



COVID Killers

Impact Predictor for California Residents



COVID Killers - Project Overview

The COVID Killers team will provide context to what factors increase the spread of the virus in California and provide predictions for future cases in the state.

This Presentation will:

- Describe The Project's Focus
- Explain why the Project was chosen
- Review data sources
- Identify Question to Answer with the Data
- Describe the Project's data exploration
- Describe the Projects' data analysis
- Display the Project's data application



COVID Killers - Purpose

Reason for selection: The spread of COVID-19 is actively reshaping society.

COVID-19 opinions, models, and datasets are plentiful, yet collectively we are unable to answer with high confidence even the most basic questions.

The Project intends to help provide insight to the COVID-19 pandemic in California and seeks to model future impacts of COVID-19 in the state: We will provide context to what factors increase the spread of the virus in California and a Future prediction of new cases.



COVID Killers - Data Sources

The Project data includes COVID cases for the State of California aggregated with virus testing, mobility, and vulnerability data.

Data was sourced from web research from the following sites:

State daily case counts (New York Times):

<https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv>

State daily testing:

<https://covidtracking.com/api/v1/states/daily.json>

State daily mobility

<https://www.google.com/covid19/mobility/>

<https://www.apple.com/covid19/mobility>

<https://github.com/descarteslabs/DL-COVID-19>

Race/Age/Ethnicity data (L.A. Times):

<https://github.com/datadesk/california-coronavirus-data>

Vulnerability index (Surgo Foundation): ...



COVID Killers - Questions to Answer with the Data

What will be the future number of California COVID-19 cases between June 15 and June 30, 2020?

1. What factors increase/decrease risk of spreading COVID-19?
2. What counties are at highest risk?
3. Can we provide county specific insights for users to make data augmented/guided decisions about their own risk?





COVID Killers - Machine Learning

1. Variable Identification – Identified the Target output(s) from the data and Features (inputs) of the predictive model:

(List: Targets and Features TBD)

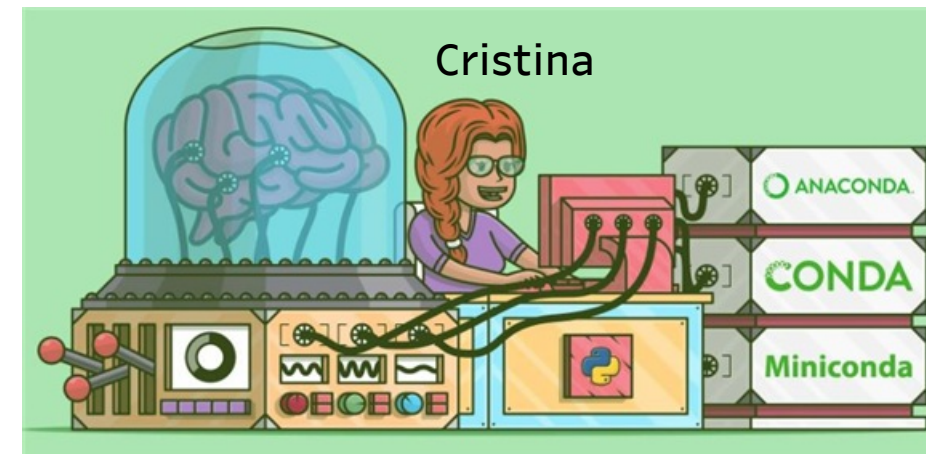
2. Univariate Analysis – Calculated variables, Categorical Variables

3. Bi-variate Analysis: identify relationships between variables with by plotting data and using regression modeling.

4. Missing values treatment – remove missing data

5. Outlier treatment – remove outliers in the data (if actual)

6. Variable transformation – scale the data..





COVID Killers - Machine Learning - ARIMA

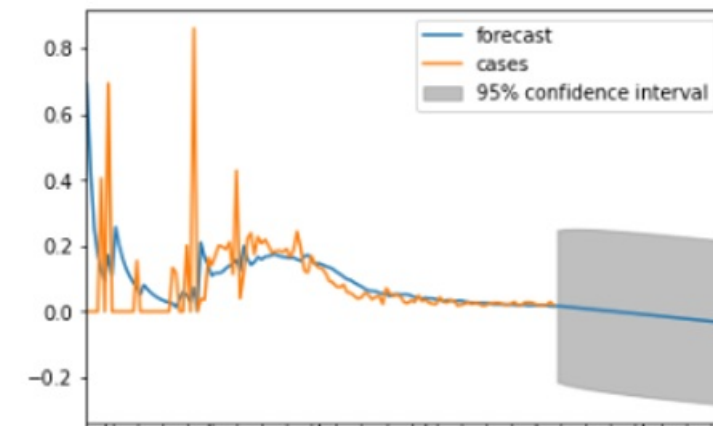
ARIMA - AutoRegressive Integrated Moving Average; a forecasting algorithm based on the premise that past values of a time series, alone can be used to predict the future values.

A Univariate Time Series Forecasting model has been developed using previous daily new cases to predict future case counts within a 95% confidence interval.

The model predicts that California cases will trend downward through July.



```
# predict data  
results.plot_predict(1,180)
```



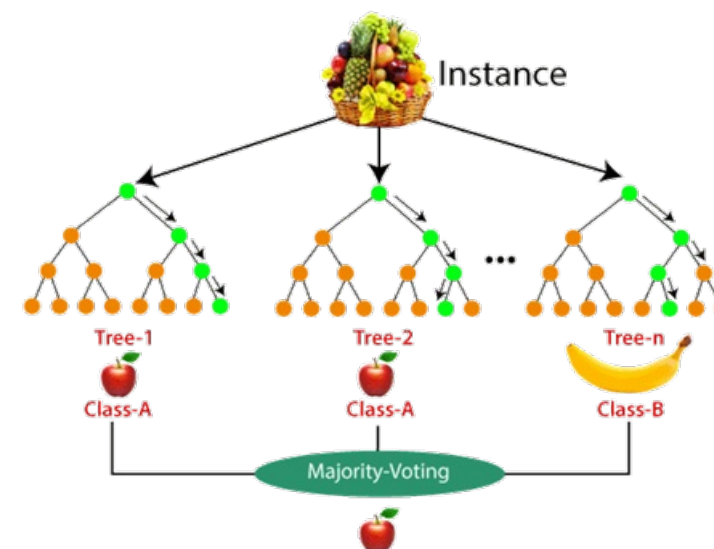


COVID Killers - Machine Learning - RandomForest

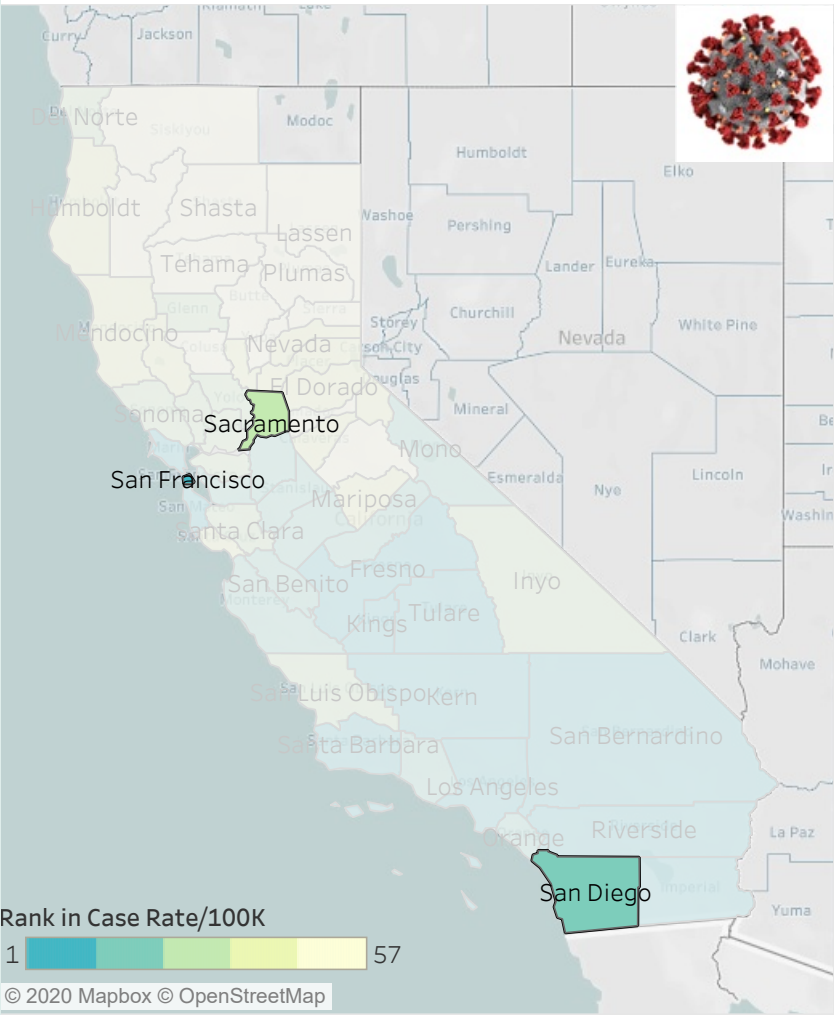
Random Forest - a classification algorithm comprised of many decision trees. It uses feature randomness to build individual trees intended to create an uncorrelated "forest" of trees whose collective prediction is more accurate than that of an individual tree.

A Multivariate Random Forest model has been developed using previous daily new cases and state mobility data, and to predict future case counts within a 95% confidence interval.

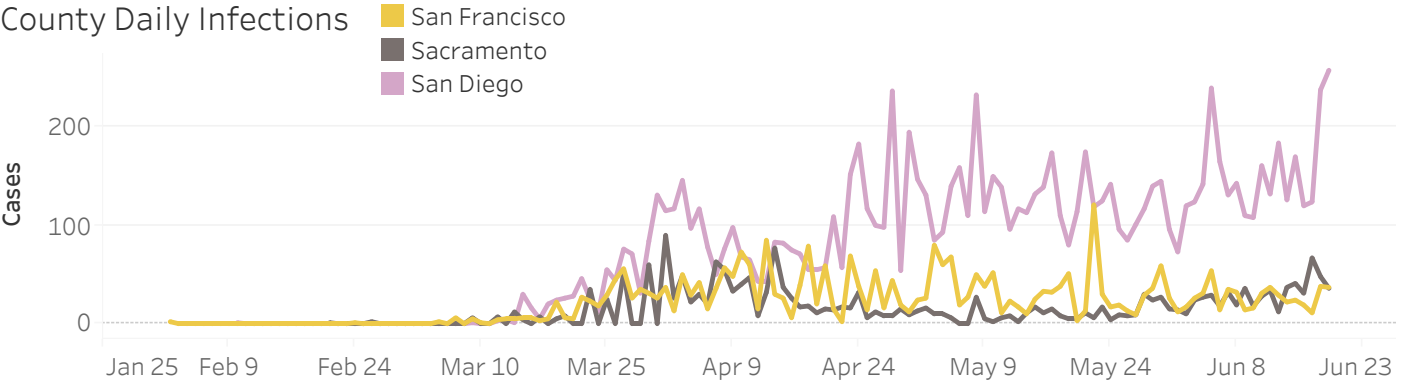
The model predicts that California cases will trend downward through July.



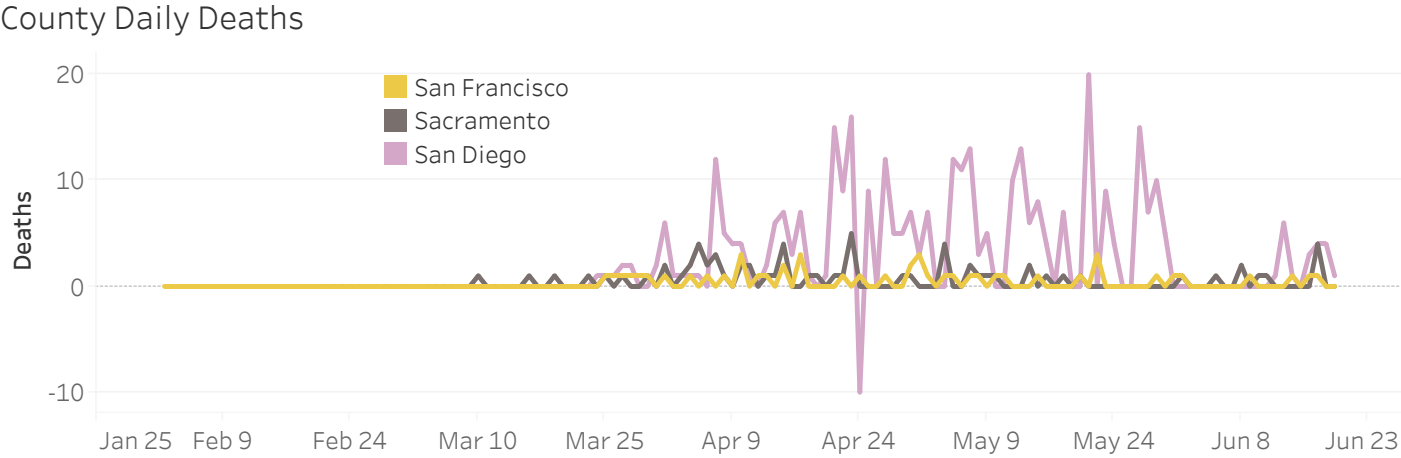
All California Counties



County Daily Infections



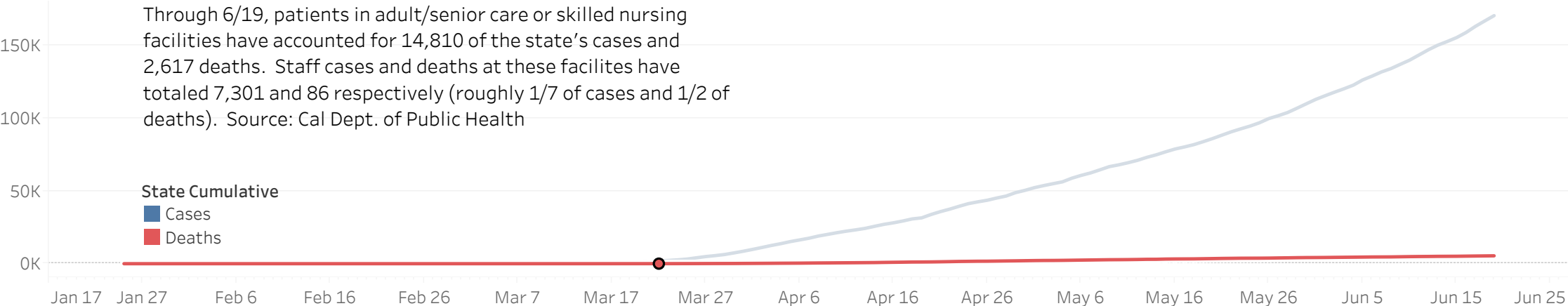
County Daily Deaths



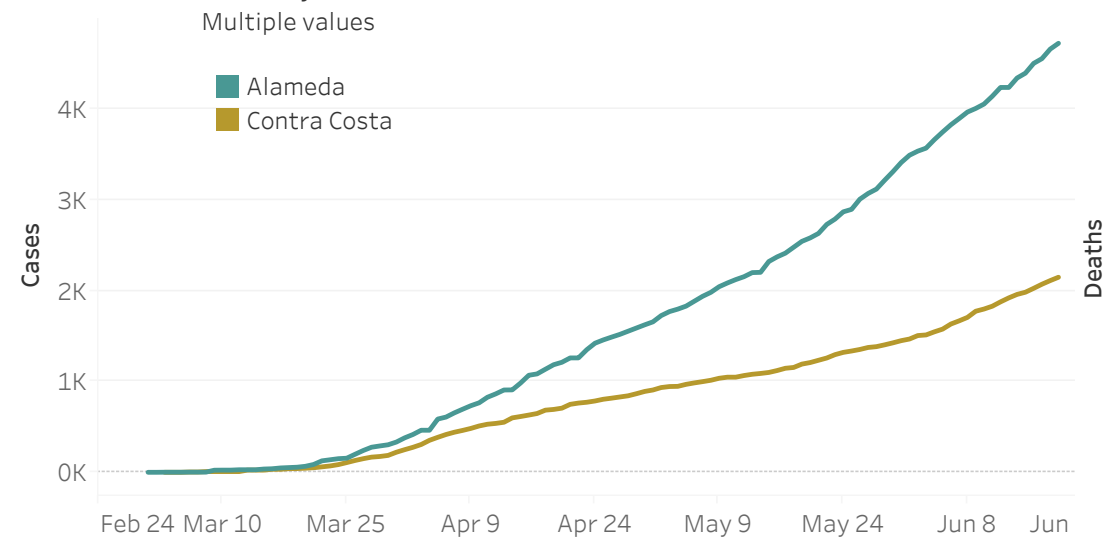
State	County	Case Rank	Cases/100K	Cases	Deaths	Deaths/100K	Population
CA	San Francisco	10	347	3,060	47	5	881,549
	San Diego	13	312	10,413	332	10	3,338,330
	Sacramento	34	127	1,975	67	4	1,552,058

Cumulative Tally of Cases and Deaths for California

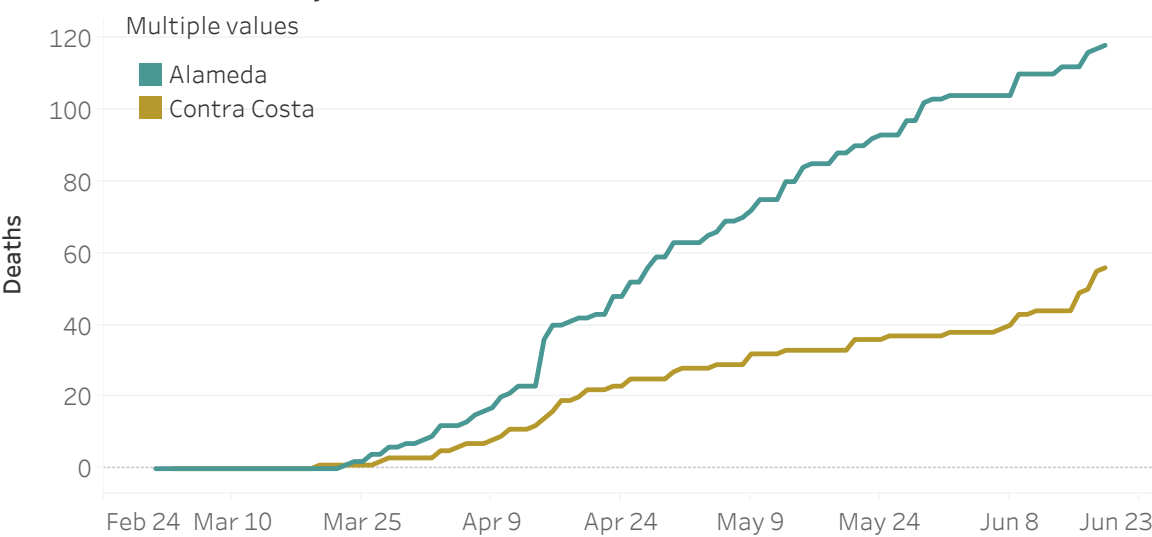
Through 6/19, patients in adult/senior care or skilled nursing facilities have accounted for 14,810 of the state's cases and 2,617 deaths. Staff cases and deaths at these facilities have totaled 7,301 and 86 respectively (roughly 1/7 of cases and 1/2 of deaths). Source: Cal Dept. of Public Health

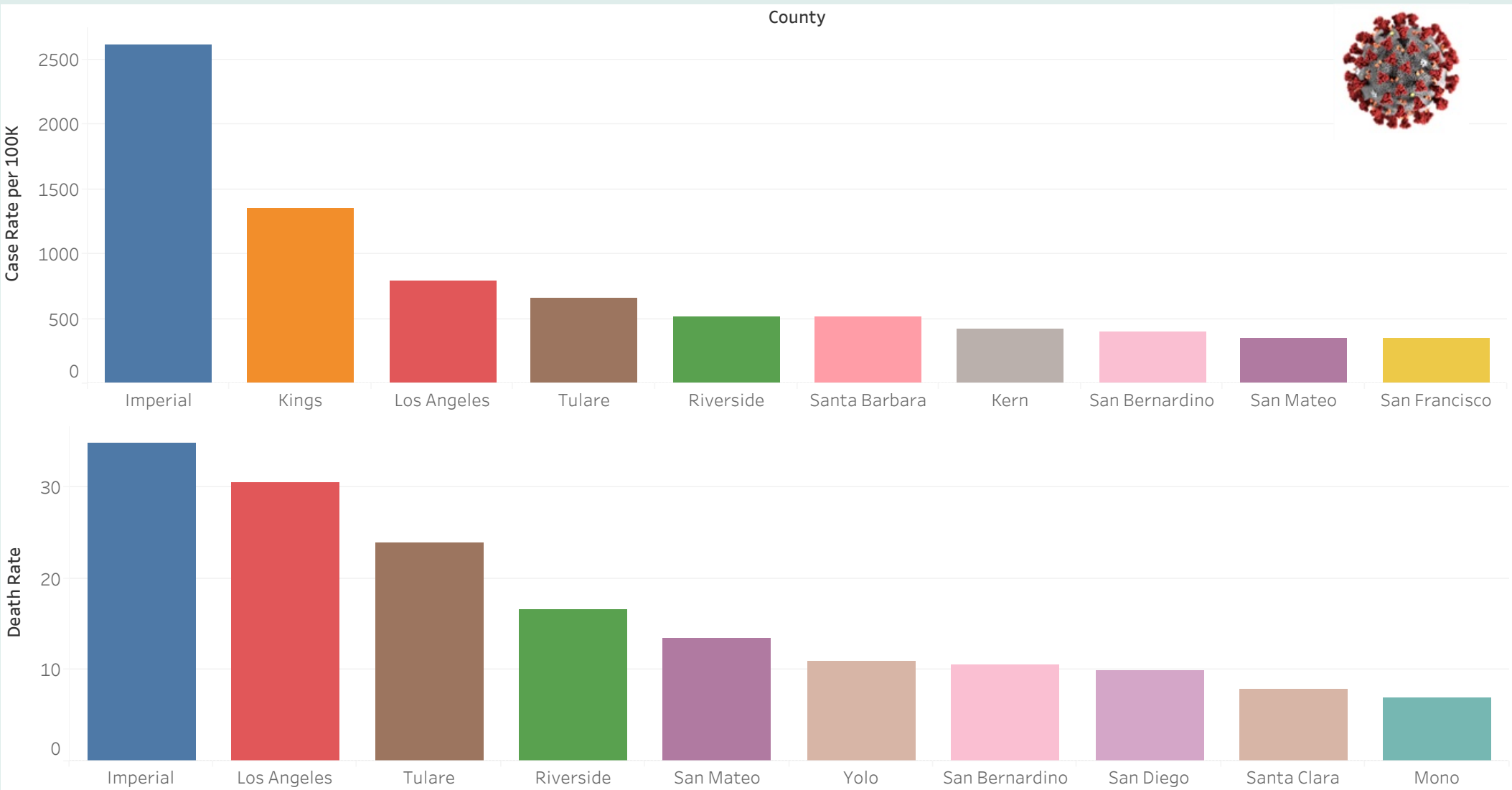


Cumulative Cases by Selected Counties



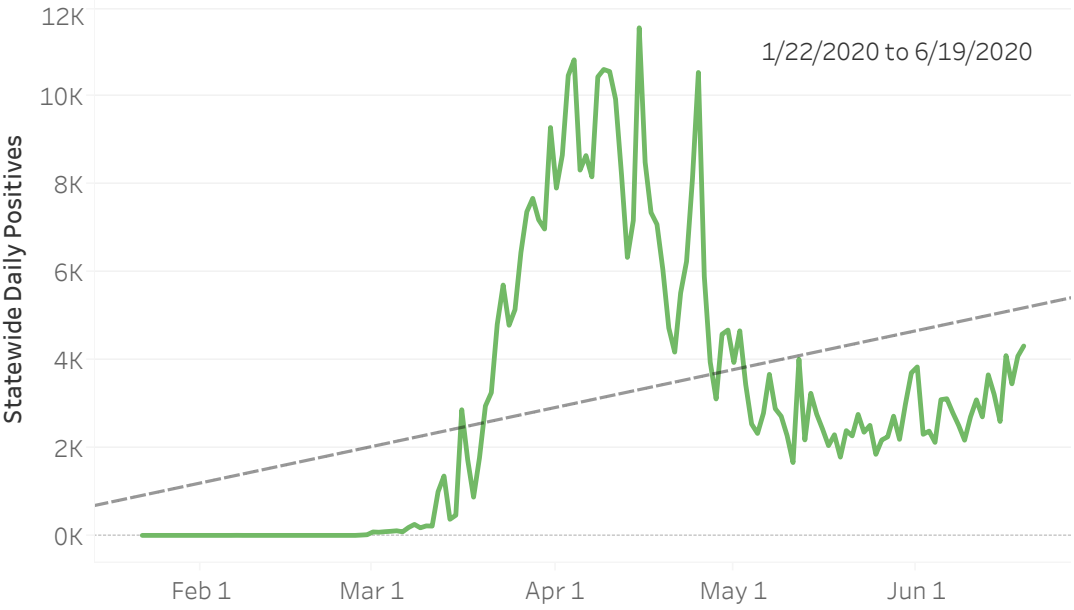
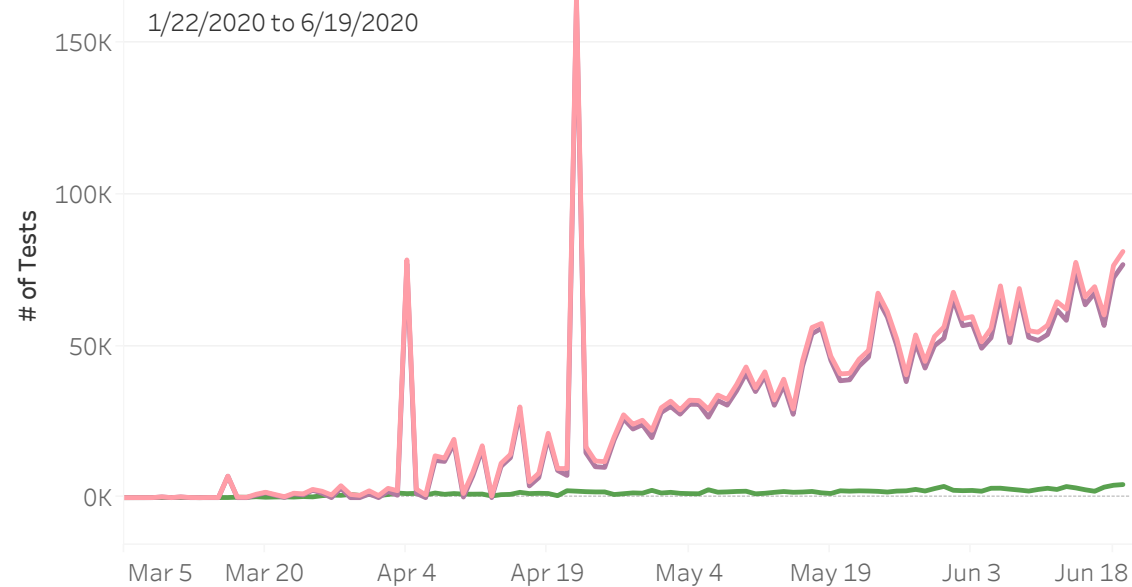
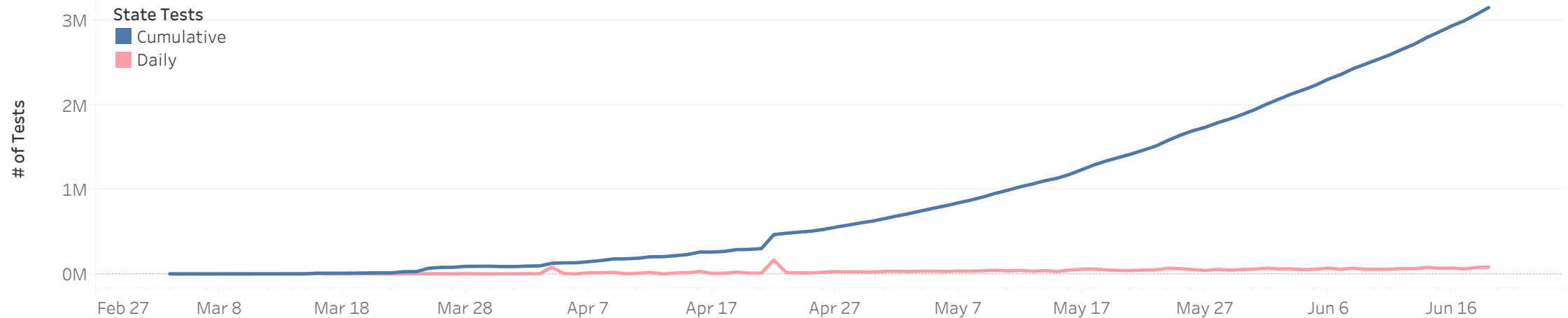
Cumulative Deaths by Selected Counties



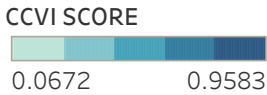


COVID Testing in CA by Date

1/22/2020 to 6/19/2020
and Null values



COVID-19 Community Vulnerability Index



Socioeconomic
0 to 0.99968163

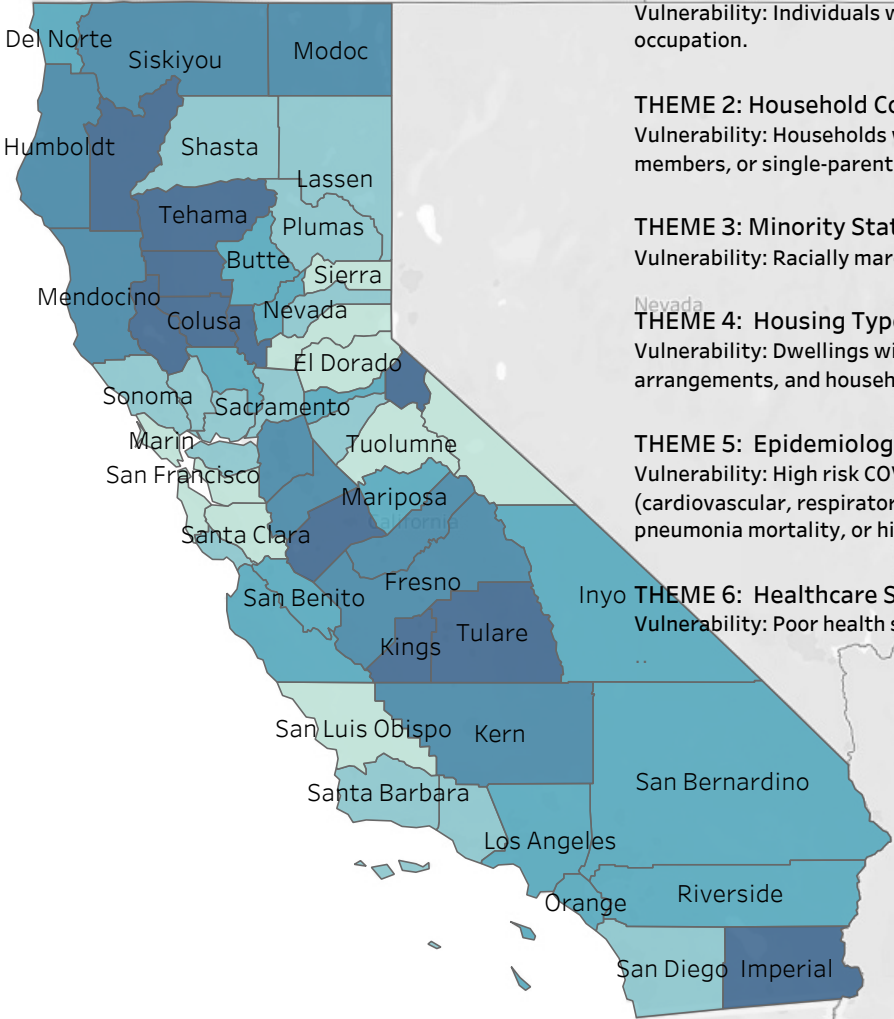
Household Composition
To Null

Minority Status & Langua..
To Null

Housing & Transportation
To Null

Epidemiological
0 to 1

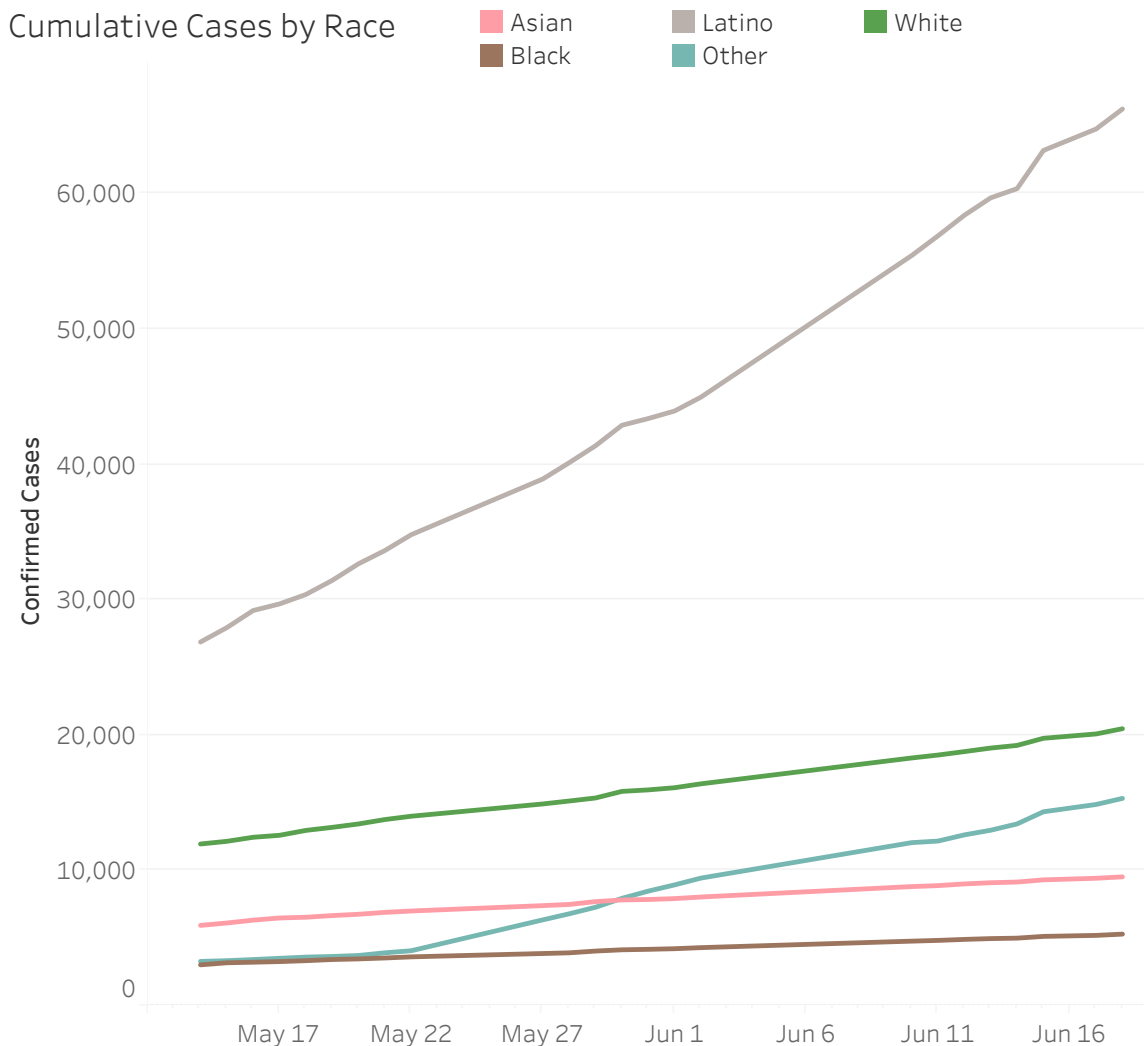
Healthcare System
To Null



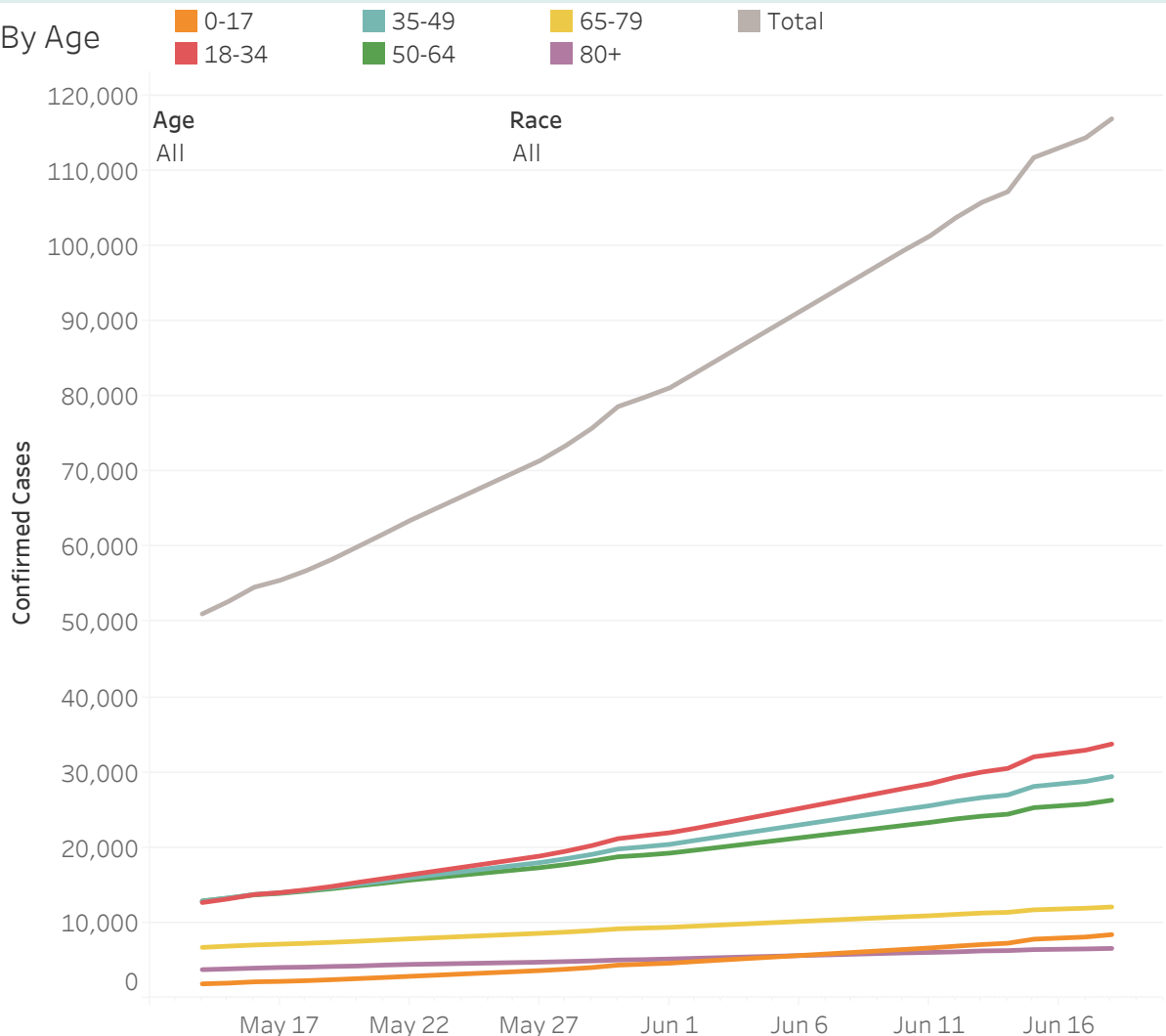
- Vulnerability: Individuals with low income, educational attainment, and no occupation.
- THEME 2: Household Composition & Disability.**
Vulnerability: Households with elderly (over 65), young (under 17), or disabled members, or single-parent homes.
- THEME 3: Minority Status & Language**
Vulnerability: Racially marginalized groups or those with limited English proficiency.
- THEME 4: Housing Type & Transportation**
Vulnerability: Dwellings with multiple units, mobile, group, or crowded living arrangements, and households without access to transport.
- THEME 5: Epidemiological Factors**
Vulnerability: High risk COVID-19 populations with underlying conditions (cardiovascular, respiratory, immunocompromised, obesity, diabetes) high flu and pneumonia mortality, or high population density.
- THEME 6: Healthcare System Factors**
Vulnerability: Poor health system capacity, strength and preparedness.



Cumulative Cases by Race

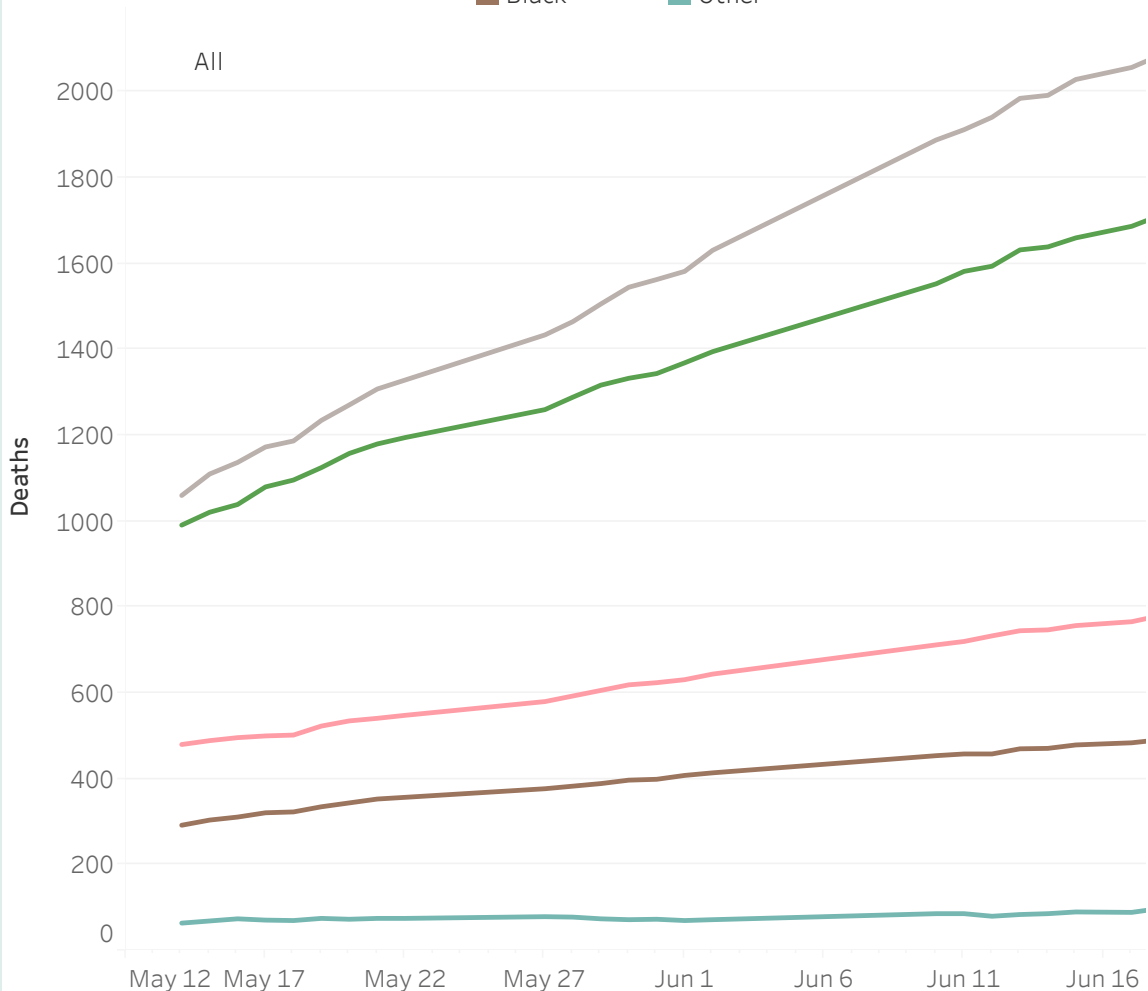


By Age



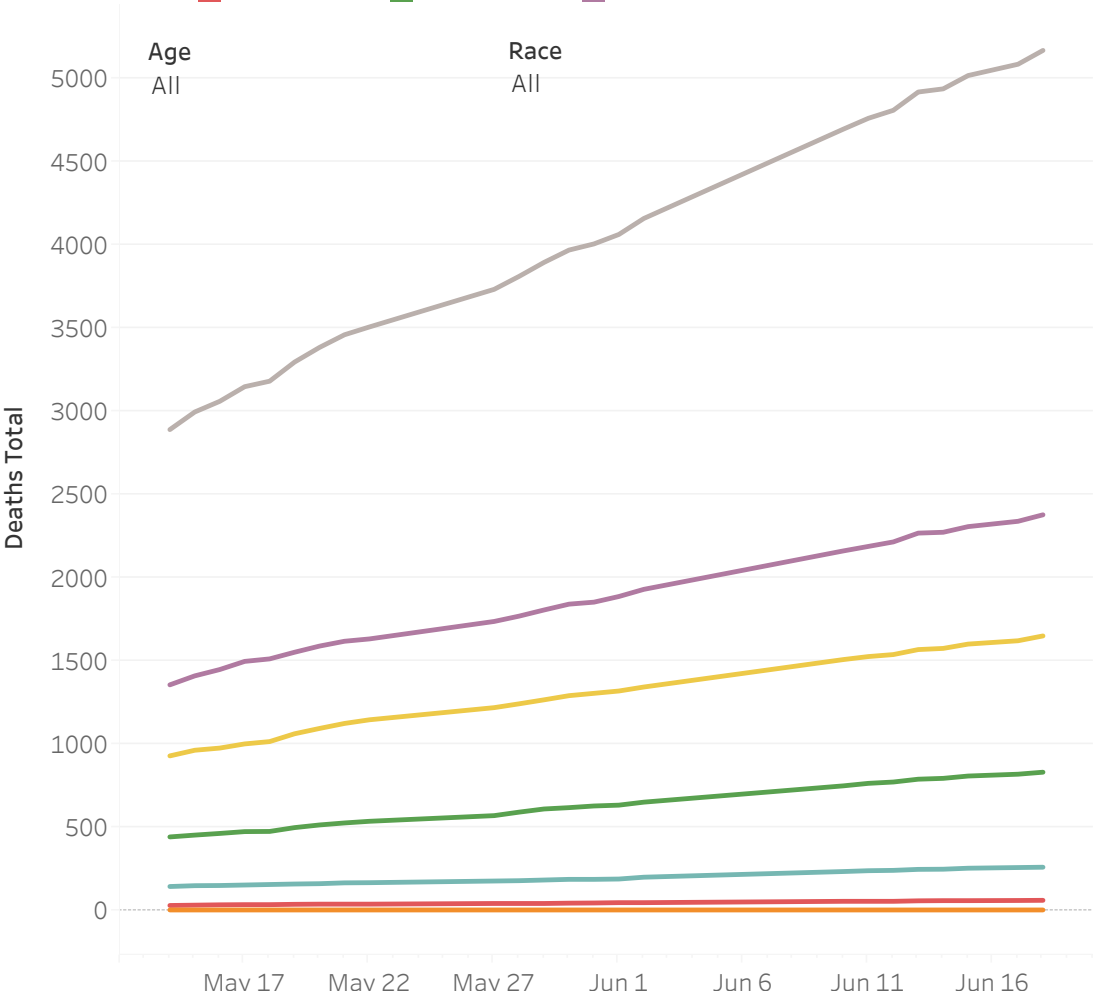
Cumulative Deaths by Race

Asian Black Latino Other White



By Age

0-17 18-34 35-49 50-64 65-79 80+ Total



Cases by Race through June 18

Confirmed Cases by Race



Age

Total

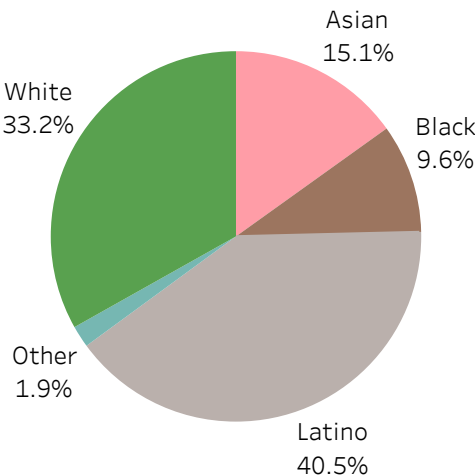
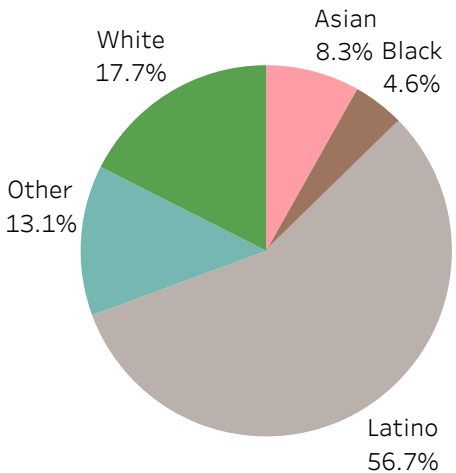
Deaths by Race through June 18

Total Deaths



Age

Total



Cases: Approximately (29%) missing race/ethnicity
Deaths: Approximately (1%) missing race/ethnicity

COLOR of COVID-19

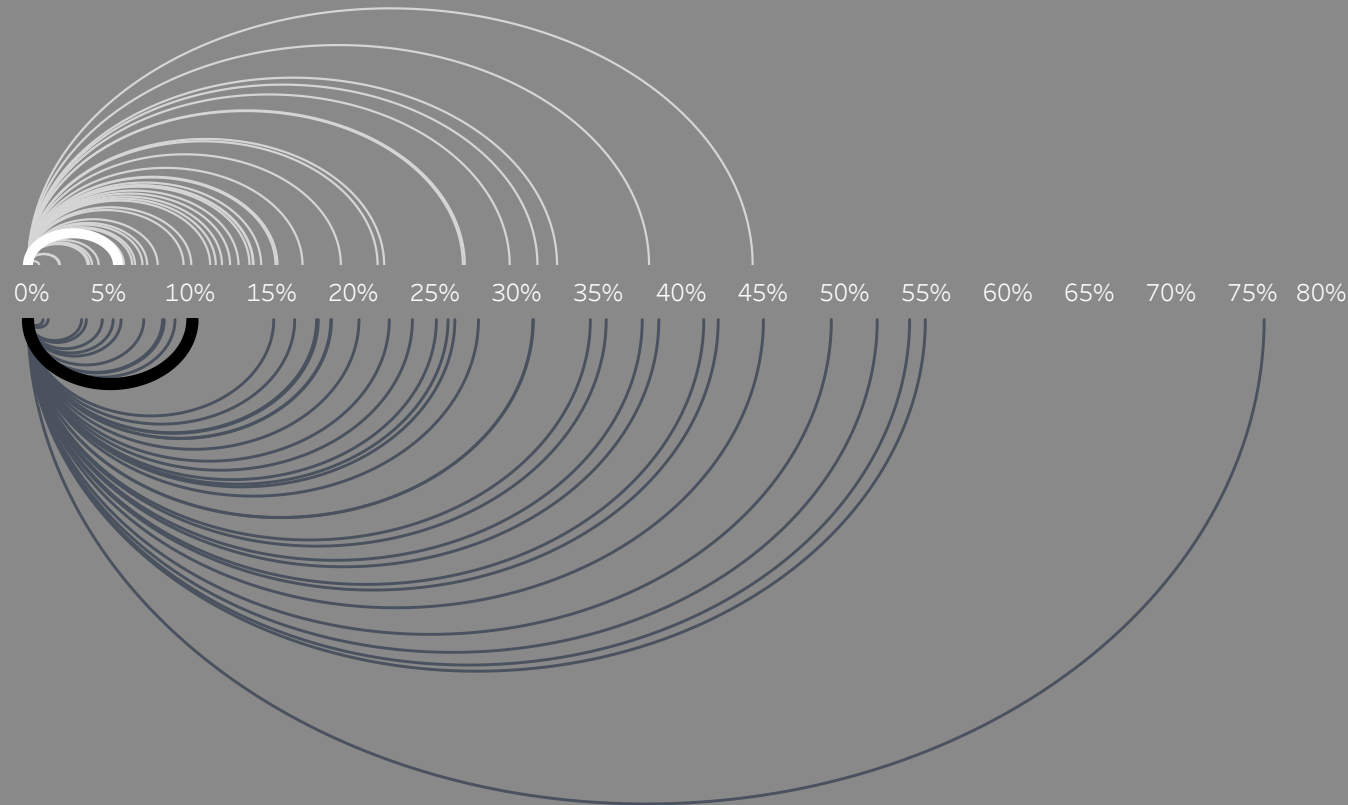
While a tragedy for everyone, COVID-19 has imposed disproportionate hardships on Afro-Am..

Blacks as % of
population in
California

5.5%

Blacks as % of
COVID-19
deaths in
California

10.1%



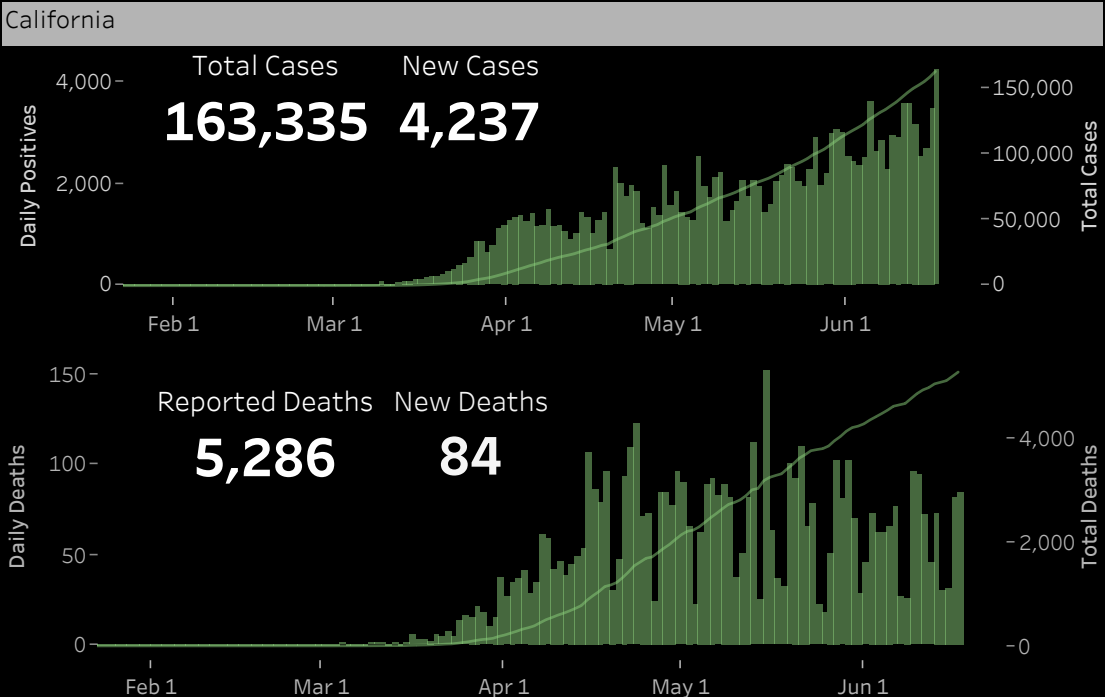
Choose a state

- ☐ Entire U.S.
- ☐ Alabama
- ☐ Arizona
- ☐ Arkansas
- ☒ California
- ☐ Colorado
- ☐ Connecticut
- ☐ Delaware
- ☐ District of Columbia
- ☐ Florida
- ☐ Georgia
- ☐ Idaho
- ☐ Illinois
- ☐ Indiana
- ☐ Iowa
- ☐ Kansas
- ☐ Kentucky
- ☐ Louisiana
- ☐ Maryland
- ☐ Massachusetts
- ☐ Michigan
- ☐ Minnesota
- ☐ Mississippi
- ☐ Missouri
- ☐ New Jersey
- ☐ New Mexico
- ☐ New York
- ☐ North Carolina

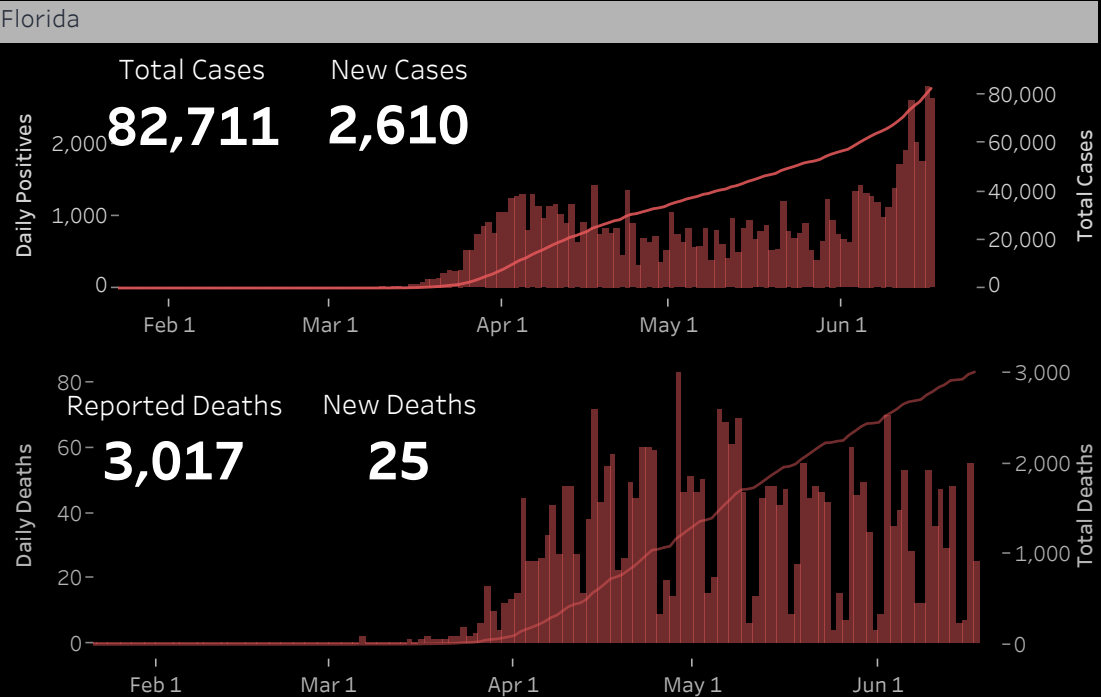
DATA VIZ: Eleonora Nazander | DATE: June 5th 2020

Compare Trends in Two States

California

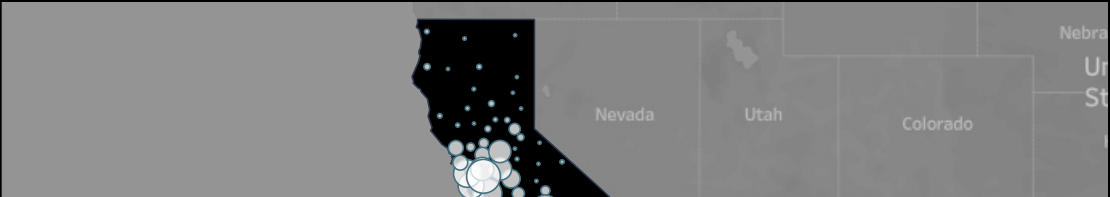


Florida



Where are the hotspots?

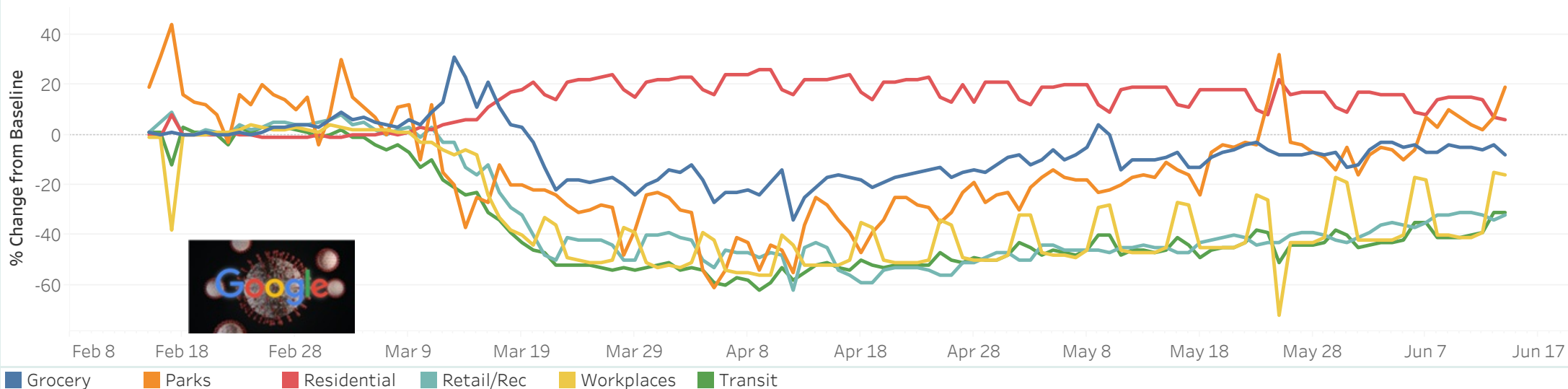
California



Florida



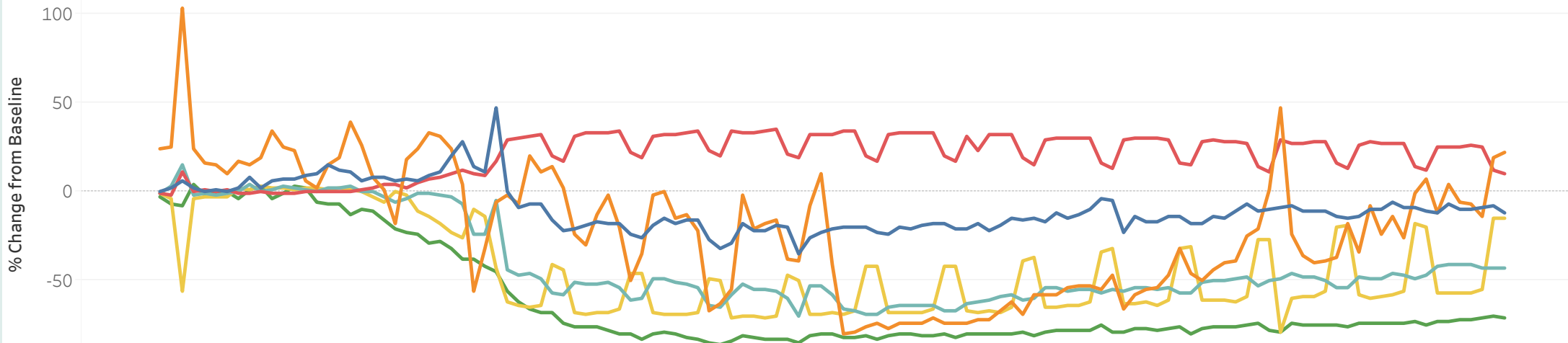
Google Mobility - Statewide



Google Mobility Report - Pick a County

San Mateo

null = State of CA



State to State Comparison as of 6/17/20

Google Mobility Data - Statewide vs.
County of Your Choice

Descartes Labs Mobility Index

Apple Mobility Data for the Bay Area

Racial Make-Up of California Counties

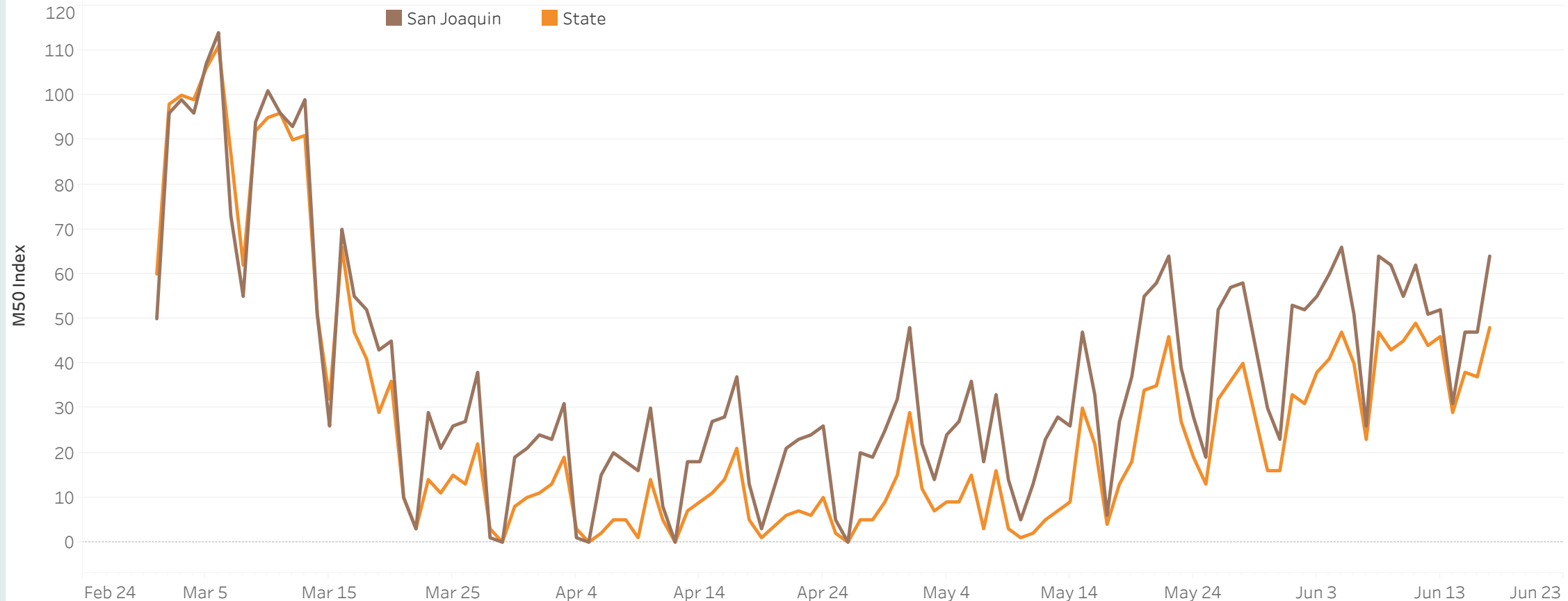


**Descartes
Labs**

Mobility Index

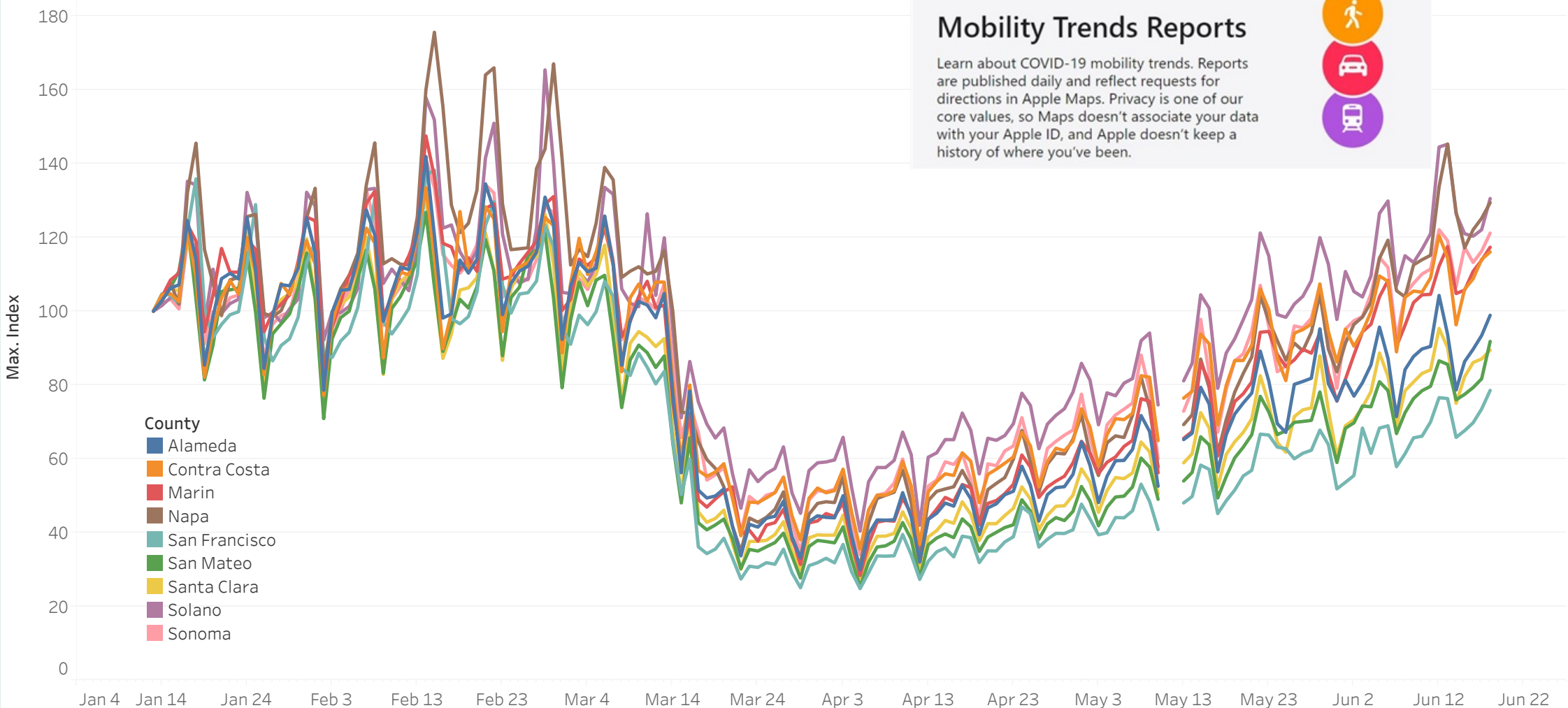
3/1/2020 to 6/17/2020

State or County
Multiple values



Descartes Labs has detected dramatic changes in mobility due to COVID-19. Mobility data at the state and county level have been made freely available via github.com/descarteslabs/DL-COVID-19. Combining information derived here with pandemic growth rates in various geographies will allow more accurate models of the interventions being made, and help save lives.

Apple COVID Mobility - Bay Area Driving



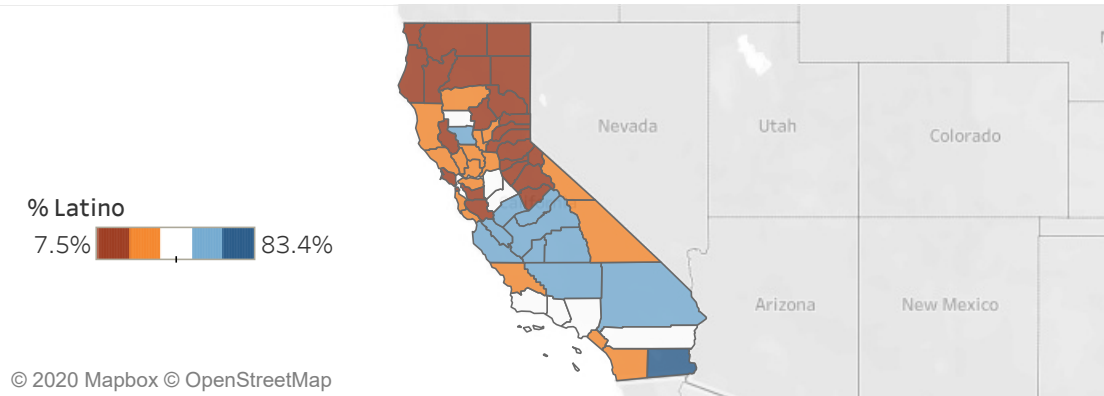
Apple Maps

Mobility Trends Reports

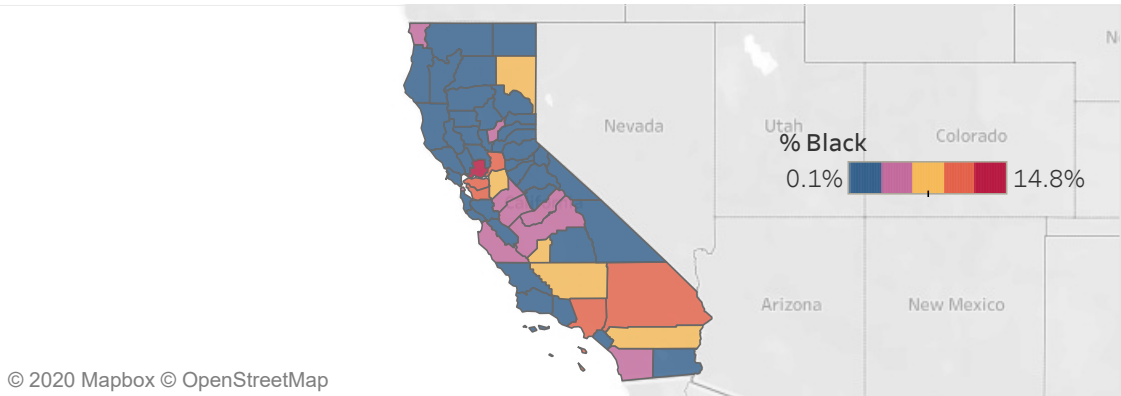
Learn about COVID-19 mobility trends. Reports are published daily and reflect requests for directions in Apple Maps. Privacy is one of our core values, so Maps doesn't associate your data with your Apple ID, and Apple doesn't keep a history of where you've been.



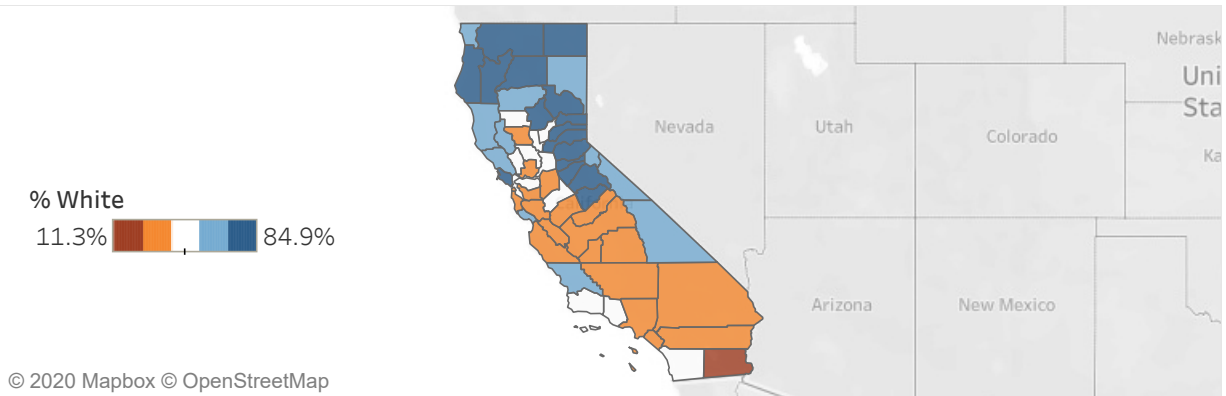
Latino Population



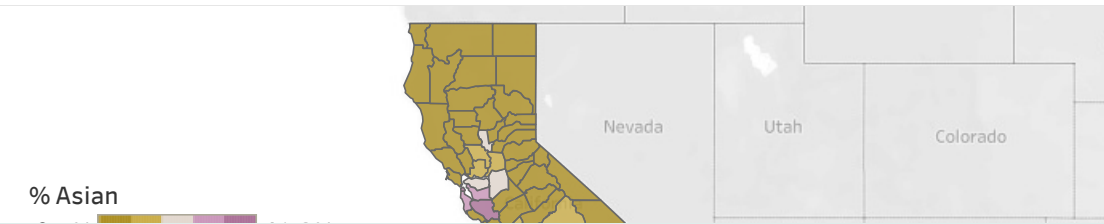
Black Population



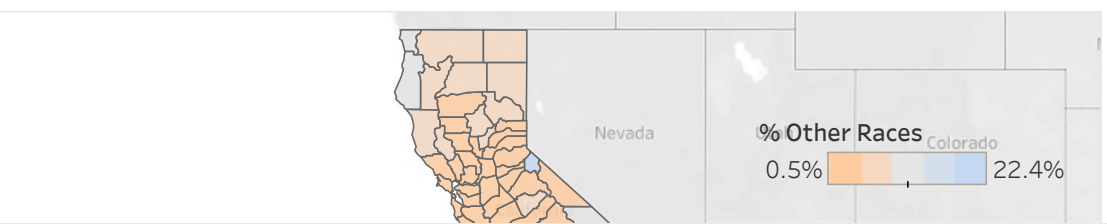
White Population



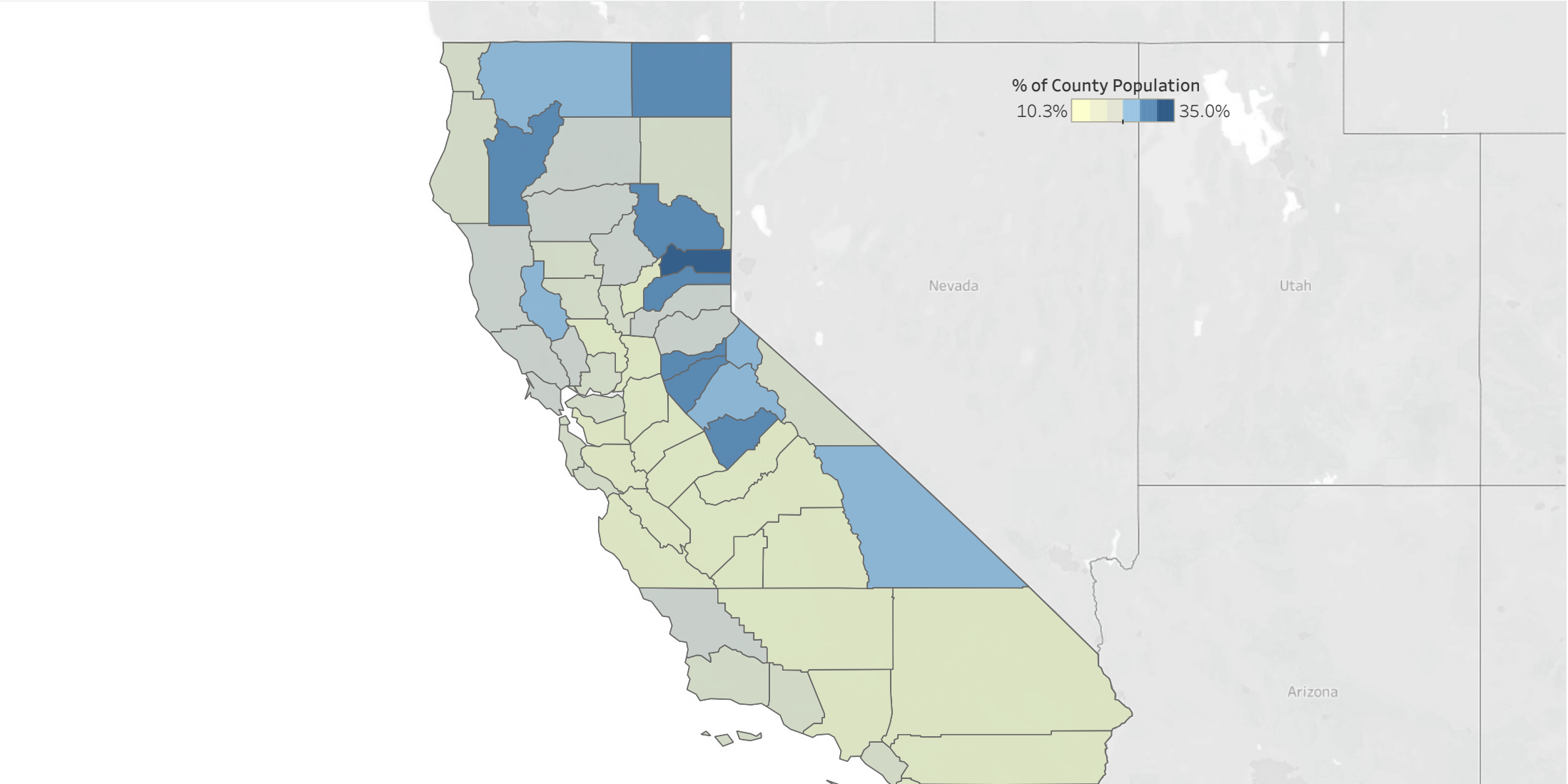
Asian Population



Other Races



Senior Citizen (>=65) Population



California Prison Clusters



COVID Killers - Group 3 - Tues/Thurs

Cristina, Samira, Drew, Doug

