Visualizations for
US COVID Demographic Classifications

by: Hope Muller

A team of students and faculty affiliated with New York University facilitated this project. We designed focus questions, then gathered data from the District of Columbia and each of the 50 states' COVID dashboards in July 2020. Classifications for each state dashboard were entered by two members of our team.

In this document, you will see data cleaning for several visualizations of the public COVID data. These visualizations were the exploratory analysis I did for our team. We hoped to recognize people groups that were potentially invisible in public COVID data in our country. Further research and analysis of that information will be on a different platform.

Variables

The variables that were analyzed from our team's dataset include:

- record_id: the unique entry id
- nice_volunteer: the name of the person who entered the data
- state: the US state, DC or a few entries listed Canada or multi-state options
- county: COVID data from specific counties, used for some large cities before some states posted state-wide data

Race & Ethnicity

- bipoc: binary option of whether or not a race/ethnicity stat or either of those options was reported
- race_ethn:binary option of whether or not a race/ethnicity stat was reported (the race/ethncity column is unique and does not overlap with the race or ethnicity columns except in specific circumstances)
- race_ethn_num: number of category options for race/ethnicity used on dashboard
- race_ethn_cat: list of category options for race/ethncity used on dashboard (for example: Black, White, Asian, Hispanic, Other)
- race: binary option of whether or not a race, alone, stat was reported
- race_num: number of category options for race, alone, used on dashboard
- race_cat: list of category options for race, alone, used on dashboard (for example: Black, White, Other)
- ethn: binary option of whether or not an ethnicity, alone, stat was reported
- ethn_num: number of category options for ethnicity, alone, used on dashboard
- ethn_cat: list of category options for ethnicity, alone, used on dashboard (for example: Hispanic, Non-Hispanic)

Gender/Sex

- gend: binary option of whether or not a gender/sex stat was reported
- gend_term: the name of this category on the state's dashboard (for example: Sex or Gender)
- gend_num: number of category options for gender/sex used on dashboard

- gend_cat: list of category options for gender/sex used on dashboard (for example: Male, Female, Other, Unknown)
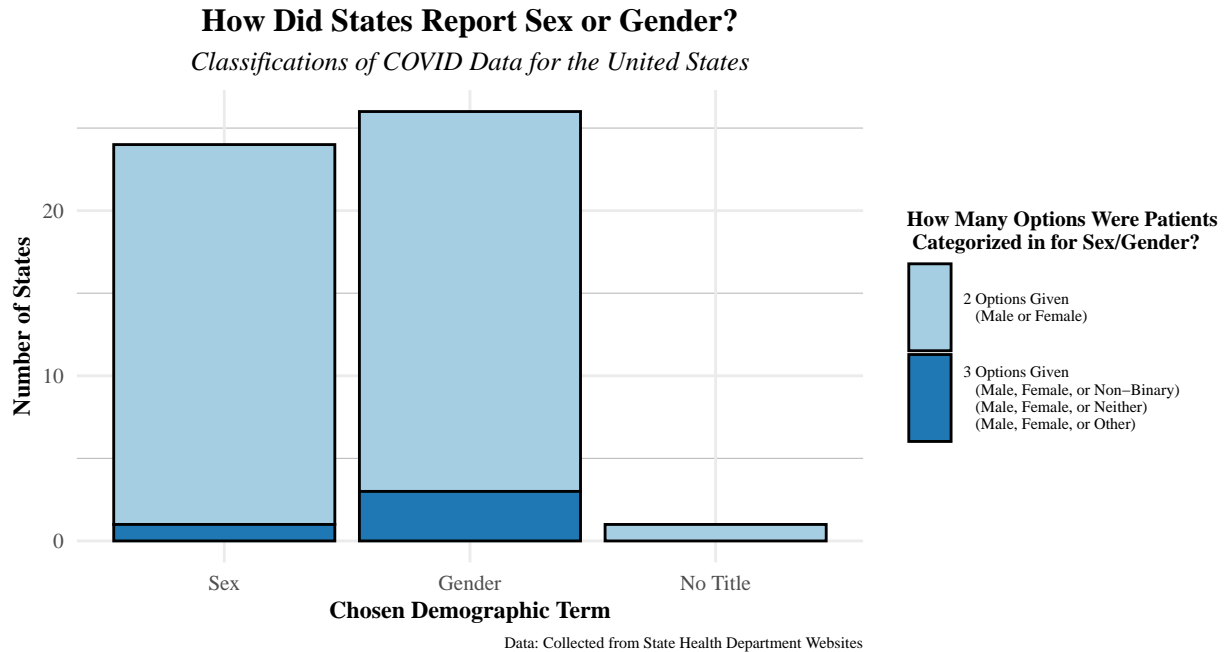
Age

- age_range: method of age reporting on dashboard (for example: individual year, decade, other)
- age_cat: ranges of age categories used on dashboard (for example: 0-17; 18-44; 45-64; 65-74; 75+)

Additional Classifications

- income: binary option of whether or not a income-level or income stat was reported
- occupation_cat: qualitative entry of aggregated employers of COVID-19 patients
- demo_terms: qualitative entry of other aggregated demographics reported for that state's COVID-19 patients
- cross_comp: binary option of whether or not cross-comparisons existed among categories (for example: sex was also aggregated by race or age)
- comorbid: binary option of whether or not aggregated comorbidity stats were reported
- comorb_desc: binary option of whether or not aggregated comorbidity categories were reported
- comorb_cat: list of category options for comorbidities used on dashboard

Sex and Gender Demographics

All states in the United States specified sex or gender for their public COVID-19 data except for the state of Hawaii. This plot shows which states used the term gender and which used the term sex to categorize COVID-19 patients. It displays the number of categories used within each. Most states used two categories. Four states, including Connecticut, Pennsylvania, Minnesota, and Oregon used 3 categories.

How Did States Report Sex or Gender?

Classifications of COVID Data for the United States

How Many Options Were Patients Categorized in for Sex/Gender?

2 Options Given
(Male or Female)

3 Options Given
(Male, Female, or Non–Binary)
(Male, Female, or Neither)
(Male, Female, or Other)

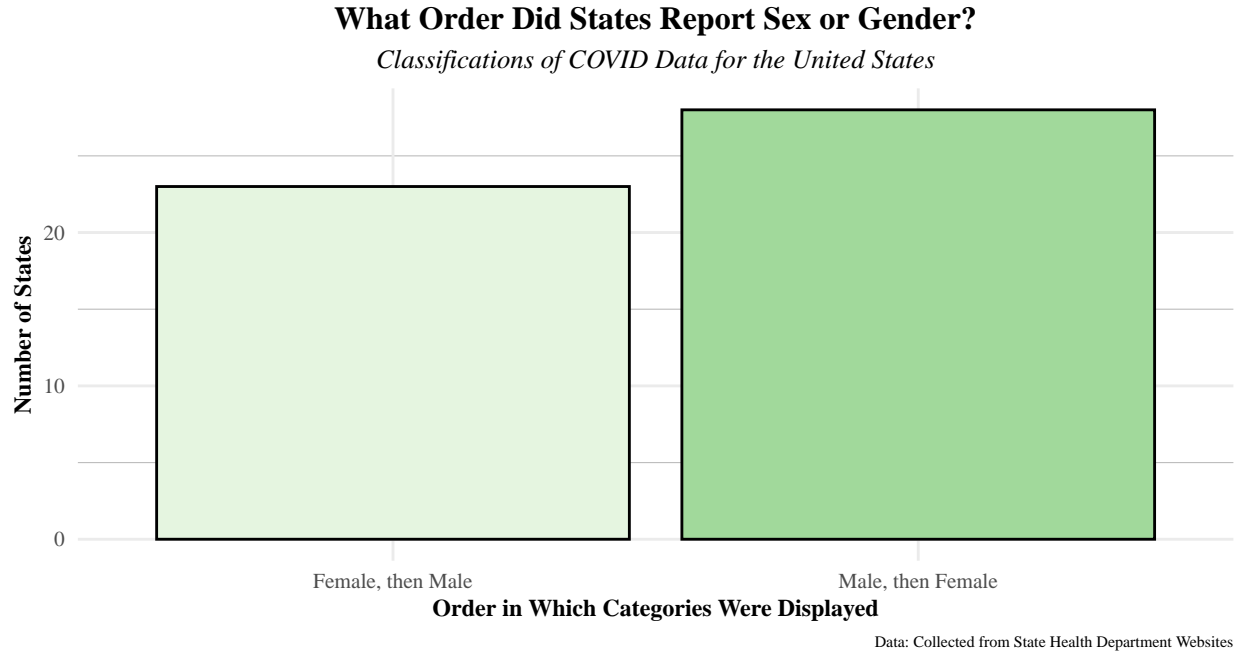Data: Collected from State Health Department Websites

The following table depicts the four states that gave more than two options to categorize COVID patients in the sex or gender demographic.

Table 1: Which States Gave a Third Option for Sex or Gender for COVID Patients?

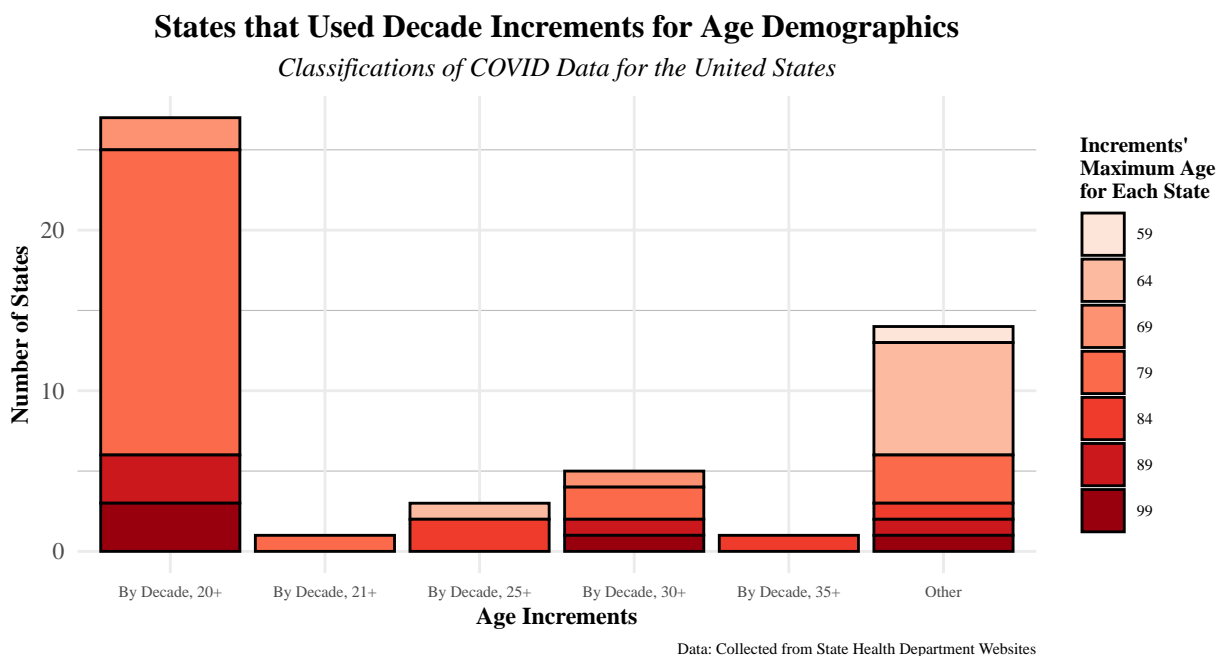| State	Demographic Term Used	Third Option for Sex or Gender Given
Connecticut	Gender	Other
Minnesota	Gender	Other
Pennsylvania	Gender	Neither
Oregon	Sex	Non-Binary, Refused, or Unknown

The following plot depicts what order each state listed male or female on their COVID dashboard.

What Order Did States Report Sex or Gender?

Classifications of COVID Data for the United States

Data: Collected from State Health Department Websites

Age Demographics

All states in the US included an age demographic with their COVID data, except for Ohio. More than half of states included typical decade increments. Other states used a variety of unique imcrements for categorizing their patients' ages.

The following plot depicts the decade and other increments that states used for their public COVID data reporting. Ohio is not included in the age demographic plots.
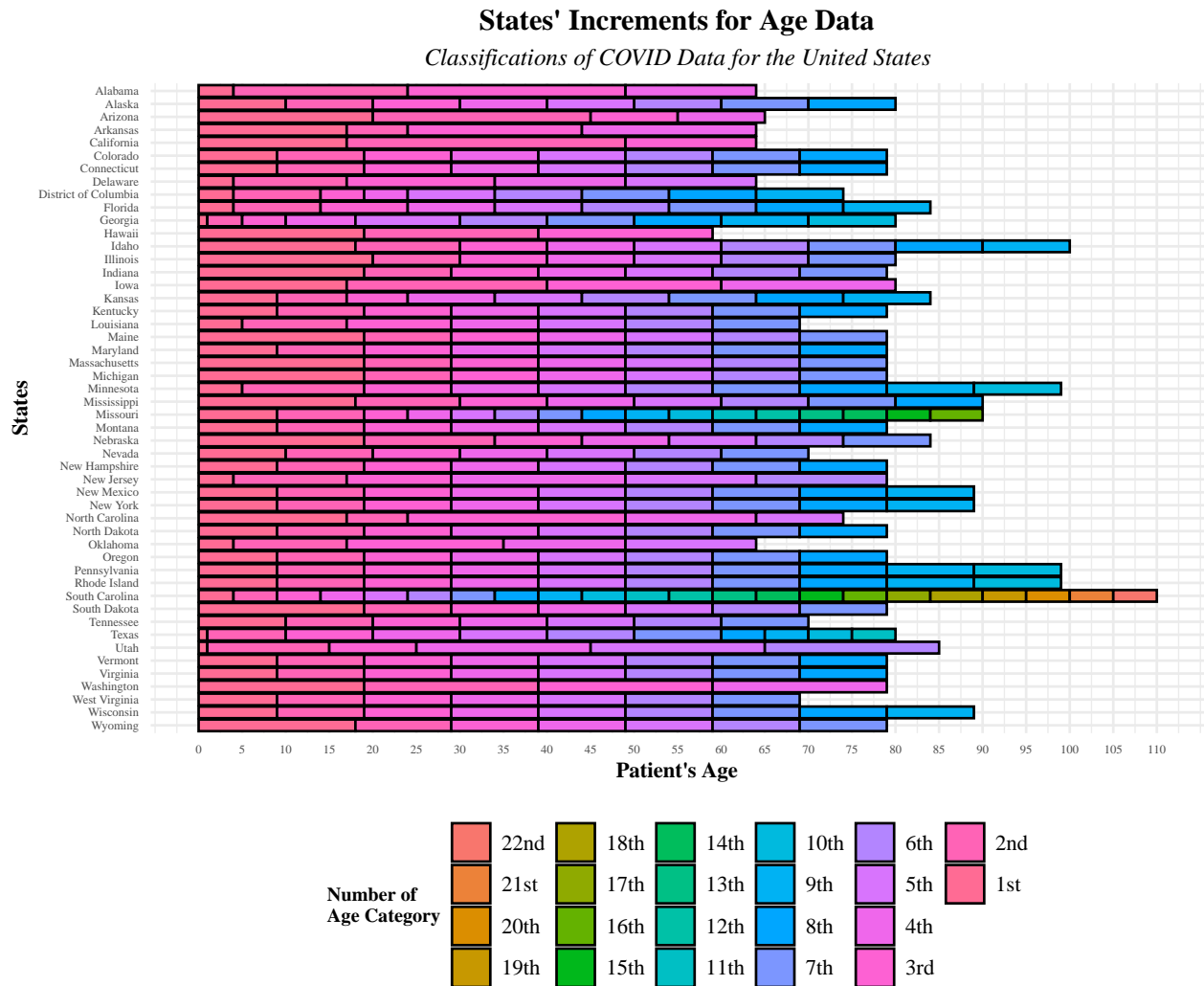
States that Used Decade Increments for Age Demographics

Classifications of COVID Data for the United States

Data: Collected from State Health Department Websites

The following table specifies the increments used by the fourteen states included in the "Other" column above.

Table 2: Other than by Decade, What Age Increments Did States Use for COVID data?

Increment Pattern	State	Increment Specifics
20 Years	Arkansas	0-17, 18-24 ... 65+
20 Years	Hawaii	0 ... 60+
20 Years	Iowa	0-17, 18-40 ... 81+
20 Years	Utah	0-1, 1-14 ... 85+
20 Years	Washington	0 ... 80+
5 Years	Missouri	0-19, ... 80+
5 Years	South Carolina	0 ... 109+
Various	Alabama	0-4, 5-24, 25-49, 50-64, 65+
Various	Arizona	0-20, 20-44, 45-54, 55-64, 65+
Various	California	0-17, 18-49, 50-64, 65+
Various	Delaware	0-4, 5-17, 18-34, 35-49, 50-64, 65+
Various	New Jersey	0-4, 5-17, 18-29, 30-49, 50-64, 65-79, 80+
Various	North Carolina	0-17, 18-24, 25-49, 50-64, 65-74, 75+
Various	Oklahoma	0-4, 5-17, 18-35, 36-49, 50-64, 65+

The following plot shows the 50 states plus the District of Columbia and the age increments each state used for their public COVID data. Each rectangle represents one increment used.
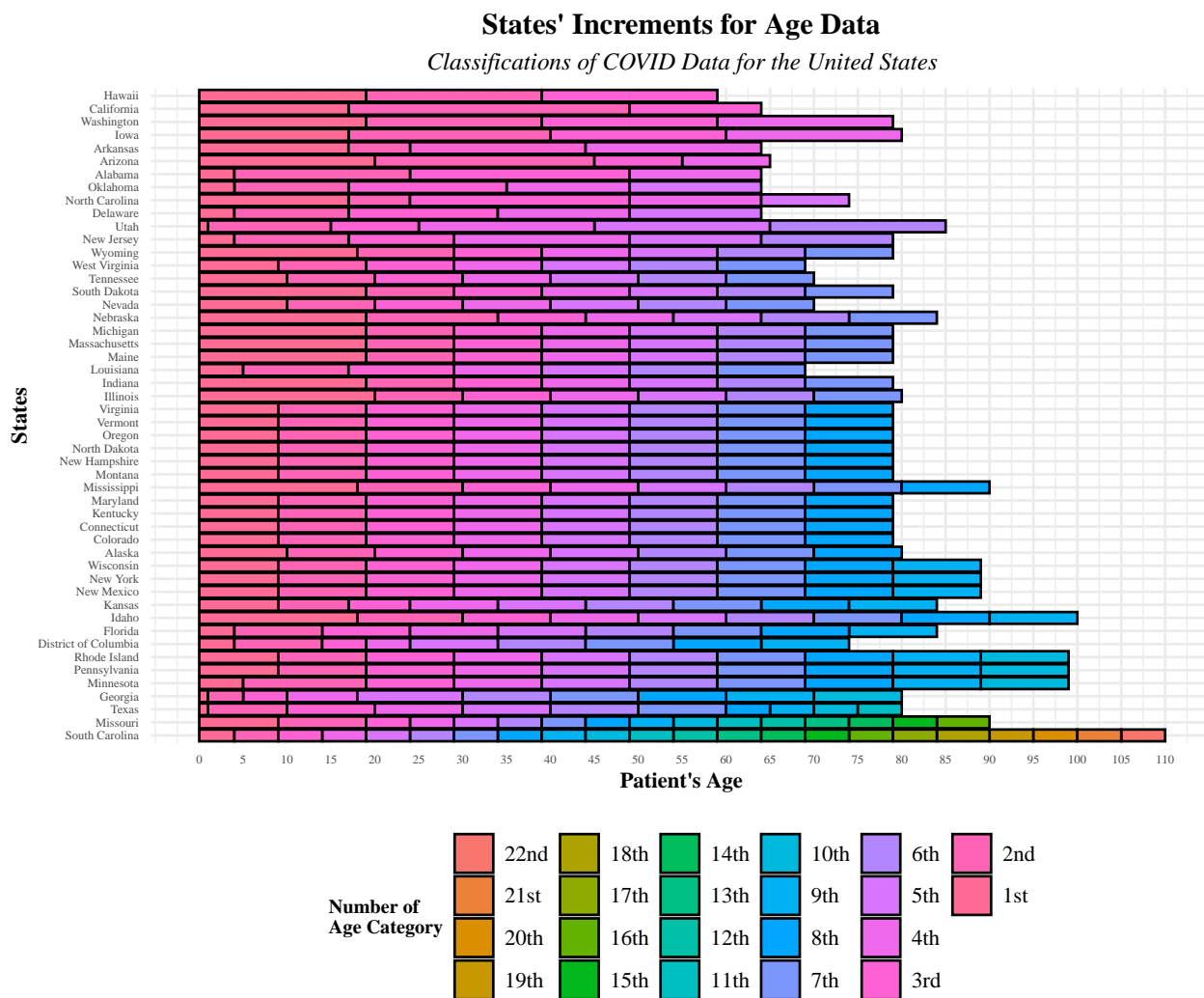
Specifically, you can see Alabama has four rectangles on the top of the plot. The four rectangles represent the four age categories used to group their COVID data on the dashboard. The first rectangle represents that they bunched ages 0-4 together ("1st increment"), the second rectangle represents they bunched ages 5-24 together ("2nd increment"), then ages 25-49 are groups together, and ages 50-64 are grouped together. Then, Alabama grouped all patients over age 65 together. The oldest, all-inclusive, age increment is not shown on this plot for any state. The rest of the states used various increments to bunch COVID patients together, as you can see below.

States' Increments for Age Data
Classifications of COVID Data for the United States

Data: Collected from State Health Department Websites

The following plot is the same exact depiction as the plot above, however, it is ordered differently (by the maximum increment used).

As shown below, there is a wide range of numbers of age increments used. The minimum number was three increments. California and Hawaii grouped patients into three age categories. This means they groups patients in with people outside of their peer group when representing COVID totals. The maximum number of increments was 22. South Carolina grouped patients into 22 age categories. They counted by fives in their age increments. So, this implies that they groups COVID patients in with a very small span of a peer group when relaying data.

States' Increments for Age Data

Classifications of COVID Data for the United States

Data: Collected from State Health Department Websites

Race and Ethnicity Demographics

All states in the United States have specified race and/or ethnicity for their public COVID dashboards.

California, Illinois, and Mississippi included a race/ethnicity section as well as a separate race section. So, they each are listed twice in the plot below. All states that included an ethnicity section also included a race section or a race/ethnicity section. So, the states in the ethnicity column are also taken into account in one of the other columns.

States included a range of 2-9 race and/or ethnicity categories in which to classify people.
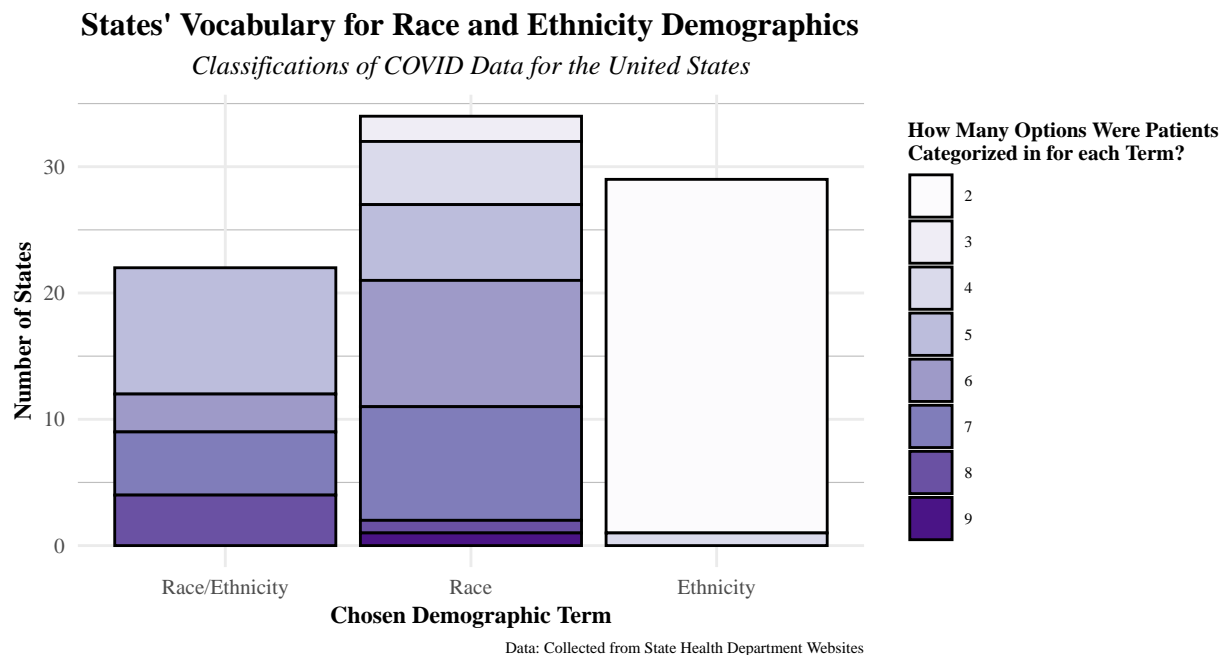
States' Vocabulary for Race and Ethnicity Demographics

Classifications of COVID Data for the United States

Data: Collected from State Health Department Websites

Table 3: What Common Race Categories Did States Use for COVID Data?

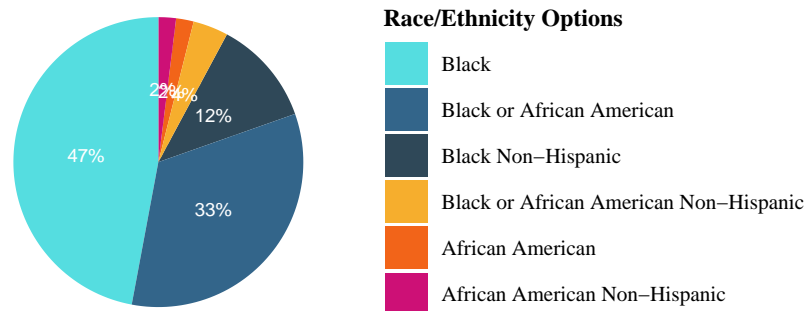| Race Categories	Percent of States
Black	100 %
White	100 %
Asian	92 %
Latino/Hispanic	92 %
Native American	71 %
Pacific Islander	53 %
Multiple Races	39 %

Subgroups for Race and Ethnicity Options Given

Below are charts that show the race and/or ethnicity terminology from the COVID data from 50 states and District of Columbia. They are grouped into continents (or areas) of descent. The last plot is for those from a mix of races. As seen in the charts below, states gave patients between 0 to 8 choices for these subgroups.

Below are the race/ethnicity options for those from African descent, who would choose non-Hispanic when given the option.

Which Race Options Were Specified by Each State?

Patients from African Descent

Race/Ethnicity Options
- Black
- Black or African American
- Black Non–Hispanic
- Black or African American Non–Hispanic
- African American
- African American Non–Hispanic

Below are the race/ethnicity options for those from European descent, who would choose non-Hispanic when given the option.

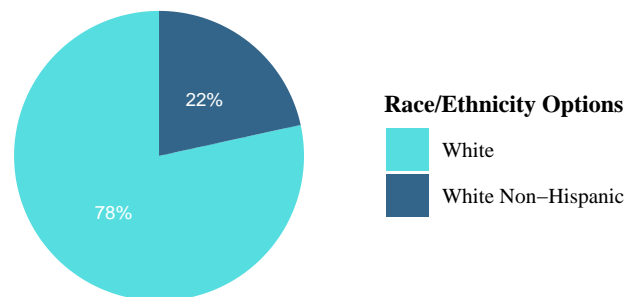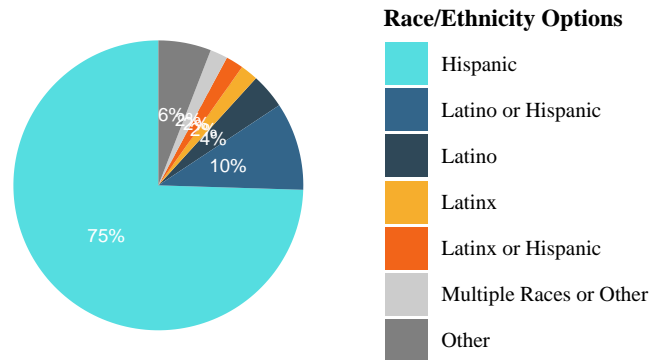### Which Race Options Were Specified by Each State?

Patients from European Descent

Race/Ethnicity Options
- White
- White Non–Hispanic

Below are the race/ethnicity options for those from Latin descent, who would opt for Hispanic when given the option. It is also combined with options collected in a separate ethnicity section, if there was one included for patients.

Which Race Options Were Specified by Each State?

Patients from Latin Descent & Hispanic

There are approximately 59,763,631 people of Latin descent who would identify as Hispanic in the United States (U.S. Census Bureau, 2018). As you can see above, 8% of states (4 states) did not publish race-specific COVID data about Hispanic patients from Latin descent.

The table below show population information from the states that did not report a Latin descent or Hispanic option for their COVID data. These four states have around 33,448 Hispanic people of Latin descent. That value is 1% of the nation's Hispanic people of Latin descent. So, are 1% of the nation's Hispanic people of Latin descent being described as "Other" in public COVID data?
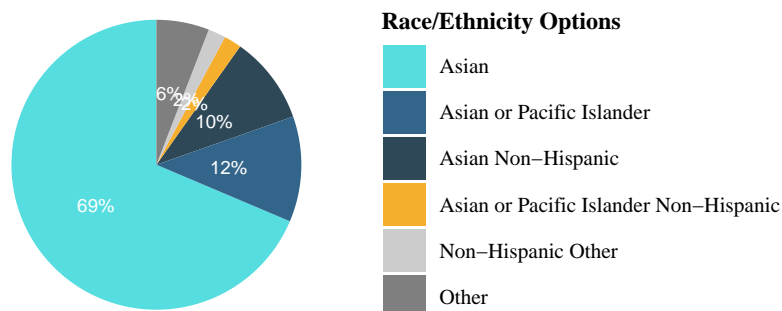
Table 4: Which States Would Report COVID Patients from Latin Descent as Other?

	State	Percent of Population	People of Latin Descent in the State
	Oklahoma	10.9	429079
	Hawaii	10.7	151356
	North Dakota	3.6	27444
	West Virginia	1.4	25569
Total			**633448**

Below are the race/ethnicity options for those from Asian descent, who would not opt for Hispanic when given the option.

Which Race Options Were Specified by Each State?

Patterns from Asian descent

Race/Ethnicity Options
- Asian
- Asian or Pacific Islander
- Asian Non–Hispanic
- Asian or Pacific Islander Non–Hispanic
- Non–Hispanic Other
- Other

There are approximately 18,193,513 people of Asian descent in the United States (U.S. Census Bureau, 2018). As you can see above, 8% of states (5 states) did not publish race-specific COVID data reflecting the number of patients that are a part of this community.

The table below show the states that did not report race-specific data about patients from Asian descent for thier COVID data. These five states have around 763,000 people from Asian descent. That value is 4% of the nation's Asian population. So, are 4% of the nation's people from Asian descent being described as "Other" in public COVID data?
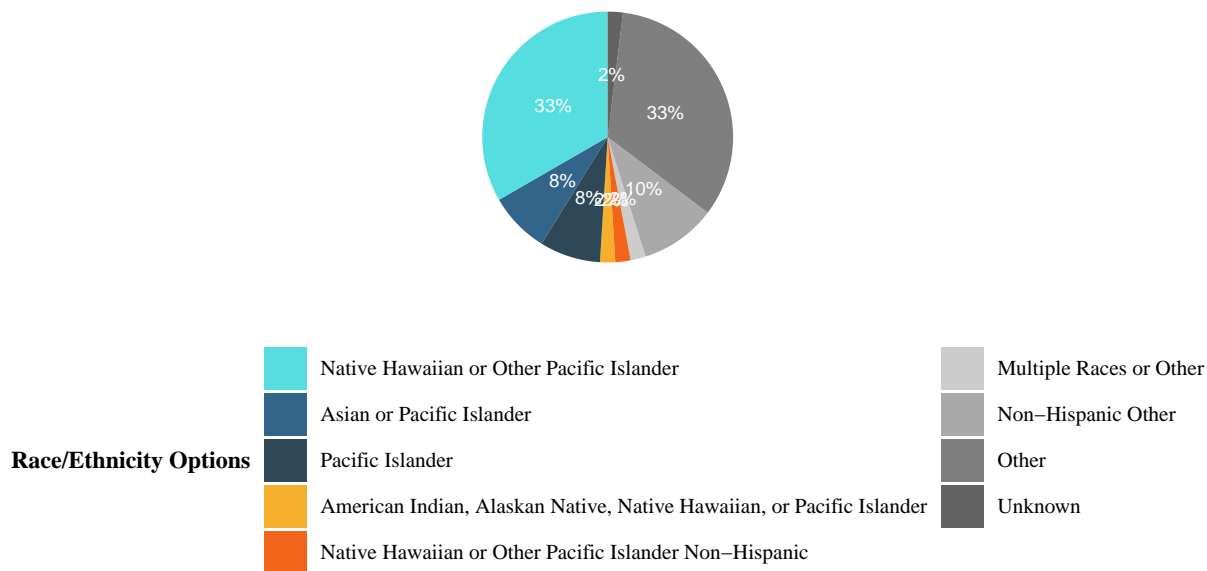
Table 5: Which States Would Report COVID Patients from Asian Descent as Other?

State	Percent of Population	People of Asian Descent in the State
Florida	2.7	580229
Missouri	2	120440
Rhode Island	3.3	35378
West Virginia	0.7	13311
Total		**749358**

Below are the race/ethnicity options for those from Pacific Island descent, who chose non-Hispanic when given the option.

Which Race Options Were Patients Given by Each State?

Patients from Pacific Island Descent

Race/Ethnicity Options

- Native Hawaiian or Other Pacific Islander
- Asian or Pacific Islander
- Pacific Islander
- American Indian, Alaskan Native, Native Hawaiian, or Pacific Islander
- Native Hawaiian or Other Pacific Islander Non–Hispanic
- Multiple Races or Other
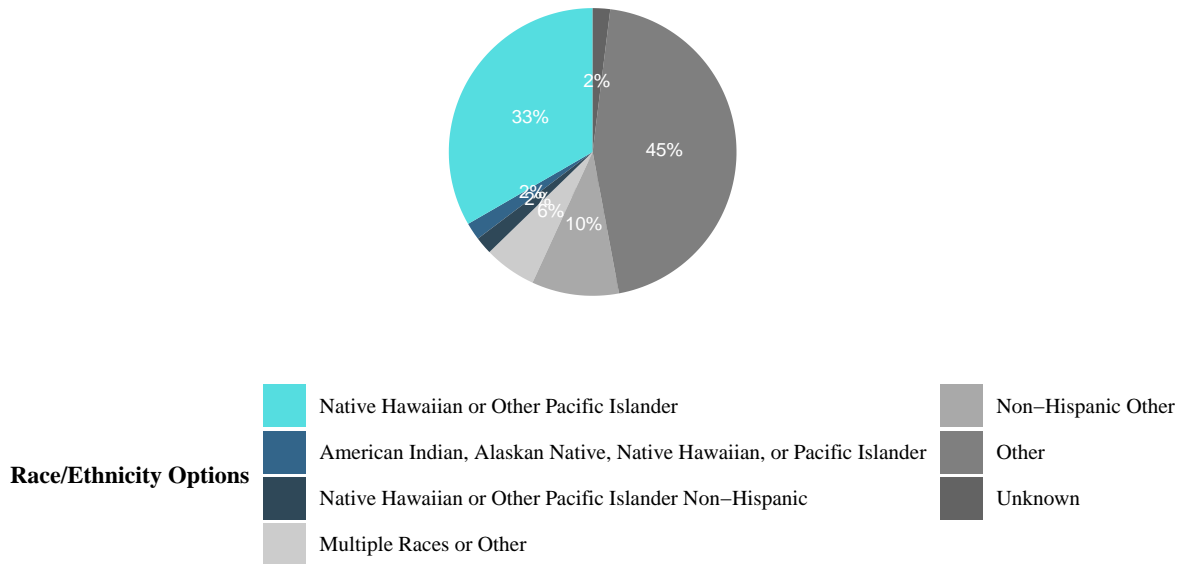- Non–Hispanic Other
- Other
- Unknown

As you can see above, 47% of states did not publish race-specific COVID data reflecting the number of patients from Pacific Island descent.

Below are the race/ethnicity options for those from Native Hawaiian Island descent, who chose non-Hispanic when given the option.

Which Race Options Were Patients Given by Each State?

Patients from Native Hawaiian Descent

Race/Ethnicity Options

- Native Hawaiian or Other Pacific Islander
- American Indian, Alaskan Native, Native Hawaiian, or Pacific Islander
- Native Hawaiian or Other Pacific Islander Non–Hispanic
- Multiple Races or Other
- Non–Hispanic Other
- Other
- Unknown

As you can see above, 64% of states did not publish race-specific COVID data reflecting the number of patients from Native Hawaiian descent.

There are approximately 559,945 people of Native Hawaiian or Pacific Island descent in the United States (U.S. Census Bureau, 2018).

The table below show the states that did not report race-specific data about patients from Native Hawaiian or Pacific Island descent. These states have around 89,049 people from Native Hawaiian or Pacific Island descent. That value is 16% of the nation's Native Hawaiian and Pacific Island population. So, are 16% of the nation's people from Native Hawaiian or Pacific Island descent being described as "Other" in public COVID data?
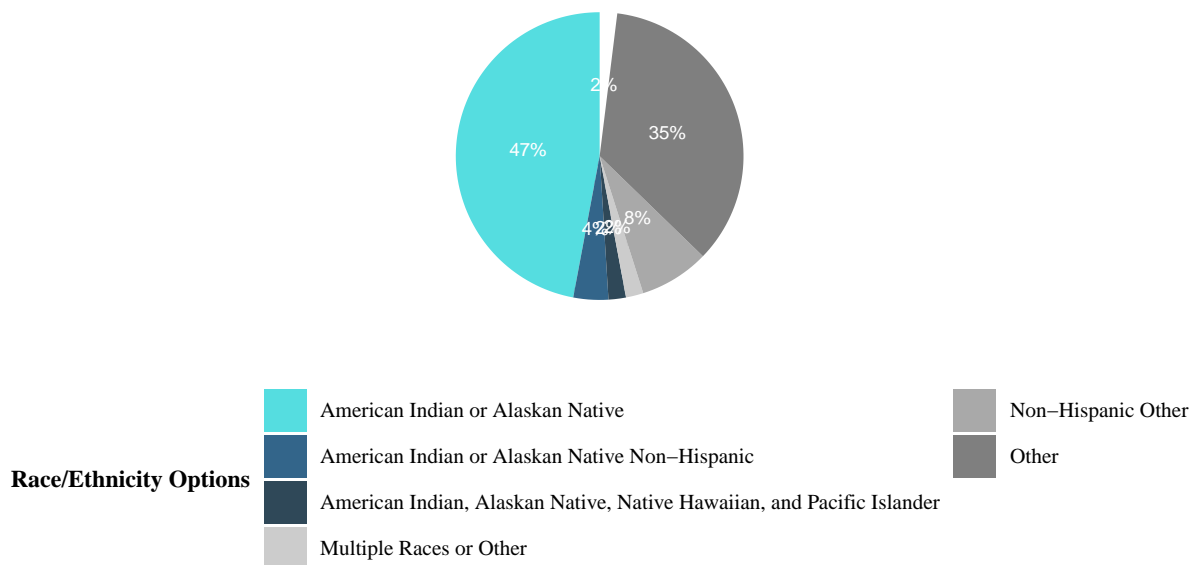
Table 6: Which States Would Report COVID Patients from Native Haw. or Pacific Is. Descent as Other?

State	Percent of Population	People of NHOPI Descent in the State
Texas	0.1	18958
Nevada	0.6	18249
Florida	0	9694
Missouri	0.1	7095
New York	0	4765
South Carolina	0.1	3685
Pennsylvania	0	3557
Michigan	0	3290
Indiana	0	3249
Kansas	0.1	2562
Maryland	0	2263
Massachusetts	0	2185
New Mexico	0.1	2030
Alabama	0	1797
New Jersey	0	1777
Mississippi	0	1179
Delaware	0.1	820
New Hampshire	0	473
West Virginia	0	451
North Dakota	0	339
Connecticut	0	244
Rhode Island	0	197
Vermont	0	139
South Dakota	0	51
Total		**89049**

Below are the race/ethnicity options for those from Native Alaskan descent, who chose non-Hispanic when given the option.

Which Race Options Were Patients Given by Each State?

Patients from Native Alaskan Descent

Race/Ethnicity Options

- American Indian or Alaskan Native
- American Indian or Alaskan Native Non–Hispanic
- American Indian, Alaskan Native, Native Hawaiian, and Pacific Islander
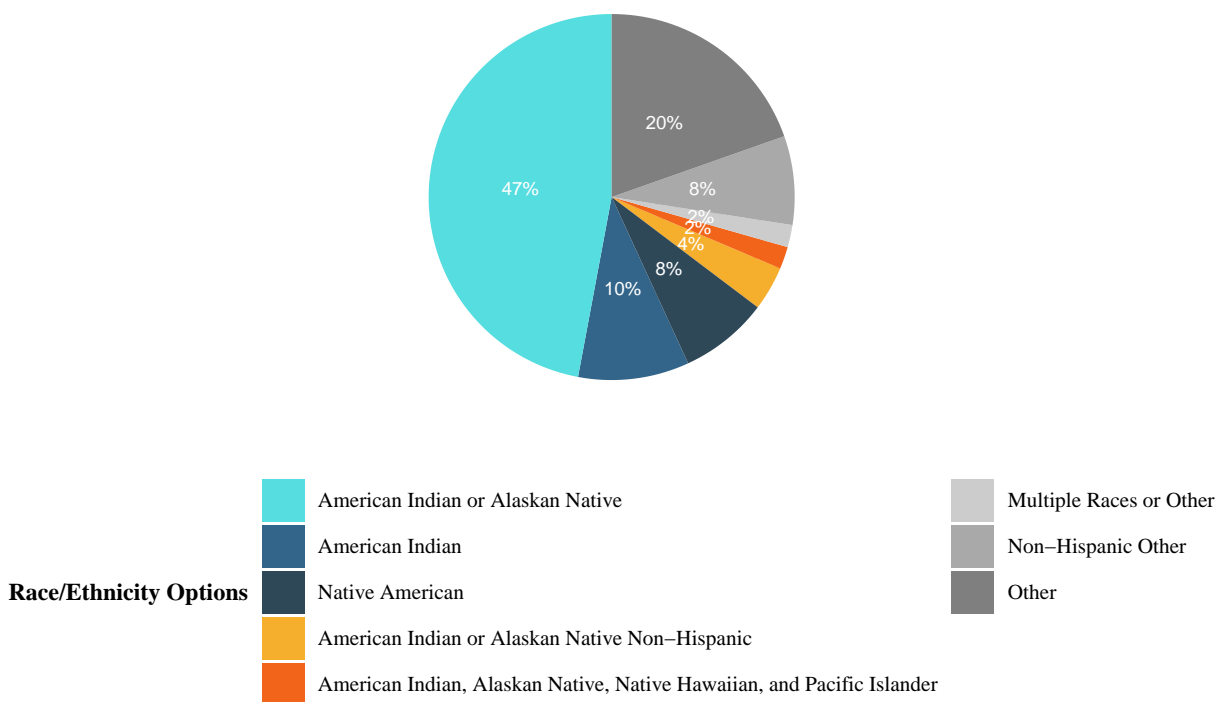- Multiple Races or Other
- Non–Hispanic Other
- Other

As you can see above, 45% of states did not publish race-specific COVID data reflecting the number of patients from Native Alaskan descent.

Below are the race/ethnicity options for those from Native American descent, and chose non-Hispanic when given the option.

Which Race Options Were Patients Given by Each State?

Patients from Native American Descent

Race/Ethnicity Options	
American Indian or Alaskan Native	Multiple Races or Other
American Indian	Non–Hispanic Other
Native American	Other
American Indian or Alaskan Native Non–Hispanic	
American Indian, Alaskan Native, Native Hawaiian, and Pacific Islander	

As you can see above, 31% of states did not publish race-specific COVID data reflecting the number of patients from Native American descent.

There are approximately 2,180,266 people of Native American or Native Alaskan descent in the United States (U.S. Census Bureau, 2018).

The table below show 16 states that did not report race-specific data about patients from Native American or Native Alaskan descent. These states have around 309,312 people from Native American or Native Alaskan descent. That value is 14% of the nation's Native American or Native Alaskan population. So, are 14% of the nation's people from Native American or Native Alaskan descent being described as "Other" in public COVID data?
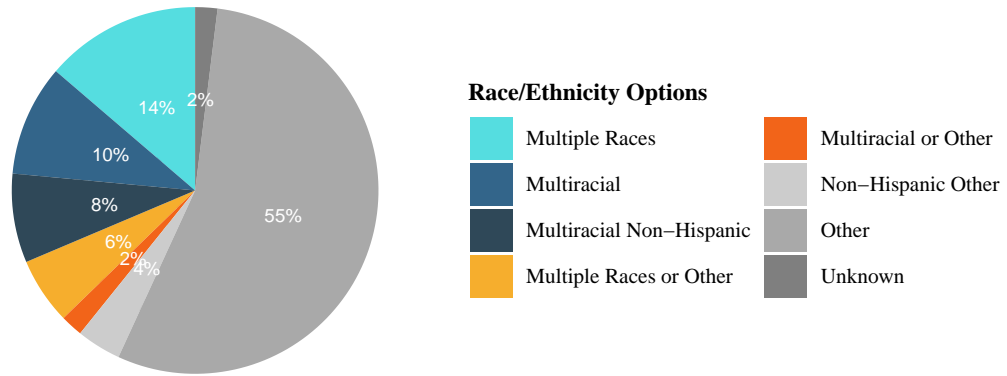
Table 7: Which States Would Report COVID Patients from Native American or Alaskan Descent as Other?

	State	Percent of Population	People of Native American or Alaskan Descent in the State
	Texas	0.2	70676
	New York	0.2	45465
	Florida	0.2	41492
	Missouri	0.4	23149
	Alabama	0.4	20450
	Pennsylvania	0.1	14367
	Massachusetts	0.2	10754
	Indiana	0.2	10560
	Maryland	0.2	9701
	New Jersey	0.1	8921
	New Hampshire	0.3	3538
	Delaware	0.4	3397
	Rhode Island	0.3	2772
	West Virginia	0.1	2535
	Hawaii	0.1	1976
Total			**269753**

Below are the race/ethnicity options for those from a mix of racial backgrounds, and chose non-Hispanic when given the option.

Which Race Options Were Specified by Each State?

Patients from a Mix of Racial Backgrounds

As you can see above, 61% of states did not publish race-specific COVID data reflecting the number of patients from a mix of racial backgrounds.

There are approximately 8,304,168 people from a mix of racial backgrounds in the United States (U.S. Census Bureau, 2018).

The table below shows 31 states that did not report race-specific data about mixed race patients. These states have around 4,365,961 people from a mix of racial backgrounds. That value is 53% of the nation's mixed race population. So, are 53% of the nation's people from a mix of racial backgrounds being described as "Other" in public COVID data?
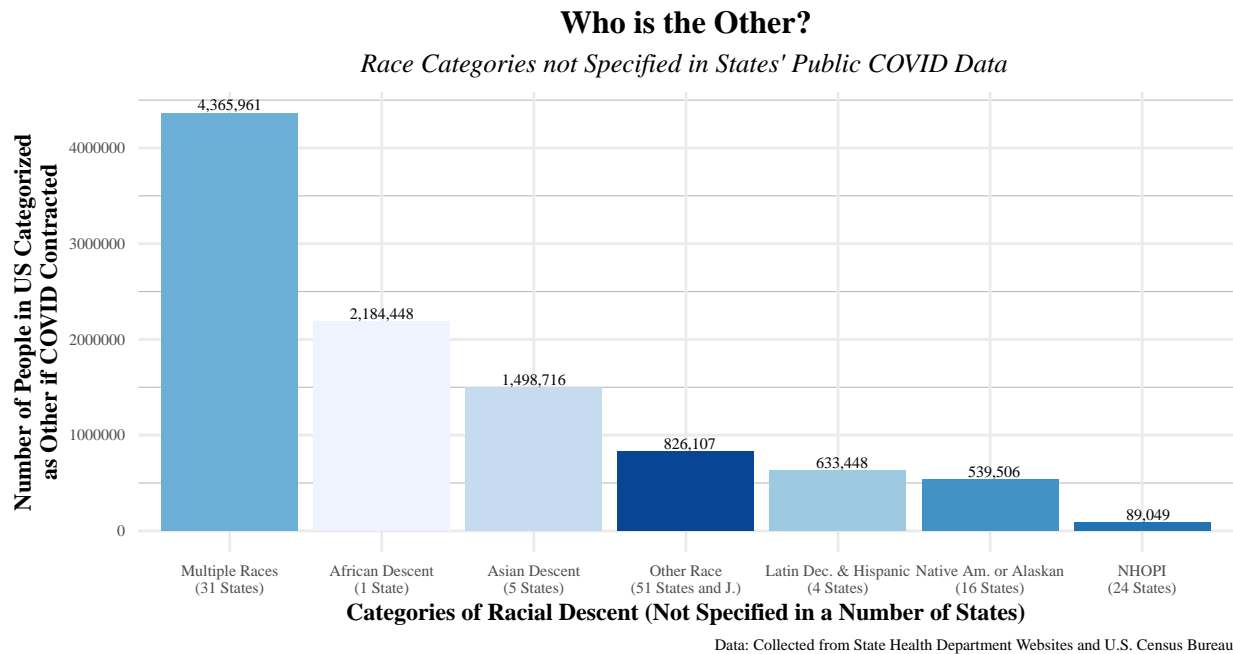
Table 8: Which States Would Report COVID Patients from a Mix of Races as Other?

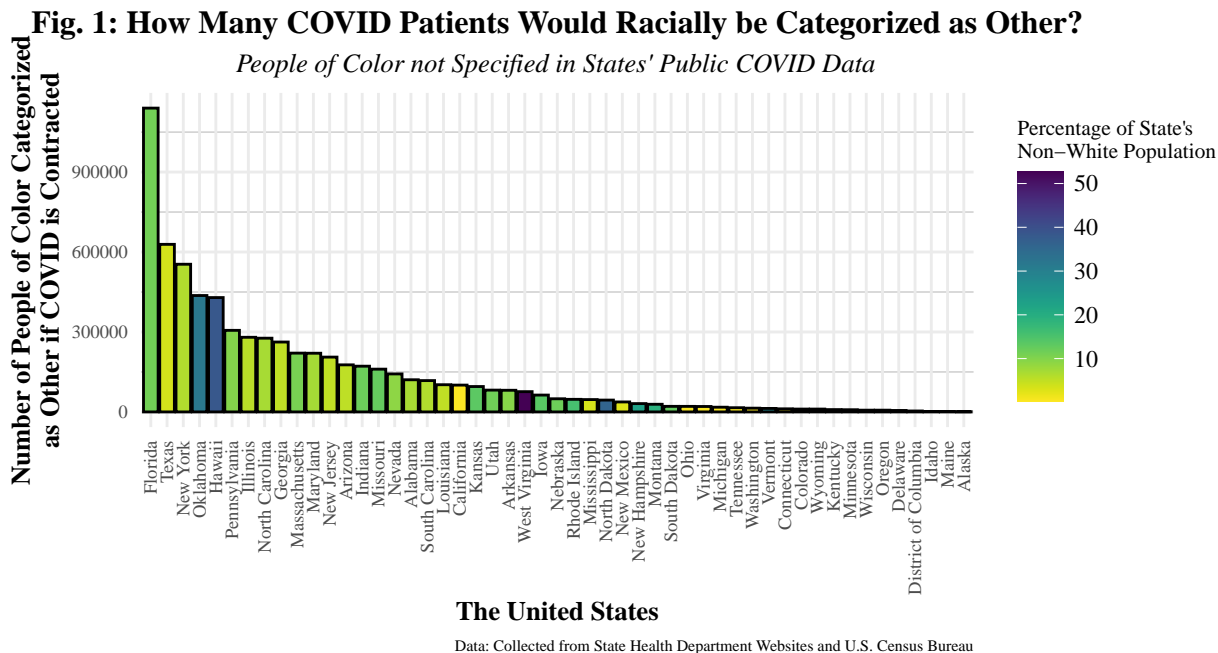|	State	Percent of Population	People from a Mix of Races in the State
	Texas	1.7	493947
	Florida	2	426855
	New York	2.1	406565
	Hawaii	19.3	274372
	Pennsylvania	2.1	266847
	Illinois	2	255782
	North Carolina	2.4	250758
	Georgia	2.2	228622
	Maryland	3	182424
	Arizona	2.3	166624
	New Jersey	1.7	154800
	Massachusetts	2.2	153883
	Indiana	2.1	142719
	Nevada	3.7	112499
	South Carolina	2	102278
	Alabama	1.9	91455
	Louisiana	2	91242
	Kansas	3.1	90708
	Arkansas	2.5	76651
	Utah	2.4	75539
	Iowa	1.9	59229
	Nebraska	2.4	47012
	Mississippi	1.4	40782
	West Virginia	1.8	31865
	New Mexico	1.5	31777
	Montana	2.6	27766
	New Hampshire	1.9	25321
	South Dakota	2.3	20554
	North Dakota	2	15016
	Vermont	1.9	11875
	Wyoming	1.8	10194
Total			**4365961**

The "Other"

With the states' public COVID data dashboards, almost 7 million people of color in the United States would be bundled into the "Other" race category if they contracted COVID. This is about 5% of the nation's non-white population (U.S. Census Bureau, 2018).
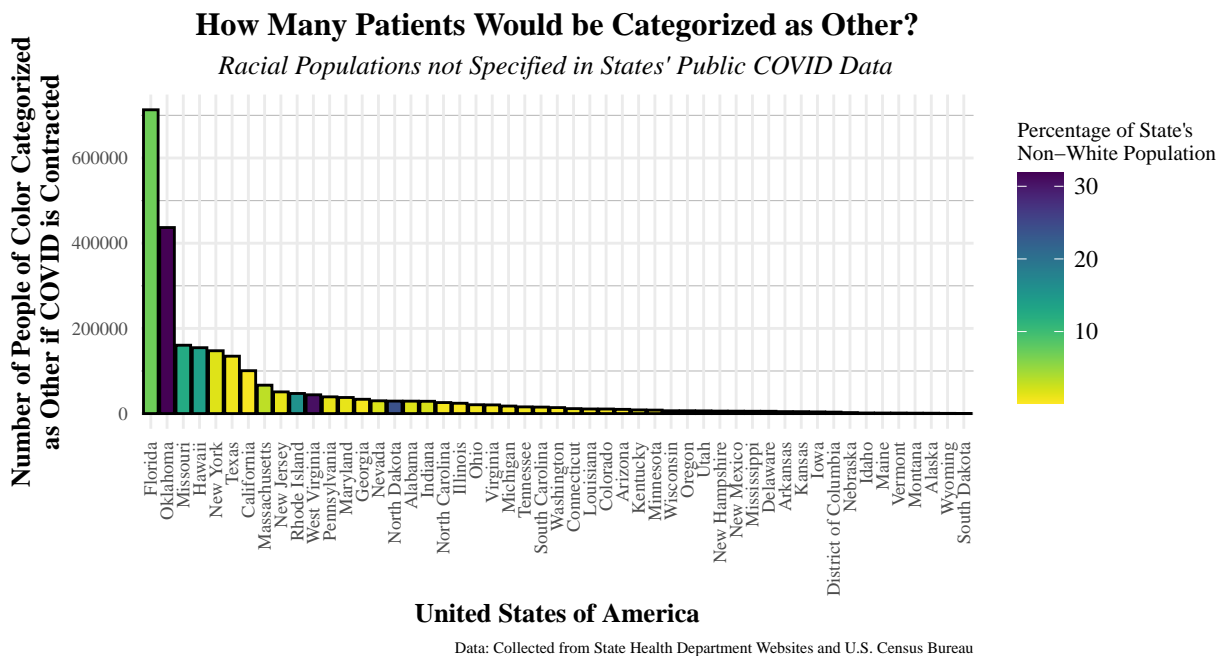
The following plot shows an approximate number of people whom would be categorized as "Other" in their state's public data if they contracted COVID.

Who is the Other?

Race Categories not Specified in States' Public COVID Data

Data: Collected from State Health Department Websites and U.S. Census Bureau
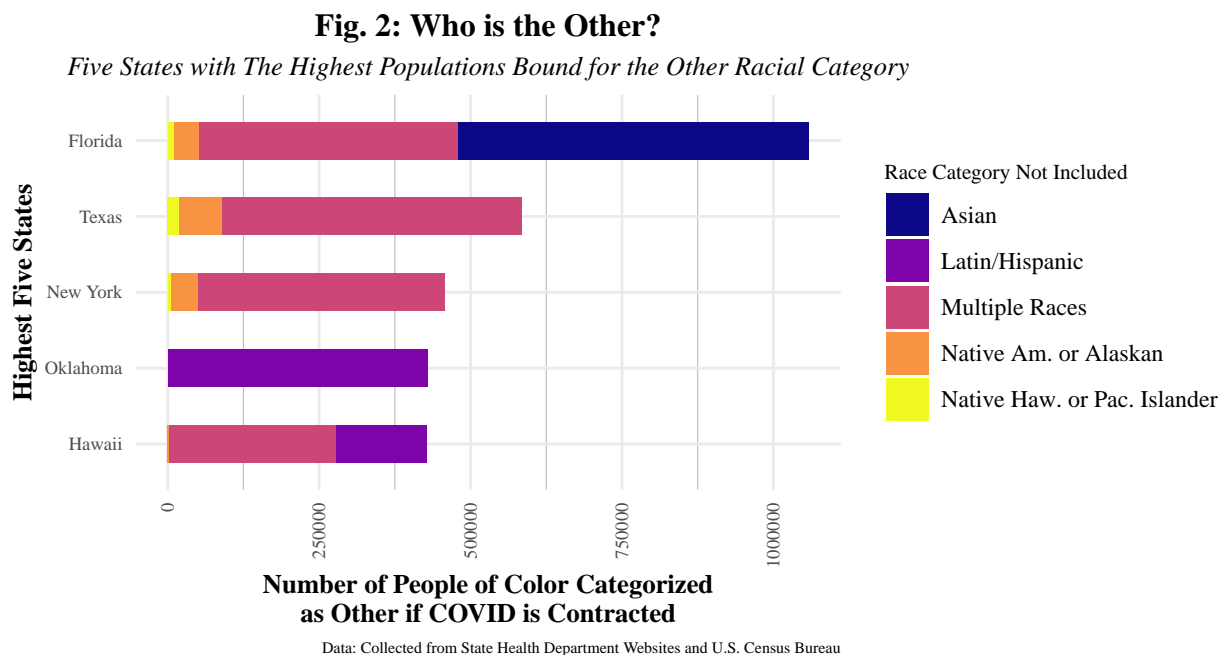
Below you can see the number of people of color from each state that would be classified as "Other" by the state's public COVID dashboard. The color notes the percentage that population represents out of all the people of color in that state. For example, you can see that Hawaii included categories to accuratley categorize about half of their population of color, while Florida did not include categories for about 25% of their population of color, which is a very large group.

Fig. 1: How Many COVID Patients Would Racially be Categorized as Other?

People of Color not Specified in States' Public COVID Data

Data: Collected from State Health Department Websites and U.S. Census Bureau

The following plot is the same as above, but, without the Multiple Races category

How Many Patients Would be Categorized as Other?

Racial Populations not Specified in States' Public COVID Data

Data: Collected from State Health Department Websites and U.S. Census Bureau

This plot demonstrates the five states with the largest population bound for the "Other" race category if they contracted COVID. For example, Florida is not reporting aggregated racial data for people of Asian descent, Native American or Alaskan descent, Native Hawaiian or Pacific Island descent, or people from a mix of racial descents. Those groups account for 1,058,270 people in the state of Florida.

Fig. 2: Who is the Other?

Five States with The Highest Populations Bound for the Other Racial Category

**Number of People of Color Categorized
as Other if COVID is Contracted**

Data: Collected from State Health Department Websites and U.S. Census Bureau

References

U.S. Census Bureau (2018). ACS Demographic and Housing Estimaes, 2018: American Community Survey 1-Year Estimates Data Profiles. Retrieved from https://data.census.gov/cedsci/table?q=race&tid=ACSDP1Y2018.DP05&moe=false&hidePreview=true.