How do COVID-19 infection patterns in California relate to pandemic responses?

INTRODUCTION

COVID-19, a novel coronavirus, has rapidly spread across the United States. Regions have varying patterns, with some states showing abnormal growth rates regarding reported COVID cases. Since California has the most population and most COVID cases in U.S (shown in Figure 1), the goal of this project was to compile and analyze COVID-19 data in 58 California counties, dating from January 2020 to February 2021. In this project, we have discovered patterns within the COVID cases data in the 58 California counties and identified factors associated with COVID cases.

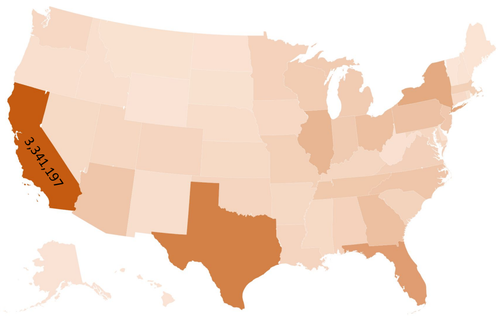


Figure 1. COVID Cases in U.S by February 1st 2021

Part I. 58 CA counties are categorized into 3 major groups based on their weekly infection percentage

Cluster Analysis

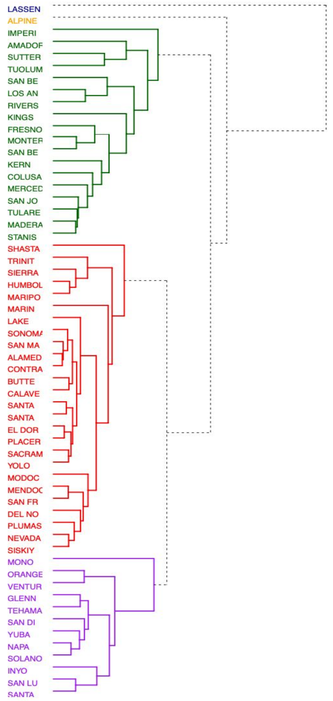


Figure 2. Cluster Dendrogram for 58 CA Counties

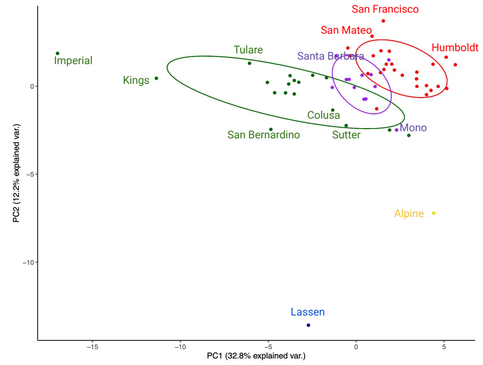


Figure 3. Principal Component Analysis for 58 CA Counties

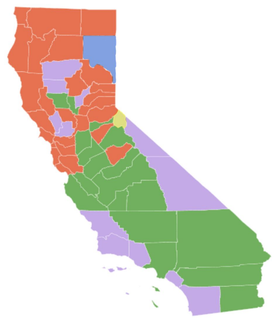


Figure 4. 5 Clusters in CA Map

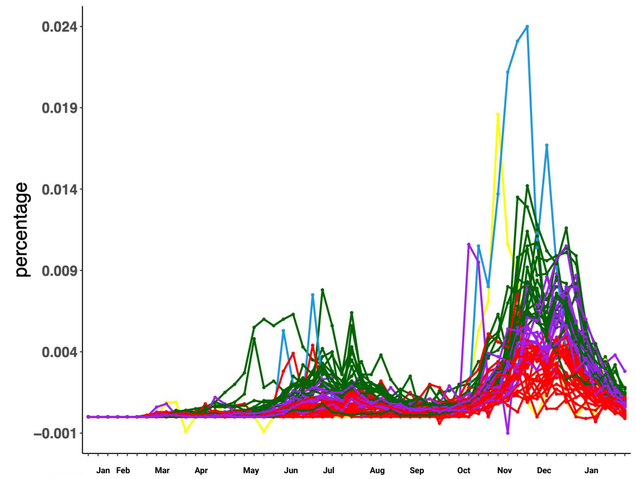
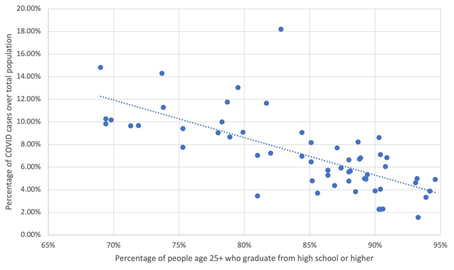


Figure 5. Percentage of Weekly Infection in Each County

- **3 major groups and 2 outlier counties are shown in Figure 2 and 3.**
- **Patterns discovered**:
  - **Neighbor effect** - Counties tend to behave similarly if they are close to each other. This is shown in Figure 4 that colors are not scattered sporadically.
  - **Difference between north and south** - The northern counties (red group) have a lower percentage of weekly infection, comparing to southern east counties (green group) which have a higher percentage of weekly infection (shown in Figure 5).

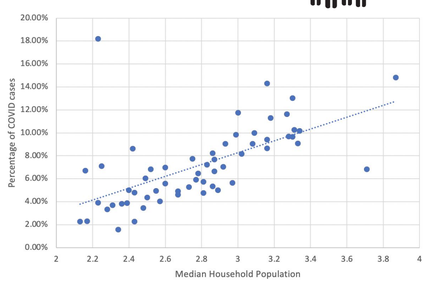
Part II. Deeper Analysis Factors Associated with Percentage of Total COVID Cases over County Population

Education Attainment



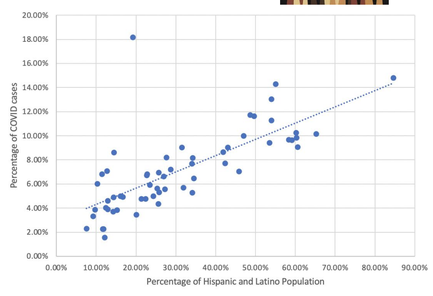
Counties with higher education attainment have a lower percentage of COVID cases --- people with a high education usually occupy white collar work (don't need to work outside during this difficult time) and they are more health-conscious.

Household Population



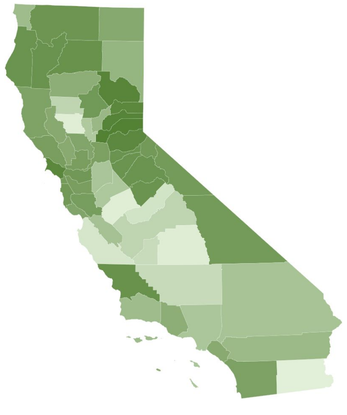
Counties with higher median household population have higher percentage of COVID cases --- crowding in the household so more people get infected.

Minority Population

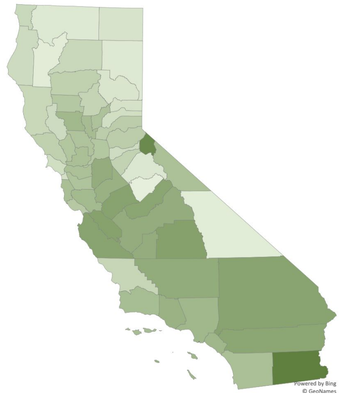


Hispanic/Latino is the largest minority groups in CA. Counties with higher Hispanic/Latino populations have higher percentage of COVID cases --- they usually occupy blue collar work and need to use public transportation

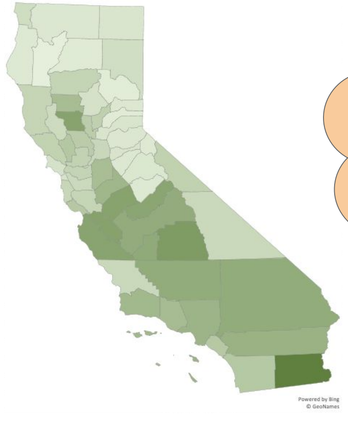
Education Attainment



Household Population



Minority Population



From the three maps on the left, we can clearly see the difference between northern and southern California in education attainment, median household population, and minority (Hispanic/Latino) population.

These factors may cause the difference between northern and southern California on the trend and patterns of COVID infection data.

Why do groups behave differently? To figure it out, we want to explore more factors associated with COVID cases.

OUTLIERS - Lassen (blue) and Alpine (yellow) have a higher percentage of weekly infection

Road to Lassen Volcanic National Park closed before the snow, so people visited there earlier which could cause an early peak in Lassen (shown in figure 5).

Roads leading to Alpine's main health clinic were closed due to weather, so citizens may not get treatment in time.

CONCLUSION

- By clustering 58 California counties, we found different patterns of COVID infection percentage: 3 major groups and 2 outliers.
- Two outliers, Lassen and Alpine, both have a relatively high increase of COVID cases and early second peaks. In Alpine, this could be because the road closures due to snow meant that residents were unable to get treated for COVID. In Lassen, road closure in December could lead to more park visits in November, resulting in the peak COVID cases around that time.
  - This shows that for future pandemics, it is crucial to have an established, well-based infrastructure, and that the government should take action to improve the infrastructure.
- Factors associated with COVID infection percentage were identified, such as education attainment, median household population, and percentage of minority (Hispanicor/Latino) population. These three factors could also be the reasons why northern and southern California behave differently.
- Our study should provide guidance for preventing pandemic spread including COVID-19.