For office use only	Team Control Number	For office use only
T1 ________________	**1901274**	F1 ________________
T2 ________________	Problem Chosen	F2 ________________
T3 ________________	**C**	F3 ________________
T4 ________________		F4 ________________

2019 MCM/ICM

Summary

The Opioid Crisis

It has long been a controversial issue in U.S. concerning the drug addiction over the last several years. Across various generations and social classes, drug use has increased. Based on the recent survey, opioid drug has exerting an exponential boom across different states, which arouses more and more public concern over the prevention of drug-related criminals.

In the first part, through data cleaning and visualization, we expect to provide the viewers with a clear acknowledgement of how the drug cases differ from the perspective of time and location. Next, the employment of the approach of K-mean clustering enables us to classify pairs of Heroin and Oxycodone abuse across each county. Based on this, not only we find out the correlation between drugs of different substances, we also pick out three counties with highest exposure rate to opioid. Lastly, the predictions of drug cases of certain locations in two years are estimated according to the time series model.

Secondly, we conducted some multivariable regressions and generalized linear model to figure out how drug-involved cases is related with several socio-economic features. Eventually, we generated a relatively plausible model to select out the statistically-dependent demographic factors.

In light of these findings, we raised a series of strategies to be implemented to improve the feasibility of drug prevention. On the basis of differential equation, we design two brand-new mathematical models, NYC and NYP respectively, to verify the effectiveness of our prior strategies and identify significant parameter bounds.

Memo

Dear Sir or Madam:

The focus of this paper is the analysis of the trend of drug cases over the time and comparison of drug abuse condition between five states. With the help of statistical methods and mathematics modelling, we intend to find out the pattern behind and figure out the correlated social factors.

Dataset: total drug reports across each year and state, substances of drugs, data of demographic features such as the number of male adults

Findings:

General trend

1. The plot demonstrates a growing trend in drug cases in Ohio State, from respective of both total number of cases and average cases per capita. Meanwhile, Pennsylvania also suffers from high risk of drug-related criminal behavior even though there seems a positive effort improving this issue recently. On the contrary, West Virginia behaves with relatively better performance with the fewest drug cases and greatest deceleration amongst these five states.
2. A linearly correlation between the drug reports of Heroin and Oxycodone across each county
3. Prediction of the total drug reports in 2018 and 2019 across three counties

County	Prediction of 2018	Prediction of 2019
Cuyahoga	**19593**	**21430**
Hamilton	**23998**	**25486**
Philadelphia	**9904**	**8741**

4. Ratio of drug addiction depends on the following demographic variables: ratio of single population, ratio of Latin American population, ratio of bachelor-degree holders, ratio of average household size.

Mathematics model:

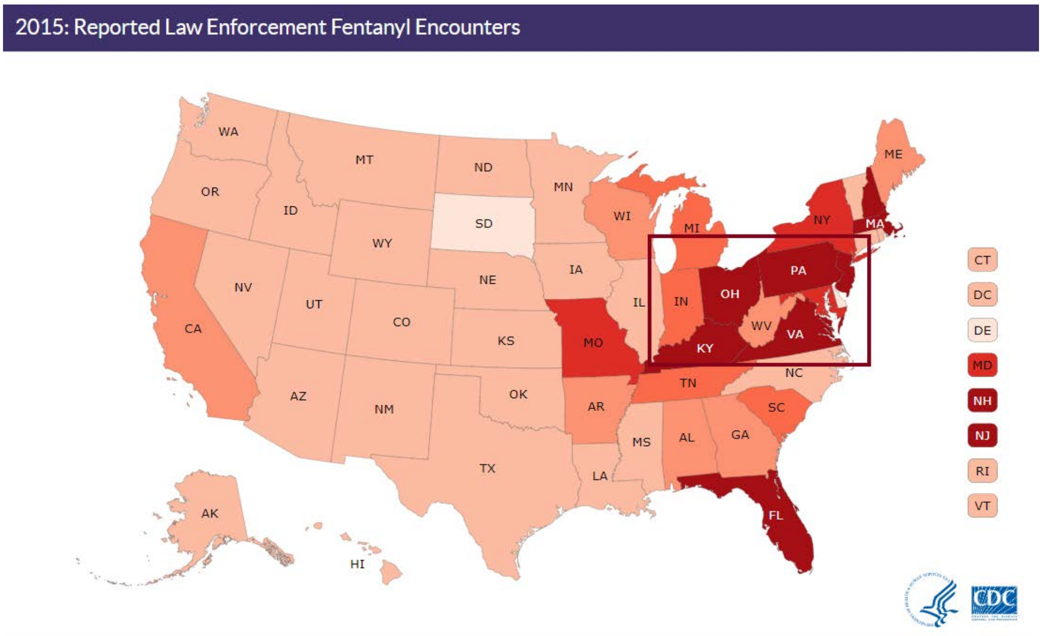
On top of that, if the velocity of spread of drugs could be controlled to be below a certain threshold, the prevention against drug abuse could be improved to a large scale. The differential equation model tells us this specific parameter threshold to verify whether the strategies raised before could succeed or not.

Overview

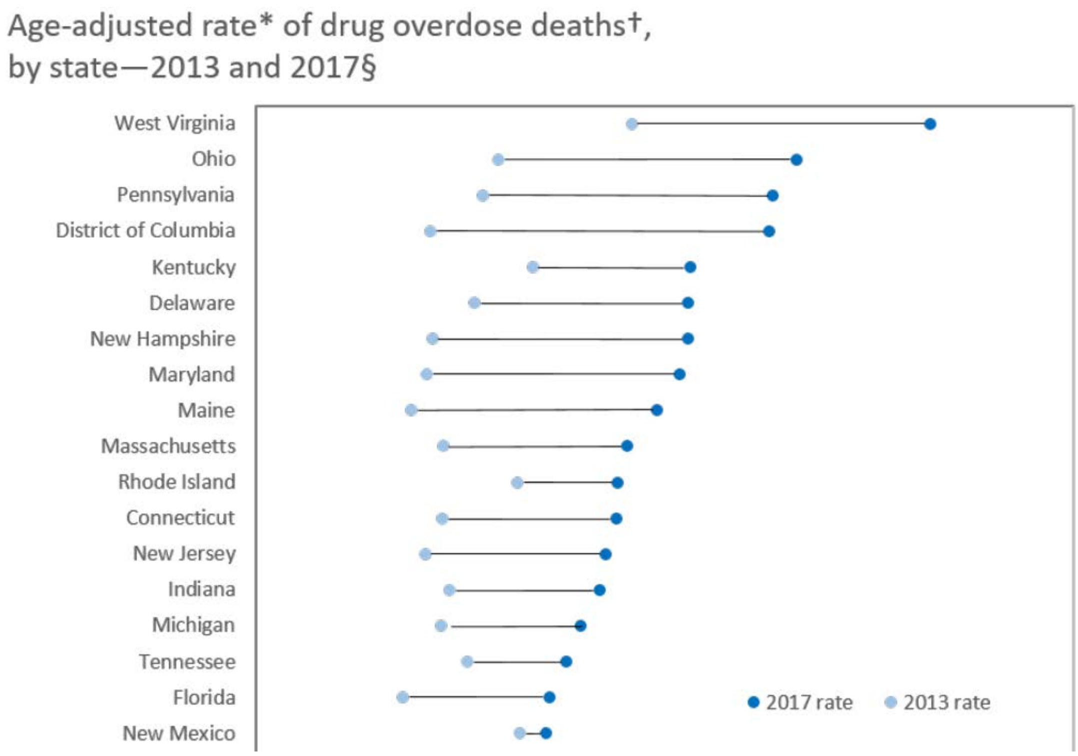
Background

The United States is experiencing a national crisis regarding the use of synthetic and nonsynthetic opioids, either for the treatment and management of pain (legal, prescription use) or for recreational purposes (illegal, non-prescription use). More and more drugs are being obtained through doctors' prescriptions or illegal ways. To make things worse, the spread of the crisis exceeded people’s expectations. Taking heroin as an example, the use of heroin has been increasing in recent years among men and women, most age groups, and all income levels. Some of the greatest increases have occurred in demographic groups with historically low rates of heroin use: women, the privately insured, and people with higher incomes, which is significantly different from the trend in previous period. In other words, heroin is spreading in all directions without dead ends. As it has been mentioned in the background, if the opioid crisis spreads to all cross-sections of the U.S. population, there might be serious economic problems.

So does fentanyl. The enclosed shows the fentanyl encounters in 2015. The darker the color, the more the number. It is quite obvious that those five states in the square is the most serious states in America, which are also the five states that we will discuss in the following part.



As the following diagram shows, the most direct problem caused by drug spread is death. However, it is much more than that. Caeti (1999) conducted an experiment on drug-related crimes in 1994-1996 in collaboration with the Houston Police Department, which showed that the rise in drug-related crimes could lead to robbery, theft and many social problems. Therefore, it is quite urgent to find an effective perspective to face the current high incidence of drug cases in the United States, which is positive.



Literature Review

Drug prevention research is always a major concern of every country. However, the different national conditions of each country cause certain limitations of those researches. Although opioid use has known risks, it is often used to treat pain and has been proved effective. However, long-term administration of chronic non-cancerous pain is controversial, so doctors do not have a clear standard for the use of opioids. For the time being, the study of opioid overdose is mature. People often study the change of total amount and the change of growth rate between different years. These results are evident on CDC and NIDA.

In addition, people also realize that drug problem is related to other factors (such as socio-economic factors). Han Xiao (2013) used analytic hierarchy process (AHP) to establish a crime probability model integrating GIS. Six factors, such as time distance, crime rate, population, police, geographical environment and victim's occupation, were identified as factors affecting the crime location, and the place where the perpetrator is most likely to commit the next crime was predicted. Zhang Ning, Wang Dawei (2018) Drug crime risk assessment and police forecasting based on risk terrain modeling, taking the high incidence of drug crimes in a city (such as gas stations) as a consideration, this paper explores ways to reduce the number of drug cases. Their research methods are reasonable, and we will study the drug cases from the perspective of social and economic data, and give corresponding prevention strategies.

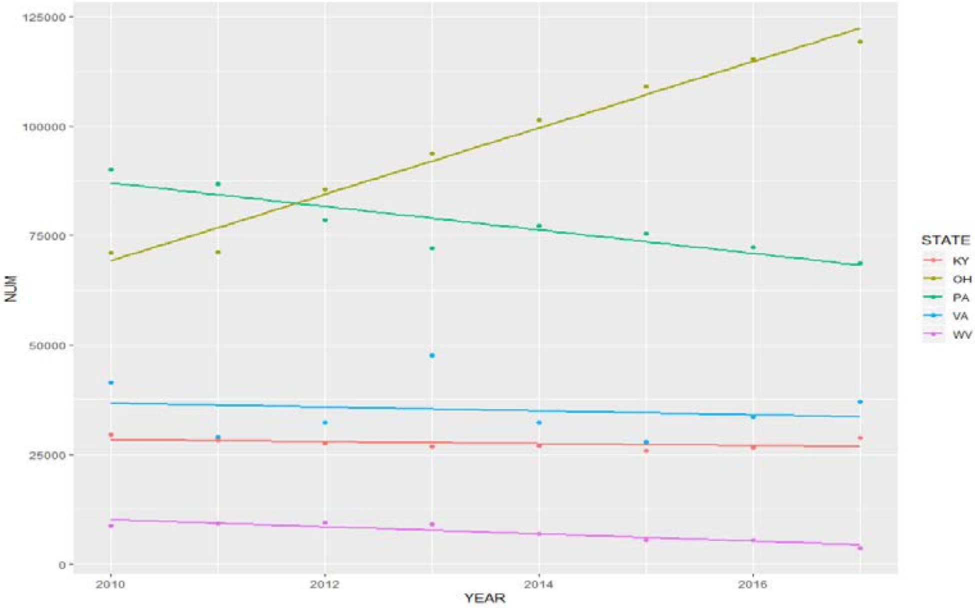
Data analysis

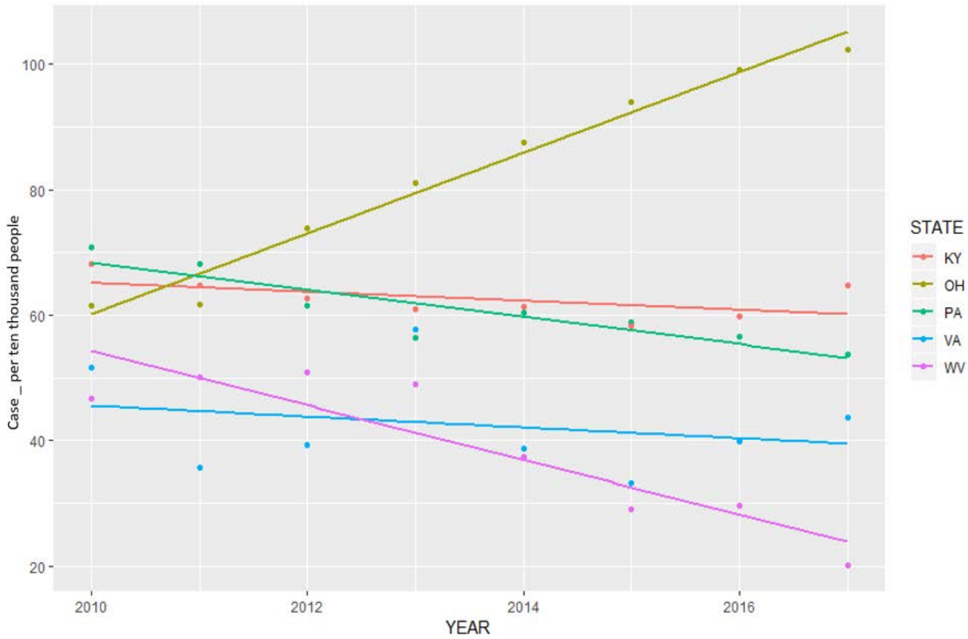
Overview and visualization of data

This part gives an overview of the qualitative analysis of various types of drug cases across five states of US (Ohio, Kentucky, West Virginia, Virginia, and Tennessee) over the recent 8 years (2010 ~ 2018).

1) The trend of drug-related cases

Aiming to interpret the data in a more efficient and straightforward way, we intend to elaborate our findings with the help of various plots. First, what we concern the most is the general trend of drug cases over these years and how they differ across each state.

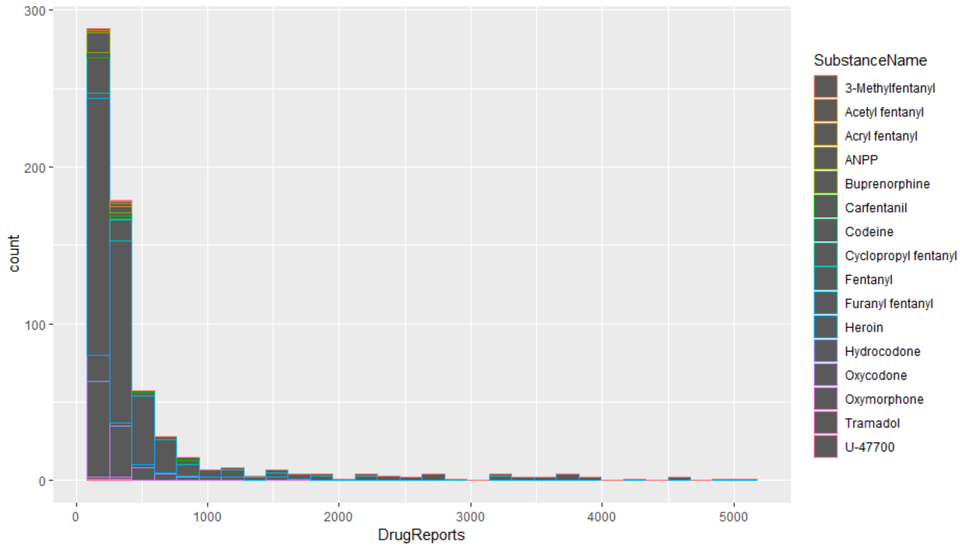




The figure above demonstrates **a growing trend in drug cases in Ohio State**, from respective of both total number of cases and average cases per capita. Meanwhile, **Pennsylvania also suffers from high risk of drug-related criminal behavior even though there seems a positive effort improving this issue recently**. On the contrary, **West Virginia behaves with relatively better performance with the fewest drug cases and greatest deceleration amongst these five states.** Thus, great efforts would be put on the State of Ohio and Pennsylvania in the following analysis.

2) The most common types of drugs

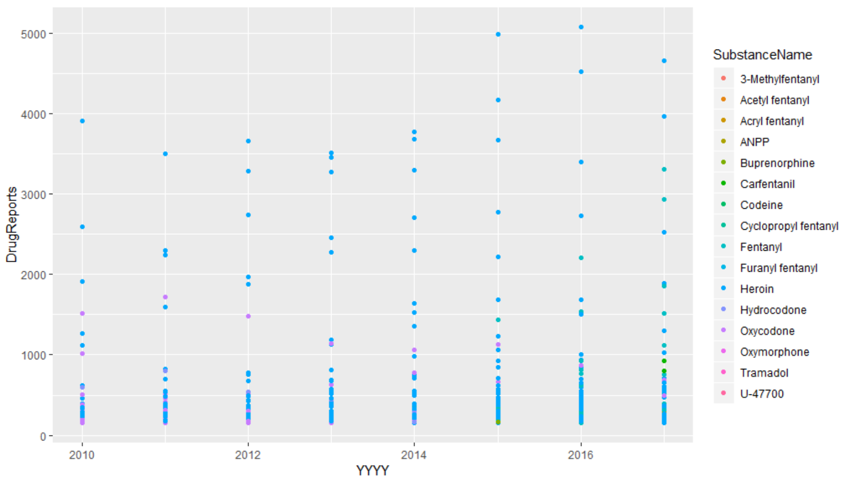
To hammer out which kinds of drugs are widely abused across the whole population, we selected the drug types with high incidence rate (those whose drug cases above 150) and labelled them with different bars in the histogram below.



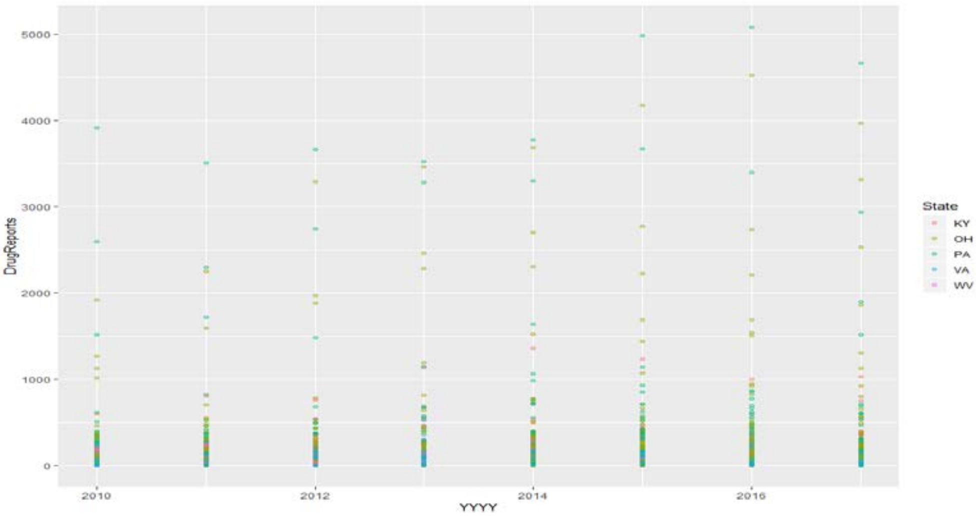
Spread of the number of counties with various drug-related reports

Judged from the shape of the histogram, it has been revealed that most counties suffer from medium-scale drug cases (the plot is negatively skewed, and the numbers of respective substance in each county are centralized in the range of below 1000).

We adopt another kind of plot with scatter points so as to scrutinize the extreme cases.



Obliviously, **heroin** turned out to be the chief culprit amongst all these drugs, almost dominating the drug market and largely overtake the second most widely-spread drug-**Oxycodone**. Therefore, the plot below illustrates how serious the abuse of Heroin shows across these states.



3) Is there a correlation between the abuse of drugs of different substances?
This idea comes from my initial hypothesis that maybe different drug-users exert a similar trend with each other or a reversed pattern, e.g. drug users of heroin tend to be more likely to addict to other drugs as well, thus driving the consumption of Oxycodone as well. The other way around, Heroin and Oxycodone maybe serve as a perfect substitution of each other. In this case, people of a county might turn to another kind of drug due to the local feature characterized as hard accessibility to Heroin.

Not only the simple correlation matrix, K-means clustering is also introduced to deal with this problem. After filtering 49 counties with high incidence rate of drug abuse, we expect to cluster pairs between the total cases of Heroin and Oxycodone respectively through approach of unsupervised learning.

Hierarchical modelling and Clustering

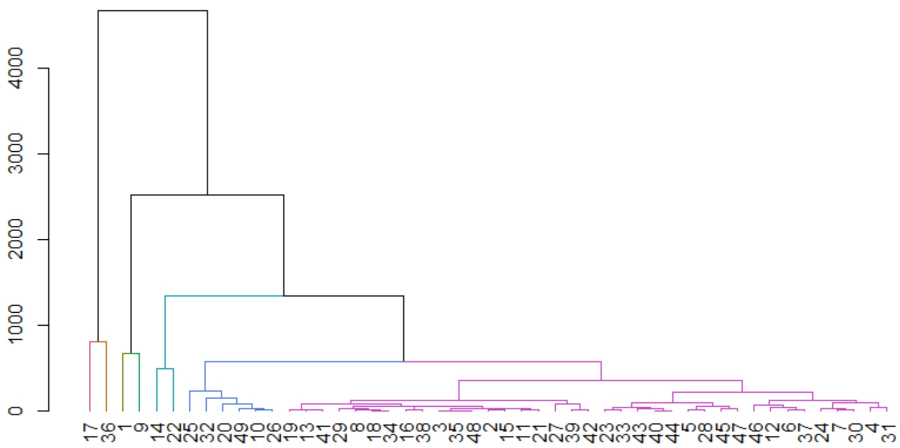
Rather than deduction from human being's intuition or naturally existing classification labels, we would like to analyze this problem in another way which let the dataset itself exhibits any pattern and be split into several clusters so that make each pair in this model taken into account. Afterwards, through observing the difference between each cluster, we hope to verify the question raised before and find out if a pattern could be concluded.

Perform the complete linkage for hierarchical clustering

1. A general introduction of the method of “complete linkage”:
Complete-linkage clustering is one of several methods of agglomerative hierarchical clustering. At the beginning of the process, each element is in a cluster of its own. The clusters are then sequentially combined into larger clusters until all elements end up being in the same cluster. The method is also known as farthest neighbour clustering.

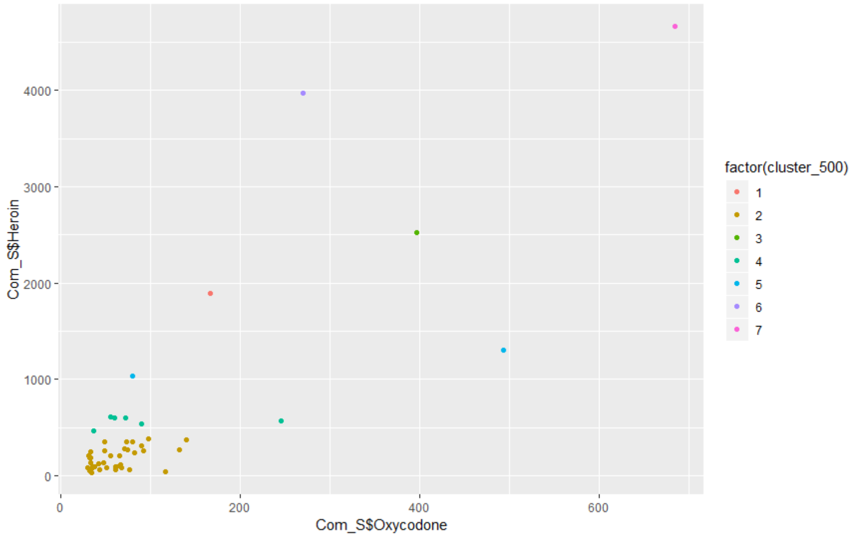
2. Reasons why I employ this approach rather than other common linkage methods:
Complete linkage tends to find the compact clusters of approximately with equal diameters. Compared to “single linkage” method, this approach could prevent the existence of the case where some clusters are forced to combine together due to the fact that only few observations are close to each other even while most of points in each cluster appear to be distant to another group. To interpret in a more comprehensive way, I would like to differ the groups to the greatest extent as to amplify the difference between each pair.

Demonstration of the result and the dendrogram of the hierarchical clustering

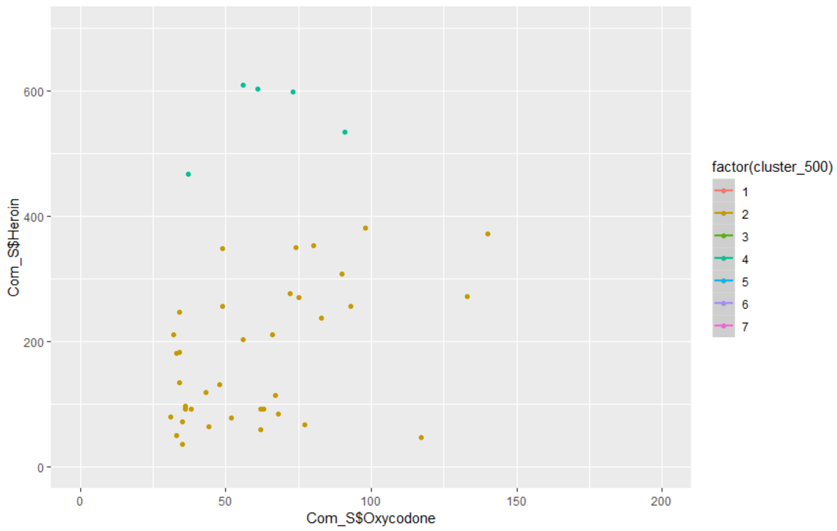


As the dendrogram above shows, the overall classification could be performed as detailed as the plot (with Y-axes interprets the height of the classification tree and X-axes shows the ID of each pair between Heroin and Oxycodone)

Eventually, we set the cutting height of the classification tree to be 500 and the dataset was divided into 6 clusters.
Visualization of the outcome of the K-means clustering:



According to the clustering outcome, most points are split into the second and the third group (yellow and green respectively). Several outliers are separated alone and made up a new cluster independent of other individuals. After checking the county names of these extreme values, **Philadelphia, Hamilton and Cuyahoga** are the top three worst counties in terms of abuse of Heroin and Oxycodone.

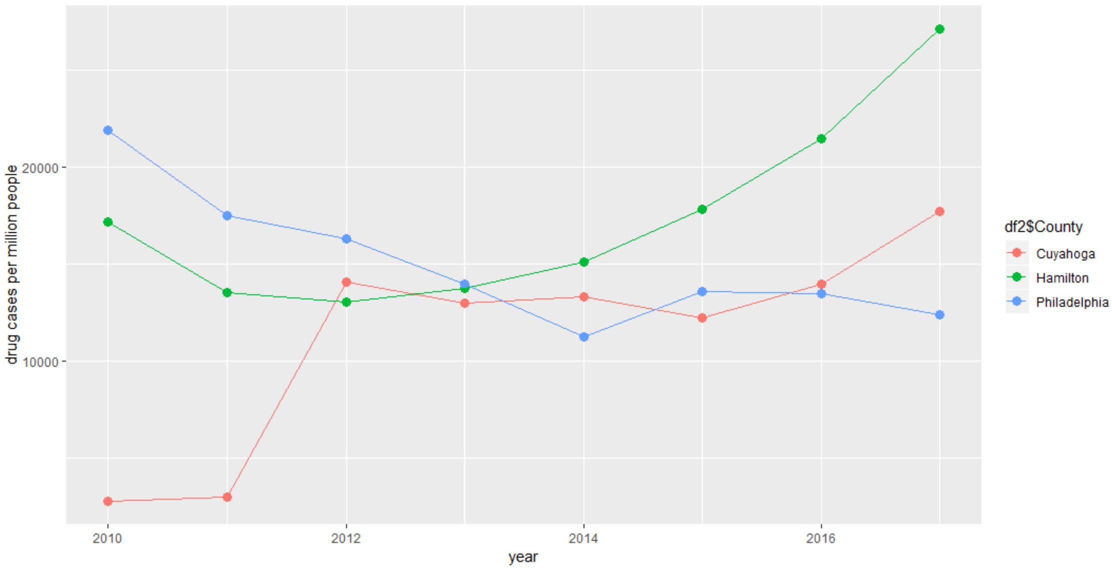


Amplify the most concentrated-clustering part of the plot, it could be observed that **there is a positive relationship between the use of Heroin and Oxycodone.** Consequently, the spread of the amount of drug consumption does not differ across various main substances. The amount of heroin-related case could serve as an important indicator interpreting the assessment criterion of the drug-controlling situation.

Forecasting

Drug cases in Philadelphia, Hamilton and Cuyahoga

Due to the fact that these three counties suffer the most from drug cases, the change of the average drug abuse (cases/million people) is shown as below.



The predicted number of average drug cases (cases/million people) in 2018 and 2019

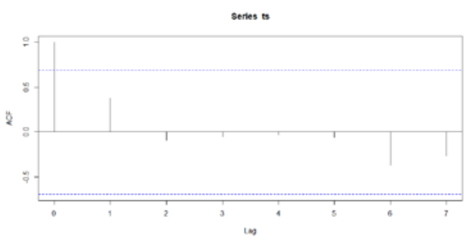
E.g. time series model of drug cases per million people

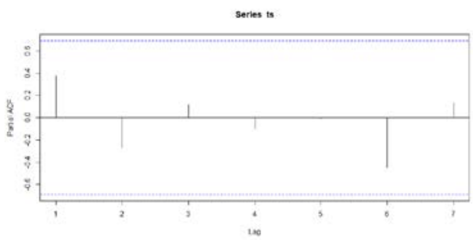
For a lack of space, only the deduction process of the time series model of the County Cuyahoga is presented in this paper. Through observing the ACF and PACF, an ARIMA (1,1,0) model proves to fit the data well.

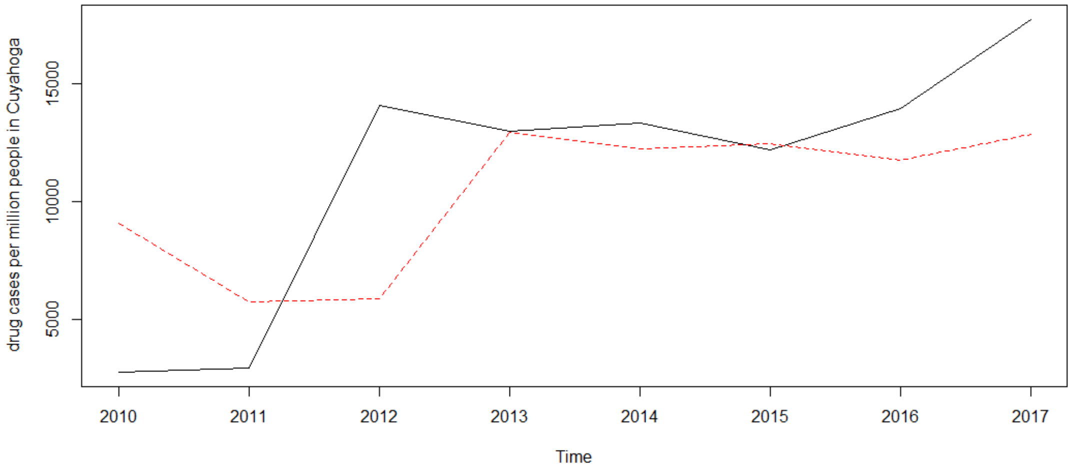
$$\Delta Y_t - \varphi * \Delta Y_{t-1} = \mu + \varepsilon_t \quad \{\varphi < 1\}$$

$$\Delta Y_t = Y_t - Y_{t-1}$$

$$Y_t == \text{drug cases per million people at year } t,\ t == \text{year}$$







County	Prediction of 2018	Prediction of 2019
Cuyahoga	19593	21430
Hamilton	23998	25486
Philadelphia	9904	8741

Regression

It is clear that drug abuse and addiction has been increasing, what remains to be found are the specific reasons behind, though there are many potential social factors might be able to explain the change of drug cases. For instance, increase in substance abuse in America proves to highly depend on individual's social class, disposable income, levels of stress and so on and so forth. Since the dataset provided for this project covers detailed information about population structure of different counties, we intend to focus more on how the incidence ratio (drug cases/population) is related with demographic features.

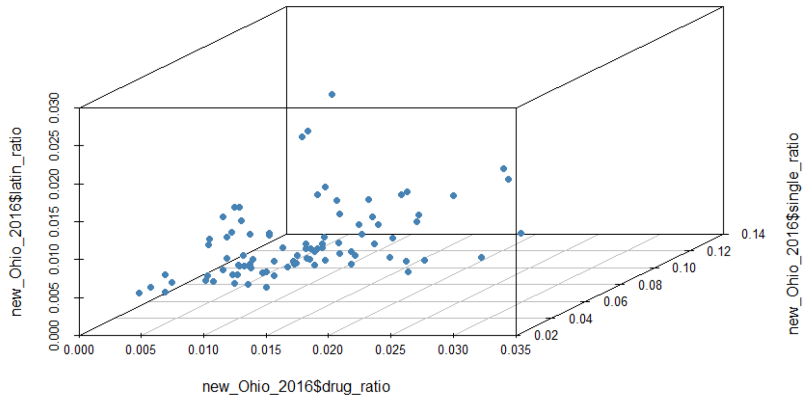
Prior to conducting the regression, we listed a series of population indicators of interest which we guess somehow would be correlated to our study.

Weight of the following demographic factor in each county	**Levels**
Gender	Male
	Female
Martial status	Single
	Married
	Divorced
If have minors in the family or not	Yes
	No
Education	Bachelor's degree or above
	Otherwise
Ethnic background	Native
	Latin America
	Asia
	Africa
Age	Above 25 years old
	Below 25 years old
Family size	(numeric)
Household size	(numeric)
Residence year	Less than one year
	Over one year

Multivariable regression (Cross sectional data)

Dataset: Number of drug cases in each county in Ohio and its relevant demographic characteristics in 2016

Based on the first round's analysis and selection of variables, it is revealed that **a higher proportion of Latin residents and the ratio of single people contributes to drive the incidence ratio of drug cases up at the statistical significance level of 5%.**



Along with the visualization of the 3-D plot above, a positive correlation between three variables are easy to be verified intuitively and the graph coincides with the result provided by the regression model.

Generalized linear model (Cross sectional data)

On the basis of the findings, we are ready to tap the dataset even further and try to find out more potential dependent factors. However, method of OLS regression sometimes appears to be too rigid to fit the data. Besides, OLS regression is also constrained to the assumption that error must follow the normal distribution. Therefore, aiming to pursue a better fitness and more flexibility on the data, we adopt the approach of generalized linear model.

Step 1

Be cautious about the issue of collinearity:

multicollinearity (also collinearity) is a phenomenon in which one predictor variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy. A multivariate regression model with collinear predictors can indicate how well the entire bundle of predictors predicts the outcome variable, but it may not give valid results about any individual predictor, or about which predictors are redundant with respect to others. **That is, this phenomenon will affect the accuracy of statistical inference though the estimates of predictors remain to be plausible.**

Here, with the help of scatter plot, we are able to find out which predictors are highly correlated, and then we only keep one of those linearly-dependent on each other and remove the rest before conducting the regression to get rid of the effect of multicollinearity.

For instance, since **average household size and average family size are linearly dependent**, the only variable to be added in the regression afterwards of these two would only be the factor — average household size.

Step 2

The result of the generalized linear regression

Method= “Gaussian”   Link function=“Identity”

Smoothing process of variables to achieve a better fitness level

Parametric coefficients:

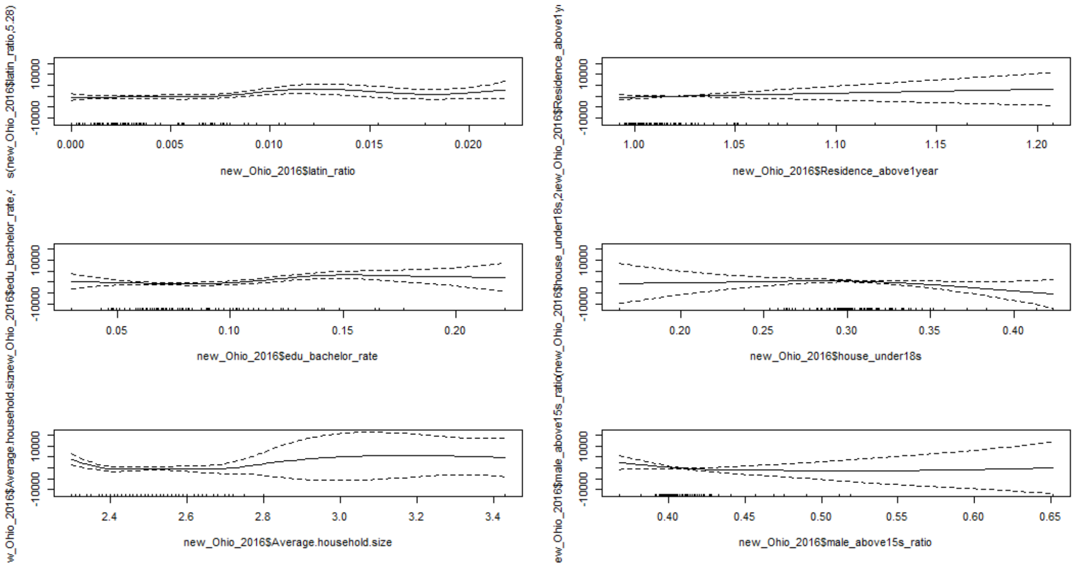
	Estimate	Std. Error	T value	Pr(>|t|)	Significance
(Intercept)	-0.03778	0.01702	-2.219	0.02996	*
Ohio_2016$single_ratio	0.05759	0.019146	3.008	**0.00374**	**

Approximate significance of smooth terms:

	Edf	Ref.df	F	p-value	Significance
s(Ohio_2016$latin_ratio)	5.285	6.343	2.026	**0.07032**	*
s(Ohio_2016$Residence_above1year)	1.000	1.000	0.762	0.38591	
s(Ohio_2016$edu_bachelor_rate)	4.136	4.997	3.919	**0.00356**	**
s(Ohio_2016$house_under18s)	2.182	2.705	1.341	0.27911	
s(Ohio_2016$Average.household.size)	5.516	6.440	2.333	**0.03982**	*
s(Ohio_2016$male_above15s_ratio)	2.777	3.314	1.026	0.35898	
R-square(ajusted)=0.618	Deviance explained=71.4%				

NB: s(variable) indicates the variable has been processed with smoothing effect

The plot below shows the smoothing effect of each variable



According to the result shown from the table above, there are four variables exerting a significant effect to the incidence rate of the drug cases, the ratio of single population, the ratio of Latin American population, the ratio of bachelor-degree holders and the average household size.

Sensitivity analysis

Parametric coefficients:

	Estimate	Std.Error	T value	Pr(>|t|)	Significance
(Intercept)	0.02936	0.01315	-2.232	0.0288	*
new_Ohio_2016$single_ratio	0.48057	0.14721	3.265	0.0017	**

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	Significance
s(new_Ohio_2016$latin_ratio)	5.266	6.343	2.385	0.03333	*
s(new_Ohio_2016$edu_bachelor_rate)	4.402	5.325	4.289	0.00121	**
s(new_Ohio_2016$Average.household.size)	6.487	7.424	3.253	0.00385	**

This is the final model we concluded after a series of variable selection and regularization for each parameter.

Strategy

In the second part, we find that there is a strong relationship between some social factors and the number of drug cases. To be more specific, these factors include ethnicity and singleness. We find that counties with a higher proportion of singles tends to have a higher number of drug cases. Therefore, the single rate can be used as a side reflection of the drug situation. In other words, in counties with high single rates, the number of drug cases is often at a high level or at a high growth rate. Even if neither of them is available, it should also be listed as the focus of surveillance and prevention. **Therefore, the potential place that we list in the previous part may not be enough. Considering the fact that not all the counties with a high single rate have a high number of drug cases, those counties (in the list) with a high single rate but not too many drug cases are also potential place that we should be careful.** Timely deployment can reduce the losses caused by drug cases. In addition, some factors that we think may be relevant do not show correlation. For example, the proportion of higher education is not negatively correlated with the number of cases. Hence, relevant people should adjust their working strategies in time.

In our study, we find that the trend of drug cases in each state is consistent with that in counties where there are more drug cases in the state, and most drug cases in the state are concentrated in several counties. This means that the occurrence of drug cases is centralized. For many reasons, some counties are suitable for the development of each link of drugs, so they become the "center" of drug cases. These "centers" not only promote the further development of drugs, but also have a certain impact on the surrounding. In other words, the development of these "centers" often determines the state's drug development. So if we can control these key areas, the number of drug crimes in the whole state will decline. But in reality, we can't and don't need to reduce the number of cases to zero when controlling drugs. So we will use the modeling method to determine how to control the drug cases in these key areas, so as to control the number of drug cases in the whole state.

In terms of specific control methods, one of the common methods in the control of illicit drugs is to control the circulation of drugs in each link. From production to the final drug dealer, they are all targets of the government. Whether it's going directly into South America to crack down on huge drug gangs or punish trafficking drug on street corner in Philadelphia, they are strict control of drug circulation. Once the circulation of drugs is controlled, people will not be able to get drugs. Following the idea of controlling drug circulation, we can give many specific countermeasures. In the latter part, we will quantify the actual value by differential equation through mathematical modeling.

Controlling opioids from legitimate sources is more complex but necessary. In practice, among new heroin users during 2000 to 2013, approximately three out of four report having misused prescription opioids prior to using heroin. Therefore, it is very important to control the number of opioid prescriptions prescribed by doctors. The best way to achieve this goal is to establish clear opioid drug use standards and regulatory system. In 2013, Pennsylvania providers wrote 88.6 opioid prescriptions per 100 persons. In the same year, the average national rate was 79.3. Since then, opioid prescriptions in the United States have declined, with a 9 percent decline in Pennsylvania from 2013 to 2015, resulting in an estimated 81.1 opioid prescriptions per 100 persons in 2015. In fact, in previous statistics, we found that the number of PA cases in this period is indeed declining.

It is also very important for the management and guidance of drug addicts. We should try to reduce the proportion of drug addicts who take drugs again. On one hand, we can further improve the community correction system. On the other hand, we should respect everyone's selection. We should do more to help those who do not want to take drugs subjectively. They are often exposed to drugs by chance, but they do not want to be addicted to drugs. As a result, such people are less likely to take drugs again once they have given up drugs than those who prefer drugs subjectively.

NYP Model and NYC Model

We create these two models based on differential equation to explore the effectiveness of our strategy. Considering the specific circumstances of previous3 drug cases, we made the following assumptions.

Assumption:

1. In a certain region, drug users and drug dealers are capable of spreading drugs to all non-drug users, all non-drug users are potential drug users and they may start taking drugs in the future.
2. In a certain region, the drugs of new drug users come from drug users or drug dealers, neglecting the possibility of homemade drugs .

Notation

t: time

n (t): The proportion of non-drug users in the total population in a region

y (t): The proportion of drug users in the total population in a region

subscript 0 represents the value of n or y when t equals 0

subscript ∞ represents the value of n or y when t equals infinity

c (t):The proportion of the total population in a region that has completely succeeded in abandoning drug habit , the probability of taking drugs again is 0.

p (t) :The proportion of the total population in a region that succeeded in abandoning drug habit , but they are potential drug users and in the future ,the probability of taking drugs again is 100%.

λ: The average number of non-drug users for per drug user to get along with each day ,non-drug users are strangers for drug user.

μ: The proportion of the total number of drug users who succeeded in abandoning drug habit each day

$1/\mu$: Average drug spread period

$\sigma=\lambda/\mu$: The average number of non-drug users for per drug user to get along with each day ,non-drug users are strangers for drug user and these non-drug users start taking drugs after the exposure to drug users ,mirrors the spread speed of the drug.

N : The total population

The NYP Model（N: non-drug users Y: drug users P: those who succeed in abandoning drug habits and become potential drug users）

Model explanation:

Considering some residents have not completely eliminated the idea of drug abuse after successful drug treatment. We can divide people into three categories. P-those who succeed in abandoning drug habits and become potential drug users are subordinate to N-non-drug users.

Model Composition and calculation: we set up the following equations to explore the parameter bound of success or failure of Strategies

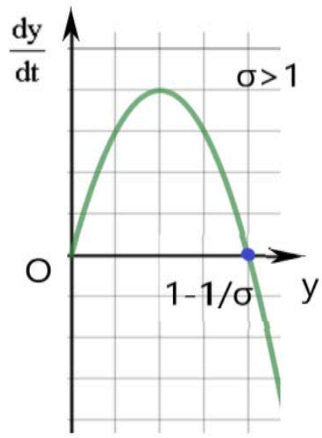
$$\begin{cases} y(t)+n(t)=1 \\ N\dfrac{dy}{dt}=\lambda Nny-\mu Ny \\ \sigma=\lambda/\mu \end{cases}$$

$$\frac{dy}{dt}=\lambda y（1-y）-\mu y \quad y（0）=y_0$$

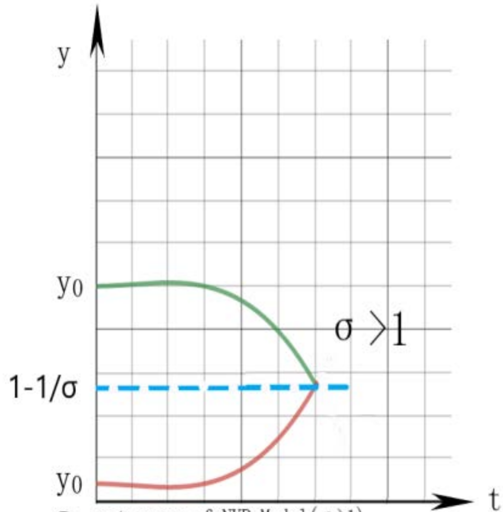
$$\frac{dy}{dt}=-\lambda y[y-(1-\frac{1}{\delta})]$$

Results analysis

Based on the results of the above calculations, we have drawn the following variable diagram

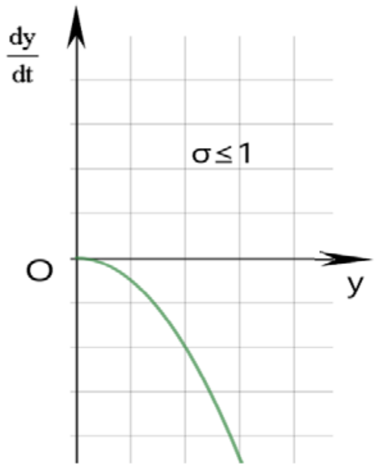


$\frac{dy}{dt}$-y curve of NYP Model (σ>1)

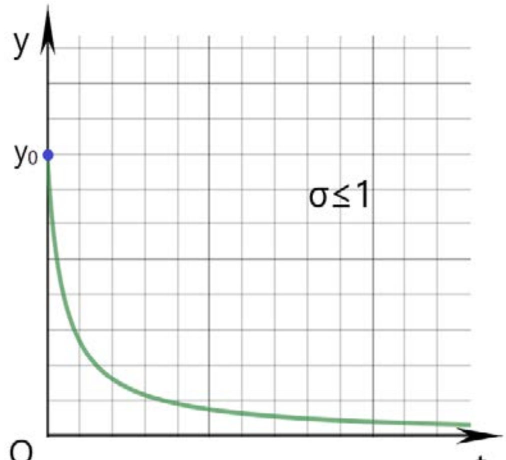


y-t curve of NYP Model(σ>1)

The green line discribes the condition of $y_0 > 1-1/\sigma$



$\frac{dy}{dt}$-y curve of NYP Model (σ≤1)



y-t curve of NYP Model(σ≤1)

Seen from above charts, it can be concluded that $\sigma=1$ is a parameter bound,when $\sigma>1$,the increase or decrease of y（t）depends on i_0 , the limit y（∞）$=1-\frac{1}{\sigma}$ increases with σ.In our strategy, we propose to control the spread of drugs by means of drug transmission. Now we can get the parameter bound of this model. When the number of unfamiliar non-drug users that a drug users or drug deals contacts daily increases, the spread of drugs tends to accelerate, and the strategy is ineffective. The total number of drug users will also rise. when $\sigma\leq1$,The strategy is effective when the number of new addicts does not exceed the number of original drug users.

The NYC Model（N: non-drug users Y:drug users C:those who has completely succeeded in abandoning drug habit）

Model Explanation:

Considering that a certain proportion of successful drug addicts are aware of the dangers of drugs .And the proper management and guidance of concerning department, such drug addicts will never use drugs. We can still divide people into three categories

Assume $n(0)=n_0>0,y(0)=y_0>0$

Model Composition and calculation:

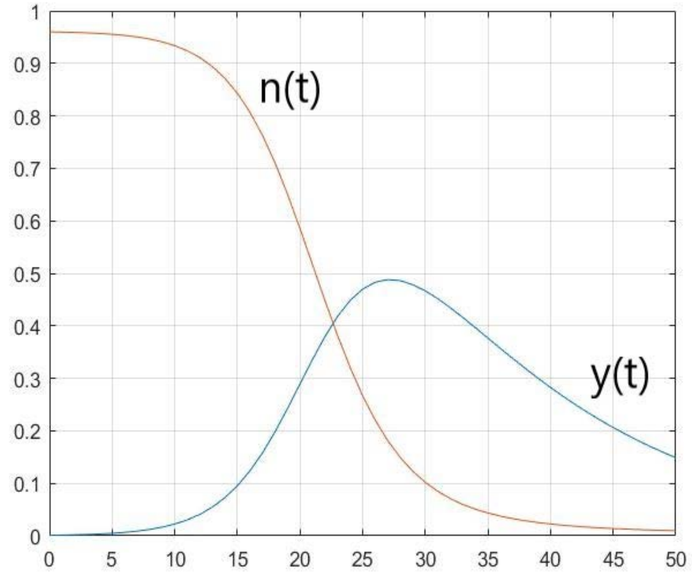
Define following equation group as Equation*:

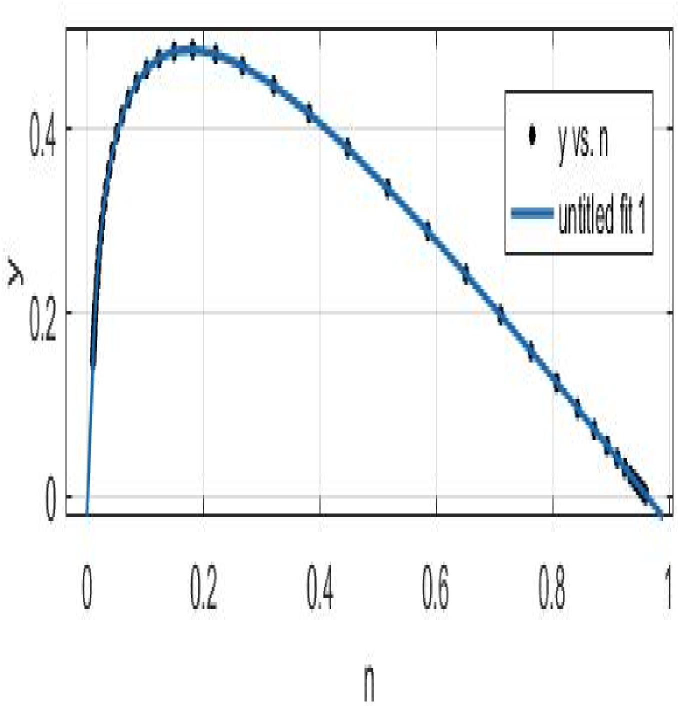
$$\begin{cases} y(t)+n(t)+c(t)=1 & ① \\ N\frac{dc}{dt}=\mu Ny & ② \\ \frac{dy}{dt}=\lambda ny-\mu y & ③ \\ \frac{dn}{dt}=-\lambda ny & ④ \end{cases}$$

σ=λ/μ

By calculating equation① and ②， we conclude equation ③ and ④.

Based on the equation*， we use MATLAB to get the arithmetic solution of the equation. suppose λ=0.4， μ=0.07， y（0）=0.001， n（0）=0.96 , we have drawn the following variable diagram .





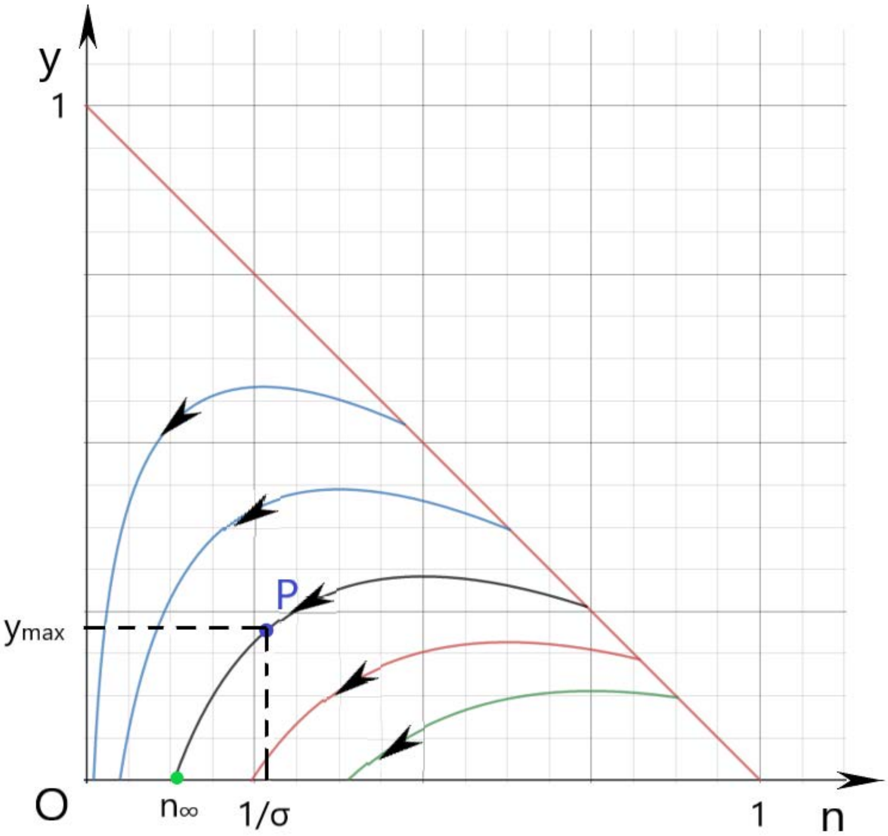
Result analysis

On the basis of numerical calculation and graphical observation, we further explore the relationship of y and n. We eliminate dt in equation* and get the following equation

$$\frac{dy}{dn}=\frac{1}{\sigma n}-1,\quad y|_{n=n0}=y_0$$

$y=n_0+y_0-n+\frac{1}{\sigma}(\ln(n)-\ln(n_0))$ ⑤

.We paint a series phase trajectory based on the equation⑤



the phase trajectory of NYC Model

Analysis of the phase trajectory

The ultimate goal is to control the number of drug users , thereby reducing the number of drug abuse cases.

Assumption: when t gradually approaches infinite y, $y_\infty=0$.

1、Each phase trajectory eventually intersects the n-axis.

2 、The proportion of non-drug users in the end is n_∞,

$$y=n_0+y_0-n+\frac{1}{\sigma}(\ln(n)-\ln(n_0)) \text{ assume } y=0$$

$$n_\infty \text{ is the answer of } n_0+y_0-n_\infty+\frac{1}{\sigma}(\ln(n_\infty)-\ln(n_0)) \text{ in domain } (0, \frac{1}{\sigma})$$

3、、if $n_0>\frac{1}{\sigma}$, y（t）will increase，when $n=\frac{1}{\sigma}$, y(t) reaches the maximum value

$$y_{max}=n_0+y_0-(1+\ln\sigma n_0)$$

4、Based on two blue phase trajectory ，if $n_0\leq\frac{1}{\sigma}$, y（t）monotonically reduced to 0，n（t）monotonically reduced to n_∞。

5、Finally, it can be concluded that when the proportion of drug users is increasing, the spread of drugs and drug cases will increase， $\frac{1}{\sigma}$ is parameter bound， if $n_0 > \frac{1}{\sigma}$， drugs will spread, and strategies will fail. If measures are implemented, σ can be effectively reduced and parameter bound $\frac{1}{\sigma}$， Then the drug will not spread continuously and the strategy is successful.

6 We can control c(t) to determine the parameter bound of our strategy that further improve the community correction system will reduce the probability of former drug users taking drugs again. $s_0=1-r_0$， when $r_0>1-\frac{1}{\sigma}$， we can control the spread of drugs, and the strategy succeeds. On the contrary, the strategy fails.

Test the NYC model

By the control variate method and setting the value of the parameters, and using the result:

$$\begin{cases} n_0+y_0-n_\infty+\frac{1}{\sigma}\left(\ln(n_\infty)-\ln(n_0)\right)=0, \text{ we get the following chart.} \\ y_{max}=n_0+y_0-(1+\ln\sigma n_0) \end{cases}$$

λ	**μ**	**σ**	**1/σ**	**n_0**	**y_0**	**n_∞**	**y_{max}**
1	0.1	10	0.1	0.96	0.04	4.4E-05	0.67382
0.8	0.1	8	0.125	0.96	0.04	0.00032	0.62017
0.6	0.3	2	0.5	0.96	0.04	0.18997	0.17384
0.5	0.3	1.66667	0.6	0.96	0.04	0.29791	0.118
0.4	0.3	1.33333	0.75	0.96	0.04	0.4798	0.06485
1	0.1	10	0.1	0.7	0.04	0.00043	0.44541
0.8	0.1	8	0.125	0.7	0.04	0.00191	0.39965
0.6	0.3	2	0.5	0.7	0.04	0.27766	0.07176
0.5	0.3	1.66667	0.6	0.7	0.04	0.39179	0.04751
0.4	0.3	1.33333	0.75	0.7	0.04	0.52679	0.04174

From the chart, we can see that for a certain n_0, the smaller the λ , the larger the μ, the greater the n_∞ , the smaller the y_{max}, and for a certain λ,μ , n_0, the larger the n_∞, the smaller the y_{max}.

Therefore, the variation rule of the parameters of our model is consistent with the actual situation of prediction, and it has certain feasibility in practical application.

The core parameter σ=λ/μ,in practical application,λ and μ are hard to estimate. Our model can get a solution of estimating σ. Firstly, in the data analysis, we conclude that i is very small, so we assume i_0 =0 , after calculation

$$\sigma=(\ln s_0-\ln s_t)/(s_0-s_t)$$

In practical application，when our strategy is applied to cope with the opioid crisis， the we can choose a certain area we want to reduce the number of drug case and choose a certain period of time，Implementing multiple experiments to get the forecasting value of s_0 and s_t， then we can calculate the value of σ and use NYP Model and NYC Model to handle the opioid crisis.

Conclusion

Strengthens:

Dealing with such a considerably large dataset, we managed to convert the tedious data into vivid plots through visualization with the help of R-studio. Furthermore, through applying statistical knowledge such as regularization and GLM, we succeeded in building several relatively reliable regression models and generated reasonable predictions. Apart from the strategies to be implemented, we also verify its feasibility by introducing newly-designed mathematical models on the platform of Matlab.

Weaknesses:

1. For a lack of time series data (we only have data from 2010 to 2017)
   The accuracy of prediction could be largely improved if we were offered with data across more time lags.
2. For a lack of data about the flow rate of the population
   We failed to design a more advanced model taking dynamic change into account
3. We assume the proportion of initial drug users is zero when calculating the value of spread velocity (σ)

Reference

1. Centers for Disease Control and Prevention. 2018 Annual Surveillance Report of Drug-Related Risks and Outcomes — United States. Surveillance Special Report 2. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. Published August 31, 2018.
2. Peirce GL, Smith MJ, Abate MA, Halverson J. Doctor and Pharmacy Shopping for Controlled Substances. Med Care 2012 Jun; 50(6):494-500.
3. Jones CM, Logan J, Gladden RM, Bohm MK. Demographic and Substance Use Trends Among Heroin Users — United States, 2002–2013. MMWR 2015; 64(26):719-725.
4. www.cdc.gov/drugoverdose/data/heroin
5. Progress in Geographic Science, 37 (8): 1131-1139. Ning Zhang, Dawei Wang.
6. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and Opioid-Involved Overdose Deaths – United States, 2013-2017. Morb Mortal Wkly Rep. ePub: 21 December 2018.
7. Multiple Cause of Death 1999–2017 on CDC Wide-ranging Online Data for Epidemiologic Research (CDC WONDER). Atlanta, GA: CDC, National Center for Health Statistics. 2018. Available at http://wonder.cdc.gov.
8. Hedegaard H, Miniño AM, Warner M. Drug overdose deaths in the United States, 1999–2017. NCHS Data Brief, no 329. Hyattsville, MD: National Center for Health Statistics. 2018.

9. Gladden RM, Martinez P, Seth P. Fentanyl law enforcement submissions and increases in synthetic opioid-Involved overdose deaths – 27 states, 2013-2014. Morb Mortal Wkly Rep. 2016; 65(33):837-43.
10. https://www.drugabuse.gov/drugs-abuse/opioids/opioid-summaries-by-state/pennsylvania-opioid-summary
11. N.P. Bhatia, G.P. Szegö, "Stability theory of dynamical systems" , Springer (1970) pp. 6, 14 A.G. [A.G. Butkovskii] Butkovsky, "Phase portraits of control dynamical systems" , Kluwer (1990) (Translated from Russian)

Appendix:

Table. Opioid Classification

Strong Agonists	Mild to Moderate Agonists	Opioids with Mixed Receptor Actions
Morphine	Acetylcodeine	Nalbuphine
3-Fluorofentanyl	Acetylcodeine	Butorphanol
3-Methylfentanyl	Acetyldihydrocodeine	Buprenorphine
4-Fluoroisobutyryl fentanyl	Codeine	
4-Methylfentanyl	Dextropropoxyphene	
Acetyl fentanyl	Dihydrocodeine	
Acryl fentanyl	Hydrocodone	
Benzylfentanyl	Oxycodone	
Fentanyl	Phenylheptylamines	
Furanyl fentanyl	Propoxyphene	
Hydromorphone		
Isobutyryl fentanyl		
Meperidine		
Methadone		
Methoxyacetyl fentanyl		
Oxymorphone		
Phenylheptylamines		

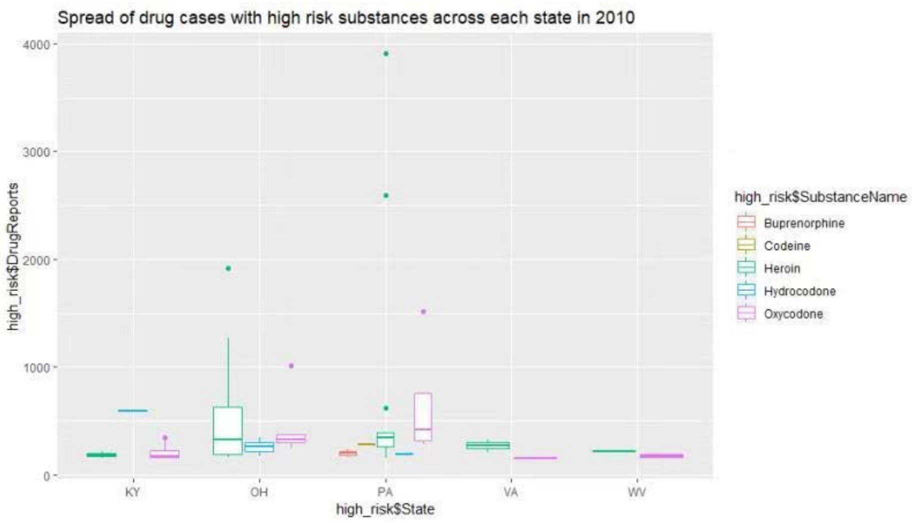
The chapter of drug ratio and single ratio

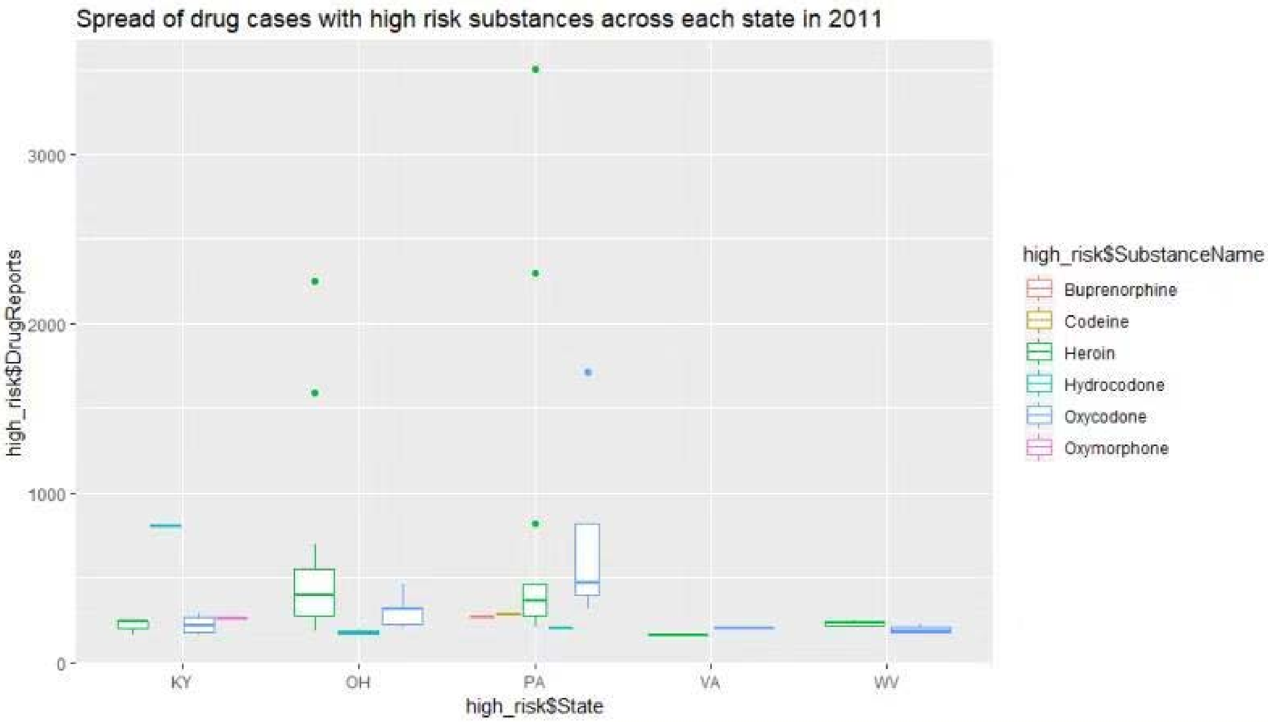
Ohio_2016_county	Ohio_2016.drug_ratio	Ohio_2016.single_ratio
Adams County, Ohio	0.013161216	0.104639412
Allen County, Ohio	0.00708028	0.109693049
Ashland County, Ohio	0.023493999	0.083075867

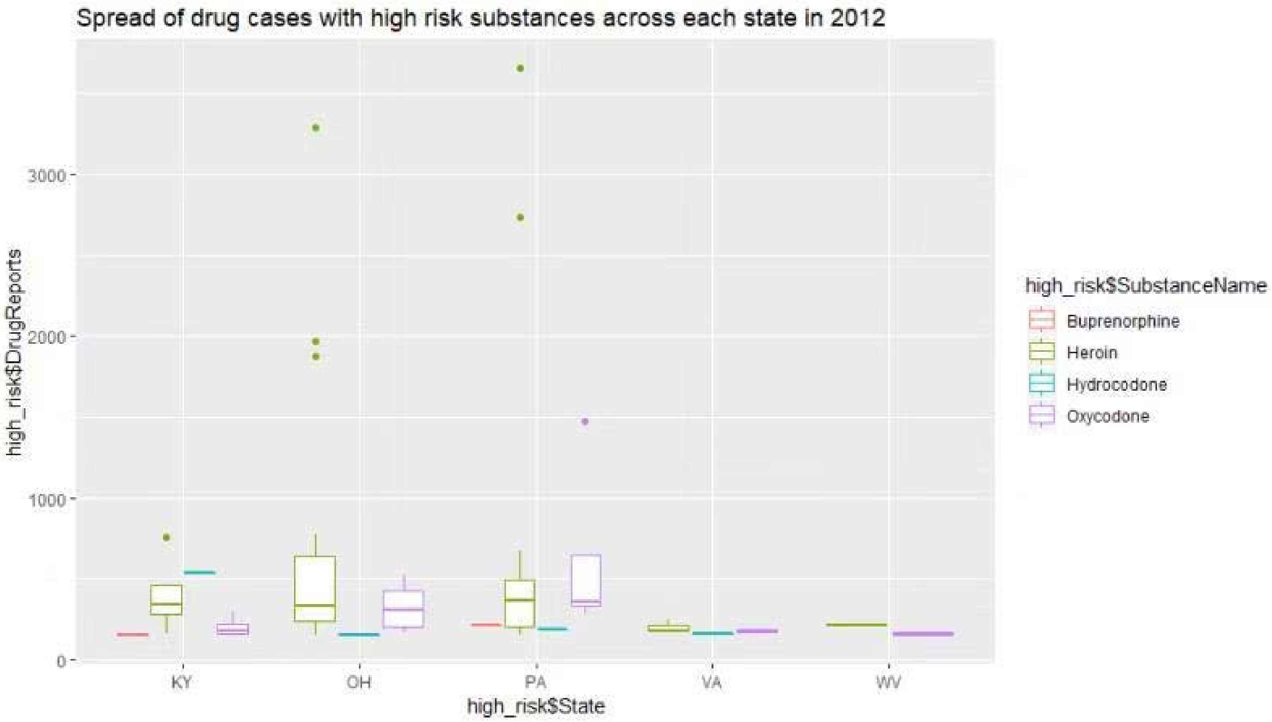
Ashtabula County, Ohio	0.012543117	0.099871134
Athens County, Ohio	0.012596777	0.079362236
Auglaize County, Ohio	0.005534852	0.068893643
Belmont County, Ohio	0.00684238	0.07911438
Brown County, Ohio	0.007819148	0.096791224
Butler County, Ohio	0.007471877	0.094974093
Carroll County, Ohio	0.002024804	0.059056205
Champaign County, Ohio	0.007180344	0.09424878
Clark County, Ohio	0.006878458	0.112817981
Clermont County, Ohio	0.006063435	0.076931256
Clinton County, Ohio	0.01332776	0.104932512
Columbiana County, Ohio	0.012970488	0.089990905
Coshocton County, Ohio	0.005634892	0.069458915
Crawford County, Ohio	0.012277983	0.090785524
Cuyahoga County, Ohio	0.014203201	0.106964807
Darke County, Ohio	0.003145875	0.08285728
Defiance County, Ohio	0.017391304	0.091039147
Delaware County, Ohio	0.004030765	0.066241885
Erie County, Ohio	0.017022248	0.092275693
Fairfield County, Ohio	0.008783004	0.098462041
Fayette County, Ohio	0.009232728	0.12077833
Franklin County, Ohio	0.007961707	0.108889389
Fulton County, Ohio	0.003428653	0.089100985
Gallia County, Ohio	0.017458975	0.084036538
Geauga County, Ohio	0.002718881	0.050214961
Greene County, Ohio	0.008528606	0.076183513
Guernsey County, Ohio	0.018088386	0.089201286
Hamilton County, Ohio	0.02206196	0.108998789
Hancock County, Ohio	0.01275327	0.085517766

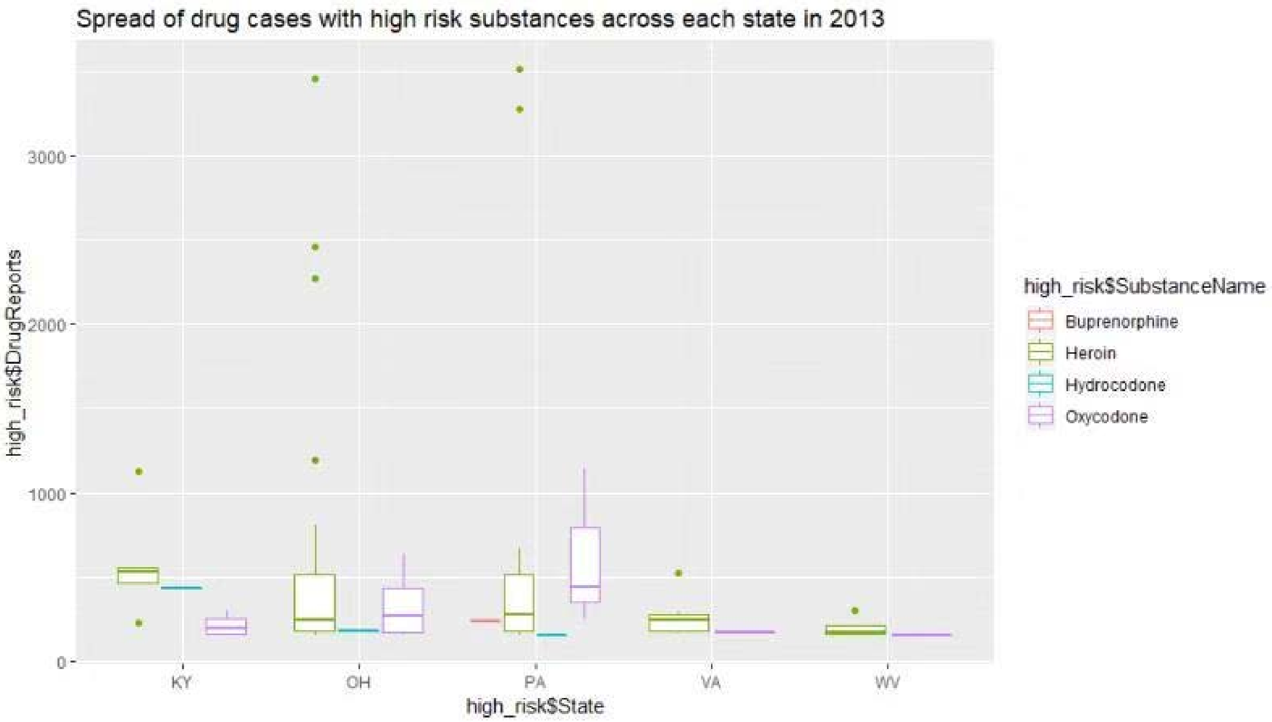
Hardin County, Ohio	0.006344658	0.069468416
Harrison County, Ohio	0.003211641	0.074495354
Henry County, Ohio	0.000983141	0.095967377
Highland County, Ohio	0.008903611	0.097986174
Hocking County, Ohio	0.010778016	0.099593855
Holmes County, Ohio	0.00172764	0.0424
Huron County, Ohio	0.008227465	0.092846676
Jackson County, Ohio	0.023142283	0.108319026
Jefferson County, Ohio	0.004113205	0.088889689
Knox County, Ohio	0.004234851	0.081940162
Lake County, Ohio	0.02587206	0.078499512
Lawrence County, Ohio	0.005426114	0.078389192
Licking County, Ohio	0.008541243	0.092372802
Logan County, Ohio	0.005733716	0.091025641
Lorain County, Ohio	0.009036748	0.105536141
Lucas County, Ohio	0.004766276	0.123366036
Madison County, Ohio	0.012169752	0.101492336
Mahoning County, Ohio	0.009539035	0.100293381
Marion County, Ohio	0.012512841	0.091326071
Medina County, Ohio	0.005138009	0.06632062
Meigs County, Ohio	0.003547633	0.093763581
Mercer County, Ohio	0.002945545	0.072692284
Miami County, Ohio	0.007211515	0.088210967
Monroe County, Ohio	0.004339609	0.094721016
Montgomery County, Ohio	0.016982247	0.113752206
Morgan County, Ohio	0.001090067	0.084962913
Morrow County, Ohio	0.007489126	0.074091702
Muskingum County, Ohio	0.007101056	0.104921294
Noble County, Ohio	0.003127113	0.038659264
Ottawa County, Ohio	0.0046507	0.078060285
Paulding County, Ohio	0.005218766	0.06763285
Perry County, Ohio	0.006402876	0.0976727

Pickaway County, Ohio	0.003637129	0.085797736
Pike County, Ohio	0.005466053	0.105977264
Portage County, Ohio	0.010420159	0.080717707
Preble County, Ohio	0.006368498	0.086677468
Putnam County, Ohio	0.001271926	0.060494959
Richland County, Ohio	0.007378113	0.097945134
Ross County, Ohio	0.007647835	0.105305921
Sandusky County, Ohio	0.013464394	0.108848137
Scioto County, Ohio	0.010471204	0.095204513
Seneca County, Ohio	0.006847496	0.101878308
Shelby County, Ohio	0.008374242	0.097568917
Stark County, Ohio	0.012853231	0.100025811
Summit County, Ohio	0.003745748	0.091573639
Trumbull County, Ohio	0.00747242	0.09758635
Tuscarawas County, Ohio	0.004350012	0.078210599
Union County, Ohio	0.005121235	0.081604426
Van Wert County, Ohio	0.004400284	0.082873895
Vinton County, Ohio	0.013968839	0.108458114
Warren County, Ohio	0.006964441	0.062466967
Washington County, Ohio	0.009754457	0.0740931
Wayne County, Ohio	0.007558316	0.075469944
Williams County, Ohio	0.001083123	0.08708629
Wood County, Ohio	0.0049583	0.072403798
Wyandot County, Ohio	0.006660625	0.081583552

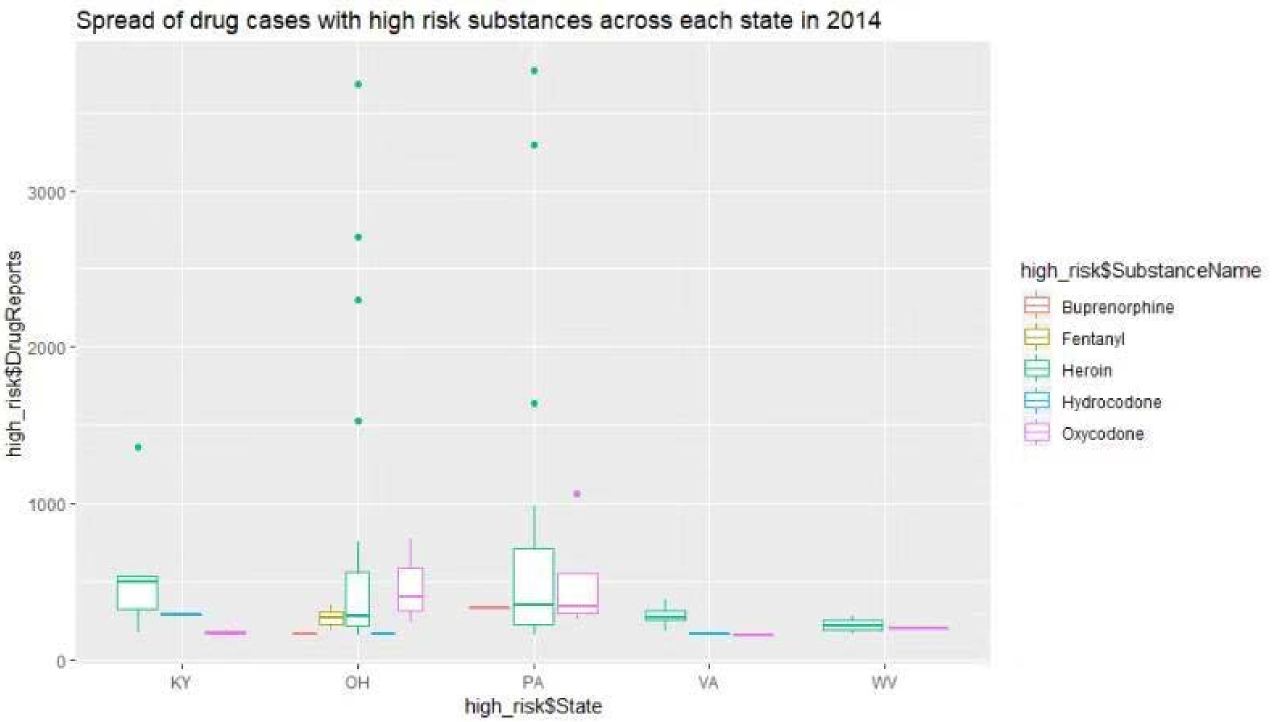

Spread of drug cases with high risk substances across each state in 2010
4000
3000
2000
1000
0
high_risk$DrugReports
KY
OH
PA
VA
WV
high_risk$State
high_risk$SubstanceName
Buprenorphine
Codeine
Heroin
Hydrocodone
Oxycodone


Spread of drug cases with high risk substances across each state in 2011
high_risk$DrugReports
0
1000
2000
3000
KY
OH
PA
VA
WV
high_risk$State
high_risk$SubstanceName
Buprenorphine
Codeine
Heroin
Hydrocodone
Oxycodone
Oxymorphone

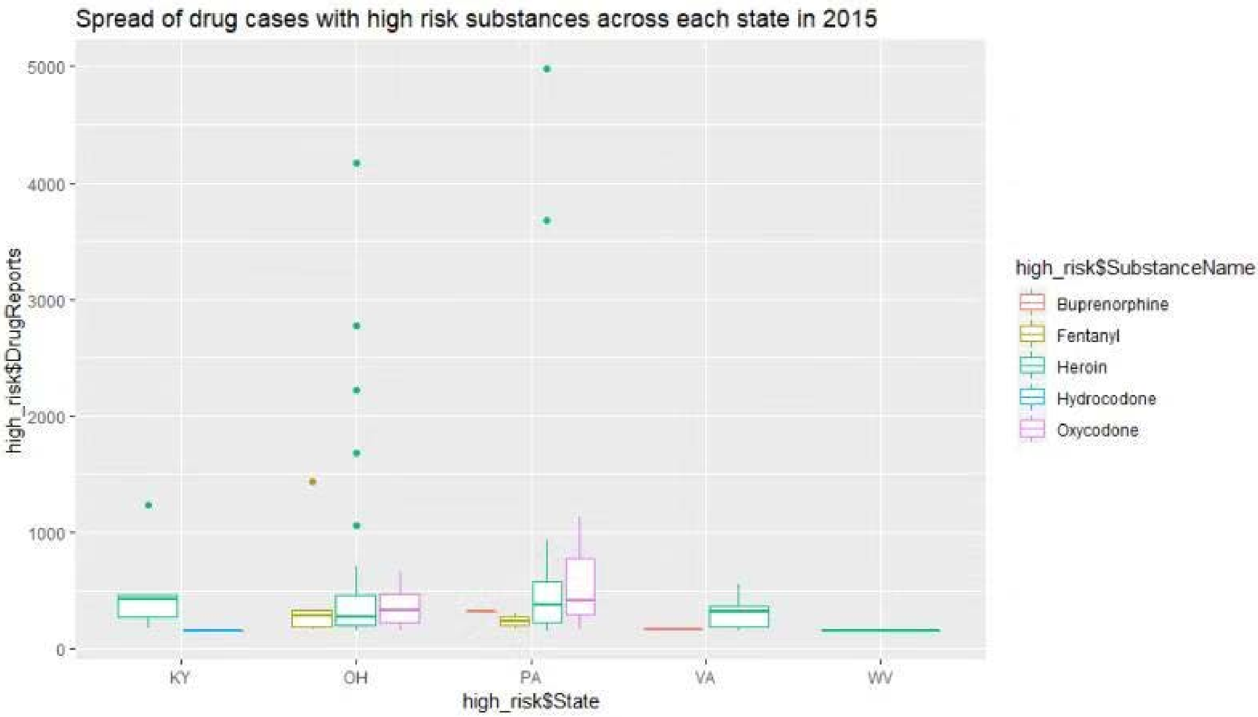

Spread of drug cases with high risk substances across each state in 2012
high_risk$DrugReports
0
1000
2000
3000
KY
OH
PA
VA
WV
high_risk$State
high_risk$SubstanceName
Buprenorphine
Heroin
Hydrocodone
Oxycodone

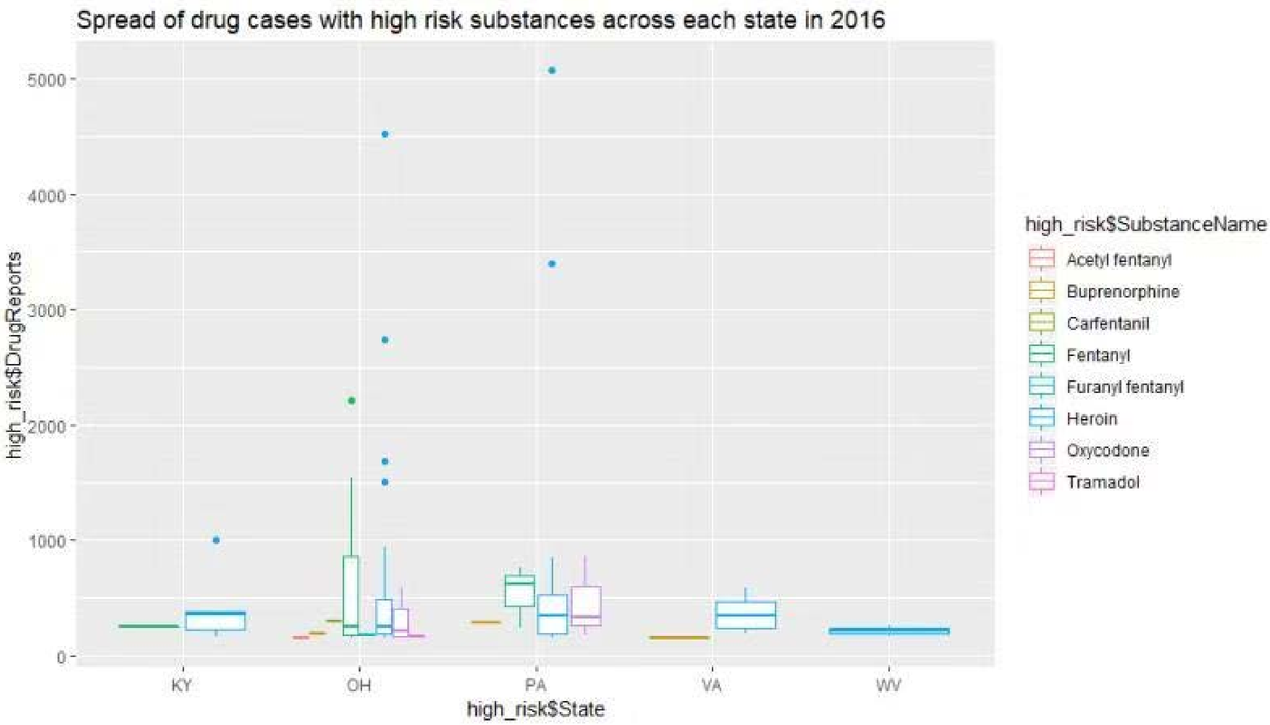


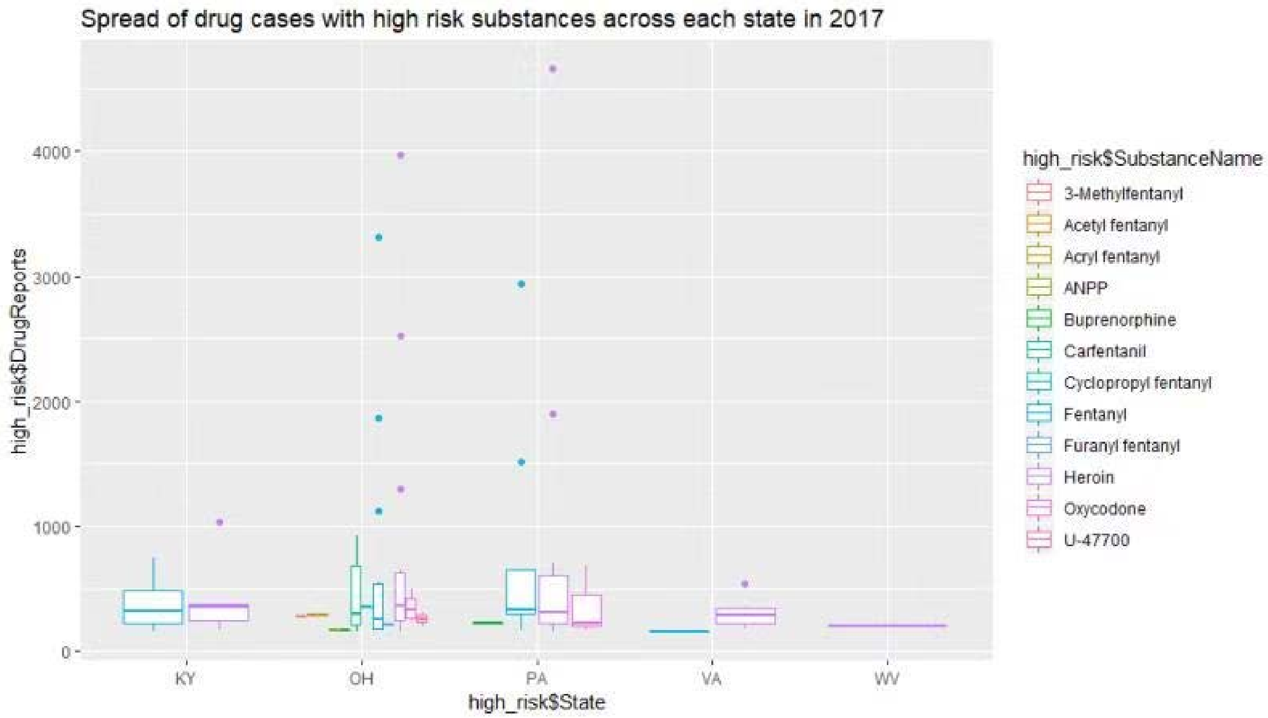
Spread of drug cases with high risk substances across each state in 2013
high_risk$DrugReports
3000
2000
1000
0
KY
OH
PA
VA
WV
high_risk$State
high_risk$SubstanceName
Buprenorphine
Heroin
Hydrocodone
Oxycodone



Spread of drug cases with high risk substances across each state in 2014
high_risk$DrugReports
3000
2000
1000
0
KY
OH
PA
VA
WV
high_risk$State
high_risk$SubstanceName
Buprenorphine
Fentanyl
Heroin
Hydrocodone
Oxycodone


Spread of drug cases with high risk substances across each state in 2015
high_risk$DrugReports
0
1000
2000
3000
4000
5000
KY
OH
PA
VA
WV
high_risk$State
high_risk$SubstanceName
Buprenorphine
Fentanyl
Heroin
Hydrocodone
Oxycodone


Spread of drug cases with high risk substances across each state in 2016
high_risk$DrugReports
0
1000
2000
3000
4000
5000
KY
OH
PA
VA
WV
high_risk$State
high_risk$SubstanceName
Acetyl fentanyl
Buprenorphine
Carfentanil
Fentanyl
Furanyl fentanyl
Heroin
Oxycodone
Tramadol



code of R-studio

```
### installation and overview of the data
summary(df)
str(df)
install.packages("ggplot2")
library("ggplot2")
install.packages("dplyr")
library("dplyr")
install.packages("tseries")
library("tseries")
install.packages("forecast")
library("forecast")
install.packages("dendextend")
library("dendextend")
install.packages("scatterplot3d")
library("scatterplot3d")
install.packages("GGally")
library(GGally)
install.packages("gam")
library("gam")


### heroin vs state
ggplot(df,aes(x=YYYY,y=DrugReports, col=SubstanceName))+
```

```
  geom_point(alpha=0.4)
heroin<-filter(df,SubstanceName=='Heroin')
ggplot(df,aes(x=YYYY,y=DrugReports,col=State))+
  geom_point(alpha=0.4)
### to be modified

### case per capita versus state

case_percapita<-Num$NUM/Num$POP/100
million_capita<-Num$NUM/100
cases_per_million_people<-Num$NUM/million_capita
Num<-data.frame(Num,million_capita)
ggplot(Num,aes(x = YEAR, y = cases_per_million_people, col = STATE)) +
  geom_point()


### time series modelling of state OH
Ohio<-filter(State,STATE=='OH')
plot(Ohio$YEAR,Ohio$NUM)
AR <- arima(Ohio$NUM,order=c(1,1,0))
print(AR)

### time series
Ohio_ts<-ts(Ohio$NUM,frequency = 1,start = 2010)
tsdisplay(Ohio_ts)
dc<-decompose(Ohio_ts)
plot(dc)

###histogram by county
high_risk_total<-filter(df,DrugReports>150)
ggplot(high_risk,aes(x= DrugReports , col = SubstanceName)) +
  geom_histogram()

### boxplot
high_risk<-filter(df,DrugReports>150 & YYYY==2017)
ggplot(data=high_risk,aes(x=high_risk$State,y=high_risk$DrugReports))+geom_boxplot(aes(col=high_risk$SubstanceName))+
  labs(title="Spread of drug cases with high risk substances across each state in 2017")

### Heroin vs Heroin
oxycodone<-filter(df,SubstanceName=="Oxycodone")
```

```
# selct the county with both Heroin and Oxycodone
com<-filter(df,SubstanceName=="Oxycodone" | SubstanceName=="Heroin")
com_high<-filter(com,DrugReports>30)
com_high_2017<-filter(com_high,YYYY==2017)
### clustering
com1<-filter(com,SubstanceName=="Oxycodone")
com2<-filter(com,SubstanceName=="Heroin")
S<-data.frame(com1$DrugReports,com2$DrugReports)
colnames(S)<-c("Oxycodone","Heroin")

dist<-dist(S,method='euclidean')
hc_dist<-hclust(dist,method="complete")
dend_cluster<-as.dendrogram(hc_dist)
plot(dend_cluster)
dend_500<-color_branches(dend_cluster,h=500)
plot(dend_500)

cluster_500<-cutree(hc_dist,h=500)
Com_S<-data.frame(S,cluster_500)
ggplot(Com_S,aes(Com_S$Oxycodone,Com_S$Heroin,col=factor(cluster_500)))+
  geom_point()+
  coord_cartesian(xlim=c(0,200),ylim=c(0,700))

### scatter plot matrix and regression

fmla <- Ohio_2016$TotalDrugReportsCounty~Ohio_2016$`HC01_VC03Estimate; HOUSEHOLDS BY
TYPE - Total households`
drug_model<-lm(fmla, data = Ohio_2016)
drug_model
summary(drug_model)

drug_model$prediction <- predict(drug_model)

# plot the results
ggplot(drug_model, aes(x = drug_model$prediction, y = Ohio_2016$TotalDrugReportsCounty)) +
  geom_point() +
  geom_abline(color = "blue")

drug_ratio<-Ohio_2016$TotalDrugReportsCounty/Ohio_2016$`HC01_VC03Estimate; HOUSEHOLDS
BY TYPE - Total households`
new_Ohio_2016<-data.frame(Ohio_2016,drug_ratio)
```

```
###scatterplotMatrix(~ Ohio_2016$ + disp + drat + wt, data=mtcars, spread=FALSE,
          lty.smooth=2, main="Scatterplot Matrix via features of household structure")


fmla2<-
new_Ohio_2016$drug_ratio~new_Ohio_2016$HC01_VC92Estimate..EDUCATIONAL.ATTAINMENT...Population.25.years.and.over...Graduate.or.professional.degree+new_Ohio_2016$HC01_VC154Estimate..YEAR.OF.ENTRY...Foreign.born+new_Ohio_2016$HC01_VC121Estimate..RESIDENCE.1.YEAR.AGO...Population.1.year.and.over...Different.house.in.the.U.S.
data_lm<-lm(fmla2, data = new_Ohio_2016)
data_lm
summary(data_lm)

fit1_LM <- stepAIC(data_lm, direction = 'backward')

drug_ratio<-Ohio_2016$TotalDrugReportsCounty/Ohio_2016$`Total Population`
new_Ohio_2016<-data.frame(Ohio_2016,drug_ratio)
fmla1<-
new_Ohio_2016$drug_ratio~new_Ohio_2016$Total.households...Nonfamily.households...Householder.living.alone+new_Ohio_2016$Population.25.years.and.over...Bachelor.s.degree+new_Ohio_2016$WORLD.REGION.OF.BIRTH.OF.FOREIGN.BORN...Foreign.born.population..excluding.population.born.at.sea...Latin.America+new_Ohio_2016$Total.households...Family.households..families....Male.householder..no.wife.present..family
reg<-lm(fmla1,data=new_Ohio_2016)
summary(reg)

latin_ratio<-Ohio_2016$`WORLD REGION OF BIRTH OF FOREIGN BORN - Foreign-born population, excluding population born at sea - Latin America`/Ohio_2016$`Total Population`
single_ratio<-(Ohio_2016$`Total households - Family households (families) - Male householder, no wife present, family - With own children of the householder under 18 years`+new_Ohio_2016$Total.households...Family.households..families....Female.householder..no.husband.present..family...With.own.children.of.the.householder.under.18.years)/new_Ohio_2016$Total.households
new_Ohio_2016<-data.frame(Ohio_2016,drug_ratio,latin_ratio,single_ratio)


### regression
fmla2<-new_Ohio_2016$drug_ratio~new_Ohio_2016$latin_ratio+new_Ohio_2016$single_ratio
reg<-lm(fmla2,data=new_Ohio_2016)
summary(reg)
```

```
### scatter plot
scatterplot3d(new_Ohio_2016$drug_ratio,new_Ohio_2016$single_ratio,new_Ohio_2016$latin_rati
o, pch = 16, color="steelblue")

scatterplotMatrix(~ new_Ohio_2016$Total.households + new_Ohio_2016$Average.family.size +
new_Ohio_2016$Average.household.size +
new_Ohio_2016$WORLD.REGION.OF.BIRTH.OF.FOREIGN.BORN...Foreign.born.population..excluding
.population.born.at.sea | cyl, data=new_Ohio_2016, spread=FALSE,
            main="Scatterplot Matrix", diagonal="histogram")

library(GGally)
S2<-data.frame(new_Ohio_2016$Total.households, new_Ohio_2016$Average.family.size,
new_Ohio_2016$Average.household.size,
new_Ohio_2016$WORLD.REGION.OF.BIRTH.OF.FOREIGN.BORN...Foreign.born.population..excluding
.population.born.at.sea)
ggpairs(S2, aes(col="blue",alpha = 0.4))

### multivariable regression
edu_bachelor_rate<-
new_Ohio_2016$Population.25.years.and.over...Bachelor.s.degree/new_Ohio_2016$Total.Populati
on
house_under18s<-
new_Ohio_2016$Households.with.one.or.more.people.under.18.years/new_Ohio_2016$Total.hous
eholds
new_Ohio_2016<-data.frame(new_Ohio_2016,edu_bachelor_rate,house_under18s)
fmla3<-
new_Ohio_2016$drug_ratio~new_Ohio_2016$latin_ratio+new_Ohio_2016$single_ratio+new_Ohio
_2016$edu_bachelor_rate+new_Ohio_2016$house_under18s
reg2<-lm(fmla3,data=new_Ohio_2016)
summary(reg2)


### generalized additive model
male_above15s_ratio<-
new_Ohio_2016$Males.15.years.and.over/new_Ohio_2016$Total.Population
new_Ohio_2016<-data.frame(new_Ohio_2016,male_above15s_ratio)

require(mgcv)
gam1<-
gam(new_Ohio_2016$TotalDrugReportsCounty~s(new_Ohio_2016$latin_ratio)+new_Ohio_2016$si
```

```
ngle_ratio+s(new_Ohio_2016$edu_bachelor_rate)+s(new_Ohio_2016$house_under18s)+s(new_Ohio_2016$Average.household.size)+
       s(new_Ohio_2016$male_above15s_ratio),
      family=gaussian(link=identity),data=new_Ohio_2016)

gam2<-
gam(new_Ohio_2016$TotalDrugReportsCounty~new_Ohio_2016$latin_ratio+new_Ohio_2016$single_ratio+new_Ohio_2016$edu_bachelor_rate+new_Ohio_2016$house_under18s+new_Ohio_2016$Average.household.size+
      new_Ohio_2016$male_above15s_ratio,
     family=gaussian(link=identity),data=new_Ohio_2016)

Residence_above1year<-
new_Ohio_2016$RESIDENCE.1.YEAR.AGO...Population.1.year.and.over/new_Ohio_2016$Total.Population
new_Ohio_2016<-data.frame(new_Ohio_2016,Residence_above1year)

gam3<-
gam(new_Ohio_2016$TotalDrugReportsCounty~s(new_Ohio_2016$latin_ratio)+new_Ohio_2016$single_ratio+s(new_Ohio_2016$Residence_above1year)+s(new_Ohio_2016$edu_bachelor_rate)+s(new_Ohio_2016$house_under18s)+s(new_Ohio_2016$Average.household.size)+
      s(new_Ohio_2016$male_above15s_ratio),
     family=gaussian(link=identity),data=new_Ohio_2016)

gam4<-
gam(new_Ohio_2016$TotalDrugReportsCounty~new_Ohio_2016$s(latin_ratio)+new_Ohio_2016$single_ratio+s(new_Ohio_2016$edu_bachelor_rate)+s(new_Ohio_2016$Average.household.size),
     family=gaussian(link=identity),data=new_Ohio_2016)

summary(gam4)
par(mfrow=c(3,2)) #to partition the Plotting Window
plot(gam4,se = TRUE)


### find out the conties with high drug cases in WV
df<-MCM_NFLIS_Data
WV<-filter(df,df$State =="WV")
WV_high_county<-filter(WV,WV$TotalDrugReportsCounty>500)
### find out the conties with high drug cases in OH
OH<-filter(df,df$State =="OH")
OH_high_county<-filter(OH,OH$TotalDrugReportsCounty>1500)
```

```
high_county<-filter(df,df$TotalDrugReportsCounty>5000)
### pick out philadelphia
Philad<-filter(df,COUNTY=="PHILADELPHIA")
### pick out Cuyahoga
Cuya<-filter(df,COUNTY=="CUYAHOGA")
### pick out Hamilton
Hami<-filter(df,COUNTY=="HAMILTON")


### drug ratio versus population
fmla5<-new_Ohio_2016$drug_ratio~new_Ohio_2016$Total.Population
model5<-lm(fmla5,data=Ohio_2016)
summary(model5)
ggplot(new_Ohio_2016,aes(new_Ohio_2016$Total.Population,drug_ratio))+
  geom_point()+
  geom_smooth()


### ggplot across years
case_population<-df2$Drug/df2$Population
df2<-data.frame(df2,case_population)
ggplot(df2,aes(df2$Year,df2$case_population,col=df2$County))+
  geom_point(size=2.8)+
  geom_line(size=0.5)+
  xlab("year")+
  ylab("drug cases per million people")

### time series of three counties
### Cuyahoga
par(mfrow=c(1,1))
ts<-ts(df2$case_population[1:8],start=2010,frequency=1)
acf(ts)
pacf(ts)
ts1_model<- arima(ts,order=c(1,1,0))
print(ts1_model)
ts.plot(ts,ylab="drug cases per million people in Cuyahoga")
ts1_fitted <- ts - residuals(ts1_model)
points(ts1_fitted, type = "l", col = 2, lty = 2)
predict_ts1 <- predict(ts1_model)
predict_ts1$pred[1]
```

```
ts1_forecast <- predict(ts1_model, n.ahead = 2)$pred
ts1_forecast_se <- predict(ts1_model, n.ahead = 2)$se
ts.plot(ts, xlim = c(2010, 2019),ylim=c(0,30000))
points(ts1_forecast, type = "l", col = 2)
points(ts1_forecast - 2*ts1_forecast_se, type = "l", col = 2, lty = 2)
points(ts1_forecast + 2*ts1_forecast_se, type = "l", col = 2, lty = 2)

### Hamilton
ts2<-ts(df2$case_population[9:16],start=2010,frequency=1)
ts2_model<- arima(ts,order=c(1,1,0))
print(ts2_model)
ts.plot(ts2)
ts2_fitted <- ts2 - residuals(ts2_model)
points(ts2_fitted, type = "l", col = 2, lty = 2)
predict_ts2 <- predict(ts2_model)
predict_ts2$pred[1]

ts2_forecast <- predict(ts2_model, n.ahead = 2)$pred
ts2_forecast_se <- predict(ts2_model, n.ahead = 2)$se
ts.plot(ts2, xlim = c(2010, 2019),ylim=c(0,30000))
points(ts2_forecast, type = "l", col = 2)
points(ts2_forecast - 2*ts1_forecast_se, type = "l", col = 2, lty = 2)
points(ts2_forecast + 2*ts1_forecast_se, type = "l", col = 2, lty = 2)

### Philadephia
ts3<-ts(df2$case_population[17:24],start=2010,frequency=1)
ts3_model<- arima(ts,order=c(1,1,0))
print(ts3_model)
ts.plot(ts3)
ts3_fitted <- ts3 - residuals(ts3_model)
points(ts3_fitted, type = "l", col = 2, lty = 2)
predict_ts3 <- predict(ts3_model)
predict_ts3$pred[1]

ts3_forecast <- predict(ts3_model, n.ahead = 2)$pred
ts3_forecast_se <- predict(ts3_model, n.ahead = 2)$se
ts.plot(ts3, xlim = c(2010, 2019),ylim=c(0,30000))
points(ts3_forecast, type = "l", col = 2)
points(ts3_forecast - 2*ts1_forecast_se, type = "l", col = 2, lty = 2)
points(ts3_forecast + 2*ts1_forecast_se, type = "l", col = 2, lty = 2)
```

MATLAB Code

```
function y=ill(t,x)
a=0.4;b=0.07;
y=[a*x(1)*x(2)-b*x(1),-a*x(1)*x(2)]';
Order
ts=0:50;
x0=[0.001,0.96];
[t,x]=ode45('ill',ts,x0);[t,x]
plot(t,x(:,1),t,x(:,2)),grid,pause
plot(x(:,2),x(:,1)),grid,
```

Results

t	0	1	2	3	4	5	6	7	8
y	0.001	0.001369	0.001873	0.002563	0.003505	0.00479	0.006542	0.008927	0.012164
n	0.96	0.959549	0.958932	0.958089	0.956936	0.955362	0.953217	0.950295	0.946326
t	9	10	11	12	13	14	15	16	17
y	0.016542	0.022442	0.030341	0.040834	0.054647	0.072548	0.095306	0.123627	0.157908
n	0.940951	0.933696	0.923963	0.910996	0.893863	0.871536	0.842937	0.806988	0.762881
t	18	19	20	21	22	23	24	25	26
y	0.197803	0.242257	0.289345	0.336163	0.37994	0.418126	0.448489	0.470103	0.483098
n	0.710568	0.650736	0.585053	0.516338	0.447472	0.381306	0.320552	0.266723	0.220321
t	27	28	29	30	31	32	33	34	35
y	0.488065	0.486043	0.478444	0.466562	0.451568	0.434416	0.415828	0.396354	0.376568
n	0.181331	0.149227	0.123052	0.101867	0.084718	0.07083	0.059628	0.050673	0.043414
t	36	37	38	39	40	41	42	43	44
y	0.356827	0.337396	0.318462	0.300145	0.282536	0.265687	0.249627	0.234368	0.219906
n	0.037488	0.032618	0.028594	0.025264	0.022485	0.020151	0.018179	0.016501	0.015068
t	45	46	47	48	49	50			
y	0.206227	0.19331	0.181129	0.169659	0.158867	0.148722			
n	0.013837	0.012775	0.011854	0.011051	0.010349	0.009731			