Differential Privacy User Implications

January, 2020

State Demography Office, Department of Local Affairs

Demography.dola.colorado.gov

COLORADO
Department of Local Affairs

Overview

- Summary review of demonstration product

- Policy Decisions
  - Privacy budget (noise) – across the board
  - Invariants
  - Persons and occupancy (occupied housing units)

- Resources

- How to submit feedback

Differential Privacy 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.

- Geographies with large Group Quarter Population issues -

Summary of findings –population– 2010 SF1 file vs DP

Average place-level percent change between 2010 SF1 and demonstration products (DP)
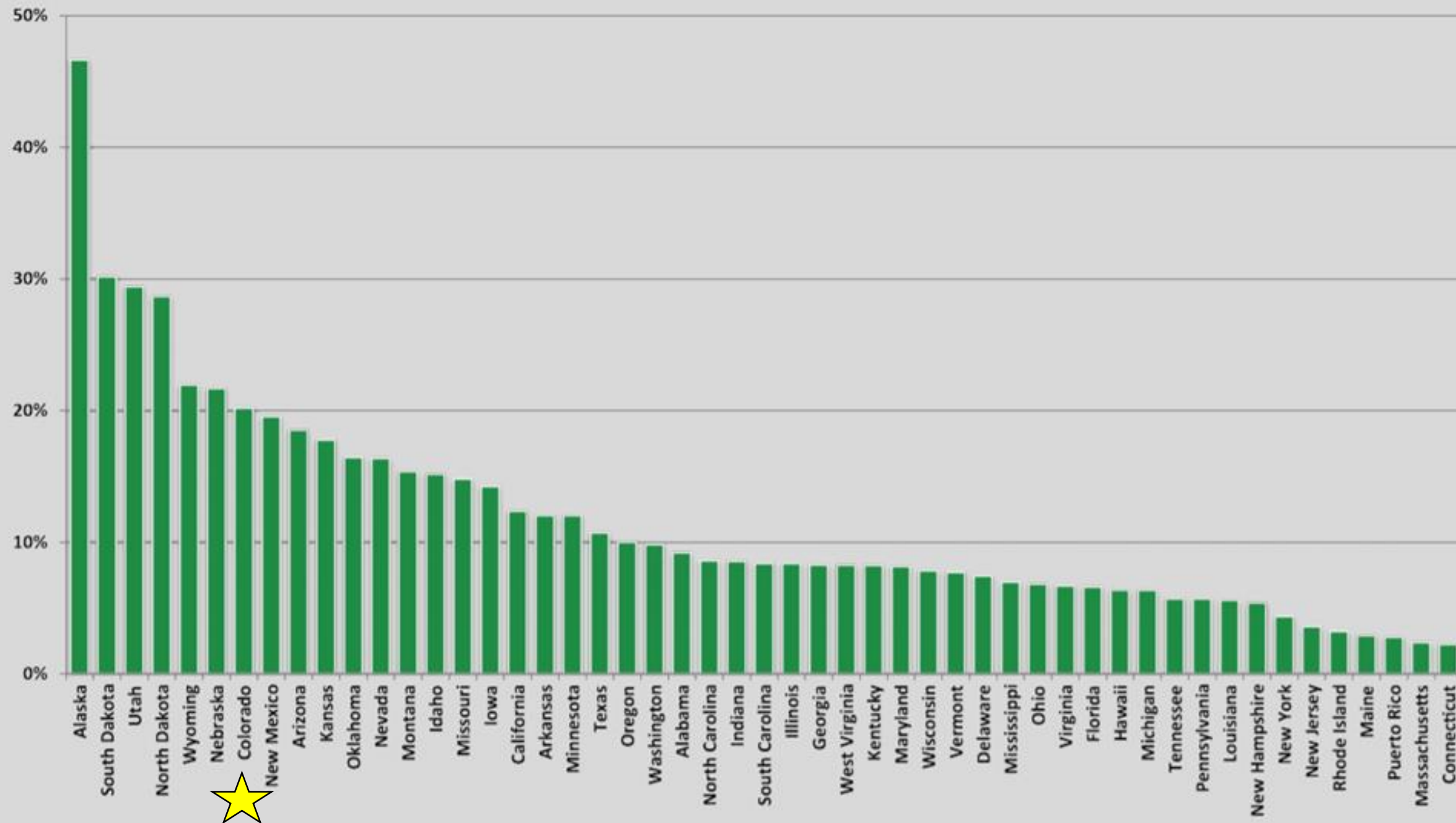
There are 30 states with populations lower than Colorado's, however, there are only 6 states (AK, SD, UT, ND, WY, NE) with higher percentage of average place-level change 2010 SF1 file vs DP due to the rural-ness of many Colorado places.
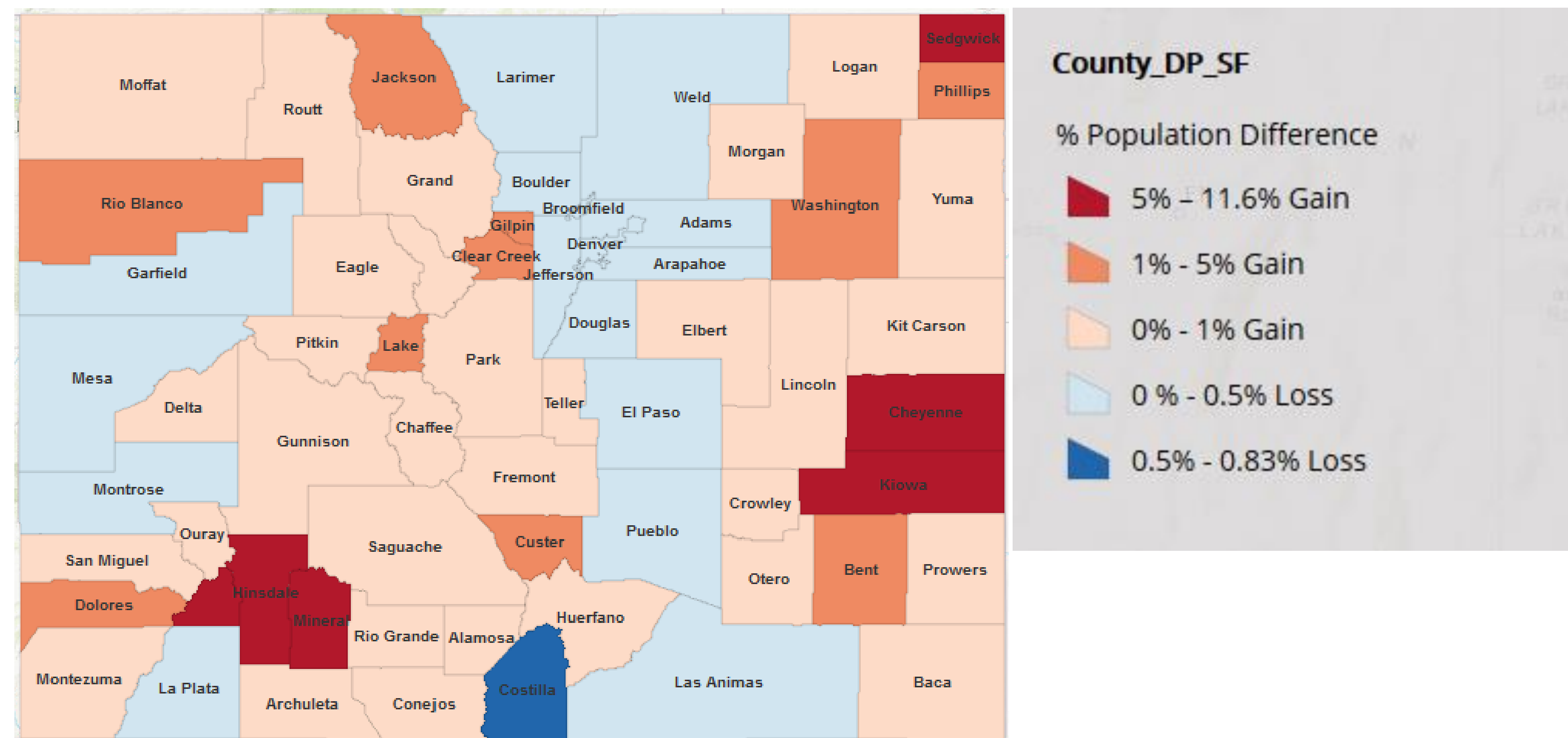
Comparison by size of incorporated area and unincorporated area.
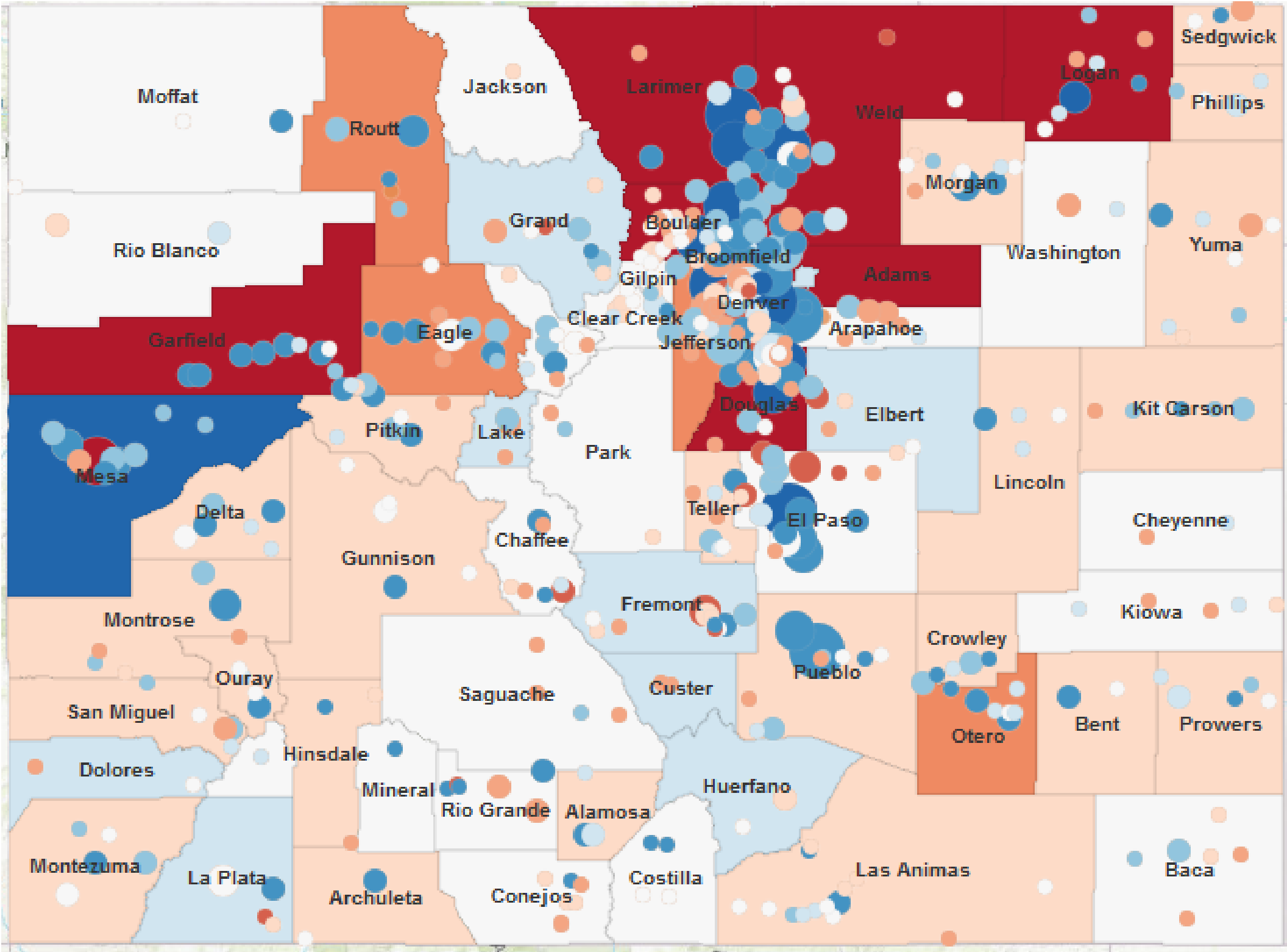
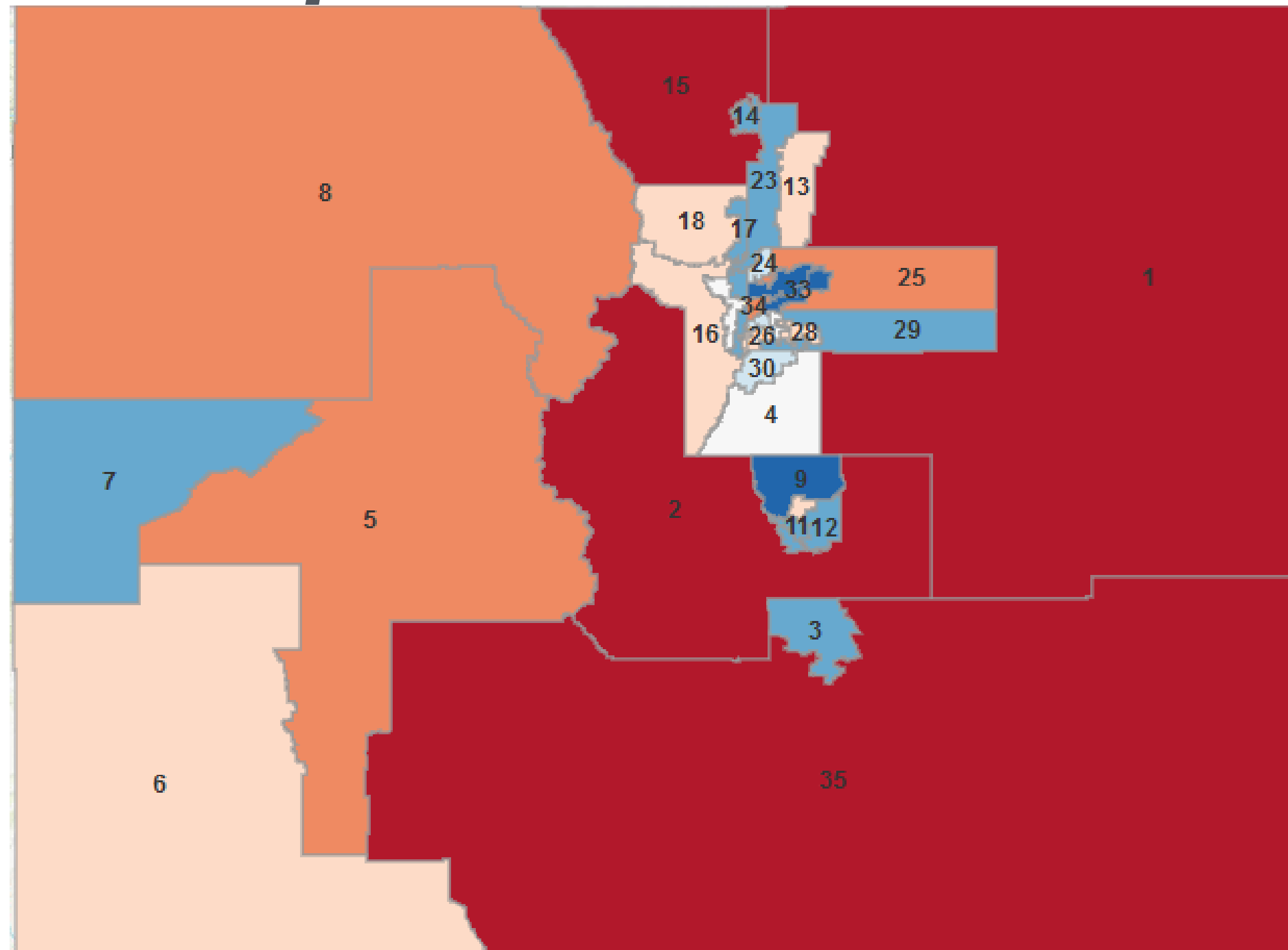

Besides an impact between urban and rural, also and impact within a county between incorporated and unincorporated.

COLORADO
Department of Local Affairs

Percent Difference in Total Population

County_DP_SF

% Population Difference

- 5% – 11.6% Gain
- 1% – 5% Gain
- 0% – 1% Gain
- 0 % – 0.5% Loss
- 0.5% – 0.83% Loss

COLORADO
Department of Local Affairs

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
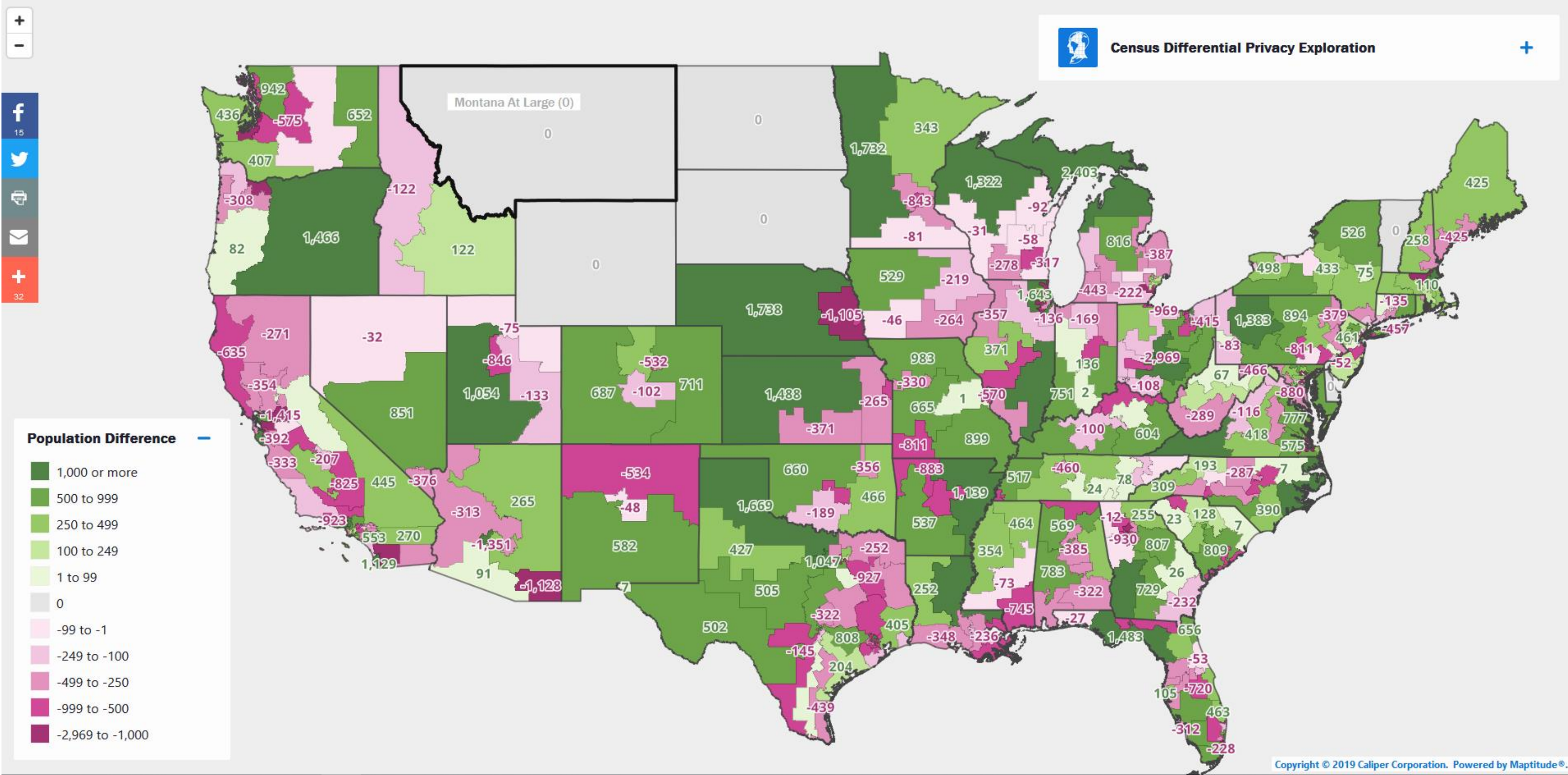- 51 - 200 Loss
- 201 - 500 Loss
- 501 - 563 Loss

- 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%
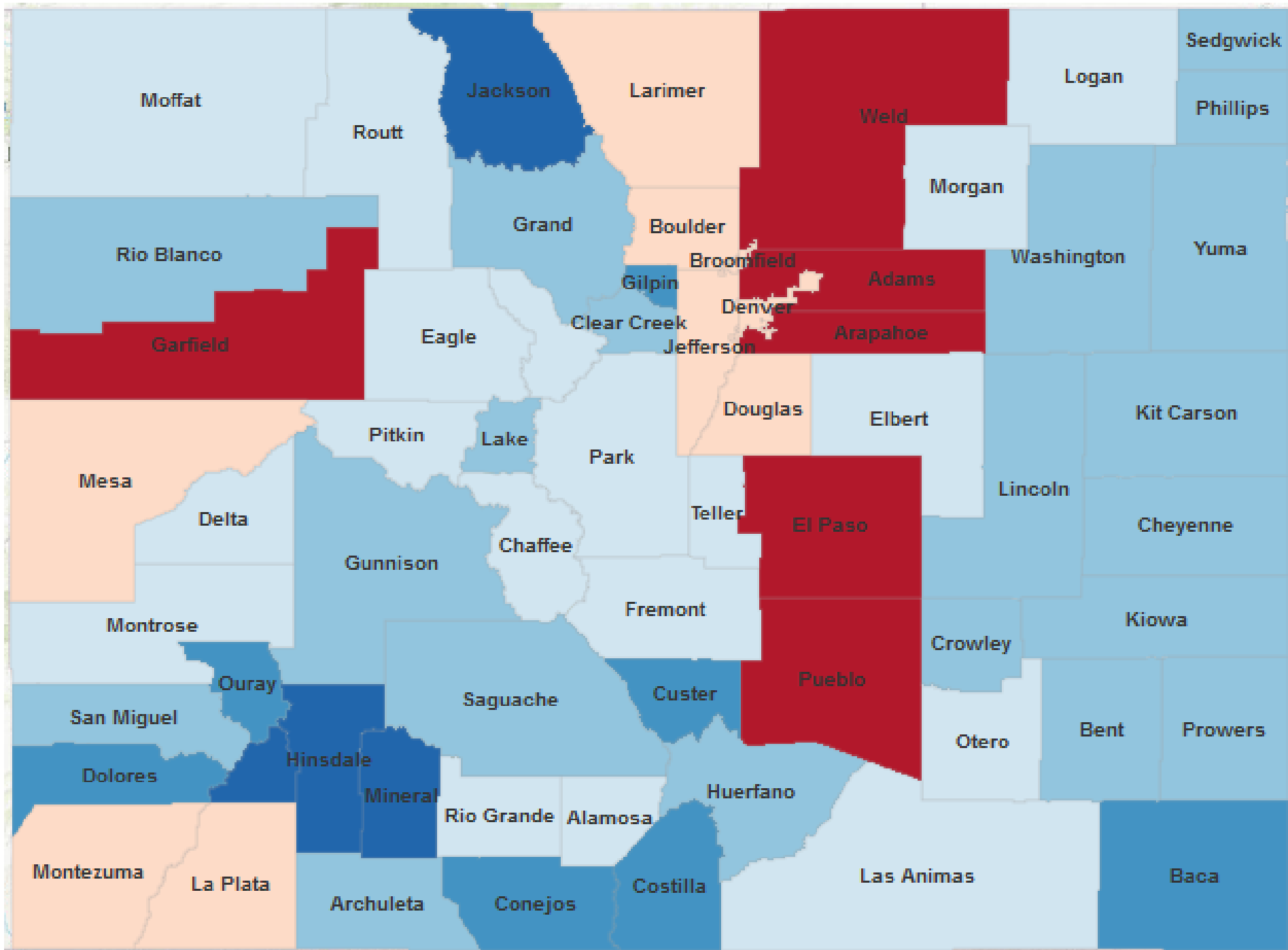- Legal implications

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

Disconnected the Relationship Between Population and Occupied Housing Units

Summary of findings –households- 2010 SF1 file vs DP

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

- Estimate of Hhds - 16,088 in SF to 24,382 in DP, 52% higher.
- Estimate of Hhds was 80% higher on average by county.
- Hhd size declined by 0.7, from 2.2 in SF1 to 1.5 in DP estimates.

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

- Estimate of Hhds - 105,558 in SF to 122,080 in DP, 16% higher.
- Estimate of Hhds was 18% higher on average by county.
- Hhd size declined by 0.3, from 2.3 in SF1 to 2.0 in DP estimates.

COLORADO
Department of Local Affairs

Summary of findings –households- 2010 SF1 file vs DP

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

- Estimate of Hhds - 182,778 in SF to 188,813 in DP, 3% higher.
- Estimate of Hhds was 4% higher on average by county.
- Hhd size declined by 0.1, from 2.5 in SF1 to 2.4 in DP estimates.

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

- Estimate of Hhds - 1.668m in SF to 1.667m in DP, 2% lower.
- 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.

COLORADO
Department of Local Affairs

Summary of findings –households with children

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

Change in the number of families with own children, 2000 to 2010
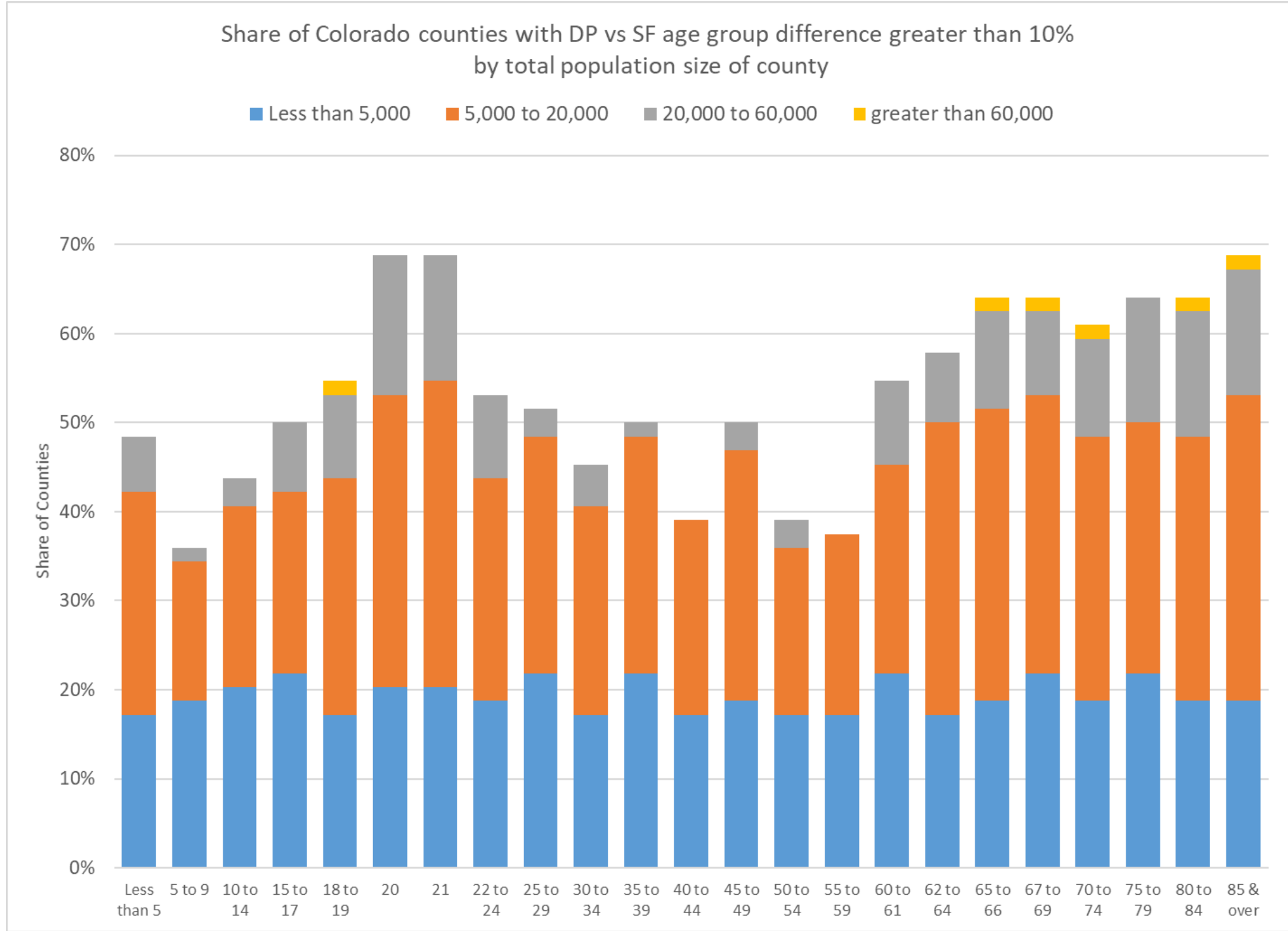
Using the SF data as released in 2010:
- 43 counties decline
- 20 counties increase, the largest increase 53% in a high growth county.

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

Share of Colorado counties with DP vs SF age group difference greater than 10% by total population size of county

■ Less than 5,000   ■ 5,000 to 20,000   ■ 20,000 to 60,000   ■ greater than 60,000

Close to 70% of counties in Colorado have a greater than 10% difference in count of 20, 21 and 85 & over year olds.

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

COLORADO
Department of Local Affairs

Summary of findings – 2010 SF1 file vs DP

Region	Based on a $20,000,000 distribution			
Region	SF 2010	DP 2010	Difference	%
REGION 6	$317,510	$334,055	$16,545	5.2%
REGION 10	$645,209	$657,696	$12,487	1.9%
REGION 12	$387,562	$398,464	$10,902	2.8%
REGION 9	$536,412	$544,799	$8,386	1.6%
REGION 5	$205,944	$212,164	$6,220	3.0%
REGION 2B	$920,163	$923,107	$2,945	0.3%
REGION 8	$307,626	$309,712	$2,085	0.7%
REGION 11	$1,095,886	$1,097,576	$1,689	0.2%
REGION 13	$489,215	$490,405	$1,189	0.2%
REGION 14	$180,027	$179,651	-$377	-0.2%
REGION 3B	$1,037,021	$1,036,087	-$934	-0.1%
REGION 1	$414,792	$412,194	-$2,598	-0.6%
REGION 7	$884,277	$880,616	-$3,661	-0.4%
REGION 2A	$1,220,089	$1,212,015	-$8,074	-0.7%
REGION 4	$2,391,012	$2,378,013	-$12,998	-0.5%
REGION 3A	$8,967,254	$8,933,447	-$33,807	-0.4%

Understanding change in the size and geographic distribution of the older adult population affects the Intrastate Funding Formula (IFF).

COLORADO
Department of Local Affairs

Summary of findings –age trends– 2010 SF1 vs DP

Understanding change in the age and sex of residents over time has implications for health statistics. Computing General Fertility Rate (GFR) (Births/Females age 15-44):

• 5 counties change from increasing GFR to decreasing. Largest change Baca County, 2000 GFR 73.7 to 2010 SF 75.0 or DP 64.0.

• 3 counties change from decreasing GFR to increasing. Largest change Sedgwick County, 2000 GFR 66.7 to 2010 SF 63.4 or DP 108.2.

• Of the counties with GFRs computed using the 2010 SF estimate in the denominator ranking in the top 10, only 5 remain in the top 10 when using the 2010 DP estimates as the denominator.

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
  - Geography
  - Variables

- Population to household/housing unit relationship

- Products

- Credibility

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 - soon!

Additional Resources

Short 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
- National Academies workshop https://sites.nationalacademies.org/DBASSE/CNSTAT/DBASSE_196518
- State network of state demographers - https://www.census.gov/programs-surveys/popest/about/fscpe/contacts.html
- State Data Centers - https://www.census.gov/about/partners/sdc/member-network.html

COLORADO
Department of Local Affairs

Questions

- Is there a step to identify where noise should be added?

- Why are small areas adversely impacted?

- Will differential privacy make it harder to do longitudinal studies, will previously released census data be re-released?

- Will additional invariants be considered? For example occupancy status?

COLORADO
Department of Local Affairs

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	2010 DP demonstration data
• Block – total population	• State – total population
• Block – voting age population	
• Block – total housing units	• Block – total housing units
• Block – occupancy status	
• Block – group quarters count	• Block – group quarters count
• Block – group quarters type	• Block – group quarters type

COLORADO
Department of Local Affairs