State University of New York at Oswego

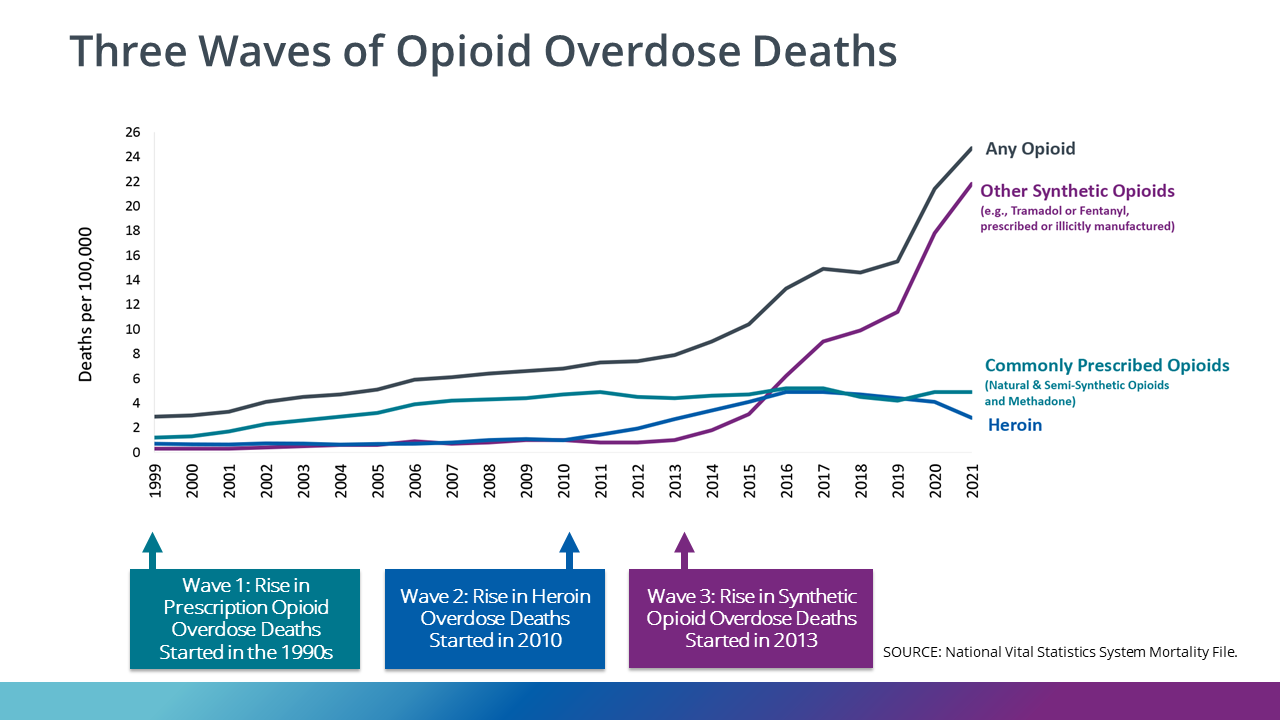
Department of Biomedical and Health Informatics

*Using Geographic Information System (GIS) to Assess and Identify Factors Leading to the Increasing Trend in Drug Overdose Mortality in New York State*

Gerald Gaitos

**Background:**

The history of the opioid epidemic phenomenon started in the 1990s with good intentions gone bad when. Initially, physicians were not aware of the addictive effects of drugs, combined with the greedy interests of pharmaceutical companies and drug cartels encouraging unsolicited opioid and opioid derivative use to the public. These factors have formed gigantic waves of addiction and deaths that are continuously increasing until today (deShazo et al., 2018). Affected individuals predominantly suffer from mental health issues, traumas, stresses, chronic diseases, and chronic pain (CDC, n. d. & Azadfard et al., 2023). These individuals need to be understood, and people need to realize that addiction is a disease and it can be cured.



**Figure 1.** Three waves of opioid overdose deaths from the 1990s to the present.

Figure 1 shows that in past decades, there have been three waves of opioid overdose mortality. Wave 1 started in 1990 when physicians were not aware of the side effects of prescription opioids. People medicated with opioids started to become addicted and dependent on the medication. Wave 2 was initiated in 2010, when people were introduced to a new type of illicit drug, namely, heroin. The number of cases of drug overdose mortality kept constantly increasing through the years until Wave 3 happened, when new synthetic opioid drugs were created and manufactured, a steep rise in the number of cases of drug overdose mortality was observed (CDC, n. d.).

The US Health and Human Services have provided strategies to fight the opioid epidemic. The strategies mainly involve improving the accessibility to health services programs assisting in the prevention, treatment, recovery, and pain management, improving the reporting and recording of public health data, supporting research to deepen the understanding of addiction and overdose, and many more (Texas Health and Human Services, n. d.). These strategies can be implemented and fortified by creating a safe online space and community primarily for individuals suffering from substance use disorder.

This project intends to understand the factors, such as socioeconomic status, demographics, etc, that contribute to the increasing trend of mortality cases using data analytics and visualization tools. Getting a better understanding of the current situation of drug overdose mortality cases will help people understand the need for policy implementation to reduce the number of mortality cases of drug overdose in New York State.

**Objective:**

The project’s main objective is to apply data analytics and visualization tools, specifically GIS, to assess the relationship between different factors and variables and the trend of drug overdose mortality rates across other states and generate an interactive cartographic representation of the data information. Insights obtained from this project will help create policies and infrastructure to reduce the increasing mortality due to drug overdose.

**Specific Aims:**

The project applies GIS to multiple datasets containing data on drug overdose mortality rates, socioeconomic factors, etc., assessing the associations between different variables and drug overdose mortality rates. This project utilizes multiple datasets from various sectors, the US Department of Agriculture Economic Research Service and the New York State. The project aims to answer the following questions:

1. Is there an association between socioeconomic factors (number of households, head of household, median household income, state of residence, etc.) and drug overdose cases? If so, which factors significantly contribute to this problem?
2. Which state/county depicts the highest/lowest count of drug overdose mortality?
3. Do the counties with the highest and lowest counts of drug overdose have similar contributing factors to this problem?
4. Is it possible to predict drug overdose death cases spatially?

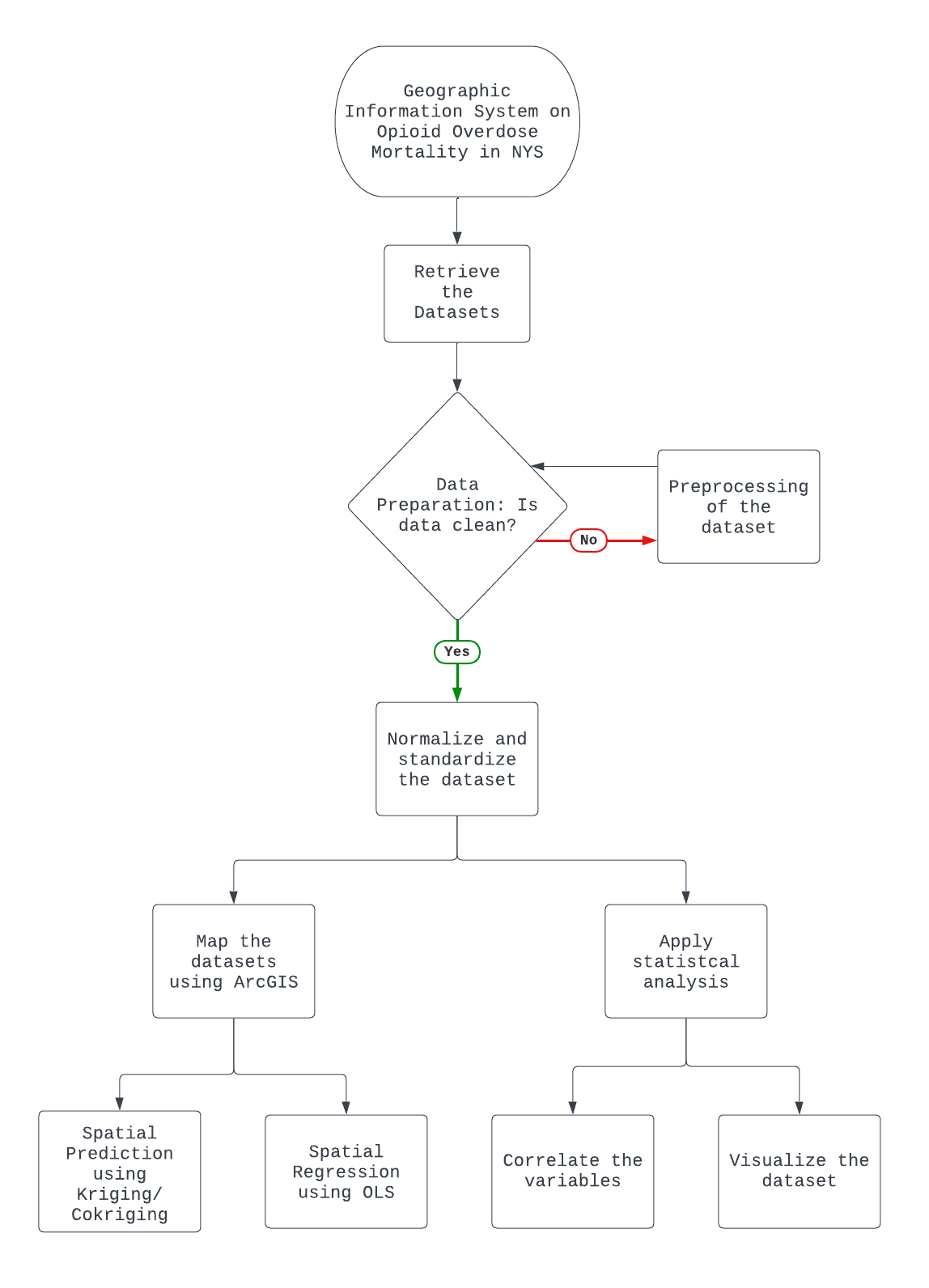
**Materials and Methods:**

The project will apply GIS for mapping and Python for other statistical analysis and data visualization. The datasets will undergo data processing to eliminate noise, inconsistencies, and missing data values, improving the result’s accuracy and reliability. Most of the datasets will be formatted appropriately to function well in the ArcGIS software.

For aim #1, statistical analysis using Python or spatial regression using ArcGIS software will determine the association among variables and the most used illicit drug among various counties.

For aims #2, #3, and #4, mapping and geospatial prediction will be done using ArcGIS to visualize the number of drug overdose mortality cases across various counties.

**Experimental Setup:**

****

**Figure 2.** The flowchart of the data analytics and visualization.

Figure 2 shows the flowchart of the data analytics and visualization applied to the project. The appropriate datasets were retrieved from different credible sources online, namely, the US Department of Agriculture Economic Research Service and the New York State. After data acquisition, all the datasets underwent data wrangling before applying different statistical analyses. The resulting merged dataset was normalized and standardized. The final dataset had 62 rows representing the counties and seven variables representing the dependent variable (Event Count/Rate) and the independent variables (Population, Poverty Rate, Unemployment Rate, Median Household Income, HS Rate, and No HS Rate). After cleaning the dataset, statistical analysis in Python and mapping in ArcGIS were performed.

**Data:**

The project will utilize different datasets from the US Department of Agriculture Economic Research Service and the New York State. The datasets from the USDA Economic Research Services will include information about the demographics and socioeconomic status of the general population in New York State in the years 2017, 2021, and 2022.

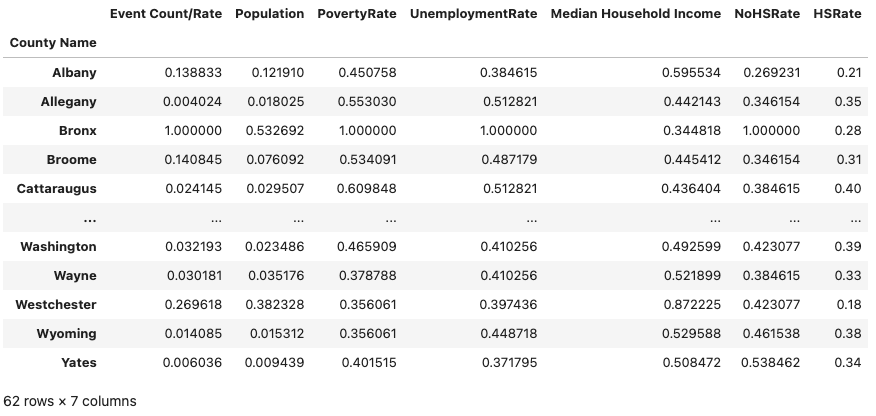
New York State Datasets:

1. New York State Opioid Data Dashboard, which can be accessed from <https://apps.health.ny.gov/public/tabvis/PHIG_Public/opioid/#dataexport>
2. Health Facility Map, which can be accessed from <https://health.data.ny.gov/Health/Health-Facility-Map/875v-tpc8>

USDA Economic Research Service Dataset:

1. County-level datasets (Poverty, Unemployment, and Median Housing Income and Education) have 62 rows and seven columns. These datasets can be accessed and downloaded in an Excel format from<https://www.ers.usda.gov/data-products/county-level-data-sets/>.

**Data Preprocessing/Normalization and Standardization**

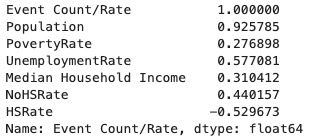


**Figure 3.** The merged dataset after data preprocessing, normalization, and standardization.

Figure 3 shows the merged dataset used in the project. The data values were normalized and standardized to ensure that all independent variables would influence the dependent variable equally. The data values ranged from 0 to 1.

**Results and Discussion:**

**Statistical Analysis using Python**

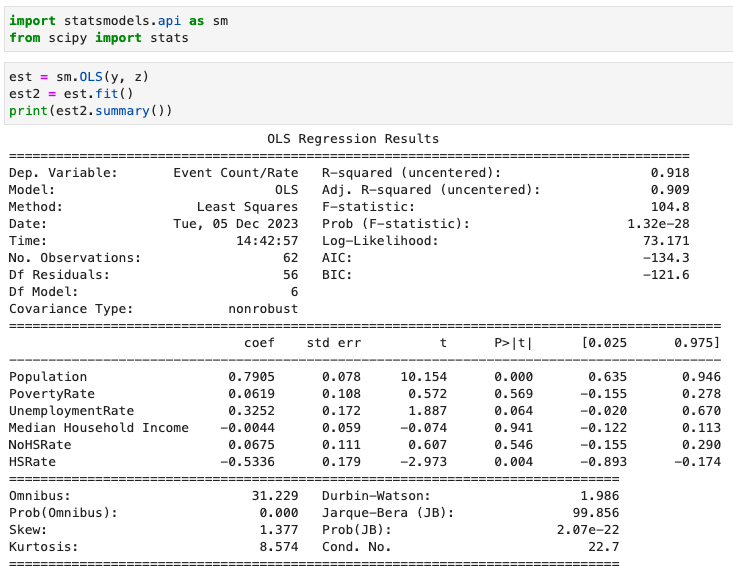


**Figure 4**. Heatmap of the dataset.

A heatmap was performed to understand the correlation among variables, as shown in Figure 4. Among the variables, Population showed the highest correlation (positive correlation) to the dependent variable, while HS Rate showed the least correlation (negative correlation). No multicollinearity was observed among variables.

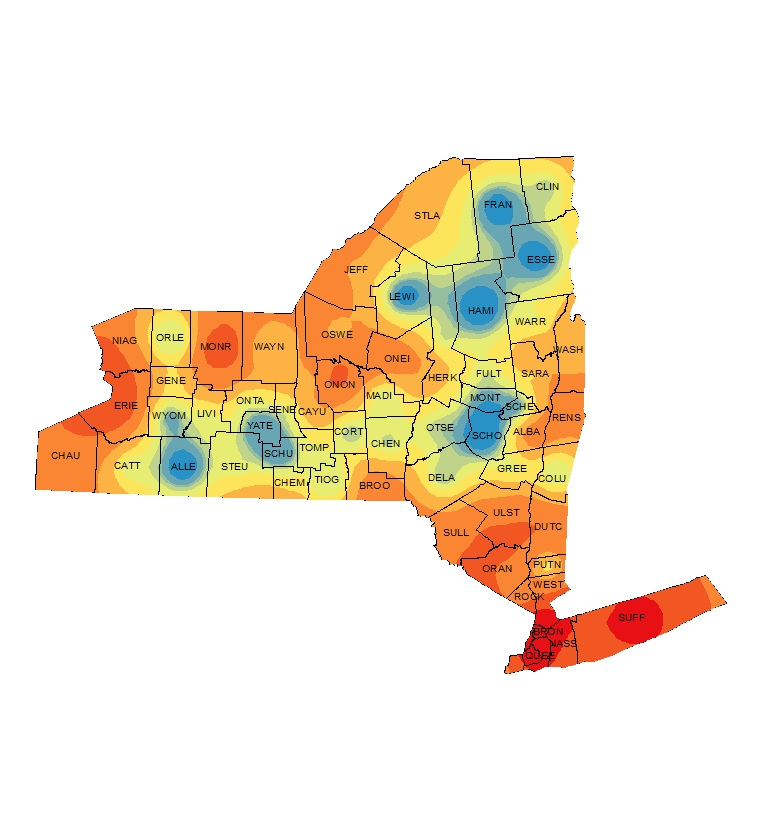
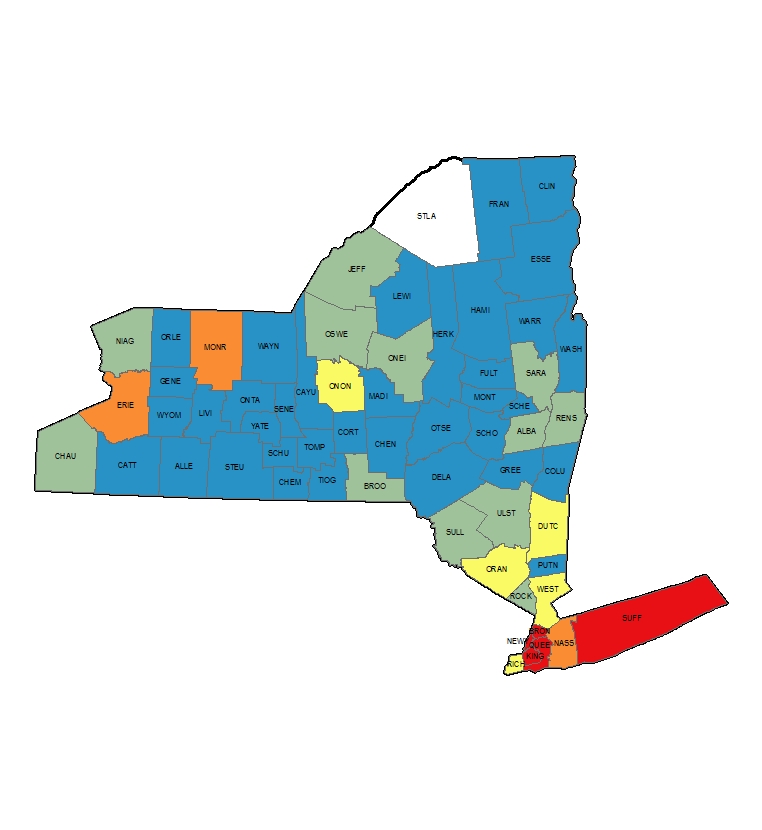
Multiple Linear Regression (MLR) was applied to see if a model could be plotted using these variables. The resulting values were interpreted into a formula with an R-value of 0.888, as shown in Figure 5. Since the significance of each variable could not be determined using MLR, Ordinary Least Square (OLS) Regression was used. The result, as shown in Figure 6, the R-value was 0.909. Among the variables, Population and HS Rate significantly contributed to the Event Count/Rate. AICc value of -134.3 represented a good model fit. The results from this analysis were then compared to the results from the ArcGIS analysis.

**Figure 5.** Multiple Linear Regression Equation.



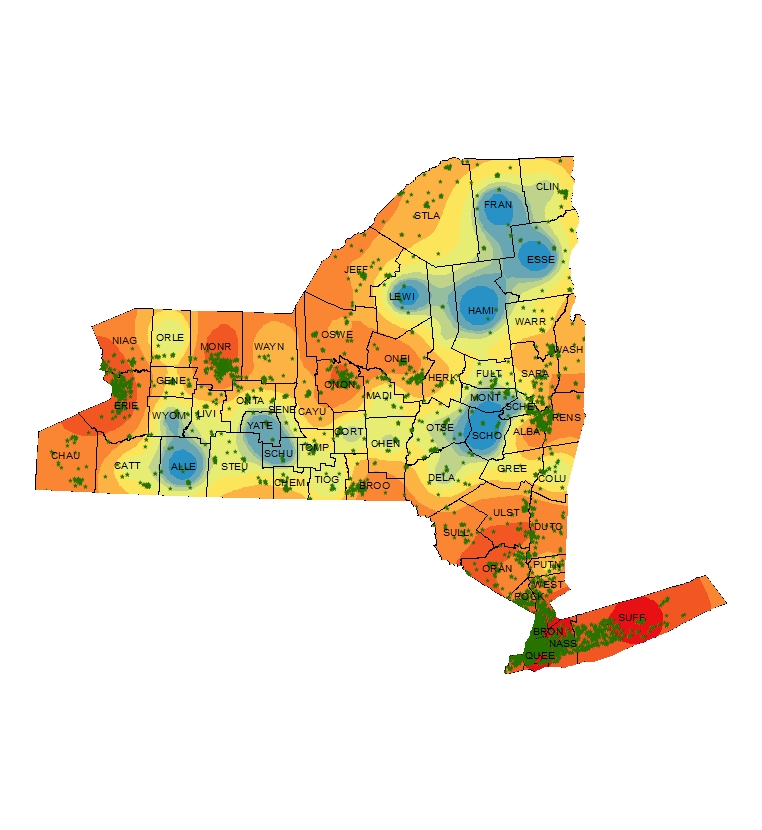
**Figure 6**. OLS Regression result using Python.

**ArcGIS Analysis**



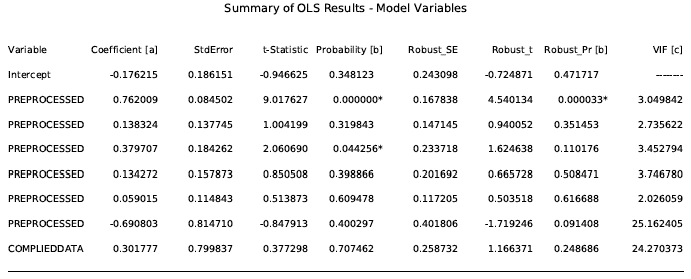
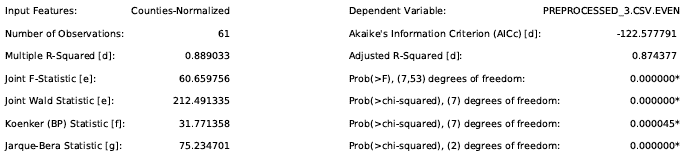
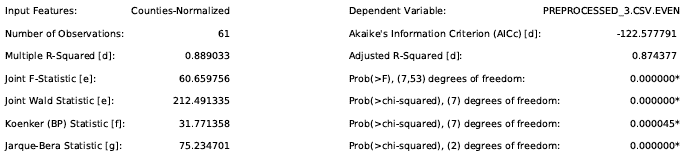
**Figure 7**. Spatial mapping of Event Count/Rate variable (left). Spatial Prediction using Kriging/Cokriging in ArcGIS (right). Red areas represent the highest number of cases of drug overdose cases, while blue areas represent the least number of cases of drug overdose.

Figure 7 shows the maps generated using ArcGIS. The left image shows the mapping of the dependent variable, Event Count/Rate. This shows that most drug overdose cases were located in the southern counties of New York State, along with Monroe and Erie Counties. After applying spatial prediction through Kriging/Cokriging, the unobserved location (the area in white in the left image) was spatially interpolated.



**Figure 8**. Superimposing the health center facilities (green dots) across New York State.

Figure 8 shows the health center facilities across New York State. Most of the health center facilities were found near the hotspots or areas with the highest number of cases of drug overdose deaths. This finding is consistent with what is expected since this project dealt with drug overdose, and the substances that this project was referring to were mostly prescription medications.



**Figure 9.** OLS Regression using ArcGIS.

Figure 9 shows the OLS Regression results obtained from ArcGIS. The values of R and AICc, 0.874 and -122.5, respectively, were comparable to the values obtained from Python analysis. The results showed different contributing variables to the increasing trend of drug overdose death cases. OLS Regression results using ArcGIS showed that Population and Unemployment Rate variables significantly contributed to the rising trend. At the same time, Population and High School Rate variables were observed using Python analysis. A possible explanation for this is the use of another variable in ArcGIS that may have resulted in a different outcome.

**Conclusion:**

This project showed that mapping through ArcGIS was effective in providing a good visualization of the statistical analysis and interpretation. It was also successful in assessing the factors contributing to the increasing trend of drug overdose cases and in identifying areas with the highest number of drug overdose cases, which were the southern counties of New York State, Monroe, and Erie Counties. It was found that areas with high population and unemployment rates had a higher chance of having cases of drug overdose. Also, education played a vital role since it showed a negative correlation with respect to drug overdose cases.

Implementing the information obtained from this project will provide an effective and informative cartographic representation of the cases of drug overdose mortality in New York state that will benefit the general public, policymakers, and healthcare providers to understand better the current and existing problems that people are currently facing with drug overdose.

**References:**

Azadfard, M., Huecker, M. R. & Learning, J. M. (2023, July 21). *Opioid Addiction*. Treasure

Island (FL): StatPearls Publishing. Retrieved September 27, 2023, from <https://www.ncbi.nlm.nih.gov/books/NBK448203/>

Centers for Disease Control and Prevention. (n. d.). *Uncovering the Opioid Epidemic*. Centers

for Disease Control and Prevention. Retrieved September 26, 2023, from <https://www.cdc.gov/museum/pdf/cdcm-pha-stem-uncovering-the-opioid-epidemic-lesson.pdf>

Centers for Disease Control and Prevention. (2023, August 8). *Understanding the Opioid*

*Overdose Epidemic*. Retrieved September 9, 2023, from<https://www.cdc.gov/opioids/basics/epidemic.html>.

Congressional Budget Office. (2022, September). *The Opioid Crisis and Recent Policy*

*Responses*. Congressional Budget Office. Retrieved September 26, 2023, from <https://www.cbo.gov/system/files/2022-09/58221-opioid-crisis.pdf>

deShazo, R. D., Johnson, M., Eriator, I. & Rodenmeyer, K. (2018, June). Backstories on the US

Opioid Epidemic. Good Intentions Gone Bad, an Industry Gone Rogue, and Watch Dogs Gone to Sleep. *The American Journal of Medicine, Vol. 131, No. 6*. <https://doi.org/10.1016/j.amjmed.2017.12.045>

McGranahan, D. A. & Parker, T. S. (2021, April). *The Opioid Epidemic: A Geography in Two*

*Phases*. U.S. Department of Agriculture, Economic Research Services. Retrieved September 26, 2023, from <https://www.ers.usda.gov/webdocs/publications/100833/err-287.pdf>

Spencer, M. R., Miniño, A., Warner, M. (2022, December). *Drug Overdose Deaths in the United*

*States, 2001-2021*. NCHS Data Brief, No. 457. Retrieved September 9, 2023, from<https://www.cdc.gov/nchs/data/databriefs/db457.pdf>.

Texas Health and Human Services. (n. d.). Quality Monitoring Program “Opioids: An Epidemic

Crisis”. Texas Health and Human Services. Retrieved September 27, 2023, from <https://www.hhs.texas.gov/sites/default/files/documents/doing-business-with-hhs/provider-portal/QMP/opioids-epidemic-crisis.pdf>

**Appendix Data Dictionary:**

1. New York State Opioid Data Dashboard

| Name | Description | Data Type | Sample |
| --- | --- | --- | --- |
| County Name |  | Allegany | Broome |
| Priority Area Number |  | Number | 1 |
| Priority Area |  | Text | Opioid Data Overview |
| Indicator Number |  | Text | op51 |
| Indicator |  | Text | Overdose deaths involving any opioid, crude rate per 100,000 population |
| Event Count/Rate |  | Number | 70 |
| Average Number of Denominator/Rate |  | Number | 192222 |
| Percentage/Rate/Ratio |  | Number | 36.4 |
| Lower Limit of 95% CI | | NaN | NaN |
| Upper Limit of 95% CI | | NaN | NaN |
| Data Comments | | Text | |
| Data Years |  | Date | 2020 |