Cynthia Begay
PM566 Final Report

Covid-19 Cases among American Indian and Alaska Natives by State

Introduction

The COVID Racial Data Tracker is a volunteer organization launched from The Atlantic and was created in collaboration between the COVID Tracking Project and the Boston University Center for Antiracist Research. On April 12, 2020, this team began collecting race and ethnicity data from every state and eventually US territories. The tracker is updated twice per week. This dataset includes the total number of COVID cases and number of reported deaths due to COVID by race and ethnicity. While this is the only COVID related dataset I could find that reports race and ethnicity on a national level, the project acknowledges this is not a comprehensive dataset on race and ethnicity.

As the COVID-19 pandemic unfolds, we are quickly learning about the disparity between race and COVID-19 cases. Currently Black, Indigenous, and Latinx people are most affected by COVID-19. According to COVIDtracking.com, Black people have the highest disparity and are dying 2.3 times the rate of white people. American Indian and Alaska Native (AIAN) have the second highest disparity and currently have a death rate of 65 per 100,000 people compared to non-Hispanic white 43 per 100,000. Historically, AIAN communities are left out of the conversation when it comes to reporting data, especially emerging data like COVID-19 cases/deaths. The current national conversation around AIAN COVID disparities has been about (my tribe) the Navajo Nation, which at its worst, had the highest infection rate in the country (higher than New York in its peak). It is important to understand the rest of the landscape in the US when it comes to tribal communities as 70% of the AIAN population live in urban cities (2010 Census). I am using this project to explore AIAN COVID rates by state to understand where the highest disparities lie in both frequency and proximity. The questions I will be exploring in this project are (1) how do AIAN COVID-19 rates/deaths compare between states, (2) which states reported the highest AIAN COVID cases in the US, and (3) which states reported the highest AIAN COVID deaths in the US?

Methods

The dataset was acquired through The COVID Tracking Project's website https://covidtracking.com/race. Data can be access here: https://docs.google.com/spreadsheets/d/e/2PACX-1vR_xmYt4ACPDZCDJcY12kCiMiH0ODyx3E1ZvgOHB8ae1tRcjXbs_yWBOA4j4uoCEADVfC1PS2jYO68B/pub?gid=43720681&single=true&output=csv. Although an API version of the dataset was advertised on the website, the race related data was only available as a .csv file. Variables assessed were date, state, state code, total COVID cases, AIAN COVID cases, total deaths, AIAN deaths. Day, month, year, and state code variables were created in the csv file prior to reading in the data. The cases total variable was converted from a character variable to an integer in excel.

Data wrangling

Data variables were assessed using the str() function to understand the dataset variable classes; rows and columns were also assessed. The "State" variable is the only character

variable; all other variable are integer variables. Before data exploration could be performed, the date variable was recreated into a date, month, and year variable in excel. All variables were checked to ensure data reflected between 4/12/2020 and 10/07/2020 when the data was retrieved. Data was limited to variables of interest: day, month, year, state, cases_AIAN, cases_total, deaths_AIAN, deaths_total. There are 2601 observations, and 8 variables present.

The head and tail of the dataset was assessed to ensure the data was read in properly. The cases total variable was validated against current reported data to ensure that the dataset did not exceed what has been reported by the CDC.

Missing values were assessed with the dim function. Missing values in this dataset are both reflected as actual missing data not reported by the state. Any "missing" data prior to the first case reported for each variable, which can be interpreted as a 0 or N/A value. All states are reported except Alabama, Delaware, Florida, Hawaii, Indiana, Massachusetts, Maryland, New Hampshire, New Jersey, New York, Pennsylvania, and Texas–these states reported N/A for AIAN Covid-19 cases for all dates reported. (Non-reports are most likely due to racial misclassification.)

A sub dataset was created to reflect the initial and most recent values as of 4/12/2020 and 10/04/2020, respectively. Covid-19 numbers reported in this dataset are addictive and not a reflection of daily updates.

Results

AIAN COVID Cases by State

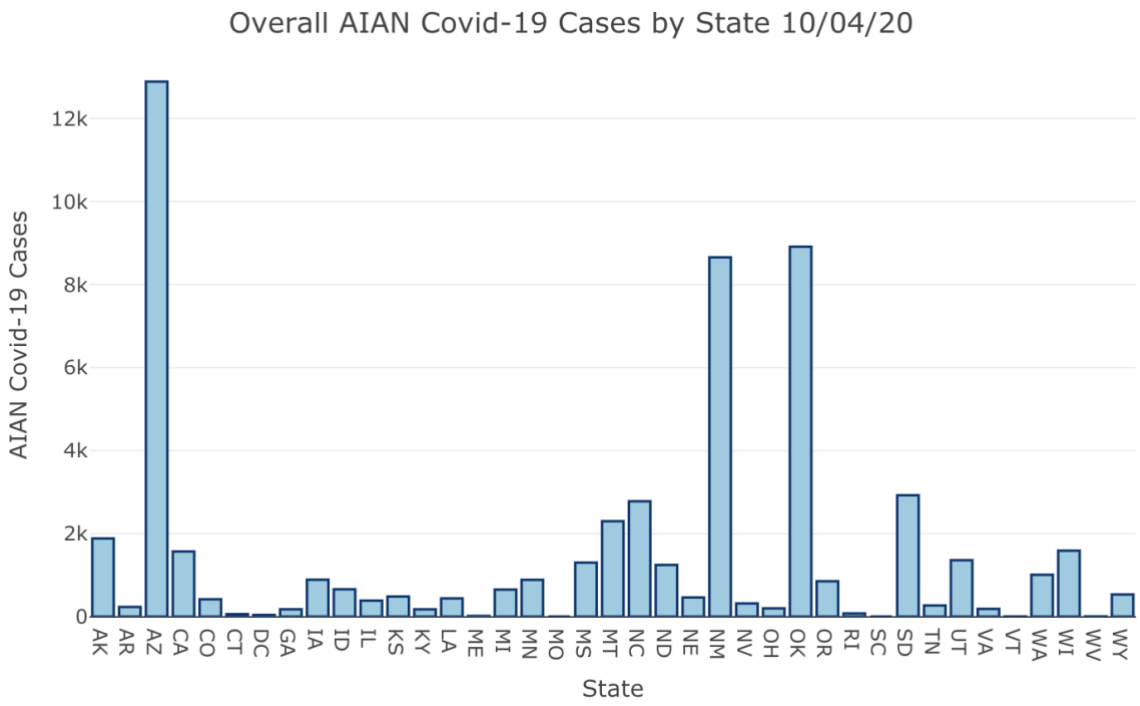


Figure 1: Histogram AIAN Covid-19 Cases by State

Heat Maps

These heat maps show AIAN Covid-19 cases and deaths over time by state. The reported numbers are cumulative by date. Although Oklahoma did not report a higher number of AIAN cases in the beginning of the pandemic, this heat map shows cases increased around August. AIAN cases were generally higher throughout the pandemic for Arizona and New Mexico.

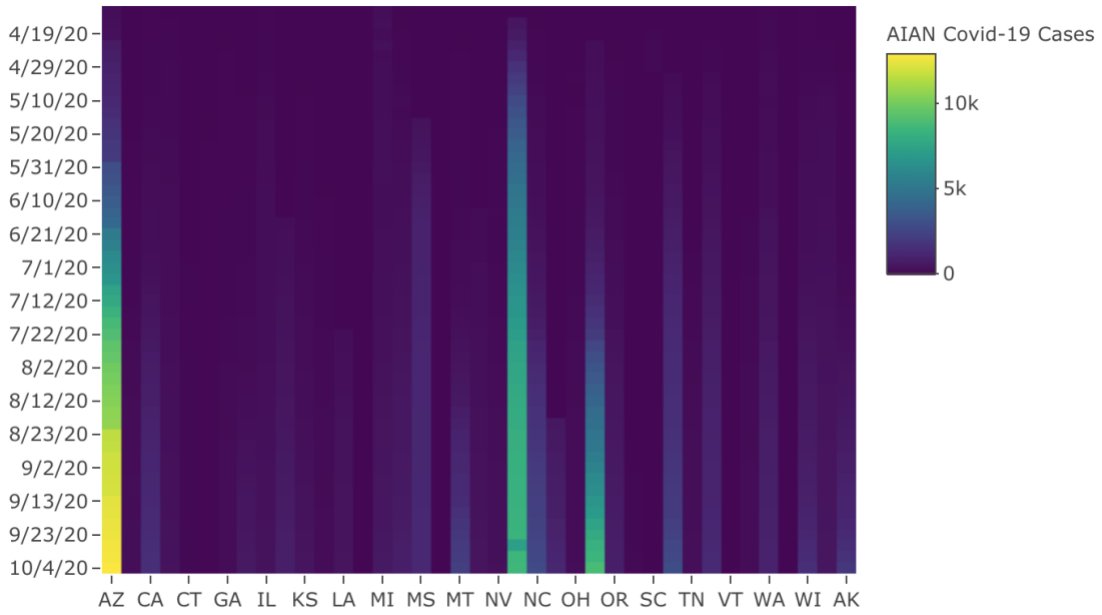


Figure 2: Heat Map of AIAN Covid-19 Cases by State

Upon initial inspection of the reported AIAN Covid-19 deaths, Illinois reported 512 AIAN which was most likely a misreporting. The corrected heat map is displayed without the outlier. Although Oklahoma had the 3rd highest reported AIAN Covid-19 cases, the deaths reported were minimal. Arizona and New Mexico reported higher AIAN deaths due to Covid-19 than any other state.

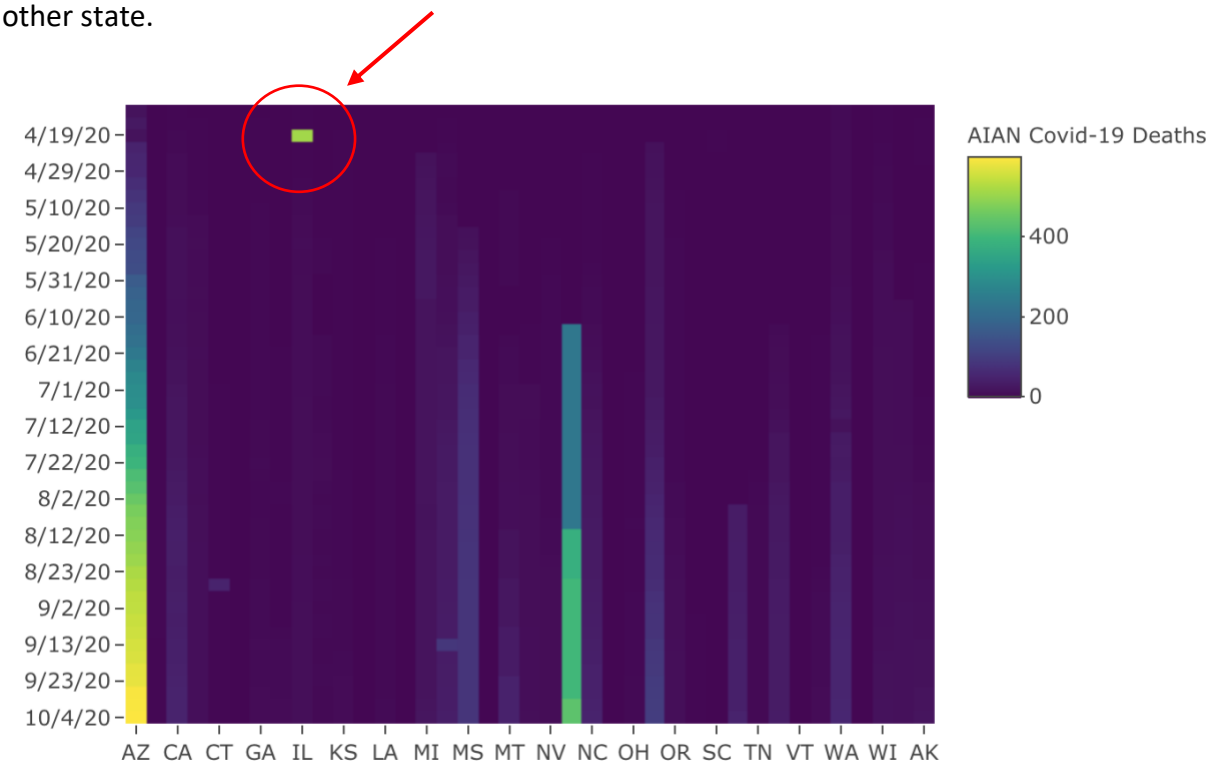


Figure 3: AIAN Covid-19 Deaths by State

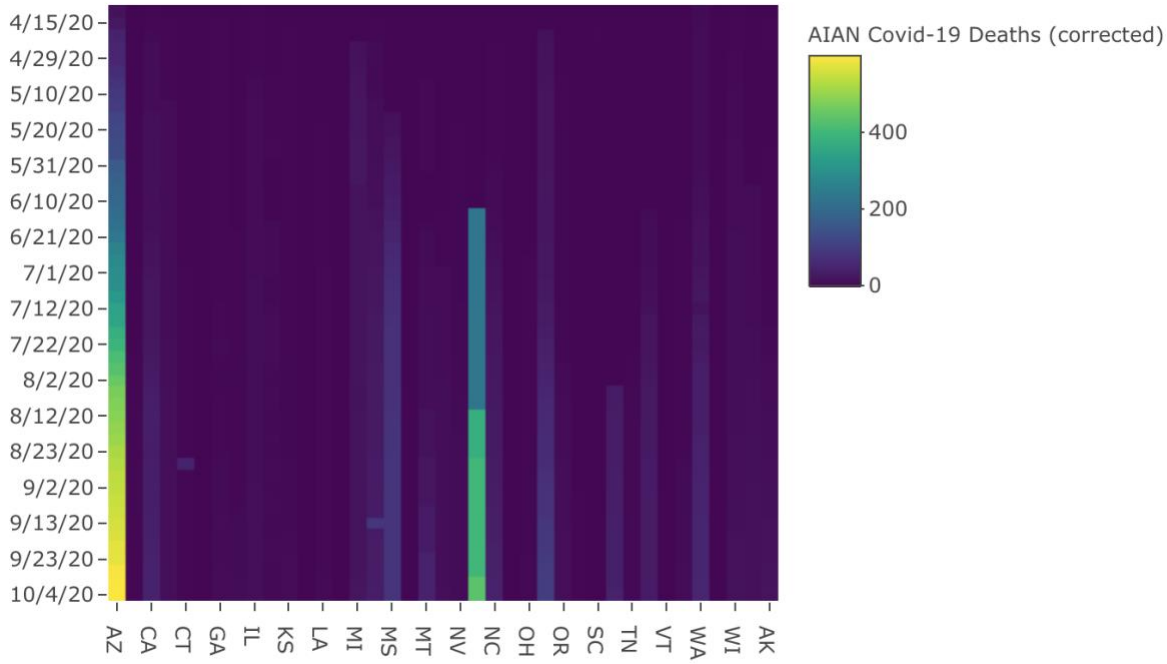


Figure 4: AIAN Covid-19 Deaths by State (Corrected)

Line Plots

According to the line plot by states, the highest cases reported were in Arizona, New Mexico and Oklahoma. All 3 states reported cases as of April 12, 2020. The lowest cases were in the Northeast and East Coast states—this may be due to the lower number of AIAN population.

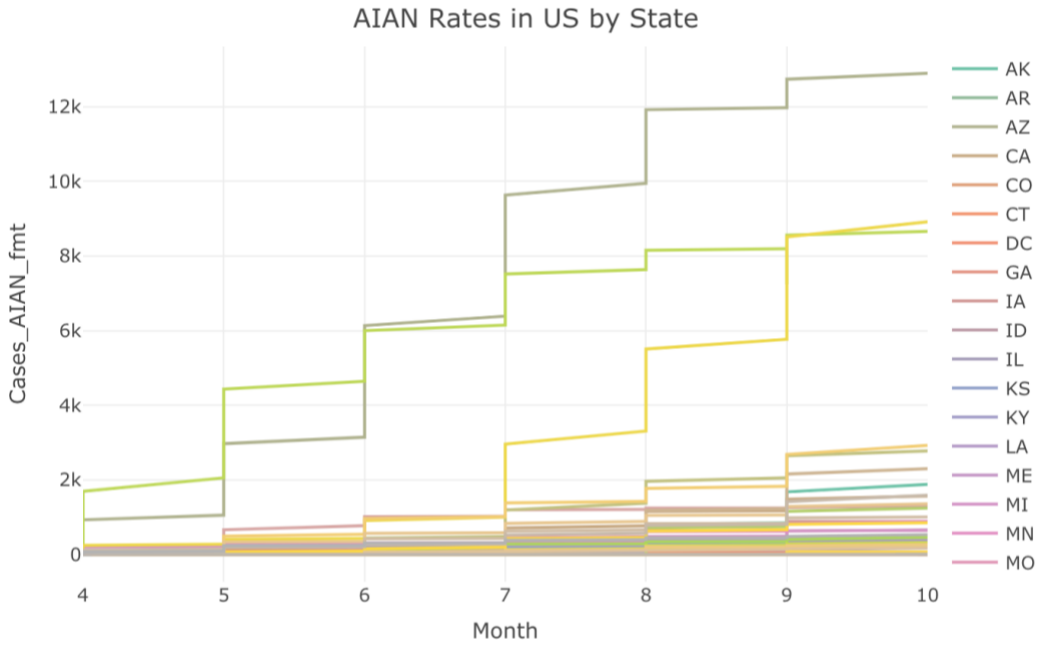


Figure 5: AIAN Covid-19 Cases by State 4/12/2020-10/04/2020

According to the individual line plot, the majority of cases reported in Arizona were between June and August. (This timeline coincides with the national reporting of the high rates on the Navajo Nation.) Arizona had the highest AIAN death rates followed by New Mexico. Mississippi reported slightly higher deaths rates just above Oklahoma.

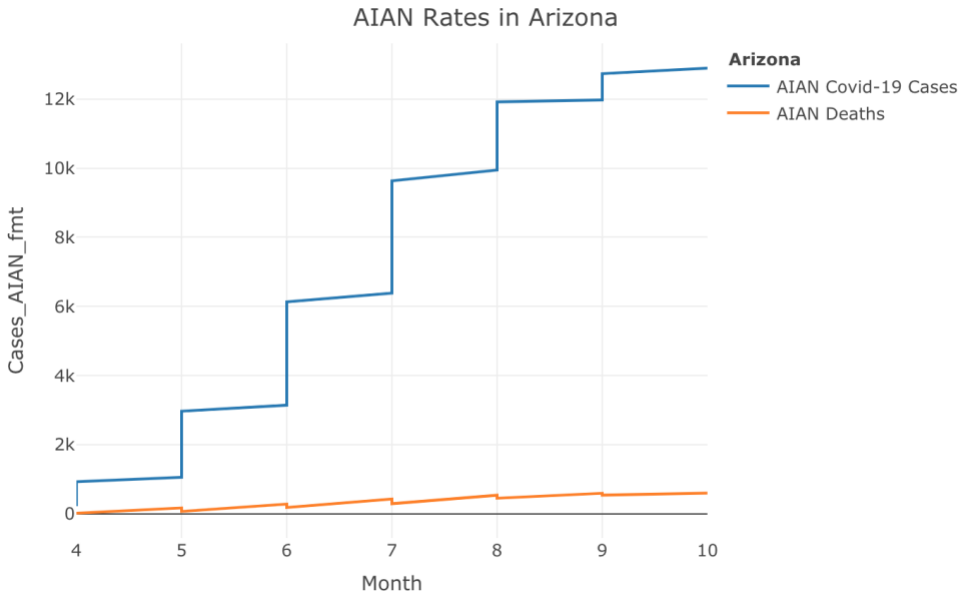


Figure 6: AIAN Covid-19 Cases in Arizona

According to the individual line plot for New Mexico, the highest reported cases were between April and August. The reported cases may have tapered off due to the contact tracing efforts organized by the Pueblo tribal communities.

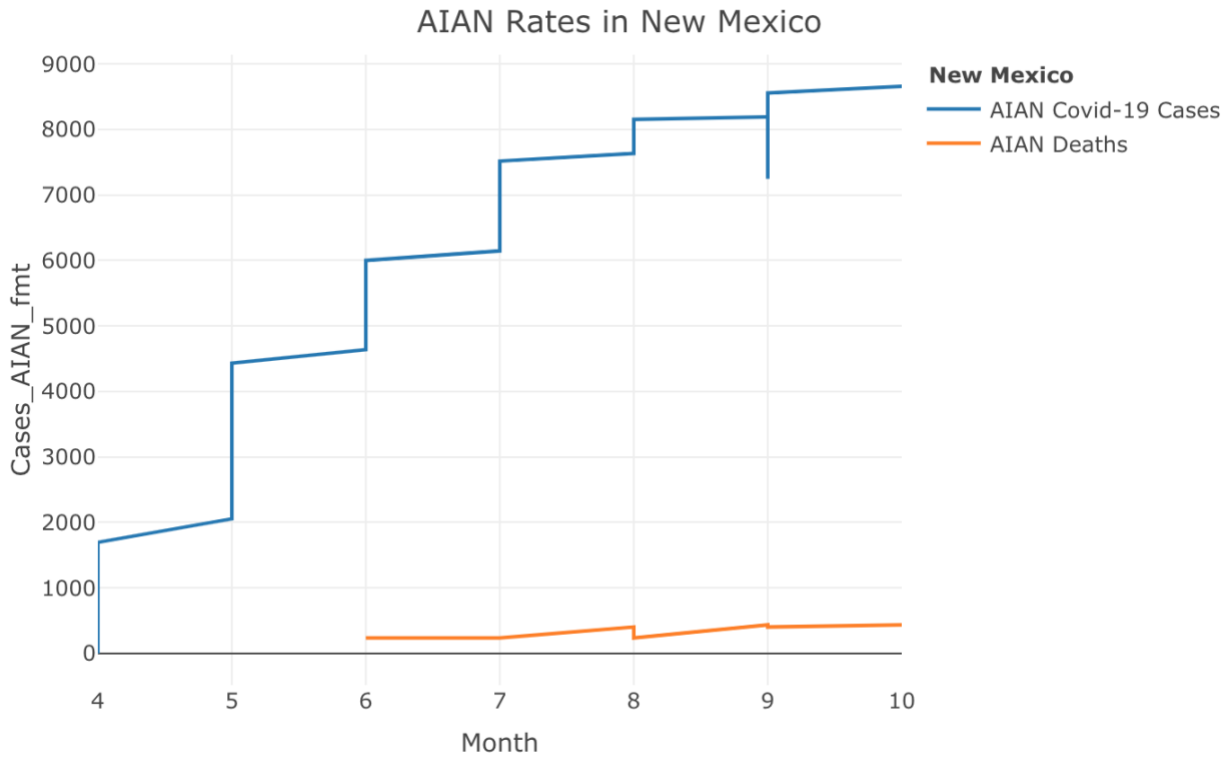


Figure 7: AIAN Covid-19 Cases in New Mexico

According to the individual scatter plot for Oklahoma, most cases were reported between July and September—there was a delay in COVID cases reaching this part of the country. Although Oklahoma reported higher AIAN cases, according to the line plot, their AIAN death rates are similar to the rest of the country.

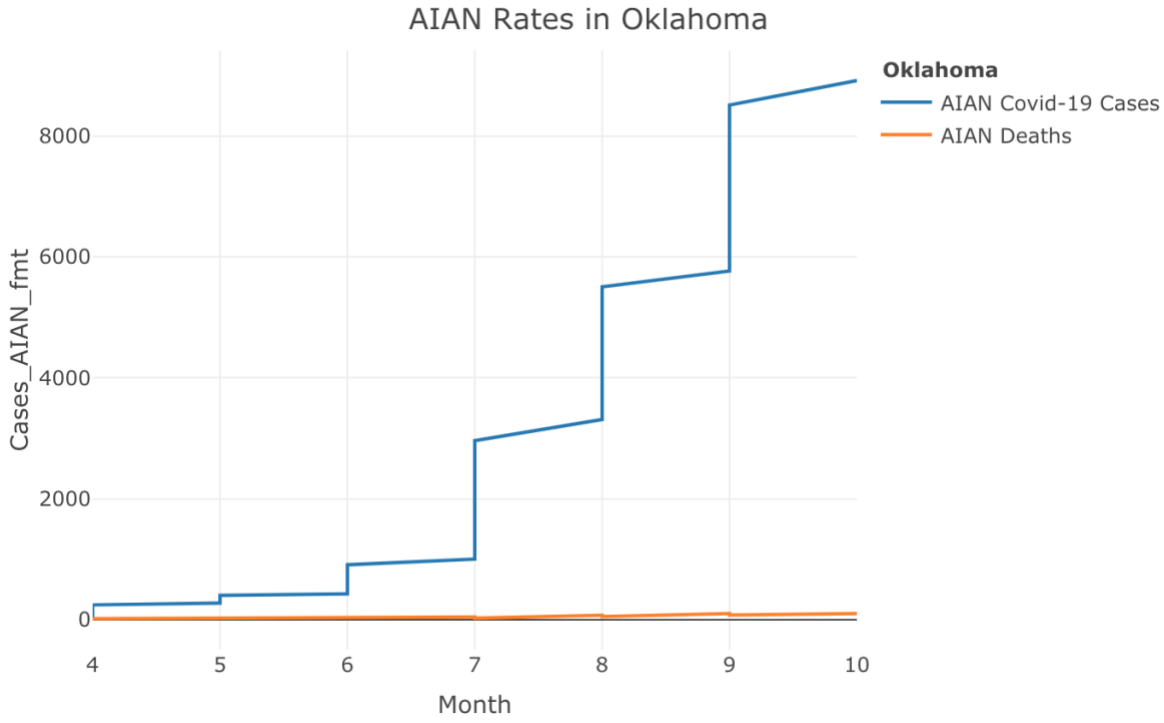


Figure 8: AIAN Covid-19 Cases in Oklahoma

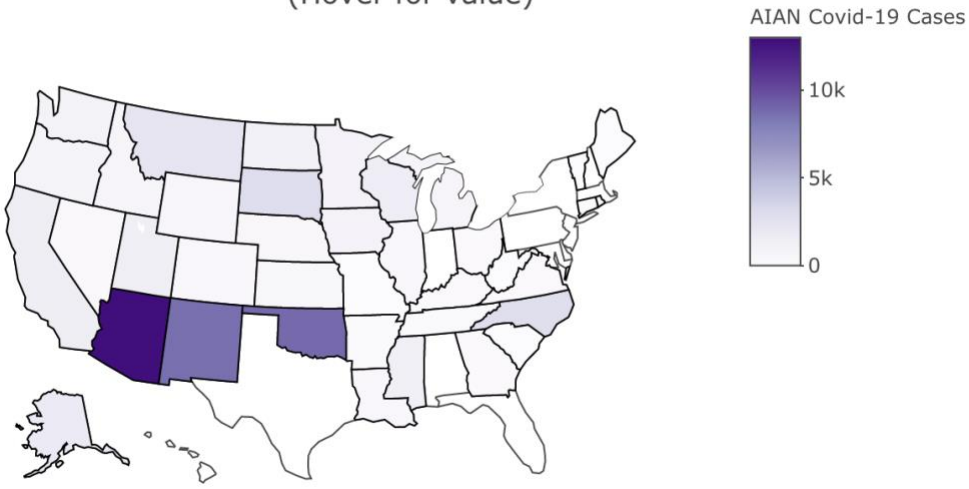
Maps

Maps show geographic proximity of reported cases by state. This can be an important indicator given each state's AIAN population has a different rural versus urban population. In addition, resources for AIAN communities and tribes vary by state. An important piece to note is Michigan initially reported 512 AIAN cases. This disparity was not detectable in the other graphs produced. However, by October 4th, we can see the disparity in cases in among Arizona, New Mexico and Oklahoma.

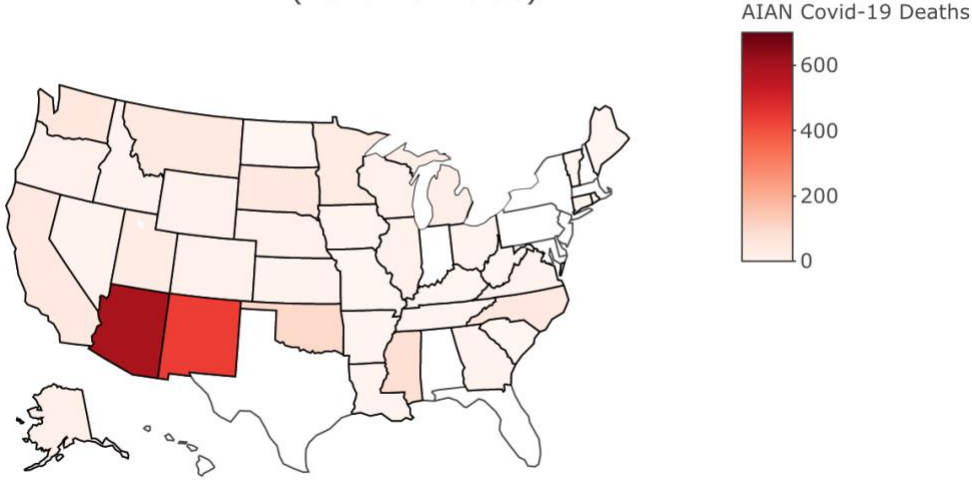
The disparity of AIAN deaths due to Covid-19 were highest in Arizona and New Mexico. Although Oklahoma reported higher AIAN cases initially, those cases did not translate into higher AIAN deaths.

AIAN Covid-19 Rates by State as of April 12, 2020
(Hover for value)

AIAN Covid-19 Rates by State as of October 04, 2020
(Hover for value)



AIAN Covid-19 Deaths by State as of October 04, 2020
(Hover for value)



Conclusion

The highest rates of COVID-19 cases among American Indian and Alaska Natives is in Arizona, New Mexico, and Oklahoma. The biggest limitation to this preliminary analysis is addressing missing values as cases were not reported as "0" but rather as missing. This project brought to light that this dataset may not provide enough accurate reporting to understand how Covid-19 has affected AIAN populations nationwide. Because states were not uniformly reporting cases and deaths, it is unknown if the states who were not included in this analysis categorized AIAN as “other” or simply did not report on the cases at all. In addition, it is unknown where the deaths are reported from. If they are reported from the general population, there may also be racial misclassification and underreporting of AIAN. Additionally, this dataset did not specify which races are 2 or more races; these cases were not included in the analysis.

As mentioned in the introduction, the Navajo nation reported the highest infection rate in April. The Navajo nation is located in Arizona, New Mexico, and a small part of Utah; this may explain the high rates reported in Arizona and New Mexico. Furthermore, there were Pueblo tribes starkly affected by the disease which are located in New Mexico. Although California and New York have the highest population of AIAN per capita, all 3 states reporting the highest COVID-19 cases have a large population of rural tribal communities. Many people in these rural tribal communities do not have access to internet or running water both, of which are necessary to receive evolving pandemic messaging in real time and to take safety precautions such as washing your hands.

The highest death rates for Covid-19 were in Arizona and New Mexico. Covid-19 death rates were high in Arizona but tapered off in New Mexico and were very minimal in Oklahoma. Additionally, death rates were not reported until June for New Mexico and Oklahoma. This may have been due to the robust contract tracing done in the Pueblo tribal communities as reported by New York Times. They were able to test tribal members quickly and get them care before cases became too serious.

This analysis helps provide insight on how states were able to address Covid-19 among the AIAN population. Understanding the frequency of cases versus deaths, also provided context given the different new sources reporting on Tribal communities. It is important to understand trends such as these given, we are entering our 3rd wave of the pandemic. This can help inform Tribal governments and state governments how to appropriately respond and reduce the burden of disease on these communities.