Can predictive modeling save lives?

An Exploration of Drug Deaths, Opioid Prescription Rates, and Supervised Learning to predict Heroin vs non-Heroin related deaths on Open Datasets

Isabel Metzger
NYU Langone Health Predictive Analytics Unit
NYC Open Data Week - March 5th 2018

Growing up in a working-class community where drugs offered comfort to many, I witnessed relatives and friends struggle with drug addiction, some until it killed them. By the time I turned 21, my uncle had died of Hepatitis C, contracted through needle sharing, and three of my friends had fatally overdosed on opiates. These losses forced me to confront the complexity of pharmacology at an early age—that drugs designed to maximize health often contribute to the undoing of vulnerable lives.

We as clinicians and health data scientists need to assume some responsibility for moving the field in a direction where clinicians are better trained to prevent and recognize drug abuse, and drug death, especially when working in resource-strapped communities.

OVER 500 DEATHS with heroin detected in their toxicology reports in Connecticut in 2016

- With national overdoses dataset from the CDC, I explore changes in overdose death rates nationally.

- With the dataset CT Accidental Drug Deaths, I analyze the majority of my data for this exploratory data analysis. Data are derived from an investigation by the Office of the Chief Medical Examiner which includes the toxicity report, death certificate, as well as a scene investigation.

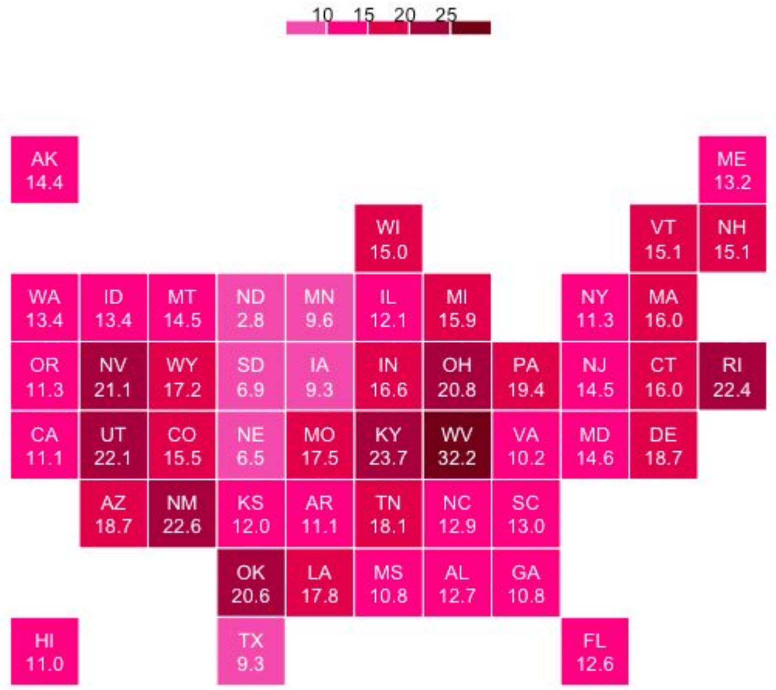
- Questions I had include: *What is the gender and age distribution of accidental drug overdoses in CT? Which drugs contribute to the majority of deaths in CT? How have the trends changed throughout the years (2012-July 2017)? What months have the highest number of accidental drug deaths? What days of the week? What dates?*

- Additionally, I looked at pharmacies in CT that prescribe naloxone/Narcan, the opiate antidote. I used machine learning techniques to build a predictive model to identify a heroin death vs a non-heroin death. The availability of naloxone pharmacies in the area was not found to significantly impact the data.

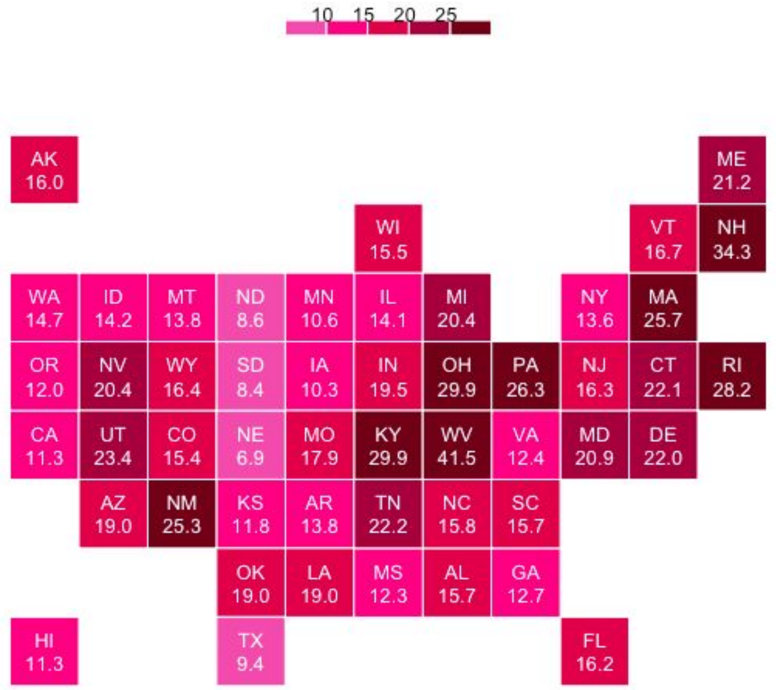
- Predictive modeling was also used on the Medicare Part D dataset to identify an Opioid Prescriber versus a Non-Opioid Prescriber and had an AUC of 0.868.

National Snapshot

Statistics from the CDC

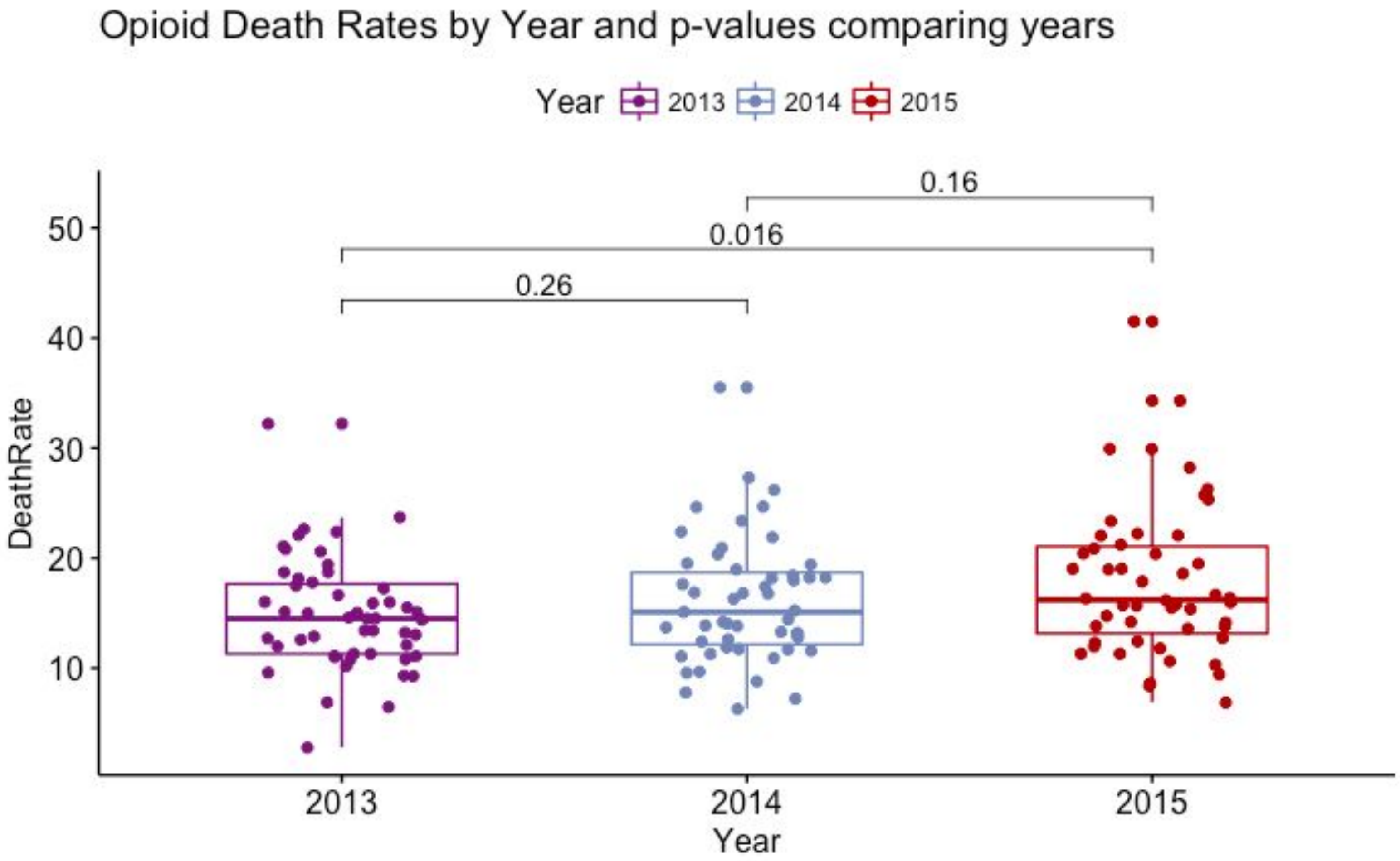


Opioid Death Rates per 100,000 people in 2013

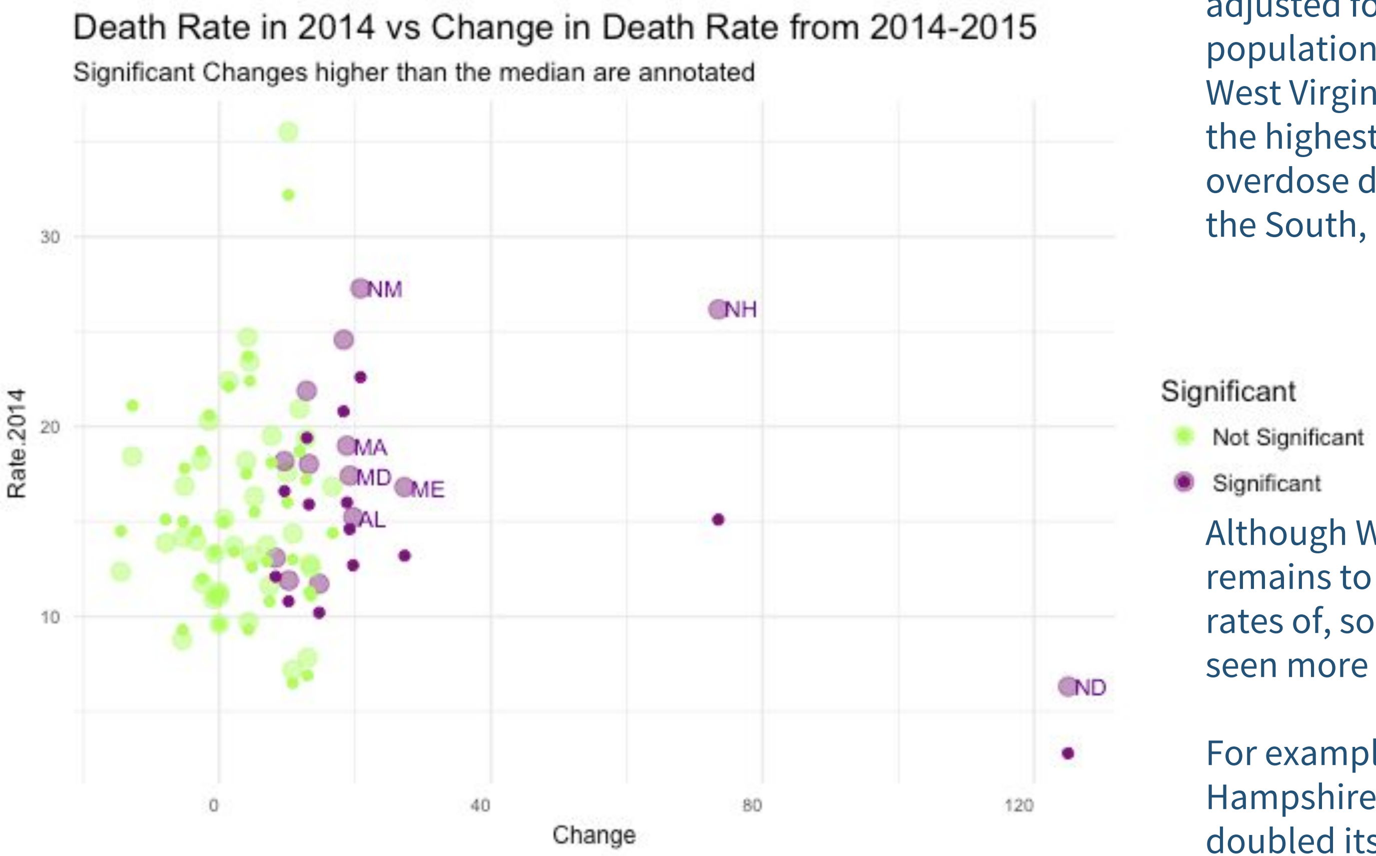


Opioid Death Rates per 100,000 people in 2015

Opioid Deaths on the Rise



Nationally, opioid death rates have **risen significantly.** These rates are age and population adjusted.



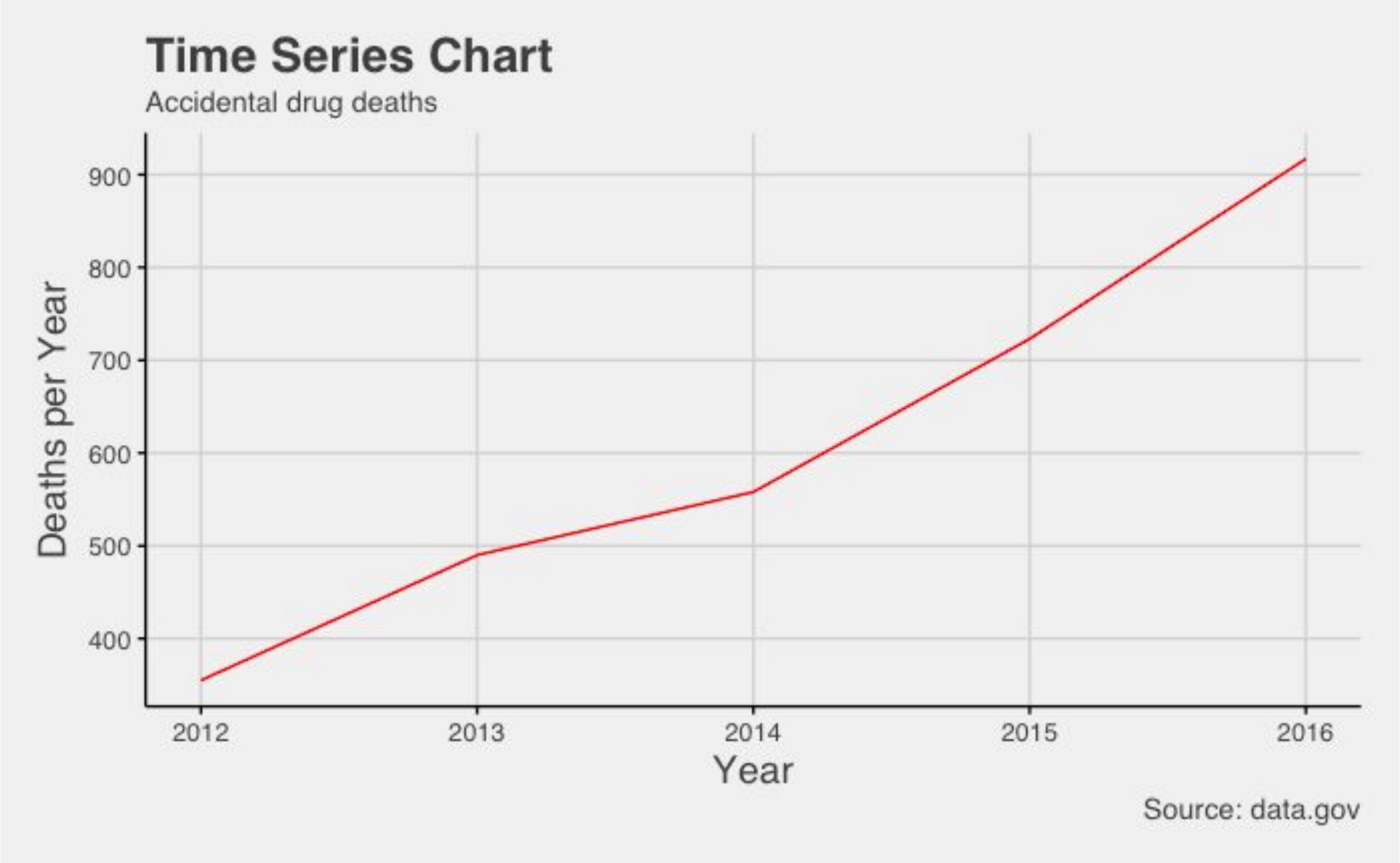
When Opioid Deaths are adjusted for age and population, we can see that West Virginia currently has the highest rates of opioid overdose deaths not only in the South, but for all states.

Although West Virginia remains to have the highest rates of, some states have seen more drastic changes.

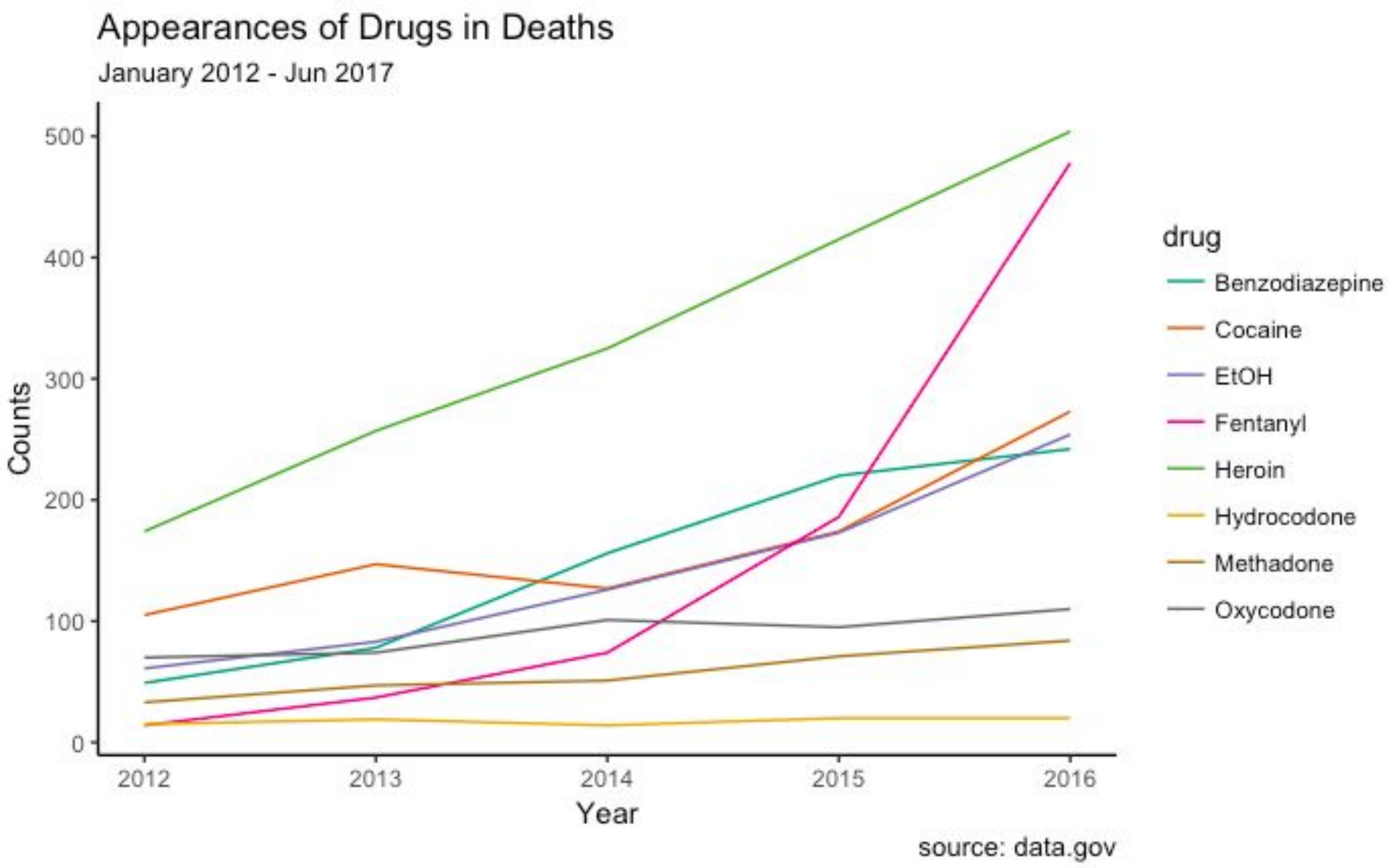
For example, New Hampshire has nearly doubled its Opioid Death Rate, with a rate of 35.5 in 2015.

Exploring the CT Accidental Drug Deaths

This comprehensive dataset consisted of over 3,500 unique observations and included data such as Case Numbers, Date of accidental drug death (from January 2012 to June 2017), Sex, Race, Age, Residence City, Death City, Injury Location, Death Location, County, Lat / Lon coordinates, Cause of Injury (Officer Notes), Heroin in Y/N format, Fentanyl Y/N format, Cocaine Y/N format, and other drugs provided in the Y/N format.



At first glance, drug deaths in CT by year appear to rise quite steadily.



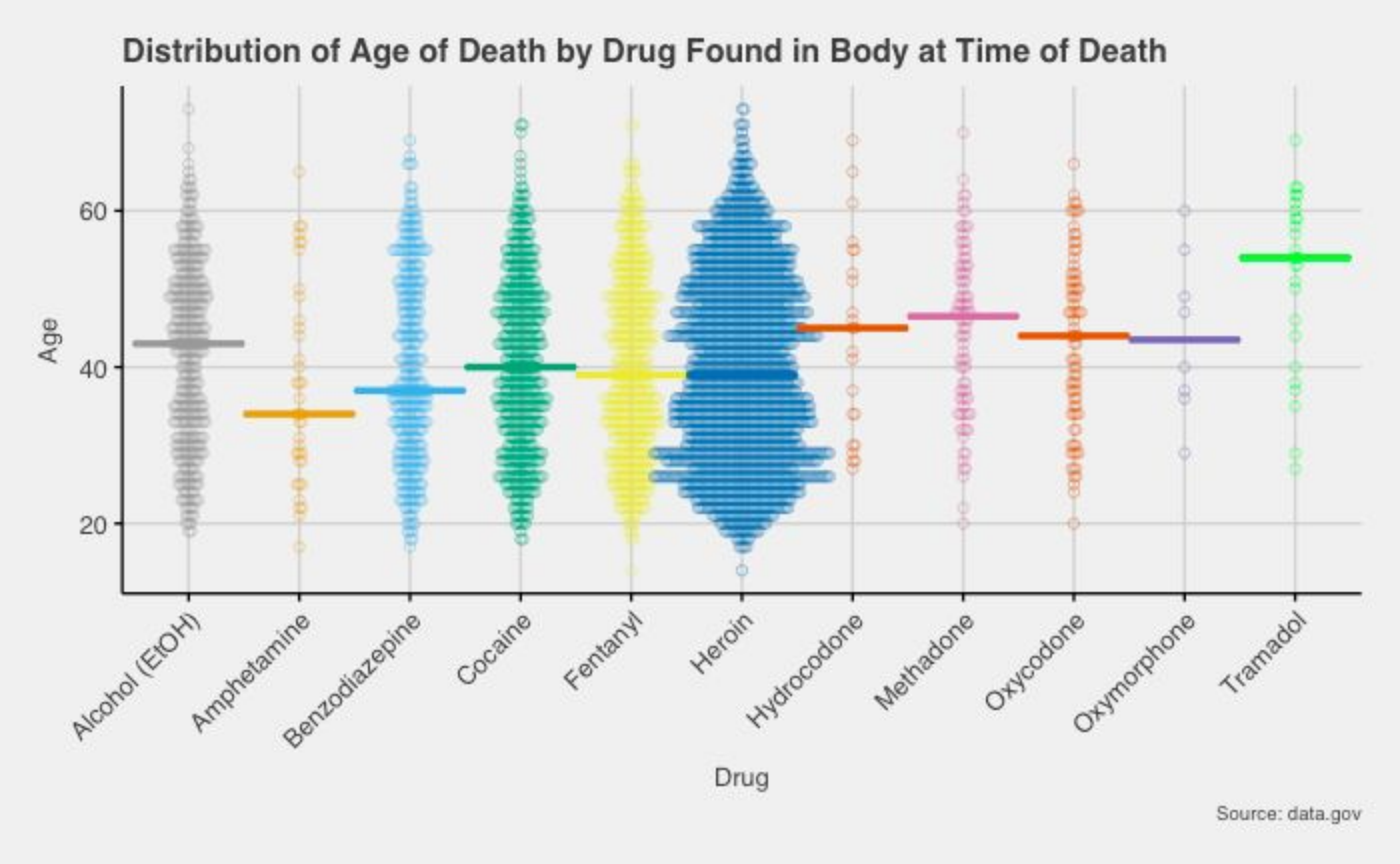
104 appearances of Heroin in Drug Deaths in 2012, 504 in 2016.

105 Fentanyl in 2012 to 478 in 2016.

49 Benzodiazepine in 2012, 242 in 2016.

Oxymorphone (*not included in this plot*) was found in only 9 deaths in 2016.

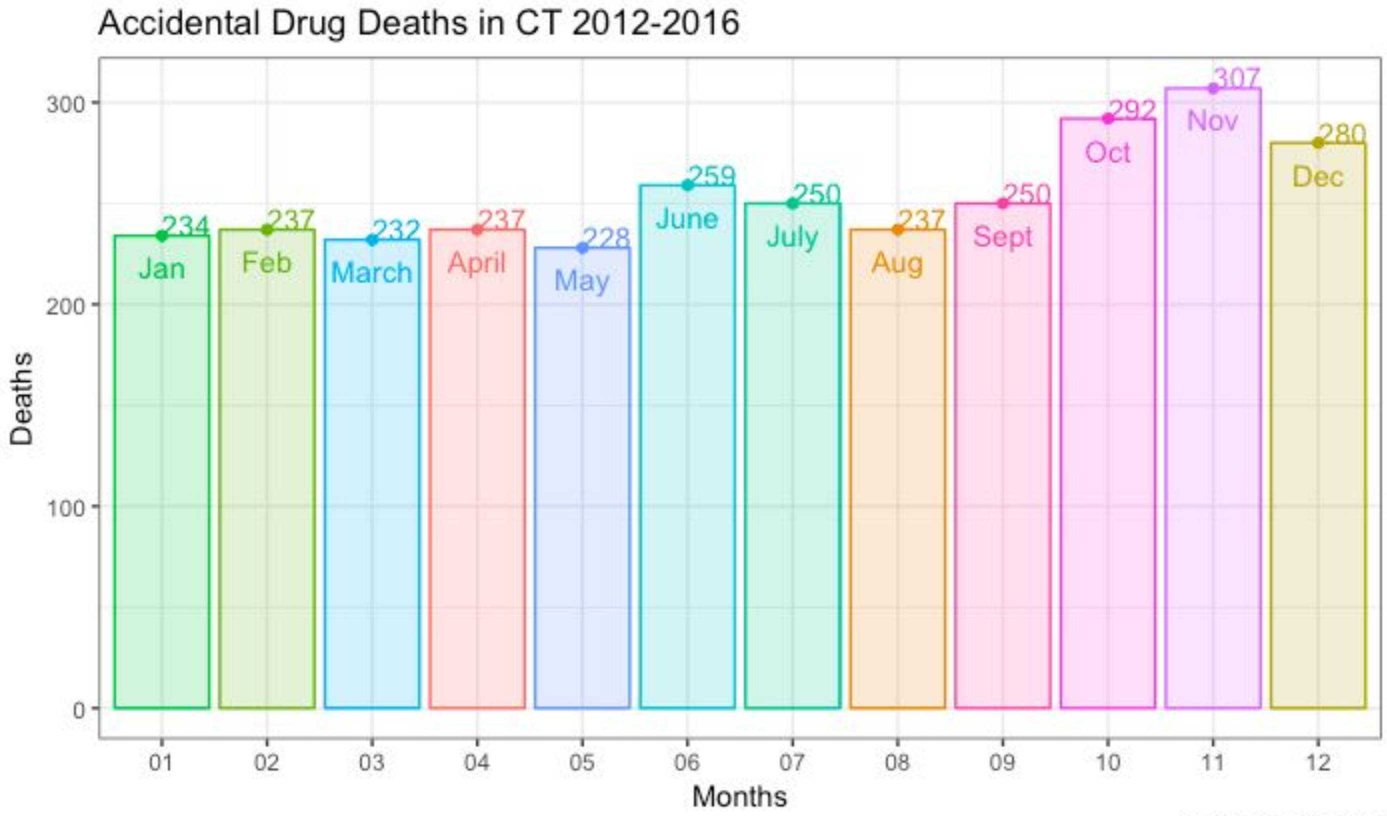
However, when deaths are separated by drug, we see that Fentanyl appearances in drug deaths are rising exponentially while Oxymorphone appearances has dropped.



Exploring the *CT Accidental Drug Deaths* dataset, we can see that the median age of a person who died with Amphetamine was the youngest at 34 years old while the median age of a person who died with Tramadol in their system was much higher, at 54 years old. The median age of death for heroin and fentanyl were both 39. The youngest person in the dataset died of **'Heroin and Fentanyl Intoxication'** at **age 14** at Hartford Hospital. The oldest heroin death age was 87.

exploratory data analysis: Accidental Drug Related Deaths 2012-June 2017

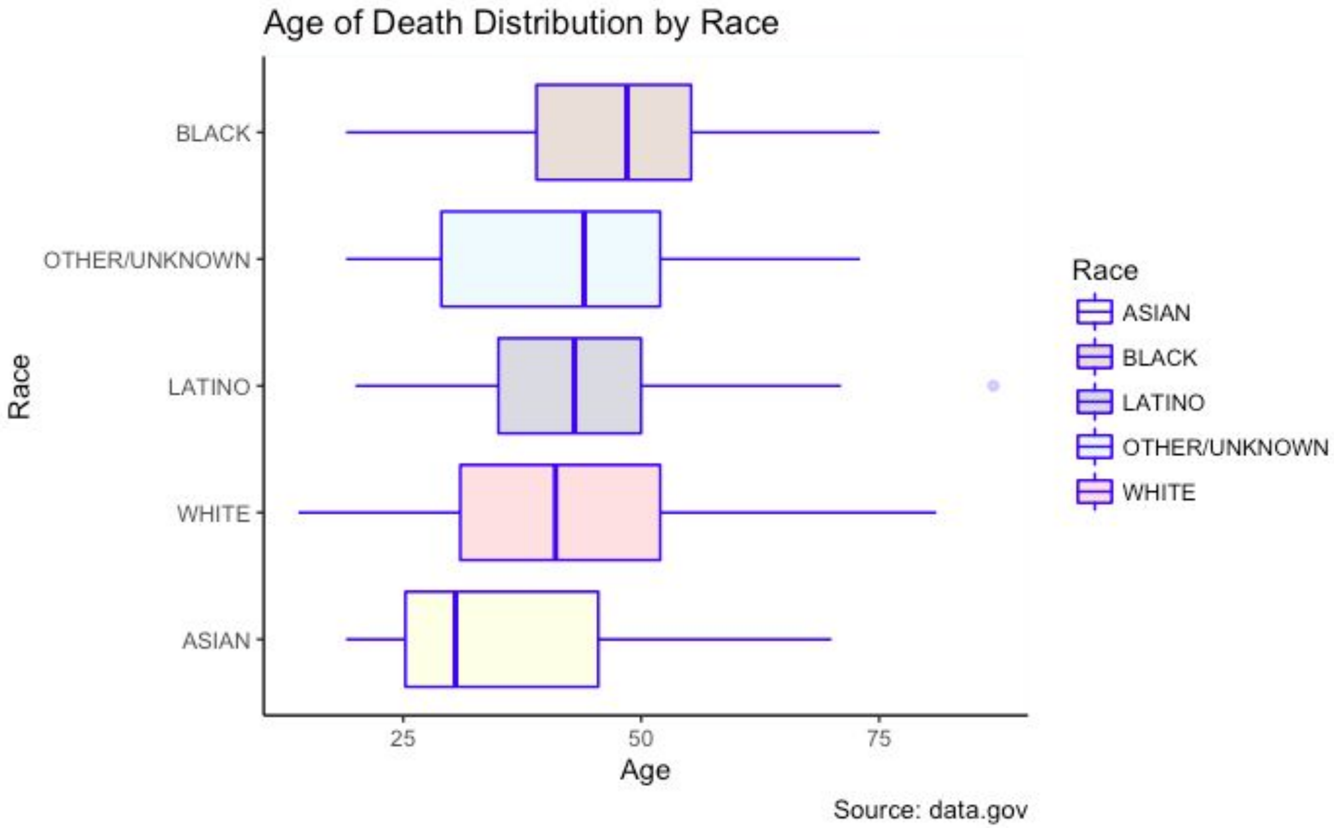
Looking through the data, findings showe that the most deaths occurred in the montl of October and November. Additionally, most deaths occurred on weekends.



source: data.gov

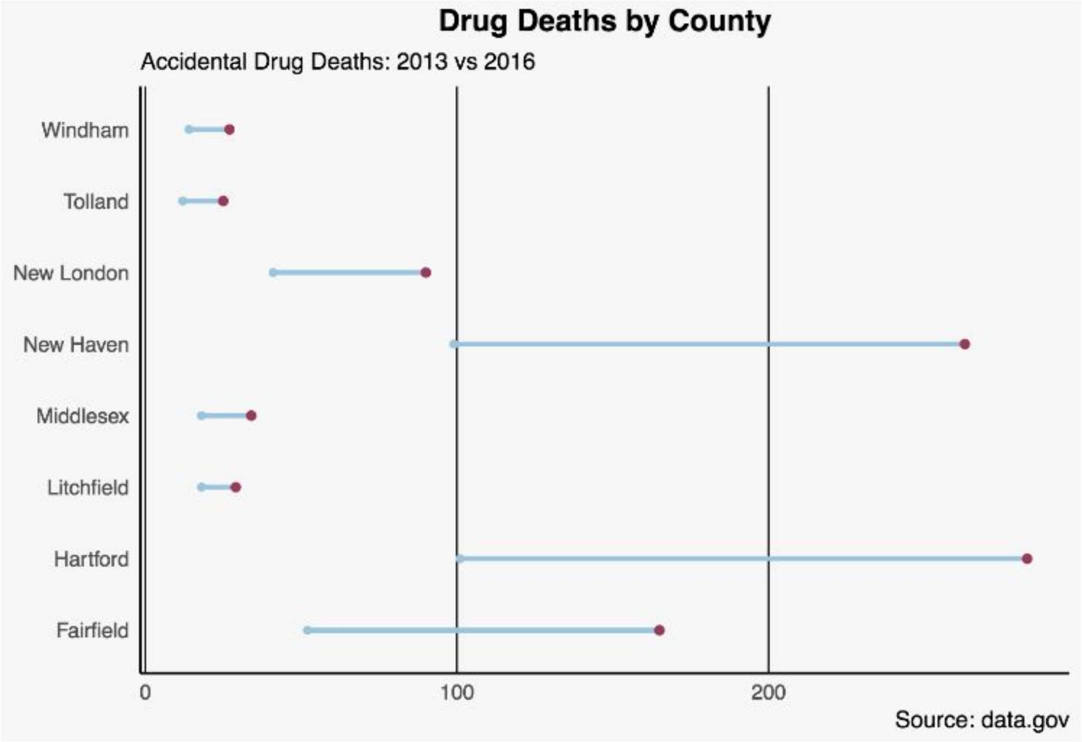
exploratory data analysis: Accidental Drug Related Deaths 2012-June 2017

A glance at the demographic data



exploratory data analysis continued

Hartford County saw the most accidental drug deaths in total throughout January 2012 - June 2017. This plot illustrates the jump in deaths from 2013 to 2016.

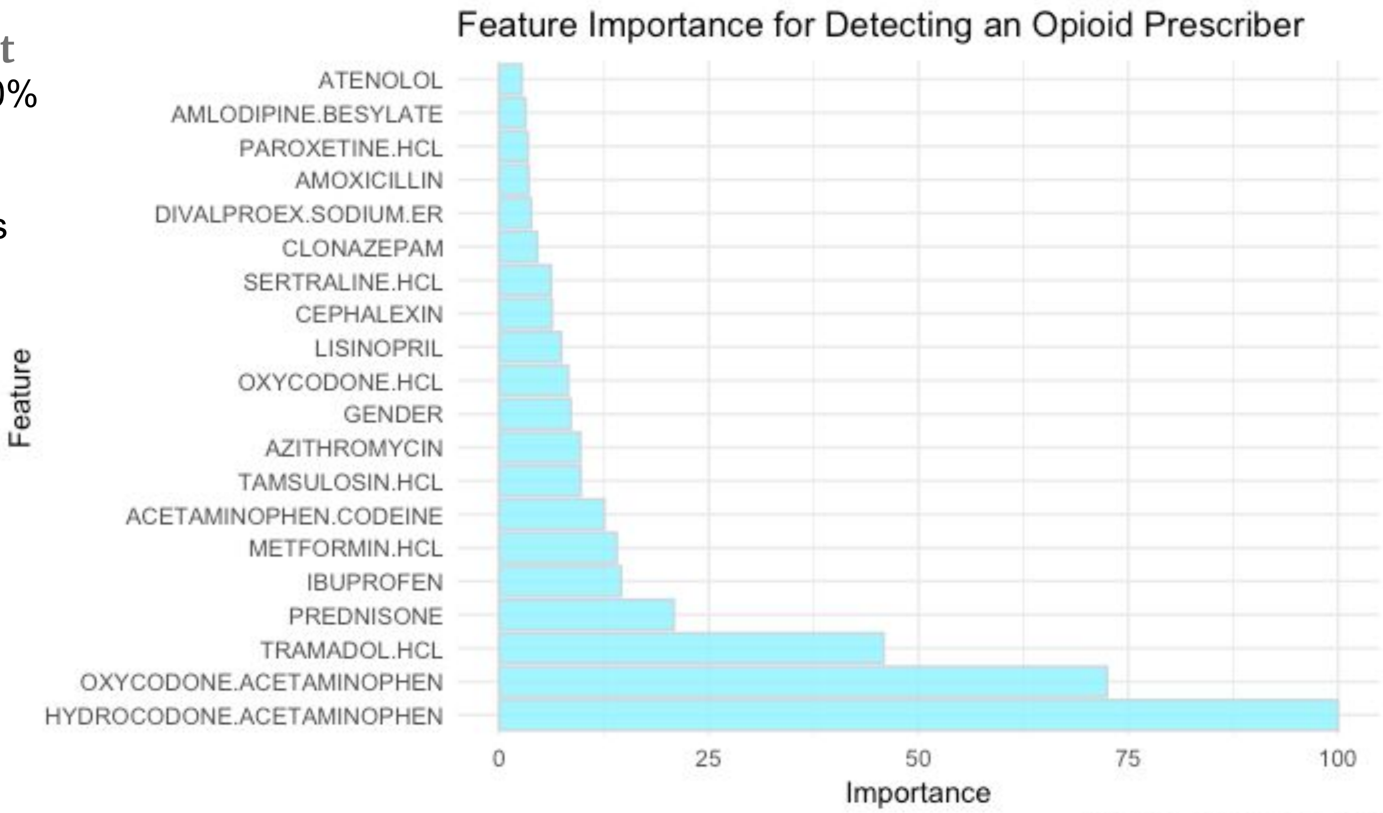


feature selection and prediction model

Part D Medicare Opioid Prescription data, subset to CT providers: Using 70% of this data and Stochastic Gradient Boosting with 280 samples and 251 predictors for two classes: 'no', 'yes' (Opioid Prescriber), these variables were shown as most important in detecting an opioid prescriber.

When tested on 30% of the data, area under the curve was **0.8681**.

Further plans include using this data to detect overprescription.



Source of data: cms.gov

other open datasets explored

Medicare Part D Opioid Prescriber Data

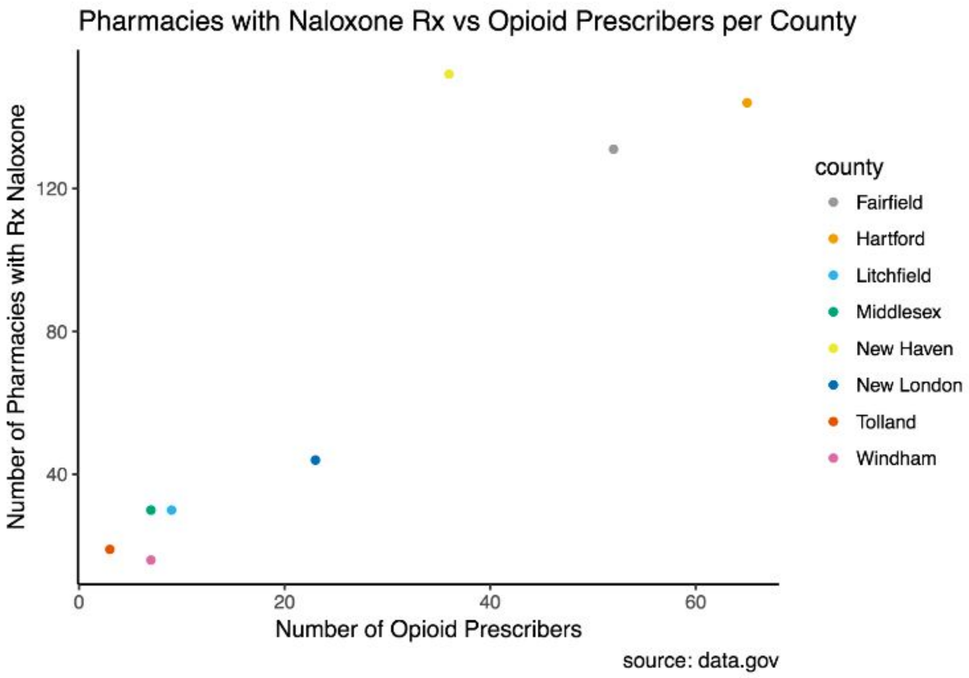
-provides NPI, NPI zip code, number of all prescriptions provided for the year, including non opioid Rx

Census Data (from noncensus package in R)

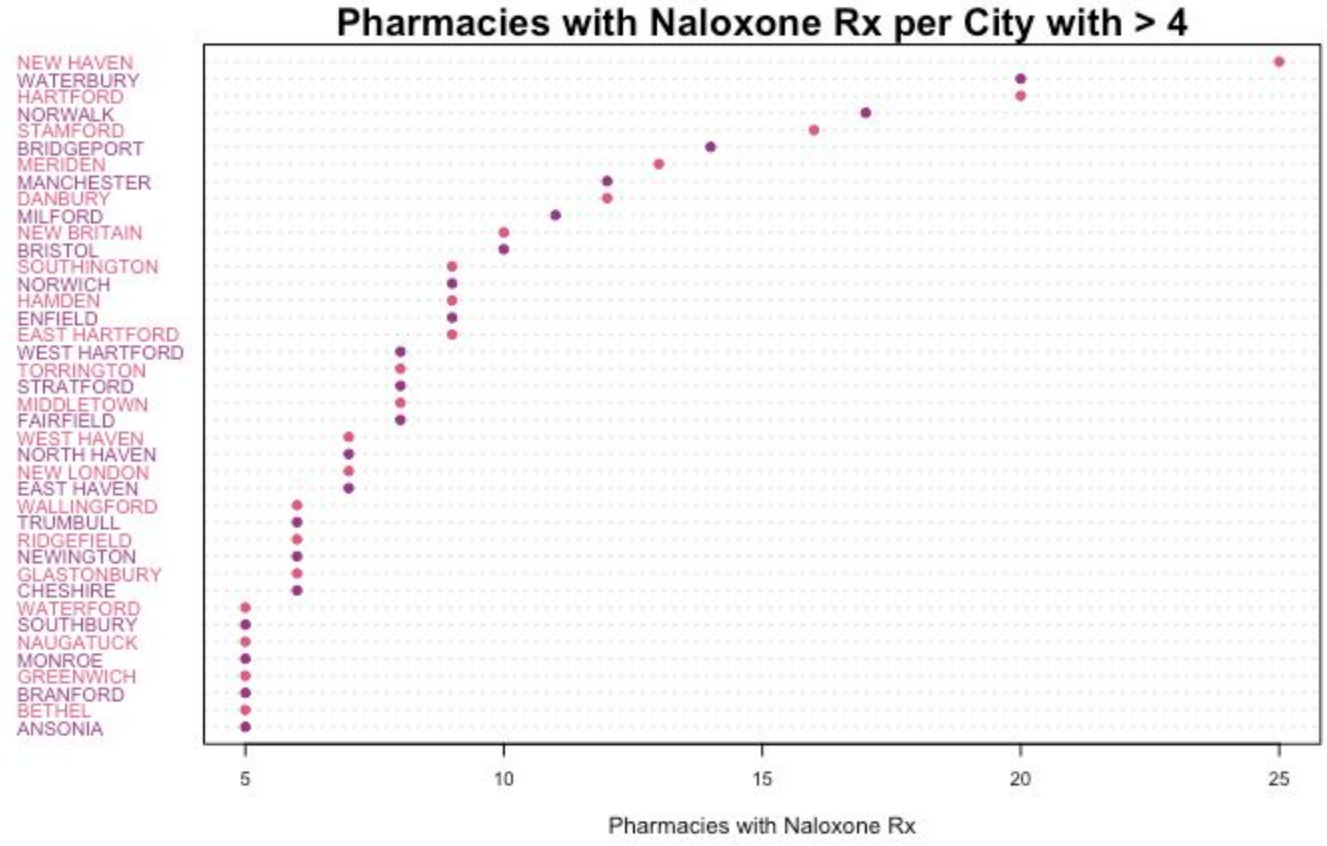
-from this could get population estimates, zip codes, counties

Pharmacies with Naloxone Rx

-List of pharmacies that provide Naloxone Rx including Lat/Lon data

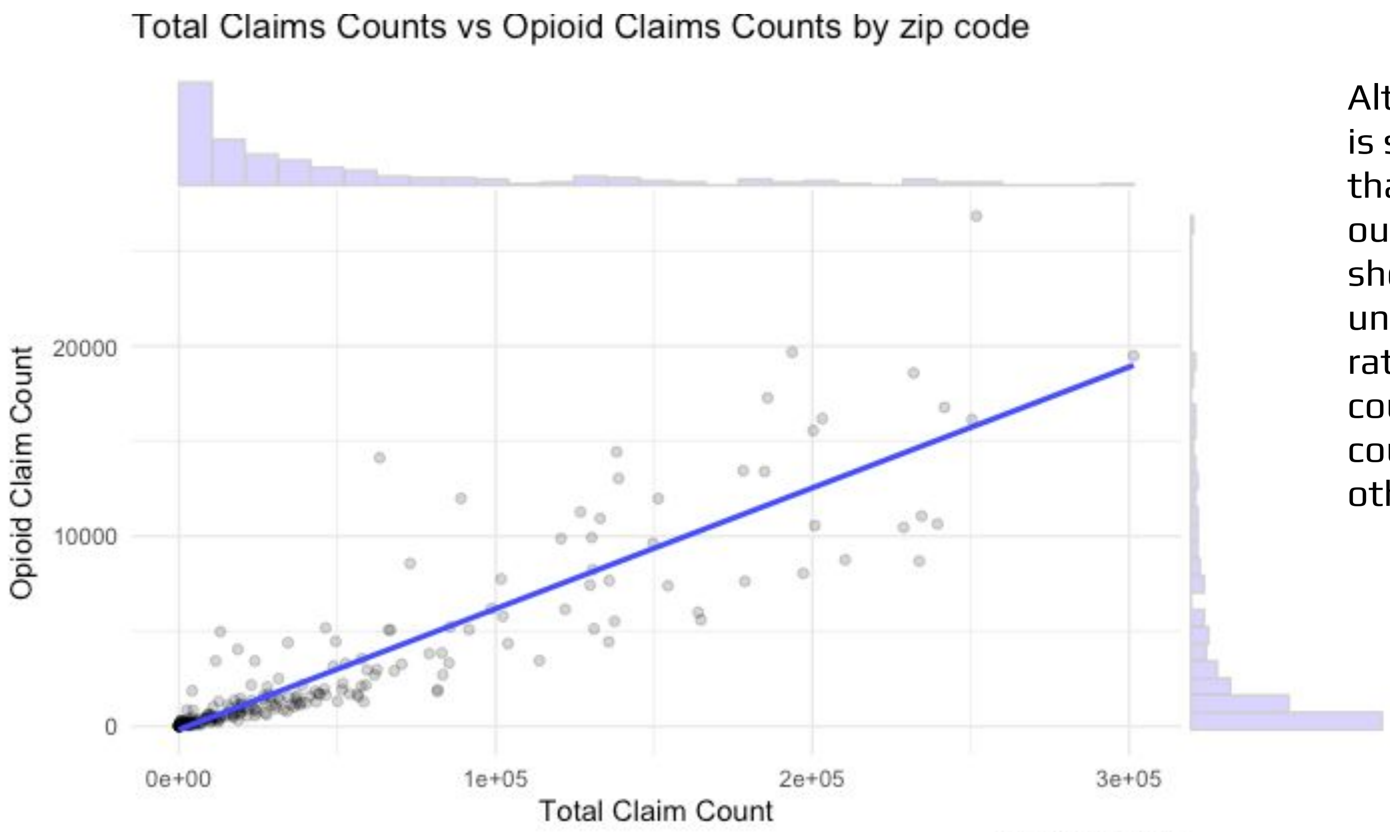


What cities have naloxone prescriptions?



In the final model for predicting a heroin death vs a non-heroin death, the presence and number of naloxone pharmacies in the area were not found to impact the prediction.

exploring opioid claims data



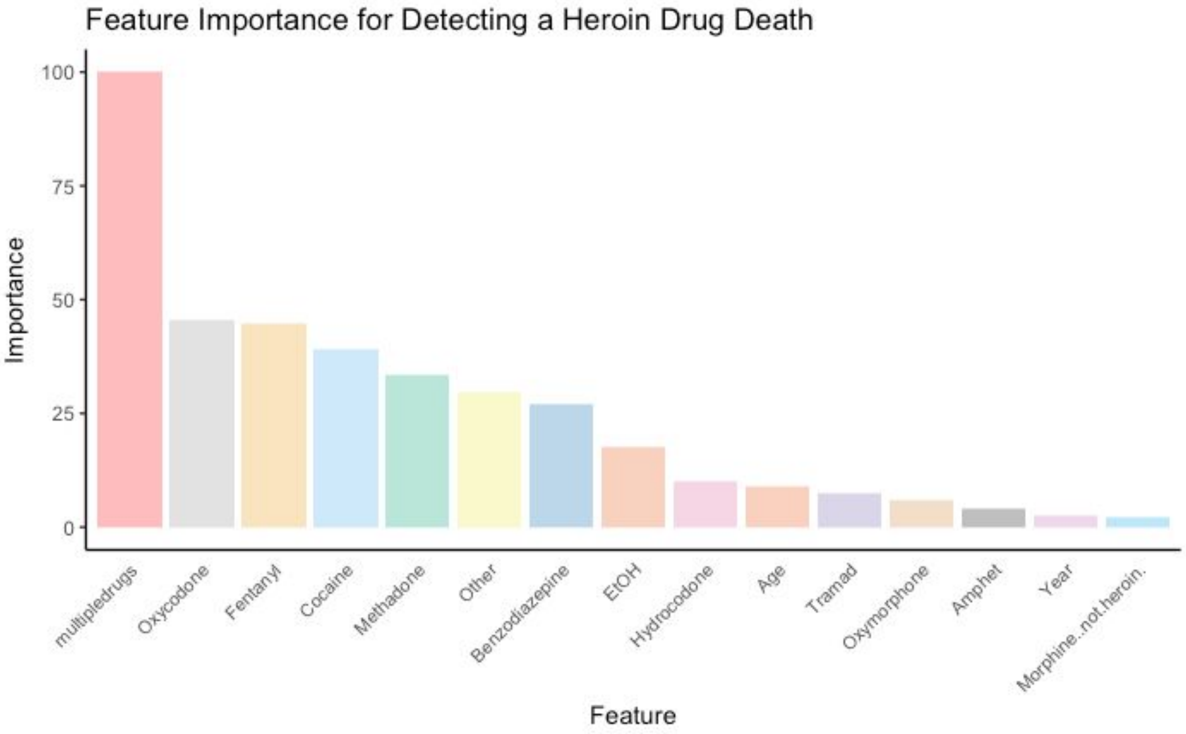
Although a linear model is shown, it can be seen that there are clear outliers. These outliers should be explored as to understanding why the ratio of opioid claim count to total claim count is unlike the others.

predictive model

machine learning methods:

Dummy encoding was used for categorical variables. The two classifiers were trained on 70% of the data (2505 observations) and used 237 predictor variables. Models were evaluated by applying 10-fold cross-validation, an established technique in supervised machine learning.

Results for the SGB model include: maximum ROC statistic of 0.9953425, a maximum Sensitivity statistic of 0.9408696, and a maximum Specificity statistic of 0.9763834

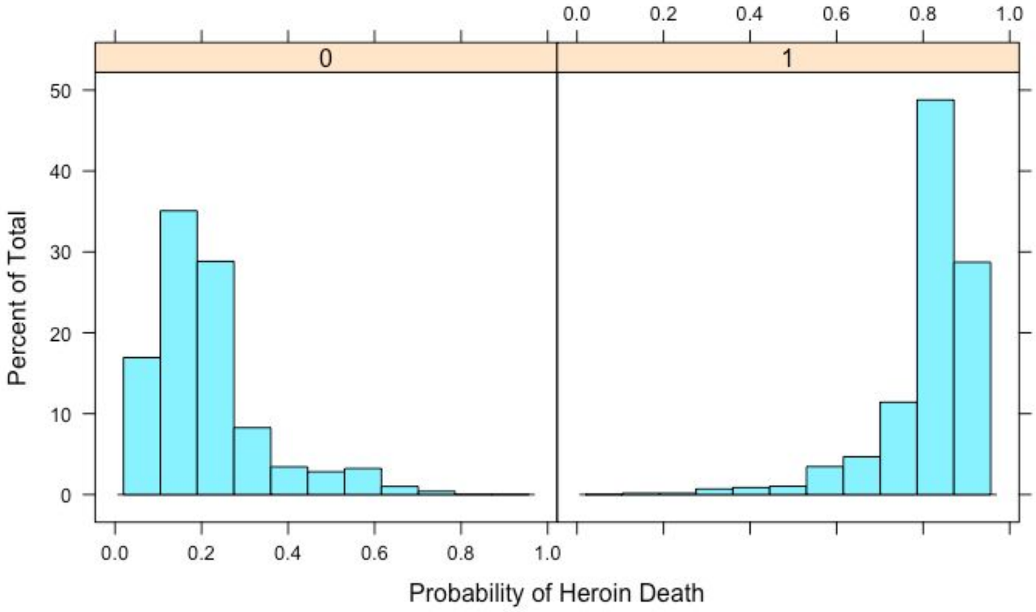


stochastic gradient boosting model

predictive model

machine learning methods:

Both the C50 model and SGB model had AUC scores over 0.99. Although performance of the models were high, limitations include that it was only using the CT Accidental Drug Deaths dataset and may not be sufficient in predicting outcomes with other state datasets. Nevertheless, my hope is that researchers can continue to use open data to find trends and to collaborate. ***Collaboration and open data are crucial in using technology to overcome some of the biggest challenges in substance abuse.***



Thank you

For any questions, please email Isabel Metzger at isabel.metzger@nyumc.org

Special thanks to my advisor Yindalon Aphinyanaphongs, MD, PhD and the Predictive Analytics Unit at NYU Langone Health!

Calendar Heatmap of CT Accidental Drug Deaths

references

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook. Computer and Information Technology: Computer Programmers. Jan 2018.

Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook. Office and Administrative Support: Customer Service Representatives. Jan 2018.

John A. Ramey (2014). noncensus: U.S. Census Regional and Demographic Data. R package version 0.1. https://CRAN.R-project.org/package=noncensus

CMS.gov. CMS 2016 Star Ratings Fact Sheet. Centers for Medicare and Medicaid Services. Available from: https://aishealth.com/sites/all/files/cms-2016-star-ratings-fact-sheet.pdf

CMS.gov. CMS Fast Facts. Centers for Medicare and Medicaid Services. February 2018 . Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Fast-Facts/index.html

"Accidental Drug Related Deaths 2012-June 2017." *Accidental Drug Related Deaths 2012-June 2017 - Data.gov*, Publisher Data.ct.gov, updated 3 Feb. 2018, catalog.data.gov/dataset/accidental-drug-related-deaths-january-2012-sept-2015

"Pharmacies Offering Narcan, Evzio and Other Brands of Naloxone." Pharmacies Offering Narcan, Evzio and Other Brands of Naloxone - Data.gov, Publisher Data.ct.gov, updated 28 Feb. 2018, catalog.data.gov/dataset/pharmacies-offering-narcan-evzio-and-other-brands-of-naloxone