Analyzing Opioid Overdose Deaths

Using supervised learning algorithms to predict opioid overdose deaths using Medicare and Medicaid prescription data

The Problem

63,632

American killed by drug overdoses in 2016

The Opioid Epidemic

The opioid epidemic is a nationwide problem involving both prescription painkillers and illegal opioids, both of which can lead to potentially fatal overdoses.

It began in the late 1990s when pharmaceutical companies began promoting prescription painkillers with promises of pain relief without addiction.

Doctors began prescribing them more frequently and opioid sales skyrocketed.

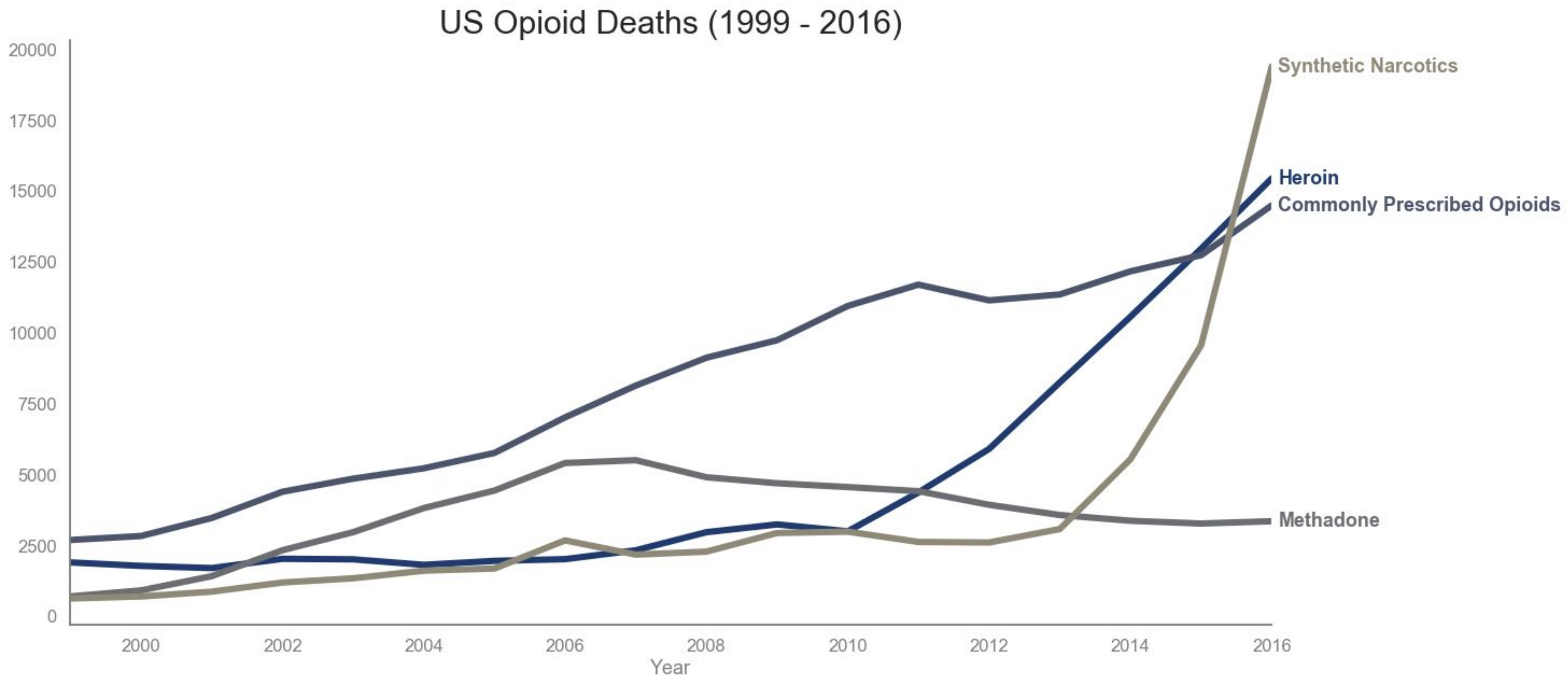
Thousands of individuals began abusing the pain relievers, many of whom became addicted or overdosed.

OxyContin Abuse

Oxycontin debuted in 1996, and was heavily marketed by its maker, Purdue Pharma, for providing 12 hour pain relief.

However, even before hitting the market, clinical trials showed patients were only getting about 8 hours of pain relief. As the effects wear off patients experience withdrawal symptoms - including intense drug cravings - which promotes addiction and dependence.

As stated in this LA Times Article over 7 million Americans have abused OxyContin - the drug that is often blamed with kicking off the opioid epidemic.



Opioid deaths in the US have increased between 1999 and 2016.

Opioid Overview

Opioids are a class of highly addictive narcotic substances, both legal and illegal that are derived from the opium poppy plant. They can be prescription painkillers (Percocet, MS Contin, Duragesic) or street drugs (Heroin).

All opioids stimulate the opioid receptors in the brain. While they vary in how powerful they are, all depress the central nervous system (slowing down body functions - including breathing) and reduce both physical and psychological pain.

Opioid Effects

Beneficial Effects

- Pain relief
- Feelings of relaxation and well-being
- Drowsiness
- Suppressed Coughing

Adverse Effects

- Constipation
- Nausea and Vomiting
- Lethargy
- Mood Swings
- Addiction
- In overdose: Respiratory Failure

Potential Impact

Learn trends that contribute to increased deaths in order to prevent them.

Improve pain management guidelines.

Reduce access to opioids.

Improve patient quality of care.

The Data

Medicaid Data

This dataset reports drug utilization data for covered outpatient drugs that are paid for by state Medicaid agencies.

The data includes the state, drug name, NDC, number of prescriptions, and dollars reimbursed.

Medicare Part D Data

The 2016 Medicare D Opioid prescriber summary file presents the rates opioid prescribing rates of providers who participate in the Medicare Part D program. It is a prescriber level dataset with information on the number and percentage of opioid claims with prescriber specialty, state, and ZIP code.

The Medicare Provider Utilization and Payment dataset contains drug level information, listed by brand and generic name, for prescriptions by Part D participating providers.

Drug Data

Drug data is from the HEDIS Medication List Directory. Drugs are listed by brand and generic names, and drug class.

Using this data I was able to filter out only drugs that are classified as opioids.

Death Data

Death data was obtained from CDC WONDER.

Data are based on death certificates and contains the single underlying cause of death and up to 20 additional multiple causes of death.

Deaths are classified using ICD–10 codes.

Drug-poisoning deaths are identified using underlying cause-of-death codes:

X40–X44 (accidental poisoning)

X60–X64 (intentional self-poisoning)

X85 (assault by drugs)

Y10–Y14 (poisoning by undetermined intent)

Death Data Continued

Drug Poisoning Deaths are further classified by ICD-10 Category Codes:

- **T40.0 Opium**
- **T40.1 Heroin**
- **T40.2 Other Opioids**
  - **Oxycodone, Hydrocodone, Morphine, Hydromorphone**
- **T40.3 Methadone**
- **T40.4 Other Synthetic Narcotics**
  - **Fentanyl, Tramadol**
- T40.5 Cocaine
- T40.6 Other and Unspecified Narcotics
- T40.7 Cannabis Derivatives
- T40.8 Lysergide (LSD)
- T40.9 Other and Unspecified Psychodysleptics (Hallucinogens)

Feature Engineering And Data Cleaning

Filtering on Morphine Equivalent Dose

Morphine Equivalent Dose (MED) is used by clinicians to compare different opioids relative strengths.

Using the drugs dataframe I dropped all drugs where MED was null, as those were not opioids.

I then used the new opioid only drug data to filter out the rows of interest in the Medicare and Medicaid datasets.

Aggregating on State

I had to combine my four data sources together to prepare for modeling.

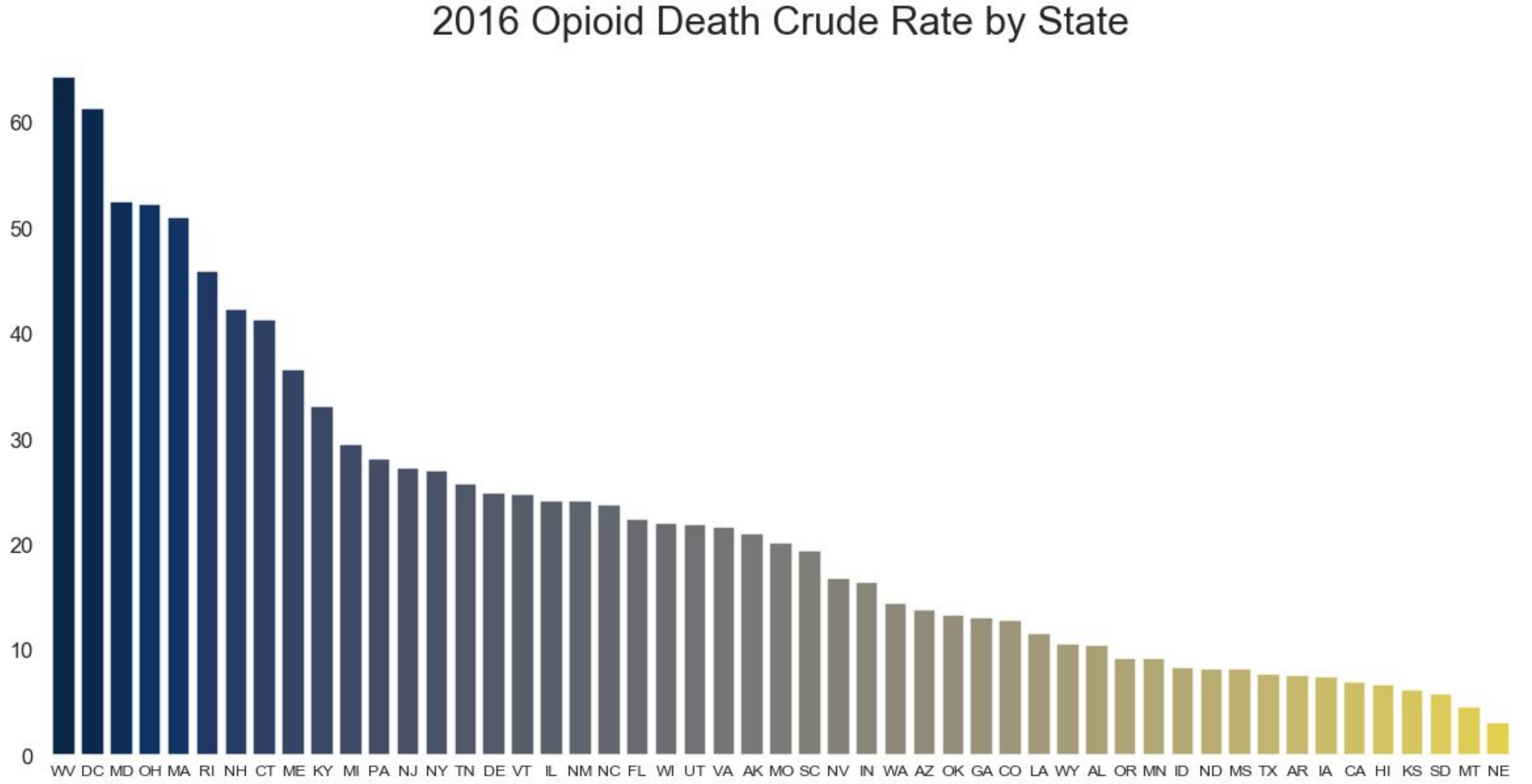
State is the common element so I merged all data into one dataframe using states as rows.

Feature Engineering

I created new features from the data:

- Claim Count Mean and Median
- 30 Day Fill Count Mean and Median
- Medicare Days Supply per Beneficiary
- 30 Day Fill Count Mean and Median
- Medicaid Units per Prescription

Exploratory Analysis



In 2016, West Virginia, DC, Maryland, and Ohio had the highest amount of opioid deaths when normalized for population.

Ohio had the highest crude rate deaths in 2016.

Could this be attributed to opioid prescribing?

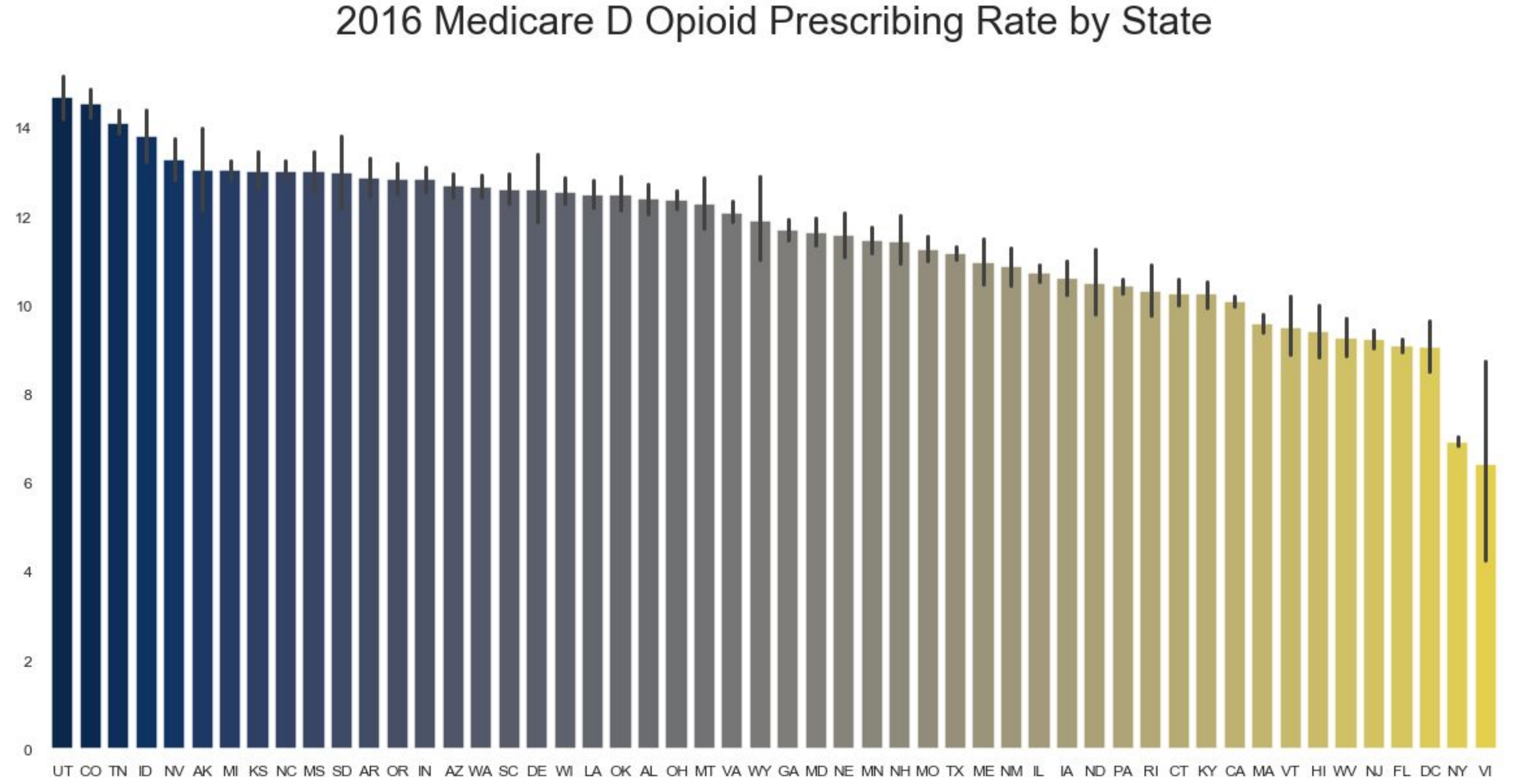
Opioid Prescribing Rate:

Ohio vs. United States

Ohio’s Medicare Part D Opioid Prescribing Rate is **significantly different** from the United States as a whole:

T - Value: -7.90

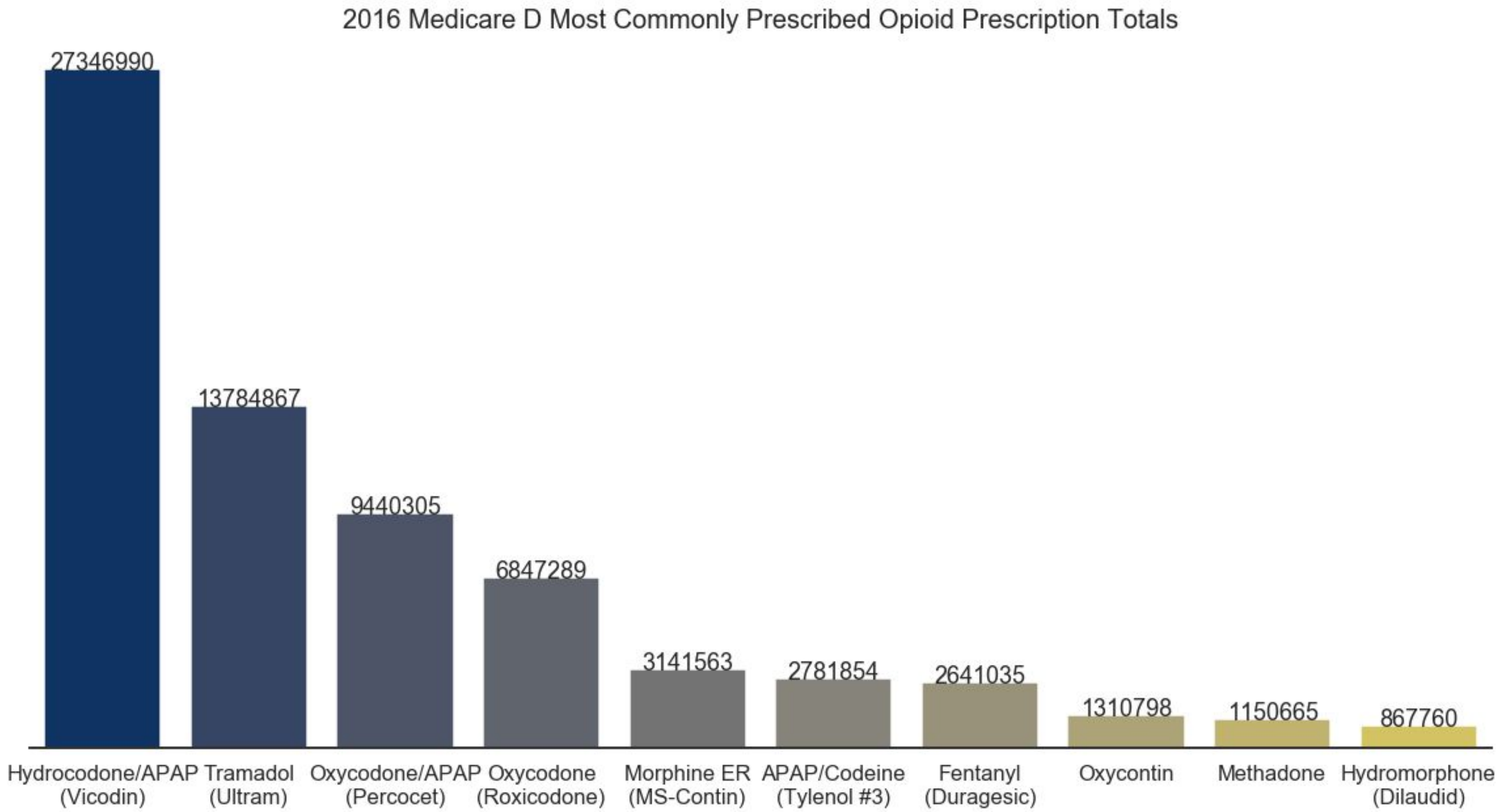
P - Value: 3.38e-15



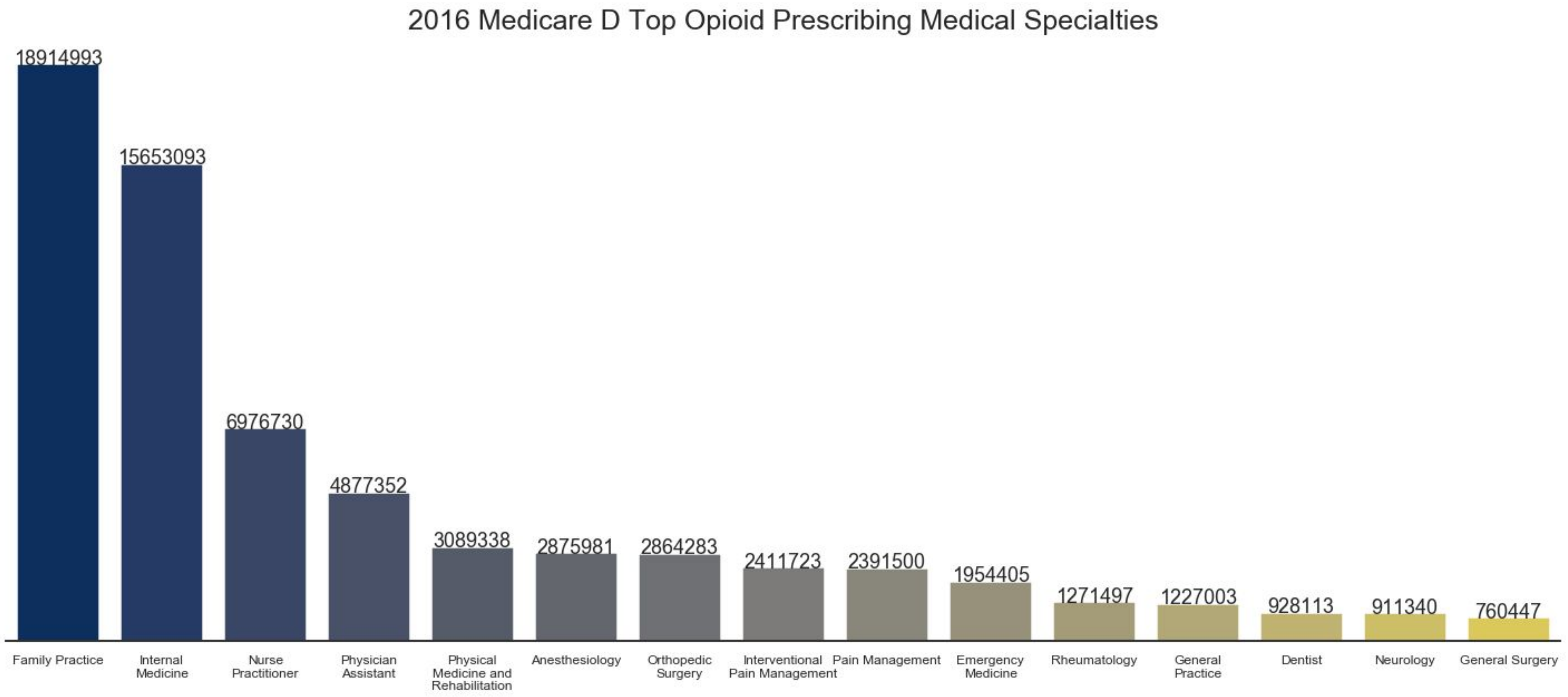
Opioid prescriptions accounted for 6.9% - 14.6% of total Medicare prescriptions.

21,772

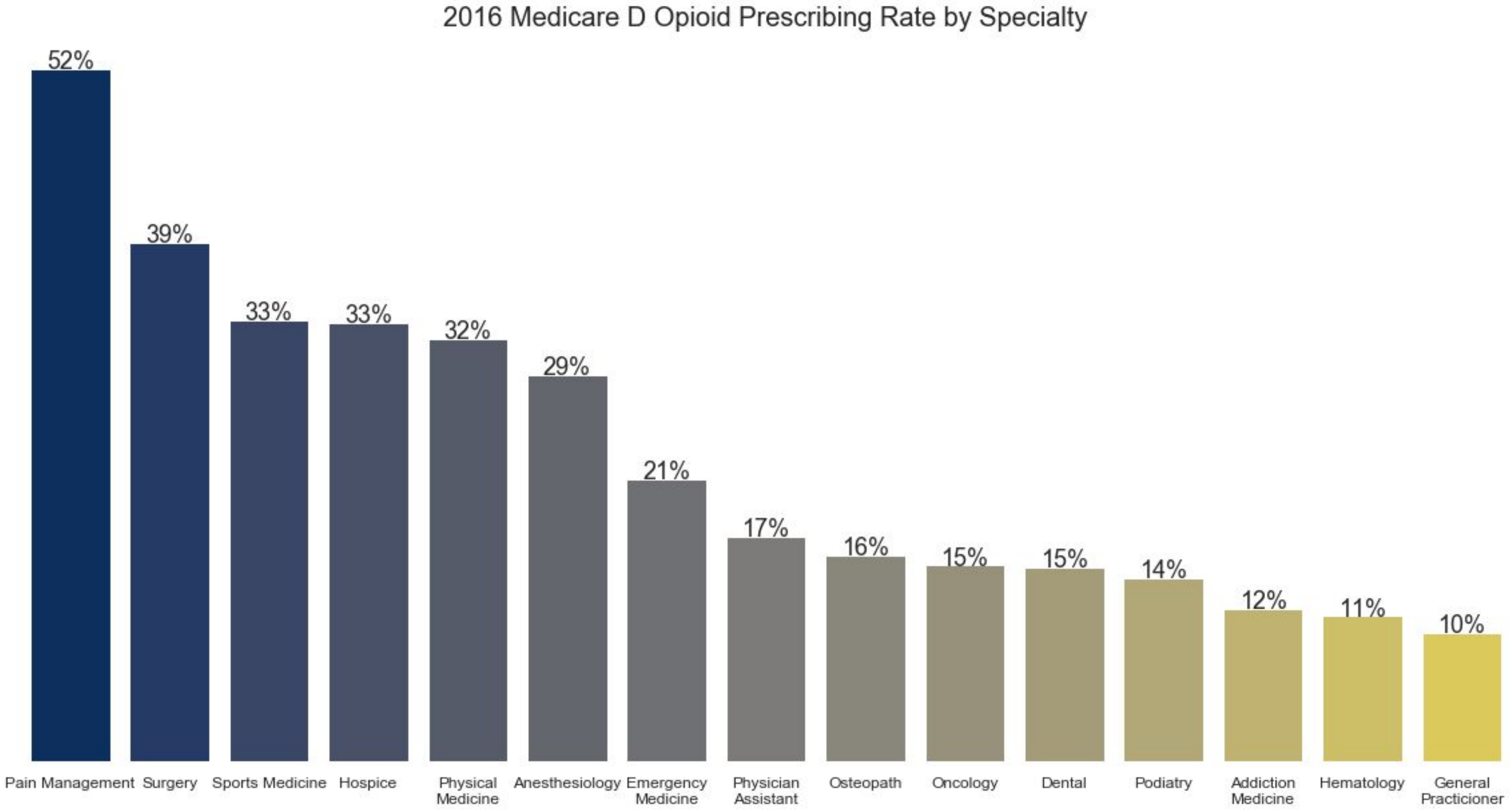
Medicare Part D opioid prescriptions written by a single doctor in 2016.



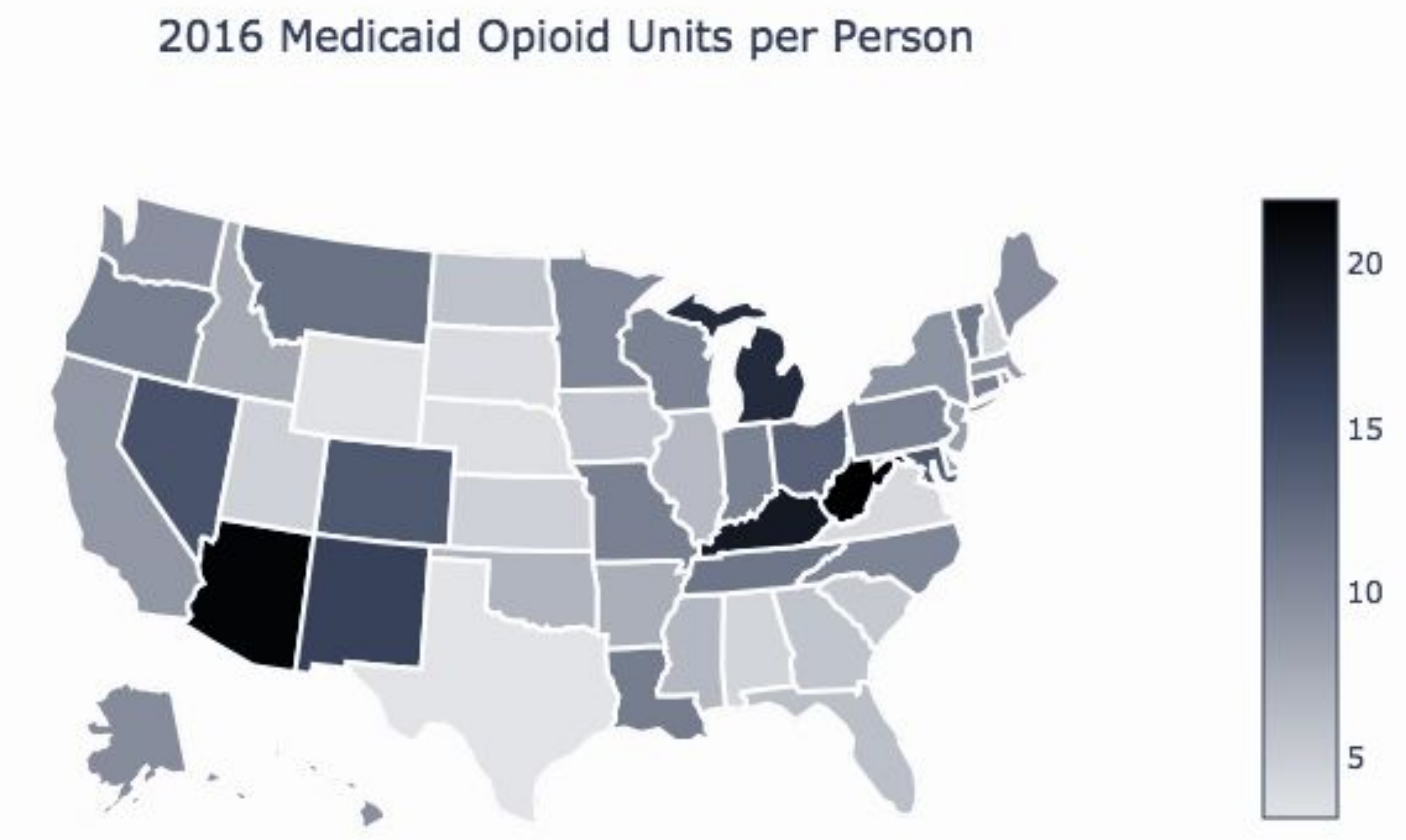
Vicodin was the most commonly prescribed opioid in 2016 with over over 27 million prescriptions.



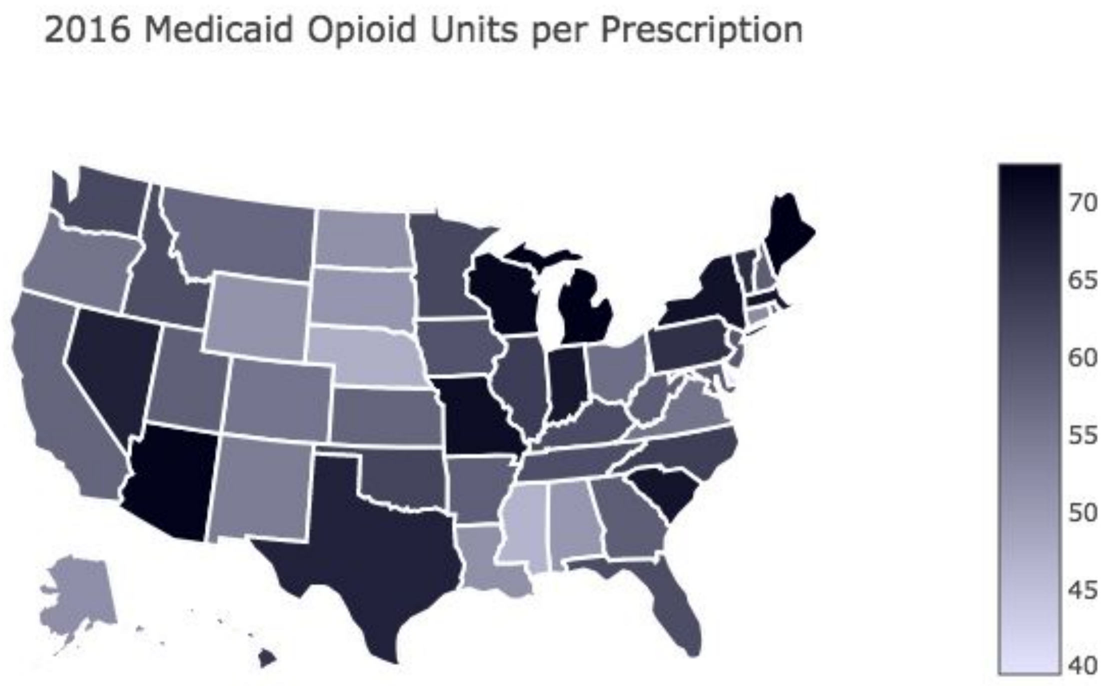
In 2016, the greatest number of opioid prescriptions were written by general practitioners: Family and Internal Medicine, Nurse Practitioners and Physician Assistants.



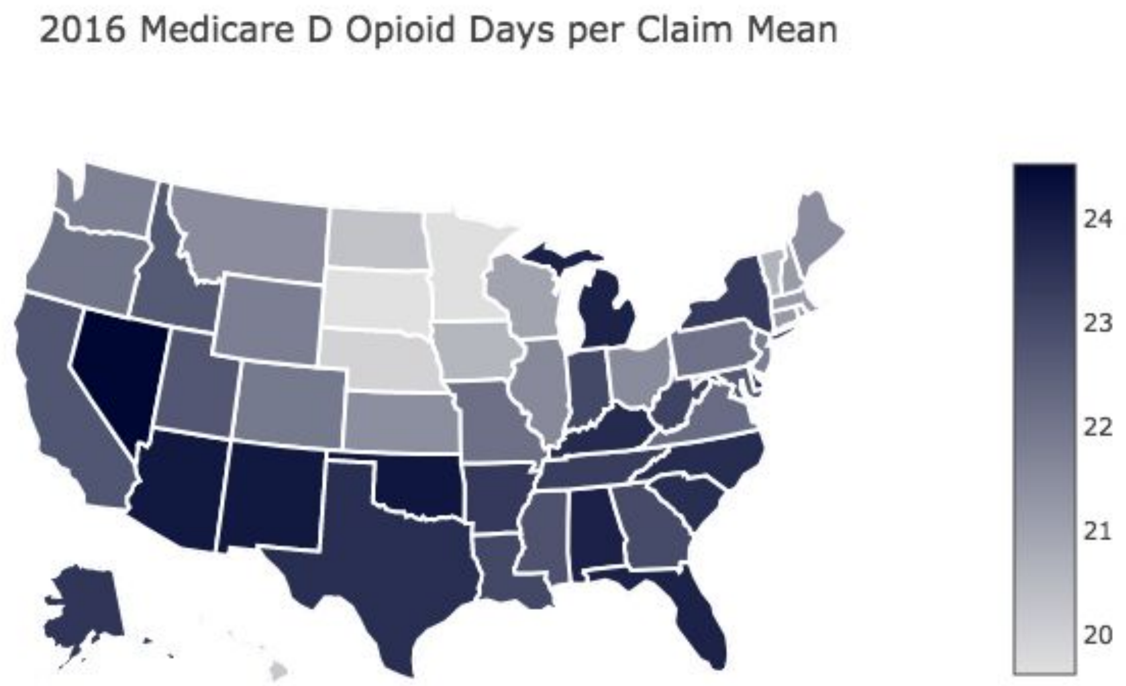
In 2016, the top opioid prescribing specialties by percentage of total prescriptions are pain management, surgery, sports medicine, and hospice.



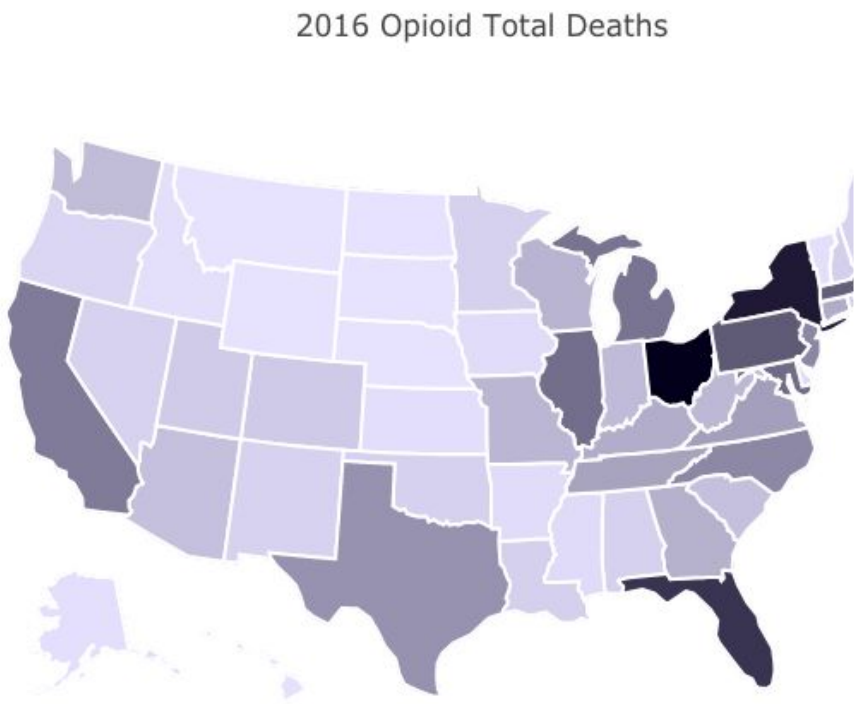
If we evenly distributed only the Medicaid opioid units dispensed in 2016 amongst the state's populations every resident of Arizona, Kentucky, and West Virginia would receive over 20 units each!

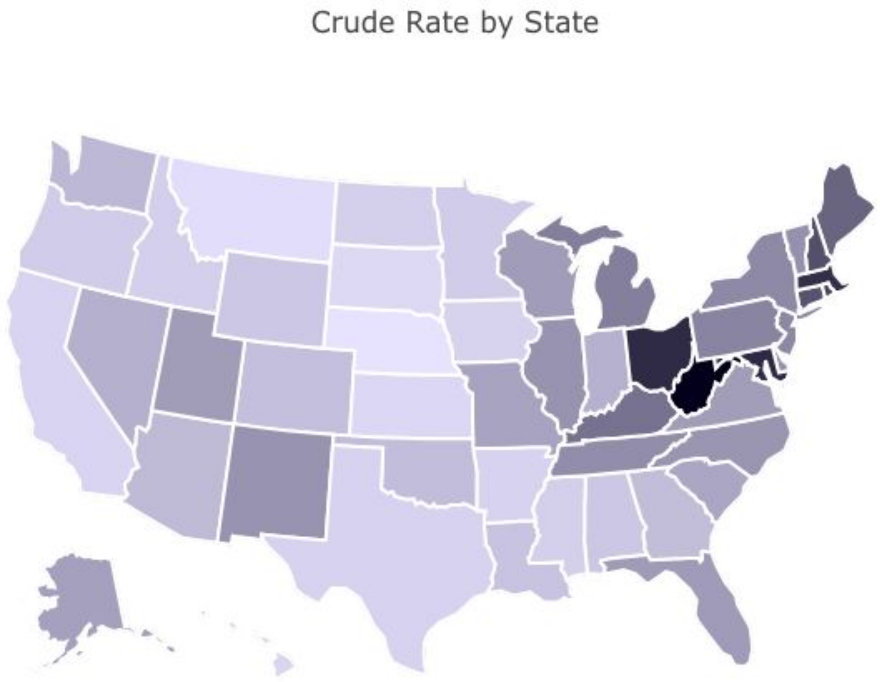


Opioid units per prescription varies widely across states with mean values between 40 to over 70 units.



The mean days supply for Medicaid prescriptions ranges from 20 to 30.





Total Opioid Deaths vs Crude Rate Deaths (Opioid Deaths per 100,000 population)

Models

Initial Modeling Results

Algorithm	Cross Validation Accuracy	Accuracy Mean
Decision Tree	0.50, 0.12, -0.02, 0.12, 0.04	0.16 (+/- 0.37)
SGD	-2.86, -1.81, -3.00, -4.47, -2.21	--2.87 (+/- 1.82)
Linear Regression	**0.86, 0.90, 0.88, 0.81, 0.77**	**0.85 (+/- 0.10)**
Bagging Regressor	**0.91, 0.89, 0.92, 0.91, 0.79**	**0.89 (+/- 0.09)**
Random Forest	**0.92, 0.86, 0.96, 0.91, 0.78**	**0.89 (+/- 0.12)**
Extra Trees	**0.90, 0.91, 0.97, 0.96, 0.87**	**0.93 (+/- 0.07)**

Relative Feature Importances

Are the models predicting deaths from the prescription data?

Modeling Results – Removing All Death Data

Algorithm	Cross Validation Accuracy	Accuracy Mean
Linear Regression	**0.14, 0.27, -0.00, -0.18, 0.08**	**0.06 (+/- 0.31)**
Bagging Regressor	0.21, 0.04, 0.63, -0.67, -0.07	0.03 (+/- 0.86)
Random Forest	0.09, -0.01, 0.43, -0.57, 0.07	00.00 (+/- 0.65)
Extra Trees	0.37, 0.20, 0.09, -0.39, -0.12	0.03 (+/- 0.53)

Removing all death data significantly affects accuracy.

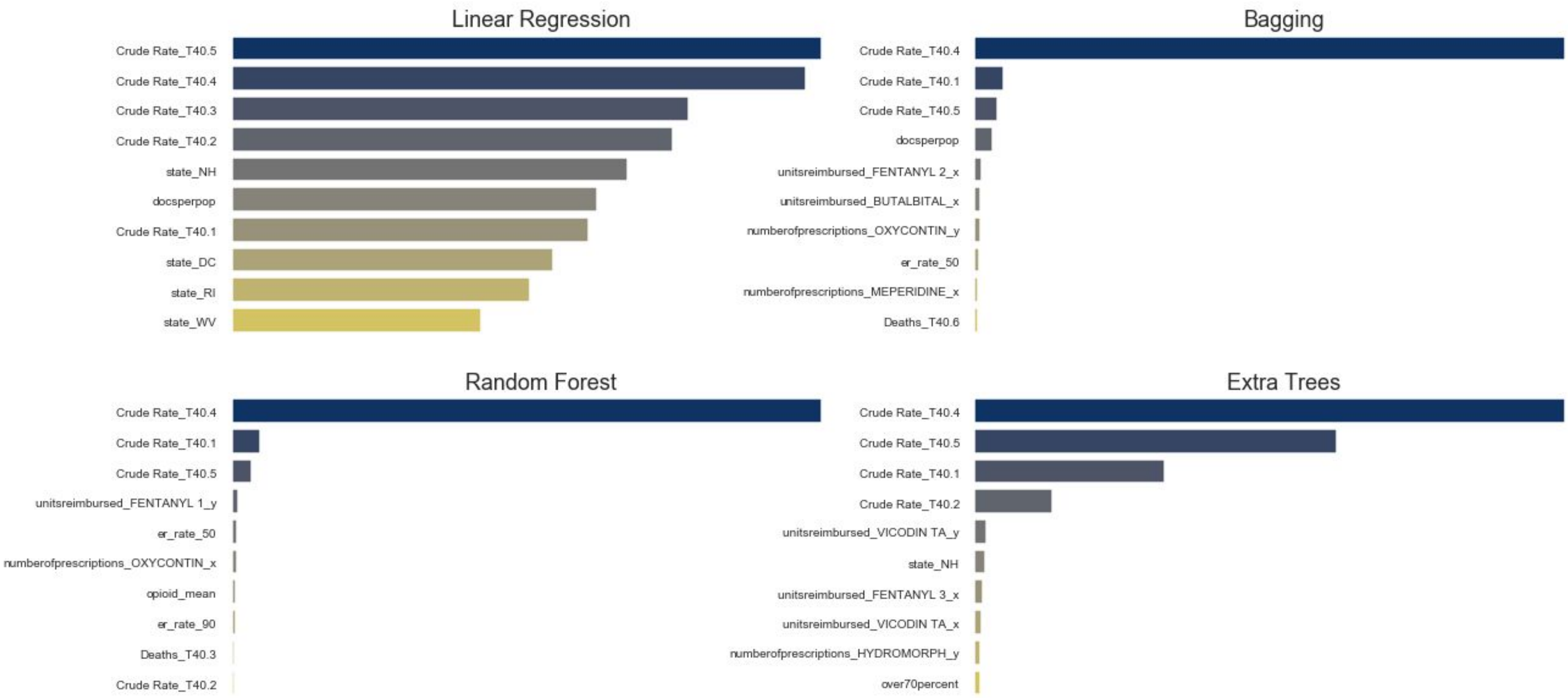
Modeling Results - Target Prescription Opioid Deaths Only

Algorithm	Cross Validation Accuracy	Accuracy Mean
Linear Regression	**-0.00 , 0.53, 0.57, -0.05, 0.39**	**0.29 (+/- 0.53)**
Bagging Regressor	-0.00 , -0.23, -0.19, -0.29, -0.04	-0.15 (+/- 0.23)
Random Forest	0.04, 0.49, 0.57, -0.57, 0.28	0.17 (+/- 0.83)
Extra Trees	0.17, 0.45, 0.61, -0.58, 0.27	0.19 (+/- 0.82)

Predicting only deaths likely due to prescription opioids improves accuracy slightly.

Modeling Results - Predicting Crude Rate with additional drug and doctor features and death data

Algorithm	Cross Validation Accuracy	Accuracy Mean
Linear Regression	0.72, 0.77, 0.66, 0.39, 0.63	0.64 (+/- 0.26)
Bagging Regressor	**0.84, 0.821, 0.92, 0.61, 0.84**	**0.81 (+/- 0.20)**
Random Forest	**0.83, 0.89, 0.95, 0.75, 0.87**	**0.86 (+/- 0.13)**
Extra Trees	**0.83 , 0.96, 0.94, 0.86, 0.83**	**0.89 (+/- 0.11)**



Relative Feature Importances

Even after adding doctor prescribing rate and drug name features, deaths are still some of the most important features.

Modeling Results – Removing All Death Data and optimizing hyperparameters

Algorithm	Cross Validation Accuracy	Accuracy Mean
Linear Regression	0.47, 0.19, 0.01 , -0.50, 0.22	0.08 (+/- 0.66)
Bagging Regressor	0.25, 0.09, 0.45, -0.07, 0.15	0.18 (+/- 0.35)
Random Forest	**0.49, 0.16, 0.26, 0.04, 0.18**	**0.23 (+/- 0.38)**
Extra Trees	0.28, 0.26, 0.38, -0.17, 0.10	0.17 (+/- 0.39)

Adding features for drug names and provider opioid rates increased the accuracy of the model, but it still performs poorly.

Limitations

Limitations of Drug Data

The drug data has significant limitations. Most notably being that the data does not represent all opioid drugs prescribed in the United States.

Medicare and Medicaid data only represent a portion of prescriptions and may not be representative of a physician's entire practice. Notably missing populations are commercially insured and self-pay patients.

While the Medicare data is more detailed, containing individual prescribers and zip codes, the medicaid data is aggregated at the state level.

Limitations of Death Data

Much of the death data obtained through CDC Wonder is suppressed. I was unable to find death data with demographics and location for the individual descendants, and was only able to use data aggregated by state.

To protect the decedent's privacy and to prevent researchers from learning individual identities, death data was suppressed in all instances where there were less than 11 decedents per category.

The ICD-10 Category codes are not mutually exclusive. Each death can be coded with up to 20 cause of death codes.

Limitations on Analysis using Medicaid Drug Data

In order to compare the drug and death datasets and to include the Medicaid drug data, the only common element to merge the data on was state.

This limited the final dataset’s number of rows to the number of states - a very small sample size.

Key Takeaways

Opioid prescribing appears to have some correlation to overdose deaths.

Certain states have a greater percentage of opioid overdose deaths, including West Virginia, DC, and Ohio.

Some prescribers write significantly more opioid prescriptions than others.

Policy Recommendations

State and/or federal level enforcement of overprescribing.

For states with crude rate deaths above average (ie: DC, OH, WV) increased monitoring and restriction of opioid prescribing; expedited license suspension for providers with death rates above average.

Require pain management training for providers with opioid prescriptions greater than 50% of total prescriptions.

Require pain management training for all prescribers writing opioid prescriptions.

Require education for patients (REMS programs) or drug testing.

Data-Driven Prevention Initiative

In 2016, Congress appropriated over $70 million to the CDC to help states improve data collection and analysis involving opioid misuse and overdose.

Further Research

It would be interesting to expand the analysis here by looking at:

- Decedent demographics
    - Race
    - Age
    - Income
- Multiple years prescription data
- Commercial and self-pay prescription data
- Opioid hospitalizations
- Narcan access and use
- National drug use survey