absolute Error (MAE);
  - Mean Numeric Error (ME) ;
  - Root Mean Squared Error (RMSE);
  - Mean Absolute Percent Error (MAPE); and
  - Coefficient of Variation (CV)

Demonstration Products – Metrics Tables - Bias

- Measures of bias.
  - Related to accuracy, but bias measures the direction of change and whether it varies with population size or some other characteristic.

- Bias measures include –
  - Mean Numeric Error (ME); and
  - Mean Percent Error (MALPE).

Demonstration Products – Metrics Tables – Examples – Accuracy

- Sample metrics table with measures of accuracy (3/16/2022 compared with the 8/25/2022 release):

Table 1.a: Total Population for county size categories - MAE, RMSE, MAPE, CV, MALPE, and outliers - 3/16/2022							Table 1.a: Total Population for county size categories - MAE, RMSE, MAPE, CV, MALPE, and outliers - 8/25/2022					
Universe: Total population							Universe: Total population					
Geography: Summary Level 050 - State-County							Geography: Summary Level 050 - State-County					
	Count of Units (N)	MAE	RMSE	MAPE (%)	CV			Count of Units (N)	MAE	RMSE	MAPE (%)	CV
All counties	3,143	1.75	2.27	0.02	-		All counties	3,143	1.75	2.27	0.02	-
Counties with total population less than 1,000	35	1.31	1.53	0.31	0.22		Counties with total population less than 1,000	35	1.31	1.53	0.31	0.22
Counties with total population 1,000 to 4,999	268	1.46	1.97	0.05	0.06		Counties with total population 1,000 to 4,999	268	1.46	1.97	0.05	0.06
Counties with total population 5,000 to 9,999	395	1.72	2.19	0.02	0.03		Counties with total population 5,000 to 9,999	395	1.72	2.19	0.02	0.03
Counties with total population 10,000 to 49,999	1,469	1.78	2.28	0.01	0.01		Counties with total population 10,000 to 49,999	1,469	1.78	2.28	0.01	0.01
Counties with total population 50,000 to 99,999	398	1.72	2.21	-	-		Counties with total population 50,000 to 99,999	398	1.72	2.21	-	-
Counties with total population of 100,000 or more	578	1.89	2.49	-	-		Counties with total population of 100,000 or more	578	1.89	2.49	-	-

Demonstration Products – Metrics Tables – Example – Accuracy, Bias, Outliers

- Sample metrics table with measures of accuracy, bias, and outliers (3/16/2022 compared with the 8/25/2022 release):

DHC Use Case Table 7.b: Household size for county size categories - MAE, RMSE, MAPE, CV, MALPE, and outliers - 3/16/2022			
Universe: Occupied Housing Units			
Geography: Summary Level 050 - State-County			
	Count of Units (N)	MALPE (%)	Count of geographies where the absolute percent difference exceeds 5%
All counties			
1-person household	3,143	(0.40)	44
2-person household	3,143	(0.50)	68
3-person household	3,143	(0.25)	264
4-person household	3,143	0.61	366
5-person household	3,143	1.64	694
6-person household	3,143	7.87	1,428
7-or-more-person household	3,143	13.53	1,920

DHC Use Case Table 7.b: Household size for county size categories - MAE, RMSE, MAPE, CV, MALPE, and outliers - 8/25/2022			
Universe: Occupied Housing Units			
Geography: Summary Level 050 - State-County			
	Count of Units (N)	MALPE (%)	Count of geographies where the absolute percent difference exceeds 5%
All counties			
1-person household	3,143	(0.49)	73
2-person household	3,143	(0.61)	100
3-person household	3,143	(0.41)	378
4-person household	3,143	0.42	499
5-person household	3,143	2.58	918
6-person household	3,143	10.24	1,707
7-or-more-person household	3,143	18.98	2,115

Demonstration Products – Metrics Tables – Examples – Accuracy

- Sample metrics table with measures of accuracy (3/16/2022 compared with the 8/25/2022 release):

Use Case Table 5.a: Single Years of Age for Population 0 to 17 Years Old for county size categories - MAE, RMSE, MAPE, CV, and MALPE - 3/16/2022						
Universe: Population 0 to 17 years old						
Geography: Summary Level 050 - State-County						
	Count of Units (N)	MAE	MAPE (%)	CV	MALPE (%)	Count of counties where the numeric difference exceeds 5%
All counties						
Under 1 years old	3,143	16.50	7.56	1.86	0.32	1,241
1 years old	3,143	16.88	7.07	1.93	0.18	1,199
2 years old	3,143	16.77	6.99	1.83	0.56	1,194
3 years old	3,143	17.07	6.90	1.88	0.19	1,144
4 years old	3,143	16.98	7.17	1.86	0.26	1,153
5 years old	3,143	17.02	7.19	1.86	0.33	1,196
6 years old	3,143	17.18	7.16	1.88	0.68	1,200
7 years old	3,143	16.69	6.96	1.85	0.45	1,190
8 years old	3,143	16.25	6.76	1.82	(0.03)	1,176
9 years old	3,143	17.35	6.78	1.87	(0.12)	1,177
10 years old	3,143	17.29	7.04	1.83	0.28	1,126
11 years old	3,143	17.34	6.72	1.84	0.48	1,165
12 years old	3,143	16.60	6.87	1.79	0.30	1,151
13 years old	3,143	16.76	7.00	1.80	0.46	1,152
14 years old	3,143	16.93	6.60	1.82	0.16	1,106
15 years old	3,143	16.25	6.14	1.73	0.33	1,063
16 years old	3,143	16.33	6.20	1.66	0.64	1,094
17 years old	3,143	16.70	5.87	1.71	0.52	1,048

Use Case Table 5.a: Single Years of Age for Population 0 to 17 Years Old for county size categories - MAE, RMSE, MAPE, CV, and MALPE - 8/25/2022						
Universe: Population 0 to 17 years old						
Geography: Summary Level 050 - State-County						
	Count of Units (N)	MAE	MAPE (%)	CV	MALPE (%)	Count of counties where the numeric difference exceeds 5%
All counties						
Under 1 years old	3,143	14.62	6.39	1.71	0.16	1,093
1 years old	3,143	14.93	6.21	1.69	0.38	1,057
2 years old	3,143	14.56	6.08	1.60	0.14	1,044
3 years old	3,143	13.15	5.80	1.41	0.48	934
4 years old	3,143	13.20	5.82	1.42	0.58	951
5 years old	3,143	16.50	6.56	1.80	0.37	1,143
6 years old	3,143	16.47	6.96	1.83	0.49	1,136
7 years old	3,143	16.42	7.01	1.78	0.24	1,184
8 years old	3,143	17.07	6.80	1.85	(0.13)	1,190
9 years old	3,143	17.01	6.78	1.82	0.04	1,141
10 years old	3,143	16.69	6.54	1.80	0.18	1,111
11 years old	3,143	16.55	6.58	1.79	(0.07)	1,161
12 years old	3,143	16.88	6.62	1.84	0.16	1,140
13 years old	3,143	16.82	6.58	1.82	0.37	1,139
14 years old	3,143	16.87	6.69	1.80	0.29	1,181
15 years old	3,143	5.11	2.77	0.50	0.23	399
16 years old	3,143	5.97	2.81	0.67	0.61	402
17 years old	3,143	6.33	2.76	0.71	0.65	406

Demonstration Products – Call for Public Feedback

- The Census Bureau has requested public feedback on the most recent demonstration products;
- The Census Bureau will accept feedback through Monday, September 26;
- Census is requesting the feedback to take the following forms:
  - Identifying the table or tables in question;
  - Analyzing the differences between the demonstration data and the 2010 Census data;
  - Noting whether the differences (along with the metrics) are acceptable or not acceptable;
  - If the differences or metrics are not acceptable, providing what difference or metric value would be acceptable; and
  - How these differences impact the use of the data.
- Submit feedback and questions to 2020DAS@census.gov

Demonstration Products – Call for Public Feedback

- Census Bureau provided two examples to provide useful feedback on the demonstration product:
  - Comparing the measures of accuracy or bias across the two DHC demonstration products for multiple characteristics; and
  - Comparing the measures of accuracy or bias with the average or count of the characteristic and geography.
- Comparing measures across the two DHC demonstration data products has the advantage of having the measures provided in the metrics tables (assuming the table and geography is represented);
- Comparing a relevant measure with the geographic counts also has the advantage of being able to use the metrics table;
- Comparing measures with the population average for a geography might require custom calculations to derive the average.

Demonstration Products – Call for Public Feedback – Example

- Comparing the measures of accuracy or bias across the two DHC demonstration products for multiple characteristics

Use Case Table 5.a: Single Years of Age for Population 0 to 17 Years Old for county size categories

Universe: Population 0 to 17 years old

Geography: Summary Level 050 - State-County

	Count of Units (N)	MAE		MAPE (%)		Count of counties where the numeric difference exceeds 5%	
All counties		3/16/2022	8/25/2022	3/16/2022	8/25/2022	3/16/2022	8/25/2022
Under 1 years old	3,143	16.50	14.62	7.56	6.39	1,241	1,093
1 years old	3,143	16.88	14.93	7.07	6.21	1,199	1,057
2 years old	3,143	16.77	14.56	6.99	6.08	1,194	1,044
3 years old	3,143	17.07	13.15	6.90	5.80	1,144	934
4 years old	3,143	16.98	13.20	7.17	5.82	1,153	951
5 years old	3,143	17.02	16.50	7.19	6.56	1,196	1,143
6 years old	3,143	17.18	16.47	7.16	6.96	1,200	1,136
7 years old	3,143	16.69	16.42	6.96	7.01	1,190	1,184
8 years old	3,143	16.25	17.07	6.76	6.80	1,176	1,190
9 years old	3,143	17.35	17.01	6.78	6.78	1,177	1,141
10 years old	3,143	17.29	16.69	7.04	6.54	1,126	1,111
11 years old	3,143	17.34	16.55	6.72	6.58	1,165	1,161
12 years old	3,143	16.60	16.88	6.87	6.62	1,151	1,140
13 years old	3,143	16.76	16.82	7.00	6.58	1,152	1,139
14 years old	3,143	16.93	16.87	6.60	6.69	1,106	1,181
15 years old	3,143	16.25	5.11	6.14	2.77	1,063	399
16 years old	3,143	16.33	5.97	6.20	2.81	1,094	402
17 years old	3,143	16.70	6.33	5.87	2.76	1,048	406

Demonstration Products – Call for Public Feedback – Example

- Comparing the measures of accuracy or bias with the average or count of the characteristic and geography.

Table P14: Age and Sex for the Population Under 20 Years
Universe: Persons
Geography: California Counties

	SF1 Population	Average SF1 Population	MAE	MAPE (%)
Under 1 years old	494,058	8,518	31.34	2.88
1 years old	497,754	8,582	31.48	4.91
2 years old	516,002	8,897	32.16	2.81
3 years old	516,611	8,907	22.91	2.87
4 years old	506,908	8,740	24.31	2.62
5 years old	505,175	8,710	29.24	2.64
6 years old	500,418	8,628	31.69	2.87
7 years old	497,030	8,569	30.34	3.54
8 years old	493,551	8,510	27.14	3.24
9 years old	509,665	8,787	35.67	2.02
10 years old	511,352	8,816	34.50	4.52
11 years old	508,004	8,759	26.40	2.65
12 years old	513,257	8,849	30.24	2.07
13 years old	523,312	9,023	35.74	3.36
14 years old	535,005	9,224	31.00	2.93
15 years old	546,806	9,428	8.57	1.37
16 years old	556,657	9,598	15.95	1.27
17 years old	563,475	9,715	19.88	1.22

Questions/Discussion

Resources – Census Bureau

- Basics of Differential Privacy –
  - Differential Privacy: An Introduction For Statistical Agencies - https://gss.civilservice.gov.uk/wp-content/uploads/2018/12/12-12-18_FINAL_Privitar_Kobbi_Nissim_article.pdf
  - Differential Privacy: A Primer for a Non-technical Audience - https://salil.seas.harvard.edu/files/salil/files/differential_privacy_primer_nontechnical_audience.pdf
- Census Bureau –
  - Disclosure Avoidance and the 2020 Census - https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html
  - 2010 Demonstration Products - https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/newsletters/New-Demonstration-Data-DHC-Webinar-August-31.html
- Github Python repositories –
  - DAS 2020 Redistricting Production Code - https://github.com/uscensusbureau/DAS_2020_Redistricting_Production_Code
  - DAS 2010 Demonstration Data Products Disclosure Avoidance System Release - https://github.com/uscensusbureau/census2020-das-2010ddp
  - DAS E2E Release - https://github.com/uscensusbureau/census2020-das-e2e
  - Disclosure Avoidance Repository - https://github.com/uscensusbureau/census-dp

Resources – Outside Analysis and Data Products

- IPUMS –
  - Changes to Census Bureau Data Products - https://ipums.org/changes-to-census-bureau-data-products
  - Demonstration Data For U.S. Census Bureau Disclosure Avoidance System -https://www.nhgis.org/privacy-protected-2010-census-demonstration-data
- National Academy of Sciences 2020 Census Data Products: Workshop on the Demographic and Housing Characteristics Files - https://www.nationalacademies.org/event/06-21-2022/2020-census-data-products-workshop-on-the-demographic-and-housing-characteristics-files
- National Academy of Sciences Committee on National Statistics (CNSTAT) December 11-12 workshop on the 2010 Demonstration Data Products - https://sites.nationalacademies.org/DBASSE/CNSTAT/DBASSE_196518?#
- Department of Finance – 8-25-2022 Demonstration data sets for California –
  - https://web-services.dof.ca.gov/dru/dhc2_8-25-2022/CA_sf1_dhc2_8-25-2022_demo.zip
  - https://web-services.dof.ca.gov/dru/dhc2_8-25-2022/CA_das_dhc2_8-25-2022_demo.zip

Contact Information

- Jonathan Buttle – jonathan.buttle@dof.ca.gov

- California Department of Finance
- Demographic Research Unit
- (916) 323-4086