For office use only	Team Control Number	For office use only
T1 _____	1923326	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	C	F4 _____

2019

MCM/ICM

Summary Sheet

Research on the opioid abuse crisis in the United States —Based on visual data processing generalized linear model

Summary

The United States is experiencing a national crisis of the opioid abuse, which has caused enormous and growing damage to the physical and mental health of the American citizens, as well as to the American social economy. Therefore, to find out the occurrence characteristics of opioid abuse and the corresponding control methods is an urgent task in today's American society. The main content of this paper is to use the visual data processing model and generalized linear model, combine the data of the census bureau given in the title to make corresponding explanations of the drug abuse situation in corresponding states and counties, and finally give the best strategy to fight the opioid crisis.

The model provided in this paper can be used to better understand the relationship between drug abuse and the situation of local residents, and to generate targeted countermeasures to curb the outbreak of regional drug abuse crisis.

Key words: Opioid abuse crisis

Occurrence characteristics

The census data

Generalized linear mode

1、Introduction

The United States is experiencing a national crisis regarding the use of synthetic and non synthetic opioids, either for the treatment and management of pain (legal, prescription use) or for recreational purposes (illegal, non-prescription use). Federal organizations such as the Centers for Disease Control (CDC) are struggling to save lives and prevent negative health effects of this epidemic, such as opioid use disorder, hepatitis, and HIV infections, and neonatal abstinence syndrome.There are also implications for important sectors of the U.S. economy as well. For example, if the opioid crisis spreads to all cross-sections of the U.S. population (including the college-educated and those with advanced degrees), businesses requiring precision labor skills, high technology component assembly, and sensitive trust or security relationships with clients and customers might have difficulty filling these positions.

For this problem, we focus on the individual counties located in five (5) U.S. states: Ohio, Kentucky, West Virginia, Virginia, and Tennessee.

The first file Supplied with this problem description (NFLIS Data) contains drug identification counts in years 2010-2017 for narcotic analgesics (synthetic opioids) and heroin in each of the counties from these five states. The additional seven files are zipped folders containing extracts that represent a common set of socio-economic factors collected for the counties of these five states during each of the year

According to the situation above, this paper focuses on solving the following problems.

1 ） Using the NFLIS data provided, analyze the spread and characteristics of the reported synthetic opioid and heroin incidents in and between the five states and their counties over time. Also, identify any possible locations where specific opioid use might have started in each of the five states.

2）Using the U.S. Census socio-economic data provided, find the U.S. Census socio-economic data associated with the use or trends-in-use of heroin, and modify your theory from Part 1 to include any important factors from this data set.

3）Identify a possible strategy for countering the opioid crisis. Use the theory to test the effectiveness of this strategy.

2、Assumption

1) The sample data are all true and accurate and ignore possible data fluctuations.

2) The county location data are correct as provided.

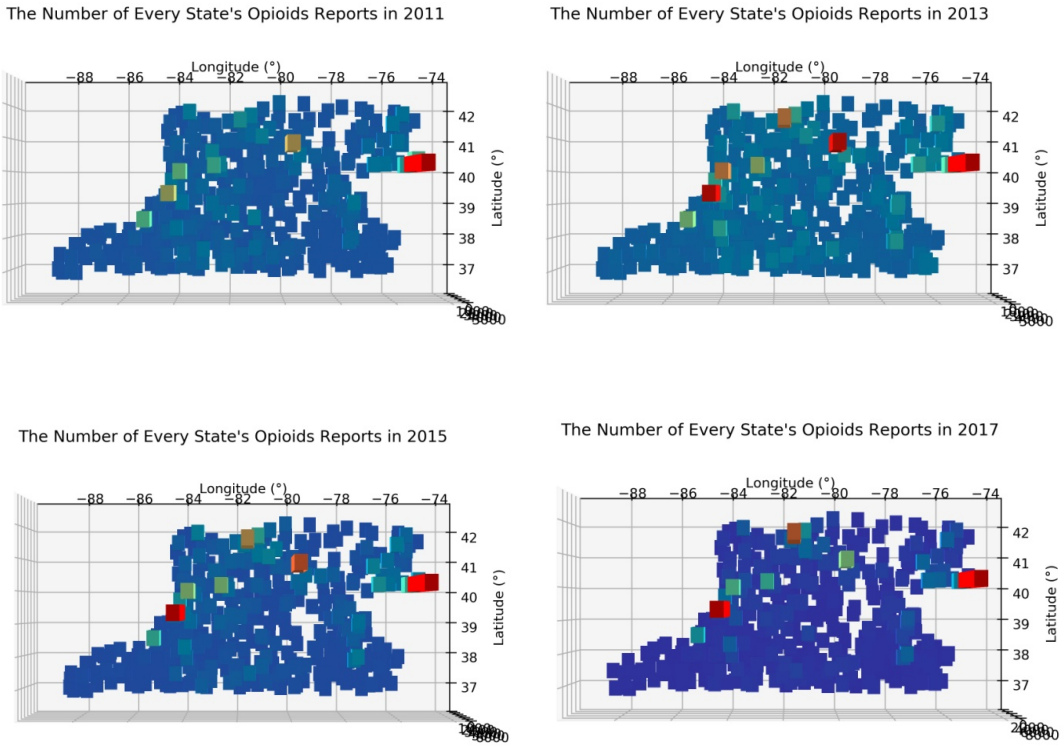
3) Ignore the effect of population movements between states and counties on outcomes.

4) The natural rate of population growth is zero. (the population remains constant)

3、Model

3.1 Spread and characteristics study based on visual data processing

The total number of reports of all kinds of drug abuse incidents in all counties in 2011, 2013, 2015 and 2017 was calculated based on the data. Combined with the geographical coordinates of all counties, it was plotted as follows:

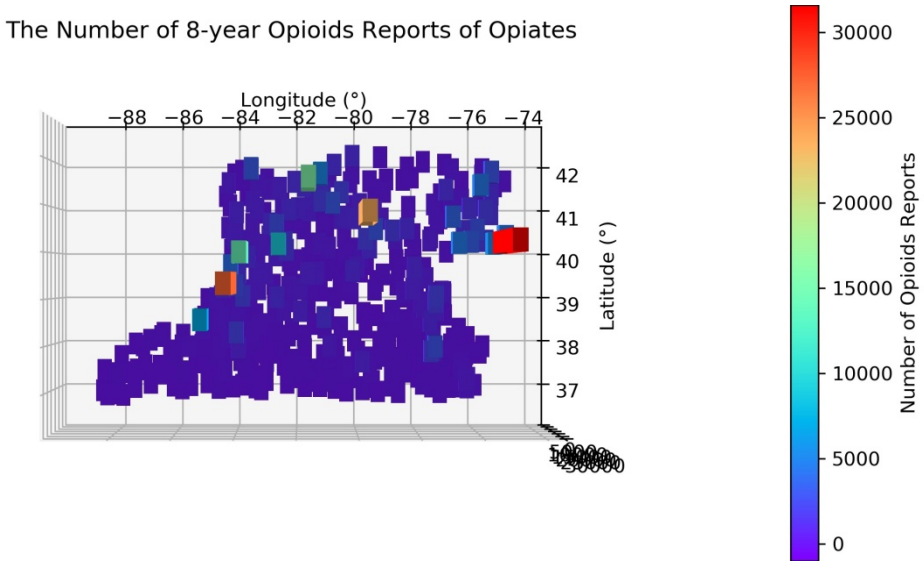


It can be clearly seen from the figure that cases of synthetic opioid and heroin abuse are concentrated in a few specific counties and there are obvious geographical characteristics. Their transmission obviously depends on geographical distance.

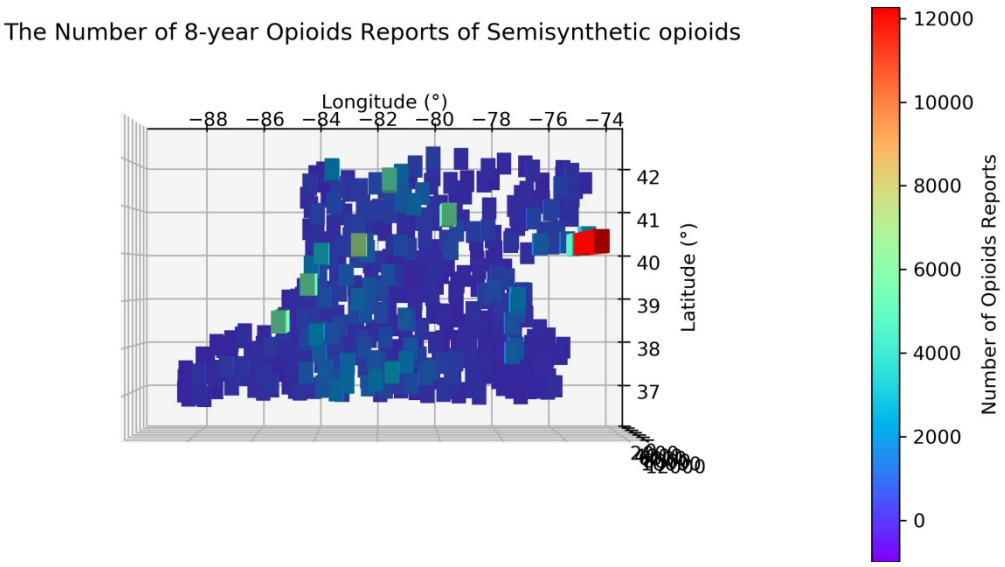
3.2 Specific opioid use in each of the 5 states

Classify opioids and obtain four categories of opioids: Optiates, Semisynthetic opioids, Synthetic opioids, and Unsorted. There are 69 substances in the excel sheet. Optiates includes heroin and different kinds of opiates. 10 are Semisynthetic opioids, mainly codeine and morphine. There are up to 46 Synthetic opioids, mainly different types of fentanyl and drugs named in the u-xxxxx format. The remaining 11 species are not yet classified. The four categories are described below for each of the five states.

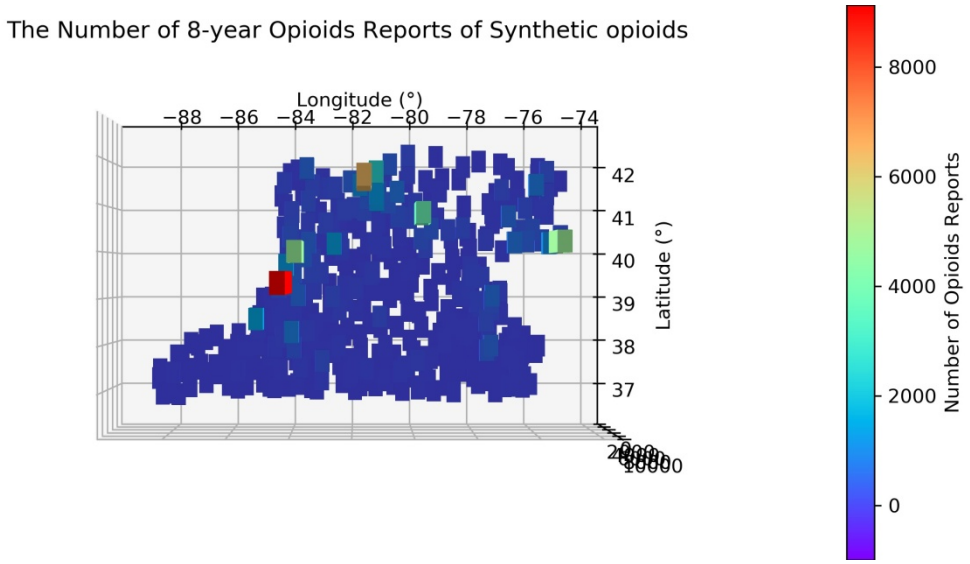
For Optiates, the following figure is drawn by taking the number of drug reports as the index and combining with the geographical location.



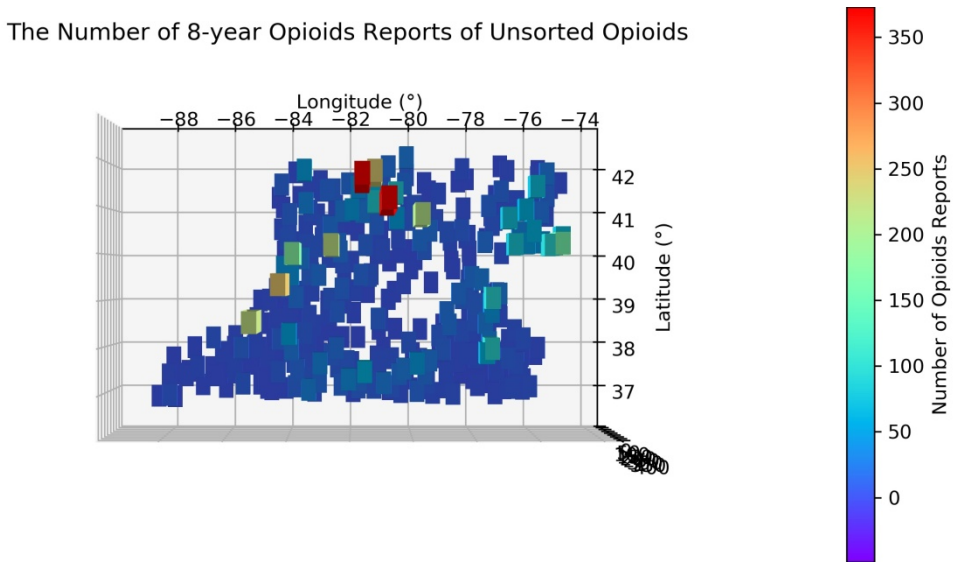
For Semisynthetic opioids, the following figure is drawn by taking the number of drug reports as the index and combining with the geographical location.



For Synthetic opioids, the following figure is drawn by taking the number of drug reports as the index and combining with the geographical location.



For Unsorted, the following figure is drawn by taking the number of drug reports as the index and combining with the geographical location.



From the figure, the counties with the highest number of reported drugs in each state are shown in the chart below, which shows possible locations where specific opioid use might have started in each of the five states.

Table 1 Targeted counties FIPS

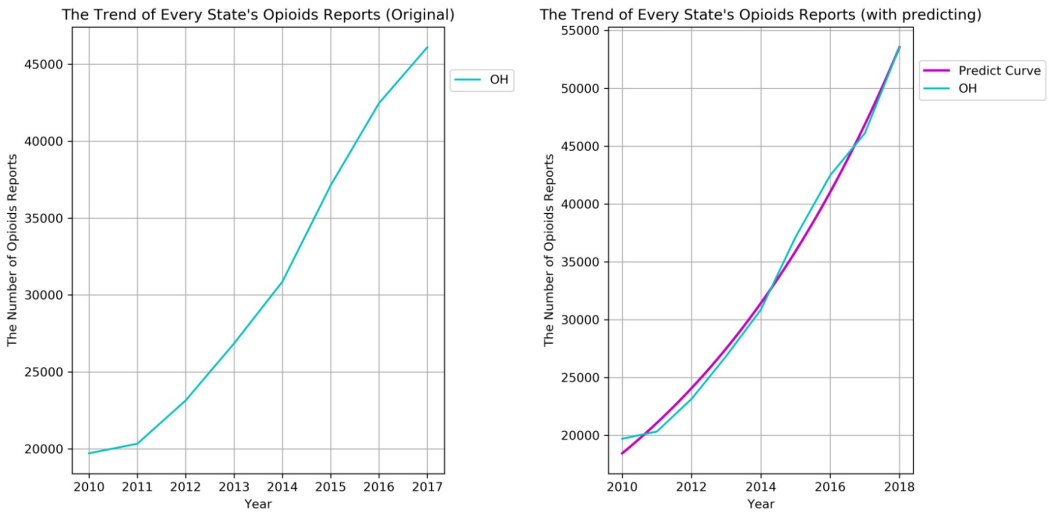
	Optiates	Semisynthetic opioids	Synthetic opioids	Unsorted
KY	21111	21111	21111	21111
PA	42101	42101	42101	42003
VA	51087	51185	51059	51087
WV	54039	54055	54039	54039
OH	39061	39049	39061	39099

3.3 Drug report thresholds

To study the thresholds of drug reporting, taking two states as examples, the summation of the number of reports of various drug abuse

incidents given in the title as the dependent variable, and time (year) as the independent variable, the graph was drawn. And the approximation function was given by using curve fitting method on the computer.

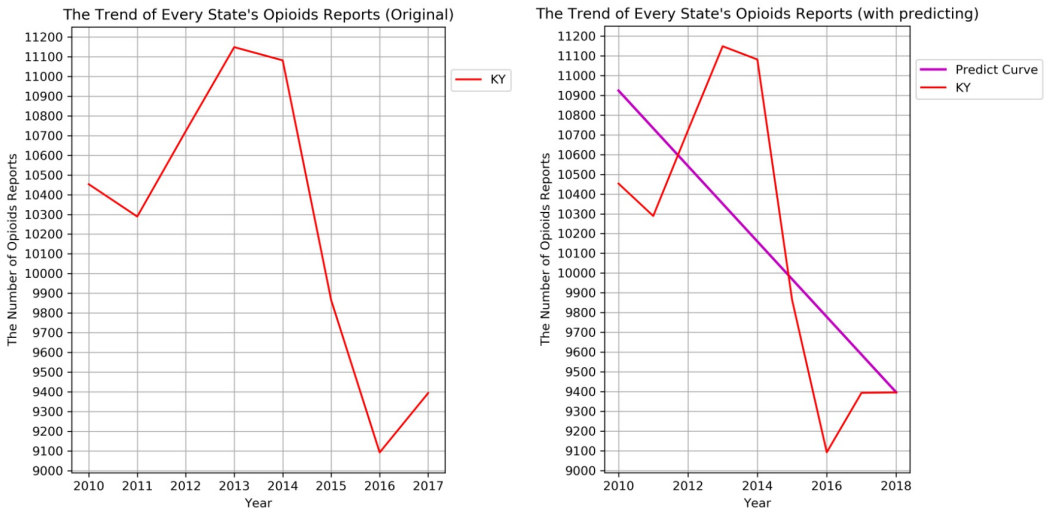
1) the dotted line graph of the number of drug abuse incident reports and time (year) in Ohio is the following figure left.



It can be seen that the figure conforms to e^{kx+b} type, and it can be obtained by least square fitting with the computer: K≈0.13321456，b≈9.82256307, the comparison between the predicted curve and the original broken line diagram is the figure above right.

According to the fitting curve, the drug abuse crisis in Ohio is becoming more and more serious, and the number of reports of drug abuse events is increasing exponentially, so the local government must take some measures. Without external intervention, it is estimated that abuse cases will exceed 100,000 by 2020.

2) the dotted line graph of the number of drug abuse incident reports and time (year) in Kentucky is the following figure left.



It can be seen that the figure conforms to kx+b type, and it can be obtained by least square fitting with the computer: K≈-190.84523810, b≈10923.58333333. The comparison between the predicted curve and the original broken line diagram is the figure above right.

According to the fitting curve, the situation of drug abuse crisis presents a stable trend in Kentucky, and the government does not need to use special means to intervene for the time being, but it still needs to pay close attention to prevent regional outbreaks.

3.4 Social factor study based on generalized linear model

The data table given contains 18 kinds of social factors, each of which gives the total number of people corresponding to various situations. The correlation analysis was made on computer between the total number of each population and the number of drug abuse incidents in the corresponding region, and the factors with high correlation were screened out to obtain the data in the following table.

Table 2 Summary of relevant social factors id

KY	HC01_VC156	HC01_VC17	HC01_VC18	HC01_VC03
OH	HC01_VC156	HC01_VC17	HC01_VC18	HC01_VC65
PA	HC01_VC17	HC01_VC173	HC01_VC175	HC01_VC30
VA	HC01_VC177'	HC01_VC206	HC01_VC47	HC01_VC71
WV	HC01_VC156	HC01_VC17	HC01_VC18	HC01_VC65

(Id comes from the data table.)

Table 3 Correlation coefficient (for Ohio)

	HC01_VC156	HC01_VC17	HC01_VC18	HC01_VC65
Pearson correlation coefficient	-0.95985492	-0.92269494	0.99379957	-0.92873027

(Generally speaking, if the correlation coefficient meets **|r|>0.95**, it is a significant correlation; if it meets **r|≥0.8**, it is a high correlation.)

Taking Ohio as an example, the influence pattern of selected social factors on the number of reported drug events was studied below. The selected social factors are as follows:

Table 4 Social factors

HC01_VC156	WORLD REGION OF BIRTH OF FOREIGN BORN foreign-born population, excluding population born at sea - Latin America

HC01_VC17	HOUSEHOLDS BY TYPE households with one or more people under 18 years
HC01_VC18	HOUSEHOLDS BY TYPE households with one or more people 65 years and over
HC01_VC65	Grandparents number of grandparents living with own grandchildren under 18 years - years responsible for grandchildren - less than 1 year

The chi-square test was carried out for the four factors.

Table 5 chi square test result

Chi-square value	P-	DOF
28727.093222658335	0.0	18

It can be seen that the chi-square value is large and the p-value is small, the four independent variables can be considered highly significant.

A generalized linear model is established for the number of substance abuse incidents and social factors. Suppose Y_{ijkl} is the actual number of substance abuse incidents and μ_{ijkl} is the predicted number of substance abuse incidents. α_i, β_j, γ_k and δ_l are four factors respectively. According to the generalized linear model, the following formula can be obtained:

$$\begin{cases} E(Y_{ijkl}) = \mu_{ijkl} \\ \eta_{ijkl} = \mu_0 + \alpha_i + \beta_j + \gamma_k + \delta_l \\ \mu_{ijkl} = g^{-1}(\eta_{ijkl}) \end{cases}$$

Based on the correlation coefficient of the correlation factors and the actual number of drug abuse incidents in Ohio, **kx+b** was selected as the correlation function here. The factor and the target function presented a linear relationship. The following table is obtained by fitting multivariate linear regression model and calculating parameters on the computer.

Table 6 Factors with parameter numerical

Factor	μ_0	α_i	β_j	γ_k	δ_l
Parameter Numerical	-6756.2421290	-0.2750167	-0.0155402	0.10463894	2.3119595

Therefore, the predictive value of the number of substance abuse incidents can be expressed as:

Year Opioids Reports = 0.10463894 HC01_VC18

+ -0.2750167 HC01_VC156

+ 2.3119595 HC01_VC65

+ -0.0155402 HC01_VC17

-6756.24212900

3.5 Possible strategy for the opioid crisis

Based on the above models, there is no doubt that the four social factors have a significant impact on the frequency of the drug abuse events, so the corresponding crisis response strategies can be proposed from these aspects.

Taking "HC01_VC18" as an example, it can be seen that families with elderly people over 65 years old are more prone to drug abuse. The possible reasons for this situation are: drug use caused by the elderly's lack of care, and drug dependence caused by the elderly's improper use of dependent drugs in the treatment of diseases.

According to this situation, the local government can accelerate the implementation of old-age welfare policies, improve the happiness of the elderly, strengthen community construction and implement the volunteer system, so that the elderly life gets more attention. In addition, medical substance supervision should be strengthened to avoid dependence and abuse of prescription drugs.Combined with the above model, when the government takes measures to reduce the drug abuse of the elderly above 65 by 10%, the corresponding factor influence is also reduced by 10%.

3.6 to the Chief Administrator of DEA/NFLIS

In our research on the opioid abuse crisis in the United States, we find a lot including 3 points as follow:

Using the MCM_NFLIS_Data provided in part 1, Our model predicts

that there are several counties might be the severely afflicted area in each of the states we study. For the sake of convenience, we divide all the opioid abuse in Opiates, Synthetic opioids and Unsorted. The results list is provided following. Hence, we should pay more attention to these specific counties and spare no effort to control opioids abuse.

FIPS code / State	Opiates	Synthetic opioids	Unsorted
KY	21111	21111	21111
PA	42101	42101	42003
VA	51087	51059	51087
WV	54039	54039	54039
OH	39061	39061	39099

Using the U.S. Census socio-economic data provided in part 2, we finish a correlation analysis and find that there are different factors in each of the states which influence the opioid abuse rates. The results are as follow(State KY for example). Generally speaking, if the correlation coefficient meets $|r|>0.95$, it is a significant correlation; if it meets $r|\geq 0.8$, it is a high correlation.

NO. in data	Correlation coefficient

HC01_VC156	0.6314589096729409,
HC01_VC17	-0.5560664725815637
HC01_VC18	-0.5560664725815637
HC01_VC200	-0.3211825832792921
HC01_VC66	0.2770338325090483

In addition, we find some common social factors as follow. We also should pay attention to these common social factors.

HC01_VC156	WORLD REGION OF BIRTH OF FOREIGN BORN foreign-born population, excluding population born at sea - Latin America
HC01_VC17	HOUSEHOLDS BY TYPE households with one or more people under 18 years
HC01_VC18	HOUSEHOLDS BY TYPE households with one or more people 65 years and over
HC01_VC65	Grandparents number of grandparents living with own grandchildren under 18 years - years responsible for grandchildren - less than 1 year

During our process of modeling, we used the MCM_NFLIS_Data

provided and we find that we can only realize the total count of the indicated substance and total count of all substances identified. But we do not realize the recovery rate of the opioids abuse which is pivotal. If we know that we can utilize the SIR model to fit the data provided better. We suggest that NFLIS can count up the number of people who recover after opioid abuse in the future.

References

[1] James Benneyan.modeling approaches, challenges, and preliminary results.Healthcare Systems Engineering Institute 2017

[2] Melissa Burden.penske alters drug plan to combat opioid abuse.proquest.2018

[3] Addiction Research - Substance Abuse.ProQuest.2018