For office use only	Team Control Number	For office use only
T1 ________________	**1921514**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**C**	F4 ________________

2019
MCM/ICM
Summary Sheet

Trends prediction of opioid abuse and response to the crisis in US

Summary

In order to cope with opioid crisis in America, we establish mathematical models according to the given datasets and Geographic information we find.

For the prediction of the spread and characteristics of the reported indicated drugs, initially, we extract data in the five states from 2010 to 2017 and mark the areas of serious drug addiction and abuse on Google map. We use the ***time series analysis*** to predict the tendency of total drug reports in each state. Secondly, we combine Geographic location with time to establish the opioid ***OriginSpreadModel*** to predict the distribution of drug abuse. Furthermore, we determine the threshold levels by multiple linear regression to alert the drug abuse.

In order to review the socio-economic factors impacting the spread of opioid drugs and give the explanation to the phenomenon of the addiction, we use ***Principal Component Analysis*** to find out six major factors highly related to the distribution of drug abuse including Widow, Unmarried status, Companion time in childhood, Population, Migration and Education. For further research, we use D-coefficient as variables to analyze relationship between the major factors and the distribution of drug abuse by ***multiple linear regression***.

At last, we conclude that migration and marital status are the most influential factors affecting the distribution of the opioid addiction and abuse. And the distribution tend to be concentrated and rampant in some major cities which can affect the edge city at the same time. To sum up, our model is practical and reliable for predicting where and when the number of drug reports is in reality.

Keywords: **Time series analysis ; longitude and latitude ; Origin-Spread-Model; Principal Component Analysis;multiple linear regression**

Contents

1 Introduction

1.1 Restatement of the problem

Opioids can effectively relieve pain, especially in cases where other treatments and drugs do not provide adequate analgesic effects. But at the same time, opioid analgesics are addictive. Today, millions of Americans suffer from pain and often use opioids to treat their conditions. However, the danger of overdose and prescription abuse have been a growing problem throughout the United States. It seriously affects public health and economic development, such as lost productivity, health care costs, addiction treatment and criminal justice involvement.

For this problem, we focus on five U.S. states (OH, KY, WV, VA, PA) and their affiliated counties. In data 1, drug identification counts for heroin and synthetic opioids are important indicators of drug abuse in a state or county. The abuse of opioids will spread in a certain pattern. This pattern may be influenced by many factors such as geographical location, population size, economic development level and policies and so on.

In order to relieve the crisis of opioid abuse, we must analyze the changes of opioid drug abuse between in 5 states and counties from 2010 to 2017 (data 1) and the social factor changes obtained from the census (data 2),then establish mathematical models to describe the spread of opioid abuse, and select strong factors that affect drug abuse, Finally make specific predictions to identify appropriate prevention and response activities.

1.2 Our goals

- Predict the trends of drug (opioid) abuse in the next few years (from both state and county perspectives), and introduce geographic factors into our model to predict drug abuse at specific locations and times .
- Determine the drug identification thresholds for each state, predict what time specific states will develop into severe drug abuse areas, and provide appropriate advice to the government.
- Choose the variables that are strongly related to drug abuse among many social factors, and modify the mathematical model based on these factors.
- Propose measures to inhibit the abuse of opioids, use the above model to verify the effectiveness and rationality of the measures,and list the key parameters for identifying the success (or failure) of measures.

2 Assumptions and notations

2.1 Assumptions

- The smallest geographic unit in which drug abuse occurs is county level.
- Ignore the impact of other states on the 5 states.
- the analysis of social factors and drug abuse relationships in one state should be portable in other states, because social factors are universal.

2.2 Notations

Table 1: Notations

Symbol	Definition
α	The weighting coefficients of one-off exponential smoothing
(x_0, y_0)	The coordinates of each state origin (all listed in Table (1))
(x_i, y_i)	The actual coordinates of each county (all listed in appendix
(x_j, y_j)	The calculated relative coordinates of each county
N_j	Total drag reports in the-origin-spread-model
w	Time series of states
Δw	Time series D-value of states

3 Data 1 preprocessing

3.1 variables election

The attached data file 'MCM-NFLIS-Data1.xlsx' provides us with 8 years (2010-2017) of data in 10 variables ,and each column has about 24,000 data. In the process of establishing a preliminary model, it is considered as a primary issue to describe the degree of abuse of drugs required (synthetic opioid and heroin) by states and counties. Other types of drugs that are detected in addition to opioid analgesics and heroin are not considered. Therefore, we ignore 'TotalDrugReportsCounty' and 'TotalDrugReportsState' two columns.

3.2 Looking for latitude and longitude

In addition, we need to get the latitude and longitude information of each county to predict the spread of drug abuse.FIPS code is not easy to establish contact with latitude and longitude. So we abandon the three variables involving FIPS and choose to directly use google map to obtain the latitude and longitude information of the centers (Location center) of some representative counties.

It is not necessary for each county to be included in the scope of the study, especially in the counties with a small total number of drug reports issued in 2010-2017, which contribute little to the degree of drug abuse throughout the state. We will ignore it for the time being.

The table (11) and table (12) (In appendix) lists the latitude and longitude of the top three counties (geographic centers) in each of the states during 2010 to 2017, preparing for the establishment of our Origin-Spread-Model. Actually, more county latitude and longitude is used for each state in the process of building the model. All latitude and longitude related information is listed in the appendix.

4 Part 1

4.1 Preliminary visualization of drug abuse

The trend of the total number of drug reports over time is shown in figure 1 above. We can preliminary get the following information by analyzing each curve of state.

- OH (Ohio) and PA (Pennsylvania) have a large drug abuse base and grow at a faster rate over time.
- Compared to OH and PA, the total drug abuse in (KY) Kentucky and (VA)

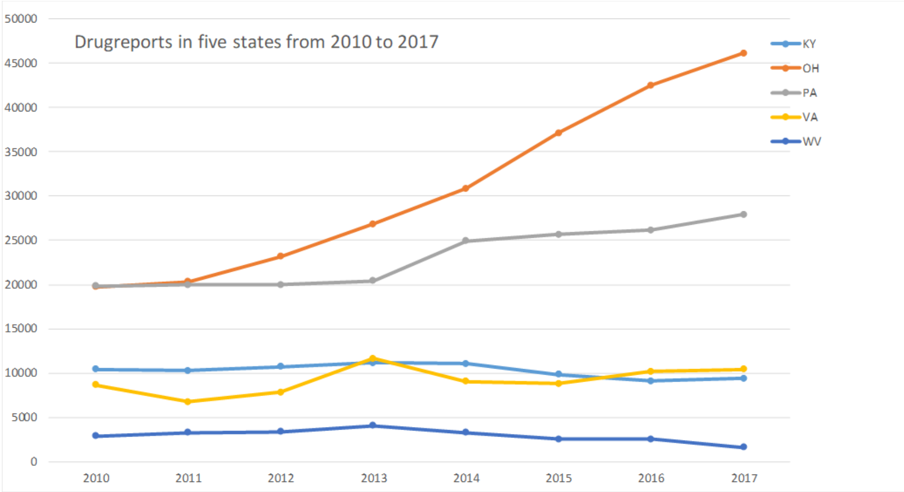


Figure 1: Drug reports In Five States (2010 - 2017)

Virginia is relatively low, and the fluctuation is relatively small, especially KY.

- WV (West Virginia) has the lowest total number of drug abuses and is relatively stable, even at the end of the decline.
- The total number of drug abuses in almost all states experienced an increase between 2012 and 2014.

We analyzed the types of drugs and concluded that the proportion of heroin abuse is relatively large, but it is decreasing year by year, while the abuse of other synthetic opioids is increasing . The total number of drug abuse is still on the rise.

4.2 The time series model

In addition to the above conclusions, figure 1 also provides a perspective to predict trends throughout the state for the next few years: time series analysis.The biggest advantage of this treatment is that there is no need to separate any factors that affect the degree of drug abuse except time. Historical data can predict the trend of the next few years from the perspective of the state.Here we use the simplest one-off exponential smoothing method to predict the trend.

The time series are expressed as $y_1, y_2, ..., y_t, ..., \alpha$ and the α is the weighted coefficient , $0 < \alpha < 1$, the equation of one-off exponential smoothing method is shown below.

$$S_t^{(1)} = \alpha y_1 + (1 - \alpha) S_{t-1}^1 = S_{t-1}^1 + \alpha \left(y_t - S_{t-1}^1 \right)$$

The above formula is improved by the moving average formula.The recurrence formula for the moving average is

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N}$$

Using $M_{t-1}^{(1)}$ as the best estimate of y_{t-N},we can get the formula below.

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - M_{t-1}^{(1)}}{N} = \frac{y_t}{N} + \left(1 - \frac{1}{N}\right) M_{t-1}^{(1)}$$

Let $\alpha = \frac{1}{N}$,Replace S_t with M_t,we can get the formula below.

$$S_t^{(1)} = \alpha y_1 + (1 - \alpha) S_{t-1}^1$$

After further derivation, we get the fomula below.

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t$$

The t index smoothing value is used as the t+1 prediction value.

For an exponential smoothing prediction method, the choice of weighting coefficients is the most critical (the weighting coefficients is expressed as α , and $0 < \alpha < 1$).We traversed 0 to 1 with an accuracy of 0.1, and finally found that when α is taken as 0.4, the sum of the residuals of the Ohio fitting is the smallest. According to the data from 2010 to 2017 in Ohio, the data for 2018 can be predicted, and the 2018 data is used as a known condition to continue the forecast. The other four states adopt the same treatment.

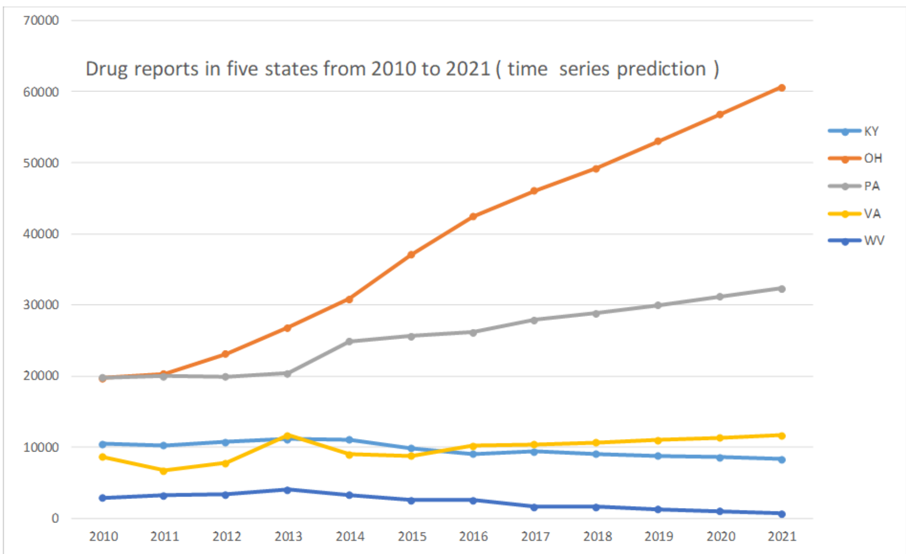


Figure 2: Drug reports in five states from 2010 to 2021 (time series prediction)

Use an exponential smoothing to predict the total number of drug abuses in each state between 2018 and 2021, as shown in figure 2. compare with figure 1, we can conclude that:

- the five states will basically follow the original trend in the next few years, and this is the result of a combination of factors.
- The US government needs to develop policies to suppress drug abuse in Ohio and Pennsylvania, and to focus on Virginia, which is on the rise.

4.3 The Origin–Spread–Model

To describe the spread of drug abuse across states, we must analyze county-level data. The impact of geographic factors needs to be added to the new model, so that we can predict the extent of drug abuse in specific locations in the future.

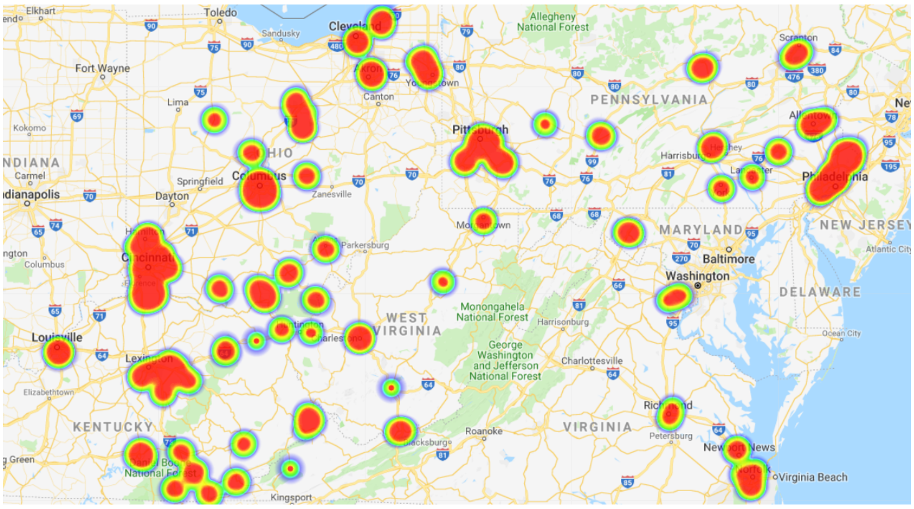
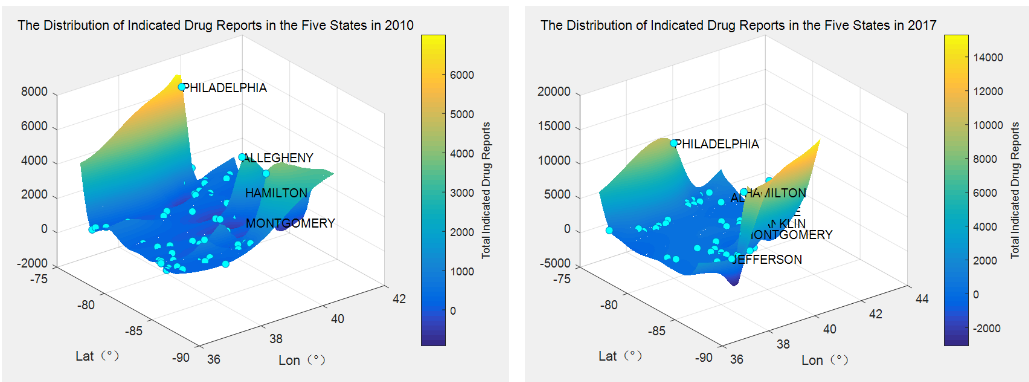


Figure 3: The heat map of total drug abuse in some counties (in 2010)



(a) The year of 2010

(b) The year of 2017

Figure 4: The surf map of total drug abuse in some counties

We selected counties with a total number of drug reports reaching more than 100 in each state in 2010. Figure 3 shows information on the latitude and longitude of these counties and the total number of drug reports. We can easily see from Figure 3:

- Centered on more serious counties such as Cincinnati, Lexington, Columbus, Pittsburgh, and Philadelphia, the severity is reduced from near to far.

Similarly, figure4 uses the surf function in matlab to fit the latitude and longitude information and the total number of drug reports in the above counties in 2010 and 2017.We can easily see from Figure 4: the average level of drug abuse in counties in 2017 has increased compared to 2010, especially around the most severe counties such as Philadelphia. We can approximate that:

- The most severe county has similar effects to assimilation or infection to surrounding counties.

Based on the above analysis, we establish a the orgin–spread–model, which uses the latitude and longitude of the counties with the highest total number of drug reports in each state from 2010 to 2017 as the reference latitude and longitude of the orgin, that is, the coordinate origin of each state.

Table 2: Selection of origin of each state

States	Conties	Latitude	Longitude
WV	BERKELEY	39.464299	-78.01382
OH	HAMILTON	39.183409	-84.52923
KY	JEFFERSON	38.189905	-85.66956
PA	ALLEGHENY	40.467319	-79.99824
VA	CHESAPEAKE CITY	36.755396	-76.28374

In summary, to judge the extent of drug abuse in a region, we must consider the time factor, but also consider the distance factor from the state origin. We give the empirical formula of the model:

$$(x_j, y_j) = (x_i, y_i) - (x_0, y_0) \tag{1}$$

Equation (1) shows the relative coordinate calculation method for each county. (x_0, y_0) indicates the coordinates of each state origin (all listed in Table (2)). (x_i, y_i) indicates the actual coordinates of each county (all listed in appendix). (x_j, y_j) represents the calculated relative coordinates of each county.

$$N_j = \frac{c * \frac{x_j}{y_j} * w}{a * {x_j}^2 + b * {y_j}^2} + d * t + e \tag{2}$$

Table 3: time series of Hamilton from 2010 to 2017

Years	2010	2011	2012	2013	2014	2015	2016	2017
The sum of drug reports	3628	2949	3880	4029	4384	6243	7825	9995

Figure (3) shows: the distribution of the drugs reported is not related to the pure distance (expressed as $\sqrt{x^2+y^2}$), so we match the coefficients for both the x^2 and y^2 of the distance factor. Figure (3) also shows: the attenuation of the drug reports along the x-axis is significantly greater than the attenuation along the y-axis, so we exemplify this feature on the numerator.In order to show the central influence of the origin, we define the parameter w. It represents the time series of the origin of the state from 2010 to 2017 (input year, output the total number of annual reports). In addition, time factors and constant terms are added, and the three are connected by a linear relationship to form the formula (2).

We used the least squares method to fit the data for each county in each of the five states. Here we analyze the orgin–spread–model of one representative state (OH).And the (3) shows time series of Hamilton from 2010 to 2017 (the value of w).

Table 4: Model Parameters (WV and OH)

States	a	b	c	d	e
OH	298.71	424.766	7.7	127.2	30.43

4.4 Threshold determination, Prediction and Conclusion

According to the conclusions obtained before: the farther away from the origin, the lower the degree of abuse. However, with the increase of time, under the effect of origin assimilation or infection, the level of drug abuse in the surrounding areas will gradually increase. After sufficient time, the degree of drug abuse will spread to the boundary of this state, and the crisis of drug abuse in the whole state at this time will be irreversible.

In order to prevent the irreversible drug abuse crisis , We need to define the drug identification threshold for each state (should be based on the dose reports). The function of this parameter is to measure the degree of drug abuse in all counties in the state. If a county exceeds this value, it is determined that the county is a serious abuse area. If the county at all borders of the state exceeds this value, then the state is determined to be a serious abuse of the area (this circumstances will be irreversible).

In Ohio, for example, the average of the total number of drug reports from 2010 to 2017 in all counties on its border is taken as Ohio's drug identification threshold.(6) we symbolically mark the four boundary points on figure5, where the

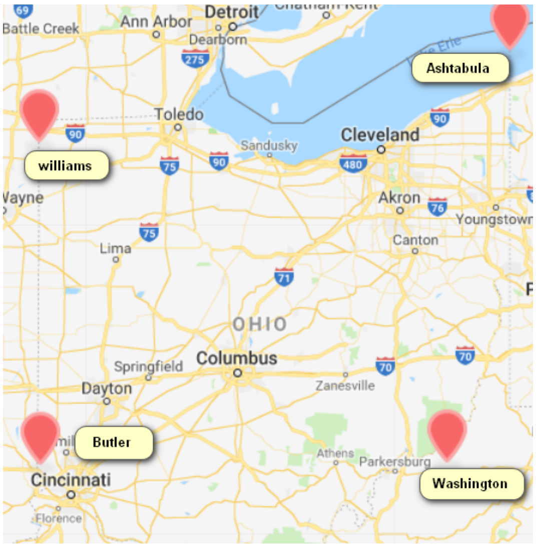


Figure 5: Select 4 counties on the Ohio border

average value of washington in the lower right is 140.75. However, the degree of drug abuse in Ohio is serious, and the final threshold is quite different from this value, so we will not Select Washington's data.

- According to the above method, we determine that the threshold of 5 states. See the Table (5) for details.

Table 5: Threshold of 5 states

States	OH	WV	KY	PA	VA
The drug identification threshold	352	82	234	321	249

Use the Origin-spread-model and state thresholds to predict a county that could be transformed into a heavily drug abuse area in 2018. Based on the distribution of geographic data, we determin the approximate range of drug abuse distribution, and then we select several points for a large number of data analysis, and finally obtained representative results of five states.

Table 6: Geographic prediction result

States	Geographic prediction result	latitude	longitude
OH	Lima	40.73	-84.09
WV	LoganChapmanville	37.93	-82
KY	Jackson	37.61	-83.45
PA	Lewis Town	40.58	-77.57
VA	New Kent	37.4	-76.85

In Ohio, for example, according to the model and threshold, at the end of 2019, Ohio's left border has a yield of 534, which has exceeded the threshold of 352. By 2020, the original geographic location with the lowest degree of abuse will reach 576. It is located on the lower left corner of Ohio, with a latitude and longitude of (39.47, -84.79) and reports will reach 576. At this point, the entire state can be considered a serious drug abuse area.

5 Data 2 preprocessing

5.1 Data 2 preprocessing

Files of Data 2 describe the census data of the five counties from 2010 to 2016 (including family, economy, education, etc., we collectively refer to social factors). The analysis and data preprocessing of these large-scale data files are shown below.

- Some of the social factor variables have missing data in some years. For such variables, we do not analyze them.
- There are many variables that are in the form of percentages, consider a situation: where if the percentage is constant, the cardinality increases and the actual value increases. Percentage class data does not have analytical value if the cardinality is unknown, we do not analyze them.
- We make a reasonable assumption: the analysis of social factors and drug abuse relationships in one state should be portable in other states, because social factors are universal.

In order to link with previous research and obtain the appropriate amount of data, we selected 138 social factors in the counties of Ohio in 2010 for further analysis.

6 Part 2

We need to screen out variables that have strong correlations with drug abuse among the 138 social factors screened, and process these variables as parameters to correct the previous orgin–spread–model. A conventional factor screening method is to use a regression model,but this method requires no strong correlation between the input variables.However, there are strong correlations such as family factors and marital factors among the 139 variables.So after many attempts, we choose the principal component analysis (PCA) method.

At first, we standardized the origin data as shown below.

$$\overline{\alpha}_{ij} = \frac{\alpha_{ij} - \mu_j}{s_j}, i = 1,2,...,n \; ; j = 1,2,...,m \tag{3}$$

Calculate the correlation matrix

$$r_{ij} = \frac{\sum_{k=1}^{n} \overline{\alpha}_{ki} \bullet \overline{\alpha}_{kj}}{n-1} \;, i \;, j \; = 1, 2,...,m \tag{4}$$

Calculate eigenvalues and eigenvectors. The eigenvalues are the coefficient of principal components which reflects the weight of each factor impacting distribution of opioid addiction and abuse. Then we filter out six the most influential variables.

$$y_1 = u_{11}\widetilde{x}_1 + u_{21}\widetilde{x}_2 + ... + u_{m1}\widetilde{x}_m$$
$$\vdots$$
$$y_m = u_{1m}\widetilde{x}_1 + u_{2m}\widetilde{x}_2 + ... + u_{mm}\widetilde{x}_m$$

Select several principal components, construct the model of evaluation and calculate the comprehensive evaluation value.

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_k} \;, j = 1, 2,...,m$$

$$\alpha_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{m} \lambda_k}$$

$$Z = \sum_{j=1}^{p} b_j y_j$$

Through PCA, we obtained 22 variables with strong correlations with drug abuse from 138 variables, and we classified 22 variables, as shown in Figure (6) . The size of the circle in the figure represents the size of the contribution value, and based on this, six different types of factors are selected, as shown in Table (6).

Table 7: Merged Variables

Merged Variables	Estimate; MARITAL STATUS - Widowed responsible for grandchildren - 5 or more years Total households YEAR OF ENTRY - Native 9th to 12th grade, no diploma Never married

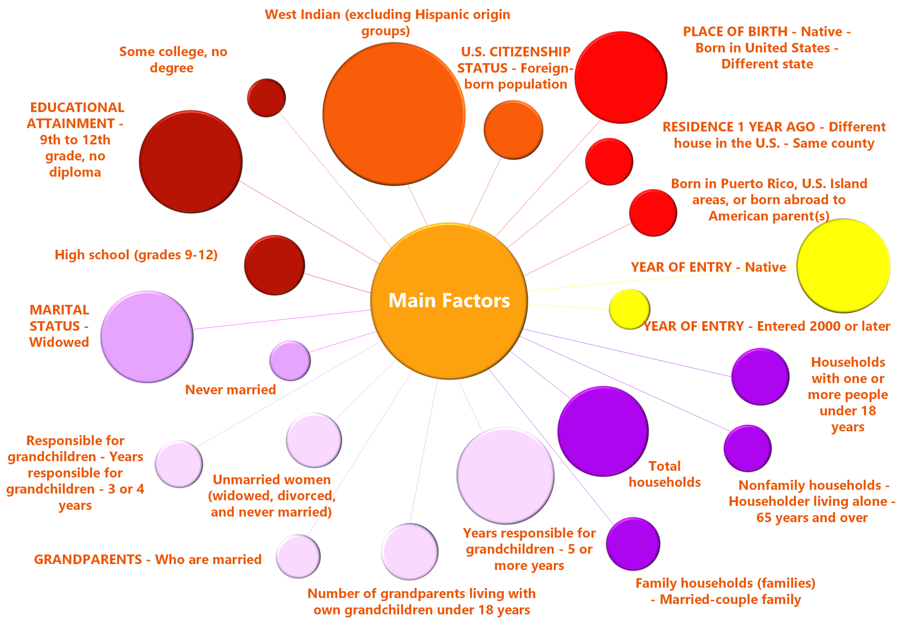


Figure 6: Select 4 counties on the Ohio border

Next, we perform linear regression on these six factors, find out the relationship between them and the drug reports, and get the weights of the six variables to facilitate the optimization of our model.

$$y_1 = u_{11}\widetilde{x}_1 + u_{21}\widetilde{x}_2 + \ldots + u_{m1}\widetilde{x}_m$$
$$\vdots$$
$$y_m = u_{1m}\widetilde{x}_1 + u_{2m}\widetilde{x}_2 + \ldots + u_{mm}\widetilde{x}_m$$

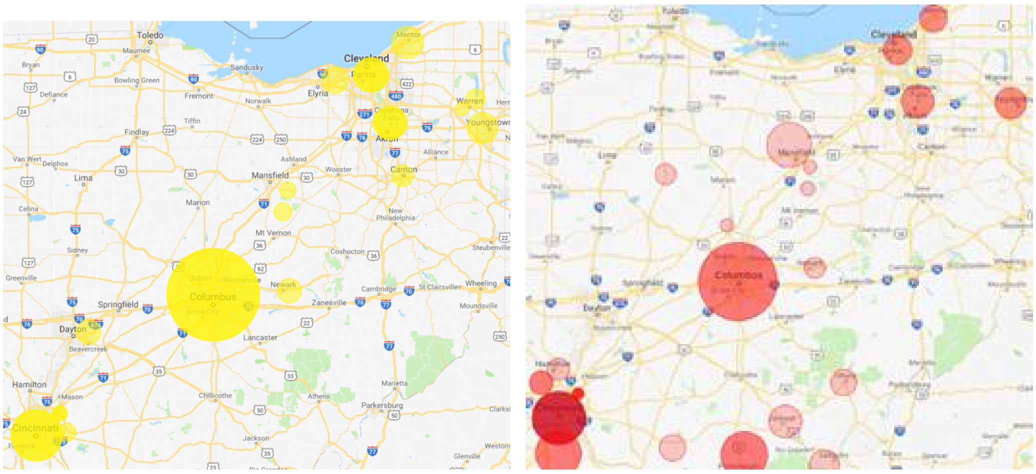
Table 8: Weights

Weights	u1	u2	u3	u4	u5	u6
value	-0.6127	-0.217	0.0566	0.0201	0.101	0.1268

After getting the linear regression expression between the drug reports and the 6 variables, we optimize the model in part 1, we replace w with the equation below.

$$\Delta w = -0.6127x_1 - 0.217x_2 + 0.0566x_3 + 0.0201x_4 + 0.1015x_5 + 0.1268x_6$$

The model of Part 1 contains six relevant factors that make our models and data more relevant and improve the accuracy.



(a) The year of 2010 (b) The year of 2017

Figure 7: Model-optimized verification

Consequently, considering the socio-economic data provided, we use our mathematical model to predict the location of HDAA and the figure (7) have confirmed the accuracy of our model. The left of The size of the each circle represents the scale of the city and the deeper the color is, the more serious situation of drug addiction is. The distribution of opioid addiction on the basis of our model are presented on the right of figure and the left one is on the basis of the real data.

7 Part 3

According to the results of the Part1 and the Part2, it is recognized that the use of opioids in some areas has gradually expanded in both space and time.

From the first part, the spread of opioid has a tendency to concentrate on diffusion. Therefore, the area of the most serious opioid addiction and abuse in each state in the past few years should be marked, like HAMILTON, CUYAHOGA, JEFFERSON. These county should strengthen the drug supervision.

From the second part, the widespread use of drugs has a complicated relationship with many factors. Many of these factors also have a relationship with each other which means it is not applicable to use Multiple linear regression analysis. We use clustering and principal component analysis to find out 22 elements related to drug report and filter out six major factors from them to facilitate the formulation of governance strategies.

The advice on the basis of six major factors are as followed:

- Control the entry of migrants and impose restrictions on those who enter the country. Those who have bad records about drugs abuse should be carefully considered. Try to keep the number of foreigners as close as possible to 2017.

- Strengthen the education of students in school, especially in grade 9 to 12, and clarify the harmfulness of such drugs while reducing the dropout rate of students in school.
- Strengthen the awareness of the elderly to develop and look after young people.
- Implement appropriate policy to control population growth rate in each state, especially OH, PA.
- Promote the development of social welfare and family harmonious.

Table 9

The reduction of 9th to 12th grade, no diploma amount	2000	4000	6000	8000
The reduction of w	43	83	112	156
The reduction of cases	9	14	15	16

Table 10

The reduction of YEAR OF ENTRY - Native amount	1000	3000	5000	7000
The reduction of w	20	61	100	140
The reduction of cases	2	13	15	16

After modifying the model, we select two main factors and calculate the change again. The conclusion is : The reduction in population mobility and The reduction of 9th to 12th grade, no diploma amount can indeed reduce the number of cases. It can be seen that if the growth of these factors can be effectively controlled, it can be changed little in the early stage, but the stage corresponding to the increase is prevented, and the growth of drug report is controlled to achieve better results.

At least, the most easily controlled factor by the government is the demographic factor. We chose a typical location: Jackson, and defined a specific parameter boundary: It was calculated that Jackson's population inflow cannot exceed 2000 every year. Only in this way, Jackson's state will not become a serious drug abuse area until 2021.

8 Strengths and weaknesses

8.1 Strengths

- This model covers a wide area and has many objects. It can analyze the changes of drug report from multiple angles and has better accuracy.
- From the perspective of time and space, the model is combined time, geographic information with socio-economic factors based on three-dimensional observation.
- Through the analysis of a large amount of data, not only makes the prediction model more accurate, but also can comprehensively select the factors affecting drug report, facilitate the comprehensive stereo analysis of the development process of drug report and the formulation of the corresponding management strategy.
- Predict the location and the number of drug reports accurately.

8.2 Weaknesses

- For the certain latitude and longitude, it is meaningless to analyze the situation of a certain point because the model is essentially a discrete model. We often need to select two or three reference points.

9 MEMO

To: Dear government

FROM: Team # 1921514

DATE: 2019.1.28

SUBJECT: The conclusion and suggestion

As is reported in a great amount of news in the nearly few years, the past decade has seen a growing epidemic of opioid addiction and abuse emerge in the United State.

We establish The Origin-Spread-Model Mathematical model to predict the location of the area of serious opioid addiction and abuse and the number of drug reports to help government combat opioid.

For requirement 1, we preliminarily combine time with location to establish a three-dimensional model named The Origin-Spread-Model to predict the specific location. We calculate the thresholds of the five states:352 (OH), 82 (WV), 234 (KY), 321 (PA), 249 (VA). And predict that Lima (40.73, -84.0953),OH, Jackson, (37.6136,-83.45),KY,Lewistown(40.58,-77.576),PA, New Kent(37.408,-76.8536),VA will break out disaster of drug addiction in 2018.

For requirement 2, we choose Principal Component Analysis to filter out the six critical socio-economic factors Widow, Unmarried status, Companion time in childhood, Population, Migration and Education. which is essential to the increasement of opioid crisis and we modify the model in Part one. The distribution of opioid addiction on the basis of our model are presented on the right of figure (7) (The left one is on the basis of the real data).

For requirement 3, we find that the distribution of the opioid addiction and abuse tend to be concentrated and rampant in some major cities which can affect the cities nearby at the same time. The reasons accounting for the phenomenon is complicated which may include population migration and family.

Taking the factors mentioned in the paper into consideration, we propose that:

- The state government can formulate the policy seriously about anti-drug and be stricted with carrying it out, especially in the major city.
- The University and College can put forward courses related to opioid and
- Strengthen the governance of migration.

- Strengthen education in ones childhood and accompany children as long as possible.
- Provide unmarried people (widowed, divorced, and never married) with help and concern

References

[1] Ma M.J.,et al, Dynamics Analysis of Drug Model, 2017.4:p.75-93.,

[2] https://www.cdc.gov/drugoverdose/data/index.html

[3] https://www.drugabuse.gov/publications/misuse-prescription-drugs/what-classes-prescription-drugs-are-commo

[4] https://en.wikipedia.org/wiki/Opioid_epidemic

[5] https://en.wikipedia.org/wiki/Heroin

[6] https://www.rehabspot.com/opioids/addiction/

[7] https://www.annualreviews.org/doi/full/10.1146/annurev-publhealth-031914-122957

Appendices

Table 11: Add caption

States	Conties	Latitude	Longitude	Sum of Drag Reports
WV	BERKELEY	39.464299	-78.013824	3046
	CABELL	38.405956	-82.247692	2126
	KANAWHA	38.357943	-81.635631	1974
	MERCER	37.389034	-81.105677	1732
	MONONGALIA	39.622188	-79.964084	1399
	RALEIGH	37.745833	-81.248344	499
OH	HAMILTON	39.183409	-84.529227	27262
	CUYAHOGA	41.454456	-81.668185	15890
	MONTGOMERY	39.248122	-84.353655	13965
	FRANKLIN	39.969976	-82.986991	9189
	LAKE	41.704496	-81.221104	5186
	Butler	40.587716	-82.425092	4019
	MAHONING	41.035129	-80.790377	3205
	LUCAS	40.704799	-82.420418	3134
	DELAWARE	40.296779	-83.076433	2586
	TRUMBULL	41.345258	-80.78818	2108
	CLERMONT	39.051345	-84.170695	2065
	LICKING	40.117818	-82.488739	1661
	SUMMIT	41.124658	-81.534788	1643
	WARREN	41.238738	-80.818447	1523
	SCIOTO	38.824546	-83.022007	1188
	RICHLAND	40.790959	-82.541415	1170
	Athens	39.331636	-82.110965	808
	ROSS	39.310588	84.650526	784
	GALLIA	38.816984	-82.344514	646
	ADAMS	38.856839	-83.478117	626
	JACKSON	39.054787	-82.638134	496
KY	JEFFERSON	38.189905	-85.669563	13979
	FAYETTE	38.045489	-84.490011	5231
	KENTON	38.948954	-84.534543	4855
	MADISON	37.724233	-84.294408	2744
	PERRY	37.261341	-83.190519	2081
	HARLAN	36.824947	-83.318095	2000
	BELL	36.718443	-83.672627	1886
	BOYD	38.353402	-82.674177	1866
	LAUREL	37.107475	-84.089137	1610
	MONTGOMERY	38.047824	-83.911017	1474
	PULASKI	37.128469	-84.595596	1442
	PIKE	37.468889	-82.377692	1416
	JESSAMINE	37.862701	-84.569252	1391
	WHITLEY	36.759151	-84.122278	1370
	KNOX	36.901026	-83.857513	1181
	CARTER	38.32094	-83.035719	1065
	CLARK	37.98532	-84.156617	1020
	ROWAN	38.191324	-83.437781	1000

Table 12: Add caption

PA	ALLEGHENY	40.467319	-79.998238	53158
	BERKS	40.33786	-75.917753	35945
	BLAIR	40.490688	-78.335835	13844
	BUCKS	40.340026	-75.095886	9418
	DAUPHIN	40.41287	-76.816743	9350
	DELAWARE	39.913831	-75.409794	8234
	INDIANA	40.623135	-79.15319	7550
	LACKAWANNA	41.411881	-75.607499	7526
	LANCASTER	40.06647	-76.306586	6967
	LEHIGH	40.607733	-75.587769	5963
	LUZERNE	41.207973	-75.941252	4705
	LYCOMING	41.284604	-76.999094	4223
	NORTHAMPTON	40.727595	-75.271199	4061
	PHILADELPHIA	39.999256	-75.151001	3936
	WASHINGTON	40.190627	-80.242736	3898
	WESTMORELAND	40.295694	-79.456391	3515
	YORK	39.960752	-76.731579	2251
VA	CHESAPEAKE CITY	36.755396	-76.283738	4900
	FAIRFAX	38.856389	-77.30167	4673
	HENRICO	37.570246	-77.367677	3767
	NEWPORT NEWS CITY	37.127114	-76.529479	2136
	NORFOLK CITY	36.892022	-76.246151	1769
	PORTSMOUTH CITY	36.841366	-76.35357	1743
	PRINCE WILLIAM	38.688484	-77.517663	618
	RICHMOND	37.537676	-77.472987	577
	ROANOKE	37.273736	-79.948788	574
	WISE	36.976908	-82.577591	550