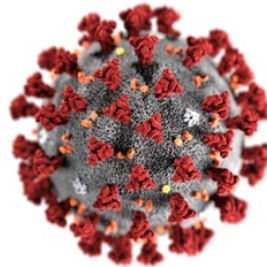# COVID Killers
Impact Predictor for California Residents
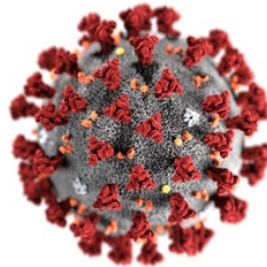
COVID Killers - Project Overview

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 project's data analysis
• Review conclusions - summary of learnings and insights

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.
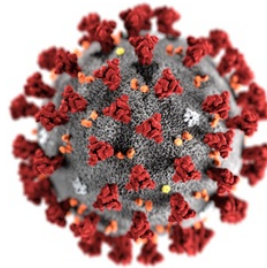
COVID Killers - Data Sources

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:
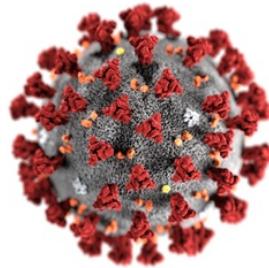
State daily case counts (New York Times):
https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv

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

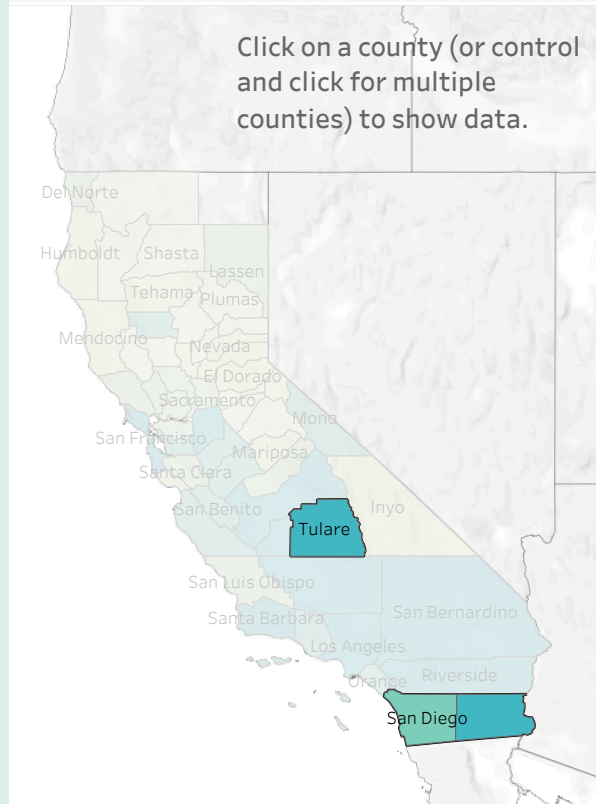Vulnerability index (Surgo Foundation): ..

COVID Killers - Questions to Answer with the Data

What will be the future number of California COVID-19 cases during the month of July 2020?
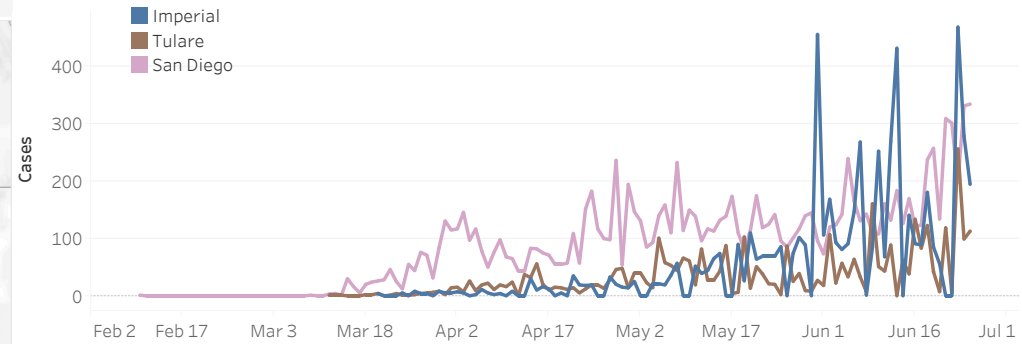
1. What factors increase/decrease risk of spreading COVID-19?
2. What counties are at highest risk?
3. Can we provide county specific insights for users to make data augmented/guided decisions about their own risk?
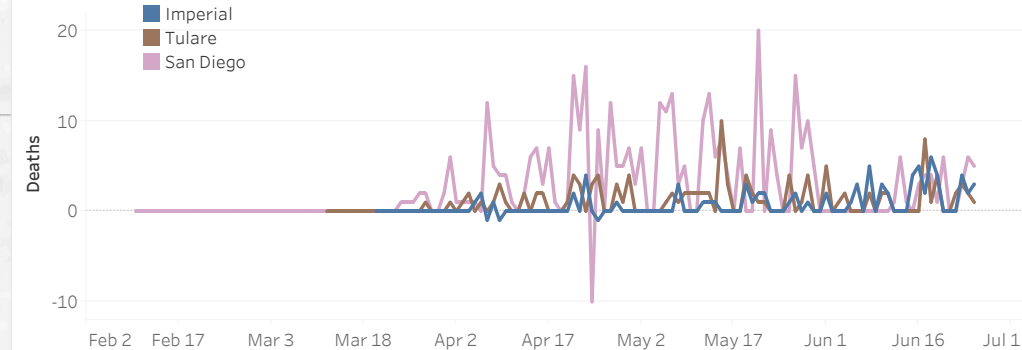
All California Counties

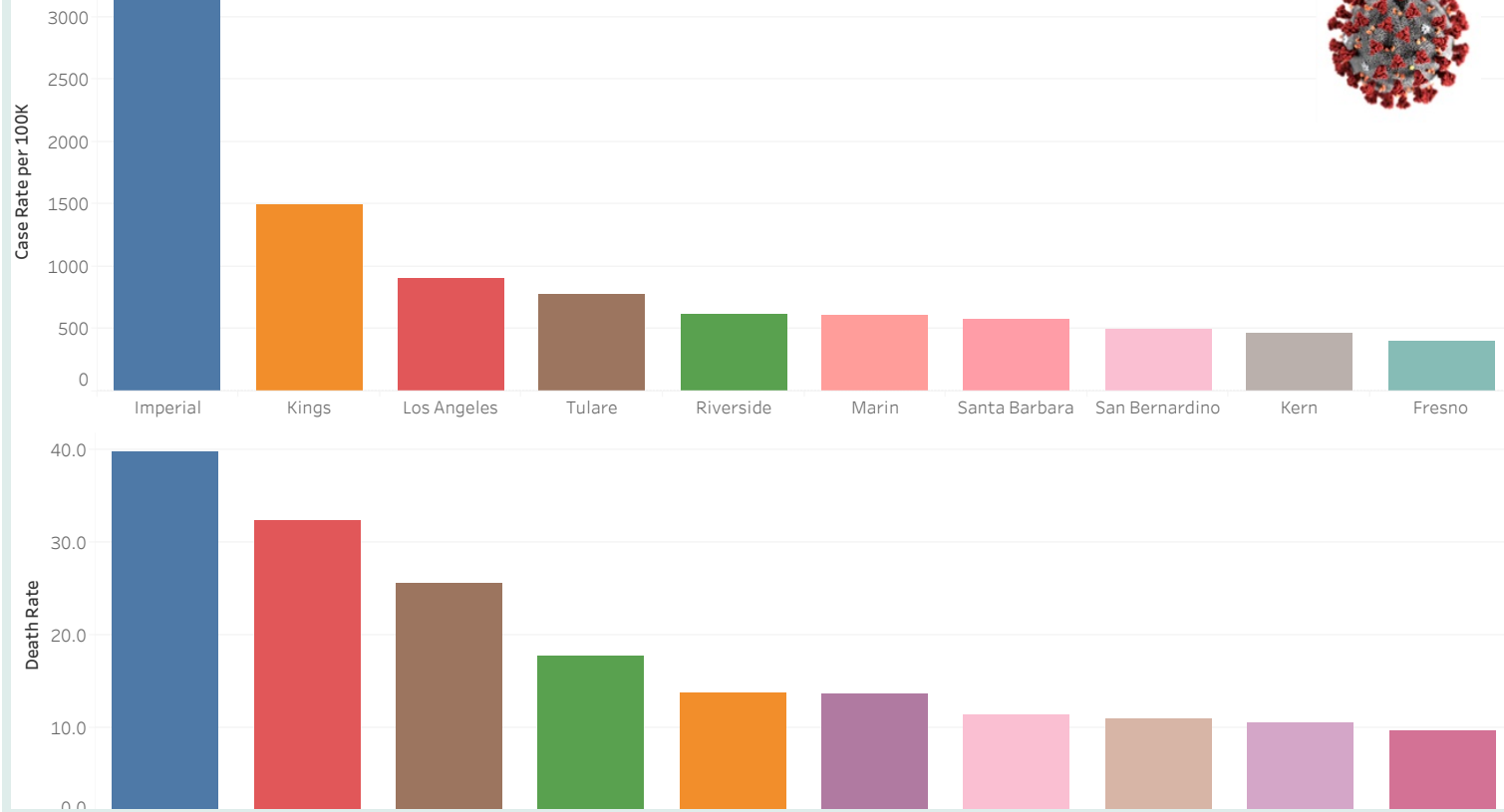Click on a county (or control and click for multiple counties) to show data.

Del Norte, Humboldt, Shasta, Lassen, Tehama, Plumas, Mendocino, Nevada, El Dorado, Mono, Sacramento, Mariposa, San Francisco, Santa Clara, San Benito, Inyo, **Tulare**, San Luis Obispo, Santa Barbara, San Bernardino, Los Angeles, Orange, Riverside, **San Diego**

County Daily Infections

Legend: Imperial, Tulare, San Diego

County Daily Deaths

Legend: Imperial, Tulare, San Diego

State	County	Case Rank	Cases/100K	Cases	Deaths	Deaths/100K	Population
CA	Imperial	1	3,170	5,744	72	39.7	181,215
	Tulare	4	782	3,646	119	25.5	466,195

County

Case Rate per 100K

| Imperial | Kings | Los Angeles | Tulare | Riverside | Marin | Santa Barbara | San Bernardino | Kern | Fresno |

Death Rate

COVID-19 Community Vulnerability Index

CCVI SCORE

0.0672   0.9583

Socioeconomic
0.006 to 0.99968163

Household Composition
To Null

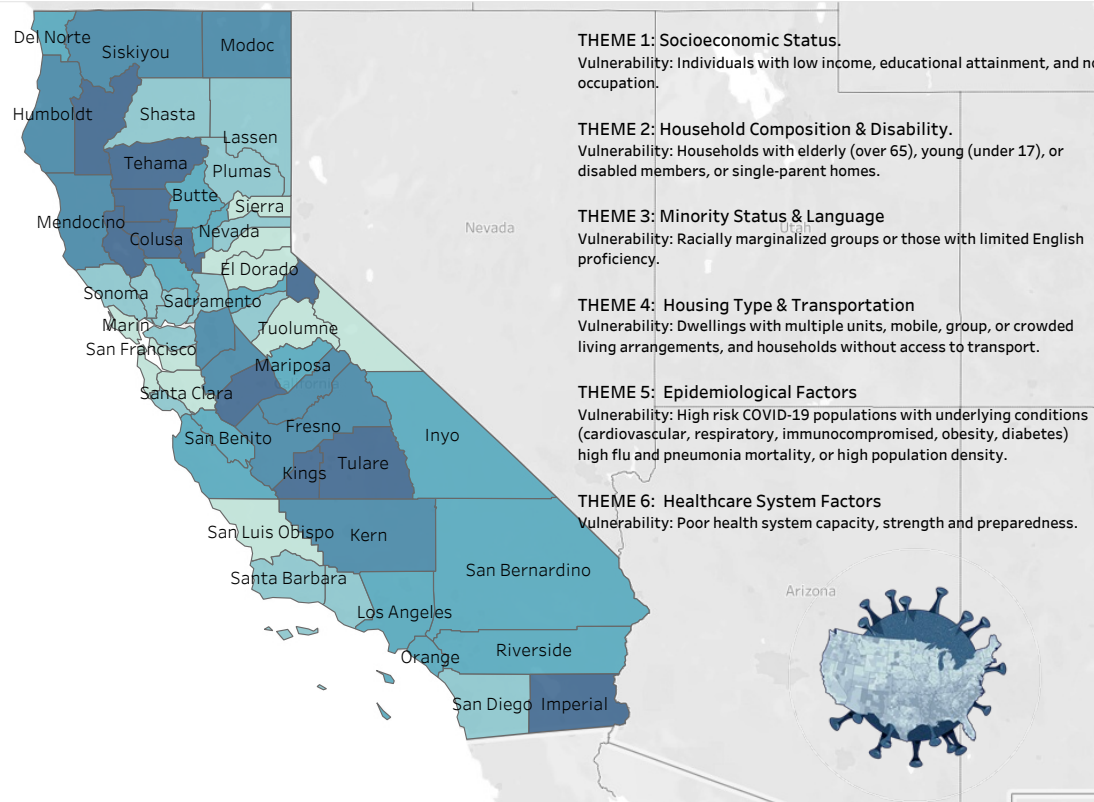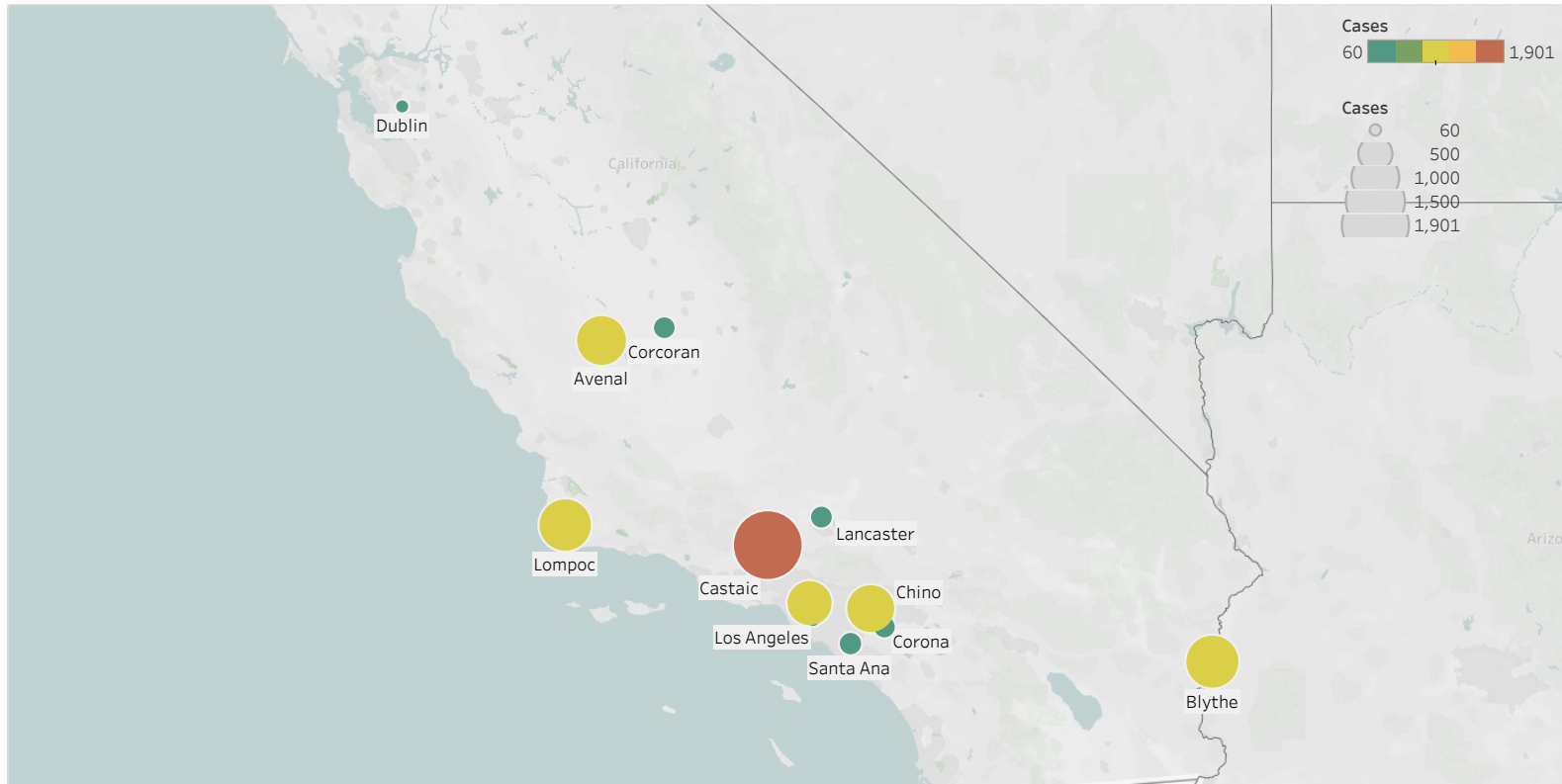**Minority Status & Langua..**
0 to 1

Housing & Transportation
0 to 1
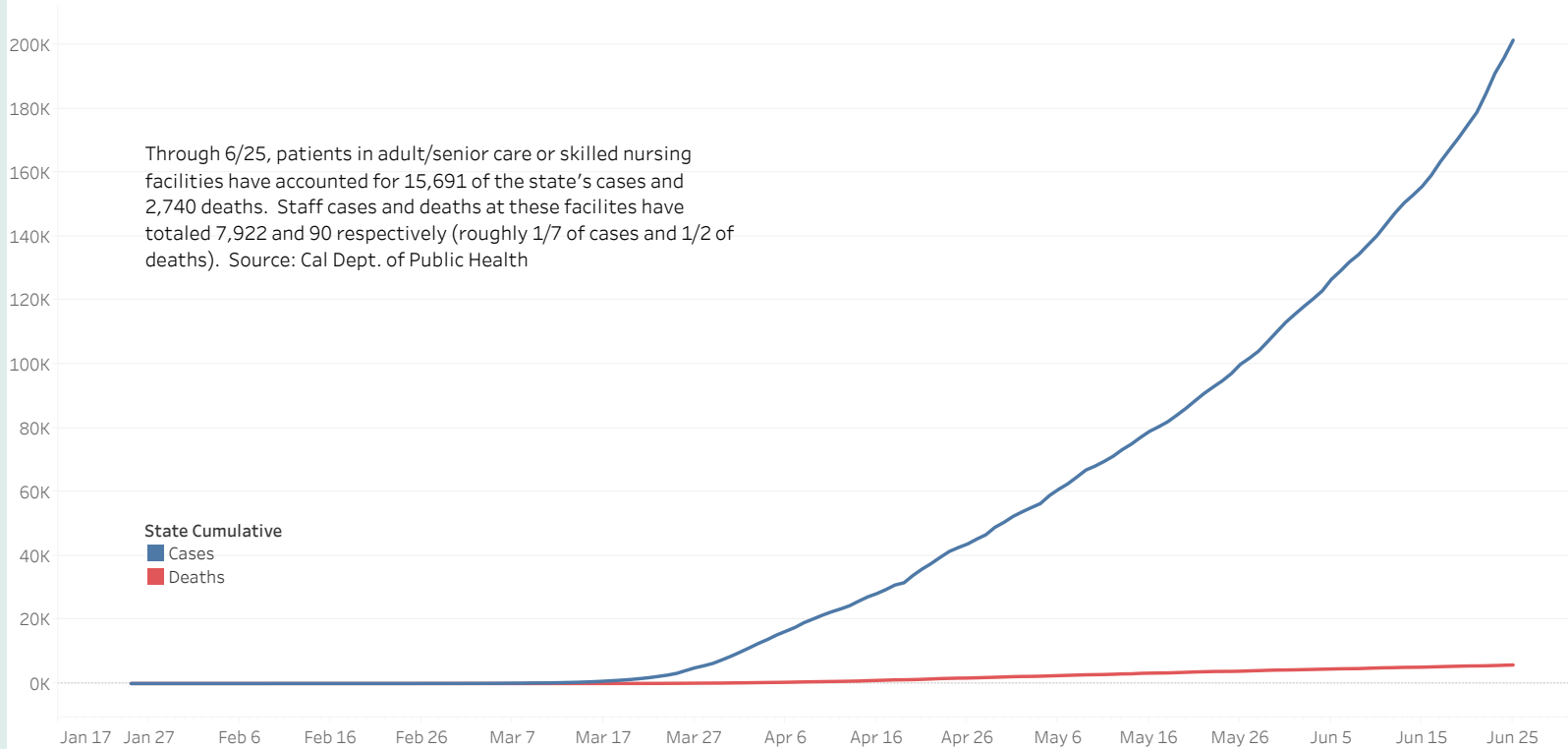
Epidemiological
0 to 1

Healthcare System
0 to 1

THEME 1: Socioeconomic Status.
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.

© 2020 Mapbox © OpenStreetMap

California Prison Clusters

Cumulative Tally of Cases and Deaths for California

Through 6/25, patients in adult/senior care or skilled nursing facilities have accounted for 15,691 of the state's cases and 2,740 deaths. Staff cases and deaths at these facilites have totaled 7,922 and 90 respectively (roughly 1/7 of cases and 1/2 of deaths). Source: Cal Dept. of Public Health

State Cumulative
- Cases
- Deaths

Compare Trends in Two States

California

California

Total Cases
195,889

New Cases
4,850

Reported Deaths
5,728

New Deaths
91

Arizona

Arizona

Total Cases
60,207

New Cases
1,810

Reported Deaths
1,467

New Deaths
69
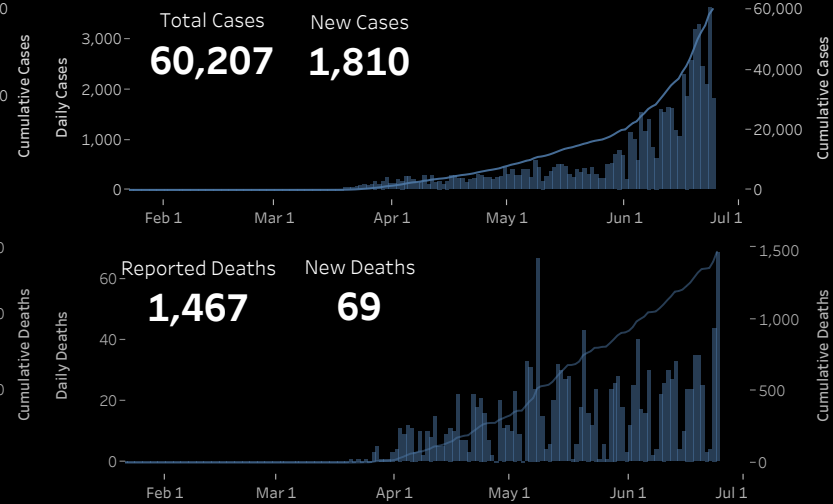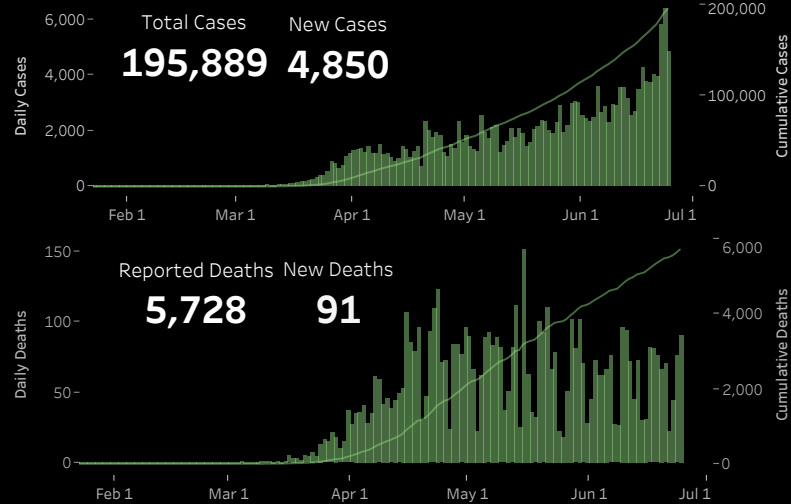
Where are the hotspots?

California

Arizona

Population by Race

7.5%   83.4%
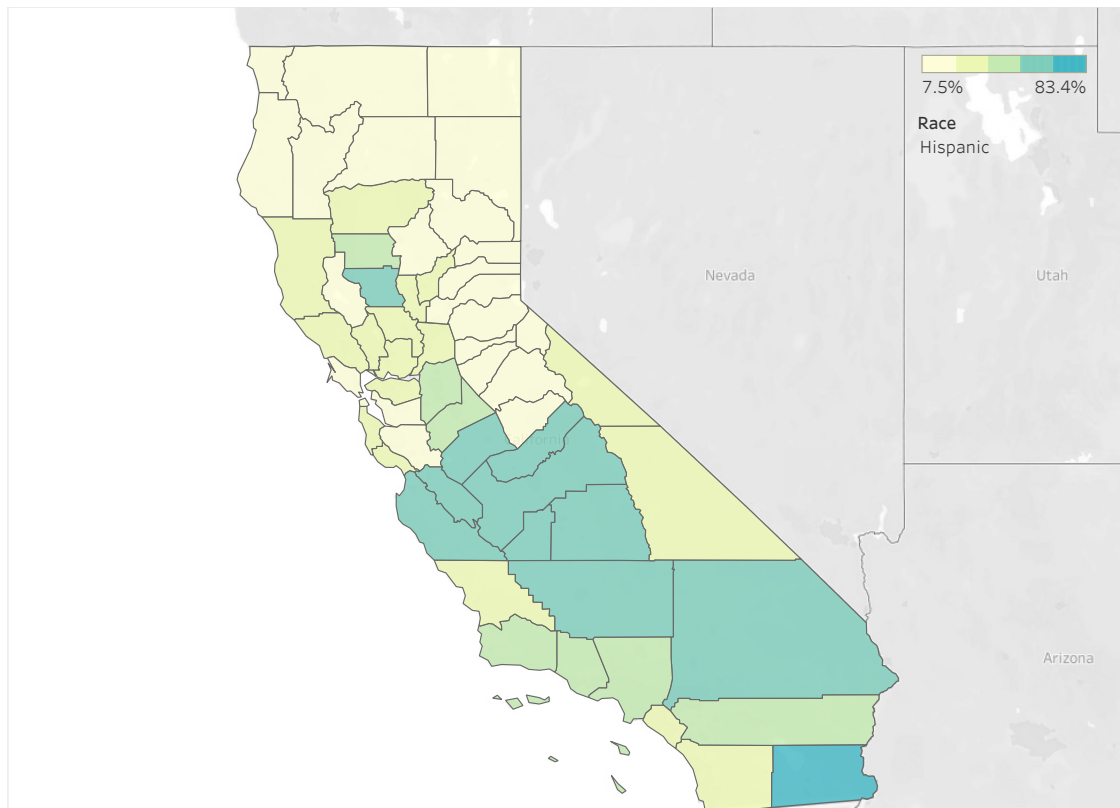
Race
Hispanic

Cases by Race through June 24

Confirmed Cases by Race 130,910

Age Total

- white 17.7%
- Asian 8.3%
- Black 4.6%
- Latino 56.8%
- Other 14.6%
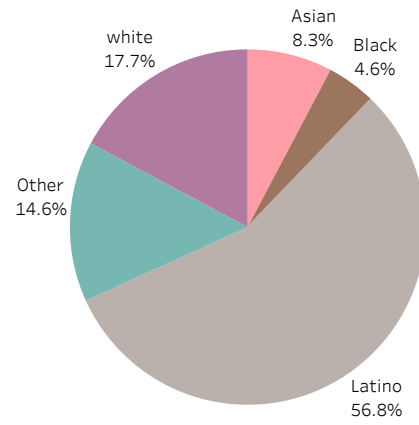
Deaths by Race through June 24

Total Deaths 5,435

Age Total

- white 33.3%
- Asian 15.1%
- Black 9.6%
- Latino 41.0%
- Other 1.9%
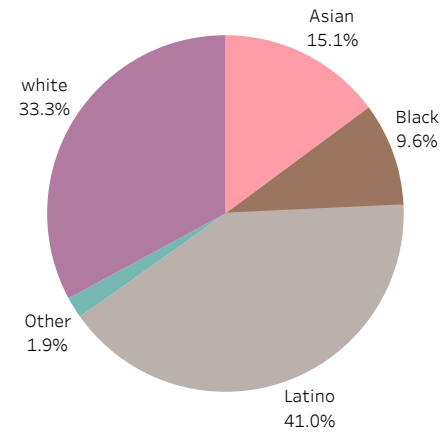
Cases: Approximately (29%) missing race/ethnicity
Deaths: Approximately (1%) missing race/ethnicity
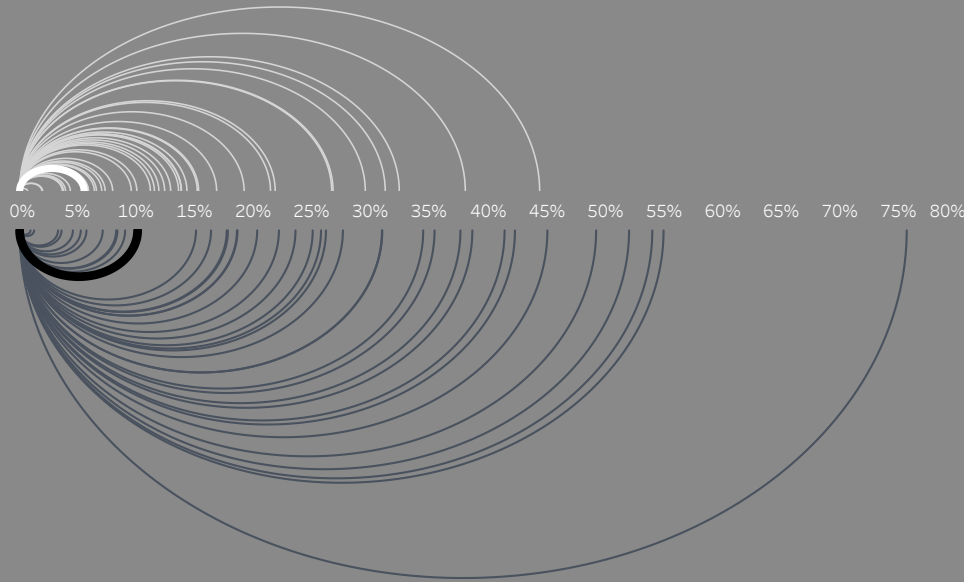
COLOR of COVID-19

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

Blacks as % of population in California
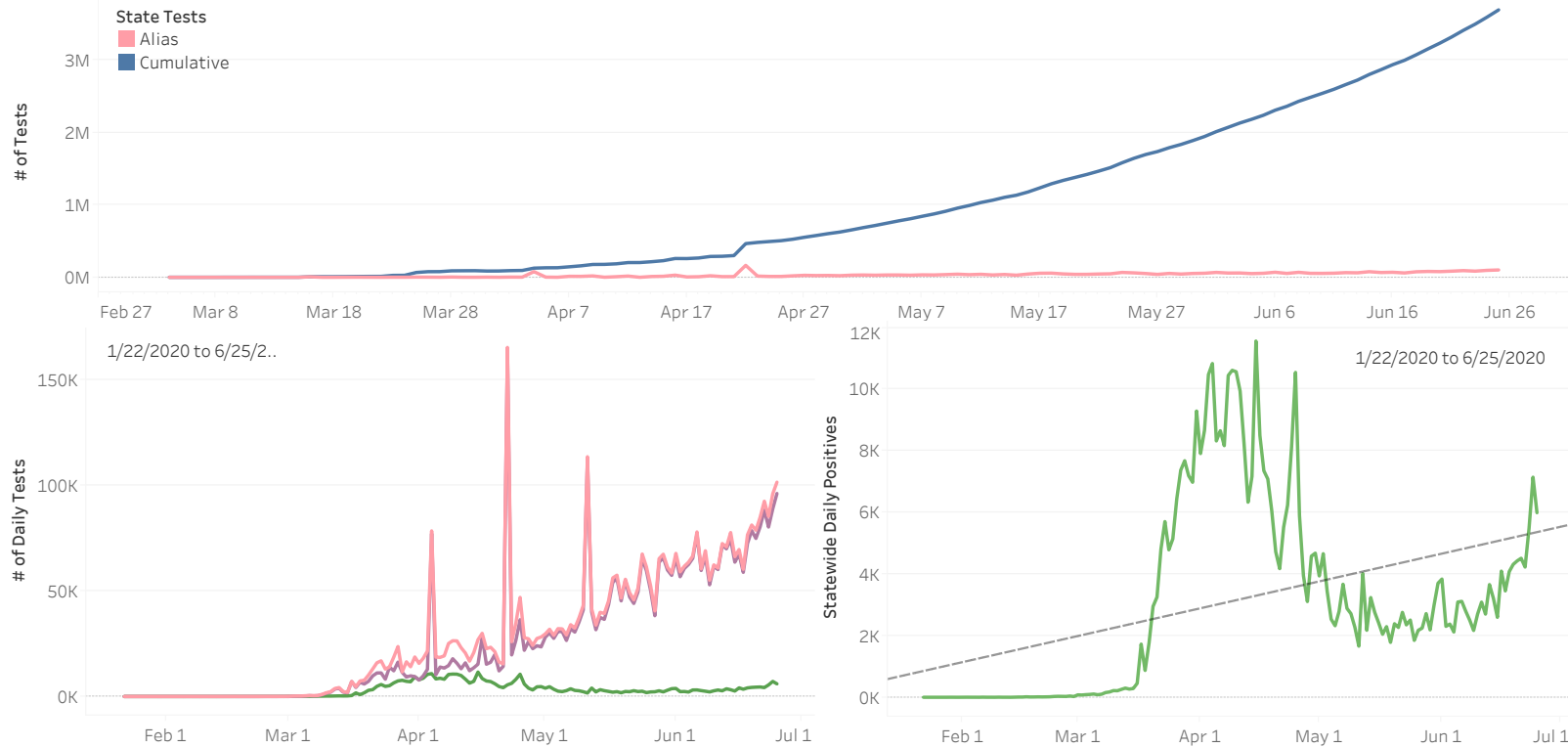
5.5%

Blacks as % of COVID-19 deaths in California

10.1%

0% 5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80%

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

COVID Testing in CA by Date

1/22/2020 to 6/25/2020 and Null values

State Tests
- Alias
- Cumulative

1/22/2020 to 6/25/2...

1/22/2020 to 6/25/2020

Google Mobility - Statewide

Legend: Grocery And.. | Parks Perce.. | Residential .. | Retail And R.. | Workplaces .. | Transit Stat..
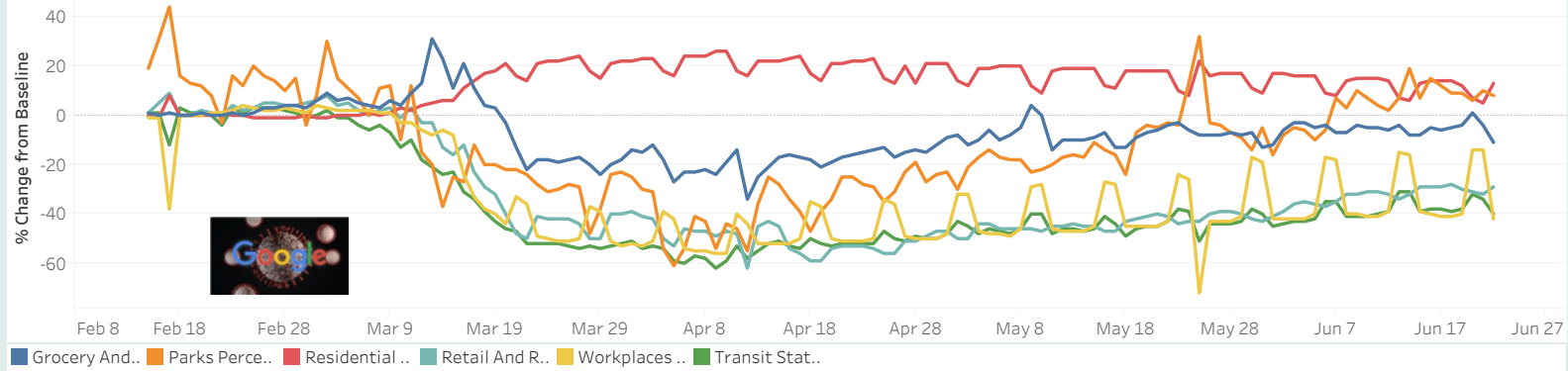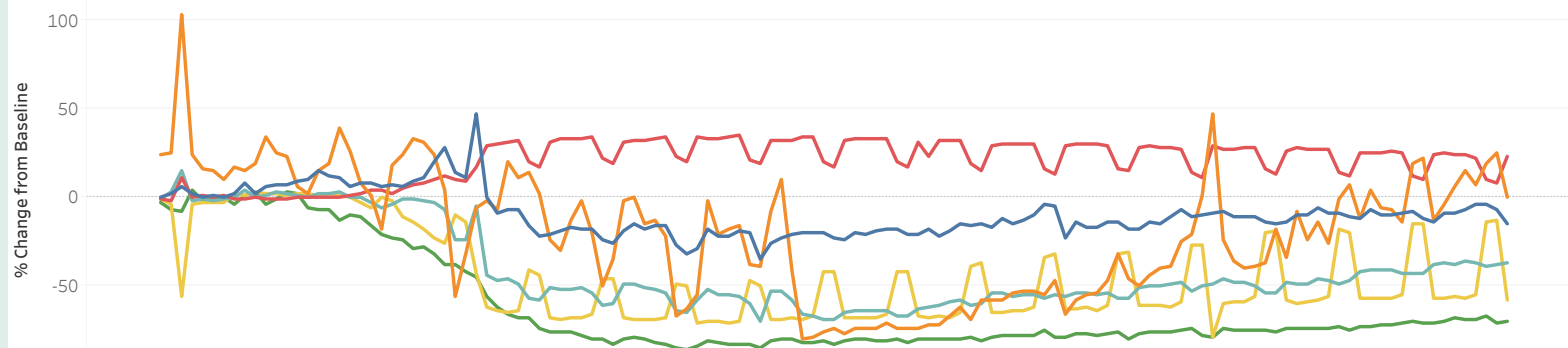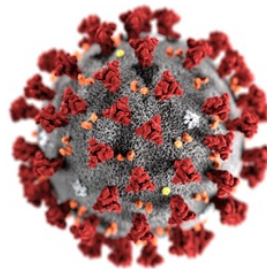
Google Mobility Report - Pick a County

San Mateo   null = State of CA

COVID Killers - Machine Learning

Time series forecasting
* Time add an explicit order dependence between observation: a time dimension
* Fit model on historical data to predict future data

Data analysis and feature engineering

* Render data stationary (mean and variance are constant, trend removed)
  - Dickey-Fuller to test stationarity
  - Add differencing
  - Log transforms
* Remove autocorrelation
  - ACF and PACF to plot autocorrelation
  - Add lags to remove autocorrelation
* Causality for multivariate models
  - Grangers to test causality
 - Removing for feature with a lesser significance
  - Drop feature..

Online Data Analytics Boot Camp

Cristina
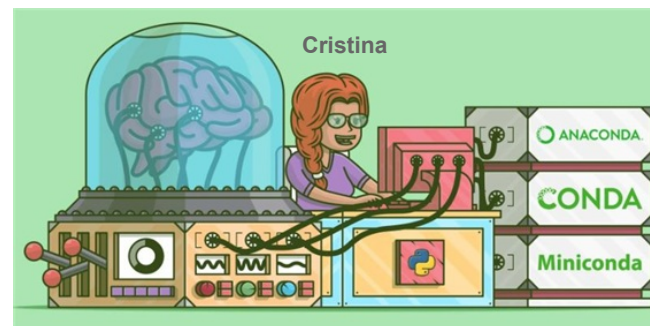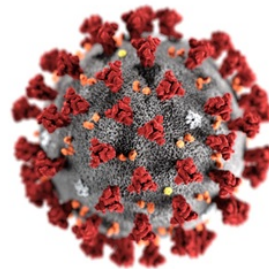
COVID Killers - Machine Learning

ARIMA - Lags of the stationarized series in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series

- Data needs to be stationary
- Case data only
- 81% accuracy

Vector Autoregression (**VAR**) is a multivariate forecasting algorithm that is used when two or more time series influence each other.
- Data needs to be stationary and causal
- Cases and mobility potentially a good fit..
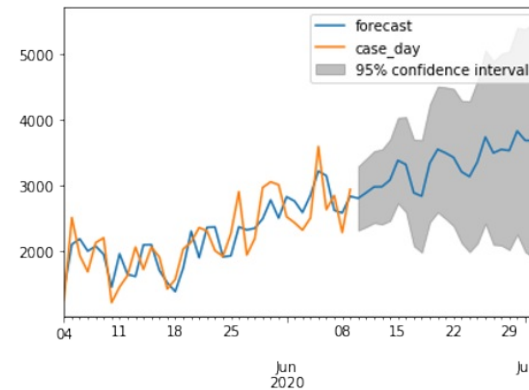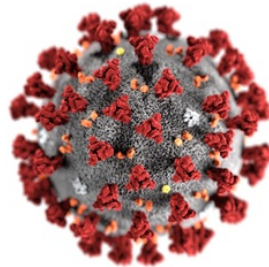
COVID Killers - Machine Learning

Future enhancements

* Use a deep learning model that will take multivariate data and external regressors to make a prediction (LSTM as a candidate)
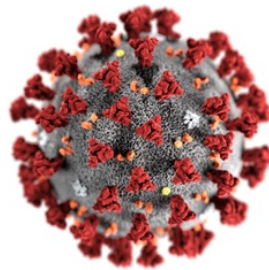* Make the modelling data agnostic and use it to forecast cases for different state and counties
* Retrieve the data from an API instead of database to always get the freshest data

Online Data Analytics
Boot Camp

COVID Killers - Learnings and Insights

1. The future number of California COVID-19 cases in July 2020 will unfortunately continue to trend up

2. Factors increasing/decreasing risk of spreading COVID-19 are multi-varied:
   * The highest risk of contracting and dying from COVID-19 exists within assisted living facilities and prisons; residents/inmates and workers
   * Other factors:
     - Population density
     - Age
     - Race/Ethnicity
     - Mobility; adherence to shelter-on-place

3. Southern and Central California counties appear to be most at risk, with Imperial County (on the southern border) showing the most rapid increase in new cases and the highest rate per 100K population

4. The COVID Killers Dashboard provides a broad range of interactive data applications identifying county specific insights for users to make data augmented/guided decisions about their own risk. "Bonus Features" slides at end provide visualizations not included in our presentation.
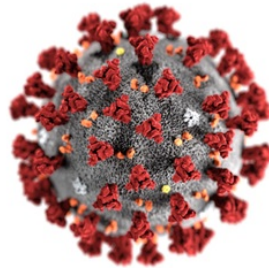
COVID Killers

Questions and Answers

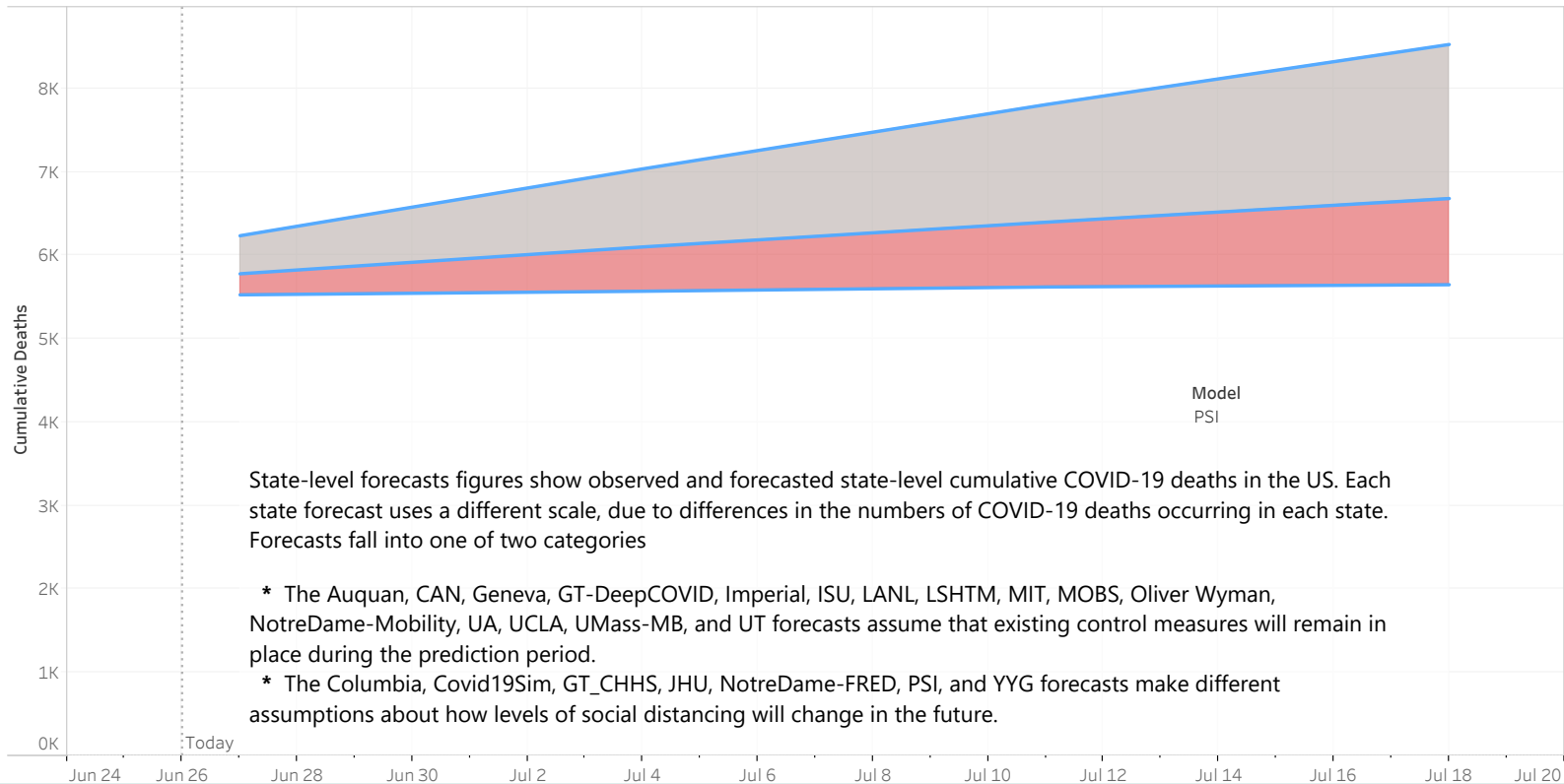https://github.com/menlodoug/Final_Project

COVID Killers - Group 3 - Tues/Thurs

Cristina, Samira, Drew, Doug

CDC Forecast Models with 95% CI for CA COVID Deaths - 4 weeks starting 6/20/20

Model
PSI

State-level forecasts figures show observed and forecasted state-level cumulative COVID-19 deaths in the US. Each state forecast uses a different scale, due to differences in the numbers of COVID-19 deaths occurring in each state. Forecasts fall into one of two categories

  * The Auquan, CAN, Geneva, GT-DeepCOVID, Imperial, ISU, LANL, LSHTM, MIT, MOBS, Oliver Wyman, NotreDame-Mobility, UA, UCLA, UMass-MB, and UT forecasts assume that existing control measures will remain in place during the prediction period.
  * The Columbia, Covid19Sim, GT_CHHS, JHU, NotreDame-FRED, PSI, and YYG forecasts make different assumptions about how levels of social distancing will change in the future.

Cumulative Deaths (y-axis): 0K, 1K, 2K, 3K, 4K, 5K, 6K, 7K, 8K

x-axis: Jun 24, Jun 26 (Today), Jun 28, Jun 30, Jul 2, Jul 4, Jul 6, Jul 8, Jul 10, Jul 12, Jul 14, Jul 16, Jul 18, Jul 20

Apple COVID Mobility - Bay Area Driving

 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.

County
- Alameda
- Contra Costa
- Marin
- Napa
- San Francisco
- San Mateo
- Santa Clara
- Solano
- Sonoma

Descartes Labs

Mobility Index

3/1/2020 to 6/23/2020

State or County
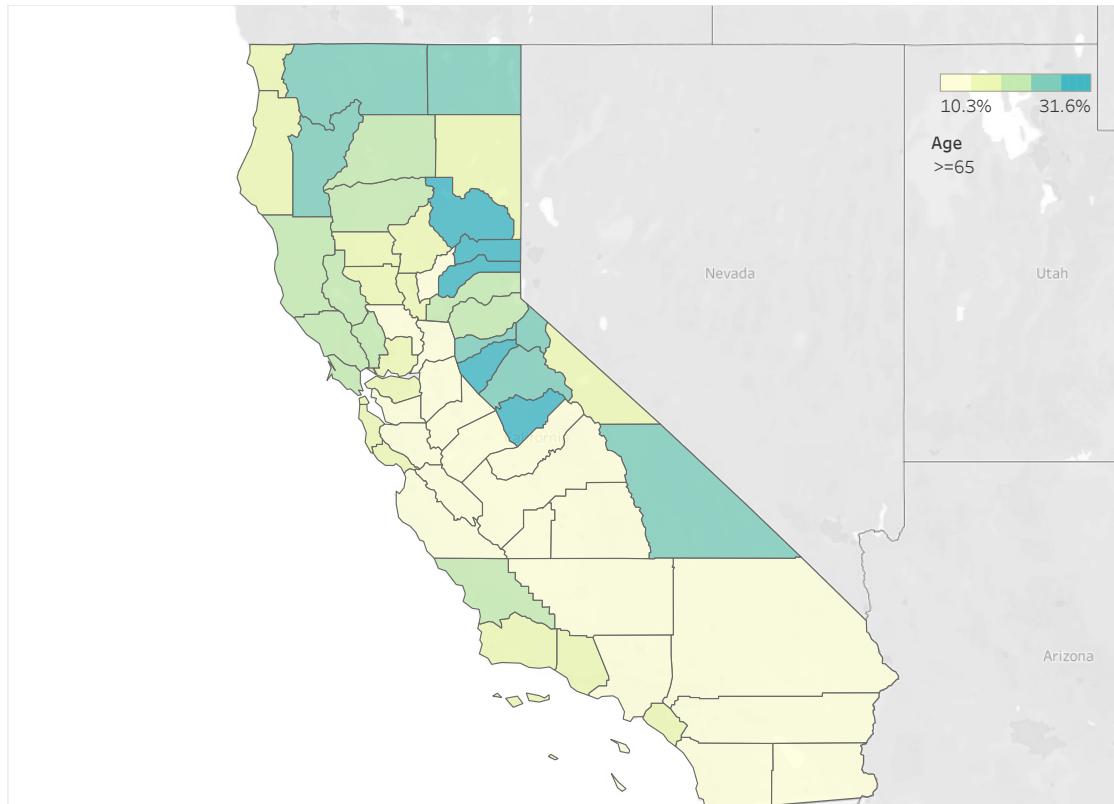Multiple values

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.

Population by Age

10.3%   31.6%

Age
>=65

Population by Gender

48.2%　62.2%

Sex
Male