Analyzing the Opioid Epidemic:

Using R and SPSS Modeler to Identify Trends in Prescribing Data

By

Kyle Capuani, Adit Kulkarni, and Laharika Ravula

Abstract

This report analyzes a Kaggle dataset containing opiate prescriptions throughout the United States in 2014. The study aims to understand both the dataset and the opioid epidemic as a whole. Outside research finds that the Kaggle dataset may not be a good representation of the crisis and that better data collection and business understanding is necessary to grasp the complexity of the widespread epidemic.

Classification models, such as Artificial Neural Networks (ANN) and Decision Trees, are created in SPSS Modeler using the dataset to predict Opioid.Prescribers, which is a Boolean variable used to judge if a prescriber has prescribed opioids ten or more times during 2014. After perceiving this statistic as being a vague predictor and inspired by the exploratory data analyses and opioid-related research, the team restructures the data using R programming and models in order to better represent the uniqueness of each physician specialty and whether or not they are overprescribing when compared to their direct peers.

The report concludes that better data is required if we hope to use predictive analytics and modeling to solve the opioid epidemic.

What is the Opioid Epidemic?

Drug overdose has been the leading cause of accidental death the past number of years (Addictions and Recovery, n.d.). But how did this epidemic begin? “The opioid crisis began in the 1990s with over-prescription of powerful opioid pain relievers. They quickly became the most prescribed class of medications in the United States, exceeding antibiotics and heart medication” (Addictions and Recovery, n.d., para. 12). Opioids can come in both legal forms (such as prescription painkillers) and illegal ones (like heroin and street fentanyl) (Ornstein, 2018). According to Addictions and Recovery, “20 to 30 percent of patients who are prescribed opioids for chronic pain will misuse them … [and] about 80 percent of people who use heroin began by first misusing prescription opioids” (n.d., para. 15). In 2016 alone, there were over 64,000 deaths caused by drug overdoses and the opioid known as ‘fentanyl’ was involved in more than 20,000 of those fatalities (Addictions and Recovery, n.d.).

More recently, a “major focus of health officials and prosecutors has been to reduce the overprescribing of opioids for chronic pain” (Ornstein, 2018, para. 8). For this project, we found a dataset containing prescription records written by 25,000 licensed medical professionals during 2014. Our objective was to explore the data and build a model that could classify whether or not a prescriber is writing opioid prescriptions multiple times in a year.

U.S. Opiate Prescription Data Set

This specific data set contained three excel files: one contained a list of common opioids, another had the opioid related death totals per state and the state’s population and the third contained 25,000 different prescribers, the state they operated in, their specialty and medical credentials, a list of around 250 common opioid and non-opioid drugs and how many of each a given prescriber has prescribed, along with a Boolean variable called “Opioid.Prescriber” which “indicated whether or not that individual prescribed opiate drugs more than 10 times in the year” (Pryor, 2016, para. 4). We want to use that Boolean value to build a model that can predict if a prescriber will prescribe opioids more than 10 times in a year. This does not necessarily define a prescriber as abusing their prescription power; however, it can be seen as a red flag if we see a correlation between this Boolean variable in U.S. States with high death totals. We can also figure out which specialty is prescribing the most opiate drugs.

Since 25,000 rows might be too much for our SPSS Modeler, we decided to take a sample of 2,500 from the original dataset.

Overview Exploratory Data Analysis

The first question we had was which states had the most opioid related deaths in the dataset? We created a bar chart counting the number of deaths by state:

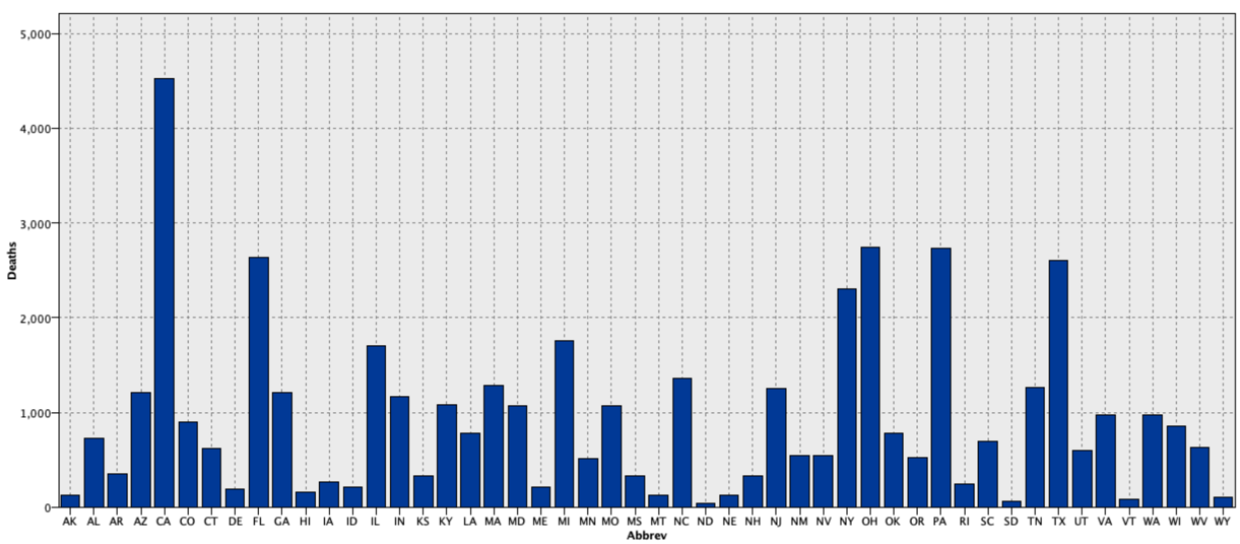


Figure 1: Opioid-related Death Totals by State

The top five states with the highest death counts were California, Ohio, Pennsylvania, Florida and Texas, respectively. While this provides a good reference point for when we will explore prescribers and their respective state of operations, we must note that some of these high death totals might be due to larger populations. For example, California and Texas are two of the highest populated states in the country. Our next question was, what would this graph look like if we took the percentage of opioid-related deaths by the state's population?

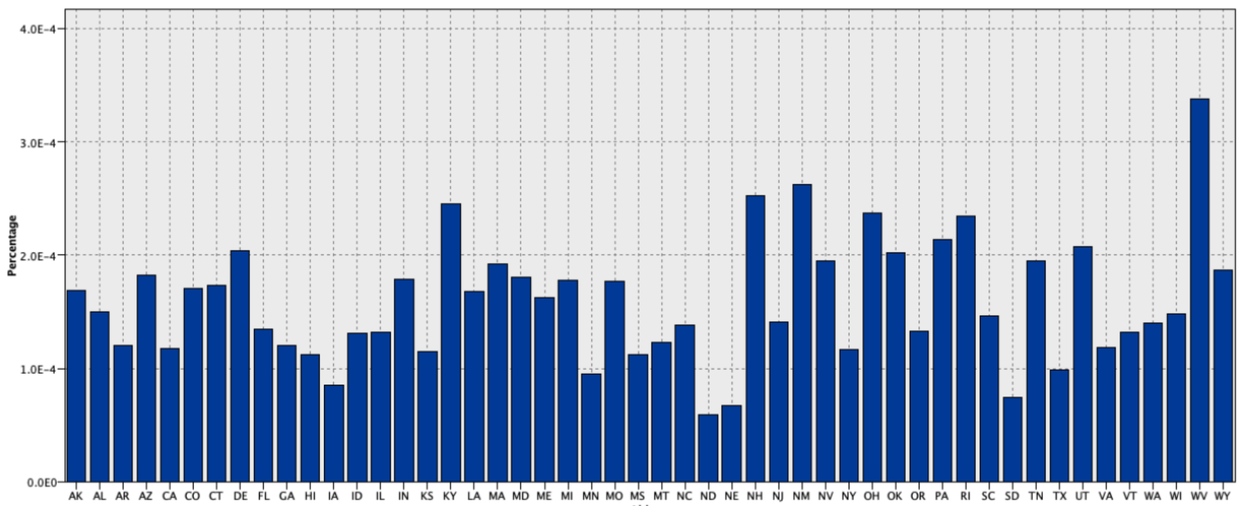


Figure 2: Percentage of Deaths by State Population

The states with the highest percentage of opioid-related deaths were West Virginia, New Mexico, New Hampshire, Kentucky and Ohio; in that order. This is the second time we noticed the state of Ohio as having high death totals related to opioids. They are number five in percentage and second in total deaths. Is there an issue of over-prescribing in this state? Or do other factors come into play?

Next, we wanted to see a scatter plot of death by population:

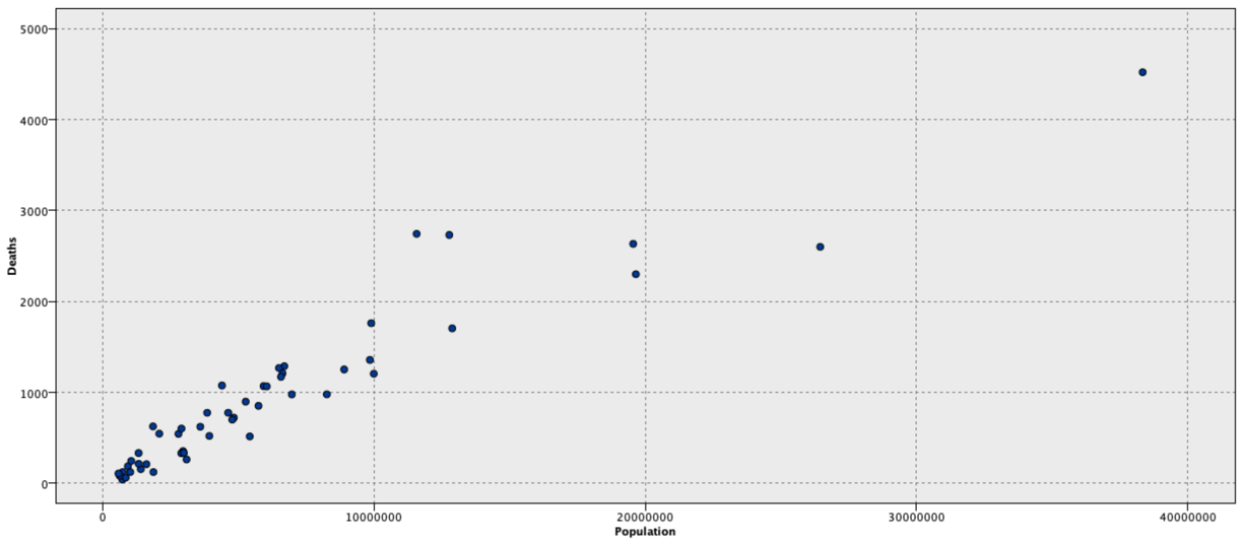


Figure 3: Deaths plotted on Population; semi-obvious linear relationship is noticed

This seems to be a linear trend. While it is obvious that a higher population has a good chance to see higher death totals, we still wanted to confirm this linear relationship. Note that there are some states with higher death totals than states with higher populations.

Now that we know which states see the most opioid-related deaths and which have the highest percentage, we moved to analyzing the Opioid.Prescriber Boolean variable:

Value	Proportion	%	Count
0.000		41.62	1040
1.000		58.38	1459

Figure 4: Distribution of Opioid.Prescribers in the entire dataset sample

This is as simple as this analysis is going to get: 58% of the prescribers in this sample have prescribed opioids ten or more times in this year. That's over 1,400 medical professionals.

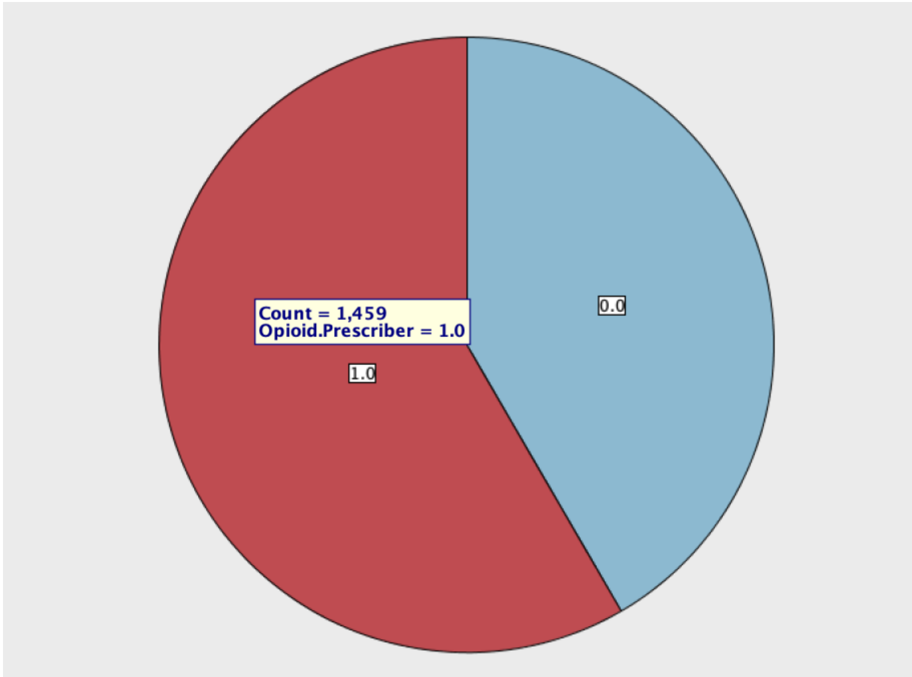


Figure 5: Pie Chart of the same relationship. A large percentage of prescribers have prescribed opioids ten or more times in this 2014 sample.

The above pie chart is a better visual showing the large percentage of prescribers who are labeled as ‘Opioid Prescribers’. This is a larger than expected percentage and it raises a number of other questions: is this figure representative of the wide-spread opioid epidemic? Is the ‘over-prescribing’ of opioids ten or more times in a year causing more opioid related deaths? And does the prescribing of opioids fuel the national crisis? We’ll explore these questions by further analyzing the data and researching a number of different sources that go beyond this specific dataset.

Next, we analyzed the Opioid.Prescriber Boolean variable in a distribution chart by State.

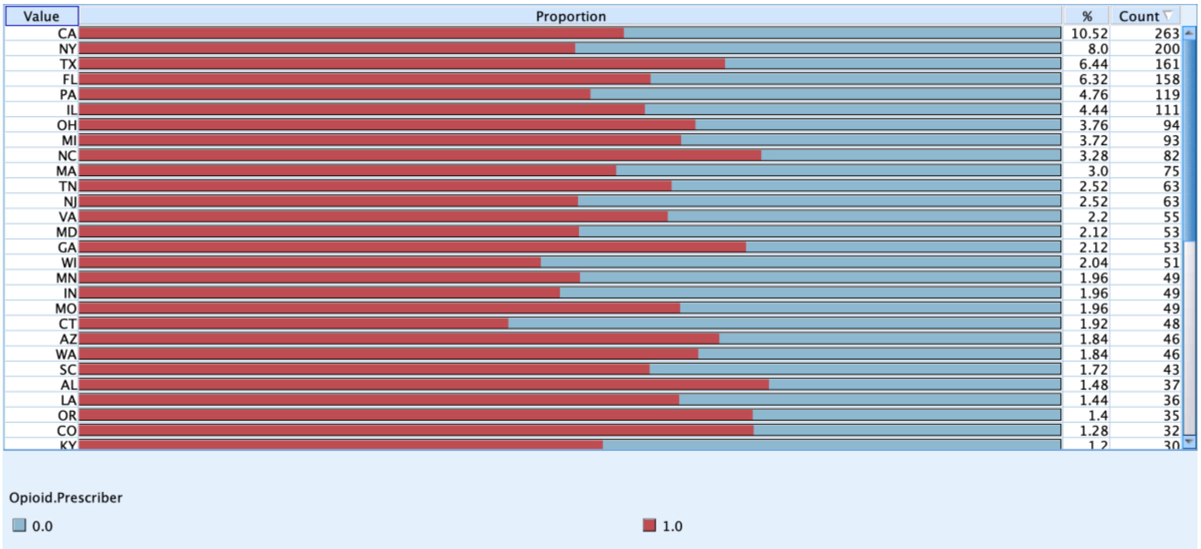


Figure 6: Opioid.Prescriber variable by State. Distribution is normalized so each bar is up to 100%. Red means '1', Blue means '0'

We normalized the distribution chart to clarify what percentage of the prescribers operating in each state were Opioid.Prescribers. The states with the most prescriber are California, New York, Texas, Florida, Pennsylvania, Illinois, Ohio, Michigan and North Carolina; these nine states account for roughly 50% of all the prescribers in the sample. Of these top states, North Carolina has the largest percentage of Opioid.Prescribers, followed by Texas; both of these states see a percentage over 60% of their total prescribers.

After those two, we see Ohio, Michigan and Florida rounding out the top five in highest Opioid.Prescriber percentages, somewhere around 50 and 60%. California, Pennsylvania, Illinois and New York all appear to be over 50% as well. From our previous analysis on opioid-related death totals, we know that Texas has the fifth highest total. California, Ohio, Florida are also present on this list and we see that all three of these states have prescribers who prescribe opioids ten or more times in a year.

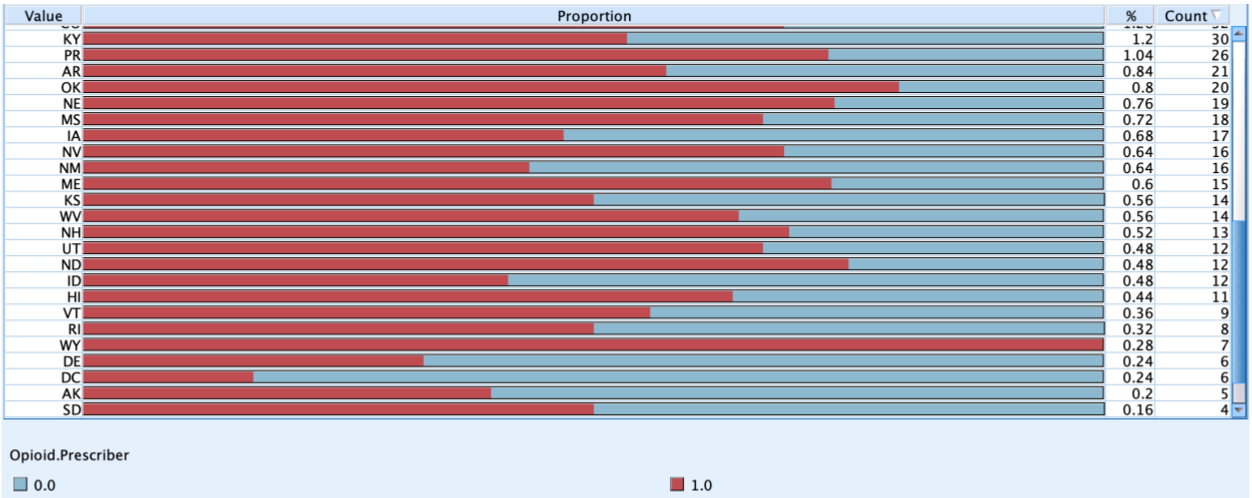


Figure 7: Opioid.Prescriber by State; further down the list

Moving further down the list, we find our 'high percentage of death' states: West Virginia, New Mexico, New Hampshire and Kentucky. West Virginia looks to have around 60% of their prescribers in this sample as Opioid.Prescribers, New Mexico appears to be under 50%, New Hampshire around 60% and Kentucky around 50%. However, we must take notice of the very small count of prescribers from these states in this sample; most have under 20 prescribers. So, while the percentage of deaths per population is still an interesting statistic, we do not have a definitive sample size to understand if the Opioid.Prescriber variable affects death percentage. We then took a look at how this Boolean variable was aligned by prescriber's specialty:

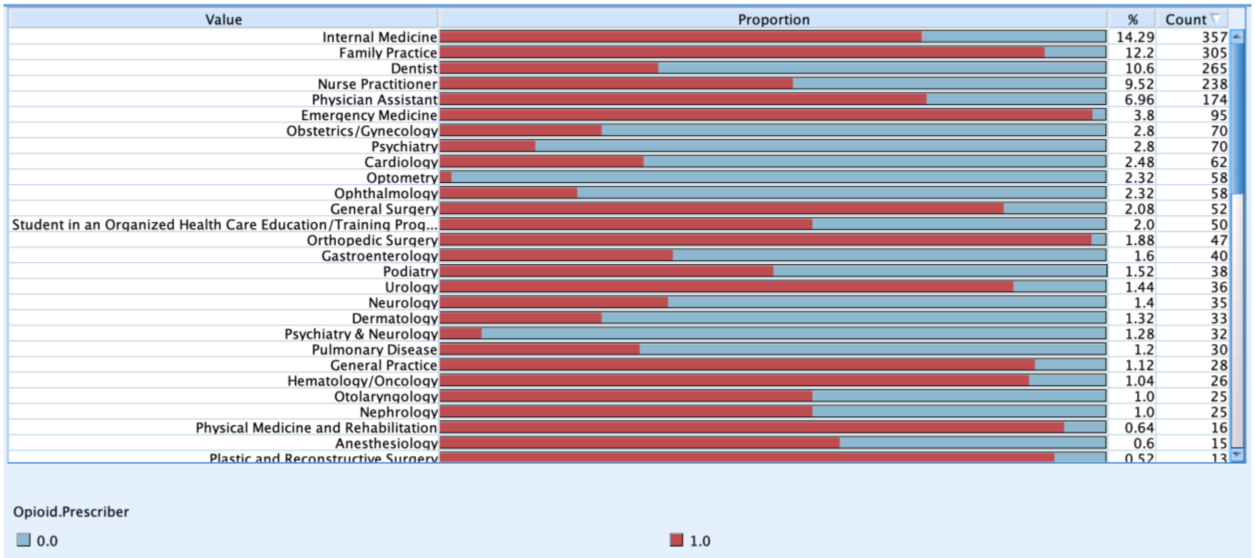


Figure 8: Opioid.Prescriber by Specialty

According to the above distribution chart, Internal Medicine is the largest specialty in this data set with 357 individuals. They prescribe opiate drugs at least ten times this year at a ~60% rate. The second largest specialty group was Family Practice at 305 individuals; they prescribed opiates 10 or more times this year at a +90% clip. This is a very large percentage.

Rounding out the rest of the top specialties: Dentist is pretty low in regards to the Opioid.Prescriber variable; Nurse Practitioner is around 50%; Physician Assistant is around 75% and Emergency Medicine is close to 99%. Specialties regarding surgery are pretty high but that can be expected given the need to prescribe pain meds post-surgery or during surgery. It’s tough to precisely comment on the rest of the specialties because they have small sample sizes. However, one of these, General Practice, with 28 individuals, looks to have an 80% Opioid.Prescriber rate.

Along with the specialties, we have a field which states the prescriber’s medical credentials.

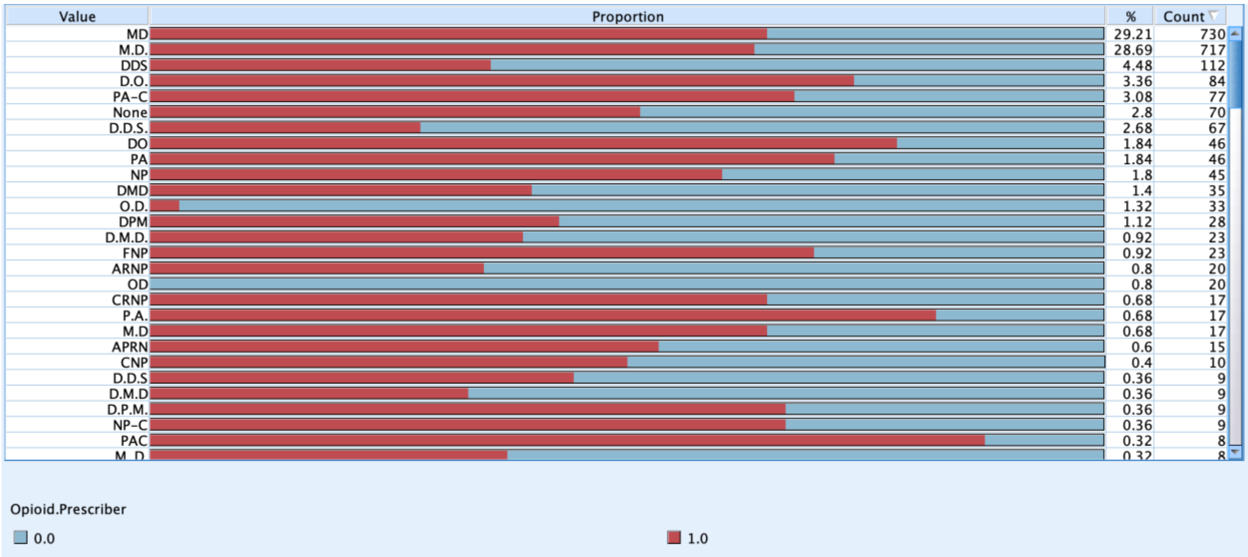


Figure 9: Opioid.Prescriber by Credentials

This field is a bit of a mess and will need to be cleaned up to be useful moving forward. There are many different credentials and some prescribers have multiple (e.g. MD and PH.D.). We sorted through the credentials with the highest counts. Atop the chart, we see MD and M.D.; we can assume that these mean the same thing. They also appear to have similar percentages of the Opioid.Prescriber Boolean variable (around 60%).

MD stands for Doctor of Medicine. “A MD works to maintain or restore human health through the study, diagnosis and treatment of disease or injury” (CentraCare Health, n.d., para. 2). A DO, the third highest credential on our list, stands for Doctor of Osteopathic Medicine; this group “belongs to a separate, but equal, branch of medicine as a MD. A DO takes in account the musculoskeletal system, physical, mental, emotional and spiritual health of a patient” (CentraCare Health, n.d., para. 3). Note that D.O. or DO in the distribution graph shows Opioid.Prescriber between 75 and 80% of the time, which is far higher than most other credentials.

A DDS or D.D.S. is Doctor of Dental Surgery and appears in around 6% of the data. It also shows a lower than 50% Opioid.Prescriber percentage, which is the lowest rate among the top credentials in the data sample.

A PA or Physician Assistant, is “a licensed health care provider who can provide a broad range of health care services that are traditionally performed by a physician, such as conduct physical exams, diagnose and treat illnesses, order and interpret tests, counsel on preventative health care, assist in surgery and write prescriptions” (CentraCare Health, n.d., para. 10). PA appears in the data set 46 times and has a ~75% rate. A PA-C is also a physician assistant in which “the ‘C’ indicates that a PA is certified by the National Commission of Certification of PAs. This certification requires 100 hours of continuing medical education every two years, along with passing a national recertification exam every six years to maintain that very important ‘C’” (Family Care Network, n.d., para. 1).

Finally, we see a ‘None’ credential appear 70 times in this sample. This means that this particular prescriber does not have noted medical credentials. They have a 50% Opioid.Prescriber rate.

As of now, we’ve found the states with the most deaths, the rate that states prescribed opioids ten or more times (our ‘overprescribed’ threshold in this dataset) in 2014 and which specialties prescribed opioids at the highest rates. Is it sufficient to claim that high opioid prescription rates by prescribers are to blame for the opioid crisis? Not necessarily. There are a number of complexities involved and we’re soon going to find out that we can’t boil the entire epidemic into one statistic or even a single dataset. The most important takeaway here is that, while we might have an understanding of our data, it might not encompass the scope of the problem. And therein lies the complexity of the opioid crisis.

Different Factors that Contribute to the Opioid Crisis

The Relationship between Pharma and Physicians

According to an article from ProPublica, there is a unique relationship between some physicians who frequently prescribe opioids to patients and the drug makers who produce them.

"Prosecutors have charged the founder of Insys Therapeutics and several of the company's sales representatives and executives for their roles in an alleged conspiracy to bribe doctors to use its fentanyl spray for unapproved uses" (Ornstein and Jones, 2018, para. 2). Purdue Pharma, the producers of OxyContin, a powerful narcotic, also are being accused of marketing the drug even though it knew that their drugs were being misused (Ornstein and Jones, 2018). As public attention increases, however, pharmaceutical company payments to physicians related to opioid drugs have decreased; "in 2016, drug makers spent $15.8 million to pay doctors for speaking, consulting, meals and travel related to opioid drugs. That was down 33 percent from $23.7 million in 2015 and is 21 percent less than the $29.9 million [it] spent in 2014" (Ornstein and Jones, 2018, para. 4).

Dr. Scott Hadland, who is an assistant professor of pediatrics at Boston University School of Medicine, has studied these trends in opioid-related marketing. Hadland reports that, according to his and his colleagues' findings, "that for every meal a physician received related to an opioid product in 2014, there was an increase in opioid claims by that doctor for Medicare patients the following year" (Ornstein and Jones, 2018, para. 10). Furthermore, a report put out from the New York State Health Foundation "found that physicians who received payments from opioid makers prescribed more opioids to Medicare patients than doctors who didn't receive the payments" (Ornstein and Jones, 2018, para. 10). Essentially, we are seeing drug maker sponsorships being handed out to physicians who, in turn, will prescribe more of that company's products. This might have a strong effect on our data set, pointing to a likelihood that many physicians were paid to prescribe more opioids in 2014. As for the dip in marketing by opioid-manufacturers in more recent years, Dr. Hadland "said it's difficult to pinpoint a single reason behind the drop, but 'it's possible that the pharmaceutical companies voluntarily reduced their marketing, realizing that they may have been contributing to overprescribing'" (Ornstein and Jones, 2018, para. 9).

While this is a rather shocking finding and might have a strong correlation with our studied dataset, it is not uncommon for physicians to be paid to market certain drugs. Doctor Brian Goldman is an emergency room physician of twenty-plus years at Mount Sinai Hospital in Toronto, Canada and a medical journalist with expertise in prescription drug abuse and chronic pain management. In his book *The Night Shift*, Dr. Goldman admits that, on many occasions in the early 1990's, he was paid by pharmaceutical companies to lecture health professionals at conferences and dinner meetings (Goldman, 2010). The majority of the time, his travel expenses, hotel accommodations and dinner tabs were comped on drug company dime (Goldman, 2010). He even appeared in educational videos on pain management and prescription drug abuse which were supported by educational grants from certain drug companies (Goldman, 2010). However, Dr. Goldman makes a point to state that "the companies that sponsored my talks had no direct input into the opinions I expressed" and he recounts a particular recording session where a drug company executive forced him to do a number of different takes, wanting him to say a certain statement promoting their drug, and he refused to do it (Goldman, 2010, pg. 85).

Dr. Goldman notes that not all physicians were comfortable with taking drug company money. Two peers of his, Joel Lexchin and Alan Cassels, an emergency physician and a drug policy researcher, respectively, wrote in the *Canadian Medical Association Journal*:

> "The people who run pharmaceutical companies don't give gifts; rather, they make investments, on which they expect a return. In the case of CME [Continuing Medical Education], the total 'gift' in the United States is in the range of U.S. $700 million annually. Gifts such as direct or indirect financial assistance to attend CME are part of the culture of reciprocity so important in physician-industry relations, and such gifts can create unconscious obligations in physicians that industry knows will be repaid in one way or another." (Goldman, 2010, pg. 87).

This relationship between pharma companies and physicians is not just opioid-related; it might be the culture of health care. While Goldman claims he did not let pharma payments influence his opinion, his experience might not be representative of physicians as a whole. As we know from the *ProPublica* article, major opioid-producing pharma companies were paying physicians to prescribe their drugs and, according to Dr. Hadland, it seemed to work.

The Insurer's Role

So far, we've learned that, in at least 2014, there was a financial incentive for physicians to prescribe opioids paid by the opioid-manufacturers. We also know, from our data, that nearly 60% of the physicians in our sample dataset prescribed opioids ten or more times that year and that death totals were highest when the Opioid.Presciber percentage was over 50% in a given state. However, there are even more factors that we must understand to explain the opioid prescribing rate.

According to another article from *ProPublica*, "at a time when the United States is in the grip of an opioid epidemic, many insurers are limiting access to pain medications that carry a lower risk of addiction or dependence, even as they provide comparatively easy access to generic opioid medications. The reason, experts say: Opioid drugs are generally cheap while safer alternatives are often more expensive" (Thomas, 2017, para. 1). When analyzing Medicare prescription plans for 35.7 million people in 2017, *ProPublica* and *The New York Times* found that only one-third of the people covered had any access to Butrans, which is a painkilling patch with buprenorphine, a less-risky opioid, while drug plans that covered lidocaine patches, which are not addictive but are more expensive, required prior approvals for patient prescriptions (Thomas, 2017). This means that it is easier for a patient, through their Medicare insurance, to be prescribed an opioid-related drug than a non-addictive drug, based on cost for the insurer. "In contrast, almost every plan covered common opioids and very few required any prior approval" (Thomas, 2017, para. 4).

One patient in the article suffered from stabbing pain in her abdomen and took Butrans to subside the pain. Her insurer, UnitedHealthcare, "stopped covering the drug, which had cost the company $342 for a four-week supply. After unsuccessfully appealing the denial, [the patient] and her doctor scrambled to find a replacement that would quiet her excruciating stomach pains. They eventually settled on long-acting morphine, a cheaper opioid that UnitedHealthcare covered with no questions asked. It cost her and her insurer a total of $29 for a month's supply" (Thomas, 2017, para. 7). "The Drug Enforcement Administration places morphine in a higher category than Butrans for risk of abuse and dependence… UnitedHealthcare, the nation's largest health insurer, places morphine on its lowest-cost drug coverage tier with no prior permission required… And it places Lyrica, a non-opioid, brand-name drug that treats nerve pain, on its most expensive tier, requiring patients to try other drugs first" (Thomas, 2017, para. 9).

For people on Medicare, this increases their likelihood of being prescribed an opioid-related drug and may further stimulate the opioid epidemic. "Dr. Thomas R. Frieden, who led the Centers for

Disease Control and Prevention under President Obama, said that insurance companies, with few exceptions, had 'not done what they need to do to address' the opioid epidemic. Right now, he noted, it is easier for most patients to get opioids than treatment for addiction" (Thomas, 2017, para. 13). Leo Beletsky, who is an associate professor of law and health sciences at Northeastern University, claims that the current insurance system is "'one of the major causes of the crisis' because doctors are given incentives to use less expensive treatments that provide fast relief" (Thomas, 2017, para. 14). This compounds on our previous analysis of drug companies financially influencing physicians and shows heavy incentives for a physician to prescribe an opioid drug over more expensive, non-addictive treatments.

In 2017, the New York state attorney general's office called out three major pharmacy companies, CVS Caremark, Express Scripts and OptumX, asking how they were planning to address the opioid crisis (Thomas, 2017). The Department of Health and Human Services is also studying whether "insurance companies make opioids more accessible than other pain treatments. An early analysis suggests that they are placing fewer restrictions on opioids than on less addictive, non-opioid medications and non-drug treatments like physical therapy, said Christopher M. Jones, a senior policy official at the department" (Thomas, 2017, para. 15).

As in the case of the patient who couldn't get her Butrans prescription filled and needed to go on a long-acting morphine to deal with chronic pain, there is a very high risk that this exposure can steamroll into life-threatening addiction; "20 percent of patients who receive an initial 10-day prescription for opioids will still be using the drugs after a year, according to a study by researchers at the University of Arkansas for Medical Sciences" (Thomas, 2017, para. 18).

Dr. G. Caleb Alexander, a co-director at the Center for Drug Safety and Effectiveness at Johns Hopkins University, states that "this has been a volume-driven epidemic, and the injuries and deaths are highly correlated with overall opioid volume on the market" (Ornstein, 2018, para. 16). With that being said, what major opioids are being prescribed, in what states are they most common and do they have a correlation with death totals based on volume?

Opioids and Other Drugs

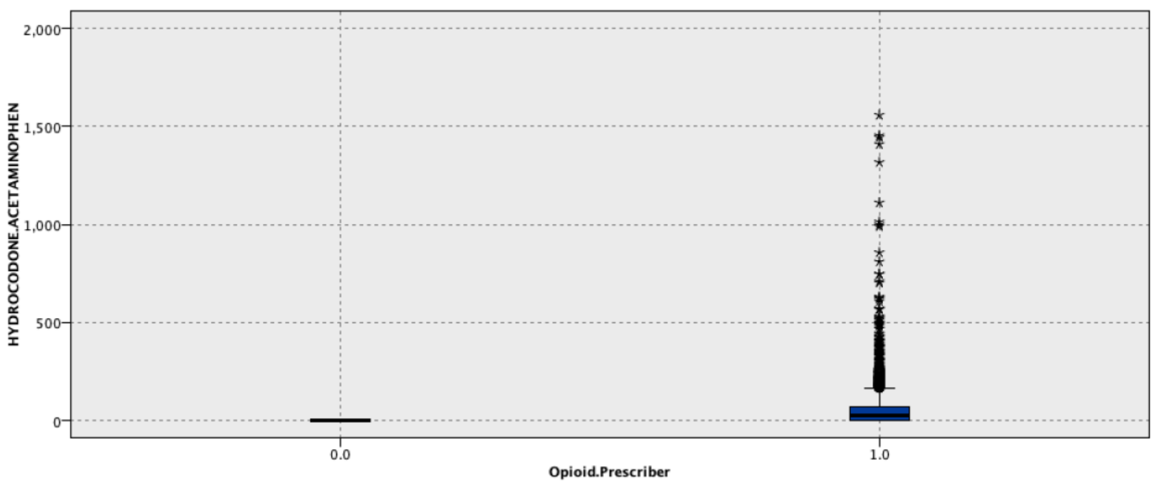
For this section, we did research on the different types of drugs being prescribed. We had to narrow down our scope since there were 250 different drugs in our dataset. We did this in a couple different ways: first, we applied a statistics node in SPSS Modeler and set it up so that we could find the correlation strength between each drug and the Opioid.Prescriber variable. We chose this because, if we are building a classification model to determine which prescriber is an Opioid.Prescriber, surely the amount of a certain drug prescribed by that prescriber will affect the model.

Since not all of the drugs in the dataset are opioids, another approach we took was to cross reference our prescriber excel sheet with our provided Opioid List file. This gave us a definitive list of which drugs were opioids and, thus, which ones we should explore deeper. Using these two methods, we narrowed down our scope and chose the following drugs to analyze:

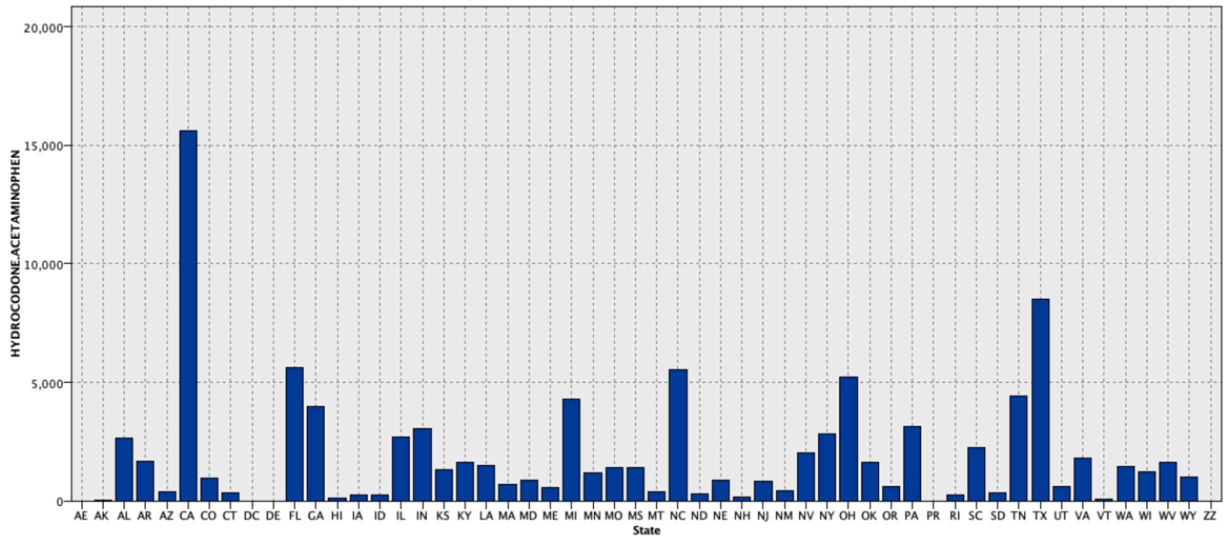
Hydrocodone Acetaminophen

Also known as Vicodin, Lortab and Norco, "hydrocodone/acetaminophen is prescribed to relieve moderate to moderately severe pain. Researchers believe that the ingredient hydrocodone relieves pain by changing how the brain and nervous system respond to painful stimuli" (Eustice, 2018, para. 4). As of 2014, "the U.S. FDA … applied restrictions to hydrocodone-containing products. The reason: to take action against the abuse of prescription pain medications and make the medications safer" (Eustice, 2018, para. 3). According to our dataset, Hydrocodone Acetaminophen has a 0.297 correlation strength with the Opioid.Prescriber variable; this is one of the highest correlation strengths from our statistics node. This drug was prescribed 100,462 times in our sample dataset of 2,500 physicians in 2014, which is the 7th highest prescription total among all of the drugs, opioid or non-opioid.

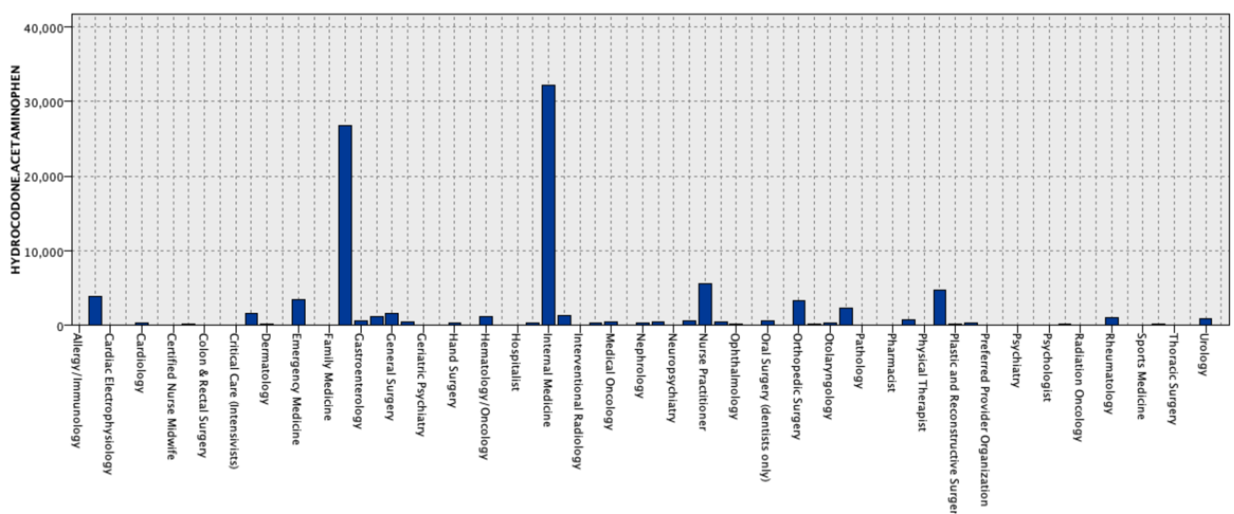
We decided to first create a boxplot to understand the variance in how each prescriber prescribed the drug, plotted on the Opioid.Prescriber variable.



Okay, this is most certainly an opioid. Nearly all physicians who prescribed this drug ended up being labeled as an Opioid.Prescriber. We see an interquartile range around 0 to maybe 25 or 50; very tough to eyeball. What we can see, though, are the very large extremes and outliers in the boxplot to the right. Some physicians prescribed Hydrocodone Acetaminophen, a known opioid, well over a thousand times.



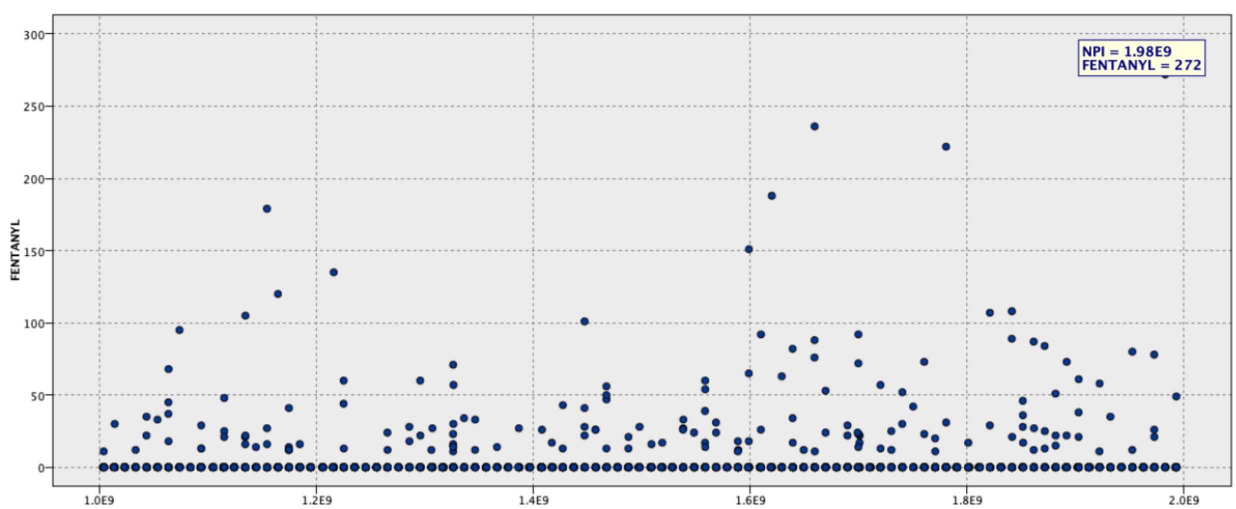
Plotting the prescriptions of this drug by state in which the physician operates in, we see that California has written over 15,000 prescriptions. That's roughly 15% of all of the prescriptions for Hydrocodone Acetaminophen in our sample dataset. Texas has written the second most prescriptions, followed by North Carolina, Florida and Ohio. Four of these states have the top five highest opioid-related death counts and North Carolina has the second highest Opioid.Prescriber rate, just behind Texas.



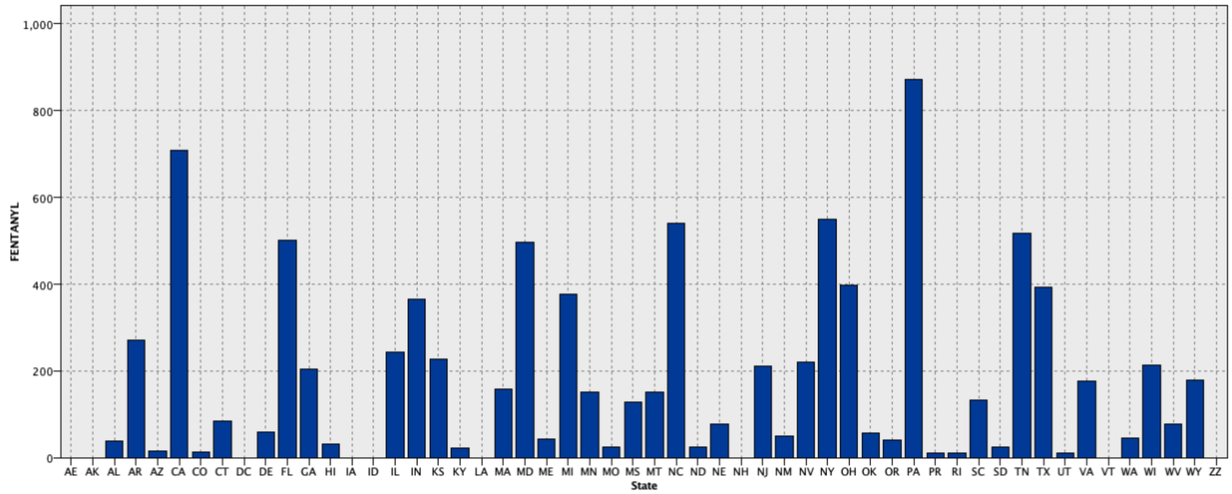
In graphing the drug prescriptions by specialty, we see that Hydrocodone Acetaminophen is most commonly prescribed by Internal Medicine and Family Medicine. This drug and its prescription volume appear to, on the surface, relate the both the Opioid.Prescriber variable and the death totals. We should see it be a major factor in our predictive model later on.

Fentanyl

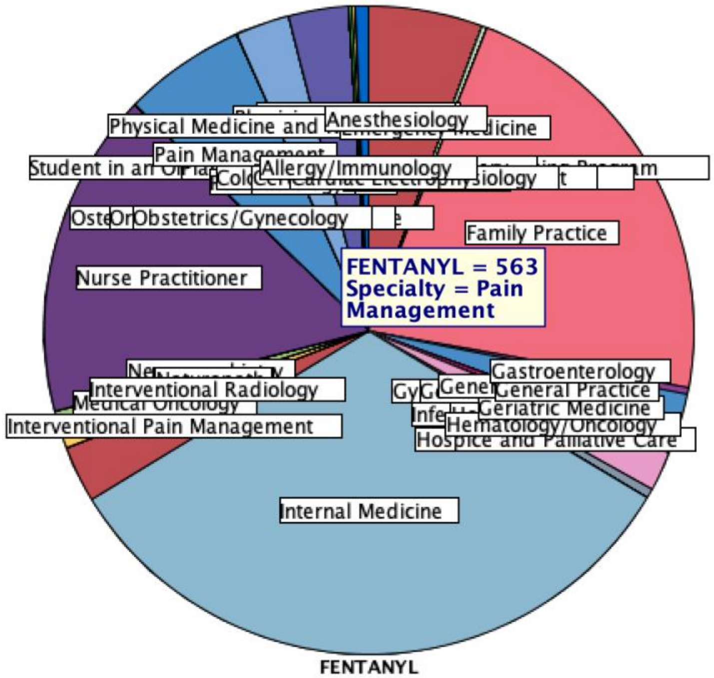
The second drug we are going to explore is Fentanyl, a well-known opioid that is most associated with the opioid epidemic. According to Addictions and Recovery, "fentanyl is so strong that it can easily cause an accidental overdose. It is 50 to 100 times more potent than morphine. About two milligrams of fentanyl – equivalent to six or seven grains of salt – is a lethal dose… In one-third of fentanyl overdoses, the individual died within seconds of taking fentanyl" (n.d., para. 18). This makes fentanyl one of the easiest opioids to misuse. Fentanyl was prescribed only 9,170 times in our sample, one of the lower totals, which may explain its lower correlation strength to Opioid.Prescriber (0.163) when compared to the other opioids listed. A 0.163 strength is pretty high though given its 97 out of 250 ranking for most prescribed drug.



This scatter plot shows Fentanyl on NPI, which stands for each prescriber ID (so a total of 2,500 in this sample size). One prescriber issued 272 prescriptions of Fentanyl in 2014, which is by far the highest total for this particular drug. There are a couple others in the 200 range as well but the majority sit near zero or below 50 prescriptions.



This is interesting. So, Fentanyl, one of the most lethal opioids, is most prescribed in Pennsylvania, California, New York, North Carolina, Tennessee and Florida, respectively. In our previous analysis of Hydrocodone Acetaminophen, or Vicodin, Pennsylvania was not very high on the prescription list. However, for Fentanyl, it is by far the most prescribed state. Pennsylvania also ranks third for the highest opioid-related death totals in the country. So, while Fentanyl is not as often prescribed, it does seem to have a strong relationship with death totals.



Looking at the prescriptions of fentanyl by specialty in a pie chart, the specialties that prescribe the most are Family Practice, Nurse Practitioner and Internal Medicine, followed by Pain Management and Anesthesiology.

Before judging each specialty for their prescribing of fentanyl, we should note the drug's uses in legal practice. According to *WebMD*, fentanyl is typically used to treat pain after surgery and as an anesthetic (Tate, 2018). While it is about 50 times more potent than many forms of heroin, it does have its use for treating patients in hospital care and "who need long-term, around-the-clock relief from severe pain" (Tate, 2018, para. 5). Its danger truly comes in its misuse and, "despite the relatively low rate of fentanyl prescriptions… illegal versions of fentanyl were largely responsible for the tripling of overdose deaths from synthetic opioids in just two years – from 3,105 in 2013 to 9,580 in 2015" (Tate, 2018, para. 10).

That’s yet another layer of complexity in the opioid epidemic and why, even though high volumes of prescribing is a factor in fueling the crisis, it is not the only cause of the high rate of addiction, overdoses and deaths. Even though fentanyl is sparingly prescribed outside of surgeries and a need-to-have basis, people can still find the opioid on the ‘street’ and this drastically increases the danger of the drug as its misuse is very likely in this situation. So, even though we can make an accurate model and label prescribers under a ‘yes or no’ variable, we probably aren’t solving the right problem, asking the right questions or even analyzing good enough data.

“Treatment for fentanyl addiction, like any opioid use disorder, includes the use of FDA-approved medications – methadone, buprenorphine, or naltrexone –prescribed and managed by a health care professional” (Tate, 2018, para. 13); we’ll explore a couple of these substitute opioids in a bit but we wanted to touch upon an issue that is typically overlooked and can’t be solved using numbers: the doctor/patient relationship.

Pain is Subjective

Remember Doctor Brian Goldman, the emergency room physician who discussed taking drug company money earlier? In his book, he also goes in depth discussing how his practice of prescribing painkillers such as OxyContin for use in pain management has changed over the course of his career.

In 1980, Dr. Goldman previously had been of the belief that doctors were “prescribing pain medication too liberally,” but, in 1990, his philosophy changed after many discussions with fellow medical professionals and a U.S.-based course that revolved around the textbook *Pain Management in Emergency Medicine* by Paul Paris and Ronald Stewart (Goldman, 2010, p. 90).

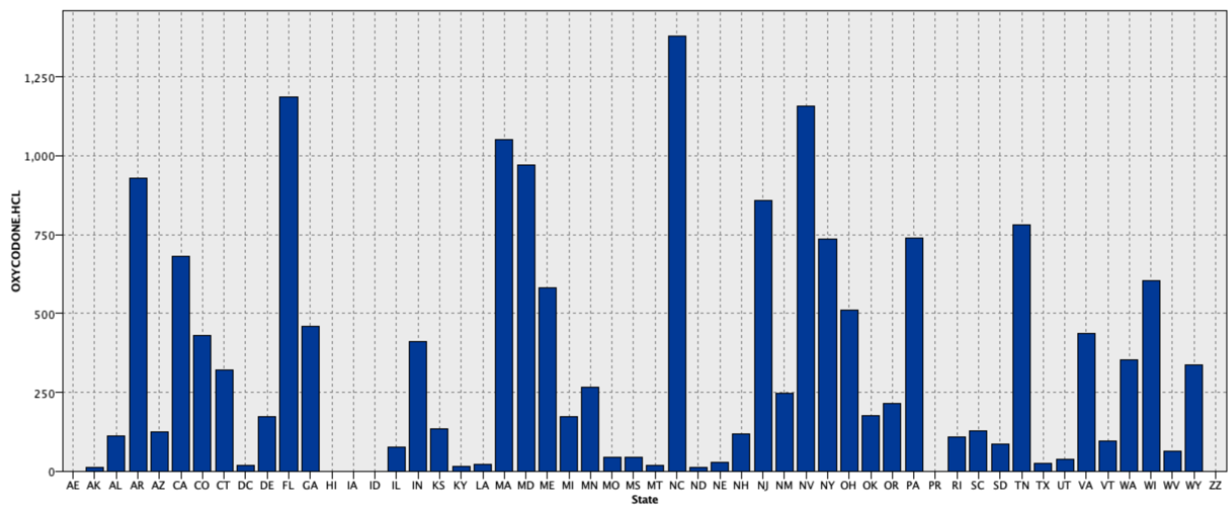
Goldman began to explore "the medical profession's inability to treat chronic pain effectively and compassionately" (Goldman, 2010, p. 91). Goldman concluded that an "underlying reason… was a kind of 'narcophobia.' [They] assumed, as a profession, that if [they] gave patients a lot of narcotics they'd become addicted" (Goldman, 2010, p. 92). Goldman's reasoning is that chronic pain is nigh impossible to verify; "pain is still a subjective complaint" (Goldman, 2010, p. 84). Painkillers might be the only thing getting patients through the day and, if they don't get the pills from their typical doctor, they will get them somewhere else, whether that be from jumping ER to ER (known as 'moonlighters') or off the streets. In Goldman's eyes, at least there would be some jurisdiction to how painkillers would be doled out.

Goldman does note that "unfortunately, [over the last number of years, this] liberalized approach to pain management has led in a growing number of instances of doctors being too willing to prescribe narcotics when they aren't necessary" (Goldman, 2010, p. 94). We can't create a predictive model that does justice to the physician/patient relationship and we certainly cannot build a model that can evaluate a patient's 'need' for pain medication. The data we have is not enough to encompass the scope of the opioid epidemic and it'd be borderline negligent to say that we have all we need to understand a subject just because our model is accurate. This is the major reason why we felt it was necessary to research the issues further, to not assume what the data was telling us was absolute and seek understanding of the factors that both compound and complex the relationships between deaths, volume of prescriptions, addiction and the health care system as a whole.

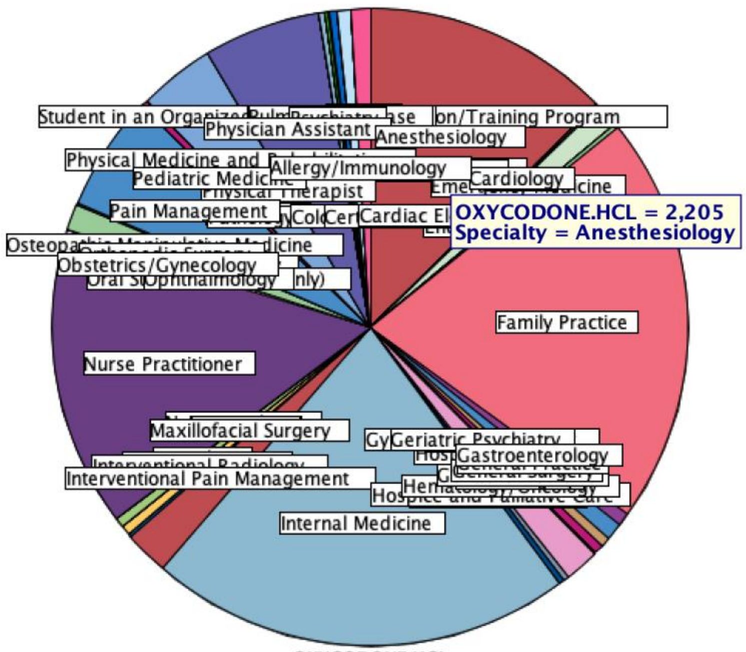
We're still going to analyze the data set to its fullest and build an accurate model but we know that this is just one aspect of the opioid epidemic. As we'll discuss in more detail, we might have stumbled upon the discovery that data miners and analysts need to start collecting data of better quality if they want to model this crisis correctly and find proper answers and sustainable solutions.

Oxycodone HCL

Also known as OxyContin, Oxycodone HCL is a highly addictive opioid used to treat moderate to severe pain. It was prescribed 17,502 times and has a correlation strength of 0.149 with the Opioid.Prescriber variable.



Oxycodone HCL was prescribed most frequently in North Carolina, one of the states with the highest Opioid.Prescriber rate. Nevada, Florida, Massachusetts, Arizona and Tennessee were also relatively high prescribed states for this drug. Interestingly, this is the first time we're seeing Massachusetts and Arizona among the top states for drug prescriptions.



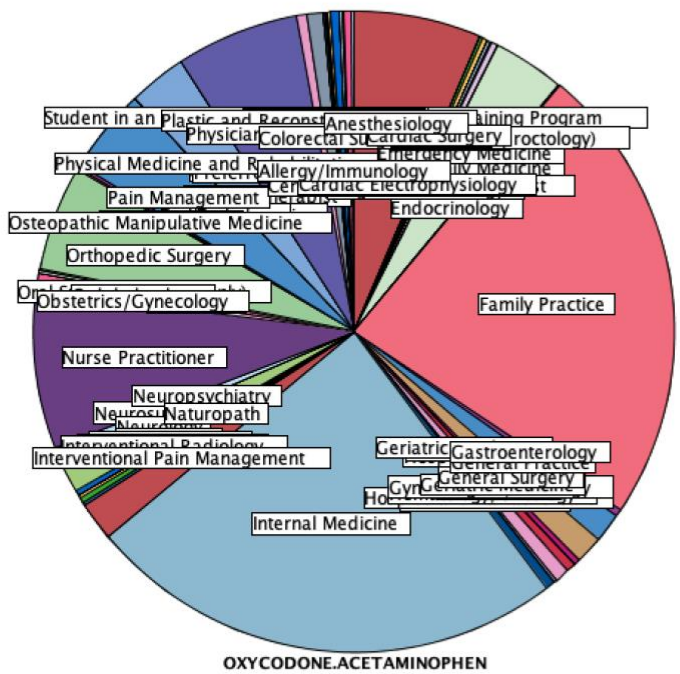
OXYCODONE.HCL

The specialties that prescribed Oxycodone HCL the most were Internal Medicine, Nurse Practitioner, Family Practice and anesthesiology. Internal Medicine, Nurse Practitioner and anesthesiology have appeared a couple times already yet only Internal Medicine has a high rate of Opioid.Prescriber (around 70%). The latter two specialties have around a 50% rate but, at least in the case of anesthesiology, may have to do with pre-surgery and post-surgery (an anesthesiologist puts patients 'under' in preparation for surgery). A Nurse Practitioner might fall under this similar reasoning and why they're prescribing a large majority of the opioids we've discussed.

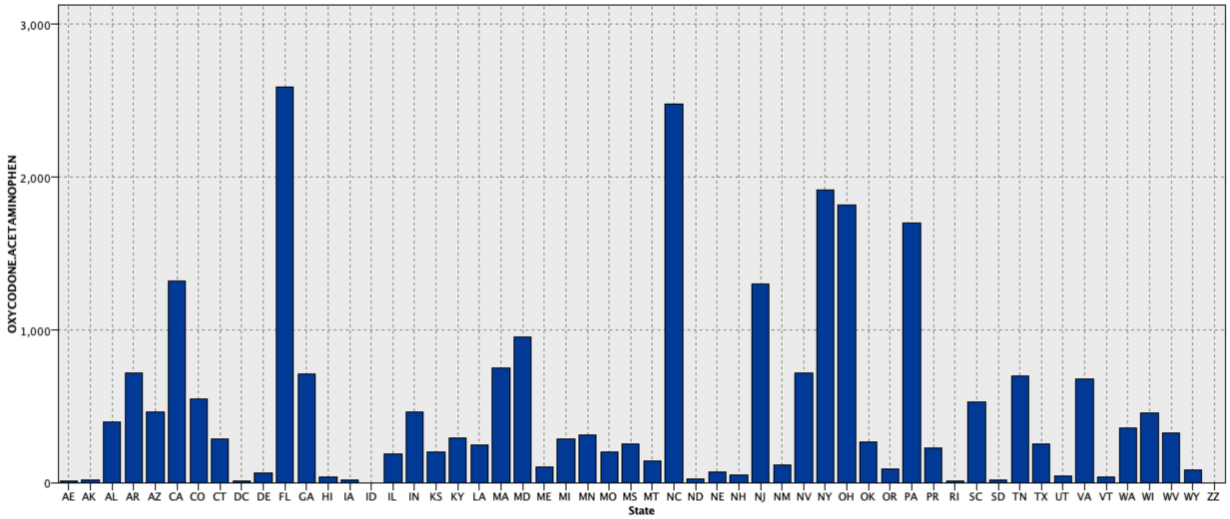
Physicians in Family Practice, however, have a 90% rate of Opioid.Prescriber which means that they are prescribing a large number of opioids. There are 305 prescribers who work in Family Practice within our sample dataset which is the most of our opioid analysis. This might be a specialty where prescriptions rates must be looked into.

Oxycodone HCL / Acetaminophen

This is a combination of two drugs we've already analyzed and is best known as Percocet. "[This] combination is used to relieve pain severe enough to require opioid treatment and when other pain medicines did not work well enough or cannot be tolerated" (Mayo Clinic, n.d., para. 1). Oxycodone HCL / Acetaminophen was prescribed 25,880 times among 2,500 physicians in 2014. Its correlation strength with the Opioid.Prescriber variable is 0.219.



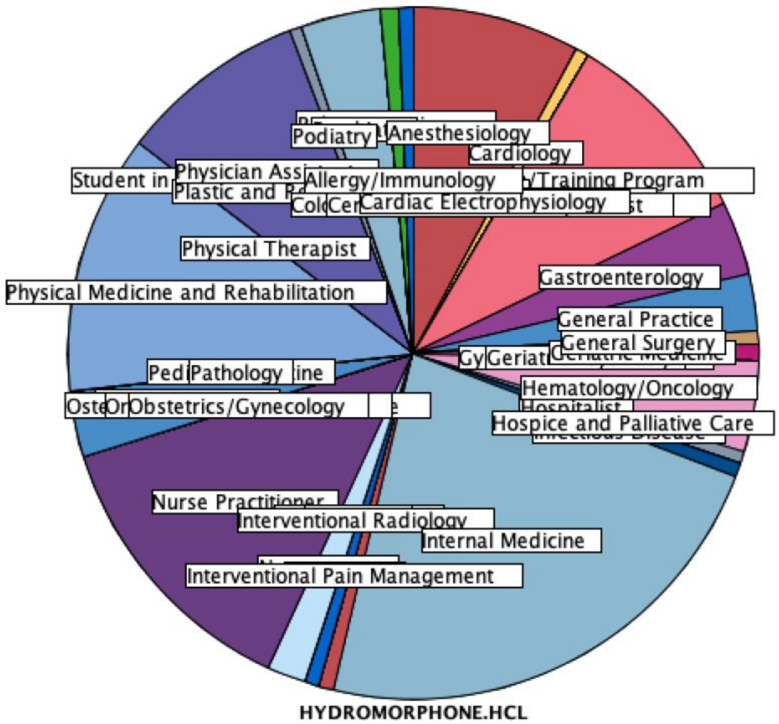
We again see Internal Medicine, Family Practice and Nurse Practitioner as the specialties that most frequently prescribe this drug. While we're not excusing Nurse Practitioner for its prescribing habits of opioids, we can envision some reasons why these totals might be high (as previously mentioned). Perhaps the same assumption can be made for Family Practice and Internal Medicine. However, Nurse Practitioner has a relatively low Opioid.Prescriber rate while the other two have relatively high percentages and are the two largest specialty groups in our sample data. If these two specialties are prescribing a lot of opioids, they will definitely be a factor in building our future predictive model and may have an influence on the overprescribing side of the opioid crisis.



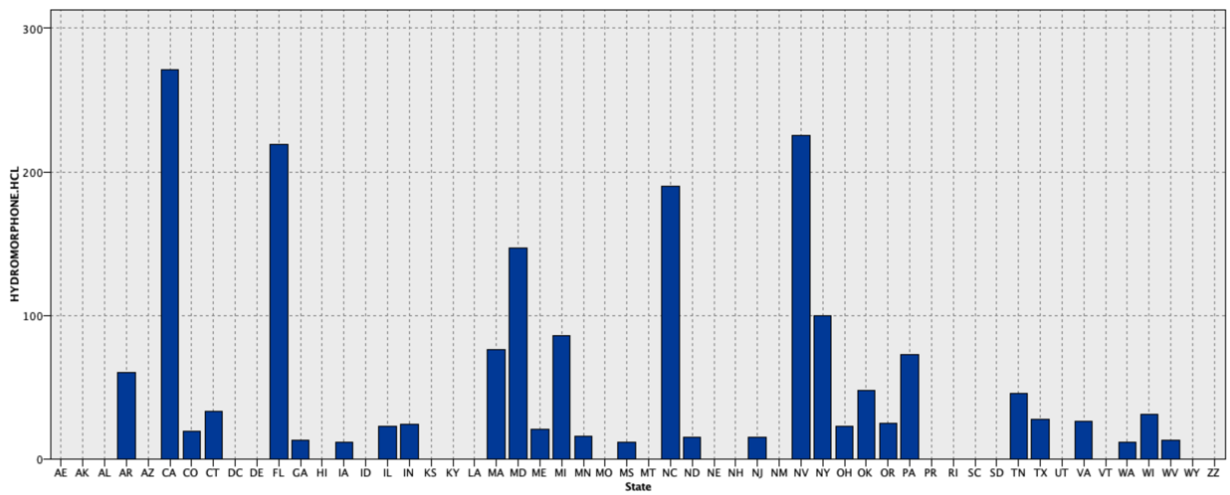
Oxycodone / Acetaminophen is most prescribed in Florida and North Carolina. These are two states that were high in death totals and had a high percentage in our Opioid.Prescriber analysis. The drug is also prescribed frequently in New York, Ohio and Pennsylvania; Ohio and Pennsylvania rank second and third in total opioid-related deaths and Ohio ranks fifth in percentage of deaths per state population.

Hydromorphone HCL

Hydromorphone is a painkiller “derived from morphine and known by the trade name Dilaudid… [it’s] a powerful opioid used to treat severe pain associated with cancer, surgery, trauma and more” (Sailor, 2018, para. 2). Hydromorphone is used more frequently than other opioids such as fentanyl for surgeries and other pain treatments because it is a little easier to manipulate and is slightly less potent (Sailor, 2018). For example, “the drug is preferred to morphine for patients with kidney problems because it doesn’t have quite the frequency of [adverse effects]” (Sailor, 2018, para. 5). Its injectable form is typically used for surgery (Sailor, 2018). It was prescribed 1,902 times within our sample data and has a correlation strength of 0.095. It is one of the least prescribed drugs in our data set, ranking 241 out of 250 in total amounts.



The drug is most prescribed by Internal Medicine and Nurse Practitioner. Also, we see the specialties of Physical Medicine and Rehabilitation become prominent for the first time in our analysis.



Even though we are looking at a small prescription size, the states which prescribed Hydromorphone HCL the most were California, Nevada, Florida and North Carolina.

Most recently, due to increased government regulations on opioid prescriptions, a nationwide shortage of the liquid form of Dilaudid has developed, forcing hospitals to adjust their procedures and use alternative drugs (Sailor, 2018). “While shortages have not affected the oral form of hydromorphone, pills cannot be easily substituted for the injectable form… [and] oral medicines are not an option for surgery patients or those who must use breathing tubes and cannot swallow pills” (Sailor, 2018, para. 7).

A spokesperson for Providence St. Peter Hospital in Olympia, Washington, one of the areas where this shortage is affecting, says that they’ve asked their “medical and nursing staffs … to reserve the use of hydromorphone for those patients who absolutely need it, and use a substitutes such as fentanyl or morphine where it is appropriate” (Sailor, 2018, para. 13). However, substituting different drugs can be risky “because different drugs have differing strengths and side effects, physicians, nurses or pharmacists must convert an order for one painkiller into the correct dosage for another. Factors including a patient’s metabolism, tolerance and history must be taken into account” (Sailor, 2018, para. 25).

Hydromorphone isn't the only painkiller in short supply around the nation; "the Federal Drug Administration maintains a list of drugs that are in shortage [and] it currently has more than 70 listed" (Sailor, 2018, para. 26).

The Opioid Shortage

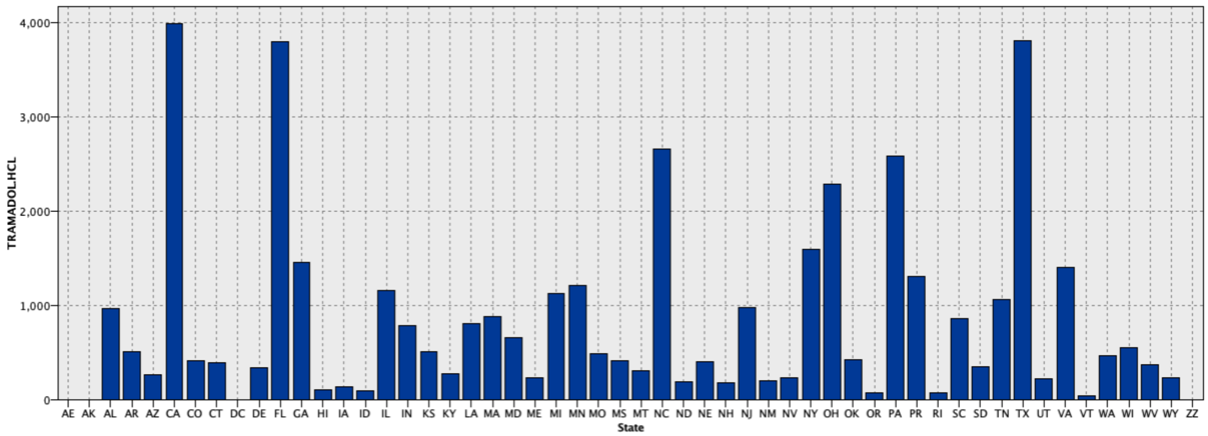
In recent years, government regulations have tried to reduce opioid addiction by restricting drug production (Bartolone, 2018). The thinking here is: less drug production will lead to less opioid prescriptions and less chance of addiction and overdose. This is also what our dataset is pointing too as well. However, as we've discussed in our research, there are other factors in play and hospitals are now facing "a dangerous shortage of the powerful painkillers for patients in acute pain, according to doctors, pharmacists and a coalition of health groups" (Bartolone, 2018, para. 1). This shortage has left many hospitals and surgical centers throughout the nation scrambling for enough injectable morphine, Dilaudid and fentanyl; all drugs necessary for surgery and treatment of severe pain (Bartolone, 2018). "The shortfall, which has intensified since last summer, was triggered by manufacturing setbacks and [the previously mentioned] government effort to reduce addiction by restricting drug production" (Bartolone, 2018, para. 2).

In response to this shortage, "hospital pharmacists are working long hours to find alternatives, forcing nurses to administer second-choice drugs or deliver standard drugs differently. That raises the risk of mistakes – and already has led to at least a few instances in which patients received potentially harmful doses, according to the nonprofit Institute for Safe Medication Practices, which works with health care providers to promote patient safety" (Bartolone, 2018, para. 3). In states such as California, Illinois and Alabama, "some patients are receiving less potent medications like acetaminophen or muscles relaxants as hospitals direct their scant supplies to higher-priority cases… The American Society of Anesthesiologists confirmed that some elective surgeries, which can include gall bladder removal and hernia repair, have been postponed" (Bartolone, 2018, para. 5). A coalition of professional medical groups including the American Hospital Association, the American Society of Clinical Oncology and the American Society of Health-System Pharmacists informed the U.S. Drug Enforcement Administration that the shortages are increasing the risk of medical errors and are potentially life threatening (Bartolone, 2018).

Restricting the production of opioids is creating shortages in areas where the painkillers are used legally and usefully. Simply cutting down on drug production is not the answer here yet might be the solution we've come up with based on analyzing data that didn't encompass the entirety of the epidemic.

Tramadol HCL

Also known as Ultram, Tramadol HCL is "a synthetic opioid analgesic medication used to treat moderate to moderately severe pain, such as that caused by osteoarthritis. [Other] brand names include ConZip, FusePaq Synapryn, Rybix, [and] Ryzolt. Tramadol binds to opioid receptors, decreasing the body's ability to feel pain. It is similar to morphine in that way it works but is about one-tenth as potent" (Eustice, 2018, para. 1). In August of 2014, "the U.S. Drug Enforcement Administration listed tramadol as a schedule IV controlled substance… due to the risk of addiction and overdose. Examples of other schedule IV drugs include Valium [diazepam], Xanax [alprazolam], and Ambien [zolpidem]" (Eustice, 2018, para. 2). Tramadol HCL was prescribed 43,883 times in our sample dataset and, at a correlation strength of 0.326, is the highest among all drugs in its correlation to the Opioid.Prescriber variable.

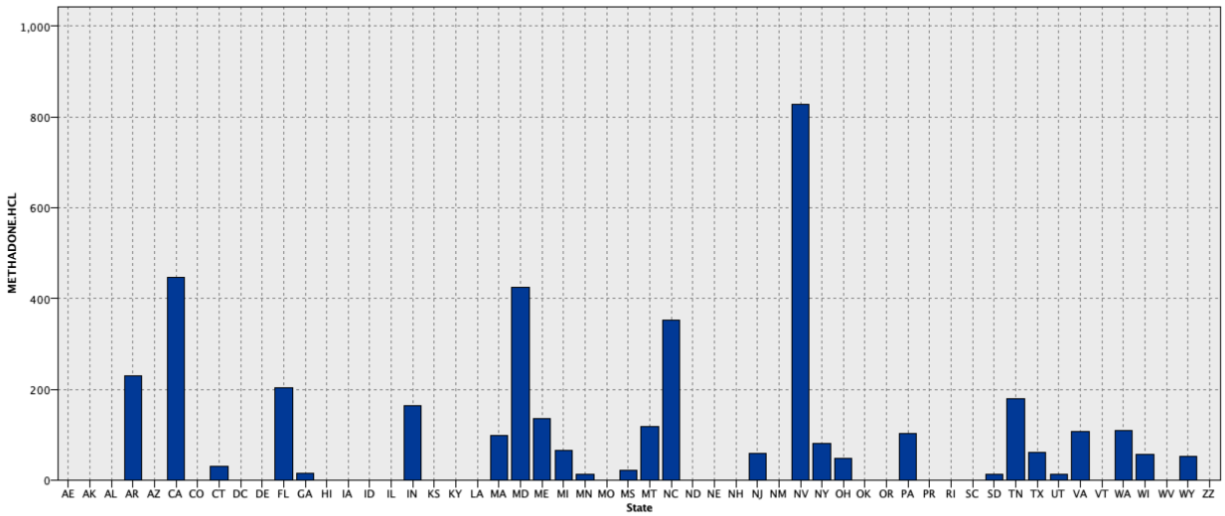


Tramadol HCL is most prescribed in California, Florida, Texas, North Carolina, Pennsylvania and Ohio.

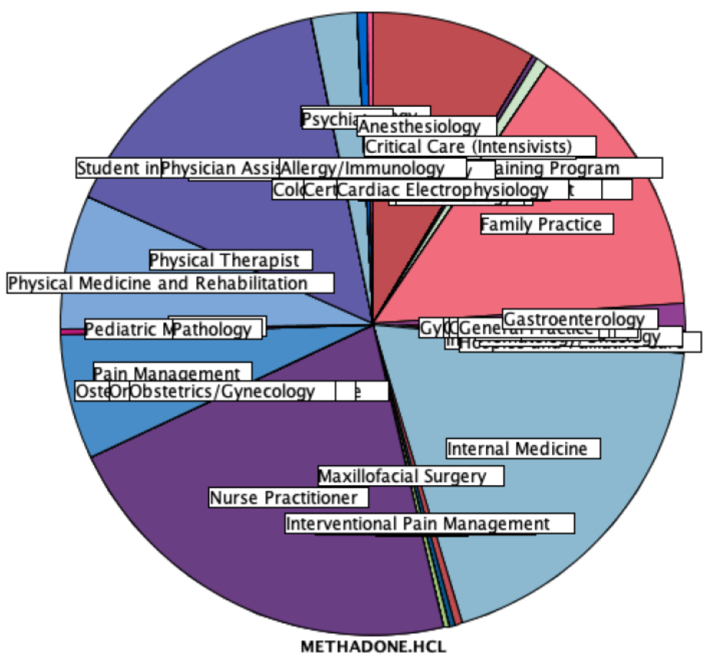
Opioids to Fight Addiction: Methadone HCL

Kind of an oxymoron, huh? Not exactly. In order to treat addiction, some physicians will prescribe less addictive opioids to replace more serious opioid addictions. One particular opioid, which was mentioned earlier in regards to getting patients off fentanyl addiction, is Methadone HCL. This drug "is used to treat addiction to opioids [such as heroin] as part of an approved treatment program… It helps prevent withdrawal symptoms caused by stopping other opioids" (Cunha, n.d., para. 1). Its brand names include Methadose and Dolophine. However, this medication is not supposed to be used to relieve mild and temporary pain; it is not for occasional use and can still be misused unless under proper medical supervision (WebMD, n.d.).

Within our sample dataset, it has been prescribed only 4,023 times and has a correlation strength of only 0.081 to the Opioid.Prescriber variable.



This drug is most prescribed in Nevada, Maryland, California and North Carolina…

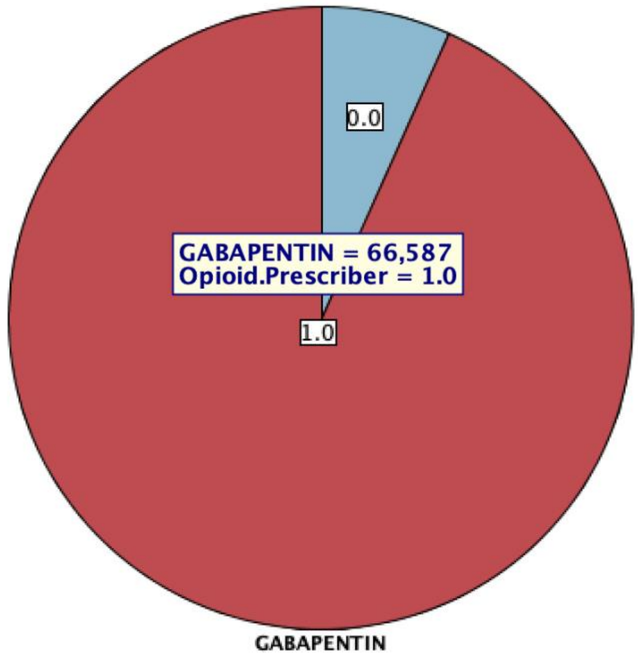


And the specialties that prescribe it most are Nurse Practitioner, internal medicine, physician assistant, family practice, pain management and physical medicine/rehabilitation. While we recognize some of these specialties from earlier opioid analyses, a couple new ones appear in areas that are more likely to deal with patients recovering from addiction or looking for non-addictive treatments.

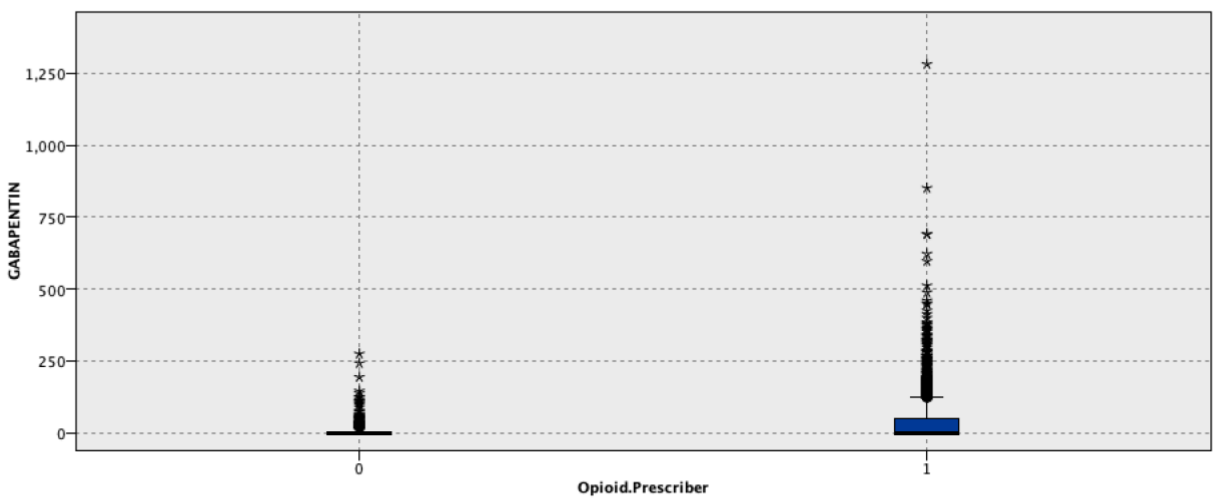
Gabapentin

Gabapentin is not an opioid yet it has a correlation strength of 0.272 to the Opioid.Prescriber variable, which is pretty high. Within this sample data, it has been prescribed 71,363 times. “Gabapentin is used with other medications to prevent and control seizures. It is also used to relieve nerve pain following shingles” (WebMD, n.d., para. 1). Also known as Neurontin, Gralise and Horizan, Gabapentin “has recently been linked with a significant percentage of drug-related deaths in Louisville, Kentucky… present in nearly one-fourth of fatal overdoses within the commonwealth’s largest city” (Health, 2018, para. 1).

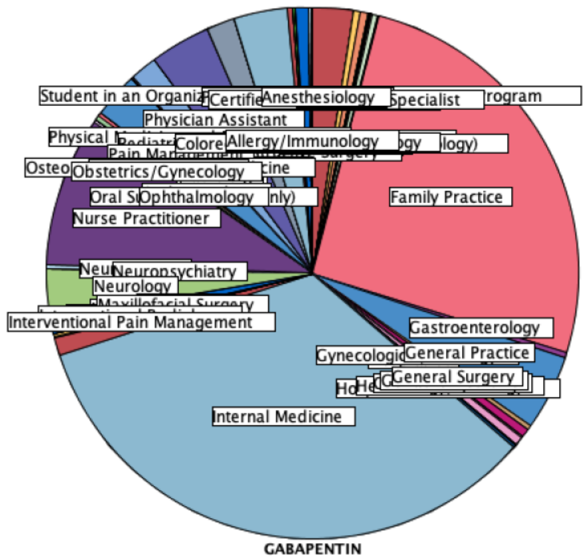
Doctors have begun prescribing Gabapentin and other opioid-free alternatives to patients with chronic pain in order to curb the number of opioids being prescribed (Health, 2018). “Since 2012, the number of gabapentin prescriptions dispensed has increased by 64 percent, making it the 10th most commonly prescribed medication in the nation” (Health, 2018, para. 2). The hope is that Gabapentin can help people with opioid withdrawals and reduce their symptoms. However, “research shows that it is likely to be misused by people with a history of opioid misuse or opioid addiction and heroin users have described it as ‘easy to obtain’. Furthermore, when used in combination with heroin, gabapentin may be especially deadly by increasing the user’s risk of experiencing an overdose” (Health, 2018, para. 4). Gabapentin has potential to be an asset in quelling the opioid epidemic but its misuse is leading to more addiction and death (Health, 2018).



Gabapentin, while not an opioid, appeared in the records of Opioid.Prescribers 66,587 times. On the other side, this drug was prescribed 4,776 times by prescribers who are *not* considered Opioid.Prescribers.



Looking at a boxplot of this drug and the Opioid.Prescriber variable, we notice that the interquartile range on the right is larger than the one to the left (Opioid.Prescriber versus Non-Opioid.Prescriber, respectively). We also see some high upper outliers on the right side. So, what does this mean? This might point to prescribers who tend to dole out opioids also prescribing Gabapentin, a drug that is being used to decrease opioid dependency. This is just an assumption but this might explain why we see Gabapentin prescriptions frequently in the records of physicians with high opioid prescription totals.



Very similar to all of the other opioids, Internal Medicine and Family Practice are, by far, the two specialties that prescribe Gabapentin the most. Nurse Practitioner also has a sizable claim in the pie chart.

To conclude our exploratory data analysis and research: overall, there does appear to be a relationship between states that prescribe opioid and other abusable opioid substitutes more frequently and states that have the highest opioid-related deaths. However, the research we conducted throughout our data analyses was meant to show the complexities in understanding what causes opioid addiction and abuse. This epidemic is not something that can be solved using a single statistic or perspective and it is dangerous to assume so.

In the Kaggle file where we found this data, the subtitle reads “can you save lives through predictive modeling?” And that might be the wrong question to be asking, given the dataset provided. Maybe it should be “is our data representative of the problem?”

While having the volume of prescriptions that each prescriber has written in a given year is great and useful data, applying the Opioid.Prescriber Boolean variable for if opioids are prescribed ten or more times assumes that all specialties are equal and that ten is some magic threshold that will lead to drug abuse by that given prescriber’s patients. It’s a dangerous assumption to model and enact regulations upon (as we’ve seen with the hospital shortages recently) and shows that we really might not be collecting enough data if we truly want to understand this crisis.

Moving forward to the modeling stage, we’re going to create a model to predict the Opioid.Prescriber Boolean. We’re also going to see if we can create a better understanding of the prescription data through restructuring the dataset and predicting how many opioids each specialty should be prescribing compared to their peers in order to create a better standard per specialty type.

Modeling

For this stage, we wanted to create a couple different classification models: an Artificial Neural Network (ANN) model, a C5.0 Decision Tree and a revision of the Opioid.Prescriber variable in which the amount of opioid prescriptions are judged within their own unique specialty (i.e. a Nurse Practitioner will not be judged by the same standard as a Family Practice physician).

Before all this, though, we needed to clean our dataset and normalize it for the ANN. We performed this cleaning using a combination of R and Excel.

All of this work can be seen in the appendix and the attached Excel files but essentially, we normalized all of the continuous variables (i.e. all of the drug prescriptions) so that they were all within 0 and 1; we did this transformation in Excel.

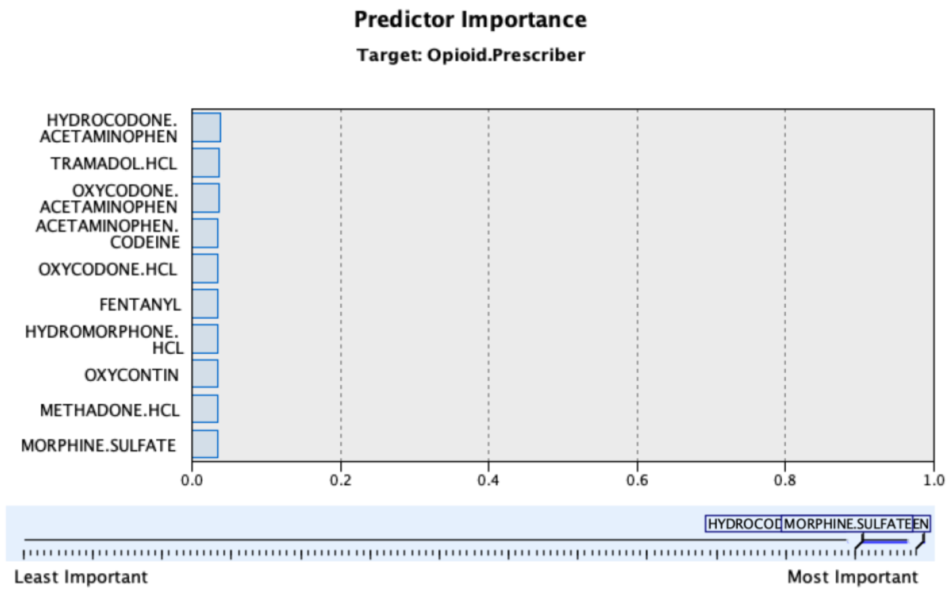
Then, we took this datasheet and imported it into R so that we could create dummy variables for all of the categorical variables (State, Gender, Specialty). We decided to drop the Credentials variable because it had many variations and a lot of overlap. For instance, there were many occasions where a single MD had multiple credentials, which caused them to warrant their own dummy variable. We assumed that this would cause an immediate overfit in the model and that Specialty would be a better input predictor anyway.

Finally, we wrote an excel file using these new transformations and imported it into SPSS Modeler. We partitioned the data into 70% training and 30% testing and ran the ANN node using all prescription and dummy variables. The training accuracy ended up at 85.19% and testing was at 81.95%. While this is a good start, we immediately see the model overfitting the data because of the many inputs.

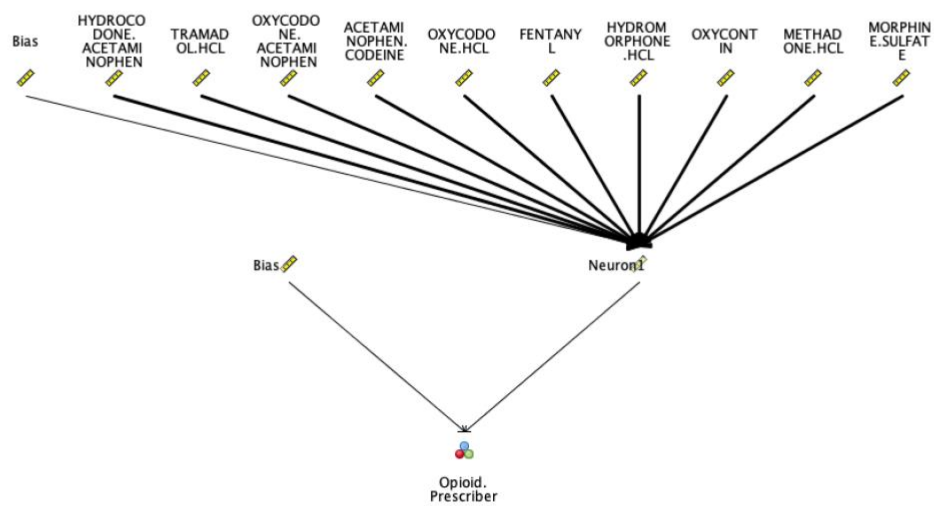
On our next attempt, we utilized the opiods_list excel file, which contained the names of all of the opioids in the data set, and removed all of the inputs that were not considered to be opioids (sans Gabopentin, since we saw a relationship between Opioid.Prescriber and this non-opioid in our EDA).

Opioids ANN Model

This model proved to be very accurate, sporting an 89.24% training accuracy and an 88.97% testing accuracy; a very close fit.



We see that Hydrocodone Acetaminophen shows a strong predictor importance. However, it appears that many of the opioids have a relatively strong importance.



This model ended up having only one neuron in its hidden layer. In our confusion matrix:

Testing	0	1
0	307	27
1	61	402

Recall = TP / (TP + FN) = 402 / (402 + 61) = 86.8%
Specificity = TN / (TN + FP) = 307 / (307 + 27) = 1 - .91916 = 8%
Precision = TP / (TP + FP) = 402 / (402 + 27) = 93.7%

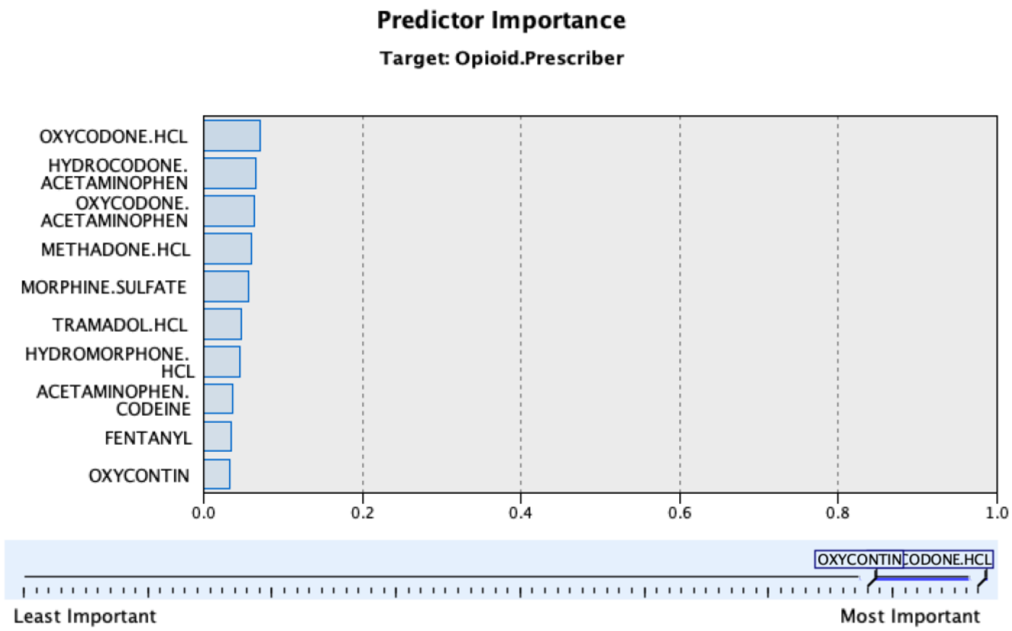
This ANN captured 86.8% of all Opioid.Prescribers in the data. The model also classified a non-Opioid.Prescriber as an Opioid.Prescriber only 8% of the time, which is a good sign which we'll explain in a moment. Finally, the ANN predicted an Opioid.Prescriber correctly 93.7%, which is great accuracy.

We like this model because it not only predicted Opioid.Prescribers very accurately, it also didn't label many non-Opioid.Prescribers as Opioid.Prescribers, which can be a damaging prediction to a physician's career if it proved to be incorrect; it's important to add context to models such as these.

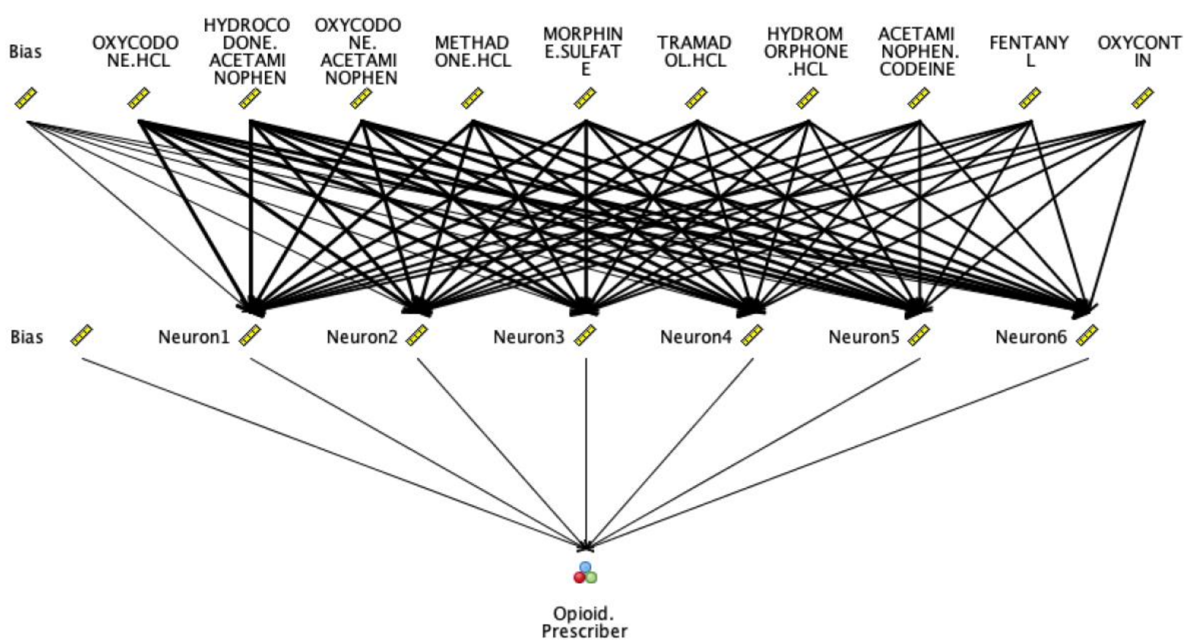
Since we noticed that none of the State dummy variables appeared in the predictor of importance chart in our latest model, we tried to make another ANN sans the State variables. So, basically, just the opioids, the specialty dummy variables and the male/female dummy variables (note: there was a version without the Gender dummy variables and accuracy went down a lot; for this reason, they're staying in).

ANN without States

This model turned out to be more accurate than our previous one: training had a 90.77% accuracy and testing had a 92.11%. This might be a slight underfit but we'll check out the confusion matrix in a moment.



We see all opioids atop the predictor of importance again albeit with the top two or three switching places. However, we do notice some regression of importance as we go down the list, which did not occur in our previous model.



Our model ended up having six neurons in the hidden layer, which is a massive leap from our last model.

Testing	0	1
0	322	12
1	51	413

Recall = TP / (TP + FN) = 413 / (413 + 51) = 89%

Specificity = TN / (TN + FP) = 322 / (322 + 12) = 1 - .964 = 3.5%

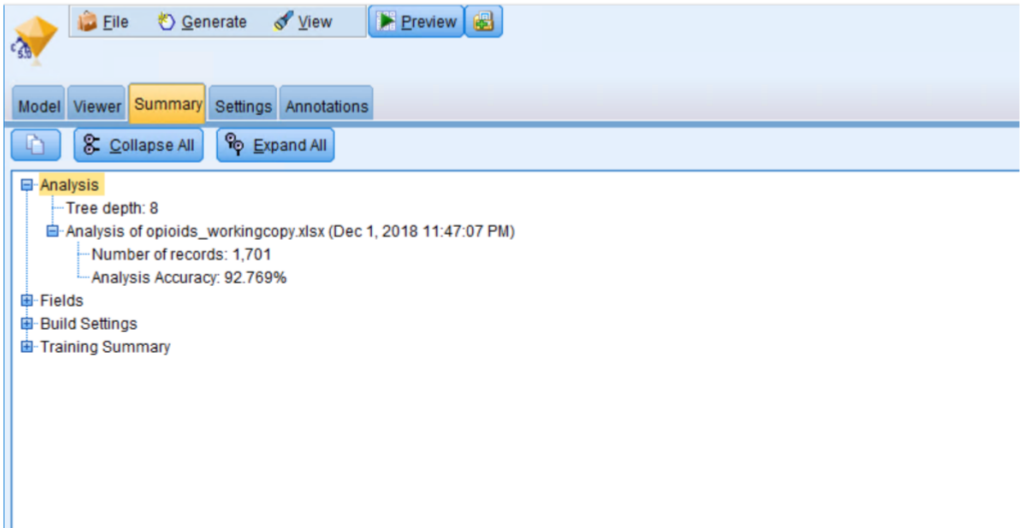
Precision = TP / (TP + FP) = 413 / (413 + 12) = 97.18%

This new ANN, without any State dummy variables, captured 89% of all Opioid.Prescribers in the dataset. Furthermore, Non-Opioid.Prescribers were classified as Opioid.Prescribers only 3.5% rate, which is a large decrease from the previous model. Also, the model predicted an ‘Opioid.Prescriber’ with 97.18% accuracy.

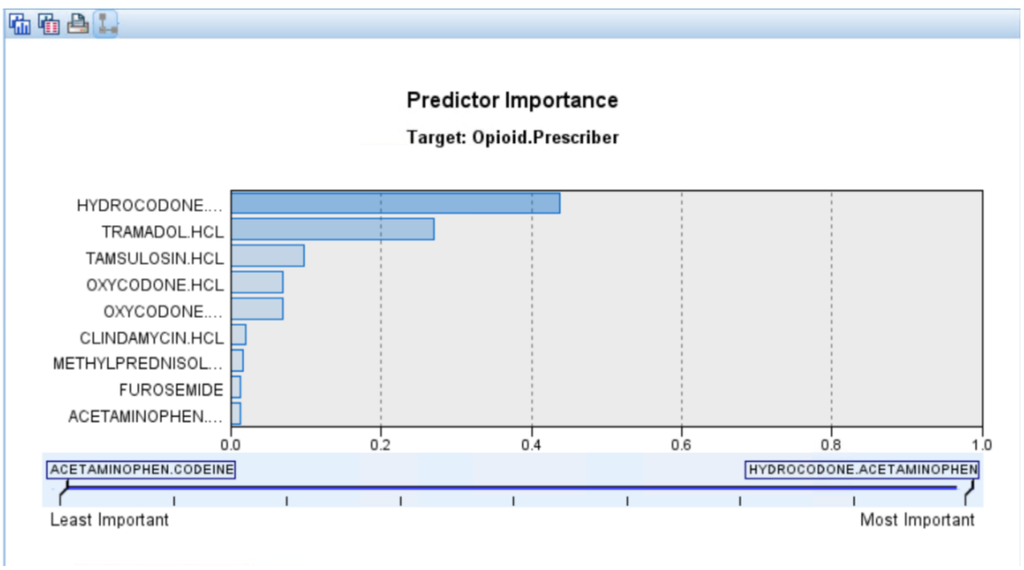
This turned out to be a far more accurate model but there is a tad worry about underfitting the data. However, we’re seeing just a 2% difference between the training and testing sets for this new model and, if we liked our last model because it didn’t label non-Opioid.Prescribers as Opioid.Prescribers, this model dropped that particular percentage by around 5% from its already low rate. So, in regards to our ANN modeling, this sans-States one is probably the most accurate option.

C5.0 Decision Tree

Our next model was a C5.0 Decision Tree in which we tried to predict the target variable of Opioid.Prescriber. This was a very similar test to the ANN albeit for using a non-normalized dataset; we didn’t need any dummy variables or standardized continuous variables.



Using all variables as inputs, our decision tree had a tree depth of 8 and a training accuracy of around 92.77%.



We notice similar predictors to our ANN model in this decision tree, with Hydrocodone Acetaminophen carrying the most significance.

You can see the decision tree above. It appears that the first branch occurs when Hydrocodone Acetaminophen prescriptions are over 5.5; this specific branch splits out 1,022 Opioid.Prescribers. From that point down, it appears that all of the variables will see splits at 5.5 prescriptions. The one interesting caveat is that Tamsulosin HCL, a non-opioid, will lead to an Opioid.Prescriber rating 69% of the time at 14.5 prescriptions or above.

Testing	0	1
0	331	3
1	41	423

Accuracy = (TP+TN) / (TP+TN+FP+FN) = (331+423) / (331+3+423+41) = 94.96%

The recall (Sensitivity) can be given as:

Recall = TP / (TP+FN) = (423) / (423+41) = 91.16%

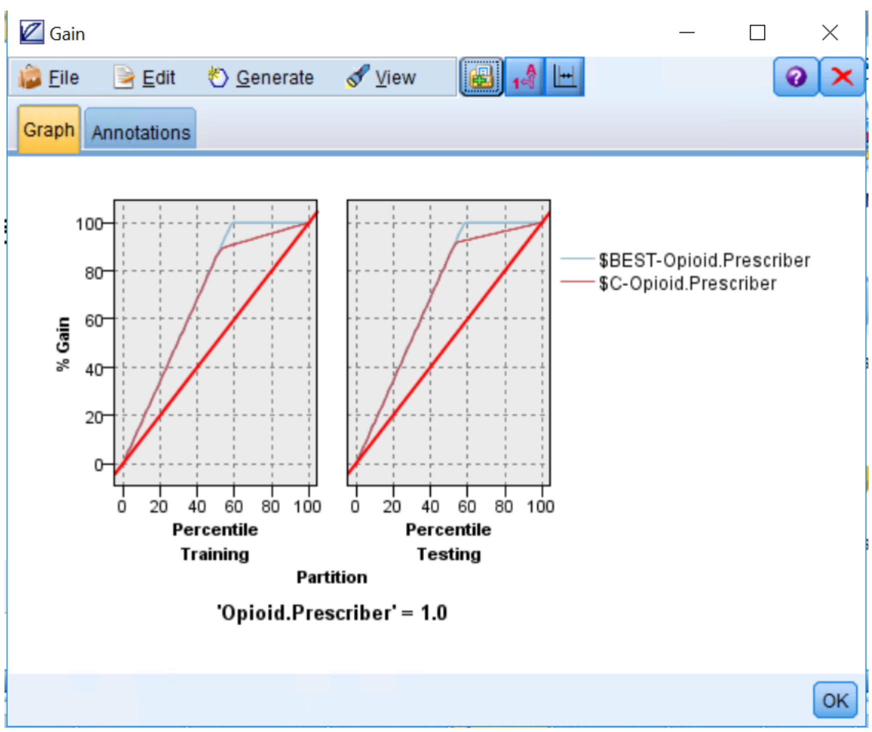
Precision can be given as:

Precision = TP / (TP+FP) = (423) / (423 + 3) = 99.29%

Specificity can be given as:

Specificity = TN / (FP+TN) = (331) / (331+3) = 99.7% = 0.3%

The model's precision, specificity and recall are better results than our ANN model. This seems to be a very accurate model.



Looking at our Gain chart, as we've gone through around 20% of the data, the at-chance line is showing that we should have correctly identified 20% of Opioid.Prescribers. However, the perfect prediction line and our model line is saying that we have correctly identified roughly 40% at this point. Once we've gone through roughly 90% of the data, we'll have correctly predicted all Opioid.Prescribers.

Opioid.Prescriber Revision

As mentioned at the onset of the paper, the Opioid.Prescriber variable labels a '1' if a prescriber has prescribed opioids ten or more times in a year. Under that assumption, the Boolean variable is arguing that 'ten' is the breaking point that will lead to overprescribing. As we've argued throughout the paper, this is a dangerous assumption to make and doesn't encompass the entirety of the situation. For this reason, we sought to recreate this Opioid.Prescriber variable with more context.

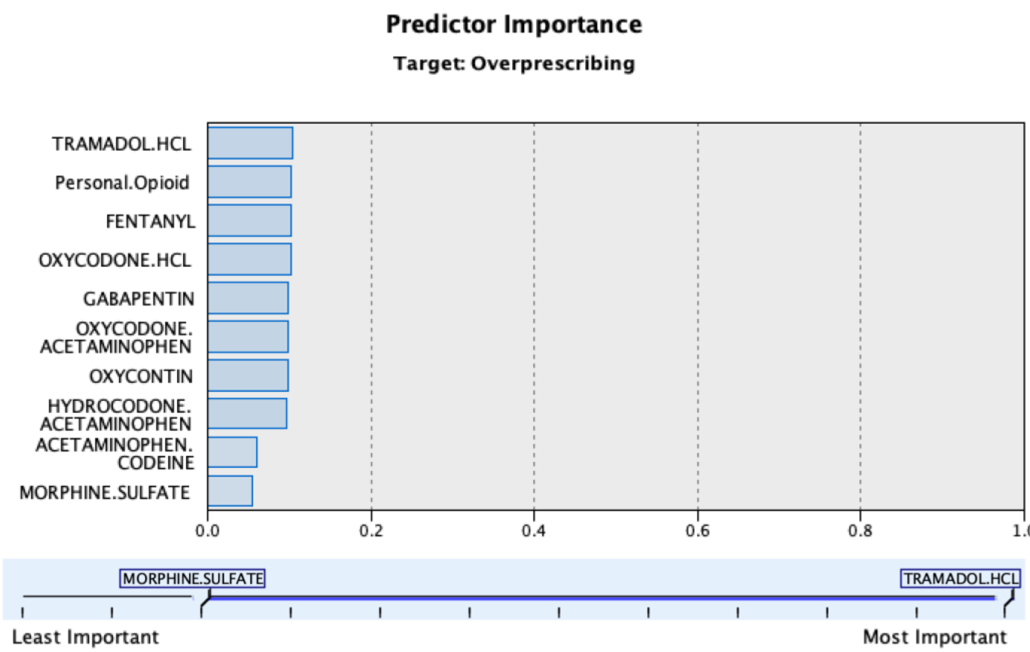
Our idea was to create unique models based on specific specialties. Our reasoning is that some specialties will require more opioid prescriptions (e.g. surgeries and hospitals) while some probably should not (e.g. family practice or others). This new Overprescription variable would be judged if a prescriber is prescribing opioids at a higher rate than average for their specific specialty. In other words, if Nurse Practitioners are prescribing 20 opioids on average in a given year, a prescriber within that specialty who prescribed 24 opioids would be labeled as 'Overprescription'.

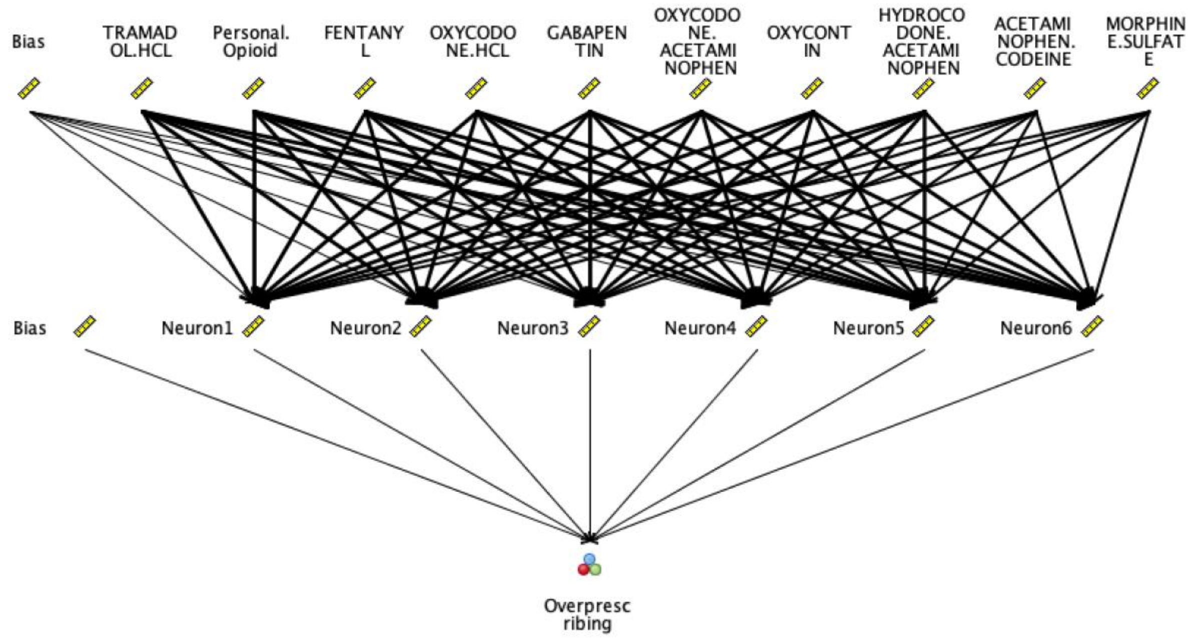
Now, this new variable does not mean that this is a good statistic for judging the opioid epidemic but it does provide more context and complexity to a model that, we feel, was too broad and vague at the onset.

To create this model, we needed to restructure the entire dataset and manipulate it to meet our needs. We used R heavily in this section to re-clean and wrangle the dataset. The R coding can be seen in the appendix (see **R code for Opioid.Prescriber Revision**).

We decided to focus on the Nurse Practitioner specialty, as it had 238 records, the third largest specialty in the data set. Our new data set filtered out any variable that was not an opioid (very similar to our previous ANN inputs), found the average total opioid prescriptions within that specialty and then found the sum of prescribed opioids given a specific prescriber. If the personal.opioid was higher than the average.opioid, the prescriber received a '1' for the Overprescription variable; a '0' if not.

We decided to us an Artificial Neural Network and used every opioid input except NPI and Specialty and made our target the Overprescribing Boolean.





Our model ended up having six neurons within the hidden layer and we saw Tramadol HCL have the highest predictor importance for modeling Overprescription for Nurse Practitioners.

The ANN model ended up with 100% accuracy in training and 97% accurate for testing; some overfitting there but we weren't able to get it much better and had limited inputs already.

Testing	0	1
0	63	0
1	2	11

Recall = TP / (TP + FN) = 11 / (11 + 2) = 84.6%

Specificity = TN / (TN + FP) = 63 / (63 + 0) = 1 - 1 = 0%

Precision = TP / (TP + FP) = 11 / (11 + 0) = 100%

This Specialty ANN, looking at just the Nurse Practitioner specialty with a new personalized Opioid.Prescriber Boolean, captured 84.6% of all Over-prescribers in the dataset. Non-Over-prescribers were classified as Over-Prescribers 0 times and the model predicted Over-prescribers at a 100% accuracy.

We might be dealing with a small sample size here but this model is very accurate. We think it's also fair because it judges all Nurse Practitioners against one another and not against all other prescribers. We don't know and can't really label how many prescriptions lead to drug abuse, addiction and death but we can make a model that looks to better understand over-prescribing more uniquely, which this model seems to do.

Conclusion

Dr. G. Caleb Alexander, the co-director of the Center for Drug Safety and Effectiveness at Johns Hopkins University, says that "there's no single data point that tells the whole story" regarding opioid abuse (Ornstein, 2018, para. 15). As we've discussed throughout this paper, we can't just blame prescribers for the opioid epidemic because they are on record for prescribing opioids. There needs to be context to those prescriptions and an understanding of the system and economy of health care. If we were to build and make decisions based on predictive models that used simple data, we would see a domino effect of other correlative problems in the future. We may already be seeing this issue with the shortage of necessary opioids in hospitals and the misuse and abuse of substitute painkillers.

Dr. Alexander and Leo Beletksy, the Northeastern professor referenced earlier, both believe that changing the opioid prescribing rate will have a less-than-desired effect in regards to quelling the crisis:

> "If you just think about prescribing, what's the right level of prescribing?" Alexander said. "Ultimately, what really matters is the quality of care that people are getting, both with respect to the management of their pain but also the identification and treatment of opioid use disorder. If, at the end of the day, we really care about how well people with

pain are doing, does any of this data really capture that?... Probably not" (Ornstein, 2018, para. 30).

As of now, 'death' is the only way that we are measuring the opioid epidemic (Ornstein, 2018). Regarding that statistic, Beletsky states that "'it really comes down to the fact that our country functions on a totally silly patchwork system' of tracking causes of death… It's a question of data speed, but it's also a question of data quality'" (Ornstein, 2018, para. 21). So 'death count' isn't even a good way to judge the opioid epidemic. Dr. Alexander recommends tracking overdoses and hospitalizations as a result of opioid use or looking at the rate of new cases of opioid use or addiction; Beletsky wants to see the percentage of people who are on maintenance treatment for addiction after a nonfatal overdose (Orenstein, 2018).

Right now, this type of data is not being collected. But instead of being critical of our collection methods of the past, we should be hopeful for the methods of the future. In regards to solving the opioid epidemic with predictive modeling, this is the beginning of discovery. If we can admit that we don't have quality data right now and start looking for a deeper understanding into the complexity of this crisis then maybe we can someday aspire to saving lives with data.

References

Bartolone, P. (2018, March 19). Opioid shortages leave US hospitals scrambling. *CNN*. Retrieved from https://www.cnn.com/2018/03/19/health/hospital-opioid-shortage-partner/index.html

Cunha, J. (n.d.). Methadone Hydrochloride. *RxList*. Retrieved from https://www.rxlist.com/consumer_methadone/drugs-condition.htm

Definitions of health care provider credentials. *CentraCare Health*. Retrieved from https://www.centracare.com/providers/definitions-of-health-care-provider-credentials/

Eustice, C. (2018, September 17). 10 things you need to know about tramadol. *VeryWellHealth*. Retrieved from https://www.verywellhealth.com/tramadol-10-things-you-should-know-190537

Eustice, C. (2018, October 2). Hydrocodone/Acetaminophen for treating pain. *VeryWellHealth*. Retrieved from https://www.verywellhealth.com/hydrocodone-acetaminophen-what-you-need-to-know-190546

Gabapentin. *WebMD*. Retrieved from https://www.webmd.com/drugs/2/drug-14208-8217/gabapentin-oral/gabapentin-oral/details

Goldman, B. (2010). *The night shift: Real life in the heart of the E.R.* Toronto, Ontario, Canada: HarperCollins Publishers.

Health. (2018, April 6). This commonly prescribed medication is contributing to the opioid epidemic. *Center on Addiction*. Retrieved from https://www.centeronaddiction.org/the-buzz-blog/commonly-prescribed-medication-contributing-opioid-epidemic

Lisinopril solution. *WebMD*. Retrieved from https://www.webmd.com/drugs/2/drug-6873-1785/lisinopril-oral/lisinopril-solution-oral/details

Methadone HCL solution. *WebMD*. Retrieved from https://www.webmd.com/drugs/2/drug-2671-2194/methadone-oral/methadone-oral/details

Opioids: Addiction, withdrawal and recovery. *Addictions and Recovery*. Retrieved from https://www.addictionsandrecovery.org/opioid-opiate-recovery.htm

Ornstein, C. (2018, March 13). Measuring the toll of the opioid epidemic is tougher than it seems. *ProPublica*. Retrieved from https://www.propublica.org/article/measuring-the-toll-of-the-opioid-epidemic-is-tougher-than-it-seems

Ornstein, C. and Jones, R.G. (2018, June 28). Opioid makers, blamed for overdose epidemic, cut back on marketing payments to doctors. *ProPublica*. Retrieved from https://www.propublica.org/article/opioid-makers-blamed-for-overdose-epidemic-cut-back-on-marketing-payments-to-doctors

Oxycodone and acetaminophen (oral route). *Mayo Clinic*. Retrieved from https://www.mayoclinic.org/drugs-supplements/oxycodone-and-acetaminophen-oral-route/description/drg-20074000

Pryor, A. (2016). U.S. opiate prescriptions/overdoses. *Kaggle*. Retrieved from https://www.kaggle.com/apryor6/us-opiate-prescriptions

Sailor, C. (2018, January 18). Painkiller shortage hits south sound. *The News Tribune*. Retrieved from https://www.thenewstribune.com/news/local/article195434574.html

Tate, N. (2018, May 1). Fentanyl: What you should know. *WebMD*. Retrieved from https://www.webmd.com/mental-health/addiction/news/20180501/fentanyl-what-you-should-know

Thomas, K. (2017, September 17). Amid opioid crisis, insurers restrict pricey, less addictive painkillers. *ProPublica*. Retrieved from https://www.propublica.org/article/insurers-limit-coverage-of-pricey-less-addictive-painkillers

What is a physician assistant? (PA-C). *Family Care Network*. Retrieved from https://www.familycarenetwork.com/provider-types/what-is-a-physician-assistant

What is acetaminophen-codeine? *Everyday Health*. Retrieved from https://www.everydayhealth.com/drugs/acetaminophen-codeine

Appendix

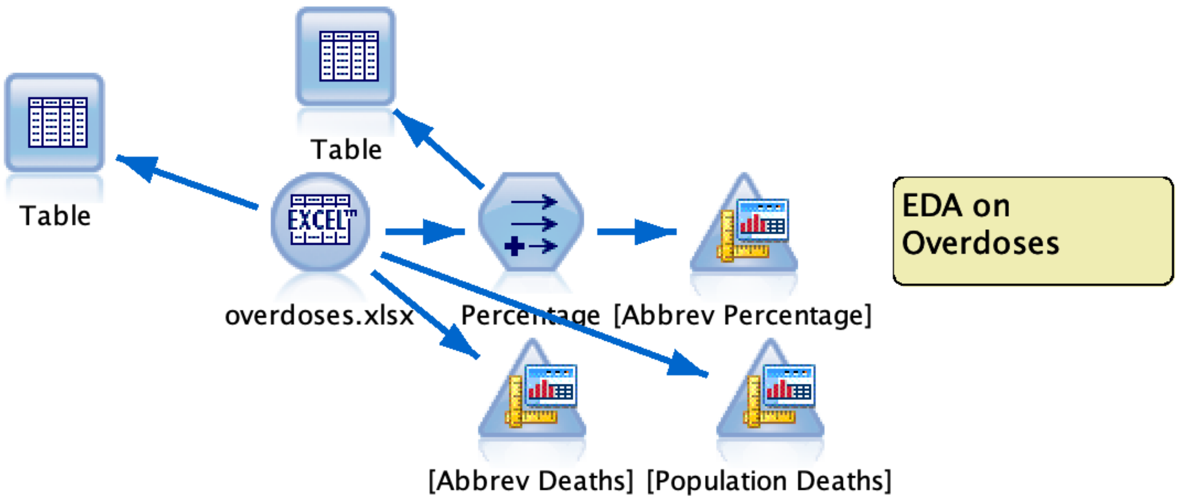
R code for dummy variable cleaning:

```
opnorm <- opioids_workingcopy

#drop credentials
opnorm <- opnorm %>%
  select(-Credentials)

#creating dummy variables for normalization of dataset
library(dummies)

#create 'state' dummy
opnorm <- cbind(opnorm, dummy(opnorm$State, sep = "_"))

#create 'gender' dummy
opnorm <- cbind(opnorm, dummy(opnorm$Gender, sep = "_"))

#create 'specialty' dummy
opnorm <- cbind(opnorm, dummy(opnorm$Specialty, sep = "_"))

#export dataframe to excel file for SPSS use
library(openxlsx)
write.xlsx(opnorm, 'opioidnormalized.xlsx')
```

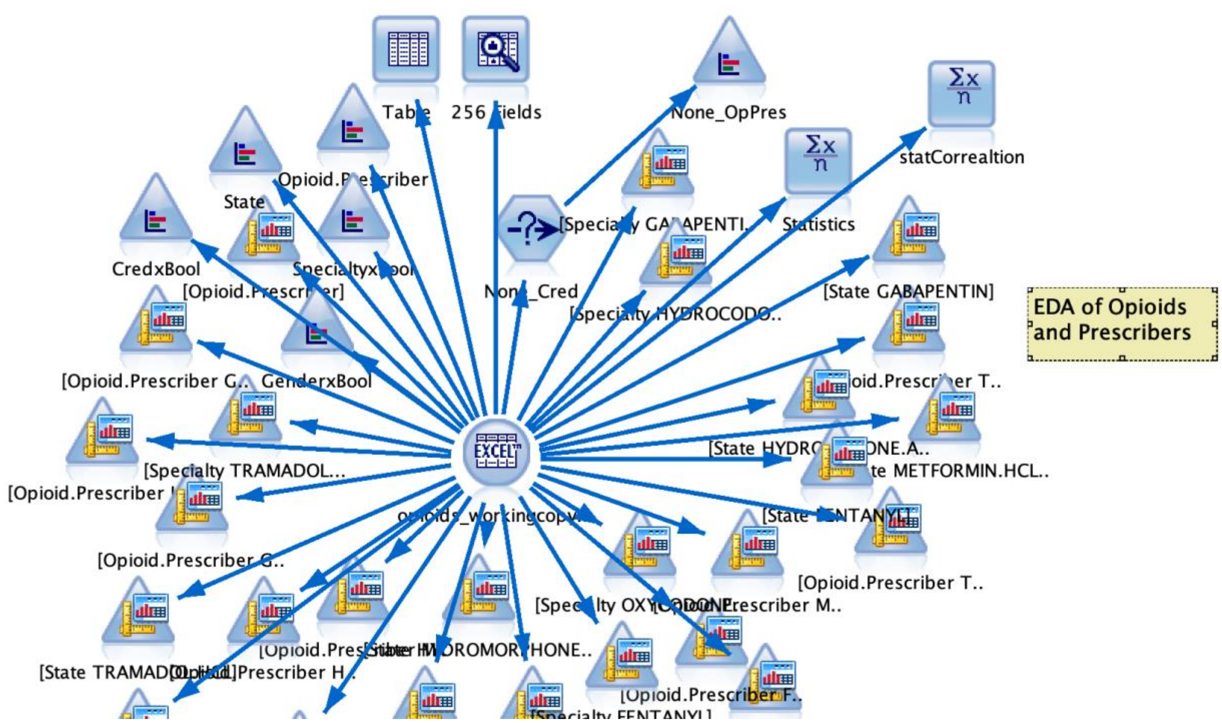
R code for Opioid.Prescriber Revision:

```
#Revisionist Opioid.Prescriber
revop <- opioids_workingcopy

#First, want to know average number of opioids per a distinct specialty
avgNP <- revop %>%
  select(-Gender, - State, -Credentials, -Opioid.Prescriber)

#Chose to look at the Nurse Practitioner specialty, which has 238 records
avgNP <- avgNP %>%
  select(NPI, Specialty, ACETAMINOPHEN.CODEINE, FENTANYL, GABAPENTIN,
HYDROCODONE.ACETAMINOPHEN,
         HYDROMORPHONE.HCL, METHADONE.HCL, MORPHINE.SULFATE,
OXYCODONE.ACETAMINOPHEN,
         OXYCODONE.HCL, OXYCONTIN, TRAMADOL.HCL) %>%
  filter(avgNP$Specialty == "Nurse Practitioner")

#Now want to know the Mean of each opioid prescription
avgOp <- avgNP %>%
  group_by(Specialty) %>%
  summarise(avg_ACETAMINOPHEN.CODEINE = mean(ACETAMINOPHEN.CODEINE),
            avg_FENTANYL = mean(FENTANYL),
            avg_GABAPENTIN = mean(GABAPENTIN),
            avg_HYDROCODONE.ACETAMINOPHEN =
mean(HYDROCODONE.ACETAMINOPHEN),
            avg_HYDROMORPHONE.HCL = mean(HYDROMORPHONE.HCL),
```

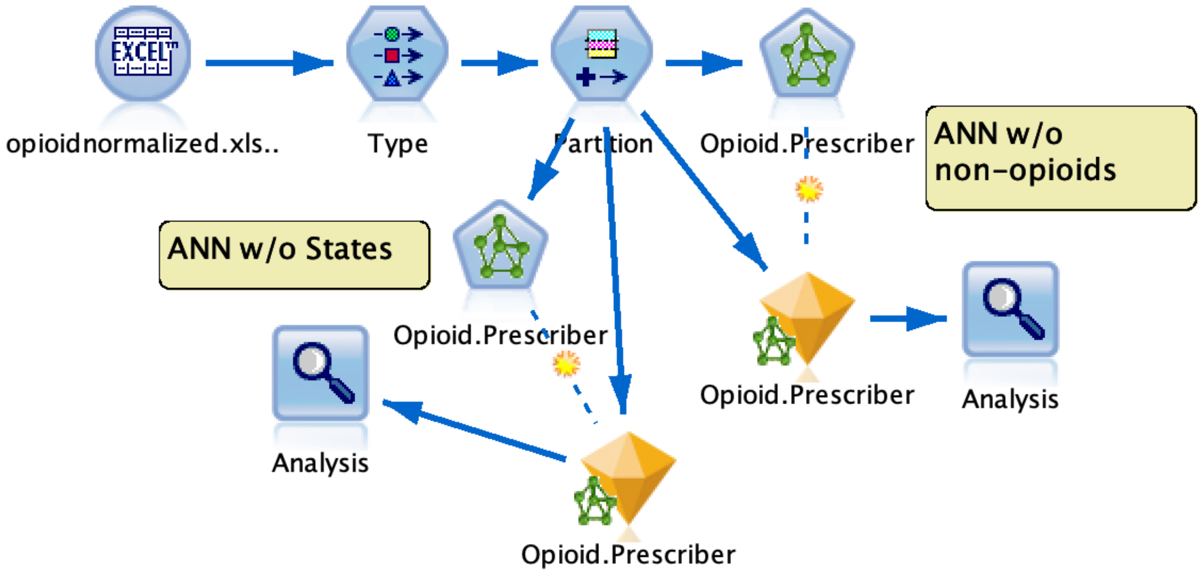
```
            avg_METHADONE.HCL = mean(METHADONE.HCL),
            avg_MORPHINE.SULFATE = mean(MORPHINE.SULFATE),
            avg_OXYCODONE.ACETAMINOPHEN = mean(OXYCODONE.ACETAMINOPHEN),
            avg_OXYCODONE.HCL = mean(OXYCODONE.HCL),
            avg_OXYCONTIN = mean(OXYCONTIN),
            avg_TRAMADOL.HCL = mean(TRAMADOL.HCL))

#Add all opioids together and get the average
average_opioid <- avgNP %>%
  select(everything()) %>%
  mutate(Average.Opioid = sum(avgOp$avg_ACETAMINOPHEN.CODEINE,
                              avgOp$avg_FENTANYL,
                              avgOp$avg_GABAPENTIN,
                              avgOp$avg_HYDROCODONE.ACETAMINOPHEN,
                              avgOp$avg_HYDROMORPHONE.HCL,
                              avgOp$avg_METHADONE.HCL,
                              avgOp$avg_MORPHINE.SULFATE,
                              avgOp$avg_OXYCODONE.ACETAMINOPHEN,
                              avgOp$avg_OXYCODONE.HCL,
                              avgOp$avg_OXYCONTIN,
                              avgOp$avg_TRAMADOL.HCL))

#Then create another column to add up a prescriber's total opioid
prescriptions
NPopioid <- average_opioid %>%
  mutate(Personal.Opioid = avgNP$ACETAMINOPHEN.CODEINE +
                               avgNP$FENTANYL +
                               avgNP$GABAPENTIN +
                               avgNP$HYDROCODONE.ACETAMINOPHEN +
                               avgNP$HYDROMORPHONE.HCL +
                               avgNP$METHADONE.HCL +
                               avgNP$MORPHINE.SULFATE +
                               avgNP$OXYCODONE.ACETAMINOPHEN +
                               avgNP$OXYCODONE.HCL +
                               avgNP$OXYCONTIN +
                               avgNP$TRAMADOL.HCL)

#Now we want to use an ifelse function to create a revisionist
Opioid.Prescriber Boolean variable
Op.Pres.Revision <- NPopioid %>%
  select(everything()) %>%
  mutate(Overprescribing = ifelse(Personal.Opioid > Average.Opioid, "1",

"0"))
```



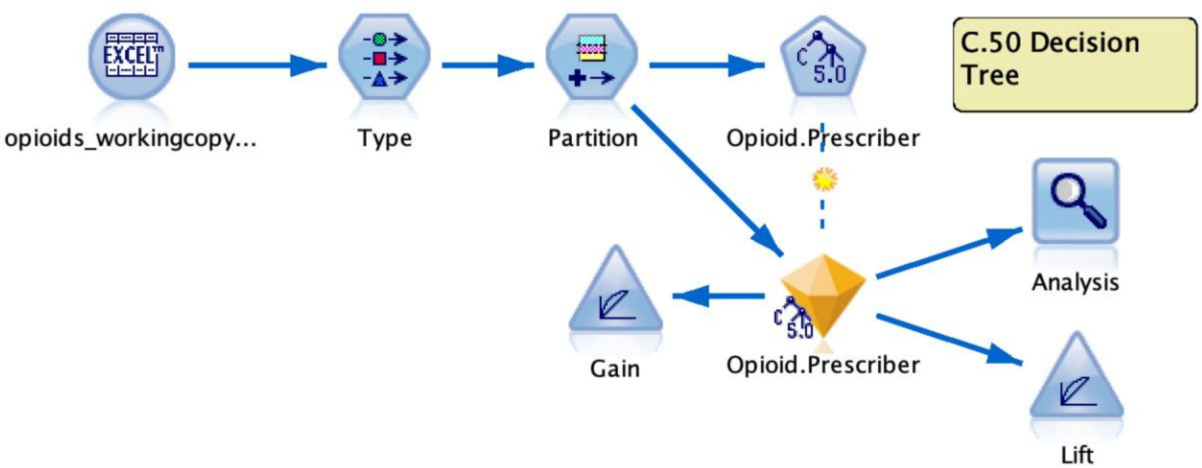
Exploratory Data Analysis on the Overdoses excel file



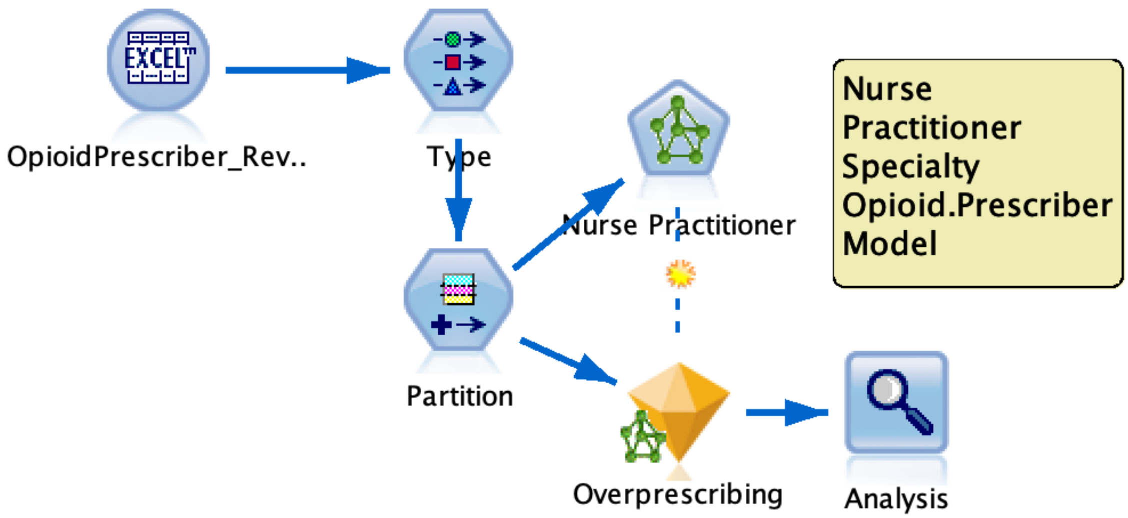
Exploratory Data Analysis on the Opioids and Prescriber Excel file



The Artificial Neural Network models; one with all inputs sans non-opioids and one without States.



The C.50 Decision Tree model



The Specialty-specific Overprescribing ANN