Analysis of Race, Region, and Income as a Factor of Ability to Pay for Medical Bills

AUTHOR

Lessly Ortiz and Sophia Silovsky

Abstract

The purpose of our analysis is to see if there exists a combined effect of income, race, and region on a person's ability to pay or having issues paying medical bills in the last 12 months. This project used data from the IPUMS National Health Interview Series from 2020. We fit three different models before deciding on one logistic regression model, where the model states that the effect of the region will change across racial groups, when controlling for the simultaneous additive effect of family income. We found evidence supporting some of our hypothesis, in terms of income being a significant predictor across all regions and racial groups, and the combination of the white group in the Northeast being a significant baseline. Additionally, we found that certain regions and race groups have significant differences from the baseline average. Specifically, the South region, Black/African American group, Asian group, and the combination of the Black/African American group in the North Central / Midwest had a significant difference from the baseline. However, we cannot completely accept or reject our hypothesis.

Introduction

The U.S. has shown a significant growth in economic disparities between different regions in recent decades. One research paper shows that as recently as 1980, "large sections of the country had incomes roughly in line with the national average", but within 33 years "the country had bifurcated into a handful of thriving metro areas and a much larger set of places with far lower incomes." [1] Moreover, healthcare access and care in the United States is not equal across citizens, plain and simple. It has already been proven by researchers that those with higher socioeconomic statuses have fewer health problems. In fact, those with a lower socioeconomic status are more susceptible to certain health conditions. [2] Additionally, a person's issues with accessing proper healthcare seems to have a relationship with their race. Research prior to 2020 shows that "adult uninsured rates and racial and ethnic coverage inequities declined in almost every state from 2013 to 2019, leading to both increased and more equitable health care access. But progress stalled nationally after 2016, and all groups have reported recent drops in coverage or access." [3]

To expand upon previous research, this project reviews the question, "is there a combined effect of income, race, and region on a person's ability to pay or having issues paying medical bills in the last 12 months?" Academics have investigated the effect race and income have on health disparities and the effect region has on income. However, not much research has been done on the effect race, income, and race has on health disparities. This project's hypothesis is that there is a difference in average odds on a person's ability to pay or having issues paying medical bills based on the combined effect of

income, race, and region. More specifically, a person's race, region, and income affects the odds of whether they will have issues paying medical bills.

Methods

Data

This project used data from the IPUMS National Health Interview Series, or the IPUMS-NHIS. The survey was conducted in 2020, and given to random households throughout the United states. The NHIS covers, on average, 45,000 households each year and asks questions about general health status or health experiences in the past 12 months. [4] In the 2020 survey, the survey reached 32,345 households. [5]

Variables

This project has one binary outcome variable, two categorical explanatory variables, and one numeric explanatory variable. The binary outcome variable, or HIPROBPAYR, indicates whether or not a person had problems paying or were unable to pay any medical bills in the past 12 months. This included bills for doctors, dentists, hospitals, therapists, medication, equipment, nursing home or home care. The levels for this variable are "Yes" and "No". A "success" in this case is a survey taker putting "Yes" to "Problems paying or unable to pay medical bills in the last 12 months". A "failure" in this case is a survey taker putting "No" to "Problems paying or unable to pay medical bills in the last 12 months". The two categorical explanatory variables are REGION and RACENEW. In other words, REGION is the region the survey taker is from, denoted by four categories: North Central / Midwest, South, West, Northeast. At first, REGION contained the four categories above, but also included two other categories, NO DATA and Unknown. To prevent any confusion while performing our analysis, we removed all instances of NO DATA and Unknown before forming our model.

RACENEW is the self-reported race of the survey taker, denoted by five categories: White only, Black/African American only, American Indian/Alaska Native only, Asian only, Other Race and Multiple Race. At first, RACENEW contained:

- White only
- Black/African American only
- American Indian/Alaska Native only
- Asian only
- Other Race and Multiple Race
- Other Race and Multiple Race (2019-forward: Excluding American Indian/Alaska Native)
- Other Race
- Race Group Not Releasable
- Multiple Race
- Multiple Race (1999-2018: Including American Indian/Alaska Native)
- American Indian/Alaska Native and Any Other Race
- Unknown-Refused
- Unknown-Not ascertained

- Unknown-Don't Know

We dropped all categories with "Unknown" and the category "Race Group Not Releasable". If we do not know the individual's race, then we cannot test our hypothesis. Additionally, we wanted to include "American Indian / Alaska Native" and "Other Race and Multiple Race" as their own racial groups. Thus, we dropped "American Indian/Alaska Native and Any Other Race", "Multiple Race", "Multiple Race (1999-2018: Including American Indian/Alaska Native)", "Other Race", and "Other Race and Multiple Race". Adding any of these categories would create multicollinearity issues. We then renamed our variable "Other Race and Multiple Race (2019-forward: Excluding American Indian/Alaska Native)" as just "Other Race and Multiple Race".

IPUMS provided the option between getting our family income values in the form of a categorical variable with income bracket levels (INCFAM07ON) or a continuous numerical value variable (FAMTOTINC). We settled with the continuous numerical variable, since it would allow a more precise measure of income. To prevent a violation in the independence assumption, we wrangled our data to select only one family member per household.

We chose HIPROBAYR over similar variables which also explained a person's healthcare payment issues because HIPROBAYR included both individuals who had issues paying medical bills or could not pay them at all within the last 12 months. It included the most number of observations that relate to our research question as a result.

Model

We had three choices for our model in this project. First, race and income had an interactive relationship when controlling for the additive effect of family income: *HIPROBPAYR ~ REGION + RACENEW * FAMTOTINC*. Second, race, income, and region had an interactive relationship: *HIPROBPAYR ~ REGION * RACENEW * FAMTOTINC*. Third, race and region had an interactive relationship when controlling for the additive effect of family income: *HIPROBPAYR ~ REGION * RACENEW + FAMTOTINC*. After comparing nested F-tests on all of our models, we decided that the third model, *HIPROBPAYR ~ REGION * RACENEW + FAMTOTINC*, was best in predicting whether or not a person had issues with paying medical bills *(p = 0.001374)*.

In other words, our chosen model states that the effect of the region will change across racial groups, when controlling for the simultaneous additive effect of family income.

Results

Our model, *HIPROBPAYR ~ REGION * RACENEW + FAMTOTINC*, produces an intercept of `-1.670`. In other words, the estimated odds of having problems paying your medical bills for a white person with an income of zero who lives in the Northeast would be `0.1882471 to 1`, on average. The intercept coefficient is statistically significant with *p-value of < 2e-16*, indicating a very small probability of observing a coefficient of this magnitude or more extreme if the actual value was zero.

Our model also produces a slope of `-9.604e-06`. In other words, for each additional $1 in a person's annual family income, the odds of them having issues paying for their medical bills decrease `0.999`

times, on average. The intercept coefficient is statistically significant with *p-value of < 2e-16*, indicating a very small probability of observing a coefficient of this magnitude or more extreme if the actual value was zero.

The only other coefficients in our model that produce significant results is one region, two racial groups, and one interaction of region and racial group produced significant differences from the average effect of race and region.

The coefficient value for *REGIONSouth* is `3.052e-01`. Thus, the odds of having issues paying for medical bills for white people with an annual family income of $0 in the South are `1.357` times the odds for the same group in the Northeast, on average *(p = 2.47e-06)*.

The coefficient value for *RACENEWBlack/African American only* is `2.865e-01`. Thus, the odds of having issues paying for medical bills for Black/African American only race group with an annual family income of $0 in the Northeast are 1.332 times the odds for a white person with an income of $0 in the Northeast, on average *(p = 0.04274)*.

The coefficient value for *RACENEWAsian* only is `-5.978e-01`. The odds of having issues paying for medical bills for Asian only race group with an annual family income of $0 in the Northeast are 0.550 times the odds for a white person with an income of $0 in the Northeast, on average *(p = 0.04274)*.

The coefficient value for *REGIONNorth Central/Midwest:RACENEWBlack/African American only* is `5.546e-01`. Thus, the odds of having issues paying for medical bills for Black/African American only race group with an annual family income of $0 in the Midwest are 1.741 times the odds for a white person with an income of $0 in the Northeast, on average *(p = 0.00264)*.

The other coefficients in our model related to region, race, and the interaction of region and race did not produce significant differences from the average effect of race and region.

Odds of Having Problems Paying Medical Bills by Race, Region, and Income

Annual Family Income in 2020 (USD)

Region — South — West — North Central/Midwest — Northeast
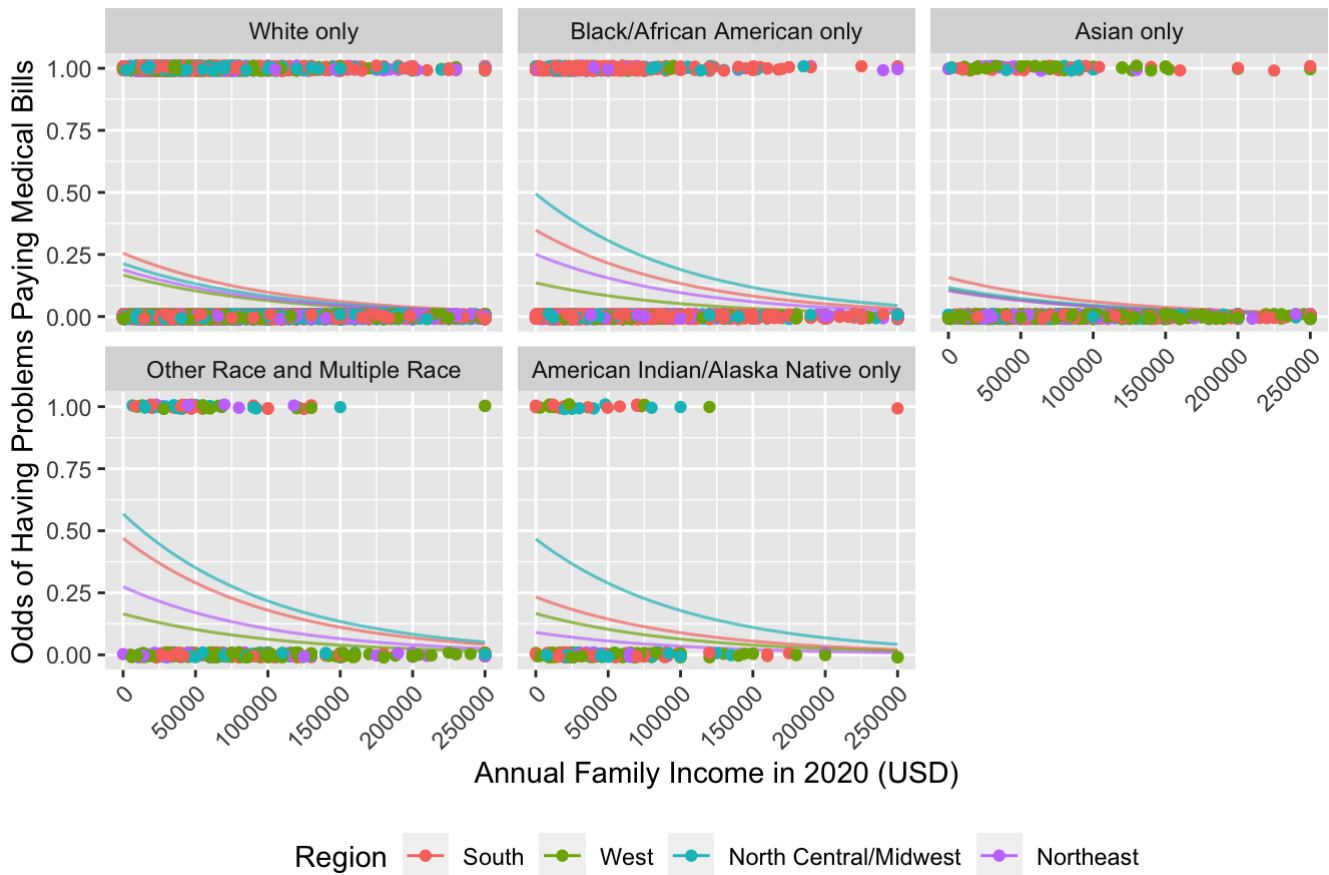
Discussion

We predicted that a person's race, region, and income combined affects the odds of whether they will have issues paying medical bills. Our prediction ended up having some evidence backing it, but we cannot fully accept our hypothesis. For regions, only the South had significant enough difference from the baseline average. For race groups, only Black/African Americans and Asian racial groups had a significant difference from the baseline average. Finally, only one combination of race and region, Black/African Americans in the Midwest, had a significant difference from the baseline. Across all groups, a person's annual income was a significant predictor of whether or not they had issues paying their medical bills.

Based on the IPUMS 2020 Health Survey, we can state that certain race groups, regions, and combination of race group and region are predictors of whether or not a person has issues paying for medical bills, but we cannot say that every racial group and region have a significant enough difference from the average. That is not to say that the racial groups and regions and their combinations do not have significance. We simply do not have the evidence to confirm or deny its significance.

It should also be noted that our data may produce less accurate results for observations related to "American Indian/Alaska Native only" & "Other Race and Multiple Race" categories. This is due to the relatively small percentage of the population, and thus our data, they make up compared to other racial

groups. Similarly, our data is limited to the fact that some U.S. regions are simply less populated than other regions. [6]

To expand on our findings in the future, it would be helpful to research why these specific regions and race groups have more difficulties in paying for medical bills. One could also look into which states individuals live, and how many people contribute to the annual family income. Do they have pre-existing conditions? Does age affect the effect of race and region? This can all be further investigated with the help of our findings.

Data Analysis Appendix

Linearity
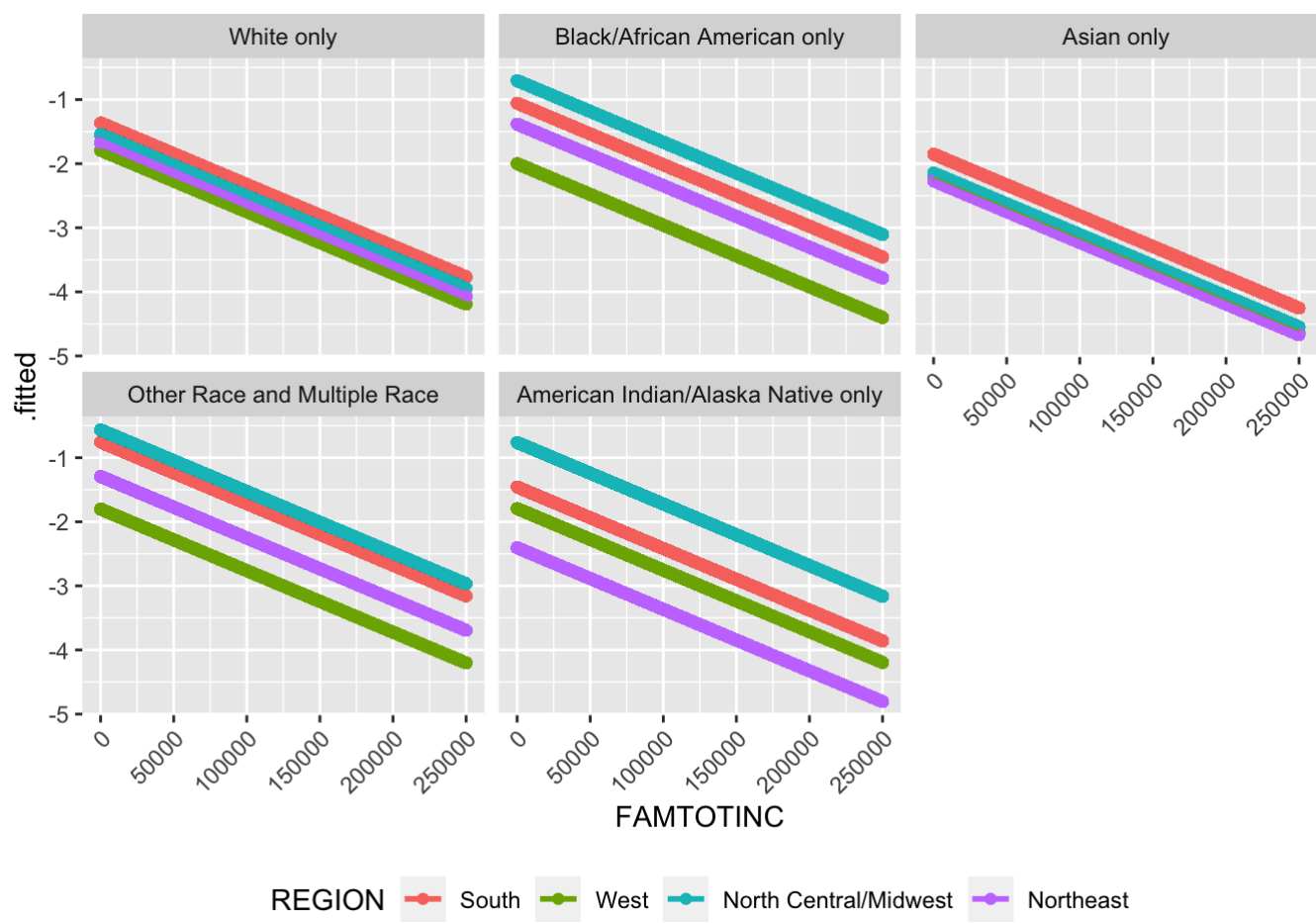
In the logit vs. predictors plot, the line is relatively linear, approving the linearity assumption.

Independence

As for our independence issue, we initially had issues with the IPUMS data because it contained multiple individuals from the same household. We were able to fix this by selecting just one individual from each household in our data. Additionally, we encountered an independence issue with our RACENEW racial categories. We found that our *Other Race and Multiple Race* variables were combinations of *Other Race and Multiple Race (2019-forward: Excluding American Indian/Alaska Native)* and *American Indian/Alaska*

Native only variables. We fixed this by using only *Other Race and Multiple Race (2019-forward: Excluding American Indian/Alaska Native)* and renaming it to *Other Race and Multiple Race*.

Randomness

This project used data from the IPUMS National Health Interview Series which gathered its data from random households throughout the United states: "IPUMS USA data can be considered representative of geographic areas of all sizes, as long as the proper weights are applied. The standard errors of statistics based on smaller geographic areas will be larger than those based on national data, but this is due only to the reduced sample size rather than to the fact that one is analyzing a small area". [7] Thus, our randomness condition for our model is sufficed.

Model Comparison

```
Analysis of Deviance Table

Model 1: HIPROBPAYR ~ REGION * RACENEW + FAMTOTINC
Model 2: HIPROBPAYR ~ REGION * RACENEW * FAMTOTINC
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     30323      19108
2     30304      19082 19   25.912   0.1327


Analysis of Deviance Table

Model 1: HIPROBPAYR ~ REGION + RACENEW * FAMTOTINC
Model 2: HIPROBPAYR ~ REGION * RACENEW + FAMTOTINC
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     30331      19133
2     30323      19108  8   25.316 0.001374 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multicollinearity

```
                      GVIF Df GVIF^(1/(2*Df))
REGION            1.973168  3        1.119938
RACENEW        9037.386383  4        3.122521
FAMTOTINC         1.030960  1        1.015362
REGION:RACENEW 13903.285569 12       1.488092
```

We performed a variance inflation factor test and all our coefficients produced values less than 5 meaning we detected no multicollinearity issues.

Full Model Summary

	Estimate
(Intercept)	−1.669834e+00
North Central/Midwest	1.243256e−01
South	3.051746e−01
West	−1.204111e−01

Black/African American only	2.865049e-01
American Indian/Alaska Native only	-7.354888e-01
Asian only	-5.978344e-01
Other Race and Multiple Race	3.770723e-01
FAMTOTINC	-9.604548e-06
North Central/Midwest:Black/African American only	5.545950e-01
South:Black/African American only	2.188111e-02
West:Black/African American only	-4.954957e-01
North Central/Midwest:American Indian/Alaska Native only	1.518721e+00
South:American Indian/Alaska Native only	6.429015e-01
West:American Indian/Alaska Native only	7.317709e-01
North Central/Midwest:Asian only	-3.760733e-03
South:Asian only	1.102828e-01
West:Asian only	1.547641e-01
North Central/Midwest:Other Race and Multiple Race	6.020369e-01
South:Other Race and Multiple Race	2.312020e-01
West:Other Race and Multiple Race	-3.888393e-01

	Std. Error
(Intercept)	6.114763e-02
North Central/Midwest	6.901502e-02
South	6.478912e-02
West	7.455326e-02
Black/African American only	1.413991e-01
American Indian/Alaska Native only	1.033237e+00
Asian only	2.716531e-01
Other Race and Multiple Race	4.107304e-01
FAMTOTINC	4.322686e-07
North Central/Midwest:Black/African American only	1.844529e-01
South:Black/African American only	1.565340e-01
West:Black/African American only	2.606293e-01
North Central/Midwest:American Indian/Alaska Native only	1.102160e+00
South:American Indian/Alaska Native only	1.076661e+00
West:American Indian/Alaska Native only	1.092869e+00
North Central/Midwest:Asian only	4.269286e-01
South:Asian only	3.496816e-01
West:Asian only	3.208622e-01
North Central/Midwest:Other Race and Multiple Race	5.155889e-01
South:Other Race and Multiple Race	4.905600e-01
West:Other Race and Multiple Race	4.994062e-01

	z value
(Intercept)	-27.30824331
North Central/Midwest	1.80142798
South	4.71027516
West	-1.61510166
Black/African American only	2.02621430
American Indian/Alaska Native only	-0.71183005
Asian only	-2.20072729
Other Race and Multiple Race	0.91805307
FAMTOTINC	-22.21893588
North Central/Midwest:Black/African American only	3.00670196
South:Black/African American only	0.13978500

```
West:Black/African American only                                    -1.90115095
North Central/Midwest:American Indian/Alaska Native only             1.37794961
South:American Indian/Alaska Native only                             0.59712507
West:American Indian/Alaska Native only                              0.66958716
North Central/Midwest:Asian only                                    -0.00880881
South:Asian only                                                     0.31538059
West:Asian only                                                      0.48233816
North Central/Midwest:Other Race and Multiple Race                   1.16766843
South:Other Race and Multiple Race                                   0.47130212
West:Other Race and Multiple Race                                   -0.77860324
                                                                       Pr(>|z|)
(Intercept)                                                        3.385706e-164
North Central/Midwest                                              7.163545e-02
South                                                              2.473826e-06
West                                                               1.062887e-01
Black/African American only                                        4.274283e-02
American Indian/Alaska Native only                                 4.765700e-01
Asian only                                                         2.775534e-02
Other Race and Multiple Race                                       3.585911e-01
FAMTOTINC                                                          2.253462e-109
North Central/Midwest:Black/African American only                  2.640986e-03
South:Black/African American only                                  8.888299e-01
West:Black/African American only                                   5.728224e-02
North Central/Midwest:American Indian/Alaska Native only           1.682188e-01
South:American Indian/Alaska Native only                           5.504239e-01
West:American Indian/Alaska Native only                            5.031210e-01
North Central/Midwest:Asian only                                   9.929717e-01
South:Asian only                                                   7.524727e-01
West:Asian only                                                    6.295657e-01
North Central/Midwest:Other Race and Multiple Race                 2.429405e-01
South:Other Race and Multiple Race                                 6.374250e-01
West:Other Race and Multiple Race                                  4.362135e-01
```

Our original model summary involved redundant category names in the coefficient variable names. This made it difficult to understand the summary of the model. Thus, we removed the category names from the variable names in this summary for ease of reading.

Data Count Summary

```
# A tibble: 4 × 2
  REGION                    n
  <fct>                 <int>
1 Northeast              5477
2 North Central/Midwest  7077
3 South                 10615
4 West                   7175

# A tibble: 5 × 2
  RACENEW                     n
  <fct>                   <int>
1 White only              24566
```

```
 2 Black/African American only           3362
 3 American Indian/Alaska Native only    244
 4 Asian only                            1750
 5 Other Race and Multiple Race          422
# A tibble: 20 × 3
   REGION              RACENEW                                  n
   <fct>               <fct>                                <int>
 1 Northeast           White only                            4552
 2 Northeast           Black/African American only            527
 3 Northeast           American Indian/Alaska Native only      18
 4 Northeast           Asian only                             318
 5 Northeast           Other Race and Multiple Race            62
 6 North Central/Midwest White only                          6318
 7 North Central/Midwest Black/African American only          463
 8 North Central/Midwest American Indian/Alaska Native only    40
 9 North Central/Midwest Asian only                           192
10 North Central/Midwest Other Race and Multiple Race          64
11 South               White only                            8007
12 South               Black/African American only           2040
13 South               American Indian/Alaska Native only      92
14 South               Asian only                             377
15 South               Other Race and Multiple Race            99
16 West                White only                            5689
17 West                Black/African American only            332
18 West                American Indian/Alaska Native only      94
19 West                Asian only                             863
20 West                Other Race and Multiple Race           197
```

Footnotes

1. https://equitablegrowth.org/how-national-income-inequality-in-the-united-states-contributes-to-economic-disparities-between-regions/↩
2. https://www.ncbi.nlm.nih.gov/books/NBK25526/↩
3. https://www.commonwealthfund.org/publications/issue-briefs/2021/jun/racial-ethnic-inequities-health-care-coverage-access-2013-2019↩
4. https://www.ipums.org/projects/ipums-health-surveys/d070.v6.2↩
5. https://nhis.ipums.org/nhis/userNotes_size.shtml↩
6. https://www.census.gov/popclock/data_tables.php?component=growth ↩
7. https://usa.ipums.org/usa-action/faq↩