

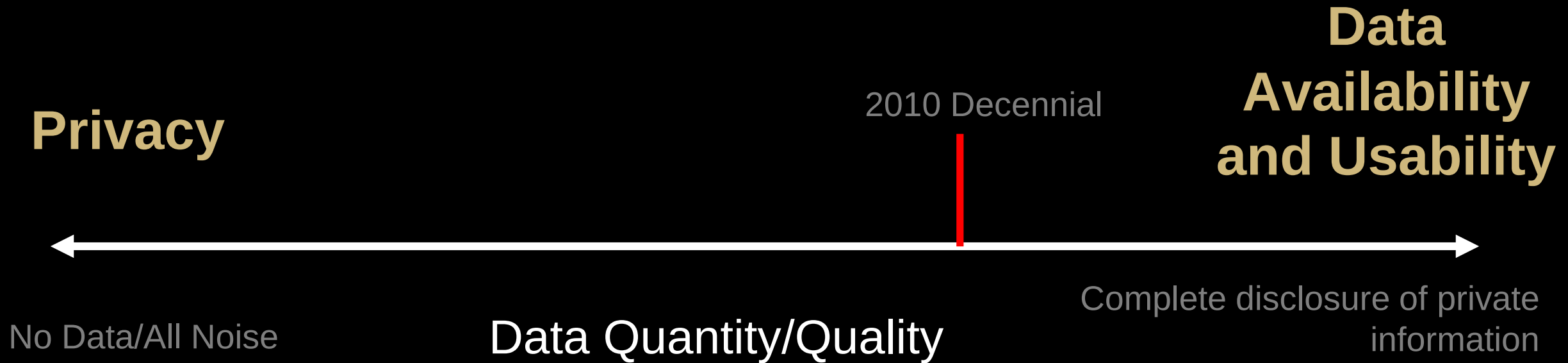
Geographic Review of Differentially Private Demonstration Data

**David Van Riper, University of Minnesota
Seth Spielman, University of Colorado**

Agenda

- ✓ Preliminary investigations of the differences between the original 2010 data and the new DP data.
 - ✓ Various types of geography
 - ✓ Metro-scale measures of inequality
 - ✓ Spatial patterns of populations within cities
- ✓ DP Products and Process.

Big Picture: Decennial Data



Big Picture: Census2020



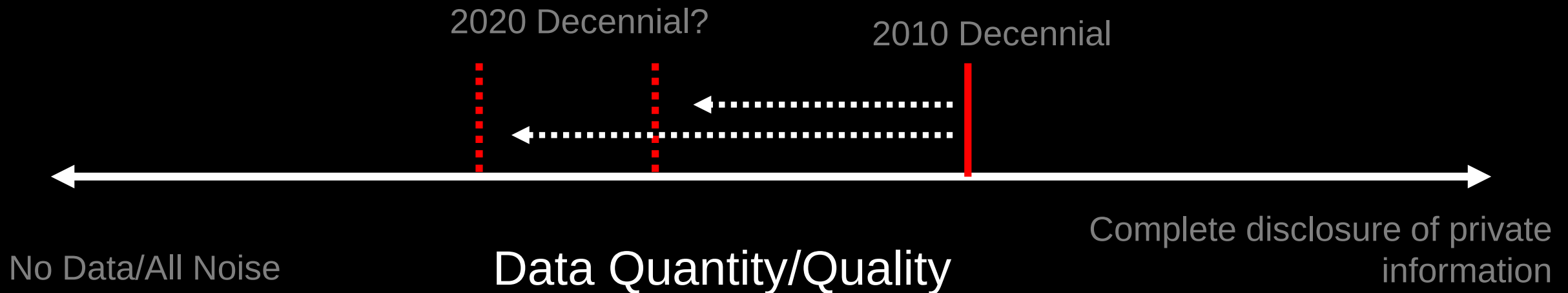
Big Picture: Census2020

Are we on the right position in this continuum?

Privacy
Private Good

To assess this question we believe you have to get under the hood and look at how these changes materially impact data users/use-cases.

Data
Availability
Public Good



Big Picture: Public Debate

What are the changes?

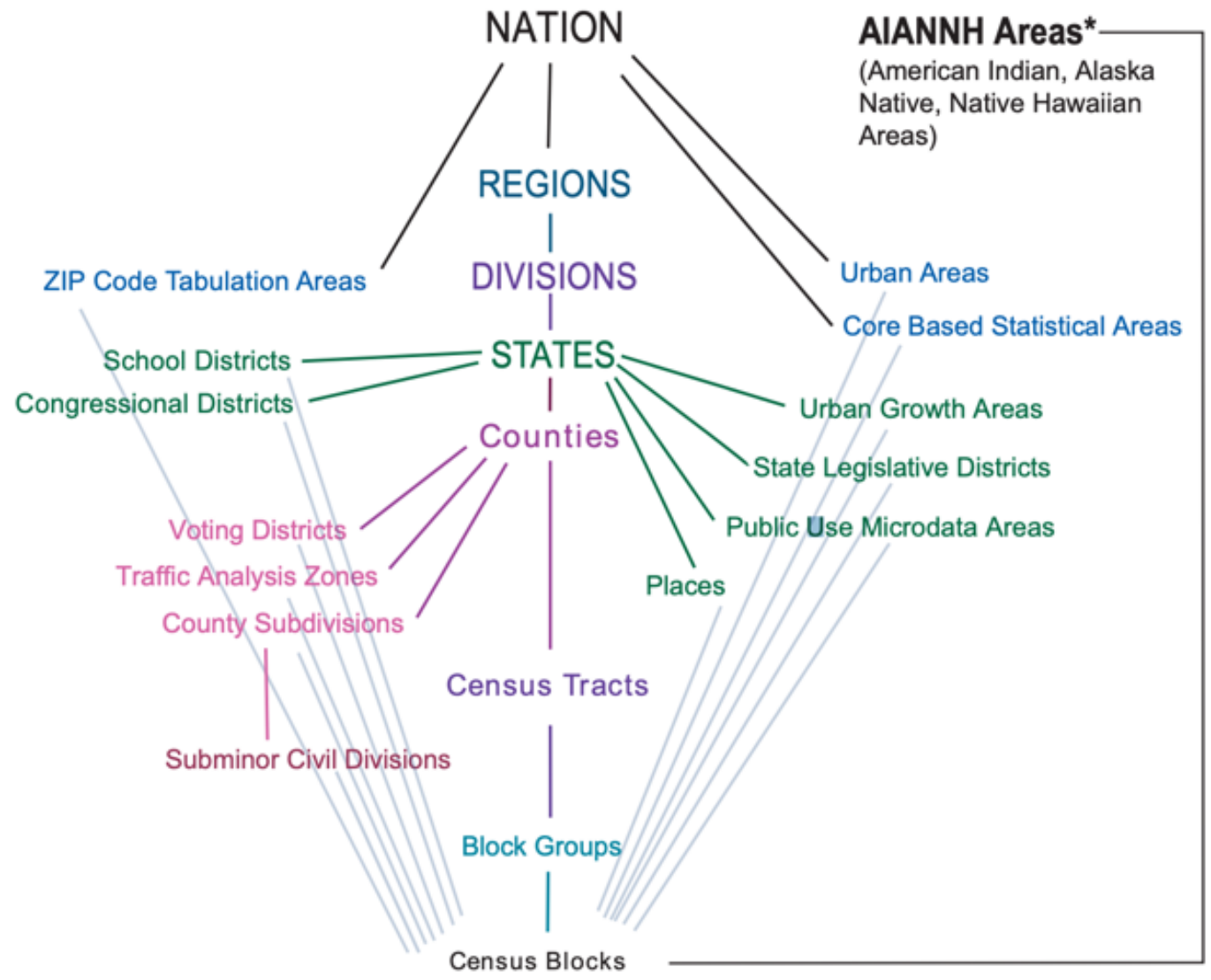
Is an empirical question. We'll share some insights, lots more to do...

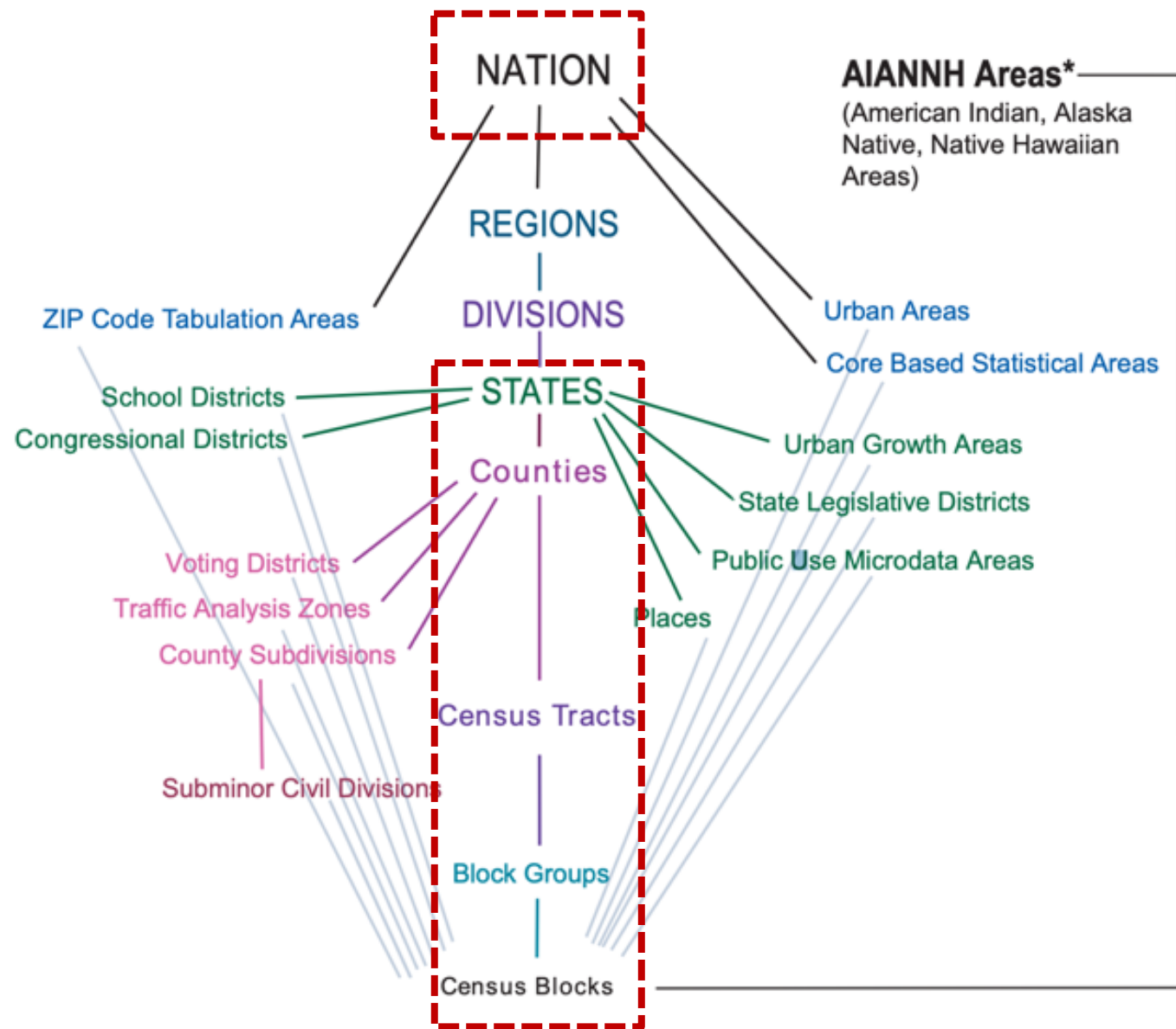
Are we in the right place on the continuum?

Is a political/personal question.

We believe its very hard to honestly answer until we understand the nature of the changes and their implications for policy, planning and governance.

Changes in
place-based
data...



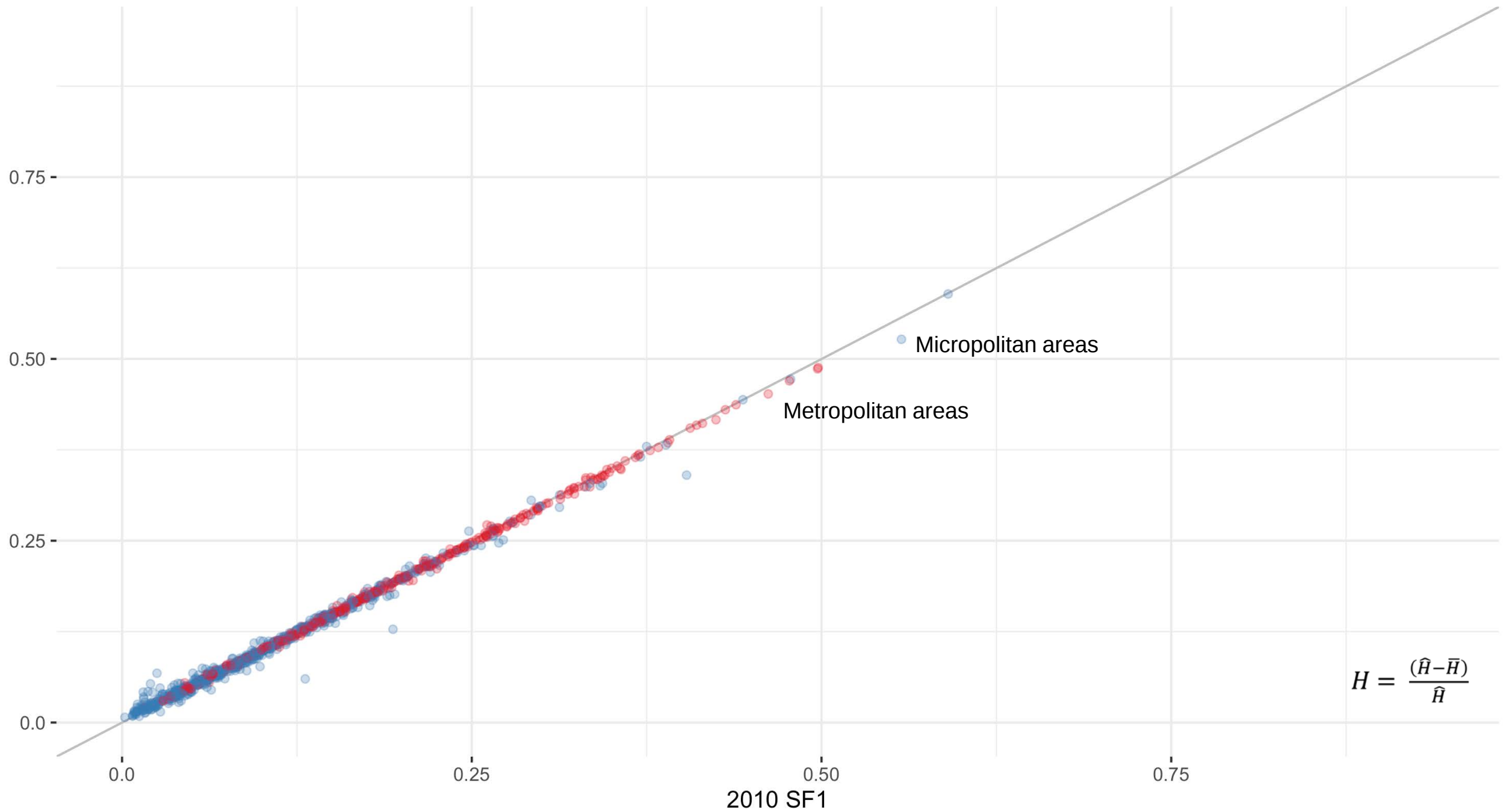


Describing Inequality

(Residential Segregation)

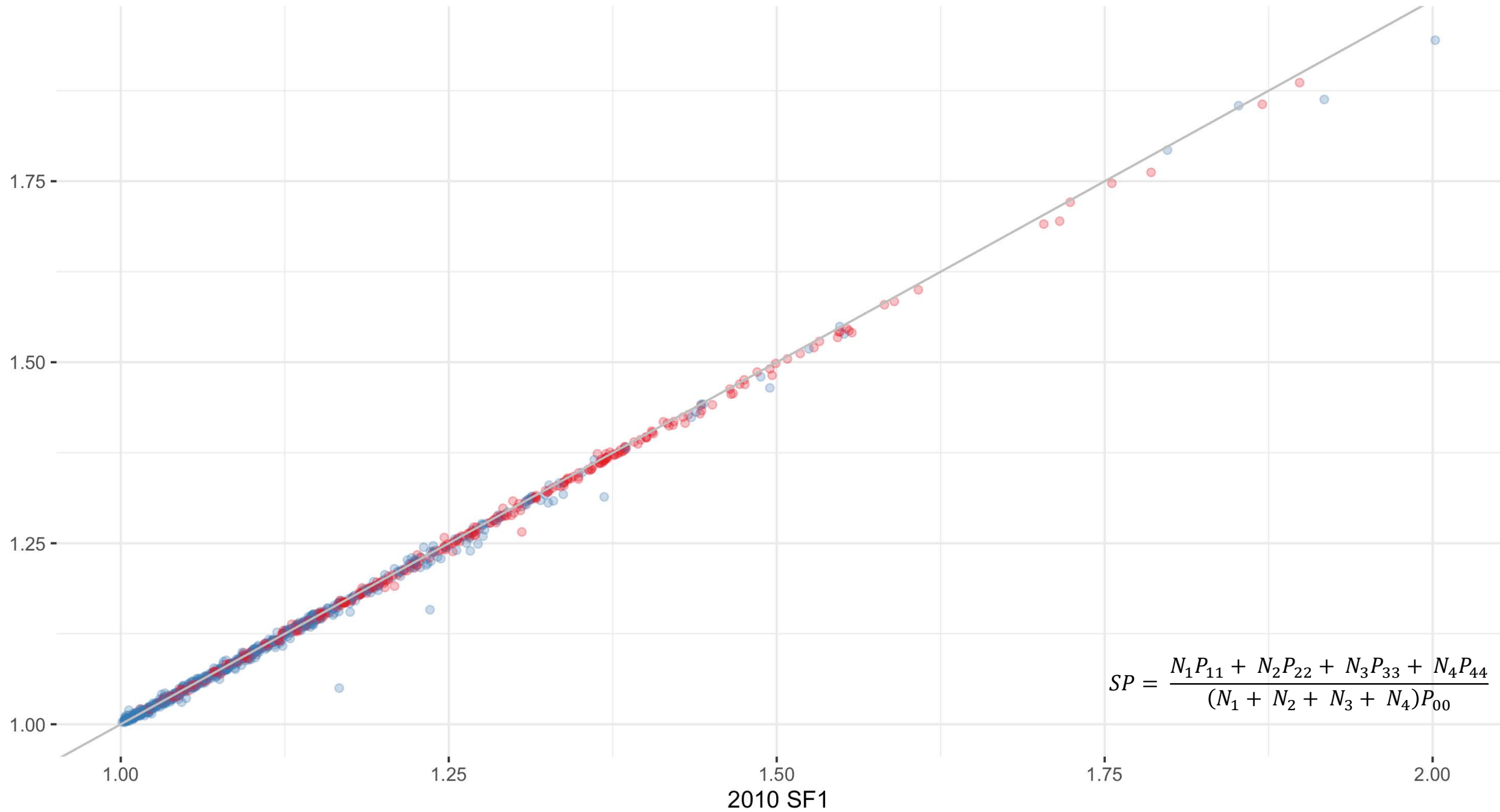
Multigroup Entropy (H), Core-based Statistical Areas

2010 Demo



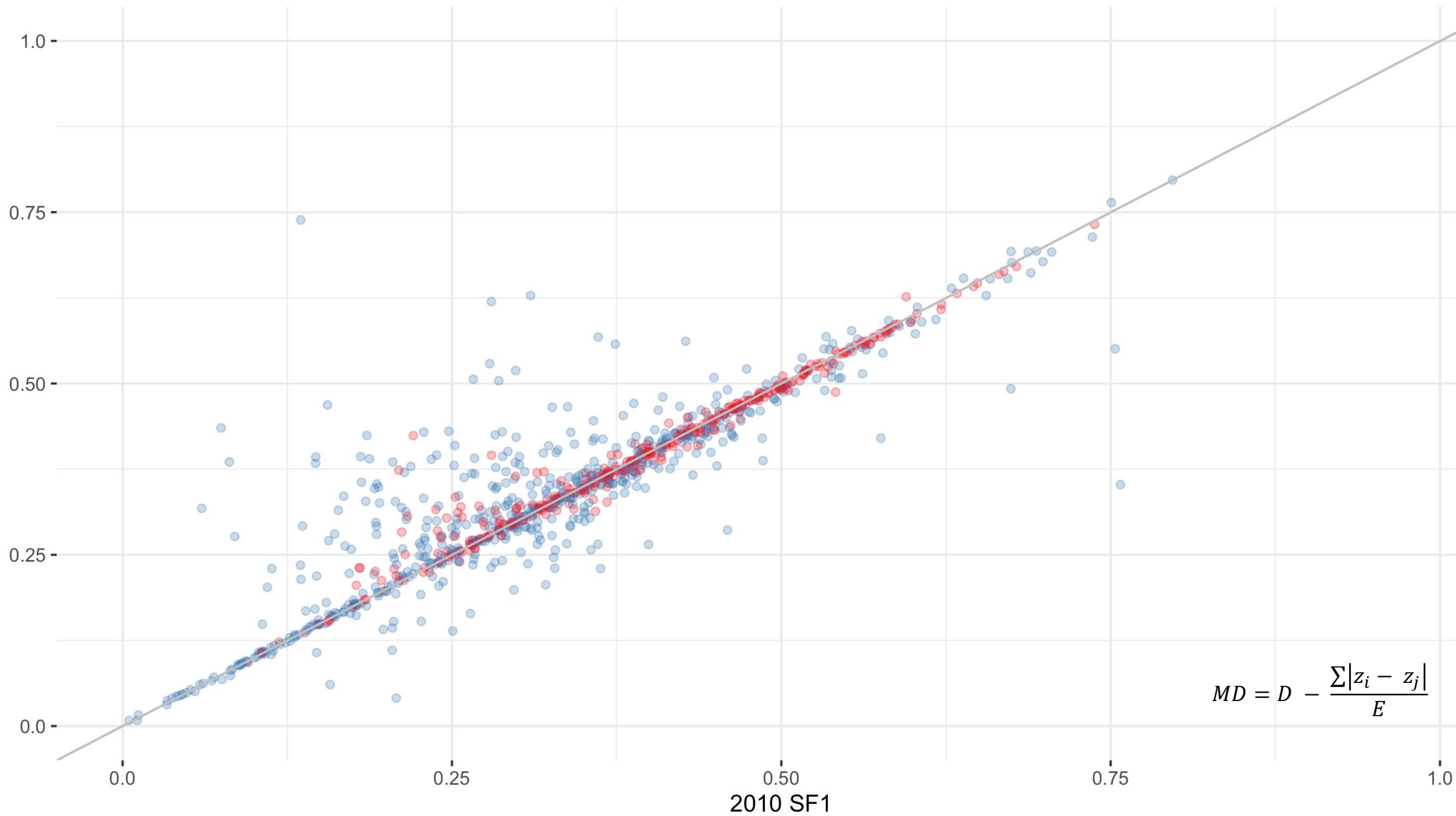
Multigroup Spatial Proximity (SP), Core-based Statistical Areas

2010 Demo

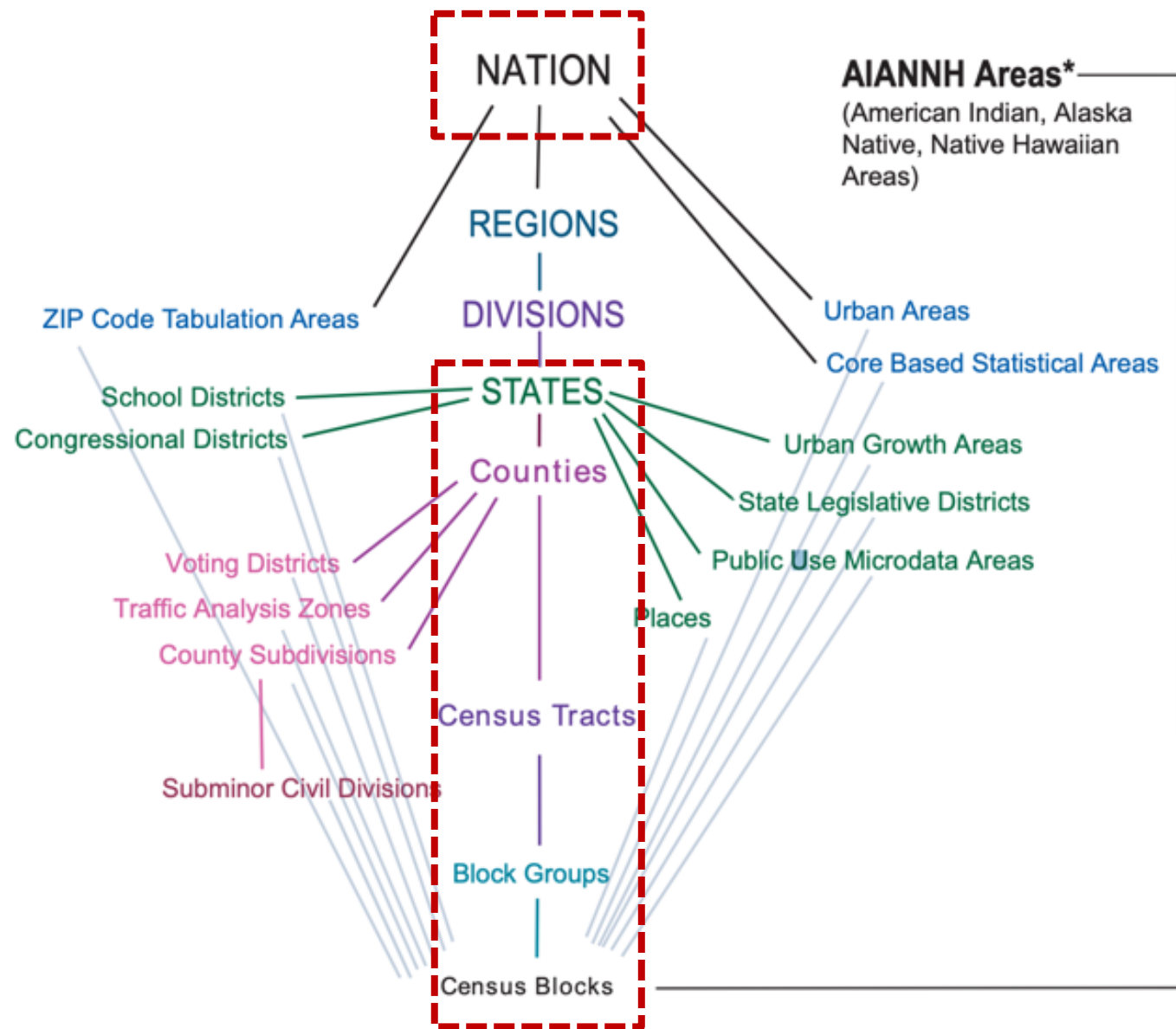


White-African American Morrill's D, Core-based Statistical Areas

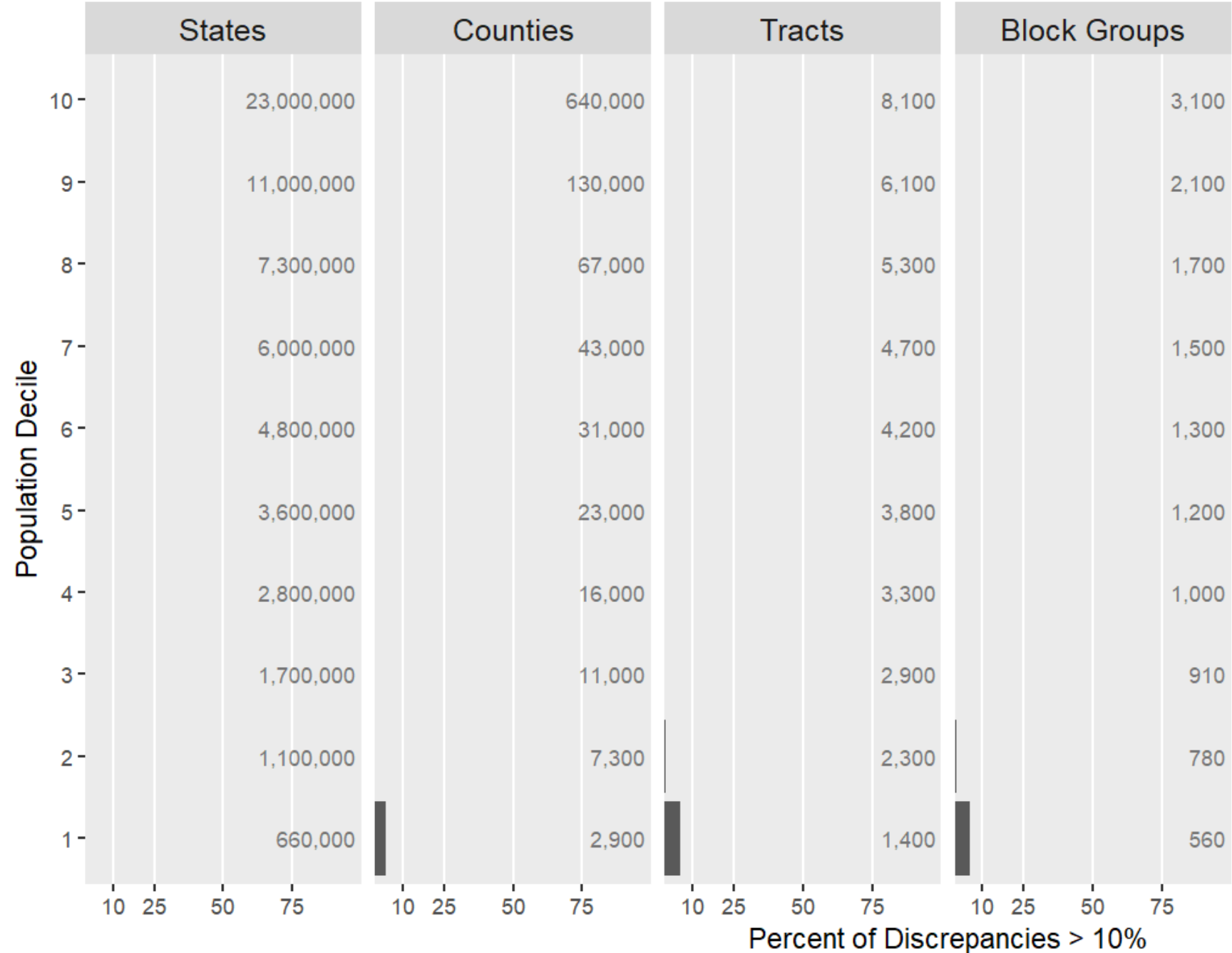
2010 Demo



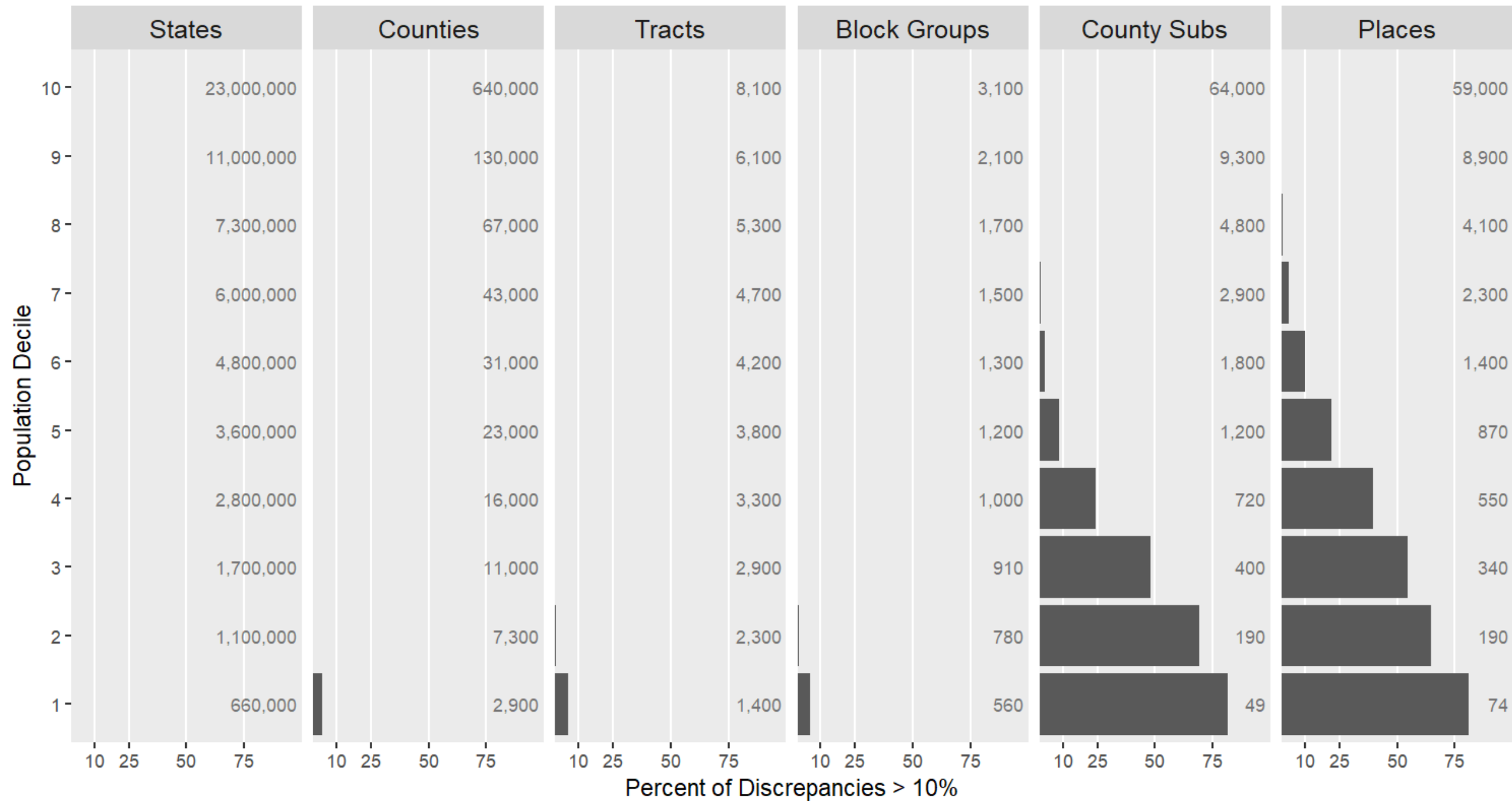
Changes along
geographic
hierarchy



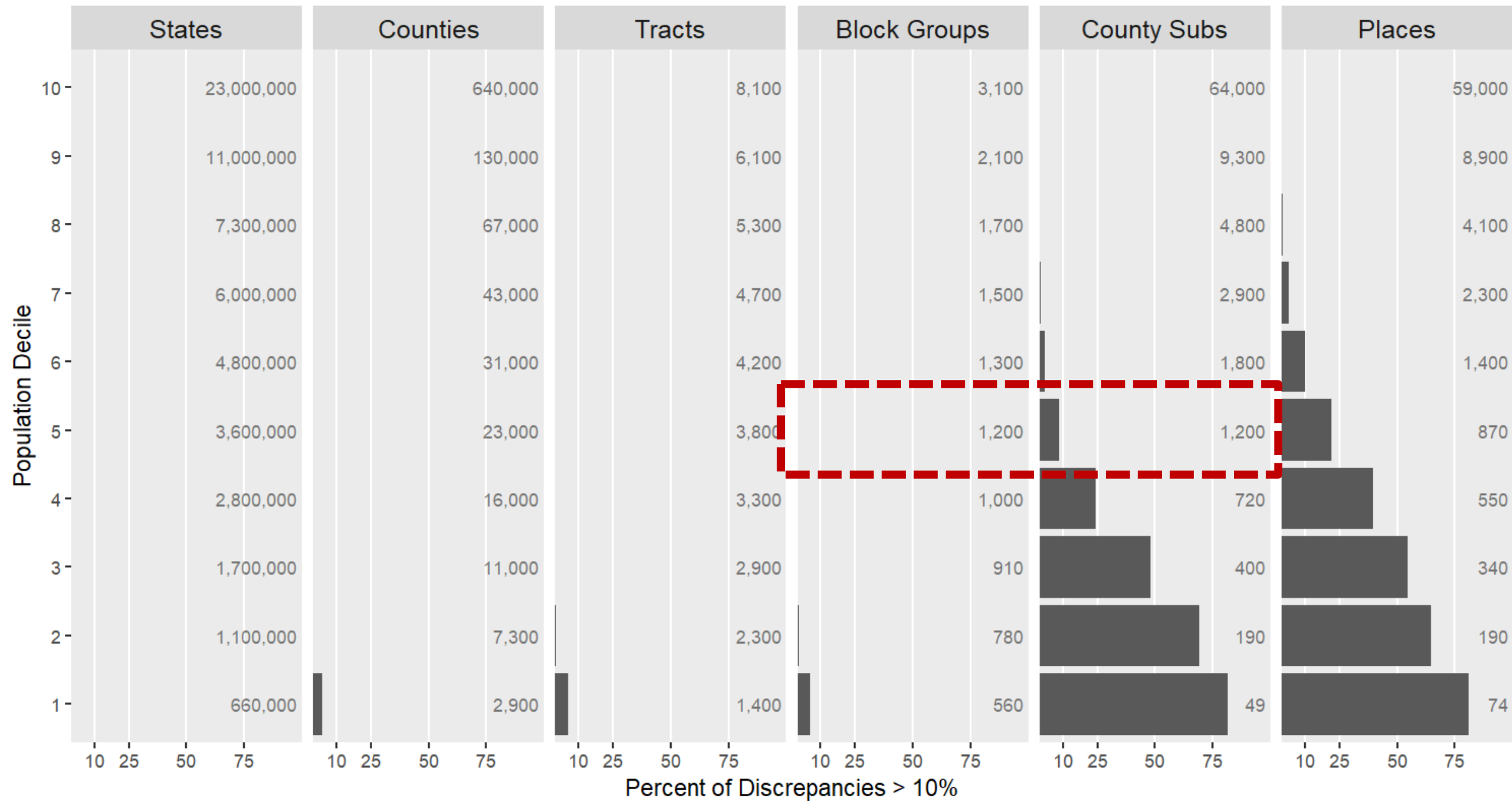
2010 SF1 vs. Demo: Total Population



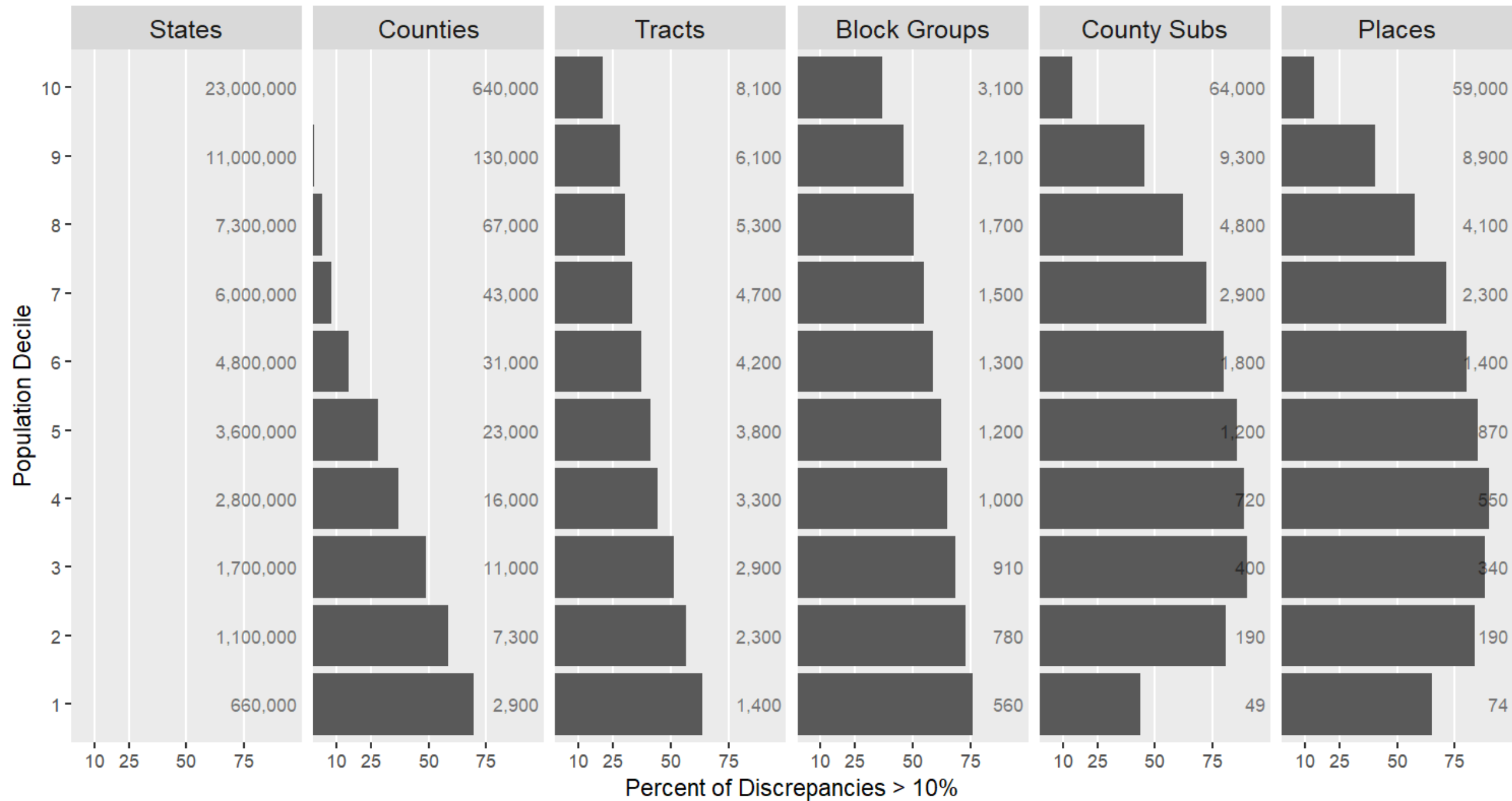
2010 SF1 vs. Demo: Total Population



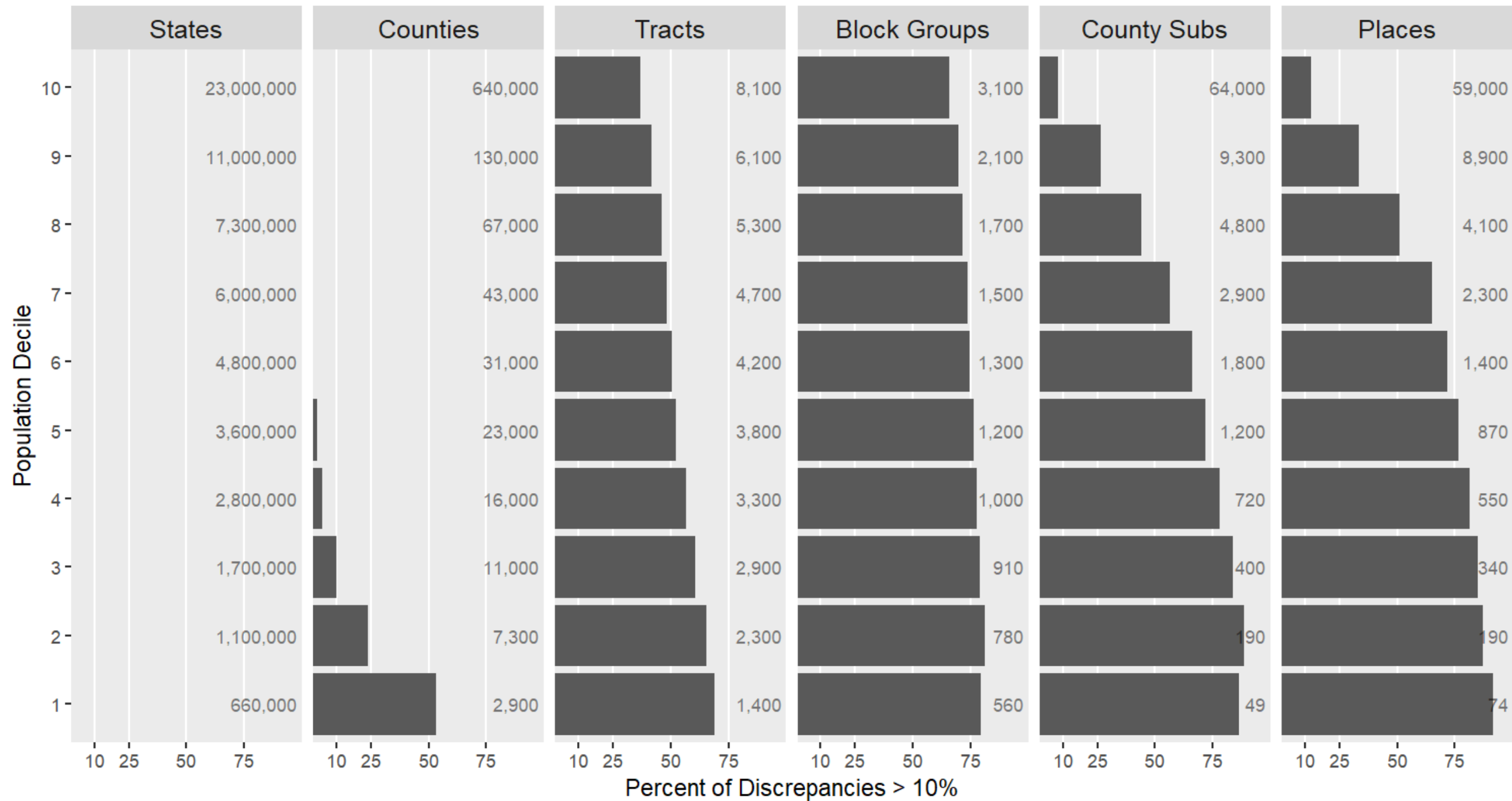
2010 SF1 vs. Demo: Total Population



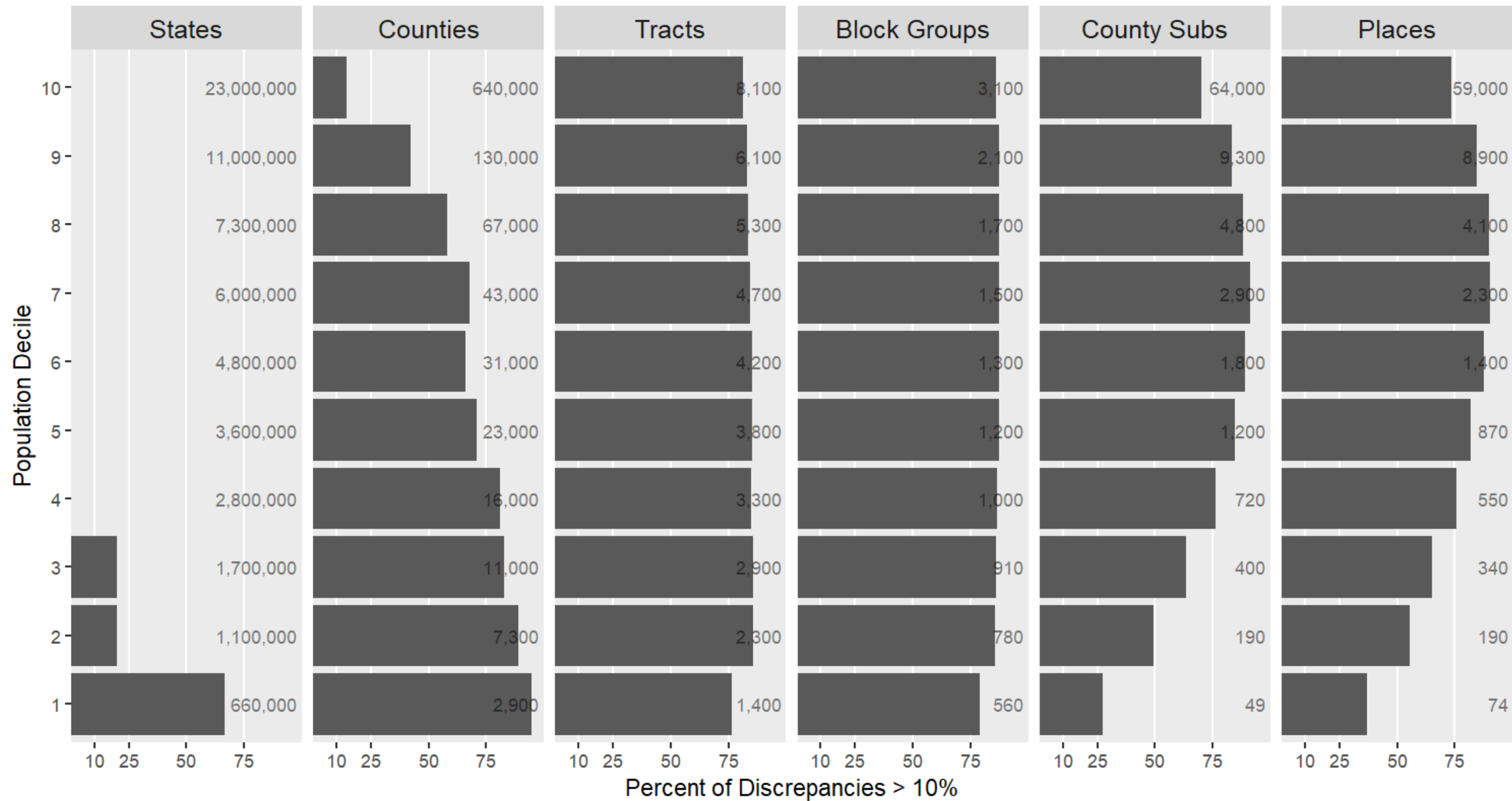
2010 SF1 vs. Demo: Hispanic/Latino Population



2010 SF1 vs. Demo: Persons 65 Years and Over



2010 SF1 vs. Demo: Black Householders

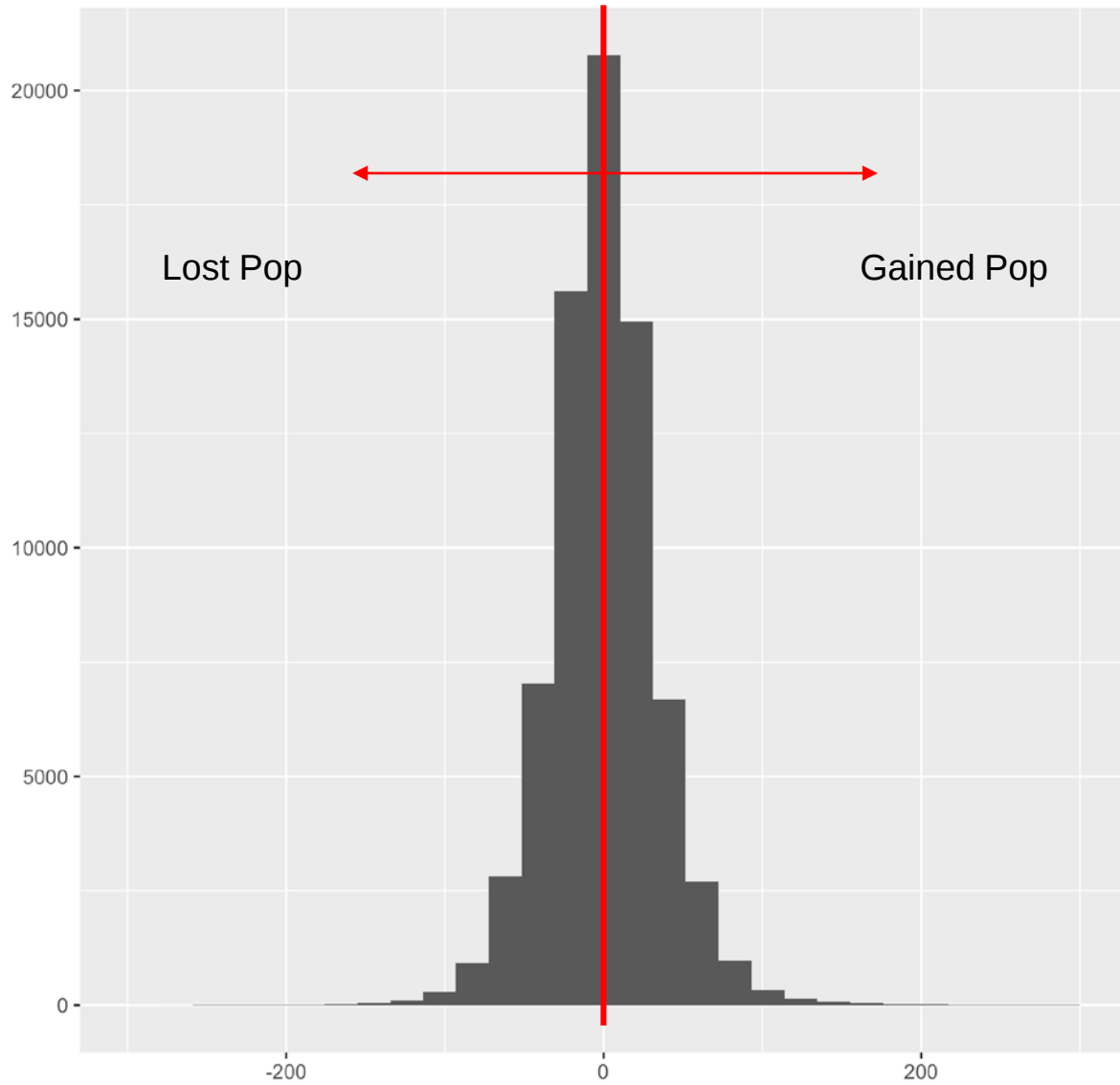


Summary: Changes by Geographic Resolution

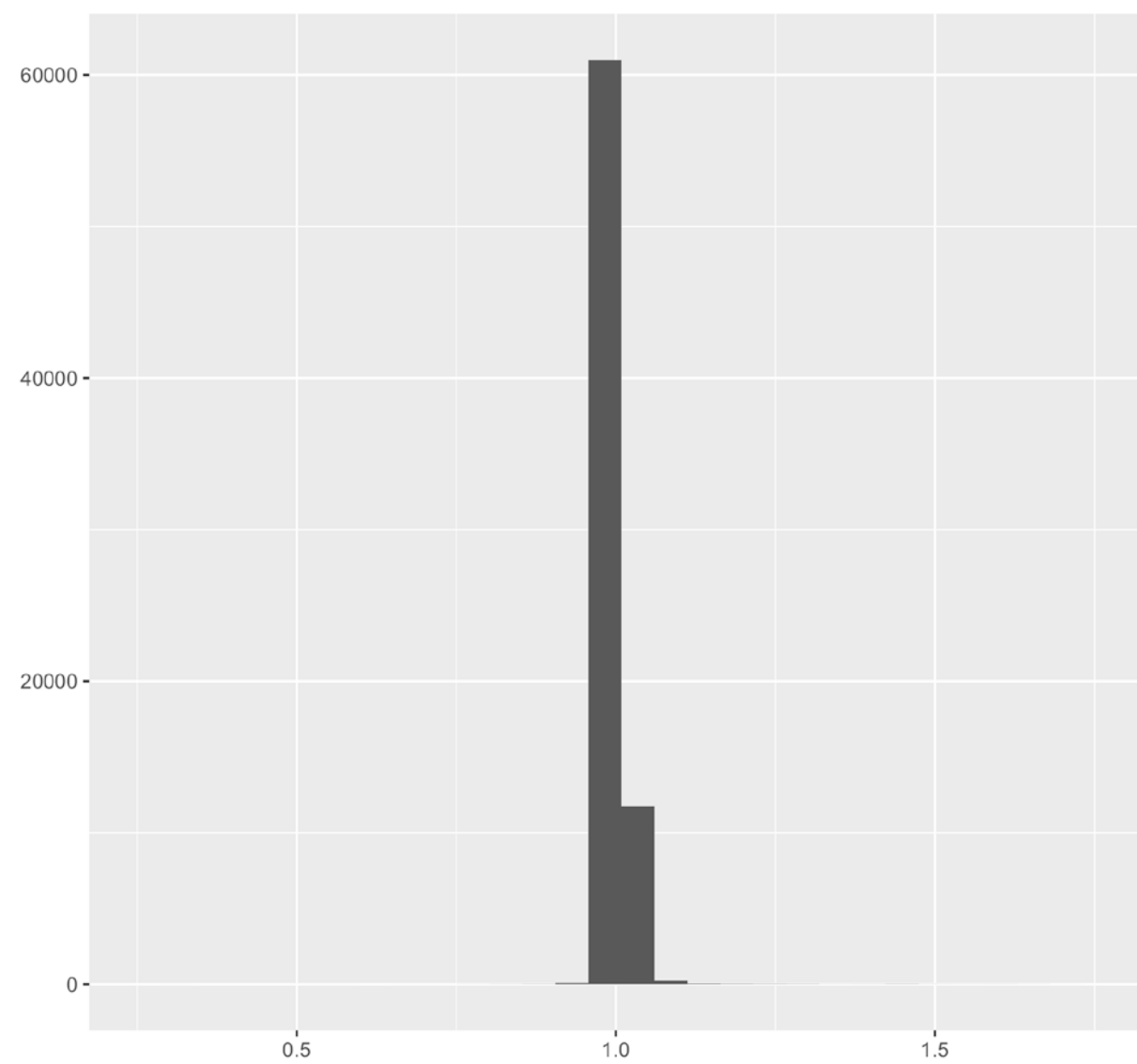
- Off-spine geographies are generally worse than on-spine geographies.
- Impact varies by place, population sub-group.

Spatial Patterns In Changes

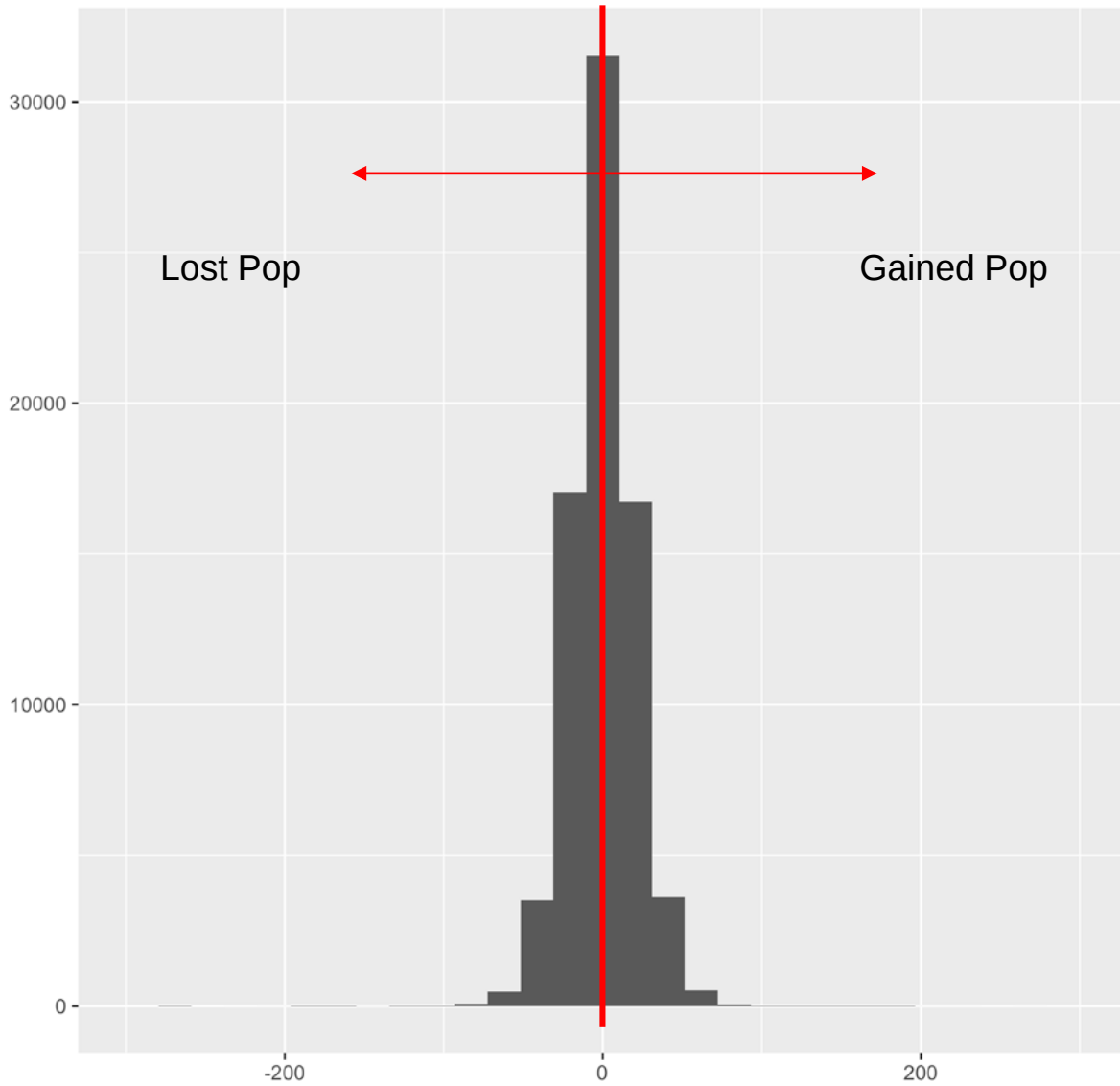
Difference Total Population: All US Tracts



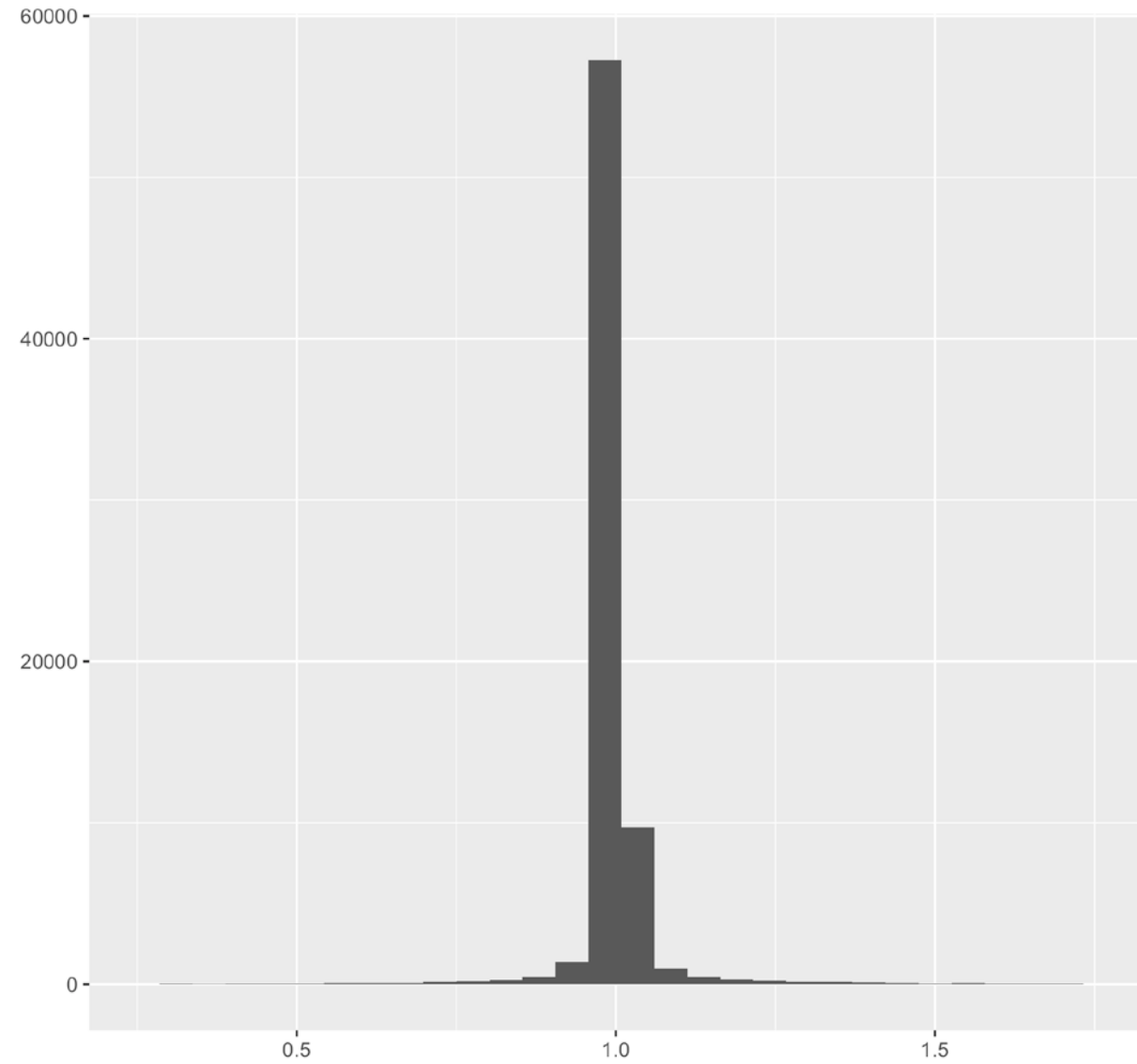
Ratio of Decennial Total Pop to DP Total Pop
All US Tracts



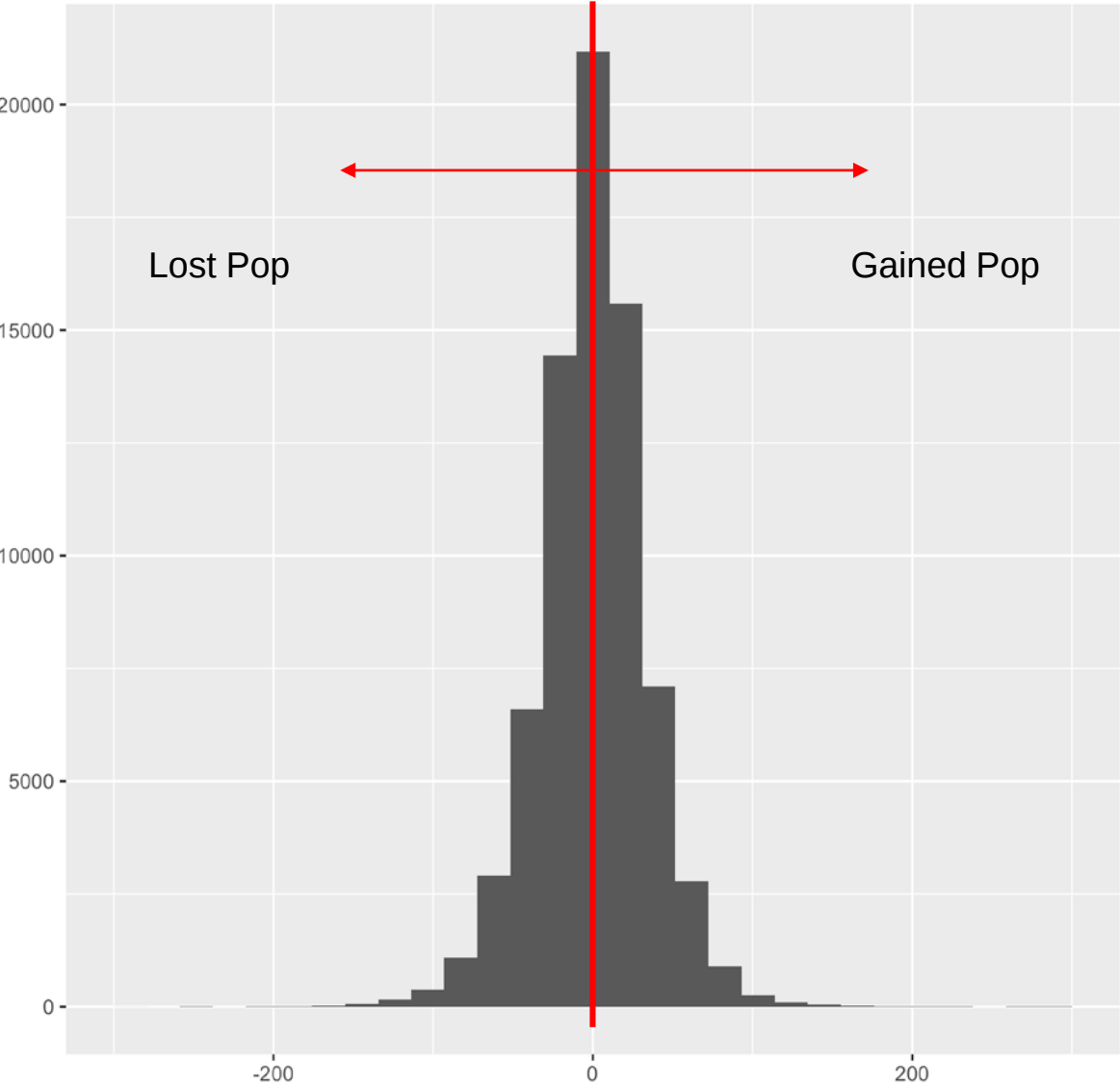
Difference White Population: All US Tracts



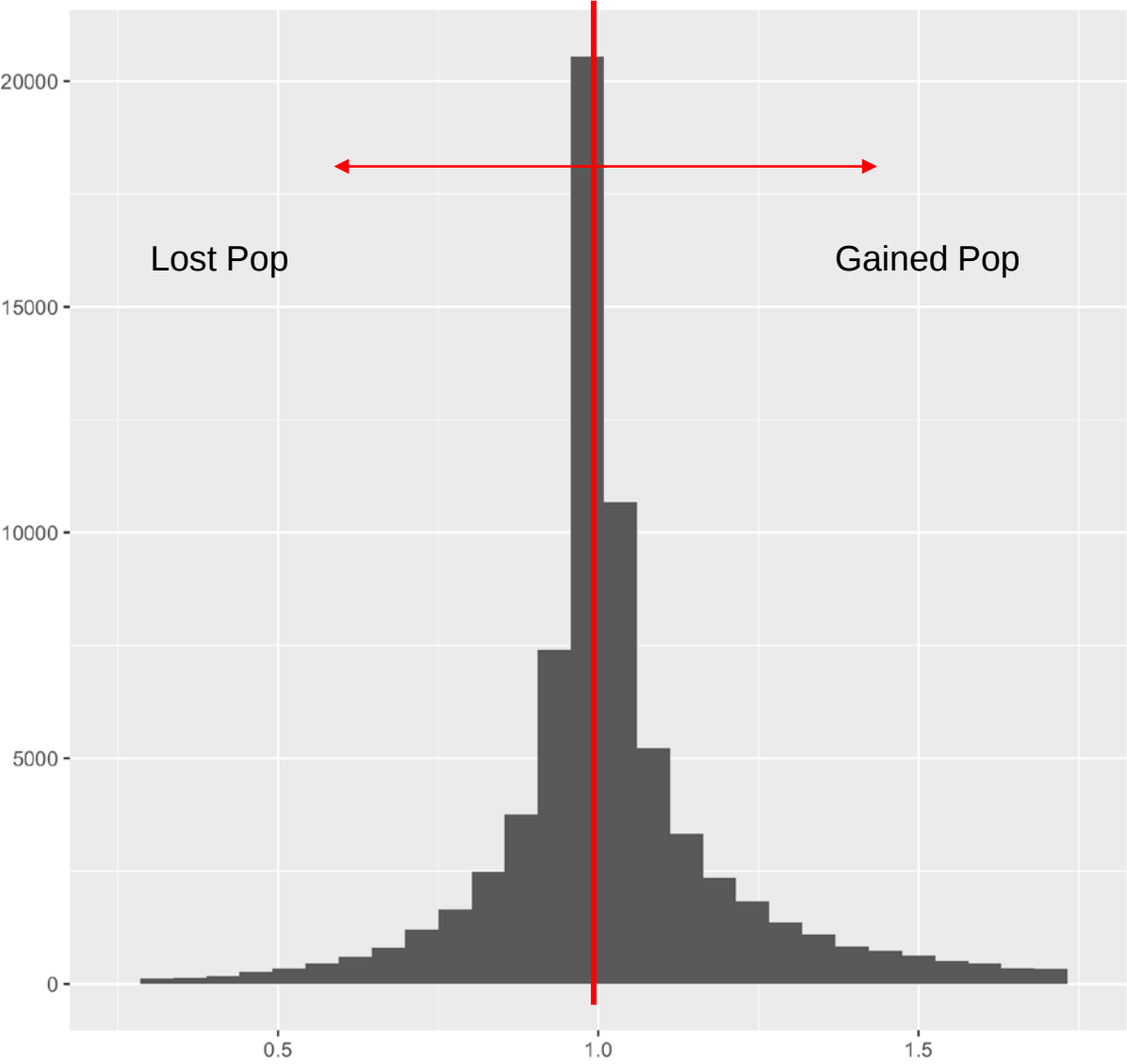
Ratio of Decennial White Pop to DP White Pop
All US Tracts



Difference Hispanic Population: All US Tracts

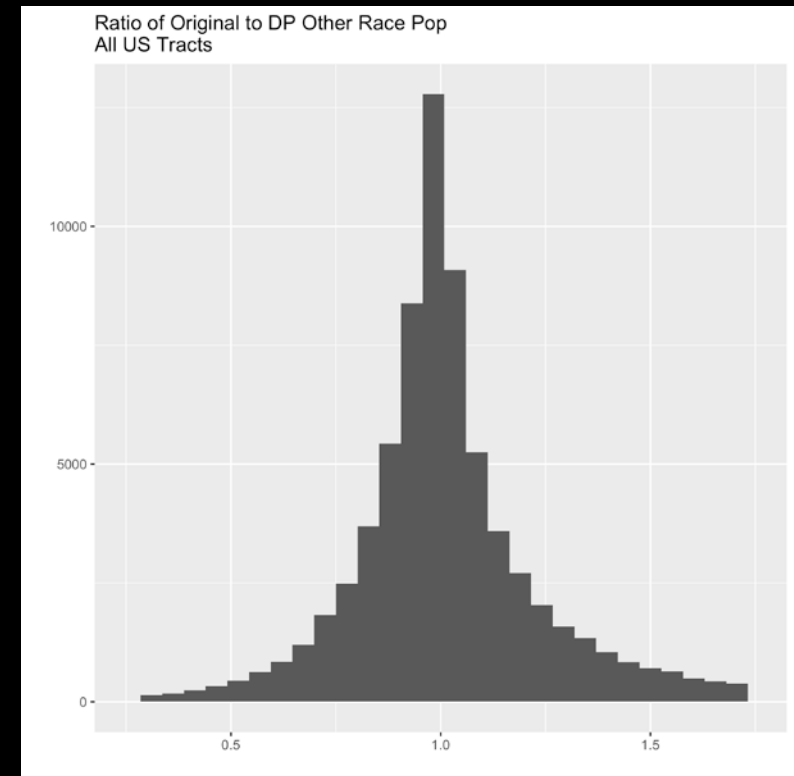


Ratio of Original Hispanic Pop to DP Hispanic Pop
All US Tracts



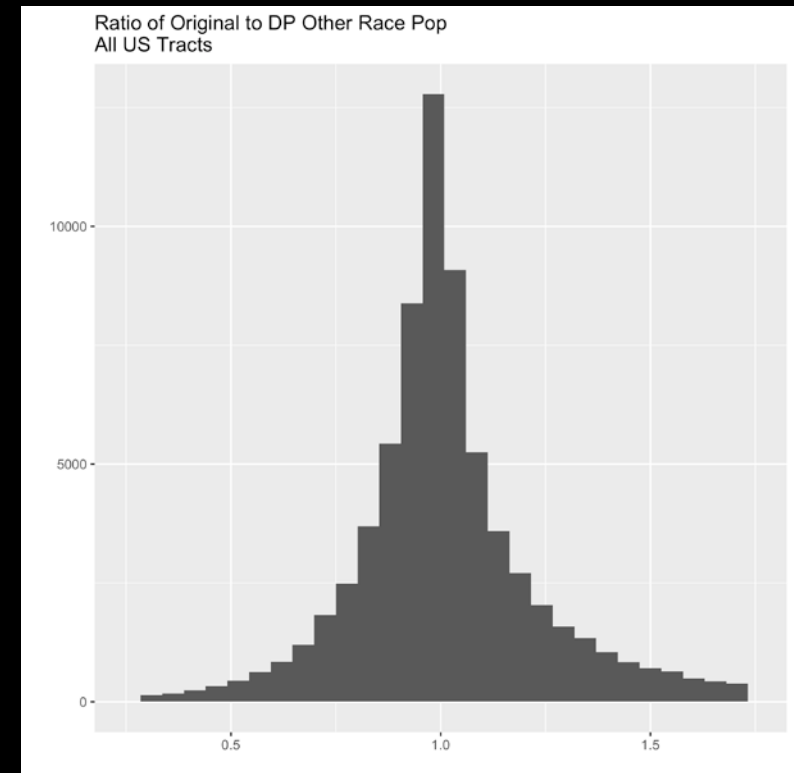
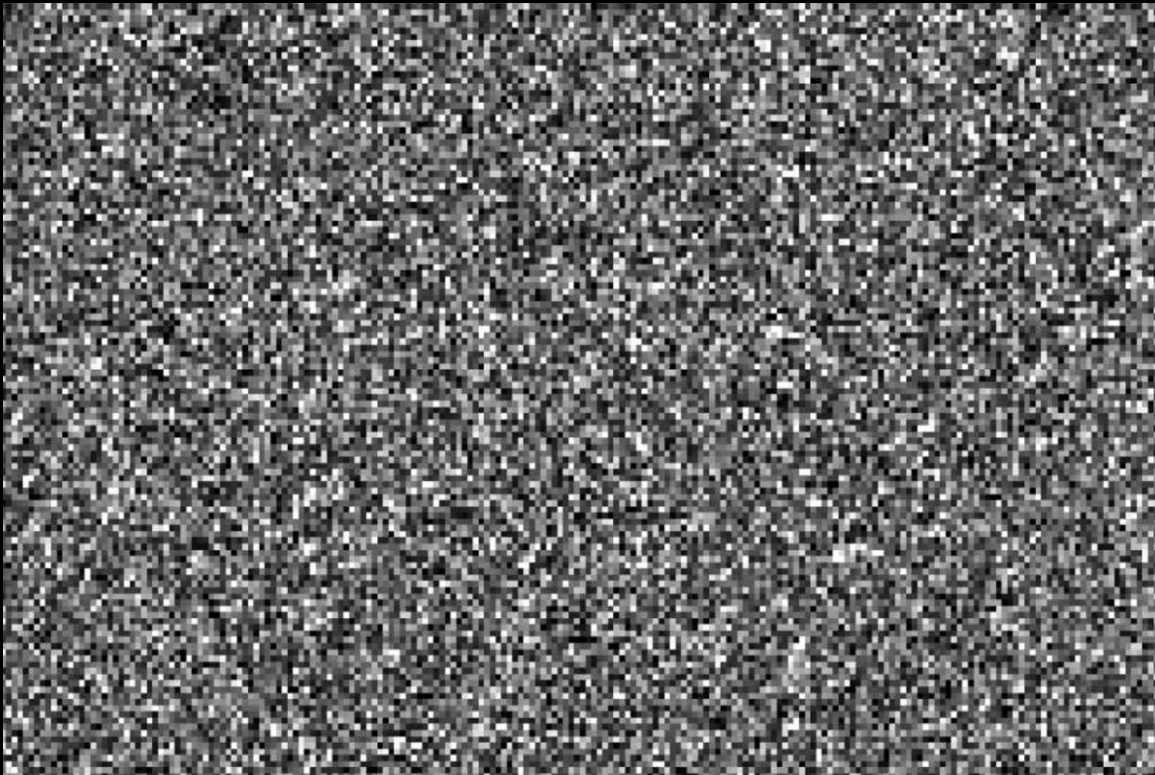
Spatial Patterns

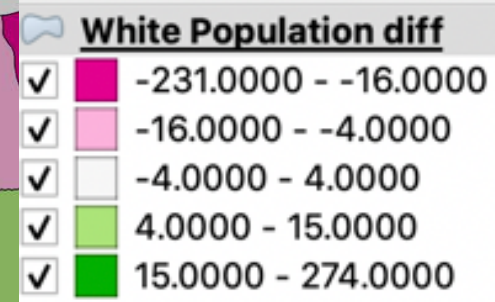
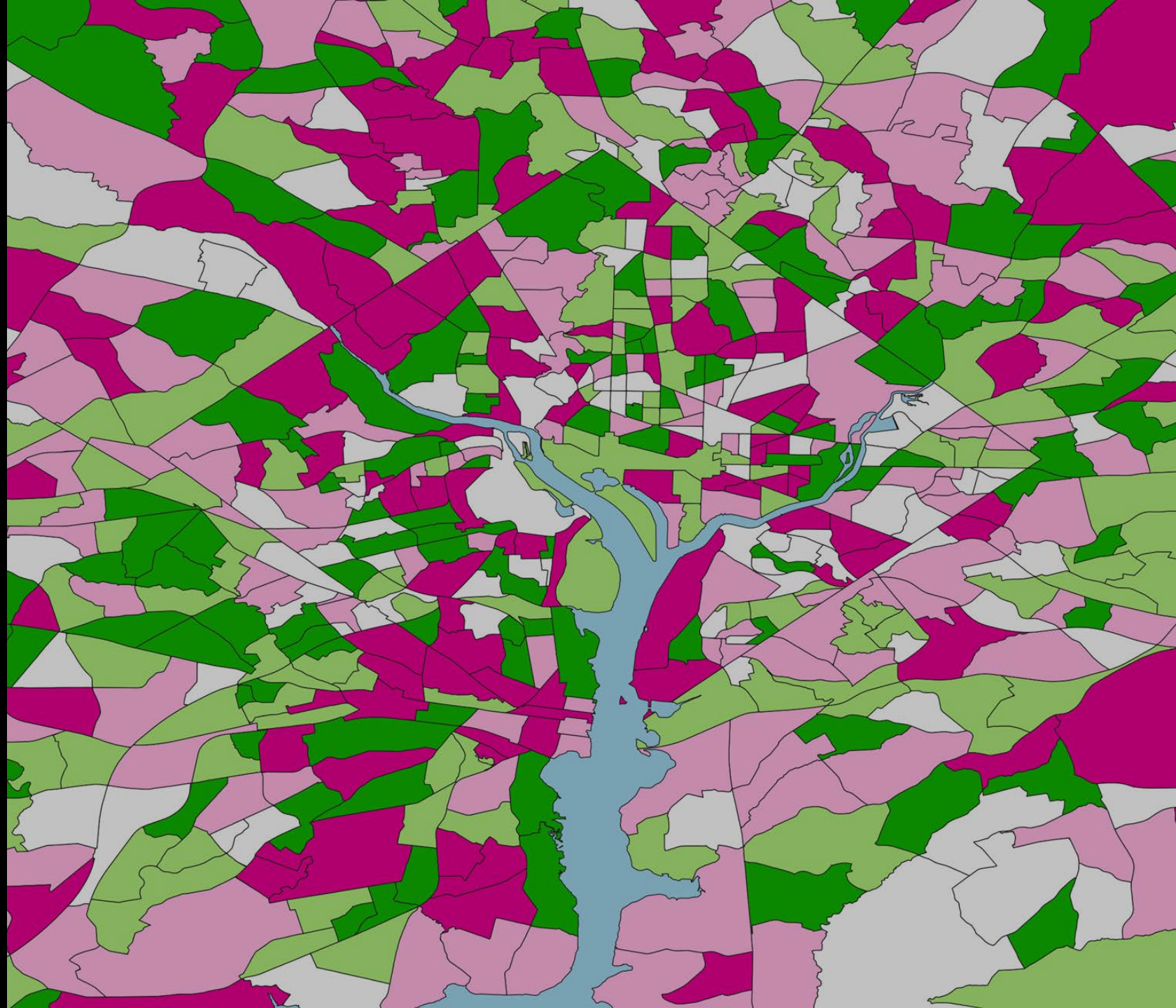
We can have a balanced histogram
biased pattern



Spatial Patterns

We can have a balanced histogram no pattern





Spatial Patterns: Moran's I

Are the changes randomly distributed in space?

Are particular parts of the map experiencing more infused noise than others?

We investigate this using a statistic called Moran's I.

It is simply the correlation between the value observed for each tract and its neighbors.

In this case the value of interest is the difference between the DP and the original population estimate.

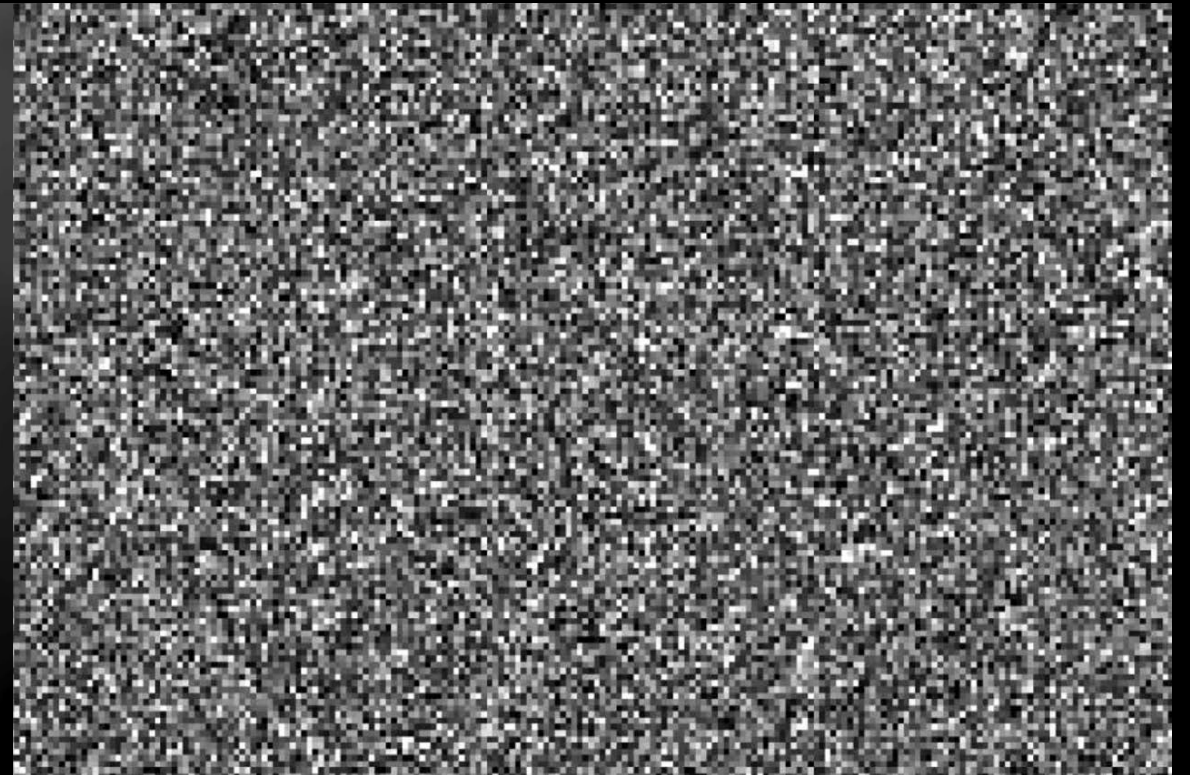
We expect these differences to be spatially random...

Spatial Patterns: Moran's I

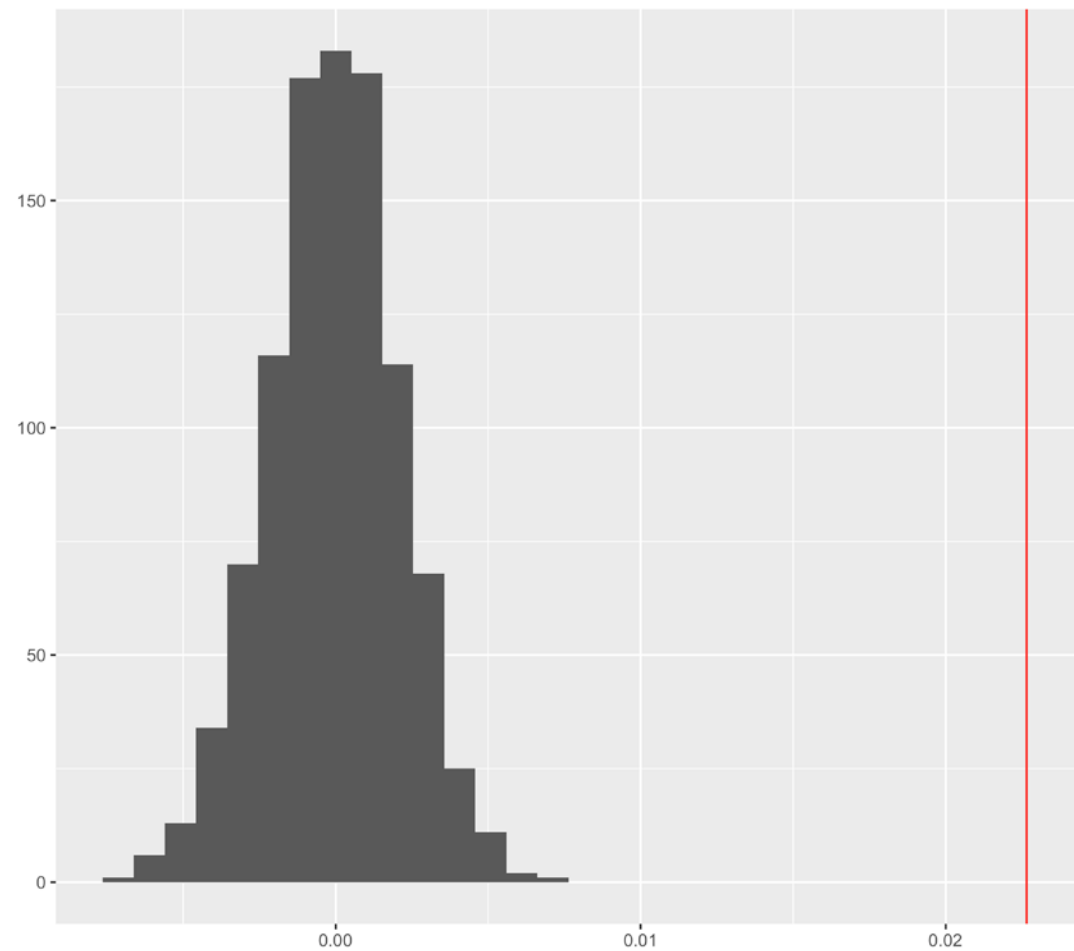
Spatial Pattern
High Moran's I



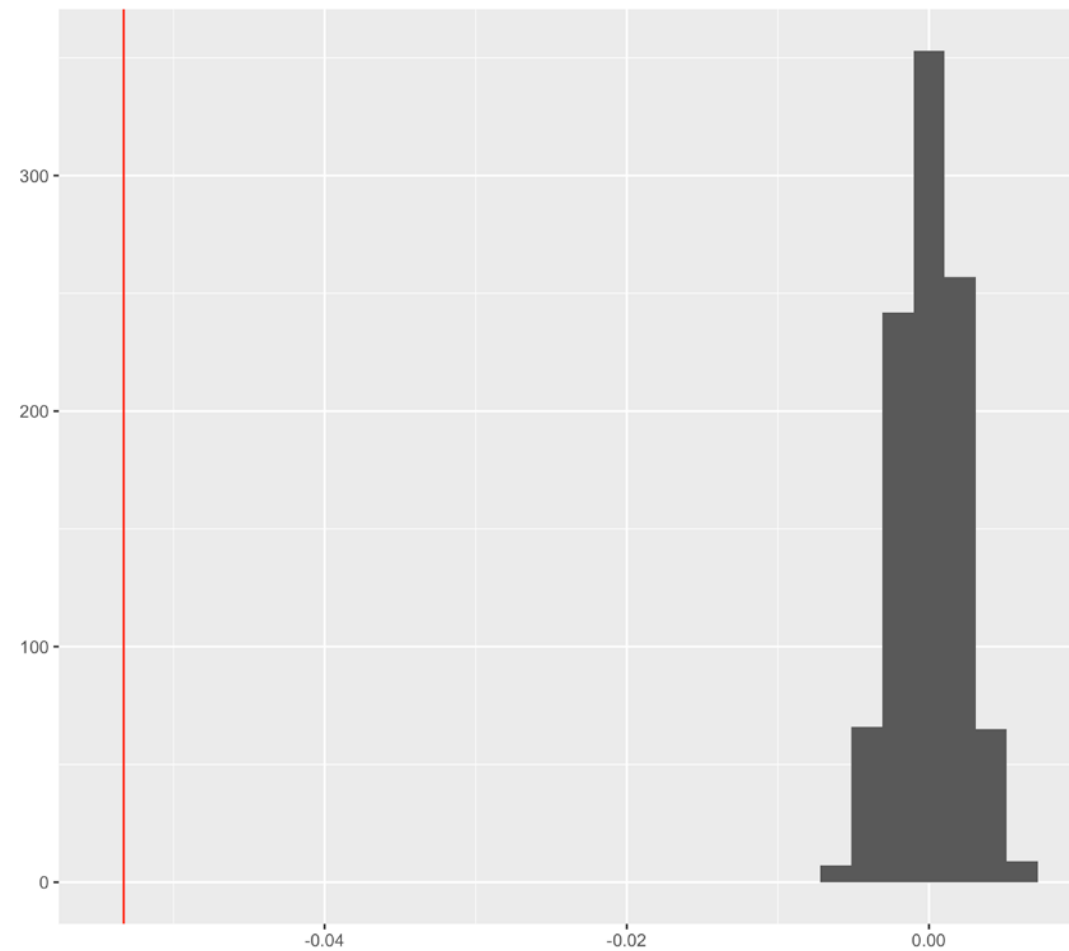
Spatial Randomness



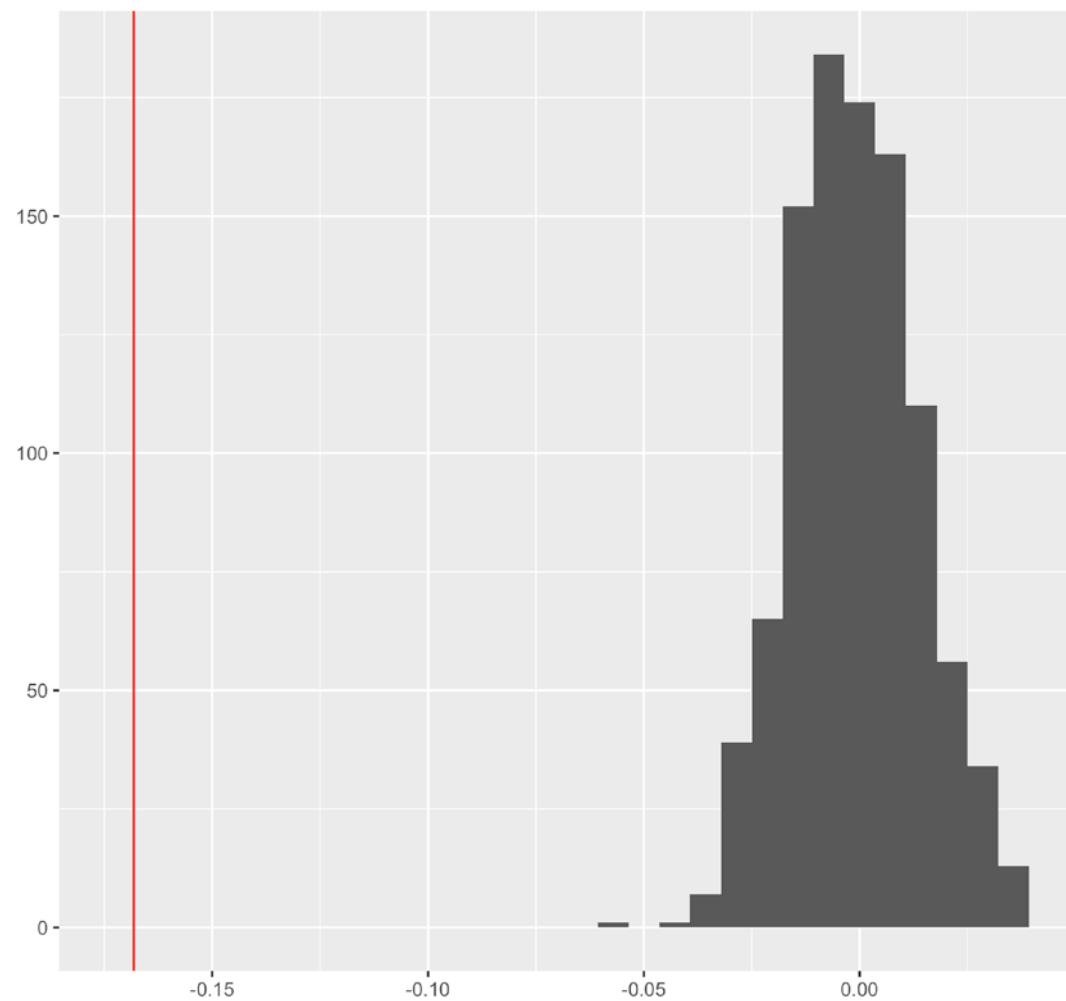
National Spatial Pattern in Differences in total pop
Observed Moran's I Compared 999 Random Maps



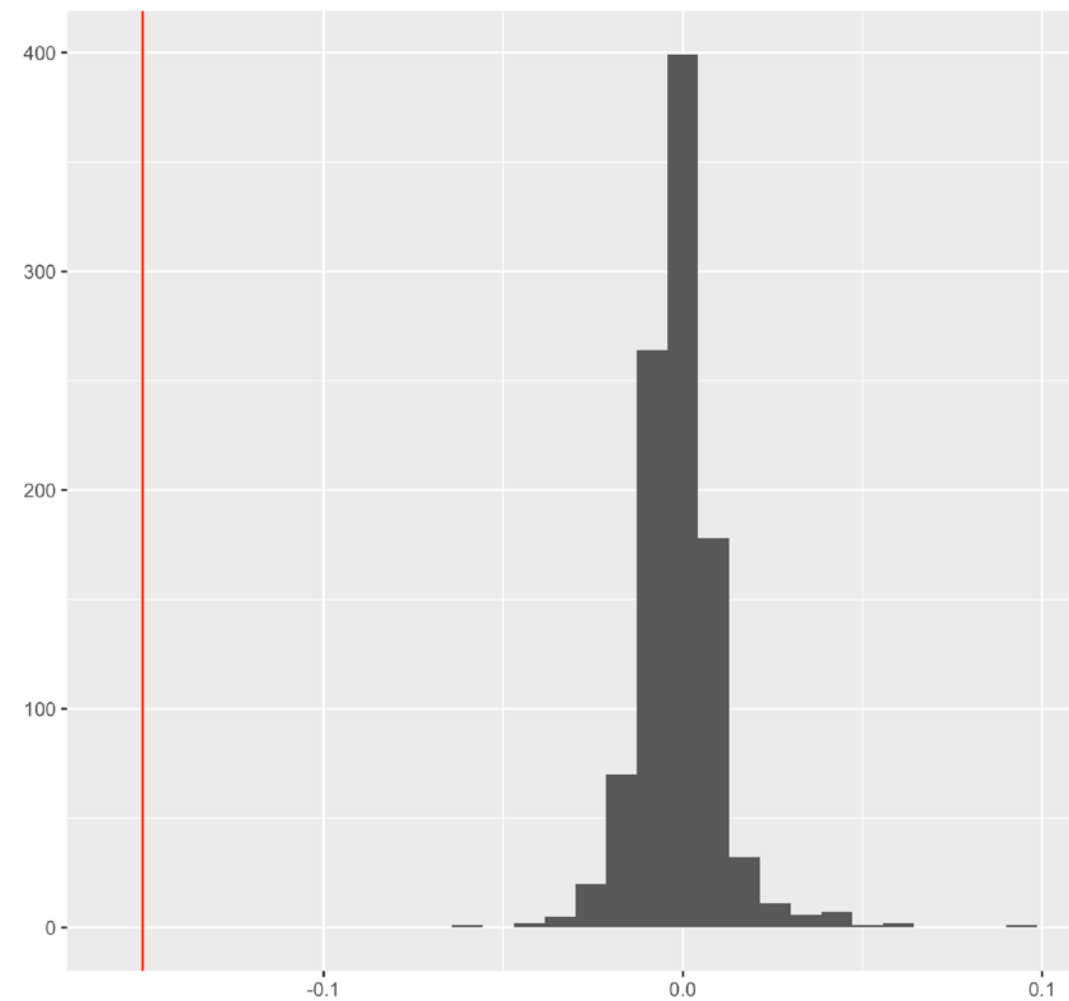
National Spatial Pattern in Differences in the White pop
Observed Moran's I Compared 999 Random Maps



Ohio Spatial Pattern in Differences in the White pop
Observed Moran's I Compared 999 Random Maps



Ohio Spatial Pattern in Differences in the Black pop
Observed Moran's I Compared 999 Random Maps



Summary: Spatial Patterns

- Tract level changes seem unbiased and normally distributed.
- We see evidence that the noise infusion is not spatially random.
- This means that for some local communities and data users impacts of differential privacy are more pronounced than we'd expect.
- Needs more investigation.

Process and Products

Process: Questions

- Going forward what is the resolution process? What are the acceptance criteria for the final DP parameters?
- How will the Bureau interact with users? What is the plan for incorporating user feedback into the DP parameters? How does the Bureau plan to educate users about these data?

Products: Wish list

- More realizations of 2010 data (or synthetic data)
 - Difficult to make recommendations based on N of 1
 - Lots of people with FSRDC access willing and able to analyze more versions
- Off-spine allocation – importance of administrative units
 - It's possible to trace top-down path through hierarchy and include off-spine levels
- More invariants
 - Block-level total population has always been invariant
 - Empirical analysis of privacy loss when block-level total pop invariant?
- Uncertainty metric(s)
 - If you want users to adopt more robust statistical techniques, then they need these

Summary: Overall

- Aim is to understand the impact of the DP changes on the data we use to study places in the US.
- We examined multiple scales and spatial patterns.
- Our initial evaluations suggest more investigation of the “trade off” is warranted.
- There is a lot more to do.

Summary: Overall

- With this differential privacy proposal there is a tension between important public and private goods. Different people will assign different values to these goods.
- To understand how much of the public good (useable/useful data) we are trading for the private good, we need to evaluate the data.
- We need a participatory decision-making process.

Questions or feedback:

`vanriper@umn.edu`

`seth.spielman@colorado.edu`

All code and data:

`github.com/geoss/CNSTAT_DIFF_PRIVACY`