Project Write-Up

Tackling California's Drug Overdose Problem

Abstract

The ultimate goal of this project is to determine what zip codes are most in need of education and outreach in order to curb the rapid increase in drug overdose cases in California. Preliminary analysis of several drug-related datasets gave us some insights into what California's current drug overdose cases involved, notably:

- We see a steady rise in Hospital Admissions and Deaths related to drug overdose over the last 5 years
- Many drug overdose cases now seem to involve prescription medications
- Initial scatterplot visualizations of several Schedule II-V prescription and demographic metrics showed some correlation with deaths by drug overdose

Further analysis will involve a linear regression model targeting the number of deaths due to drug overdose.

Design

We intend to build a linear regression model that uses features such as Schedule II-V drug prescription practices as well as demographic data, with drug overdose deaths as our target variable. This analysis at the county level will allow us to predict the number of overdose deaths at the zip code level. The analysis will also serve to identify what populations are most at risk for drug overdose so that the CDPH can provide drug education and outreach programs specifically tailored to them.

Data

For the preliminary analysis presented here, we used data from the following sources:

- Cause of death, drug type, by county and year: https://wonder.cdc.gov/ https://www.cdc.gov/drugoverdose/deaths/index.html
- Hospital and ER admissions due to drug overdose in CA: https://data.chhs.ca.gov/dataset/hospital-emergency-department-diagnosis-procedure-and-external-cause-codes
- California Schedule II-V Drug Acquisition, Prescription, and Dispensation Public Statistics: https://oag.ca.gov/cures/statistics

- "Current Data Pile" from the California State Association of Counties for demographic data: https://www.counties.org/data-and-research

All datasets can also be found on my GitHub:
https://github.com/ericajstevenson/METIS_projects

Algorithms

Cleaning & EDA

All data cleaning was carried out in Excel

Data analysis and visuals were produced with Excel and Tableau

Initial analysis of the prescription data required a lot of decoding with the included "data dictionaries"

Aggregation

We aggregated by county the data I found for Schedule II-V prescriptions in CA. I combined by ICD-10 code the reasons for hospitalization due to drug overdose because we didn't need so much detail. We created a final dataset that merged the number of deaths due to overdose per county with tables containing Schedule II-V prescription practices (2020) as well as demographic data for each county (2017-2018). Most of my county data was normalized by the county's population and calculated per 100k people.

Visualization

The interactive Tableau dashboard containing these data and analyses can be accessed on the web here: https://public.tableau.com/app/profile/erica.stevenson

Tools

- Excel for data cleaning, aggregation, and analysis
- Tableau for plotting and interactive visualizations