PennState
Health and Human Development

Differential Privacy and Mortality Rates in the United States

Alexis R. Santos-Lozada, PhD

Committee on National Statistics | Washington, DC, December 11-12, 2019

Introduction

- The U.S. Census is proposing the application of differential privacy to protect the privacy of respondents.

- Differential privacy "marks a sea change for the way that official statistics are produced and published" (Garkinkel, Abowd and Powazek, 2018).

- The U.S. Census has released 2010 Demonstration Data Products to help data users understand how differential privacy may or may not impact the data products they (we) are used to receiving.

PennState

Objective

- The purpose of this case study is to assess how differential privacy products impact the calculation of Mortality Rates in the United States, in comparison to the business as usual product.

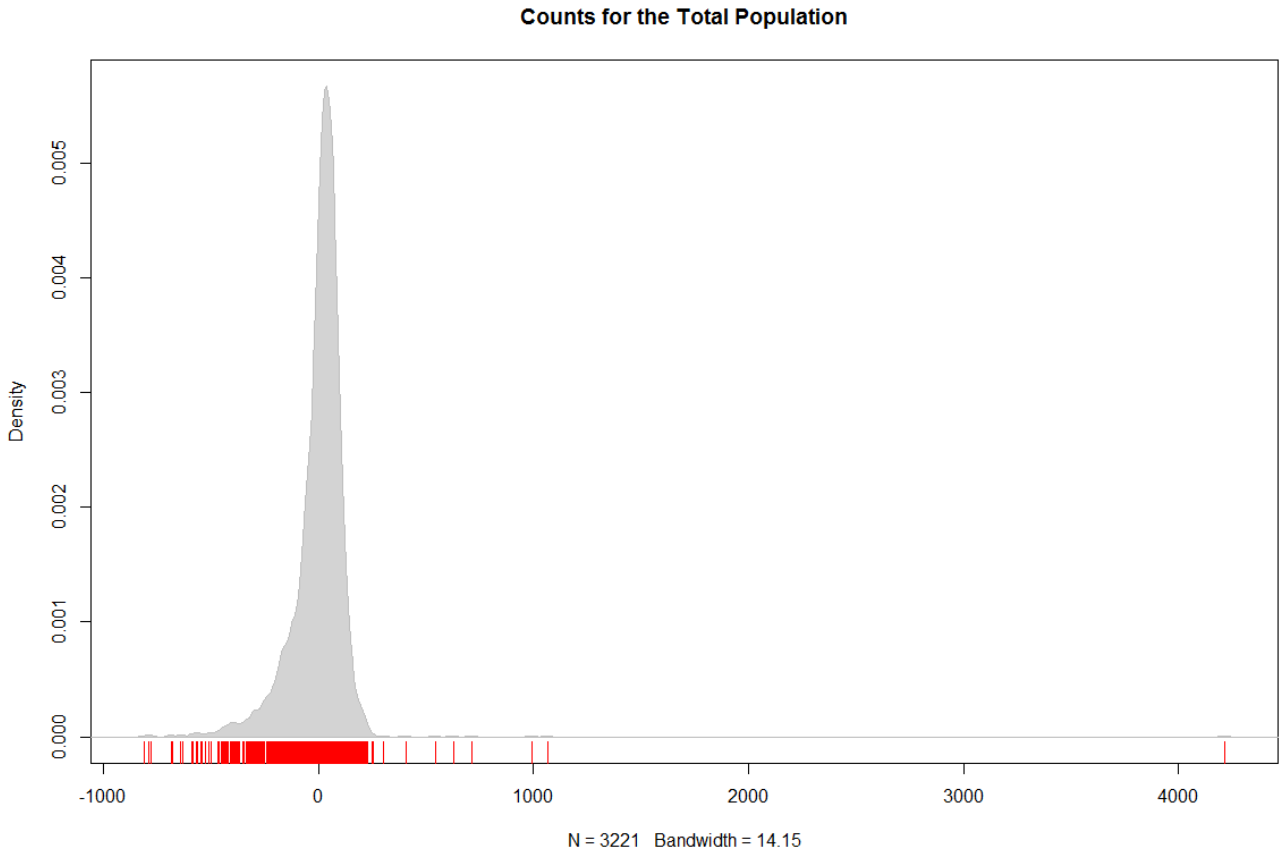
PennState

Data and Methods

- Data for this study come from:
  - CDC Wonder: Three year average for county-level death counts in 2009-2011. These will be our numerators.
  - 2010 Census: Total and racial/ethnic counts by county using the "business as usual" privacy disclosure protections. This will be the denominator for the first set of mortality rates. (M_1)
  - 2010 Demonstration Data: Same as for the previous by derived using the differential privacy algorithm. This will be the denominator for the second set of mortality rates (M_2)
  - ACS 2012 5-year file for regression analysis.
  - Rural Urban Continuum Codes to control for metro/non-metro.

PennState

Changes in the denominators



- First, I estimate mortality rates for every county in the U.S. using two different denominators
  - US Census 2010 tables
  - 2010 Demonstration Products
- Any difference in mortality rates is due to changes in the denominators.

PennState

Changes in county counts (denominators)

Group	Minimum	Maximum
Total Population	-853	4259
Non-Hispanic Whites	-321	227
Non-Hispanic Blacks	-147	289
Hispanics	-950	2,966

PennState

Changes in county percentages (denominators)

Group	Minimum	Maximum
Total Population	-2.70	695
Non-Hispanic Whites	-100	52.90
Non-Hispanic Blacks	-147.41	289.41
Hispanics	-77.25	27904.33

- Extreme values for Hispanics concentrated in small-areas:
  - Kalawao County, Hawaii
  - Rock County, Nebraska
  - Garfield County, New Hampshire

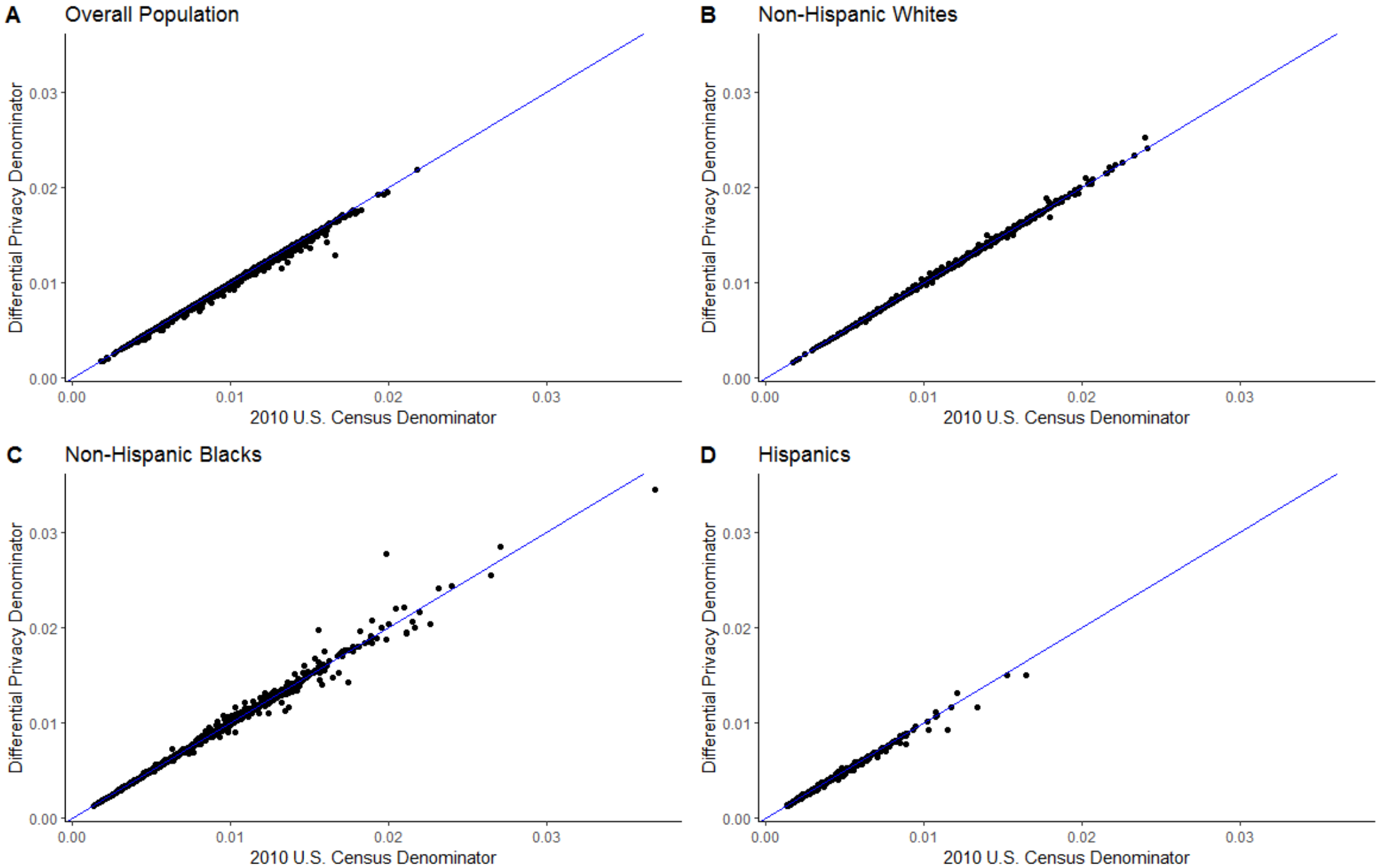
PennState

Comparison of Mortality Rates

- In the following section, I compare the mortality rates produced with two different denominators.

- Deviations from the blue line is a result of the changes in denominators.

PennState


A
Overall Population
B
Non-Hispanic Whites
C
Non-Hispanic Blacks
D
Hispanics
Differential Privacy Denominator
2010 U.S. Census Denominator
0.00
0.01
0.02
0.03

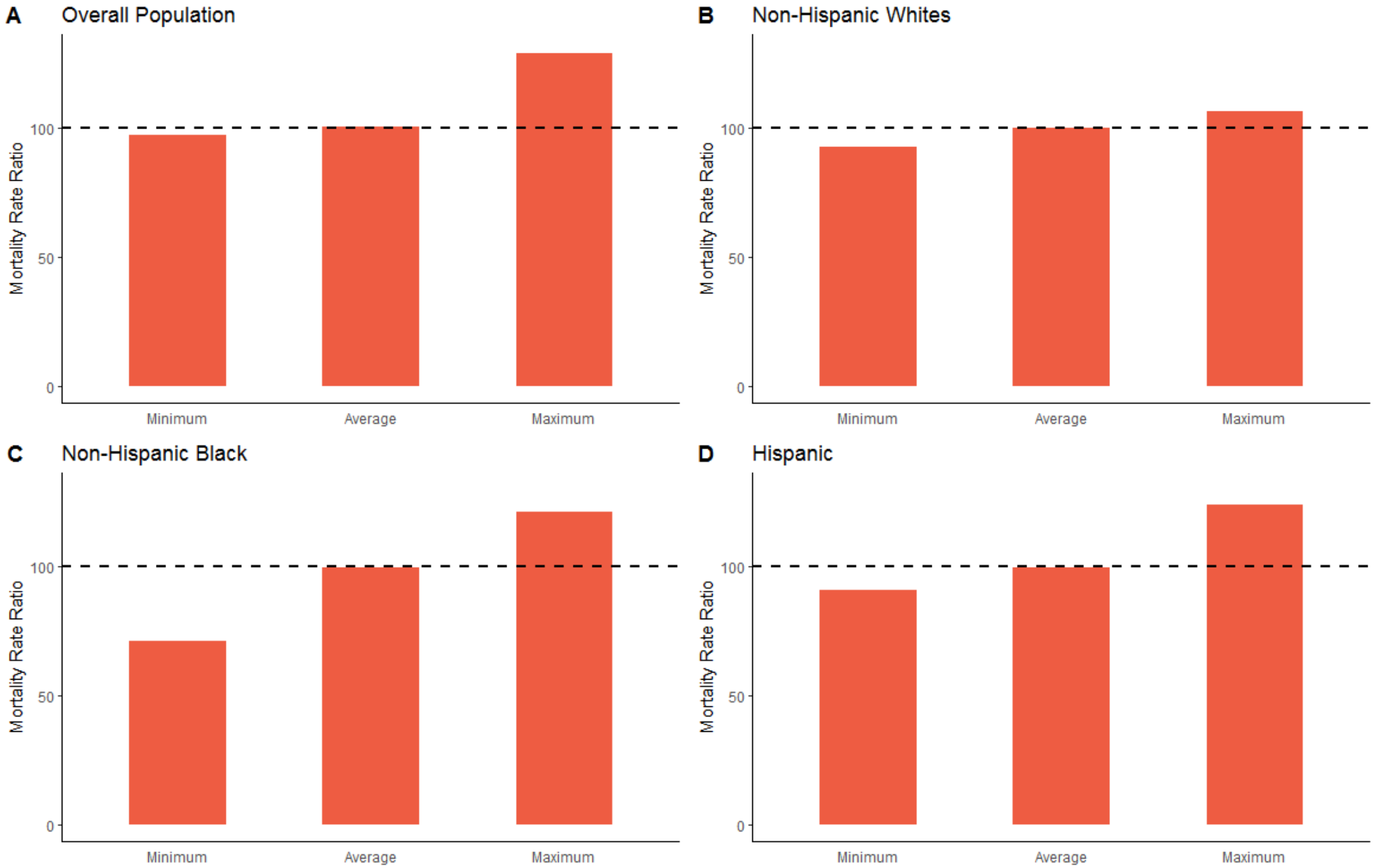
Mortality Rate Ratios

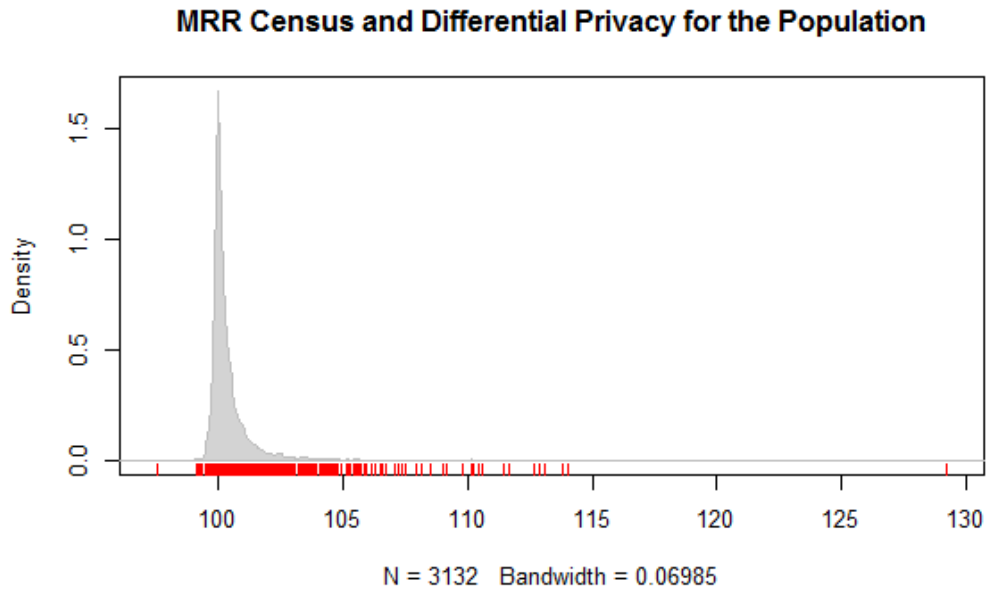
- In order to study the effect of these changes in denominators, I calculate the Mortality Rate Ratios using the following formula

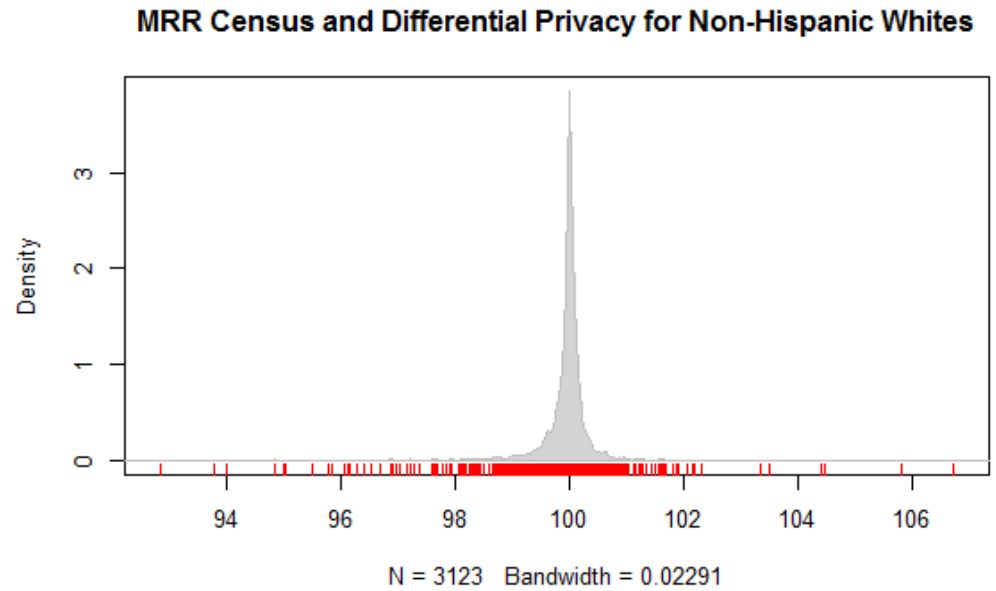
$$MRR = \frac{M_1}{M_2} * 100$$

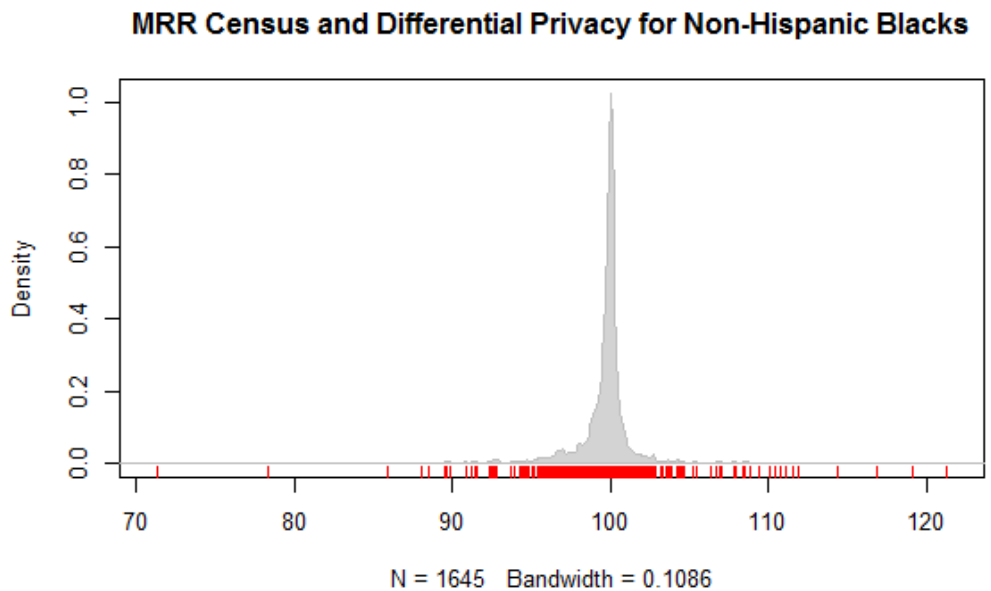
Where M_1 is the Mortality Rate using 2010 Census as the denominators and M_2 is the Mortality Rate using the 2010 Demonstration Products.

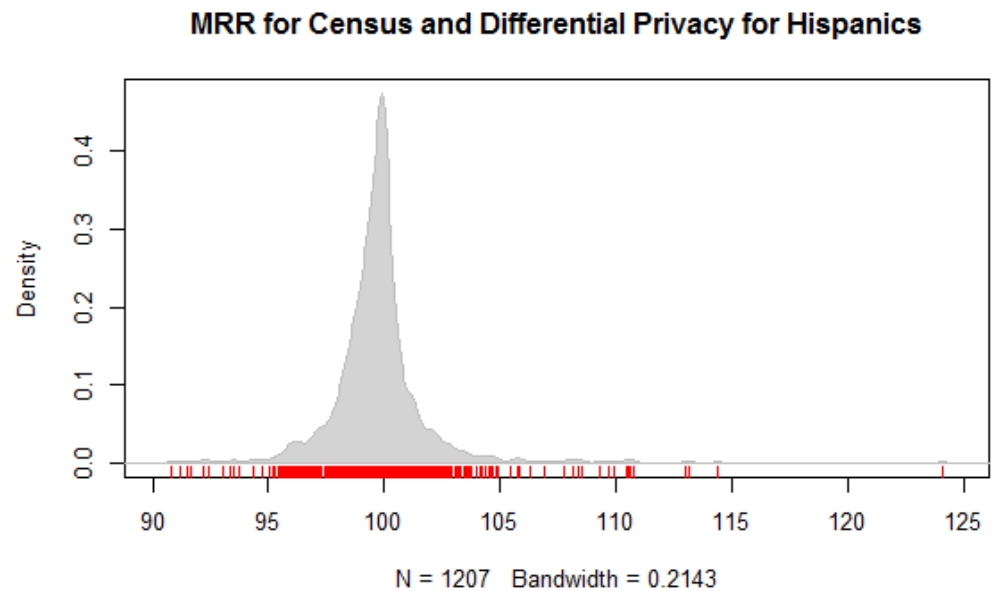
PennState

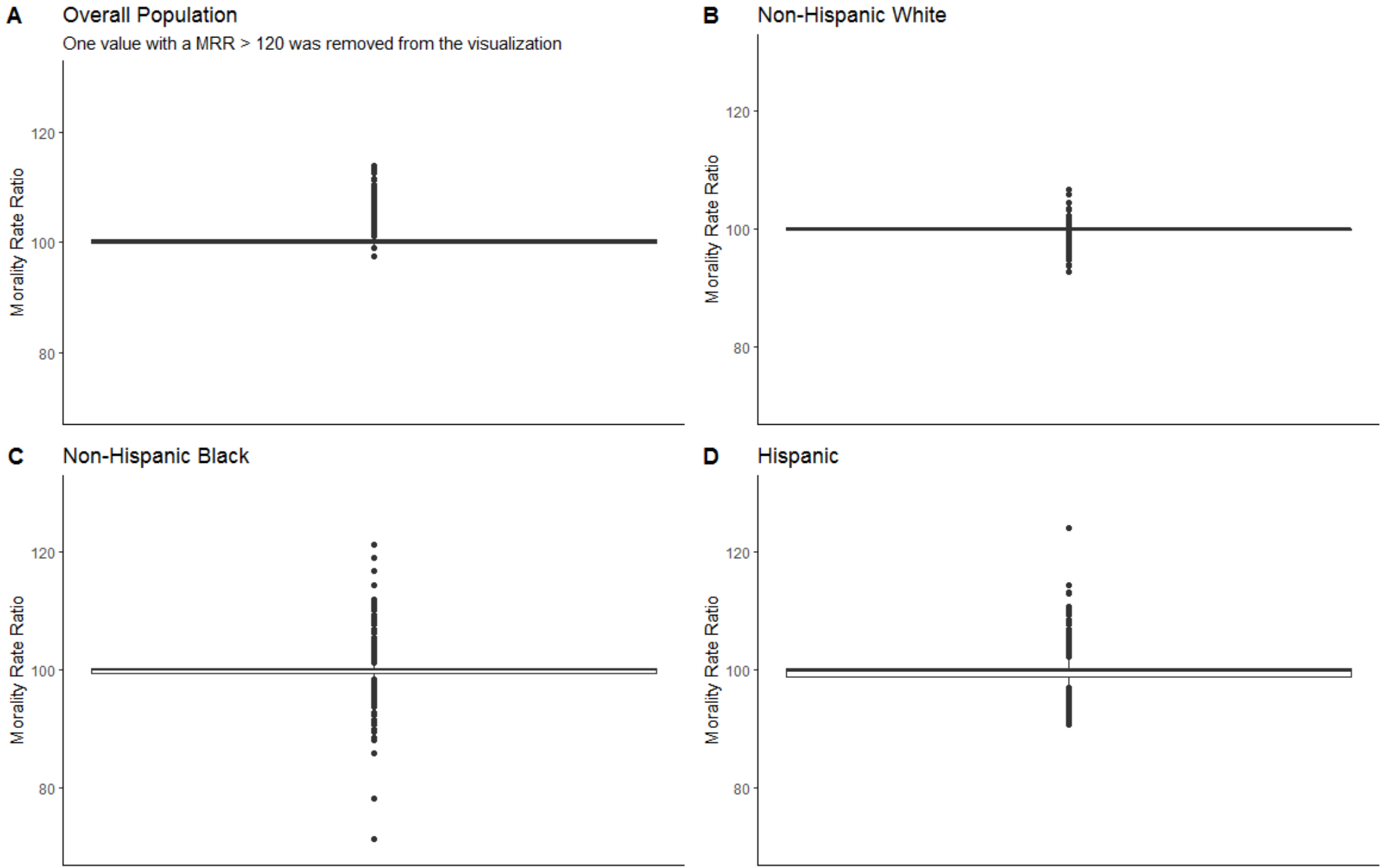

A
Overall Population
B
Non-Hispanic Whites
C
Non-Hispanic Black
D
Hispanic
Mortality Rate Ratio
100
50
0
Minimum
Average
Maximum

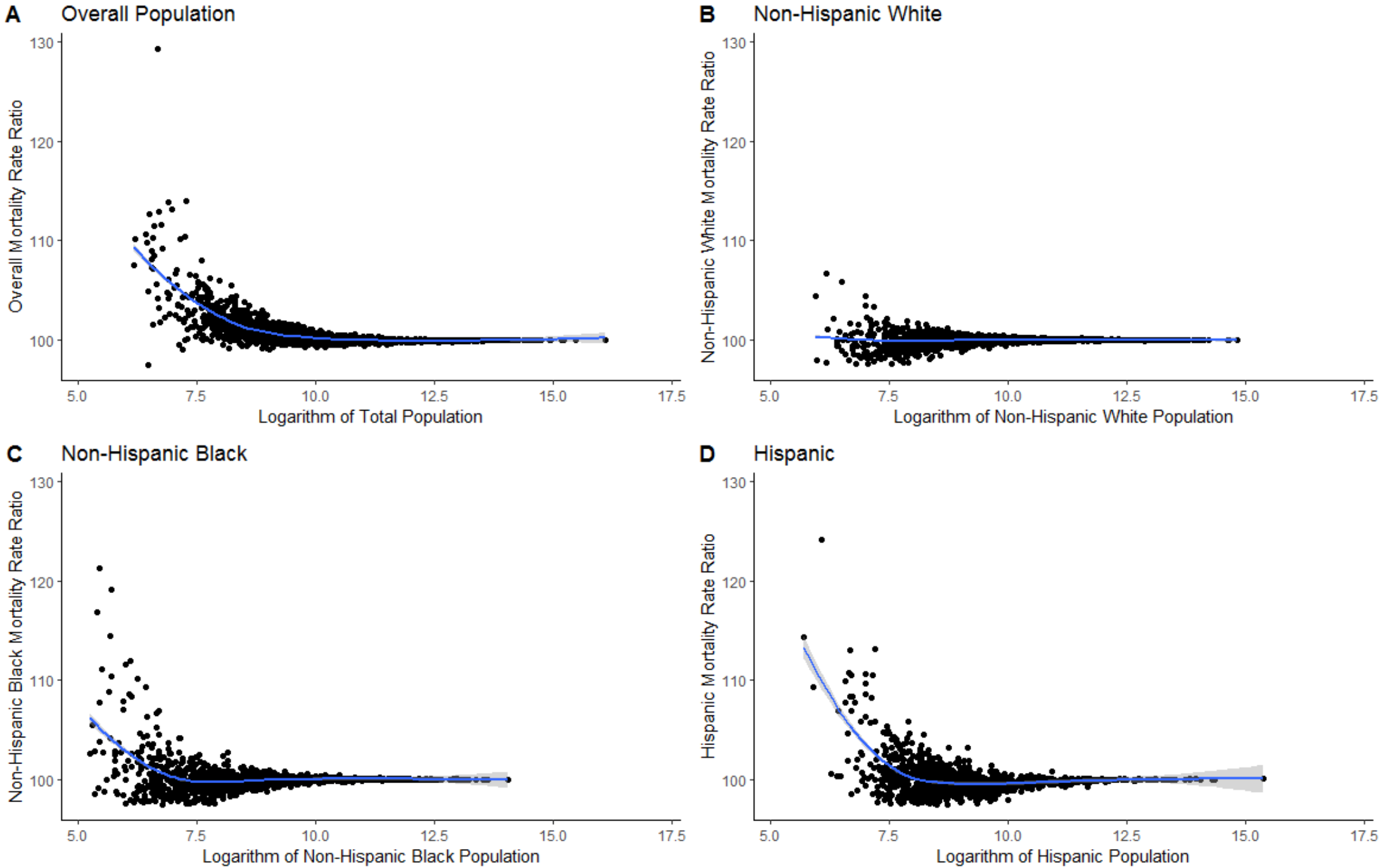

MRR Census and Differential Privacy for the Population
Density
0.0
0.5
1.0
1.5
100
105
110
115
120
125
130
N = 3132 Bandwidth = 0.06985

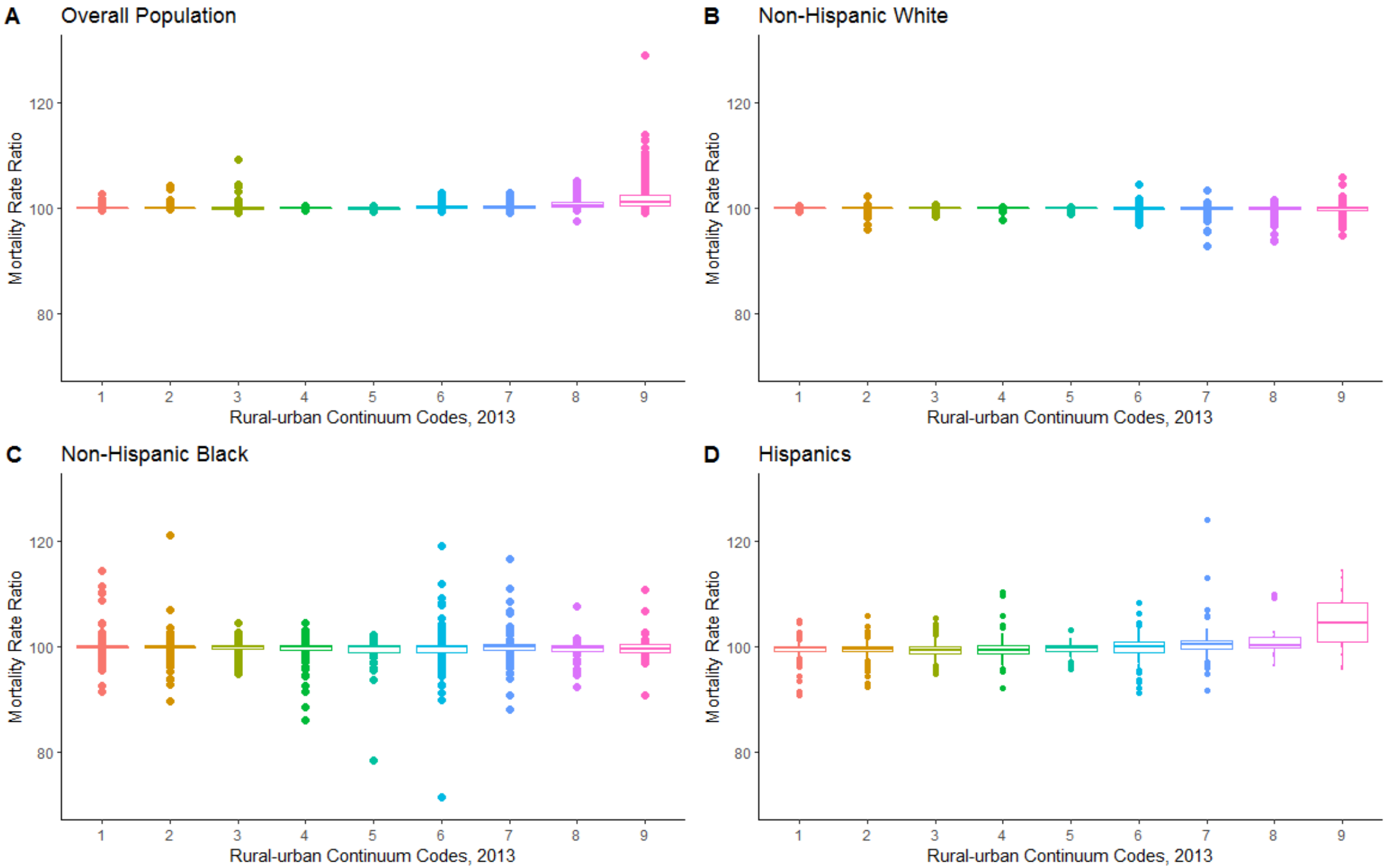

MRR Census and Differential Privacy for Non-Hispanic Whites
Density
0
1
2
3
94
96
98
100
102
104
106
N = 3123 Bandwidth = 0.02291


MRR Census and Differential Privacy for Non-Hispanic Blacks
Density
0.0
0.2
0.4
0.6
0.8
1.0
70
80
90
100
110
120
N = 1645 Bandwidth = 0.1086


MRR for Census and Differential Privacy for Hispanics
Density
0.0
0.1
0.2
0.3
0.4
90
95
100
105
110
115
120
125
N = 1207 Bandwidth = 0.2143

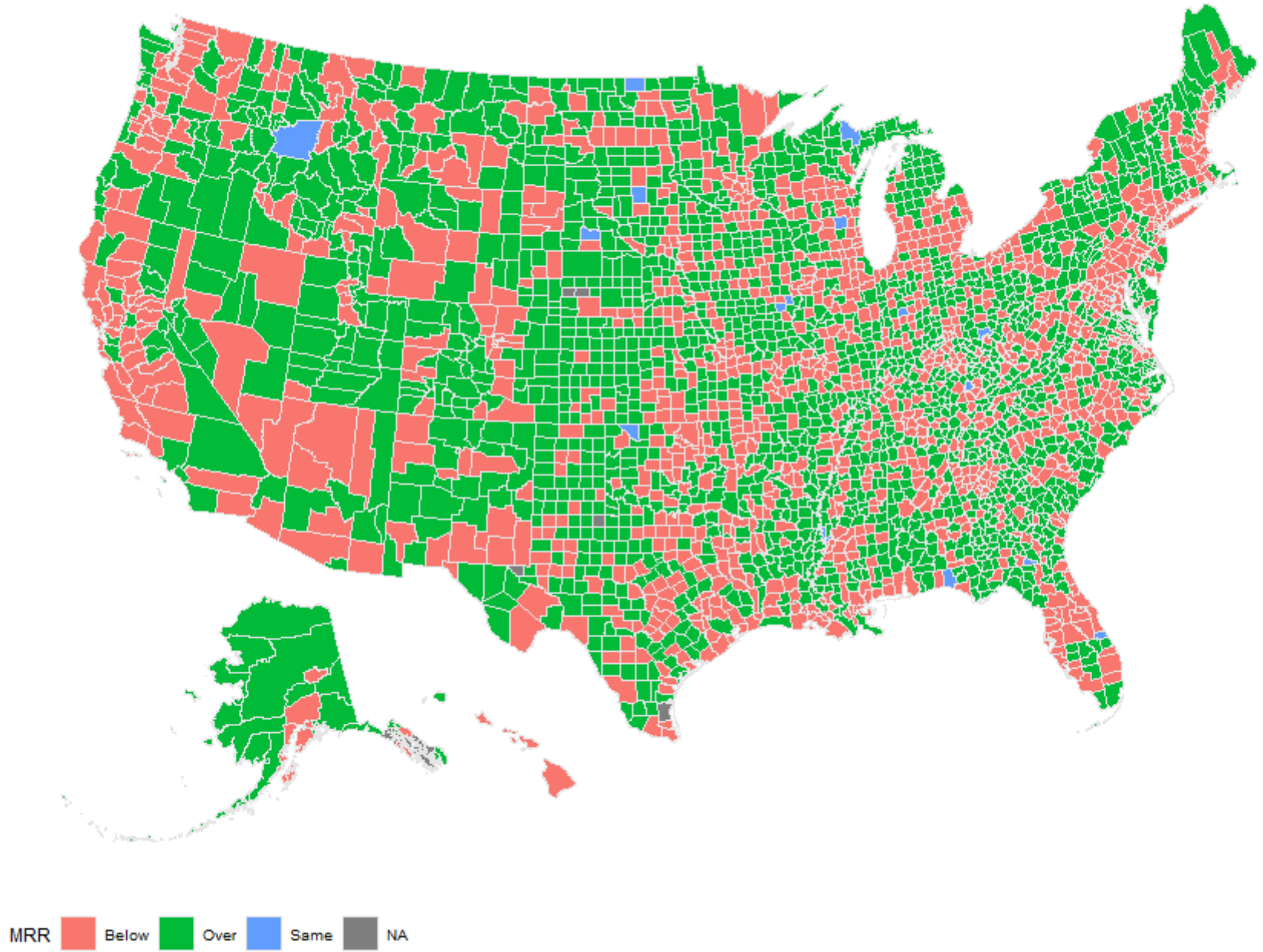

A
Overall Population
One value with a MRR > 120 was removed from the visualization
Morality Rate Ratio
120
100
80
B
Non-Hispanic White
Morality Rate Ratio
120
100
80
C
Non-Hispanic Black
Morality Rate Ratio
120
100
80
D
Hispanic
Morality Rate Ratio
120
100
80


A
Overall Population
Overall Mortality Rate Ratio
130
120
110
100
5.0
7.5
10.0
12.5
15.0
17.5
Logarithm of Total Population
B
Non-Hispanic White
Non-Hispanic White Mortality Rate Ratio
130
120
110
100
5.0
7.5
10.0
12.5
15.0
17.5
Logarithm of Non-Hispanic White Population
C
Non-Hispanic Black
Non-Hispanic Black Mortality Rate Ratio
130
120
110
100
5.0
7.5
10.0
12.5
15.0
17.5
Logarithm of Non-Hispanic Black Population
D
Hispanic
Hispanic Mortality Rate Ratio
130
120
110
100
5.0
7.5
10.0
12.5
15.0
17.5
Logarithm of Hispanic Population


A
Overall Population
B
Non-Hispanic White
C
Non-Hispanic Black
D
Hispanics
Mortality Rate Ratio
120
100
80
1
2
3
4
5
6
7
8
9
Rural-urban Continuum Codes, 2013

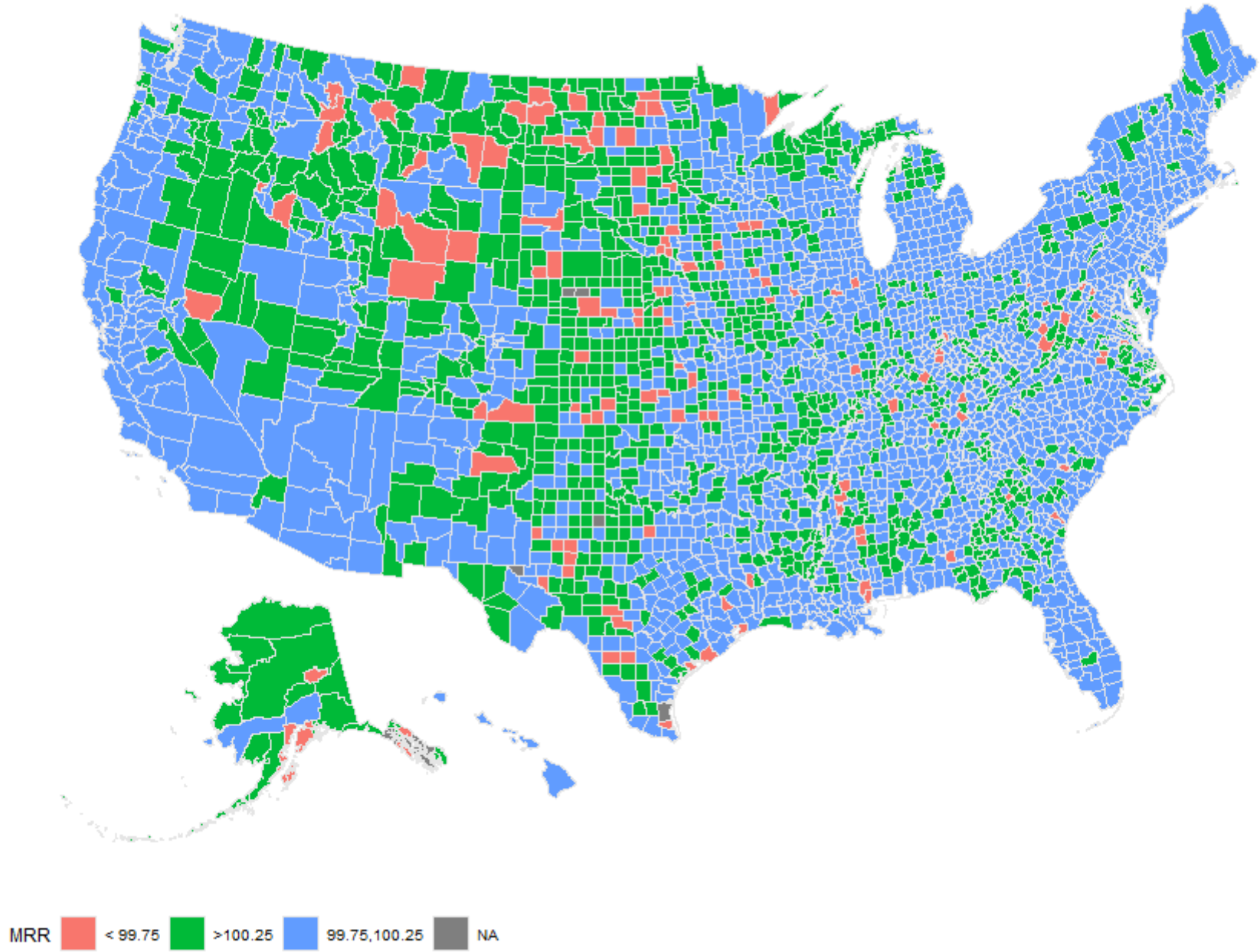
1,2,3 = Metro

2010 Mortality Rate Ratios using two denomiators



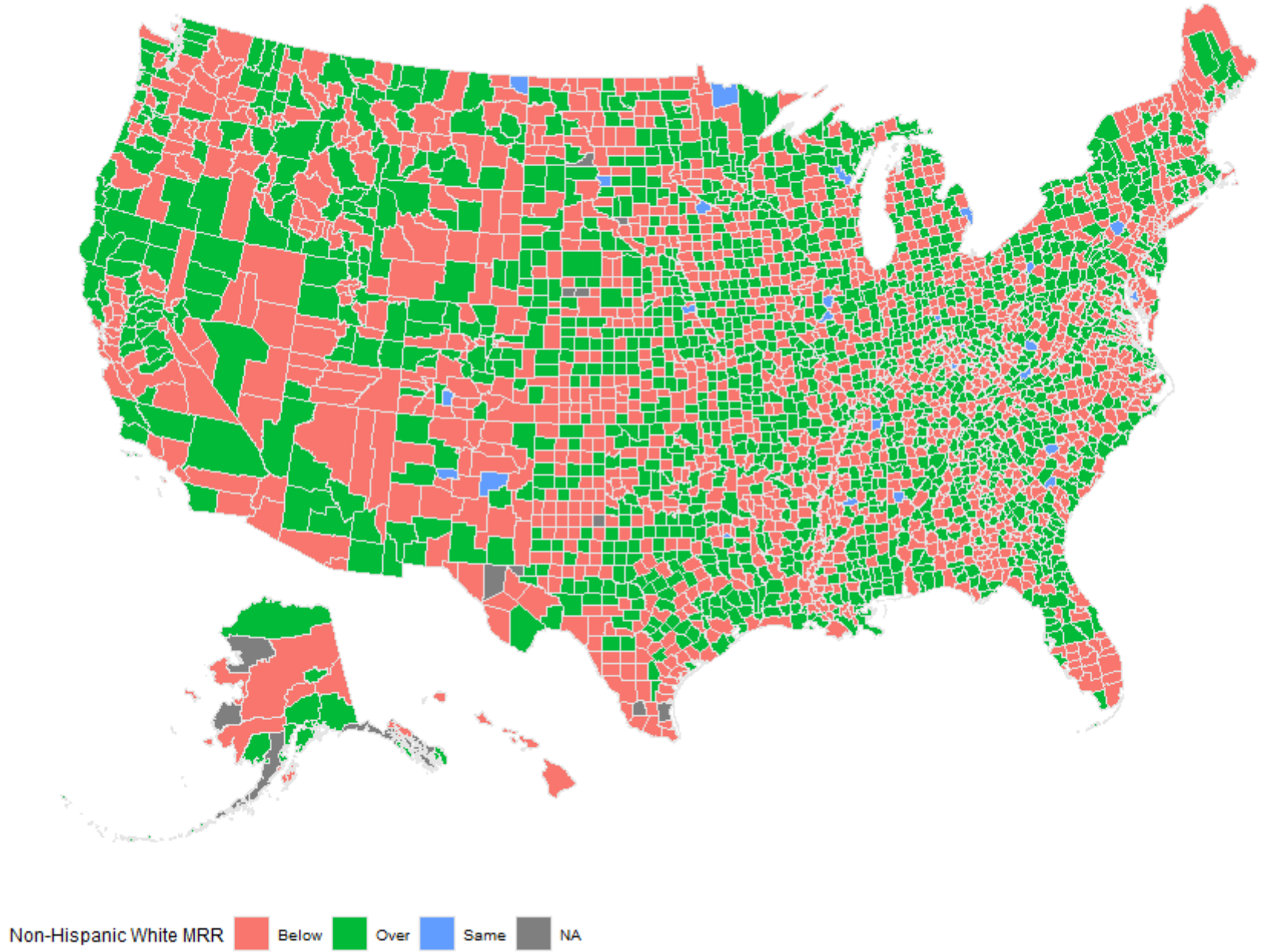
Produced by Alexis R. Santos @appdemography

2010 Mortality Rate Ratios using two denomiators, by level of accuracy



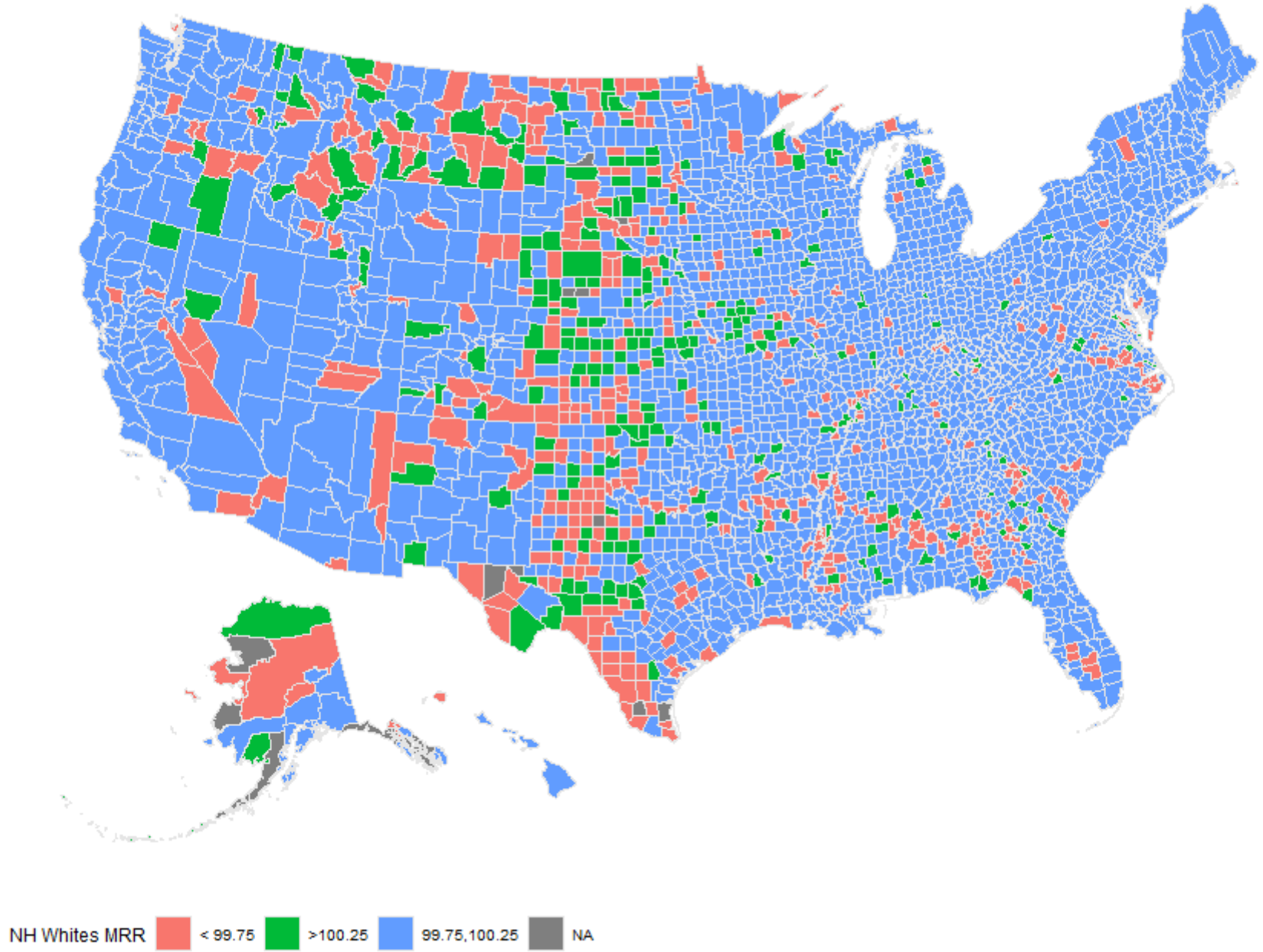
Produced by Alexis R. Santos @appdemography

2010 Mortality Rate Ratios for Non-Hispanic Whites using two denomiators



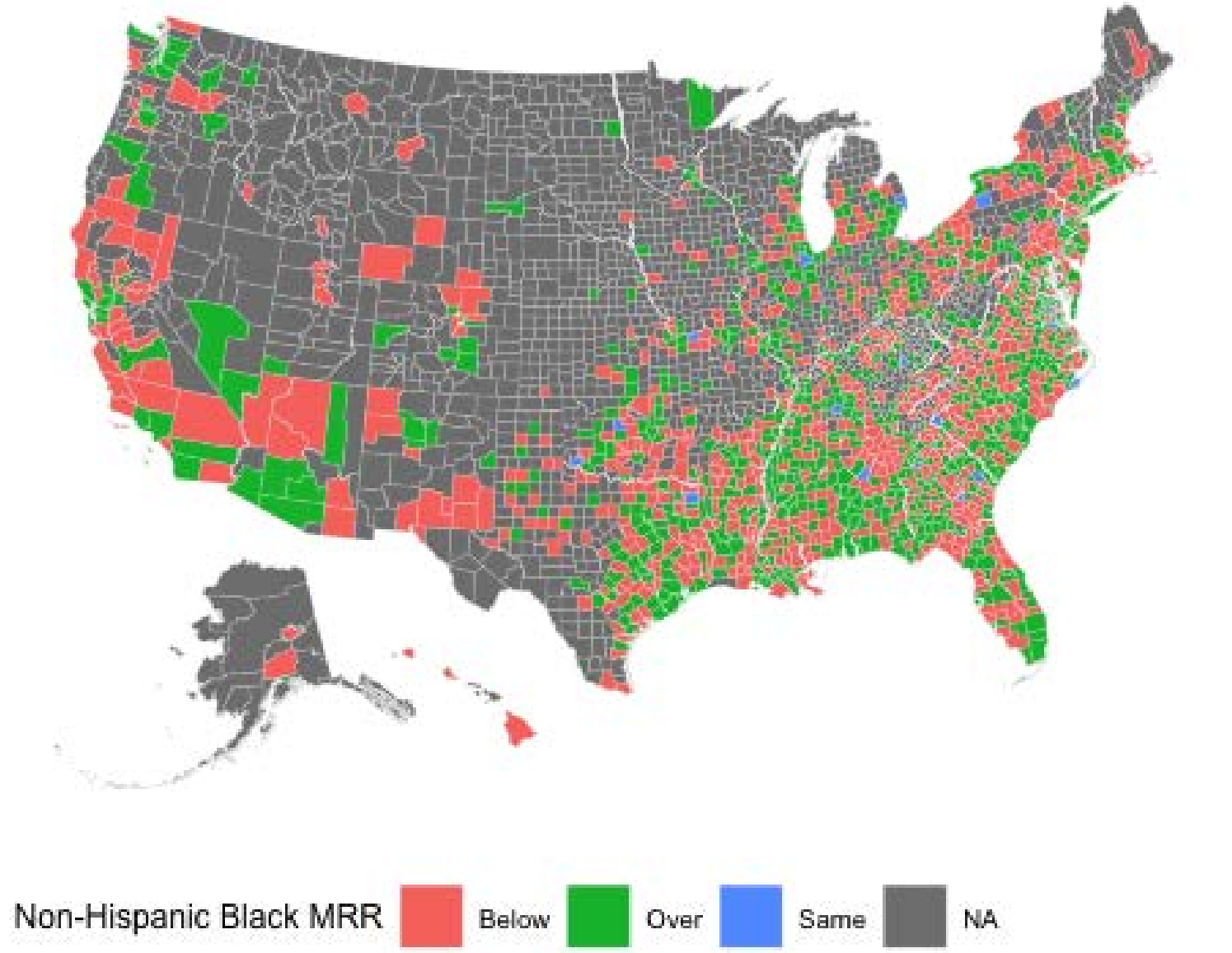
Produced by Alexis R. Santos @appdemography

2010 Mortality Rate Ratios using two denomiators for NH-Whites, by level of accuracy

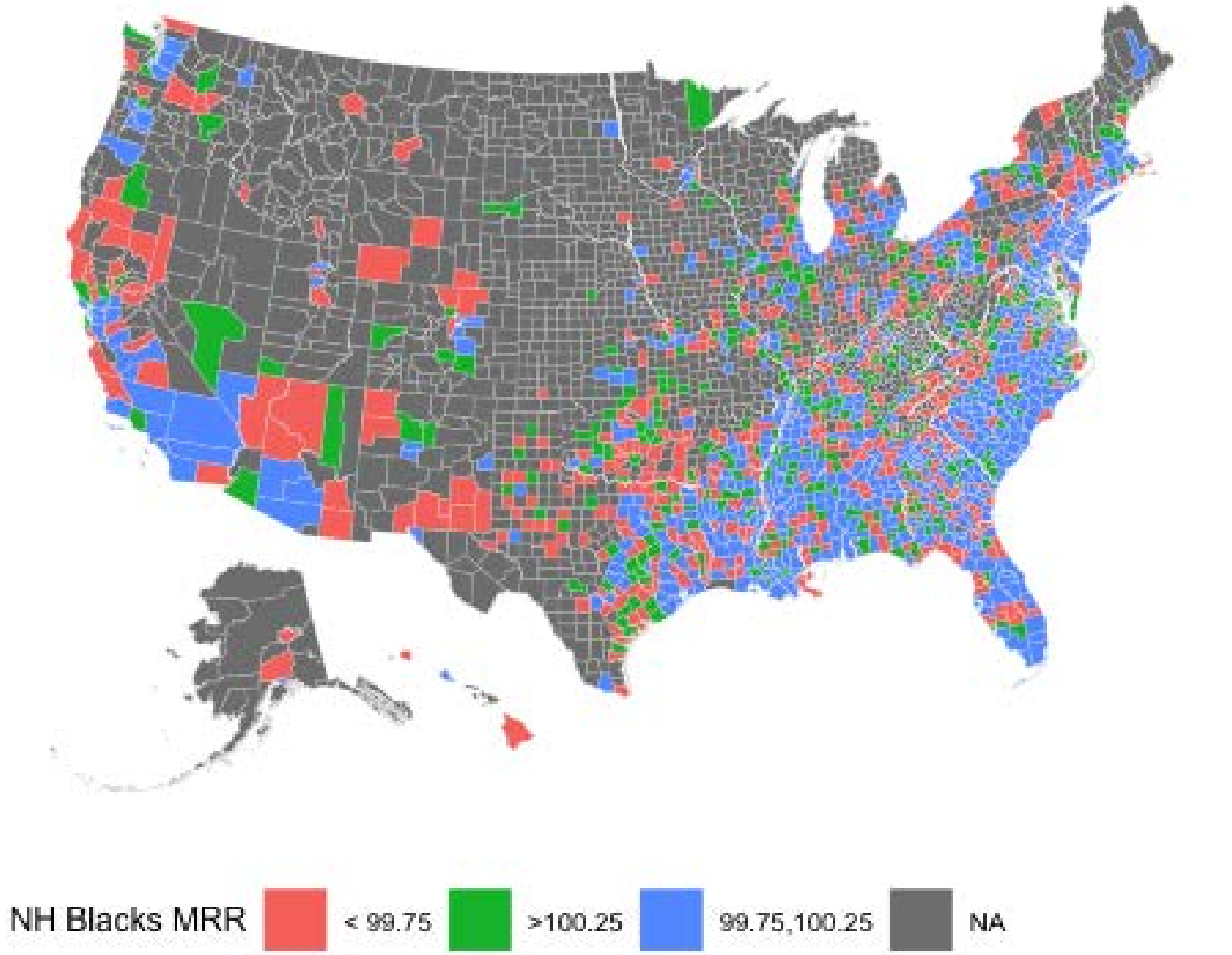


Produced by Alexis R. Santos @appdemography

2010 Mortality Rate Ratios for NH Black using two denomiators

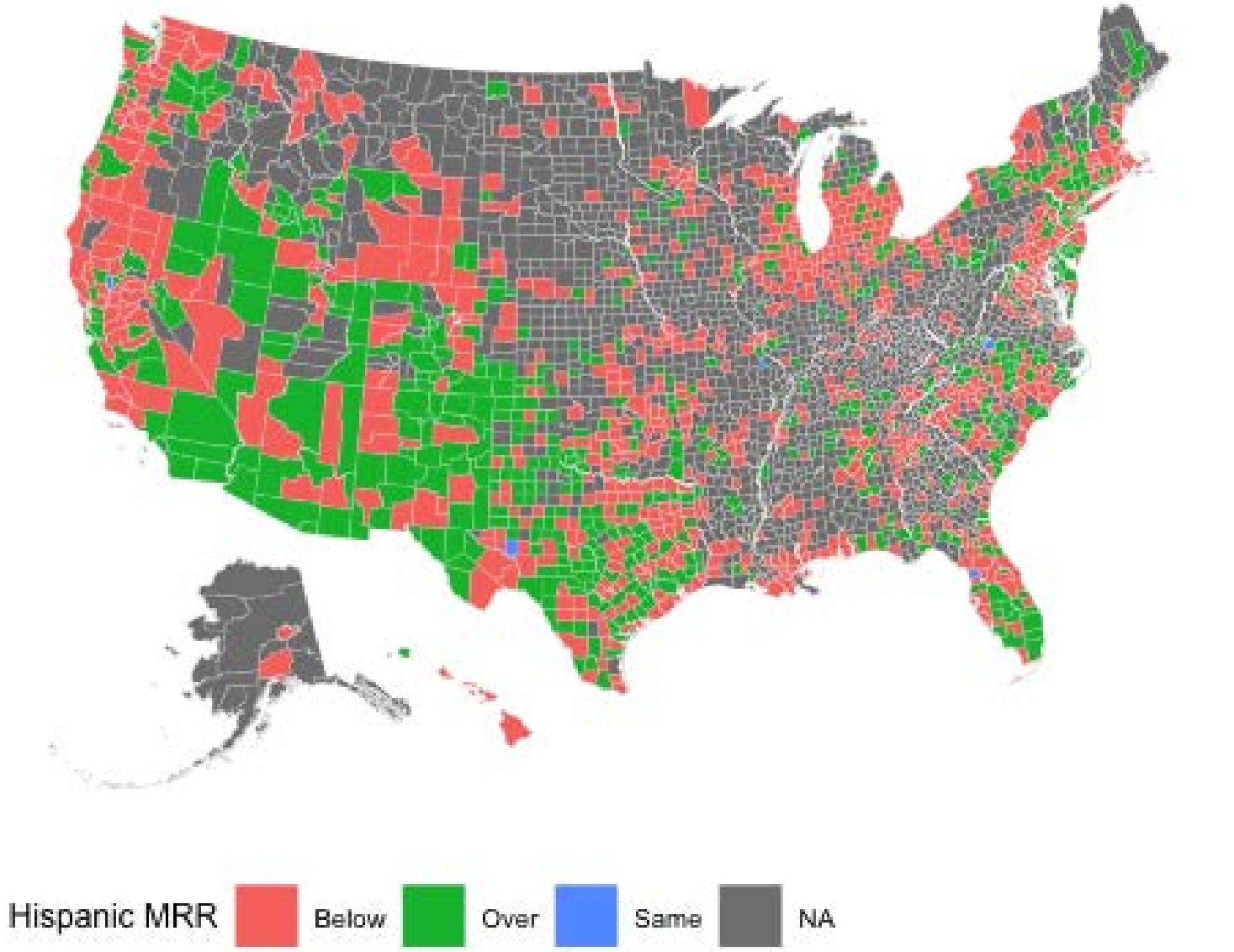


2010 Mortality Rate Ratios using two denomiators for NH-Blacks



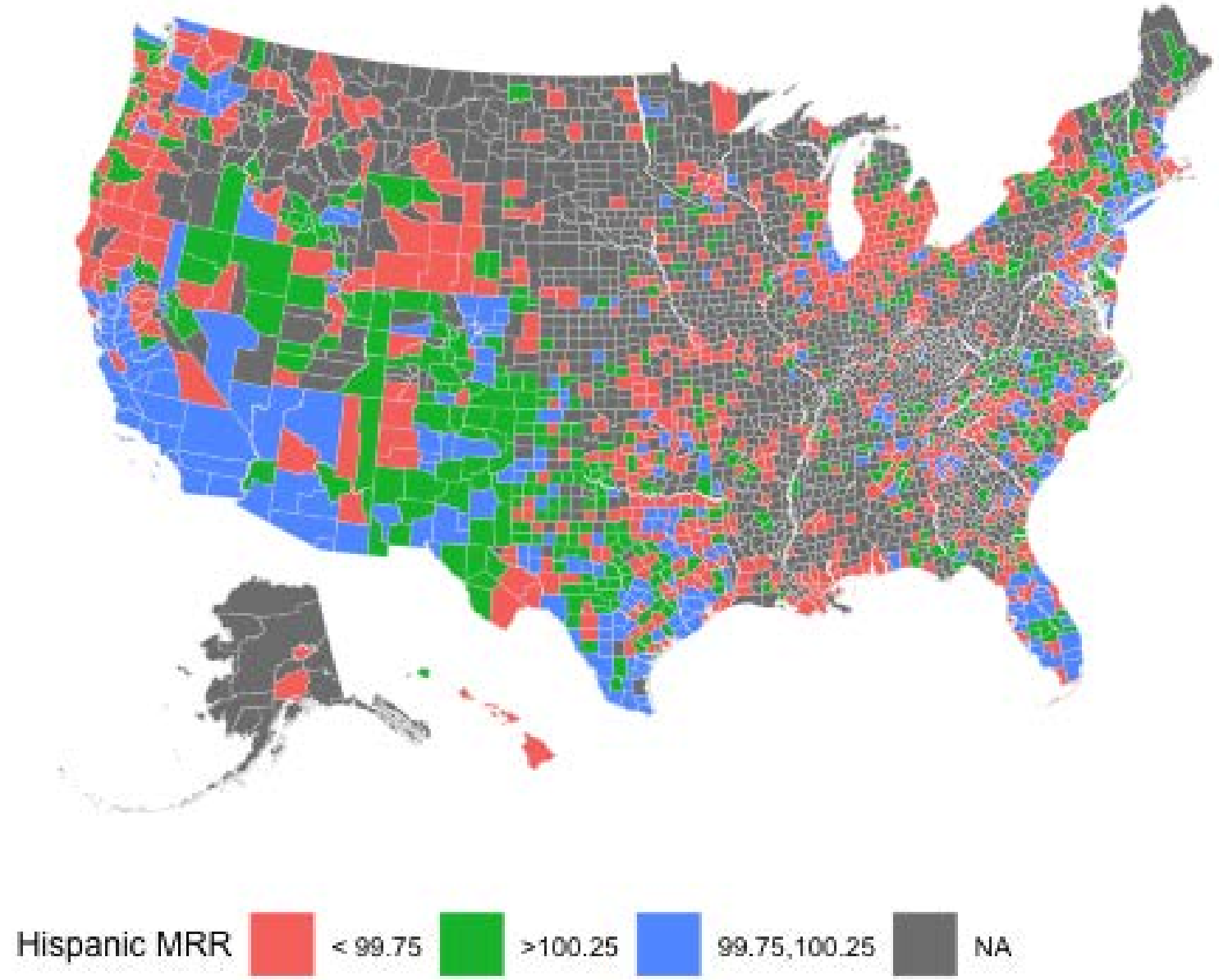
Produced by Alexis R. Santos @appdemography

2010 Mortality Rate Ratios for Hispanic using two denomiators



Produced by Alexis R. Santos @appdemography

2010 Mortality Rate Ratios using two denomiators for Hispanics



Produced by Alexis R. Santos @appdemography

What are the correlates of these differences in Mortality Rates?

- I obtain covariates of socioeconomic and demographic characteristics from the ACS to study the association between these characteristics and MRR.

- Coefficients are calculates using an OLS model while controlling for the following characteristics:

  - Metro/Non-Metro (USDA), Total Population, Percent in Poverty, Percent Black, Percent Hispanic, Percent under 18 years, Percent 65 and older, % Immobile, Percent less than HS, South, Percent Unemployed, Percent Female Employment and Percent of Female Headed Households. (Litcher and McLaughlin, 1995).

PennState

Associations between sociodemographic characteristics and then MRR

- A higher proportion of non-Hispanic Blacks, Hispanics, and persons living below the threshold is associated with higher MRR.

- A higher proportions of population below 18 years and over 65 year have higher MRR.

- Higher proportions of unemployed and immobile (have not migrated in the last year) and being a southern county are characteristics associated with lower MRR.

PennState

Conclusions

- The implementation of differential privacy introduces variation in the denominators used to construct mortality rates.
- The effect is larger for small-areas, and for racial/ethnic minorities.
- Several key socioeconomic characteristics of interest to demographers are associated with increased levels of variation in the overall mortality rates.

PennState

Implications

- Reduced understanding of the demographic profile of certain areas of the nation (i.e. rural America, emerging destinations).

- Reduced understanding of health disparities for areas where we are not doing so well.

PennState

Recommendation

- Controlling the population counts at smaller geographies for total and race/ethnicity.

- Postponing the implementation of differential privacy in the 2020 Census products until we can:

  - Complete additional case studies that require data at smaller geographies.

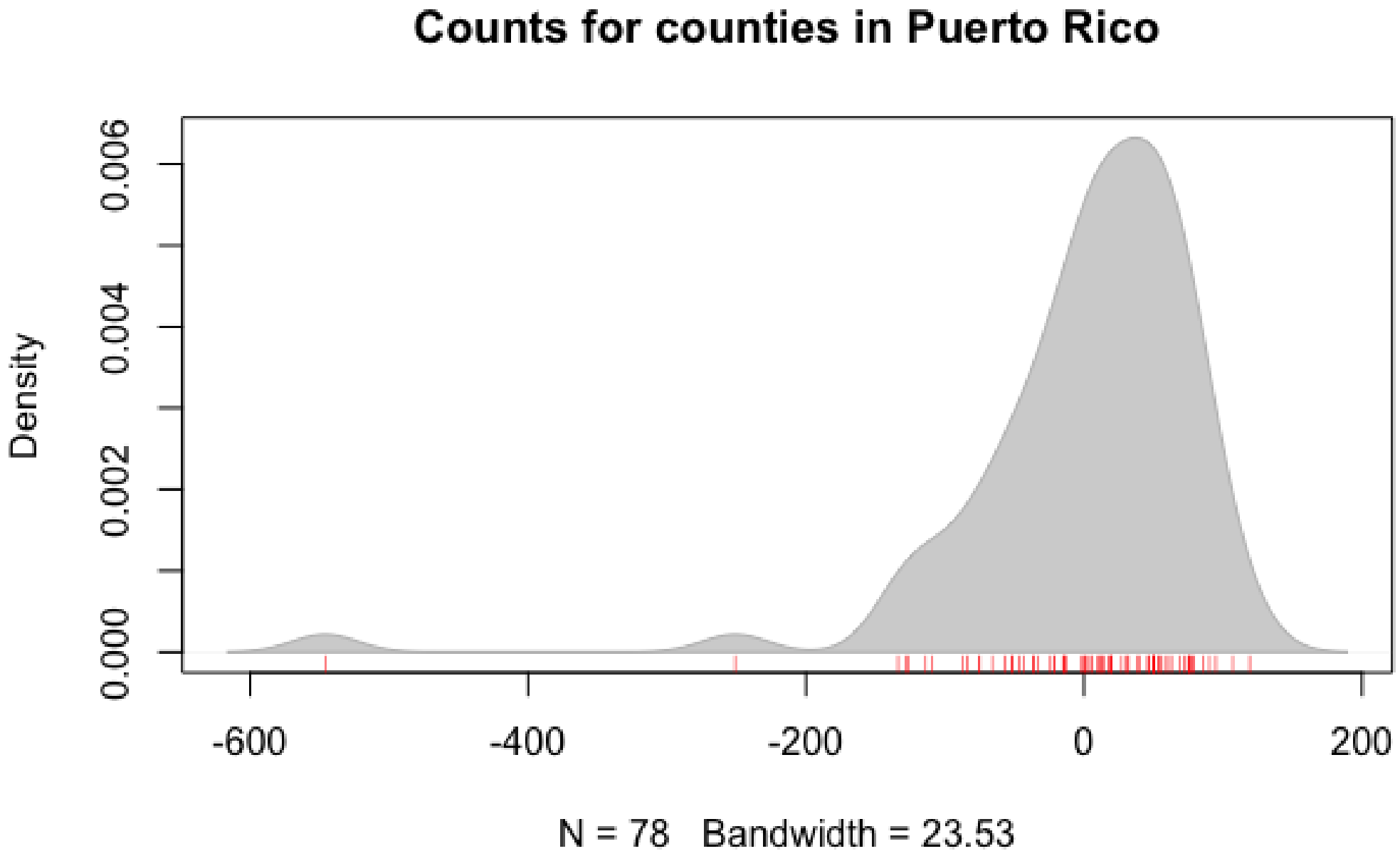
  - Determine the extent to which DP will limit our ability to work with data users and produce accurate information for policy purposes.

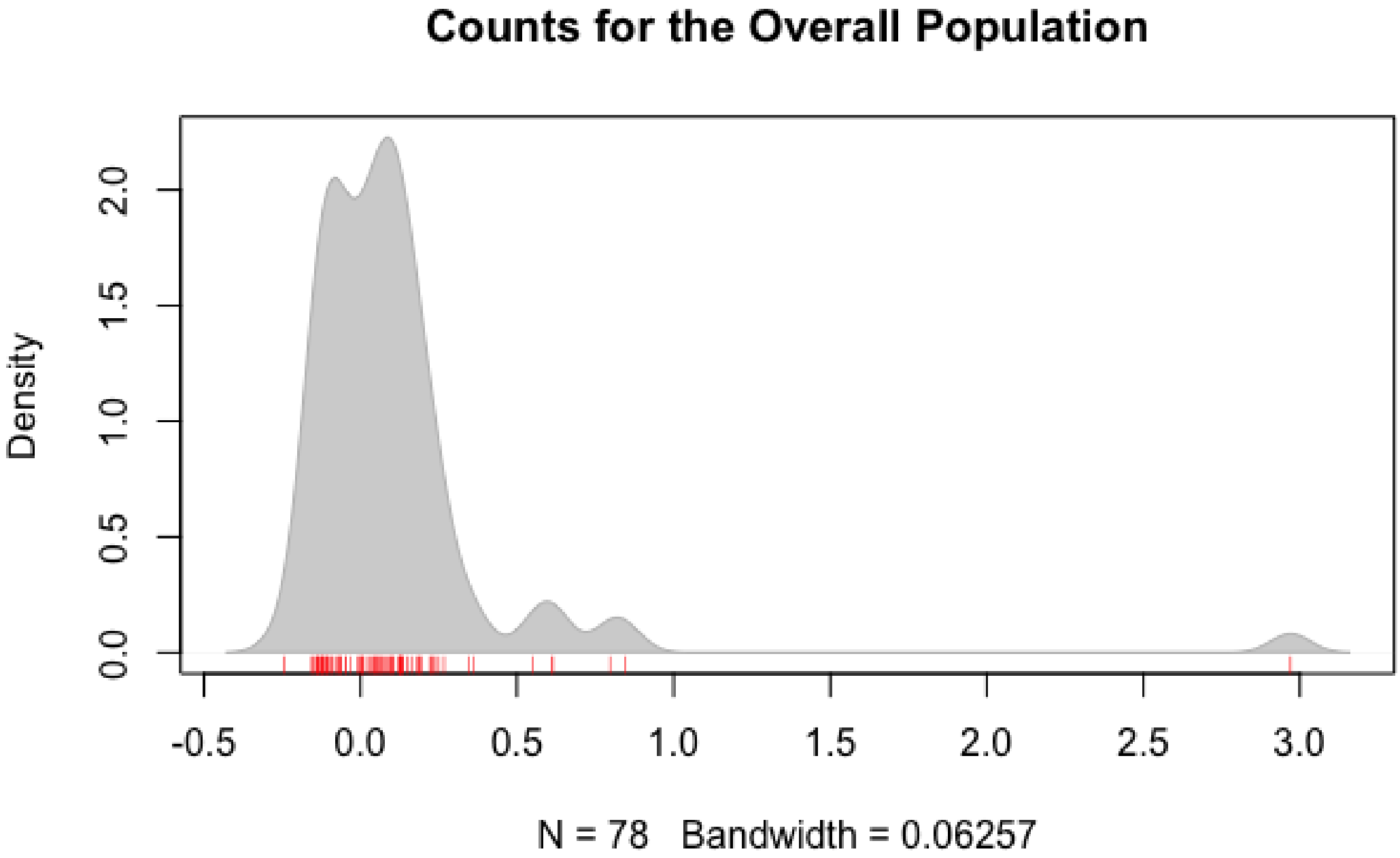
PennState

Thank you!

Any questions?

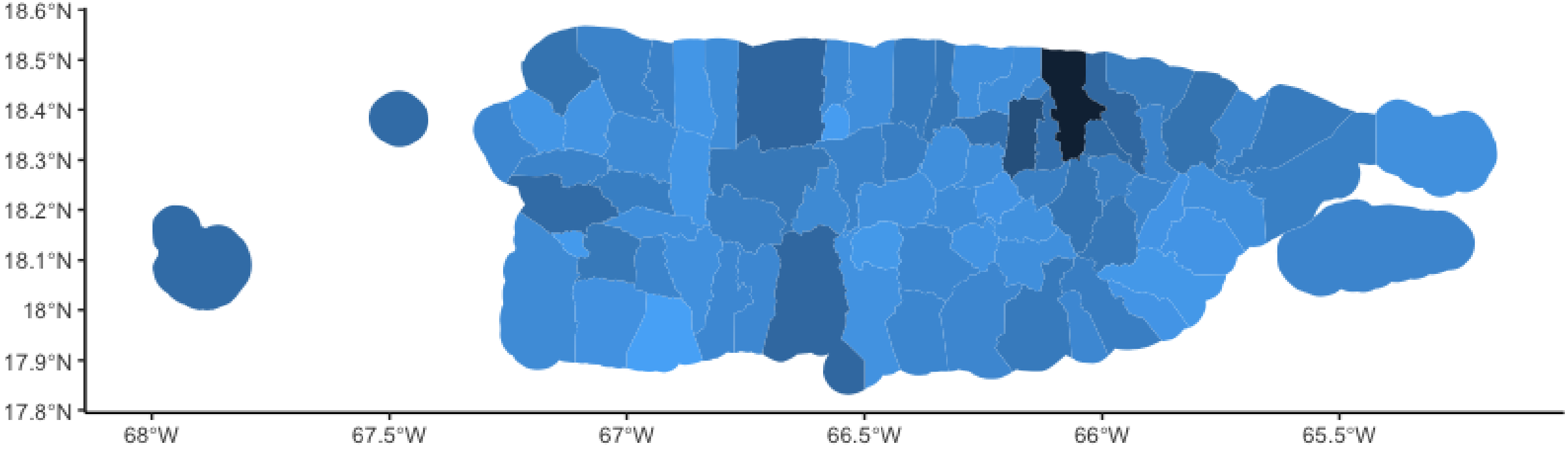
What about Puerto Rico?


Counts for counties in Puerto Rico
Density
0.000
0.002
0.004
0.006
-600
-400
-200
0
200
N = 78 Bandwidth = 23.53


Counts for the Overall Population
Density
0.0
0.5
1.0
1.5
2.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
N = 78 Bandwidth = 0.06257

Change from Census to DP

Demonstration Products



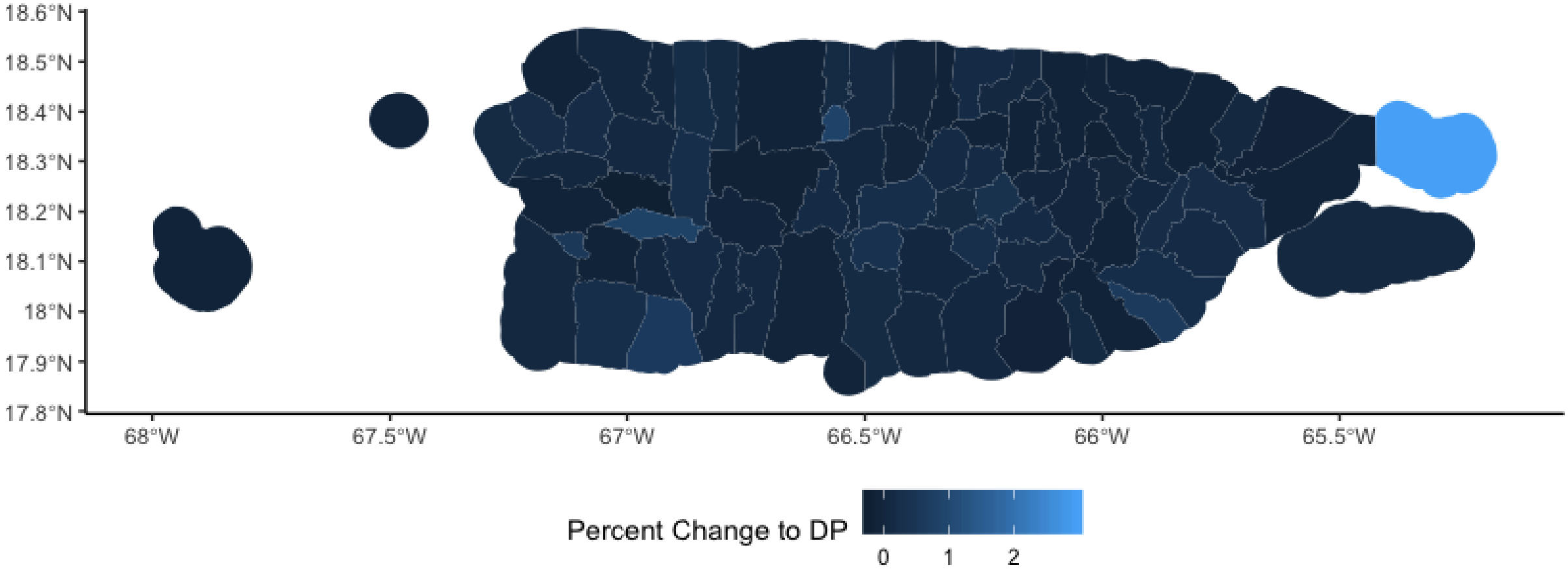
Change to DP

-400 -200 0

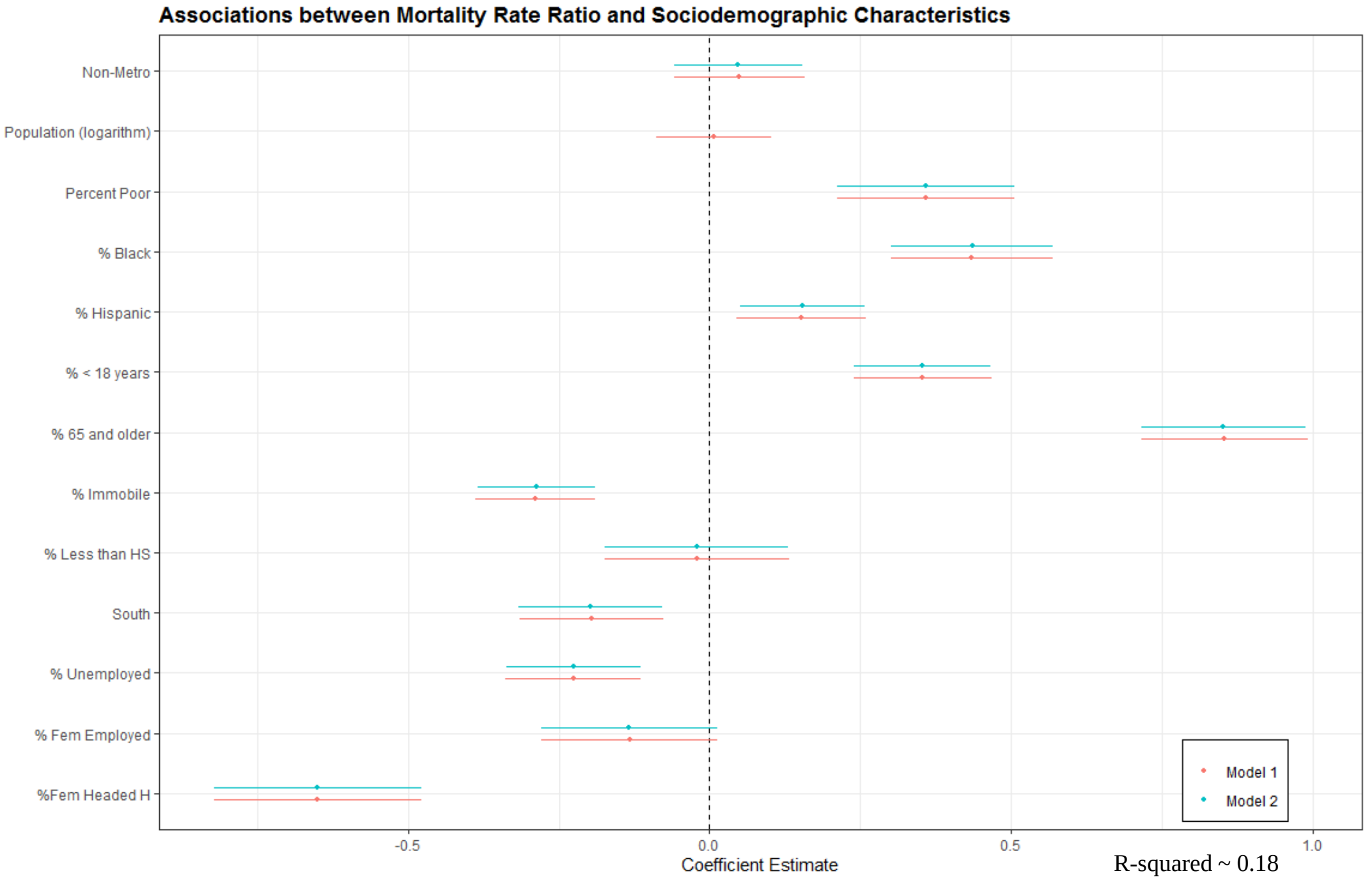
Produced by Alexis Santos @appdemography

Percent Change from Census to DP

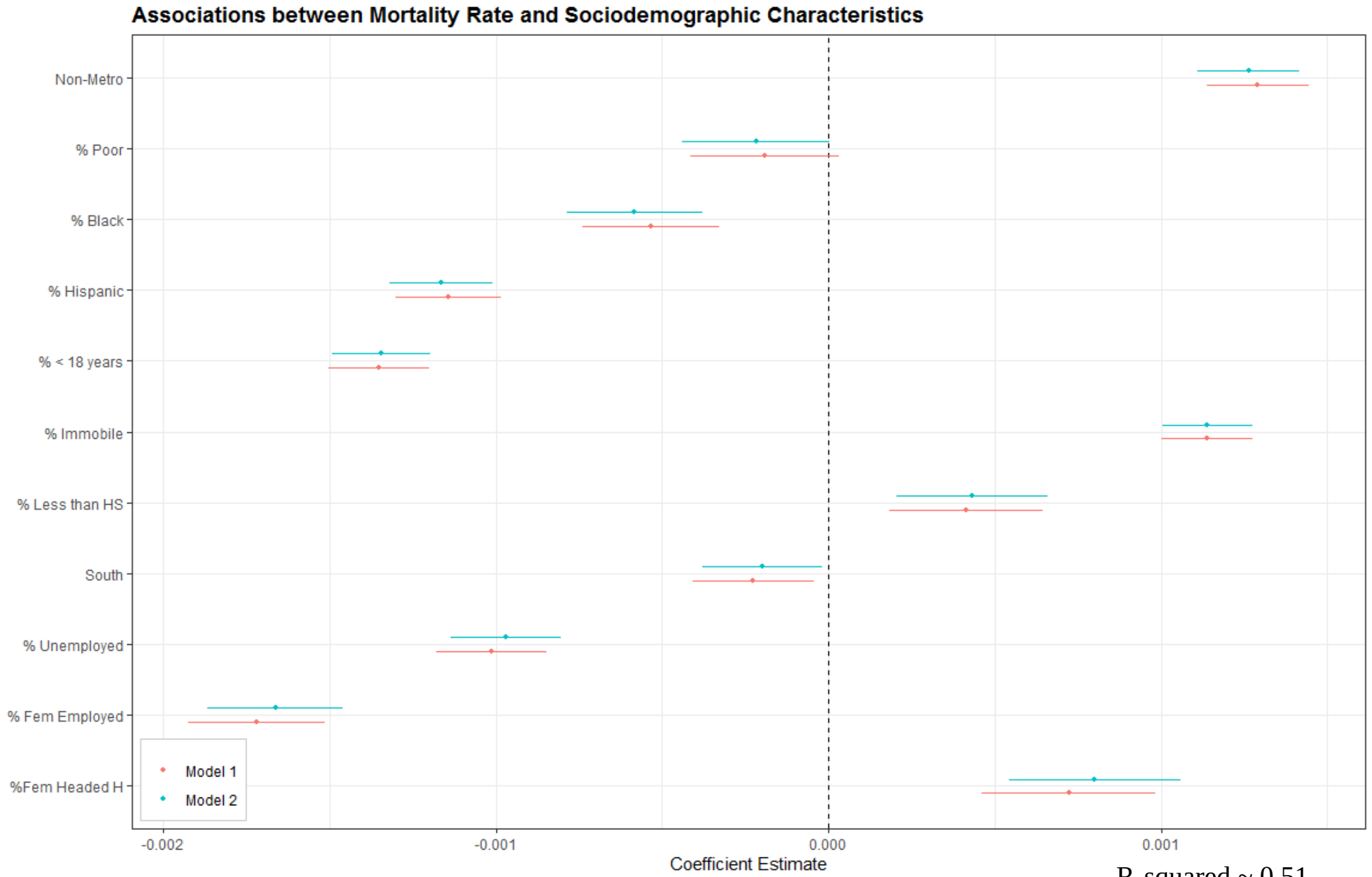
Demonstration Products



Produced by Alexis Santos @appdemography



R-squared ~ 0.18



R-squared ~ 0.51

R-squared ~ 0.71