Differential Privacy User Implications

March, 2020

State Demography Office, Department of Local Affairs
Demography.dola.colorado.gov

COLORADO
Department of Local Affairs

Differential Privacy – Why and Implications for Census 2020 Data and Future Data

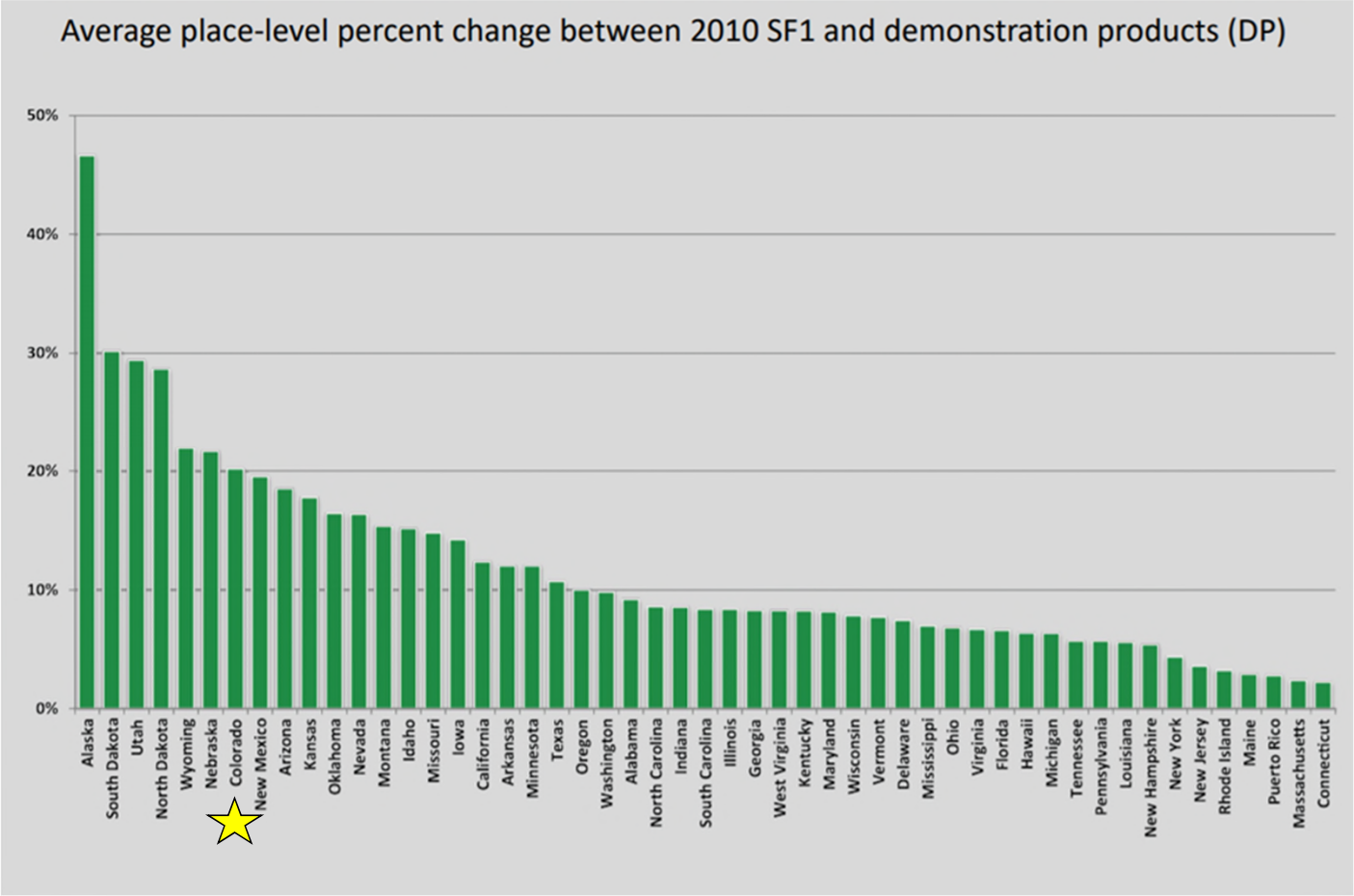
- Advances in computing power and the availability of external data sources make database reconstruction and re-identification increasingly likely. *Problem for assuring data privacy*.

- Differential privacy uses advances statistical methods to insert noise into the data to maintain the confidentiality of respondents. *Problem for assuring data accuracy and reliability.*

Differential Privacy Test Results

- Size of geography impact – differential in size within a state
- Urban to rural bias
- Unreasonable results –
  - household size <1,
  - families with children with no adults,
  - population with no housing units,
  - occupied housing units with no people, 100% occupancy.
- Significant issue with the separation of the relationship between housing units and population.
- CB is re-working methods as a result.

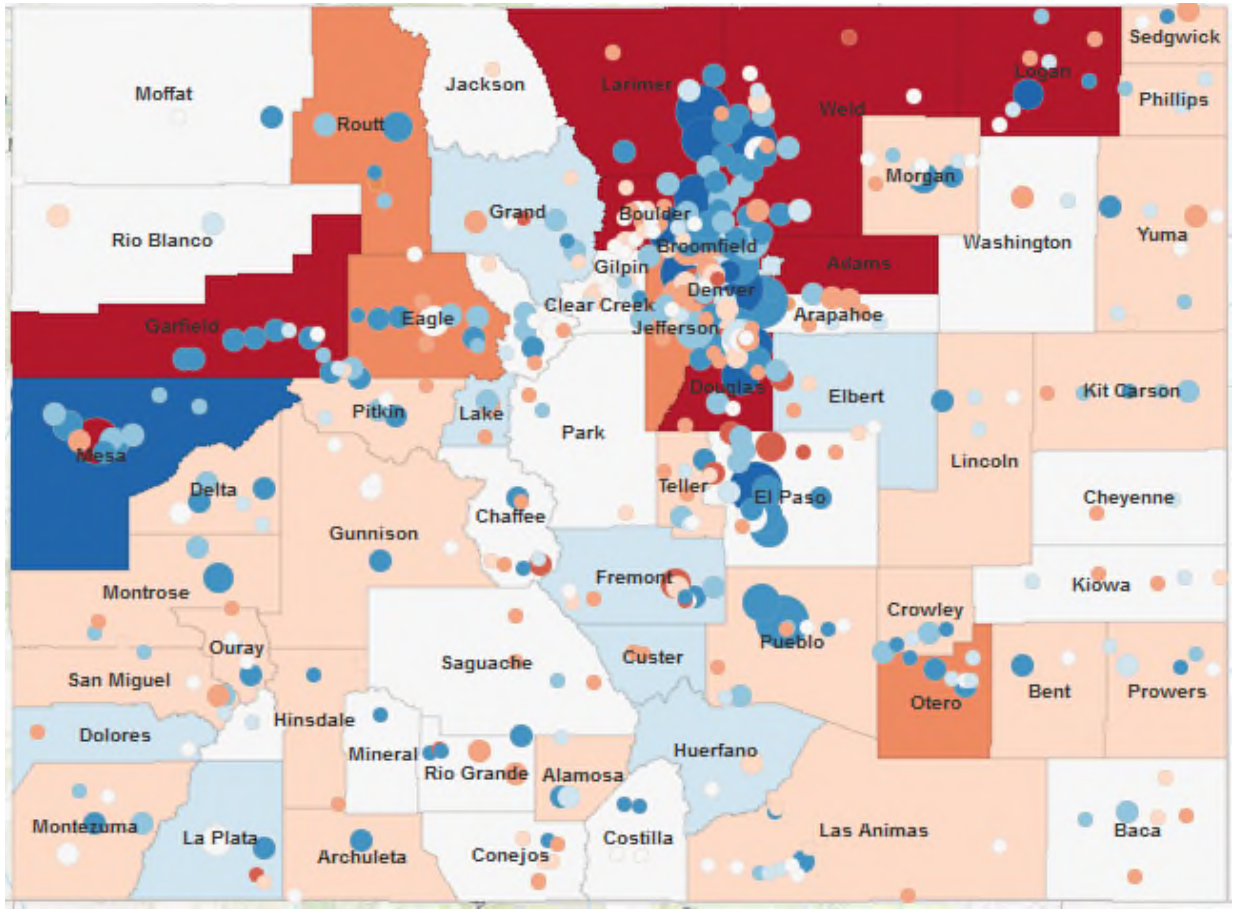
Summary of findings –population- 2010 SF1 file vs DP

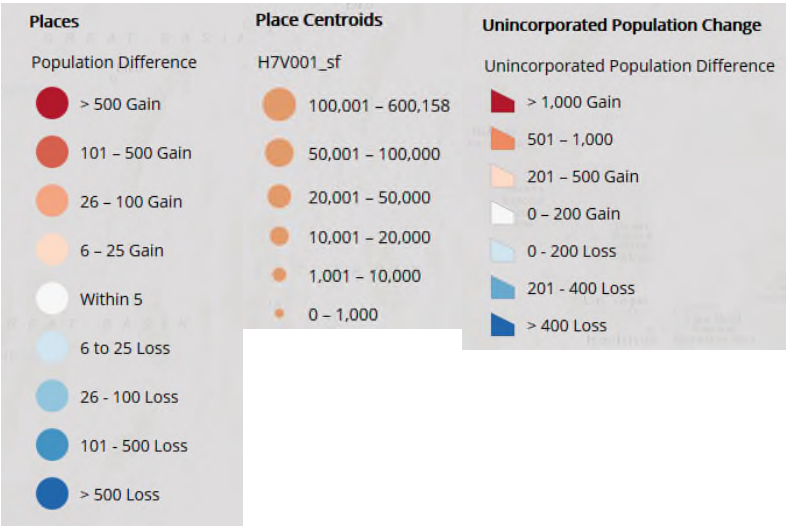


- CO 6th largest percentage change at place level (orig vs. DP test)
- Issue with rural and difference in size within a state

Source: State of Alaska

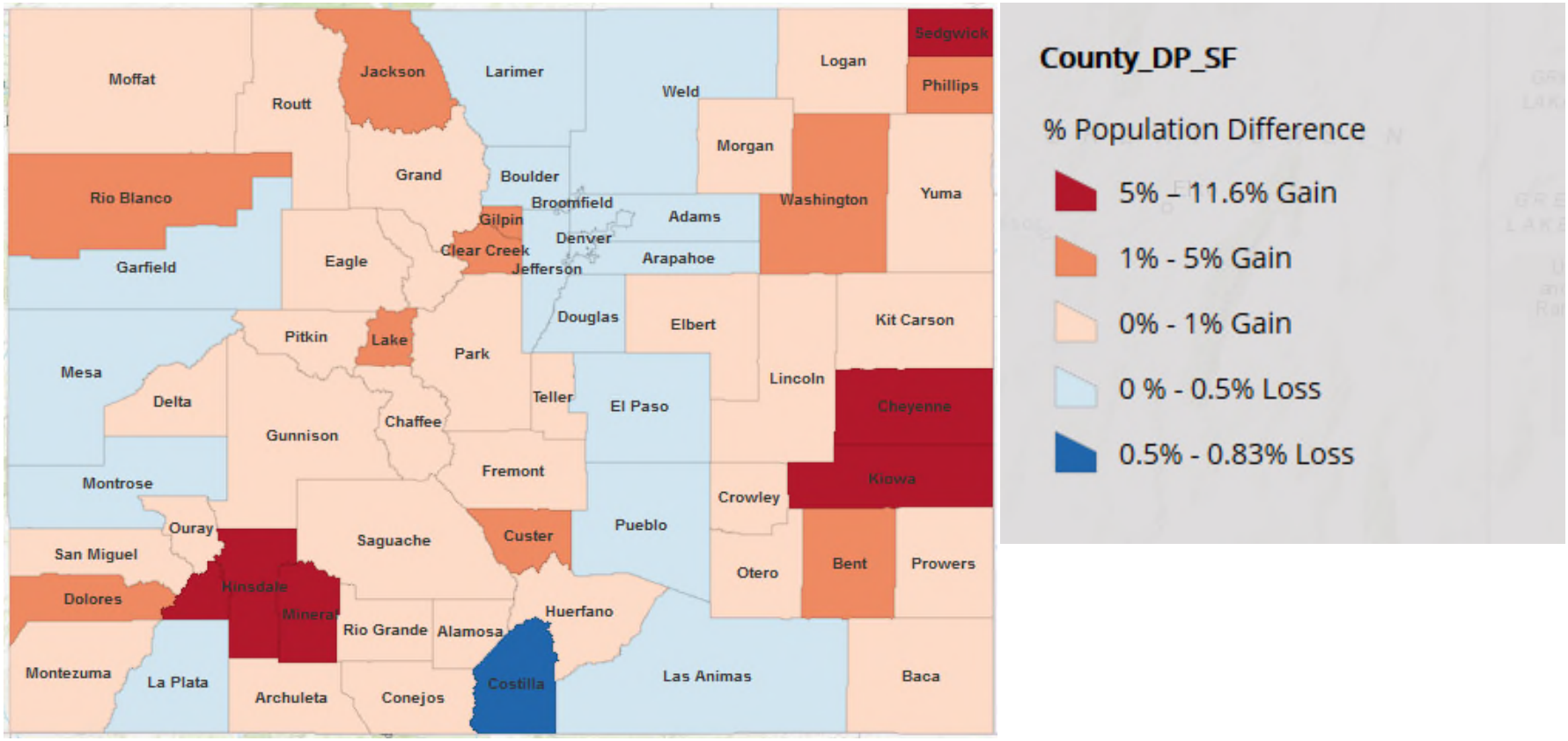
Comparison by size of incorporated area and unincorporated area.



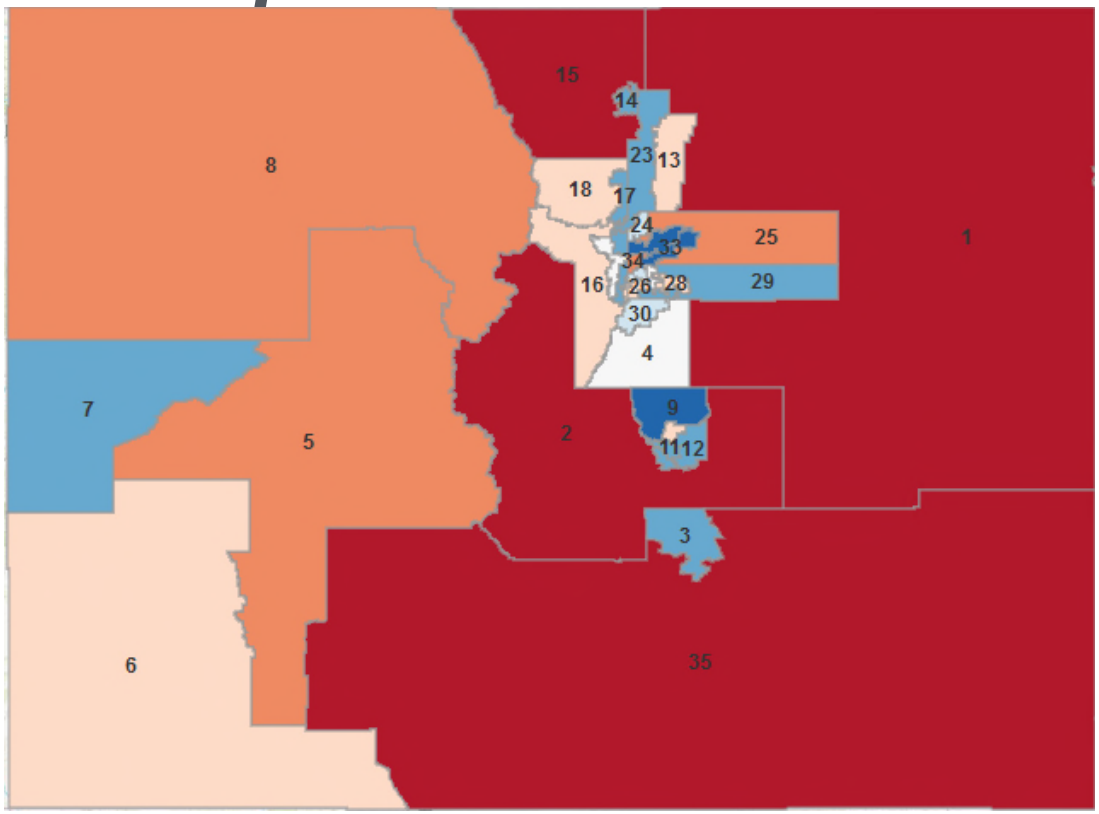


Impact between urban and rural, as well as county incorporated and unincorporated.

Percent Difference in Total Population



Impact to State Senate Districts



Changes in population after differential privacy largely reflect what has been seen in previous maps with rural districts generally gaining population while urban districts are more likely to lose population. The changes do not move any senate districts outside of a 2.8% population variance.

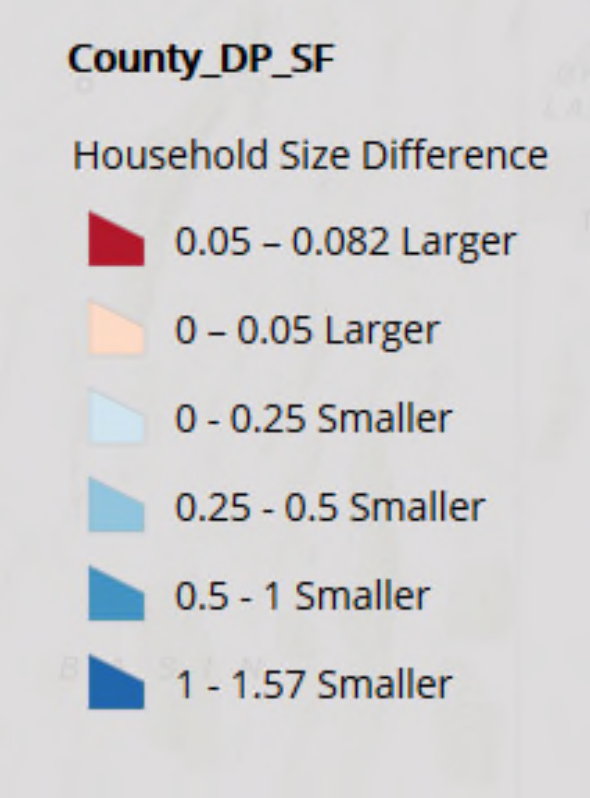
State_Upper

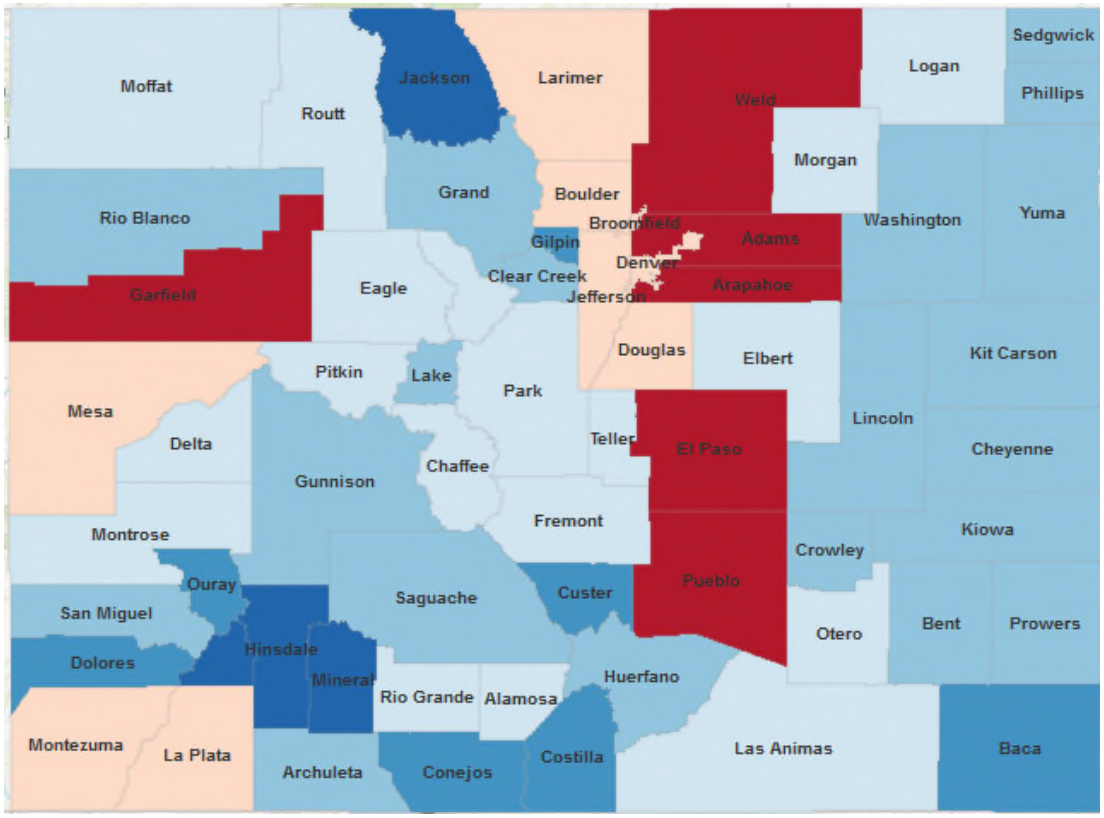
Population Difference

- 500 – 932 Gain
- 201 – 500 Gain
- 51 – 200 Gain
- Within 50
- 51 - 200 Loss
- 201 - 500 Loss
- 501 - 563 Loss

- For state senate districts, the maximum deviation from the SF district population increases from just over 2.5% to just under 2.8%
- For House districts maximum deviation increased from just under 2.5% to 3.5%
- Legal implications

Disconnects the Relationship Between Population and Occupied Housing Units





Household type - households with children

An aging population typically results in fewer households with children and more households without children.

Change in the number of households with children, 2000 to 2010

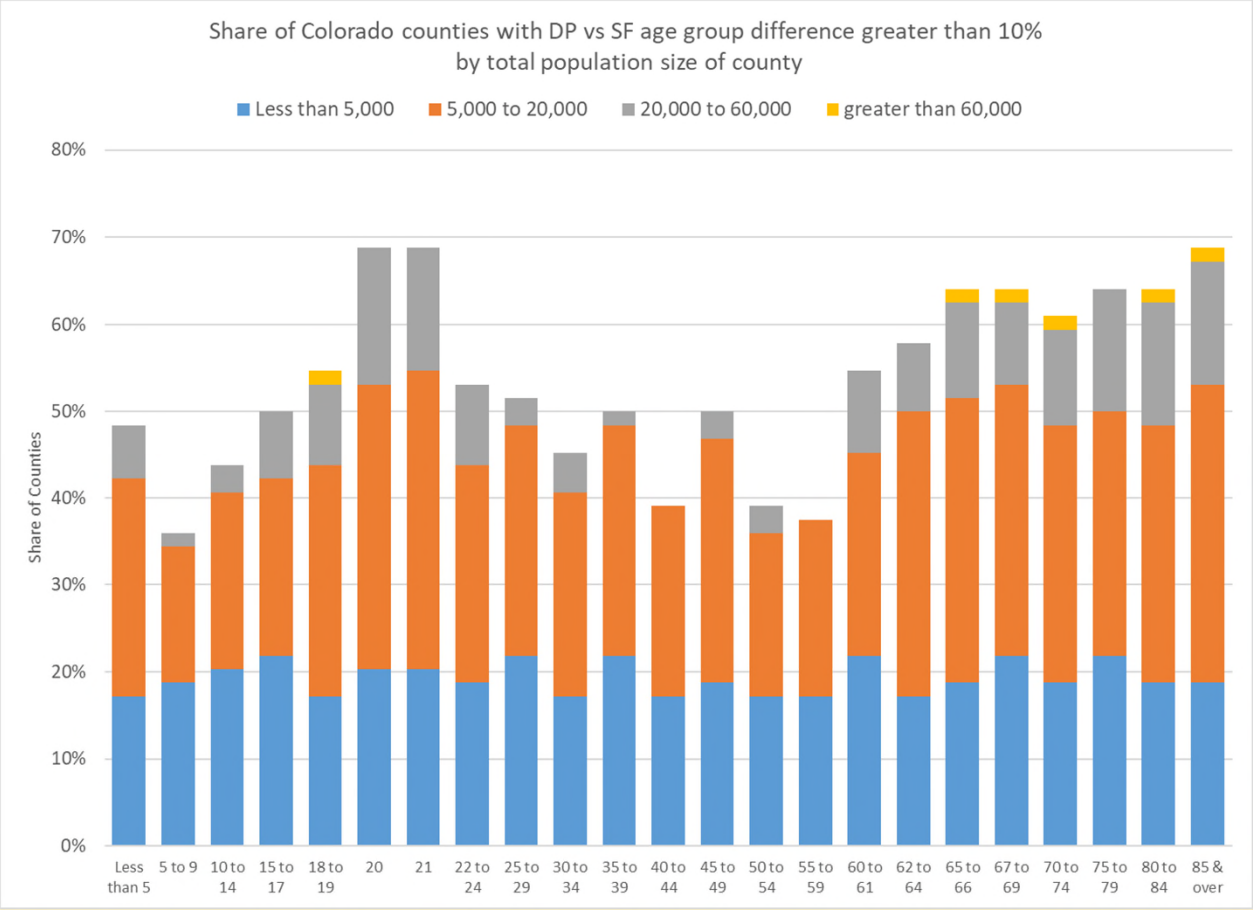
Using the SF data as released in 2010

- 43 counties decline
- 20 counties increase, the largest increase 53% in a high growth county

Using the DP data

- 11 counties decline
- 50 counties increase, 11 greater than 50%, 7 greater than 100%

Summary of findings – age - 2010 SF1 file vs DP



Over 60% of counties have a greater than 10% difference in count of 65 and older.

Potential Impacts on use of Differential Privacy

These initial analyses reveal the implementation of differential privacy is most concerning in its potential impact on data accuracy at small geographies and population size.

- Redistricting
- Financial distribution formulas
- Vital statistics rates
- Loss of population to household relationship
- Rural/Urban reallocation
- Time series work

Concerns

- Privacy vs. Accuracy
- Invariants (maintained as counted)
  - Geography
  - Variables
- Population to household/housing unit relationship
- Products
- Credibility
- Use of taxpayer dollars

Feedback on Demonstration File

- From CB: Data user feedback will help the Census Bureau's senior leadership make informed decisions about the final structure of 2020 data products and the fine-tuning of the disclosure avoidance system. Because decennial census data are widely used in ways that go beyond the Census Bureau's needs, we want to ensure that they are fit for as many data users' needs as possible.

- Feedback on the demonstration files should be sent to: dcmd.2010.demonstration.data.products@census.gov
- Formal letter to the CB from Colorado Legislature

Additional Resources

YouTube video explaining differential privacy https://www.youtube.com/watch?v=pT19VwBAqKA

For Demonstration Data Product:

- https://www.nhgis.org/differentially-private-2010-census-data (recommended)
- https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2010-demonstration-data-products.html

Other Resources

- Differential Privacy in Colorado Interactive MAP http://arcg.is/1X4afz
- https://ipums.org/changes-to-census-bureau-data-products
- CNStat workshop https://sites.nationalacademies.org/DBASSE/CNSTAT/DBASSE_196518
- State demographer network - https://www.census.gov/programs-surveys/popest/about/fscpe/contacts.html
- State Data Centers - https://www.census.gov/about/partners/sdc/member-network.html

Thank you

State Demography Office
Department of Local Affairs
Elizabeth Garner, Elizabeth.garner@state.co.us, 303-864-7750
Cindy DeGroen, Cindy.degroen@state.co.us, 303-864-7752
Demography.dola.colorado.gov

COLORADO
Department of Local Affairs

Policy decisions – Invariants (as enumerated)

2010 SF data as released in 2010

- Block – total population
- Block – voting age population
- Block – total housing units
- Block – occupancy status
- Block – group quarters count
- Block – group quarters type

2010 DP demonstration data

- State – total population
- Block – total housing units
- Block – group quarters count
- Block – group quarters type

Households- 2010 SF1 file vs DP

For counties <5,000 total population (22% of CO counties)

- DP households 52% higher than SF
- DP households 80% higher on average by county
- Household size declined from 2.2 in SF1 vs 1.5 in DP estimates

For counties 5,000 to 20,000 total population (38% of CO counties)

- DP households 16% higher than SF
- DP households 18% higher on average by county
- Household size declined from 2.3 in SF1 to 2.0 in DP estimates

Households- 2010 SF1 file vs DP

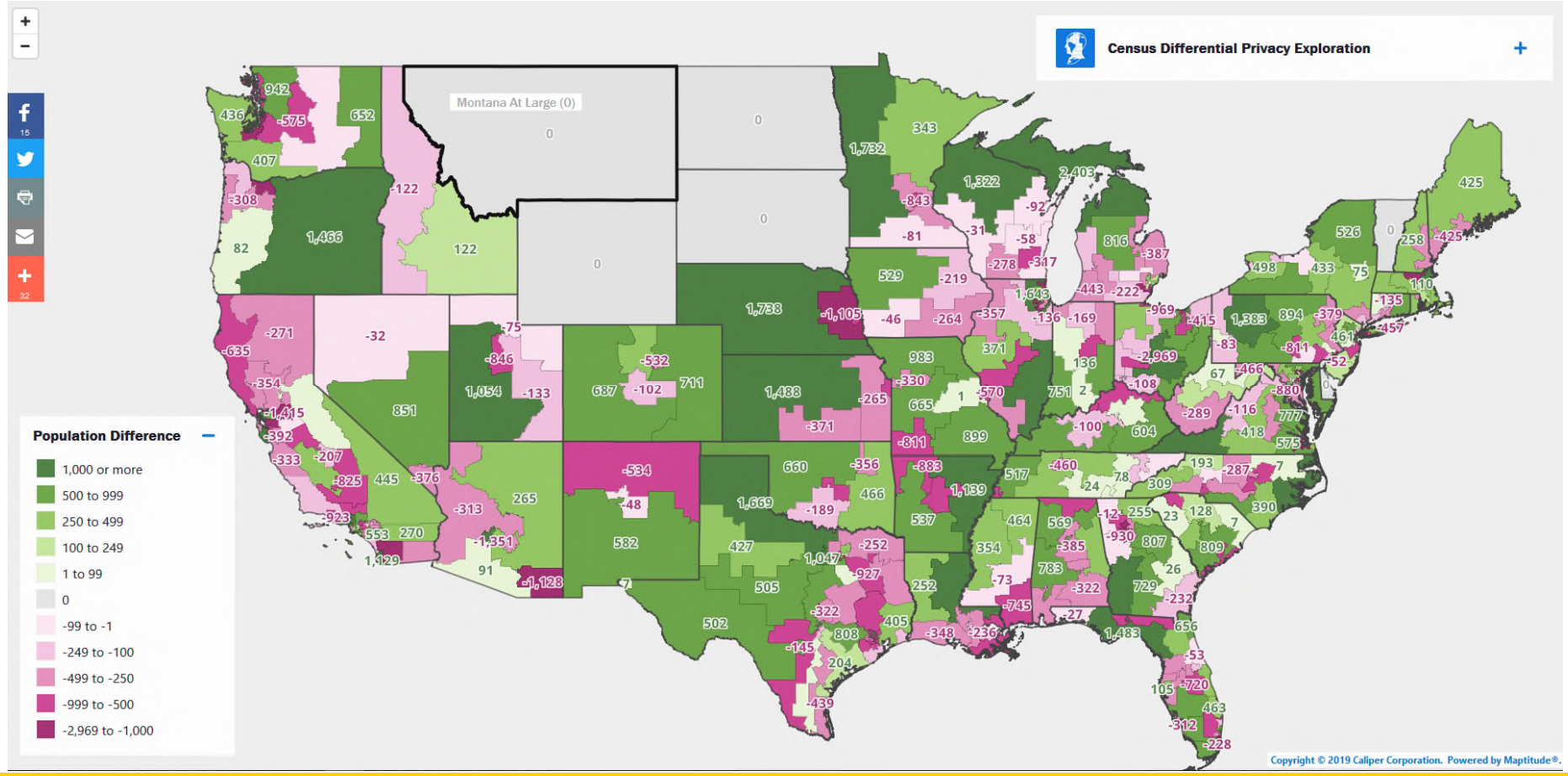
For counties 20,000 to 60,000 total population (22% of CO counties)

- DP households 3% higher than SF
- DP households 4% higher on average by county
- Household size declined from 2.5 in SF1 vs 2.4 in DP estimates

For counties 60,000+ total population (19% of CO counties)

- DP households 2% lower than SF
- Household size was unchanged from SF to DP

For blocks (201,062 blocks in Colorado in 2010)

- 18,751 blocks have more occupied housing units than population
- 11,090 blocks have occupied housing units but 0 population

Impact by current 116th Congressional District, SF vs. DP