

▼ Background

Our dataset contains transaction information on hundreds of thousands of recorded opioid sales in the United States from 2006-2012. Our data comes from the DEA's ARCOS Dataset, which kept record of all opioid transactions to pharmacies in every state for these 6 years. Our dataset does not contain any data on individual buyers, but rather contains data about where the transaction occurred, when, what was sold, and how much was sold to each retailer. Combining this data from Arcos with Census data for Ohio counties, such as average age and average income, can give us an idea of what company targeted what area, or what demographic. We can use the data we have available to us to find key characteristics that pharmaceutical companies look for when deciding who to target for opioid sales. Since the opioid epidemic is such a widespread issue throughout the United States even today, this is information the public needs to see and understand. Anyone effected by the opioid epidemic should investigate the full extent of the dangerous overprescription by these companies in both Ohio cities and the United States as a whole.

▼ Data

Our data gives us access to multiple columns of information on hundreds of thousands of opioid transactions to pharmacies around the United States. To find out what companies sold to how many people and where, we will have to thoroughly explore our data around different times and in different states and counties within states. Combining our transaction data with United States Census data about individual states and counties allows us to further explore what places the opioid crisis has affected the most. By combining data from these two datasets, for example, we can find what Ohio counties the most opioid transactions have occurred in, how many opioids there are in the county compared to how many people live there, and how many sales there were in different counties of varying economic status. We can also get a general idea of what type of person pharmaceutical companies typically target opioid sales to, as we have access to average county age and average income per person in each county.

SHOW CODE

SHOW CODE

```
↳ Mounted at /content/drive
'drive/My Drive/arcos-oh-statewide-itemized.tsv.gz'
'drive/My Drive/co-est00int-01-39.xls'
```

SHOW CODE

🔗 /usr/local/lib/python3.6/dist-packages/IPython/core/interactiveshell.py:2718: DtypeWarning

Columns (3,5,22,25,26,27) have mixed types. Specify dtype option on import or set low_

	BUYER_NAME	BUYER_ADDRESS1	BUYER_ADDRESS2	BUYER_CITY	BUYER_STATE	BUYER_ZIP
0	BACH, JOHN M DVM	WAVERLY VETERINARY HOSPITAL INC	95 WOODLAND DRIVE	WAVERLY	OH	45690
1	COUNTS, GAIL E DVM	SHAWNEE ANIMAL CLINIC	101 BIERLY RD	PORTSMOUTH	OH	45662
2	COUNTS, GAIL E DVM	SHAWNEE ANIMAL CLINIC	101 BIERLY RD	PORTSMOUTH	OH	45662
3	DRESBACH, HUGH DELNO DVM	1130 NORTH COURT STREET	NaN	CIRCLEVILLE	OH	43113
4	FARST, DAVID W DVM	510 N MAIN STREET	NaN	ARCANUM	OH	45304

SHOW CODE

🔗

	County	2006	2007	2008	2009	countyPop
0	.Adams County	28615.0	28538.0	28686.0	28500.0	28584
1	.Allen County	106716.0	106717.0	106773.0	106518.0	106681
2	.Ashland County	53231.0	53544.0	53499.0	53305.0	53394
3	.Ashtabula County	102340.0	102156.0	101826.0	101584.0	101976
4	.Athens County	64237.0	64102.0	64546.0	64618.0	64375

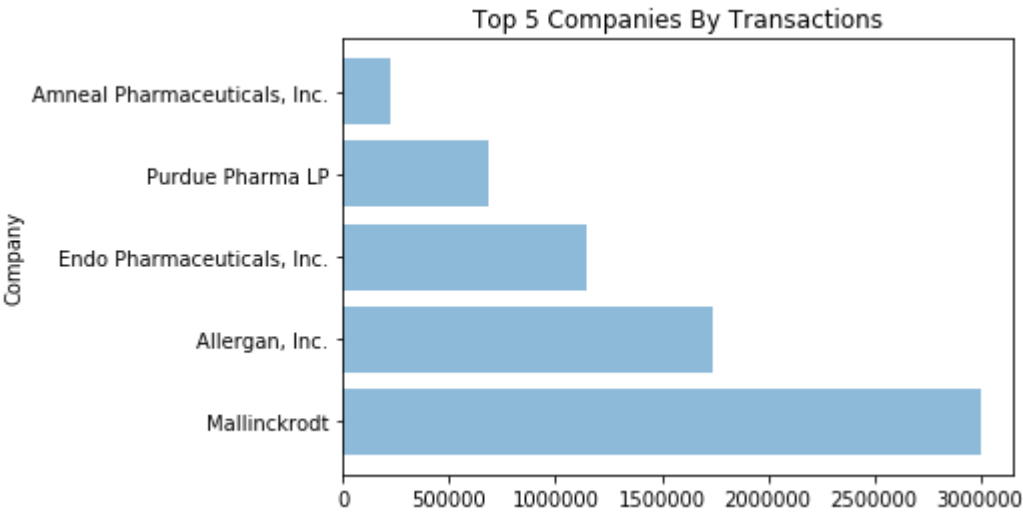
▼ Methodology

Using the data we have available from these datasets, we can explore such questions as which counties have been sold the most opioids, which drugs have been sold the most, and what companies are most responsible for these transactions.

Below is a basic visualization of the top 5 pharmaceutical companies with the most transactions within all 7 included years of the ARCOS dataset. We can use this data to get a better idea of what

companies have gained the most from overprescribing opioids in the state of Ohio. From here, we can explore what counties these companies have sold to the most, and what medication these companies are selling in these transactions. We will also be able to tell the top overprescribing companies in each county, and if these companies typically target old or young counties, or perhaps high income or low income counties.

SHOW CODE



We can also use the data available to us to compare each county in a state to see what areas have been affected most over time. We are able to see how transactions per county have changed from year to year. We can also use locational data from ARCOS to see what medications have been sold the most by county and what counties have the most transactions per the population of the county.

Below is a simple heatmap generated using plotly that shows how many transactions have occurred in each county for every year combined in the ARCOS dataset. This heatmap can be replicated for each year to show the change in amount of transactions per county over time.

SHOW CODE



```

Requirement already satisfied: plotly-geo in /usr/local/lib/python3.6/dist-packages (1
Requirement already satisfied: geopandas==0.3.0 in /usr/local/lib/python3.6/dist-packa
Requirement already satisfied: fiona in /usr/local/lib/python3.6/dist-packages (from g
Requirement already satisfied: pandas in /usr/local/lib/python3.6/dist-packages (from
Requirement already satisfied: descartes in /usr/local/lib/python3.6/dist-packages (fr
Requirement already satisfied: shapely in /usr/local/lib/python3.6/dist-packages (from
Requirement already satisfied: pyproj in /usr/local/lib/python3.6/dist-packages (from
Requirement already satisfied: six>=1.7 in /usr/local/lib/python3.6/dist-packages (fro
Requirement already satisfied: cligj>=0.5 in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: munch in /usr/local/lib/python3.6/dist-packages (from f
Requirement already satisfied: click-plugins>=1.0 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: attrs>=17 in /usr/local/lib/python3.6/dist-packages (fr
Requirement already satisfied: click<8,>=4.0 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: python-dateutil>=2.6.1 in /usr/local/lib/python3.6/dist
Requirement already satisfied: matplotlib in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.6/dist-pack
Requirement already satisfied: pyparsing!=2.0.4,!2.1.2,!2.1.6,>=2.0.1 in /usr/local/
Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: pyshp==1.2.10 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: shapely==1.6.3 in /usr/local/lib/python3.6/dist-package
Requirement already satisfied: plotly==4.1.0 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: six in /usr/local/lib/python3.6/dist-packages (from plo
Requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.6/dist-packag
/usr/local/lib/python3.6/dist-packages/pandas/core/frame.py:7138: FutureWarning:

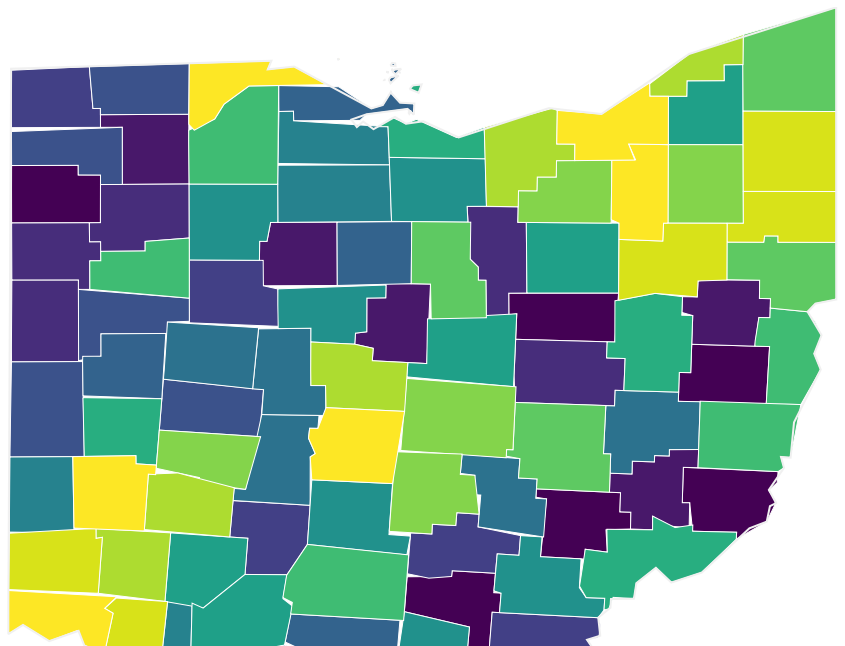
```

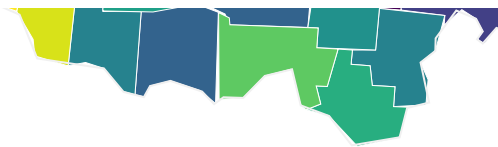
Sorting because non-concatenation axis is not aligned. A future version of pandas will change to not sort by default.

To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort=True'.

Total Opioid Transactions in Ohio

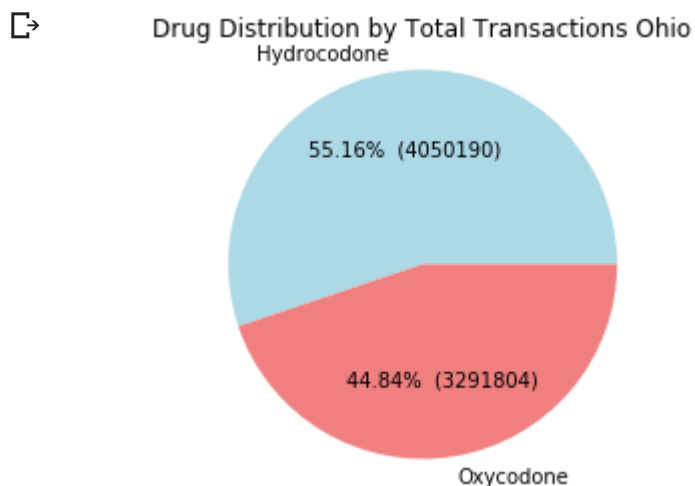




We can also use the data we have from ARCOS to find out what type of prescription medication has been sold to what pharmacy, how much of it has been sold, and we can find out the total magnitude of how much medication has been sold by exploring how much total weight of what medication has been sold. Exploring types of prescriptions can provide more insight as to what medications have been most targeted to what county and perhaps we can also tell if a medication is more likely to be sold in a poor or wealthy county, or a more or less populated county.

Below is another basic visualization created using the library Matplotlib showing the distribution of total transactions including each drug type in Ohio over all years the ARCOS database covers. This can be taken further by showing the total weight of each drug sold in each county. We can get a better understanding of just how many opioids have been purchased in each county by calculating how much of each medication is included in all transactions in a State or county, both by number of transactions and by total weight.

SHOW CODE



From here, we can use machine learning to further explore the data. By clustering the data by county and combining the data from the United States Census site with the data from ARCOS, we can see where counties fall when comparing average age or average income of a county by total transactions in each county. We can cluster this county information to create groups based on transaction information. We can find what age groups have been targeted most by pharmaceutical companies and what drug is most likely to be sold to a county of what average income.

We can also use machine learning to make predictions on whether a county is likely to have a high amount of opioid transactions based on transaction and population data we have access to. We can make predictions based on a variety of quantitative values we have available to us from this set. Perhaps by using machine learning we can also tell what county is likely to sell a certain type of drug to a certain demographic. The datasets we are exploring and analyzing can give a much clearer picture of how severe the impact of the opioid epidemic has been, and where it has had the greatest effect on the country.