# Analysis of the Opioid Epidemic

James Bonaffini, Abdou Ndong, Max Ryoo

Motivation

- There has been an Opioid Epidemic in the United States for the last decade
  - 700,000+ drug related deaths from 1999 to 2017
  - 70,000 drug-related deaths in 2017, with nearly 50,000 opioid-related
- Many actors in this crisis - we want to make sense of the data and find patterns
  - Prescribers
  - Pharmaceutical Companies
  - Insurance Companies
  - Patients
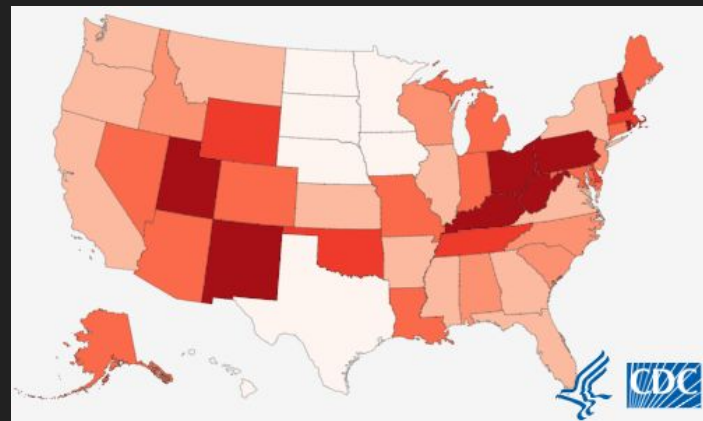
2014 Drug-Related Deaths, CDC

Data

- Data originally from 2014 Medicare Part D prescription data - nearly 25 million records
- Dr. Alan Pryor downloaded, cleaned, and collated 25,000 prescriber records; posted them on kaggle.
- Observations organized by each unique prescriber
  - State
  - Credentials
  - Gender
  - Specialty
  - Prescription records of over 200 unique drugs

ABILIFY
ACETAMINOPHEN.CODEINE
ACYCLOVIR
ADVAIR.DISKUS
AGGRENOX
ALENDRONATE.SODIUM
ALLOPURINOL
ALPRAZOLAM
AMIODARONE.HCL
AMITRIPTYLINE.HCL
AMLODIPINE.BESYLATE
AMLODIPINE.BESYLATE.BENAZEPRIL
AMOXICILLIN
AMOX.TR.POTASSIUM.CLAVULANATE
AMPHETAMINE.SALT.COMBO
ATENOLOL
ATORVASTATIN.CALCIUM
AVODART
AZITHROMYCIN
BACLOFEN

And the list goes on...

Data Cleaning

- Organization of drugs into a drug class
- Cleanup of credential variation
  - MD, M.D., MD. PhD. → MD
- One-hot encoding
  - Gender
  - Specific Specialty → General Categories

Drug
ACETAMINOPHEN.CODEINE
FENTANYL
HYDROCODONE.ACETAMINOPHEN
HYDROMORPHONE.HCL
METHADONE.HCL
MORPHINE.SULFATE
MORPHINE.SULFATE.ER
OXYCODONE.ACETAMINOPHEN
OXYCODONE.HCL
OXYCONTIN
TRAMADOL.HCL

Class	Count	Class	Count	Class	Count
antipsychotic	7	alpha.adrenergic.agonist	2	bronchodilator	3
opiate	11	diuretic	8	nitrate	3
antiviral	1	anticonvulsant	12	biguanide	3
corticosteroid	11	decarboxylase.inhibitor	1	DPP4 inhibitor	2
biphosphonate	2	anti.inflammatory	7	supplement	6
xanthine.oxidase.inhibitor	1	antimicrobial	1	laxative	2
benzodiazepine	6	anticoagulant	4	prostaglandin.analog	3
antiarrhythmic	1	antiplatelet	1	anesthetic	1
antidepressant	17	antifungal	3	antimetabolite	1
calcium.channel.blocker	11	antigout	1	prokinetic agent	1
antibiotic	12	proton.pump.inhibitor	5	leukotriene.receptor.antagonist	1
amphetamine	1	cardiac.glycoside	2	carboxylic acid	1
beta.blocker	12	antidiarrheal	1	NMDA.inhibitor	2
statin	8	carbonic anhydrase inhibitor	1	antiemetics	2
alpha.blocker	5	hormone	5	thiazolidinediones	1
muscle.relaxant	6	cholinesterase inhibitor	1	dopamine agonist	2
insulin	9	H2.blocker	2	anti.anginal	1
ACE.inhibitor	6	antilipemic	2	immunosuppressant	1
angiotensin.receptor.blocker	8	fibrate	1	protectant	1
anticholinergic	8	sulfonylurea	5	nucleoside analog	1
antihistimine	3	vasodilator	1		

Exploratory Data Analysis

Types of Clinicians

- **Most of the prescribing clinicians are doctors**
  - Pain Management, Rehabilitation, Anesthesiology Prescribed the most
- **More opiates prescribed does not equal more deaths per population**
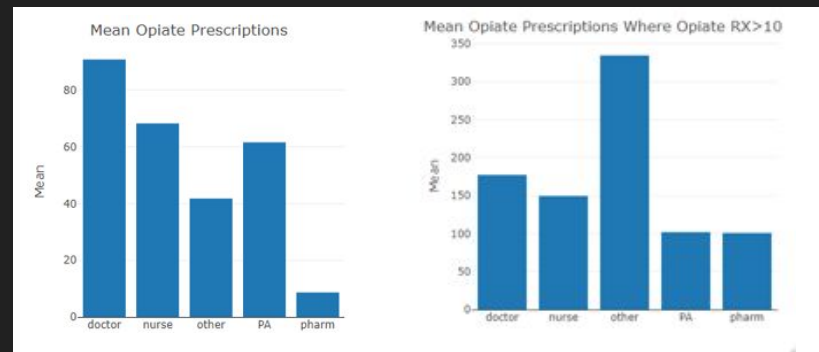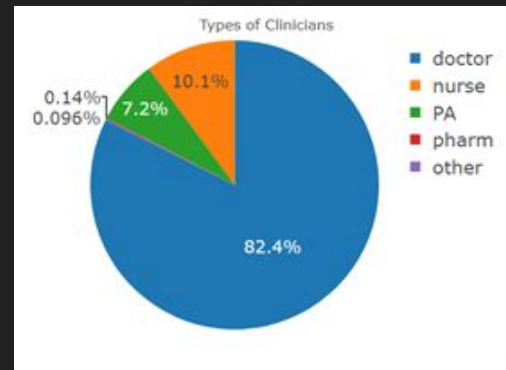
Population	Deaths	State	medOpiate	avgOpiate	deathper100k	rank.avgOpiate	rank.medOpiate	rank.deaths
4,833,722	723	AL	26	178.37	14.96	1	2	25
4,625,470	777	LA	13	170.64	16.80	2	23	23
2,991,207	336	MS	33	154.16	11.23	3	1	43
2,959,373	356	AR	22	142.69	12.03	4	5	38
6,495,978	1,269	TN	17	137.61	19.54	5	12	11

Supervised Learning

	MSE
Linear Reg	20.37189
AIC Step	20.38088
AIC Forward	20.38331
BIC Step	20.41889
BIC Forward	20.40971
Ridge Reg	20.39705
Lasso Reg	20.39326

|	MSE
$k=10$	22.06087
$k=20$	21.04554
$k=30$	20.81493
$k=40$	20.6957
$k=50$	20.6161

- 80% Training, 20% Testing Set
- Methods
  - Regression
  - KNN
- Using number of deaths per 100,000 people as the response variable
- As predicted from our exploratory data analysis, the number of opiates was not a great predictor
- Most of the prediction was from the regression intercept - very little variance in the predicted response
- Finding patterns using unsupervised learning might be more fruitful

Clustering (Unsupervised)

- K-means Clustering
  - 1:20 nodes
    - 3 - 4 true centroids (clusters)
- Evaluation methods
  - Elbow (WSS) (3 or 4)
  - Average Silhouette function (2)
  - Gap Statistics (5)
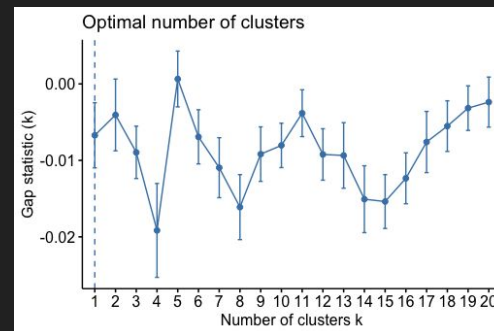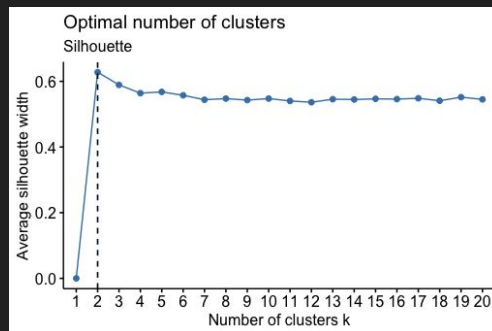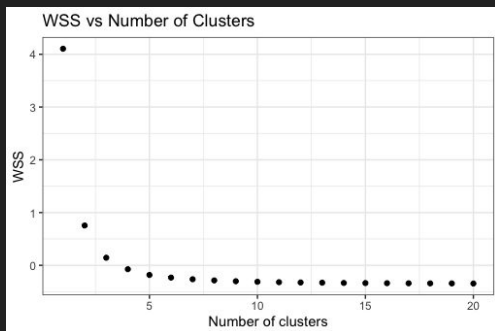
# Principal Component Analysis (PCA) (Unsupervised)

- 14 variables only are accounted for a variance of 90% and they are shown in the table below
- No correlation between Deaths and number of Opiates Prescriptions
- Classifying the drug variables into 3 groups using kmeans clustering show that there is a pattern in the data even though there is no much variance between them.
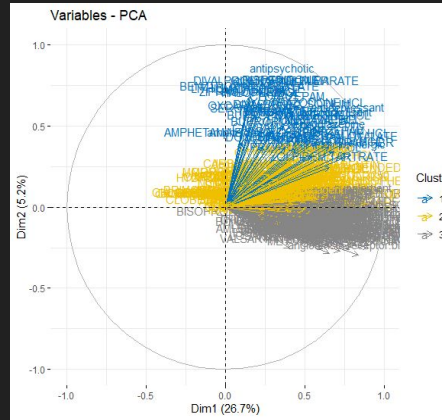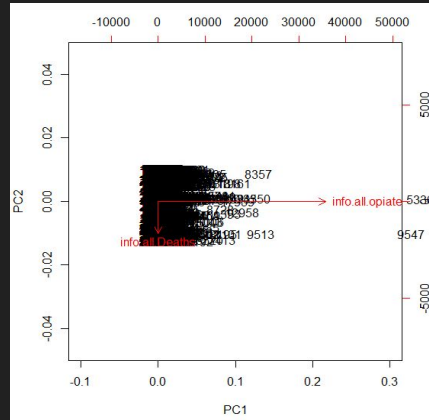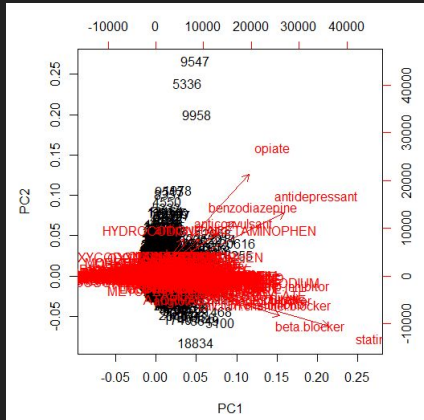
	correlation	p.value
diuretic	0.91807507	0.000000e+00
statin	0.90313305	0.000000e+00
calcium.channel.blocker	0.90255609	0.000000e+00
ACE.inhibitor	0.89334063	0.000000e+00
AMLODIPINE.BESYLATE	0.88803911	0.000000e+00
proton.pump.inhibitor	0.87512555	0.000000e+00
LEVOTHYROXINE.SODIUM	0.86886159	0.000000e+00
FUROSEMIDE	0.86740293	0.000000e+00
beta.blocker	0.86614058	0.000000e+00
supplement	0.86539285	0.000000e+00
LISINOPRIL	0.85700889	0.000000e+00
SIMVASTATIN	0.85600716	0.000000e+00
biguanide	0.85139160	0.000000e+00
OMEPRAZOLE	0.84757992	0.000000e+00
hormone	0.84599760	0.000000e+00
METFORMIN.HCL	0.83887267	0.000000e+00
angiotensin.receptor.blocker	0.83685980	0.000000e+00
POTASSIUM.CHLORIDE	0.83238723	0.000000e+00
METOPROLOL.TARTRATE	0.82379978	0.000000e+00
sulfonylurea	0.81342089	0.000000e+00
ATORVASTATIN.CALCIUM	0.79447445	0.000000e+00
anticholinergic	0.78808900	0.000000e+00

PCA (Unsupervised, cont'd)

- Creating binary variables between a certain amount of prescribed opiates/more and less the amount show a clear trend in the data. The binaries for the opiates prescriptions are 10 or more, 50 or more, 500 or more and 1000 or more.
- Some patterns related to opiates prescriptions could be found in the dataset. We have some insights in the data but further work will be needed to be done in this regards.
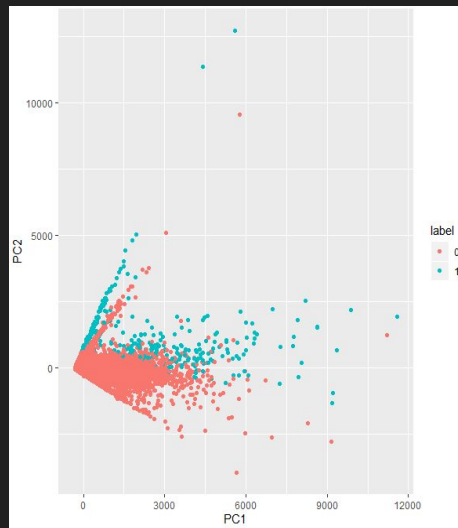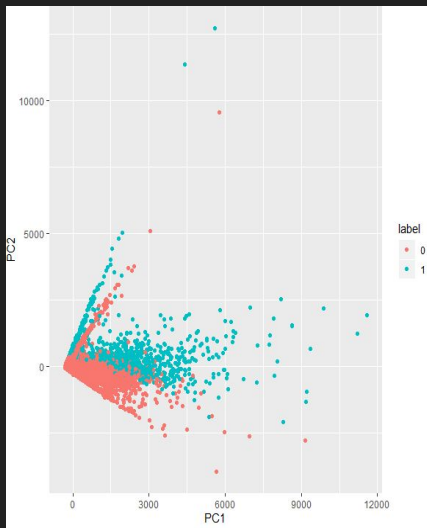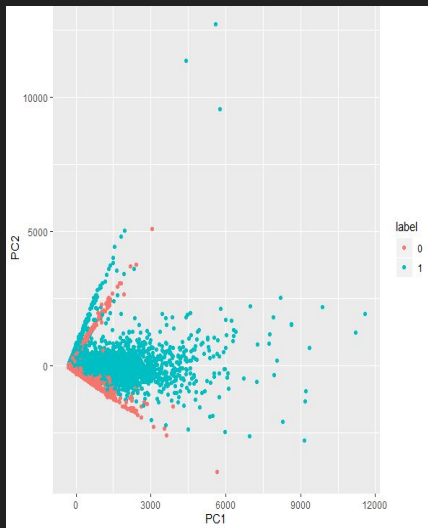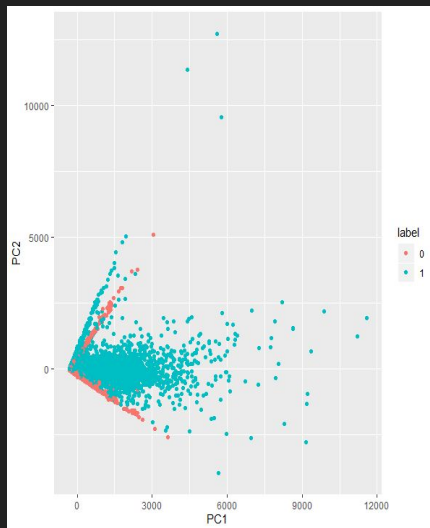
Model Pros & Cons

- Supervised
    - Known input and output
    - Unsuccessful at predicting impact of opiates on drug-related deaths
- Unsupervised
    - Clusters existed - there are patterns in the data
    - Difficult to interpret how the data is clustered
    - Can't directly distinguish a factor that causes opiate overdose.

Significance & Conclusions

- Data
  - CMS makes data easily accessible
  - Lots of work must be done to clean and interpret the data
- Models
  - There wasn't a clear indicator to predict whether a prescription will lead to more deaths
  - Patterns existed, based on unsupervised approaches
- Possible directions
  - Perhaps most opioid deaths are due to illegally obtained drugs
  - Apply these model concepts to pharmaceutical manufacturing data
- This is a challenging topic that is worth exploring further

Questions?

Resources

Kaggle

- https://www.kaggle.com/apryor6/us-opiate-prescriptions

CDC

- https://www.cdc.gov/drugoverdose/

- https://www.cdc.gov/drugoverdose/data/statedeaths/drug-overdose-death-2014.html

CMS

- https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber