Ashlyn Jew
STAT 222

Drug Overdoses in the United States

1 Problem Overview

1.1 Motivation

The drug epidemic in the United States has been a growing concern in the past years. We are currently at the peak of drug overdoses where the number of deaths caused by drug poisoning tower over the number deaths by car accidents, guns, and HIV [1]. To understand this devastating and widespread problem, we will analyze the potential factors contributing to the drug epidemic.

1.2 Problem Statement

Through this project, we aim to answer the following questions: (1) How do different aspects of counties (e.g. location, demographics, opioid dispense rates, etc.) relate to drug overdose rates?, and (2) Can we utilize the various demographic trends of each county to more accurately estimate and explain the drug overdose rates?

2 Data Description

2.1 Original Data

The county-level drug overdose mortality data was extracted from the Multiple Cause of Death (Final) Database on CDC Wonder. The database contains data collected by the National Center for Health Statistics and is based on information on death certificates filed in the United States [2]. We filtered for data with death codes defined by the CDC as death by drug poisoning.

Additional data on opioid dispensing rates [3], ethnicity [4], unemployment rates [5], incarceration rates [6], poverty rates, and median household income [7] was retrieved by our team. We also added geospatial variables to our data: average overdose death rates of adjacent counties, maximum overdose death rates of adjacent counties, and geometry data from county shapefiles.

2.2 Data Description

For this report, we are focusing on data from 2010 to 2019 since our data is more complete for those years (Figure 1). We have 8835 observations and 41 variables in our 2010-2019 dataset. Each observation represents a county in a certain year. Each variable is also at a county-level granularity.

Our dataset is missing data for many counties each year. However, the counties present in the data represent over 50% of the total U.S. population. We also aim to fill in the missing data with our explanatory model.

2.3 Variables of Interest

Since all the data and variables retrieved were self-curated, all variables are considered important potential factors contributing to the overdose death rates, albeit we expect our model to tell us which variables are more important than others. Table 1 details a list of our variables of interest and their hypothesized relationship to overdose death rates.

In our initial EDA, we have noticed that there were clusters of higher overdose death rates in certain areas of the U.S. In Figure 2 below, West Virginia and the surrounding Appalachian region seem to have a cluster of higher overdose death rates. Thus, we hypothesize that there are spatial autocorrelations and spillover effects between counties. So we have included spatmax (maximum drug overdose death rate of adjacent counties) and spatmean (average drug overdose death rate of adjacent counties) as variables of interest.

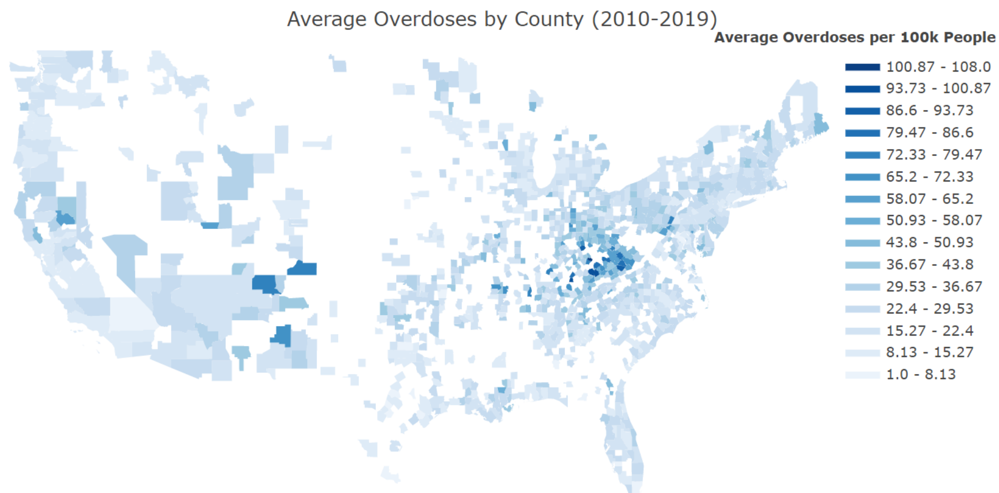


Figure 2: Average Overdoses per 100k People by County (2010 - 2019)

3 Methods

3.1 Geospatial Components

To support our idea of adding spatial components to our model, we calculated Moran's I to determine if there was a spatial autocorrelation between the U.S. counties. Moran's I is computed as follows:

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

where $z_i = (x_i - \bar{x})$ is the deviation of an attribute (i.e. our variables) for county i from its mean and $w_{i,j}$ is the spatial weight between county i and j, and n is the total number of counties [8].

The spatial weight was calculated using the notion of Queen contiguity, which determines weights depending on if they share a common vertex, and hence considered neighboring [8]. The Moran's I we computed was approximately 0.46085, which indicates that there is a positive spatial autocorrelation. Thus, we have evidence that there is some sort of spatial clustering of our variables.

Next, we look at the local Moran's I for West Virginia, which is one of the states with the most obvious clustering of drug overdose deaths.

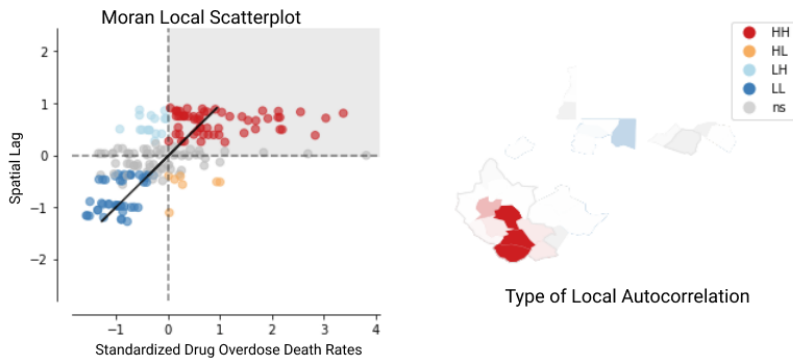


Figure 3: Moran Local Plots for West Virginia

In the Moran Local Scatterplot (Figure 3 left plot), the upper right and lower left quadrants indicate a positive spatial autocorrelation (i.e. cluster of counties) for drug overdose death rates. The lower right and upper left quadrants indicate negative spatial autocorrelation (i.e. counties are different from their neighbors) for drug overdose death rates. The slope indicates the local Moran's I for West Virginia. Since we see more points in the upper right and lower left quadrants, there is likely stronger positive autocorrelation in West Virginia. In the right plot, the red indicates High-High local autocorrelation, where a higher drug overdose death rate has neighbors that also have higher drug overdose death rates.

With this spatial autocorrelation analysis, we see that there is evidence of positive spatial correlation of drug overdose rates in our counties. Thus, we will be including spatial components within our model.

3.2 Statistical Model / Machine Learning Model

To begin, we used ordinary least squares regression as our baseline model. We performed exhaustive feature selection with 19 variables: 'Year', 'Population', 'Unemployment_rate', 'Dispense_rate', 'TOT_MALE', 'WA_MALE', 'BA_MALE', 'IA_MALE', 'AA_MALE', 'NA_MALE', 'TOM_MALE', 'NH_MALE', 'H_MALE', 'Urbanicity', 'Jail_Population', 'Incarceration_Rate_per 100k', 'PovertyCount', 'PovertyPercentage', and 'MedianHHI'. We did not include the 'XX_FEMALE' counterparts of the demographics variable since previous analysis showed that the distribution of the 'XX_MALE' and 'XX_FEMALE' variables were very similar across all races. We used mean squared error and 5-fold cross-validation as our hyperparameters in our exhaustive feature selection.

The exhaustive feature selection output 'Year', 'Unemployment_rate', 'Dispense_rate', 'AA_MALE', 'TOM_MALE', 'NH_MALE', 'Jail_Population', 'Incarceration_Rate_per_100k', 'PovertyCount', and 'MedianHHI' as the best subset of features to predict drug overdose death rates ('Cruder_Rate'). This set of variables are henceforth referred to as "BSS". Table 2 provides the five-number summary statistics as well as the variance inflation factor (VIF) of the variables chosen by the exhaustive feature selection. We recognize that the VIF of certain variables are quite high (> 10), indicating strong multicollinearity. In future models we will use methods such as principal component analysis or ridge regression to reduce the multicollinearity among our variables.

We fit OLS models with the following formula:

$$y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \epsilon_i$$

where y_i is the drug overdose death rate at county i, β_{i0} is the intercept coefficient at county i, x_{ik} is the k-th explanatory variable at county i, β_{ik} is the k-th local regression coefficient for the k-th explanatory variable at county i, ϵ_i is the random error term associated with county i, and i goes from 1 to n (number of observed counties) [9].

We fit four OLS regression models with drug overdose death rate ('Cruder_Rate') as the dependent variable: (1) Cruder_Rate ~ BSS, (2) Cruder_Rate ~ BSS + spatmax, (3) Cruder_Rate ~ BSS + spatmean, and (4) Cruder_Rate ~ BSS + spatmax + spatmean. Each explanatory variable was standardized before the regression was run.

4 Results

4.1 Model Output

The summary output of the four OLS models is displayed in Table 3:

	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	-367.99	0.0	-2593.707	0.0	-2080.257	0.0	-2083.644	0.000
Year	0.183	0.0	1.299	0.0	1.044	0.0	1.045	0.000
Unemployment_rate	0.152	0.0	0.958	0.0	1.161	0.0	1.148	0.000
Dispense_rate	0.299	0.0	2.066	0.0	1.56	0.0	1.566	0.000
AA_MALE	0.098	0.0	0.911	0.001	0.387	0.157	0.403	0.142
TOM_MALE	-0.146	0.0	-1.12	0.011	-0.532	0.208	-0.547	0.196
NH_MALE	0.222	0.0	2.034	0.0	1.941	0.0	1.938	0.000
Jail_Population	-0.1	0.001	-0.708	0.025	-0.556	0.068	-0.559	0.066
Incarceration_Rate_per_100k	0.023	0.044	0.712	0.0	0.73	0.0	0.731	0.000
PovertyCount	-0.143	0.0	-1.503	0.0	-1.357	0.0	-1.359	0.000
MedianHHI	-0.174	0.0	-2.521	0.0	-1.81	0.0	-1.834	0.000
spatmax	N/A	N/A	7.351	0.0	N/A	N/A	0.315	0.330
spatmean	N/A	N/A	N/A	N/A	8.237	0.0	7.935	0.000

Table 3: OLS Coefficient Estimates and P-values

Tables 4-7 display the detailed OLS output results.

4.2 Model Interpretation

For Model 1 - Best Subset (BSS) OLS, all variables are statistically significant (at a 0.05 significance level). This may be due to our large number of observations (7591 non-null observations vs. 10 features). Notably, Jail_Population and PovertyCount have a negative coefficient, contrasting with the hypothesized relationship we stated in Table 1. We expected an increase in the number of people in jail or an increase in the number of people in poverty to increase the number of drug overdose death rates, but the coefficients propose otherwise. This model had the lowest R-squared (0.247) but the lowest AIC value (1.910e+04).

For Model 2 - BSS OLS with spatmax, all variables are also statistically significant, including spatmax. This may also be due to the large number of observations compared to features (8835 non-null observations vs 11 features). Notably, spatmax has the greatest positive coefficient value and suggests a stronger correlation with the overdose death rates. The R-squared value (0.490) is nearly twice of Model 1's, but the AIC value is the highest out of all the models (5.258e+04).

For Model 3 - BSS OLS with spatmax, all variables are significant except for AA_MALE (% Asian Male) and TOM_MALE (% Multiracial Male). Remarkably, spatmean has the greatest positive coefficient value that is also larger than the spatmax coefficient in Model 2. This suggests that the presence of spatmean causes AA_MALE and TOM_MALE to lose explanatory power in the model. The R-squared value (0.527) is slightly higher than Model 2's, but the AIC value (5.205e+04) is comparatively higher than Model 1.

For Model 4 - BSS OLS with spatmax and spatmean, AA_MALE, TOM_MALE, and spatmax are not statistically significant variables. spatmean also has the greatest coefficient value in this model, suggesting that it may have the highest explanatory power among the variables included in the model. With spatmean in the model, spatmax is no longer significant, demonstrating that the mean drug overdose death rates of adjacent counties is a better indicator of drug overdose death rates than the maximum drug overdose death rates of adjacent counties. The R-squared value and AIC value are the same as Model 3's, indicating that the addition of spatmax does not influence model performance significantly.

We also observe that the signs of all variables do not change across all models. We also realize that OLS may not be the best-fitting model, but it serves as a baseline for comparison with our future models. The results demonstrate that the variables chosen deserve to be investigated further and included in our future models again. With spatmax and spatmean having the greatest coefficient values in Models 2-4, we have reason to believe that including more sophisticated spatial components in our future models will prove to result in a better fit.

5 Conclusions

5.1 Summary

Through our preliminary analysis and modeling, we have determined that spatial components have the best explanatory power among the different features we selected as potential factors contributing to the drug overdose death rates in the United States.

5.2 Further Steps

Recognizing that our preliminary modeling has not been adequately motivated and is mostly naive, we aim to improve our model and look at more sophisticated model performance metrics. Since we have seen the significance of spatial components, we plan to use geographically weighted regression with our Queen contiguity weight matrix. Moreover, since our data also deals with different years, and the 'Year' variable is a significant variable in our model, we will also look into implementing spatial-temporal modeling.

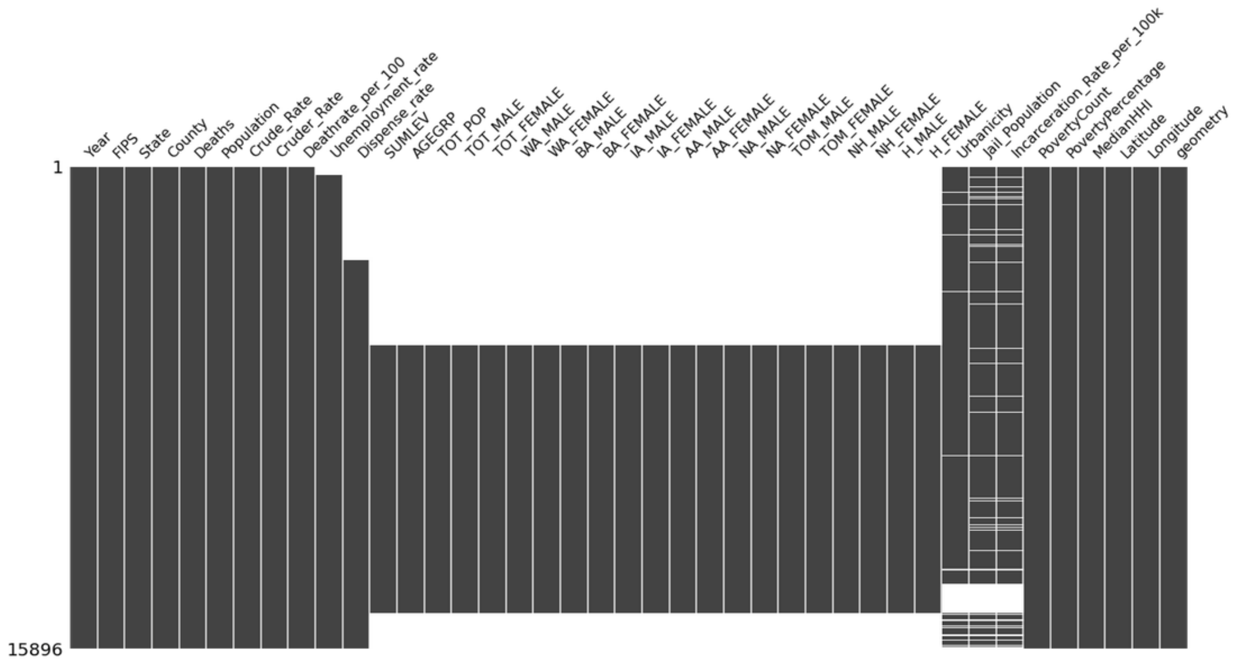
Additionally, we are planning to add more covariates, since most of the variables in our model seem to be significant. With more covariates, we may be able to remove variables with multicollinearity. Data transformations will also be implemented to correct the skewness of our variables.

Furthermore, we will use less naive model metrics, as we understand that R-squared is a metric that will improve as we add more variables. Carrying out cross-validation and train-validation-test splits will allow us to have a clearer view of model performance. Plotting residuals and analyzing outliers will also help us in knowing how to improve our models as well.

Overall, there are many directions we can take to improve our model to answer our questions. With future model refinement, we hope to have more concrete evidence on which variables are the most correlated with the rising overdose death rates in the U.S. However, our current findings motivate our hypothesis that spatial factors are an important contributor to the drug overdose death rates.

Appendix

Figure 1. Completeness of data.



Black indicates data that is present and white indicates missing data.

Table 1. Variables of Interest and Hypothesized Relationship to Drug Overdose Death Rates

Variable Name	**Variable Description** **(all variables are at county-level)**	**Hypothesized relationship to drug overdose death rates**
Cruder_Rate	Number of drug overdose deaths per 100,000 persons	
Dispense_rate	Retail opioid prescriptions dispensed per 100 persons	Dispense_rate *increase* ⟶ Cruder_Rate *increase*
Unemployment_rate	Unemployment rate	Unemployment_rate *increase* ⟶ Cruder_Rate *increase*
PovertyCount	Estimate of people of all ages in poverty	PovertyCount *increase* ⟶ Cruder_Rate *increase*
PovertyPercentage	Estimated percent of people of all ages in poverty	PovertyPercentage *increase* ⟶ Cruder_Rate *increase*

MedianHHI	Estimate of median household income	MedianHHI *increase* → Cruder_Rate *decrease*
WA_MALE	Percentage of white males	WA_MALE *increase* → Cruder_Rate *increase*
Incarceration_Rate_per_100k	Incarceration rate per 100,000 persons	Incarceration_Rate_per_100k *increase* → Cruder_Rate *increase*
Jail_Population	Jail population count	Jail_Population *increase* → Cruder_Rate *increase*
Urbanicity	Urbanicity (rural, small/mid, suburban, urban)	Urbanicity *increase* → Cruder_Rate *increase*
Population	Population Count	Population *increase* → Cruder_Rate *increase*
Year	Year of record	Year *increase* → Cruder_Rate *increase*
spatmax	Maximum Cruder_Rate of adjacent counties	spatmax *increase* → Cruder_Rate *increase*
spatmean	Mean Cruder_Rate of adjacent counties	spatmean *increase* → Cruder_Rate *increase*

** Demographic variables also include AA_MALE (Asian), BA_MALE (Black), IA (American Indian/Alaska Native), NA_MALE (Native Hawaiian/Pacific Islander), TOM_MALE (Multiracial). There are also XX_FEMALE demographic variables for the same races.

Table 2. Five Number Summary Statistics and VIF of Best Subset (BSS)

	count	mean	std	min	25%	50%	75%	max	VIF
Year	7591.0	2014.218285	2.585592	2010.00	2012.000	2014.00	2016.000	2018.00	68.784791
Unemployment_rate	7591.0	6.542590	2.725402	2.00	4.500	6.00	8.000	29.40	8.954940
Dispense_rate	7591.0	90.727335	40.130801	9.90	64.000	83.90	109.300	426.40	10.116277
AA_MALE	7591.0	8407.628903	34112.541448	5.00	296.000	1008.00	4469.000	720458.00	5.854949
TOM_MALE	7591.0	3730.123040	8132.088403	36.00	699.000	1474.00	3502.000	154085.00	14.211463
NH_MALE	7591.0	117908.745752	179201.669693	2956.00	33490.500	61169.00	127541.500	2539478.00	14.546340
Jail Population	7591.0	683.893442	1178.096923	3.00	185.000	347.00	728.500	19091.94	9.974360
Incarceration Rate per 100k	7591.0	423.926483	278.985128	4.25	251.035	358.88	522.285	4265.42	4.811989
PovertyCount	7591.0	42216.501910	91349.717261	1527.00	10491.500	18623.00	40665.500	1873522.00	13.589621
MedianHHI	7591.0	53625.945330	14747.719159	22289.00	43426.000	50612.00	60041.500	140382.00	31.760316

Table 4. Detailed OLS Output of Model 1 - Best Subset (BSS) OLS

```
==============================================================================
Dep. Variable:            Cruder_Rate   R-squared:                       0.247
Model:                            OLS   Adj. R-squared:                  0.246
Method:                 Least Squares   F-statistic:                     249.0
Date:                Sat, 19 Mar 2022   Prob (F-statistic):               0.00
Time:                        00:09:05   Log-Likelihood:                -9537.4
No. Observations:                7591   AIC:                         1.910e+04
Df Residuals:                    7580   BIC:                         1.917e+04
Df Model:                          10
Covariance Type:            nonrobust
=================================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
Intercept                  -367.9896     11.136    -33.045      0.000    -389.819    -346.160
Year                          0.1827      0.006     33.057      0.000       0.172       0.194
Unemployment_rate             0.1521      0.014     10.784      0.000       0.124       0.180
Dispense_rate                 0.2985      0.014     21.930      0.000       0.272       0.325
AA_MALE                       0.0982      0.024      4.073      0.000       0.051       0.145
TOM_MALE                     -0.1459      0.037     -3.920      0.000      -0.219      -0.073
NH_MALE                       0.2223      0.032      7.024      0.000       0.160       0.284
Jail_Population              -0.0997      0.031     -3.231      0.001      -0.160      -0.039
Incarceration_Rate_per_100k   0.0231      0.011      2.017      0.044       0.001       0.046
PovertyCount                 -0.1429      0.033     -4.313      0.000      -0.208      -0.078
MedianHHI                    -0.1742      0.015    -11.574      0.000      -0.204      -0.145
==============================================================================
Omnibus:                     3866.583   Durbin-Watson:                   1.362
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            40325.914
Skew:                           2.211   Prob(JB):                         0.00
Kurtosis:                      13.390   Cond. No.                     2.30e+06
==============================================================================
```

Table 5. Detailed OLS Output of Model 2 - Best Subset (BSS) OLS w/ spatmax

```
==============================================================================
Dep. Variable:            Cruder_Rate   R-squared:                       0.490
Model:                            OLS   Adj. R-squared:                  0.489
Method:                 Least Squares   F-statistic:                     616.4
Date:                Sat, 19 Mar 2022   Prob (F-statistic):               0.00
Time:                        00:12:05   Log-Likelihood:                -26279.
No. Observations:                7066   AIC:                         5.258e+04
Df Residuals:                    7054   BIC:                         5.266e+04
Df Model:                          11
Covariance Type:            nonrobust
=================================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
Intercept                 -2593.7068    144.006    -18.011      0.000   -2876.003   -2311.411
Year                          1.2987      0.071     18.167      0.000       1.159       1.439
Unemployment_rate             0.9582      0.184      5.195      0.000       0.597       1.320
Dispense_rate                 2.0658      0.164     12.587      0.000       1.744       2.388
AA_MALE                       0.9110      0.284      3.209      0.001       0.354       1.468
TOM_MALE                     -1.1197      0.439     -2.553      0.011      -1.979      -0.260
NH_MALE                       2.0340      0.375      5.423      0.000       1.299       2.769
Jail_Population              -0.7084      0.316     -2.243      0.025      -1.328      -0.089
Incarceration_Rate_per_100k   0.7123      0.145      4.902      0.000       0.427       0.997
PovertyCount                 -1.5026      0.388     -3.868      0.000      -2.264      -0.741
MedianHHI                    -2.5208      0.183    -13.809      0.000      -2.879      -2.163
spatmax                       7.3515      0.130     56.565      0.000       7.097       7.606
==============================================================================
Omnibus:                     2928.886   Durbin-Watson:                   1.883
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            28098.756
Skew:                           1.723   Prob(JB):                         0.00
Kurtosis:                      12.141   Cond. No.                     2.44e+06
==============================================================================
```

Table 6. Detailed OLS Output of Model 3 - Best Subset (BSS) OLS w/ spatmean

Dep. Variable:	Cruder_Rate	R-squared:	0.527
Model:	OLS	Adj. R-squared:	0.527
Method:	Least Squares	F-statistic:	715.3
Date:	Sat, 19 Mar 2022	Prob (F-statistic):	0.00
Time:	00:14:30	Log-Likelihood:	-26011.
No. Observations:	7066	AIC:	5.205e+04
Df Residuals:	7054	BIC:	5.213e+04
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2080.2567	140.294	-14.828	0.000	-2355.275	-1805.239
Year	1.0438	0.070	14.989	0.000	0.907	1.180
Unemployment_rate	1.1610	0.177	6.549	0.000	0.813	1.509
Dispense_rate	1.5605	0.159	9.795	0.000	1.248	1.873
AA_MALE	0.3874	0.274	1.415	0.157	-0.149	0.924
TOM_MALE	-0.5324	0.423	-1.259	0.208	-1.361	0.296
NH_MALE	1.9405	0.361	5.373	0.000	1.233	2.649
Jail_Population	-0.5558	0.304	-1.827	0.068	-1.152	0.040
Incarceration_Rate_per_100k	0.7304	0.140	5.220	0.000	0.456	1.005
PovertyCount	-1.3567	0.374	-3.627	0.000	-2.090	-0.623
MedianHHI	-1.8099	0.176	-10.262	0.000	-2.156	-1.464
spatmean	8.2374	0.130	63.291	0.000	7.982	8.493

Omnibus:	2958.626	Durbin-Watson:	1.964
Prob(Omnibus):	0.000	Jarque-Bera (JB):	31758.368
Skew:	1.707	Prob(JB):	0.00
Kurtosis:	12.809	Cond. No.	2.47e+06

Table 7. Detailed OLS Output of Model 4 - Best Subset (BSS) OLS w/ spatmax and spatmean

Dep. Variable:	Cruder_Rate	R-squared:	0.527
Model:	OLS	Adj. R-squared:	0.527
Method:	Least Squares	F-statistic:	655.8
Date:	Sat, 19 Mar 2022	Prob (F-statistic):	0.00
Time:	00:15:24	Log-Likelihood:	-26011.
No. Observations:	7066	AIC:	5.205e+04
Df Residuals:	7053	BIC:	5.214e+04
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2083.6440	140.337	-14.847	0.000	-2358.748	-1808.540
Year	1.0455	0.070	15.008	0.000	0.909	1.182
Unemployment_rate	1.1482	0.178	6.459	0.000	0.800	1.497
Dispense_rate	1.5663	0.159	9.824	0.000	1.254	1.879
AA_MALE	0.4029	0.274	1.469	0.142	-0.135	0.940
TOM_MALE	-0.5466	0.423	-1.292	0.196	-1.376	0.283
NH_MALE	1.9376	0.361	5.365	0.000	1.230	2.646
Jail_Population	-0.5593	0.304	-1.839	0.066	-1.156	0.037
Incarceration_Rate_per_100k	0.7311	0.140	5.225	0.000	0.457	1.005
PovertyCount	-1.3592	0.374	-3.633	0.000	-2.092	-0.626
MedianHHI	-1.8345	0.178	-10.296	0.000	-2.184	-1.485
spatmax	0.3153	0.324	0.974	0.330	-0.319	0.950
spatmean	7.9350	0.337	23.570	0.000	7.275	8.595

Omnibus:	2956.358	Durbin-Watson:	1.964
Prob(Omnibus):	0.000	Jarque-Bera (JB):	31567.161
Skew:	1.707	Prob(JB):	0.00
Kurtosis:	12.775	Cond. No.	2.47e+06

References

1. Katz, J. (2017, April 14). You draw it: Just how bad is the drug overdose epidemic? The New York Times. Retrieved March 18, 2022, from https://www.nytimes.com/interactive/2017/04/14/upshot/drug-overdose-epidemic-you-draw-it.html
2. Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death, 1999-2020 on CDC WONDER Online Database, released in 2021. Data are from the Multiple Cause of Death Files, 1999-2020, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at http://wonder.cdc.gov/mcd-icd10.html on Feb 16, 2022 1:22:58 AM
3. Centers for Disease Control and Prevention. (2021, November 10). U.S. opioid dispensing rate maps. Centers for Disease Control and Prevention. Retrieved March 18, 2022, from https://www.cdc.gov/drugoverdose/rxrate-maps/index.html
4. U.S. Census Bureau. (2021, October 8). County population by characteristics: 2010-2019. Census.gov. Retrieved March 18, 2022, from https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html
5. U.S. Bureau of Labor Statistics. (n.d.). Local Area Unemployment Statistics. U.S. Bureau of Labor Statistics. Retrieved March 18, 2022, from https://www.bls.gov/lau/#tables
6. Institute, V. (n.d.). Vera-Institute/Incarceration-Trends: Incarceration trends dataset and Documentation. GitHub. Retrieved March 18, 2022, from https://github.com/vera-institute/incarceration-trends
7. U.S. Census Bureau. (2021, October 8). Small Area Income and Poverty Estimates (SAIPE) Program Datasets. Census.gov. Retrieved March 18, 2022, from https://www.census.gov/programs-surveys/saipe/data/datasets.html
8. Rey, S. J., Arribas-Bel,, D., & Wolf, L. J. (2020). Geographic Data Science with python. Global Spatial Autocorrelation - Geographic Data Science with Python. Retrieved March 20, 2022, from https://geographicdata.science/book/notebooks/06_spatial_autocorrelation.html?highlight=moran+s+i#continuous-case-moran-plot-and-moran-s-i, https://geographicdata.science/book/notebooks/04_spatial_weights.html#contiguity-weights
9. Oshan, T., Li, Z., Kang, W., Wolf, L., & Fotheringham, A. (2019). mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. ISPRS International Journal of Geo-Information, 8(6), 269. https://doi.org/10.3390/ijgi8060269