Drug Overdoses in the United States

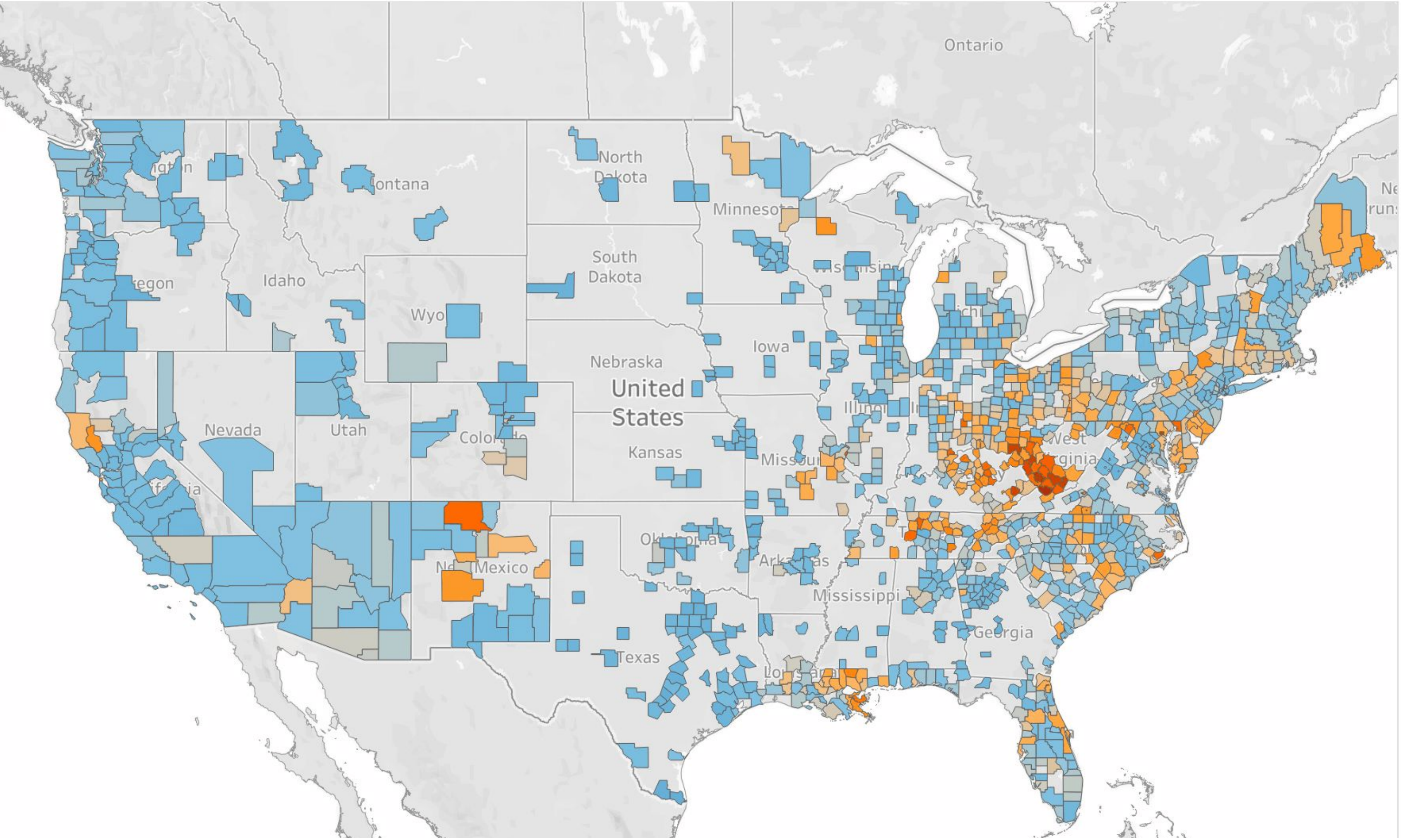
Ashlyn Jew, Meera Duggal, Rhiann Zhang, Wei Deng

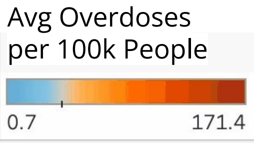
How do different aspects of counties (e.g. location, demographics and opioid dispense rates) relate to drug overdose rates?

Can we utilize the various demographic trends of each county to more accurately explain drug overdoses?

Previously...

Average Overdose Rate - 2020





What we've Considered:

-Missing Data

-More than half of population level tho!

-Increasing overdose rates throughout the years

-Possible clustering of overdose rates

Exhaustive Search

- Our preliminary best subset:
  - Year
  - Unemployment_rate
  - Dispense_rate (Number of Opioid Prescriptions per 100 people)
  - AA_MALE (Number of Males that identify exclusively as Asian American)
  - TOM_MALE (Number of Males that identify with two or more ethnicities)
  - NH_MALE (Number of Males that identify as non-Hispanic)
  - Jail Population
  - Incarceration Rate per 100k People
  - PoveryCount (Number of people below the poverty threshold)
  - MedianHHI (Median Household Income)

- All variables were standardized (except Year and our response variable Overdose Rate)

Best Subset OLS

$$y_{i,t} = \underbrace{\beta_{i0,t}} + \sum_{k=1}^{p} \underbrace{\beta_{ik,t}}\underbrace{x_{ik,t}} + \underbrace{\epsilon_{i,t}}$$

drug overdose death rate at county i, for year t

intercept coefficient at county i, for year t

k-th local regression coefficient for the k-th explanatory variable at county i, for year t

k-th explanatory variable at county i, for year t

the random error term associated with county i, for year t

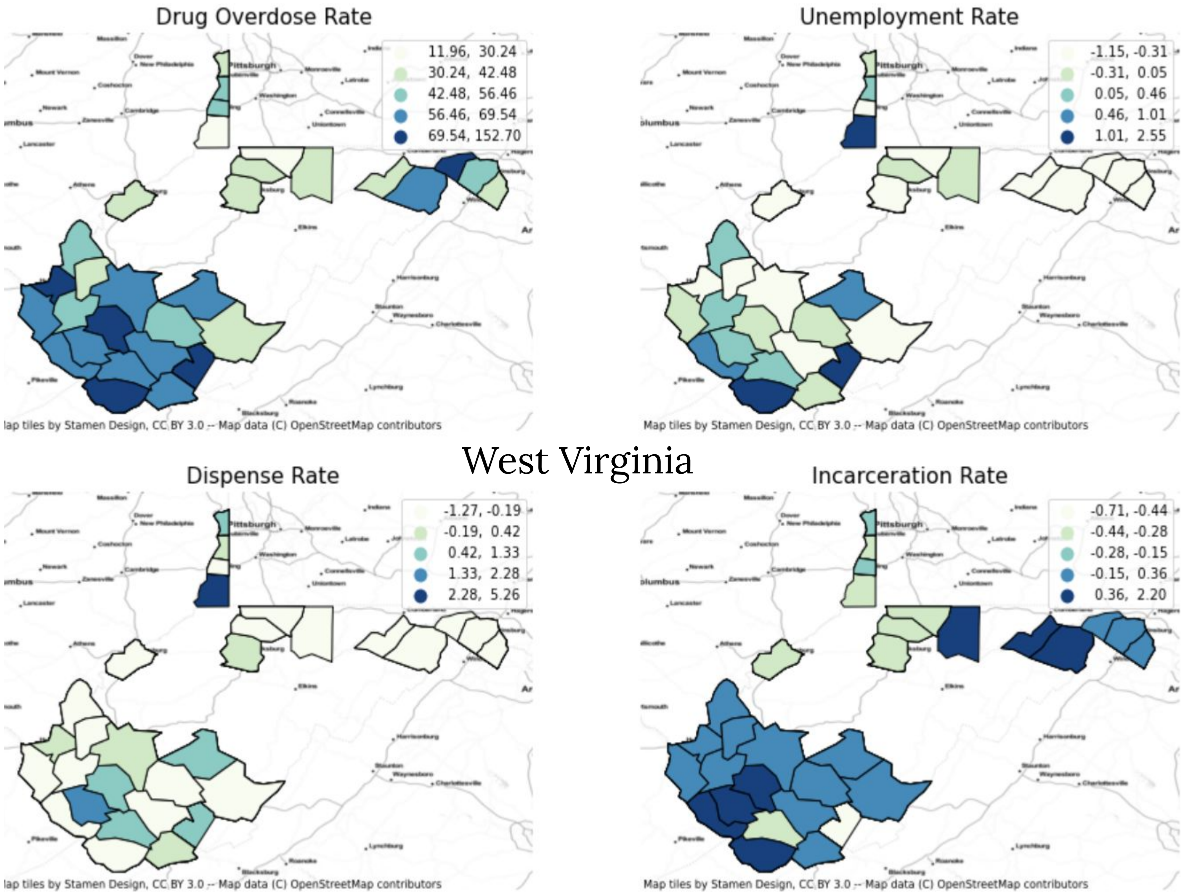
	Best Subset (BSS) OLS	
	Estimate	P-value
Intercept	-223.7	0.0
Year	0.112	0.0
Unemployment_rate	0.082	0.0
Dispense_rate	0.178	0.0
AA_MALE	0.012	0.353
TOM_MALE	-0.032	0.106
NH_MALE	0.163	0.0
Jail_Population	-0.043	0.009
Incarceration_Rate_per_100k	0.008	0.18
PovertyCount	-0.15	0.0
MedianHHI	-0.121	0.0

overdose rate (response variable) is log transformed

Note:

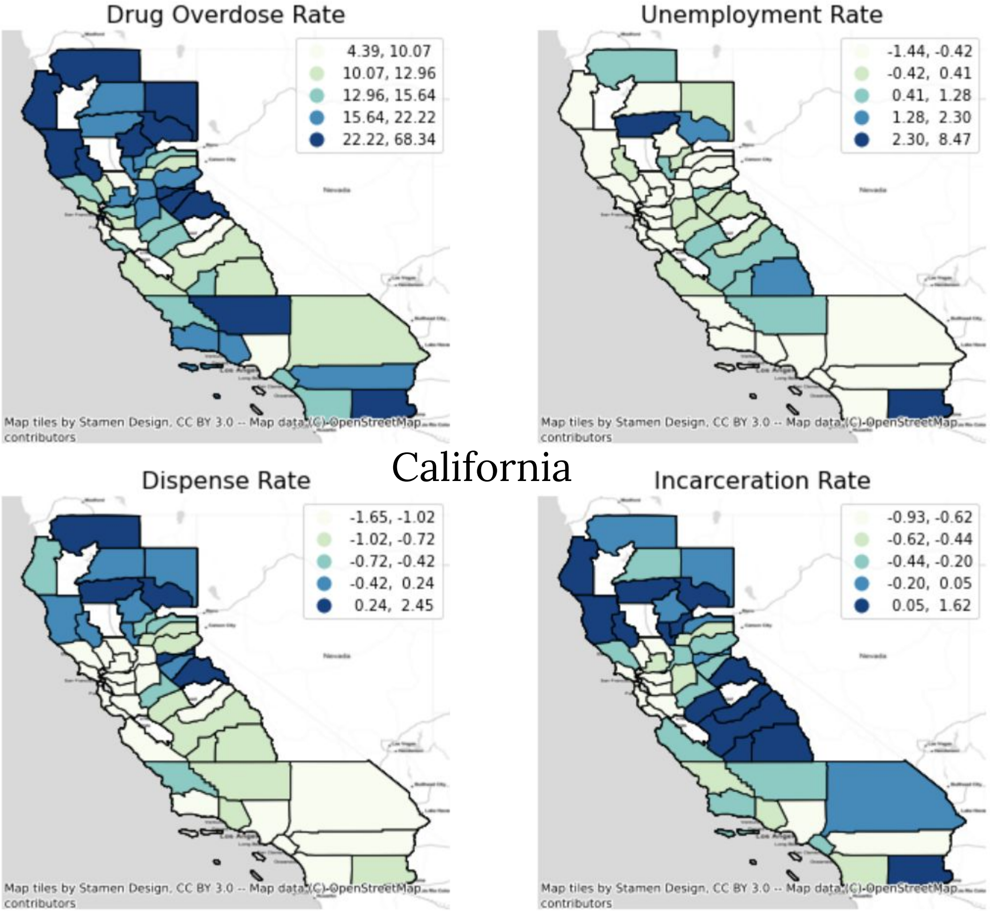
- AA_MALE, TOM_MALE, and Incarceration_Rate_per_100k are statistically insignificant

Introducing a Spatial Component



- Similar clustering patterns between certain factors
  - Ex: Lower regions of WV have higher rates of both overdose and incarceration
- **Takeaway**: There is a spatial component present in our data

Introducing a Spatial Component



- Similar to West Virginia, we have clusters in California
    - Ex: Northern California has higher Overdose, Incarceration, and Opioid Dispense Rates

Geospatial Factors: Global Moran's I

- We introduce Moran's I to determine if there is a spatial relationship in our data (for example, it measures how one county is similar to all others)

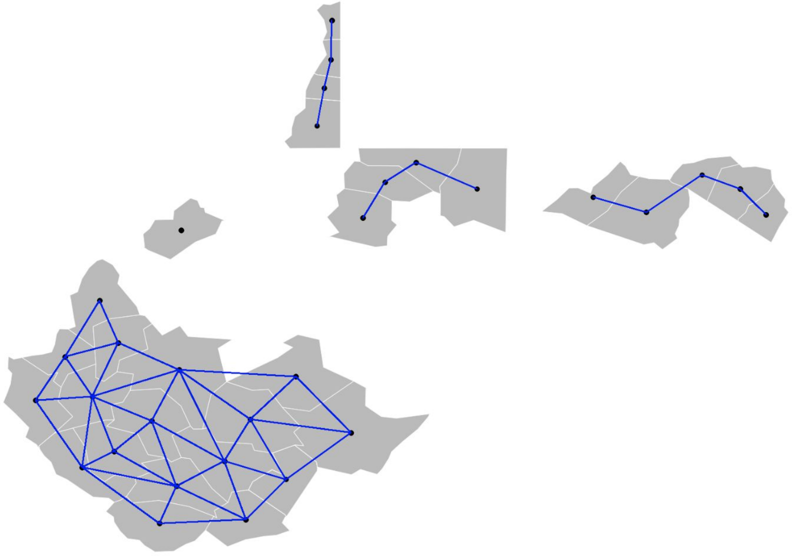
$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \sum_{i=1}^{n} z_i^2}$$

- $z_i = (x_i - \bar{x})$ is the deviation of an attribute (i.e. our overdose rate) from its mean for county i and $w_{i,j}$ is the spatial weight between county i and j, and n is the total number of counties

Moran's I

- In the previous equation, we used the queen weights (depicted on the right)
- Our Moran's I is 0.46085
  - This indicates a moderate positive spatial autocorrelation between the counties
- There is a spatial autocorrelation between the counties, so we concluded that it is important to have a spatial component in our model.

Ex. Queen Weights for West Virginia



$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous} \\ 0 & \text{if } i \text{ and } j \text{ are not contiguous} \end{cases}$$

OLS with Naive Spatial Components

Incorporating two new spatial component variables:

- **spatmax**: *Maximum* overdose rate of the counties that are adjacent to the focal county

- **spatmean**: *Average* overdose rate of the counties that are adjacent to the focal county

OLS Regression Outcomes

	Best Subset (BSS) OLS		BSS OLS w/ spatmax		BSS OLS w/ spatmean		BSS OLS w/ spatmax & spatmean	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Intercept	-223.7	0.0	-134.508	0.0	-116.857	0.0	-116.893	0.000
Year	0.112	0.0	0.068	0.0	0.059	0.0	0.059	0.000
Unemployment_rate	0.082	0.0	0.037	0.0	0.044	0.0	0.044	0.000
Dispense_rate	0.178	0.0	0.1	0.0	0.082	0.0	0.082	0.000
AA_MALE	0.012	0.353	-0.002	0.844	-0.02	0.065	-0.020	0.068
TOM_MALE	-0.032	0.106	-0.007	0.677	0.013	0.441	0.013	0.447
NH_MALE	0.163	0.0	0.126	0.0	0.122	0.0	0.122	0.000
Jail_Population	-0.043	0.009	-0.021	0.083	-0.016	0.177	-0.016	0.176
Incarceration_Rate_per_100k	0.008	0.18	0.024	0.0	0.024	0.0	0.024	0.000
PovertyCount	-0.15	0.0	-0.126	0.0	-0.121	0.0	-0.121	0.000
MedianHHI	-0.121	0.0	-0.129	0.0	-0.105	0.0	-0.106	0.000
spatmax	N/A	N/A	0.246	0.0	N/A	N/A	0.003	0.792
spatmean	N/A	N/A	N/A	N/A	0.277	0.0	0.274	0.000
	AIC	Adj R^2	AIC	Adj R^2	AIC	Adj R^2	AIC	Adj R^2
	9581	0.307	6750	0.492	6334	0.521	6336	0.521

- **Takeaways:**
  - Spatial components are able to explain more of the variability in our model
  - spatmean explains more variability in our model than spatmax does

Multicollinearity

feature	VIF
Year	68.784791
Unemployment_rate	8.954940
Dispense_rate	10.116277
AA_MALE	5.854949
TOM_MALE	14.211463
NH_MALE	14.546340
Jail Population	9.974360
Incarceration Rate per 100k	4.811989
PovertyCount	13.589621
MedianHHI	31.760316

- High Variance Inflation Factors
- Indication of multicollinearity present between our variables

In Progress...

- Adding new data to increase the number of predictors
    - Reducing multicollinearity
    - Decreasing the possibility of pre-selecting variables

Premature Deaths	Potential Years Lost	% Low Weight Births	% Smokers	% Adults Obsese	% Excessive Drinking	Vehicle Crash Death Rate	Teen Birth Rate	% Uninsured	PrimCarePhys per 100k	HS Grad Rate	% College Edu	% Children in Poverty	% 1PHH	Crir
68872.0	10189.200000	10.210000	23.600000	31.9	12.600000	25.600000	53.200000	17.100000	105.00000	67.0	55.100000	22.1	36.200000	431.
675.0	9967.400000	9.450000	27.400000	31.5	14.100000	28.300000	52.000000	15.500000	67.50000	75.0	55.400000	14.9	29.700000	256
2219.0	8321.800000	8.820000	21.900000	26.2	19.400000	23.200000	49.600000	20.900000	120.90000	70.0	61.500000	15.1	28.900000	194
403.0	9559.000000	11.350000	22.900000	37.6	8.500000	29.200000	79.900000	16.300000	57.00000	55.0	34.700000	31.9	52.500000	72
365.0	13282.900000	9.940000	33.000000	32.3	11.300000	42.500000	64.400000	19.900000	41.70000	60.0	40.300000	25.0	32.000000	164
...	...	...	...	...	...	...	...	...	...	...	...	...	...	

- Implementing elastic net to reduce multicollinearity and model complexity

Next Steps and Goals

- Incorporate the Queen's Weight Matrix using a weighted least squares model with elastic net
- Explore the idea of including interaction terms in the model
- Examine the temporal aspect of our data
- Determine which features are able to best explain overdose rates
- Estimate missing overdose rates for the counties that are missing from our data