Aaron Tae, Alex Han, Daniel Neufeldt

Part 1 - Building the Model

The Final Model, Assumptions, and Modifications

For this problem, we felt that using a modified SIR model would be the best decision. This is because we can clearly, yet accurately create a representation through the model. There were clear "susceptible, infected, and recovering" individuals, being non-drug users(S), drug users(I), and sober people(R) respectively. We referenced a website (1) to build our model. After viewing this model though, we felt that because recovering drug users are very prone to relapses, there should be an option where the sober people can go back to being drug users. This new SIR model that we built was constructed where non-drug users could enter the drug user category, but not the other way around.Additionally, drug users could enter the sober category and vice versa. Note that the sober people could never become a non-drug user.

Some assumptions for our model were:

1. Each report in the data represented one drug user.
2. Once someone goes from non-drug user to drug user, they cannot go back.
3. There are no deaths from overdose.
4. The probability that each drug is counted among all the drugs is the same (i.e., heroin and morphine have the same chance of being counted in the report).
5. The population stays constant.

Some variables for our model were:

1. S_0 is equal to the number of people minus the number of drug reports.
2. I_0 is equal to number of total drug reports.
3. R_0 is zero.
4. r is the rate that people go from being non-drug users to drug users.
5. v is the rate that people go from being drug-users to sober.
6. p is the rate that people go from sober to drug users.
7. P is the population at the starting point (which is 2010 in this case).

With all of these assumptions implemented, we can now create this SIR model through three differential equations:

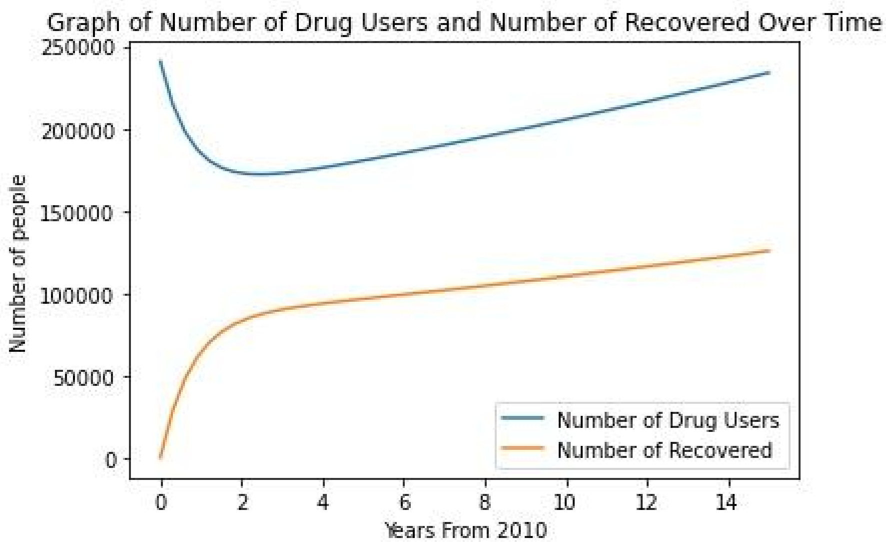
$dS(t)/dt = -r/P * S(t) * I(t)$

$dI(t)/dt = r/P * S(t) * I(t) - v * I(t) + p * R(t)$

$dR(t)/dt = v * I(t) - p * R(t)$

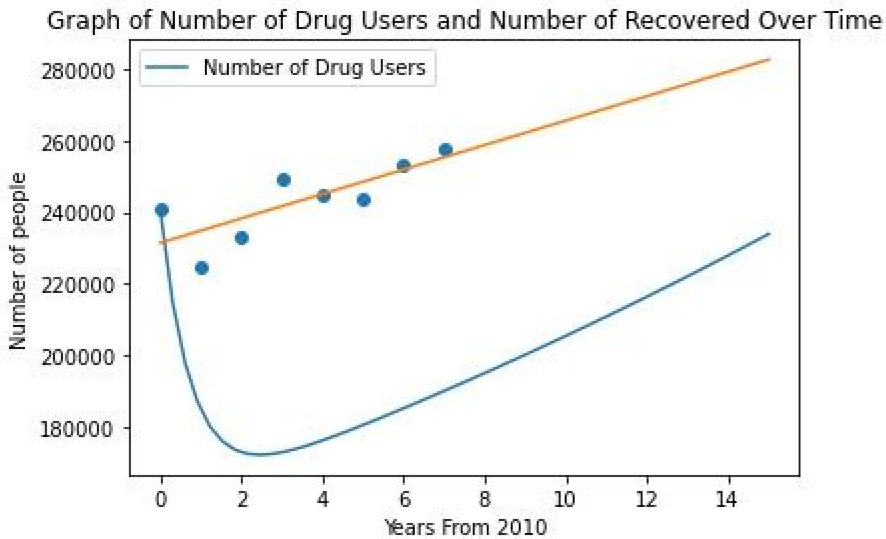
We now give values to our constant values. For the population, we decided to just use the initial population from the data - which is the population from 2010 - because it was a simpler way to get a similar trend as we believed that the population would not change drastically over the course of 2010 to 2017. We set the value of R(0) to be zero because there was no way for us to find the number of sober individuals in the given data. We decided that setting R(0) to equal zero was the best decision and to let the model adjust the R(t) value on its own. Using outside sources, (2) (3), we found the values for *v* and *p* to be 0.47 and 0.85 respectively. To find the value of *r*, we found the difference of the number of reports between 2010 and 2011. We then divided that number by the difference of populations between 2010 and 2011. We repeated this process while increasing the years by one until we got to 2017. This resulted in seven different rates from non-drug users to drug users for each year. We then took the average of all of these values and ended up with a value around 0.04031. This value of *r* was the rate that the population shifts from non-drug user to drug user; however, we wanted the rate that one person shifts from non-drug user to drug user, so we divided this number by our initial population, which was 38,162,380.

When applying our model to this problem, our model predicted that the number of drug reports would consistently increase over time with the number of people becoming sober also rising, although at a much slower rate. The graph that our model produced when we inputted the data for all five states in total was:



A large issue with the graph is that up until the x-axis reaches 2, the graph is not accurate at all. To account for this when observing trends and finding slopes, we only used the graph after x equals 2. The reason why the graph looks like this is because we assumed R(t) to be zero at the very start, so the model had to use the first two years to adjust the values before they became more accurate. We can see from this graph that the amount of drug users is not stopping overall for these five states.

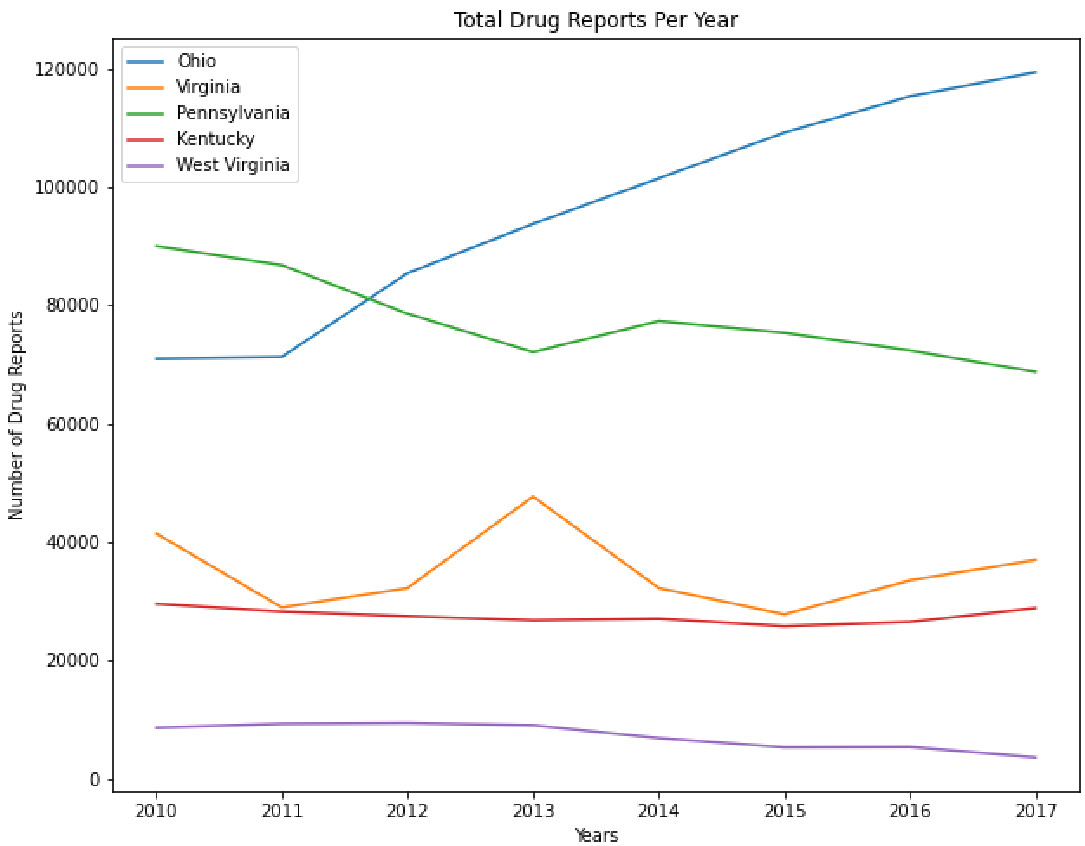
We now test our model to see if the accuracy is adequate or not. We plotted the actual total number of drug reports from all five states on a graph with the number of drug users that our model predicted and then created a line of best fit for the plotted points.



The slope for the line of best fit was approximately 3422 people per year. This number is quite different from our slope that we found for the model as that slope was approximately 4942 people per year. This may seem like a very large number, but considering the fact that the y-axis is between around 170,000 and 300,000, we can conclude that the model is fairly accurate at least up until around 25 years from 2010. We can back that previous statement because when the x-axis is around the 25 mark, the two lines intersect. Another aspect of the graph that we must address is the fact that the plotted points are so much higher than the actual line when the model begins to become accurate. This difference can be explained by the fact that the model had to adjust from having R(t) set at zero from the start. Because R(t) was set so low from the start and dI(t)/dt was dependent on R(t), it makes sense that the number of drug users would be lower than the actual data from the start.

Analyzing the Data

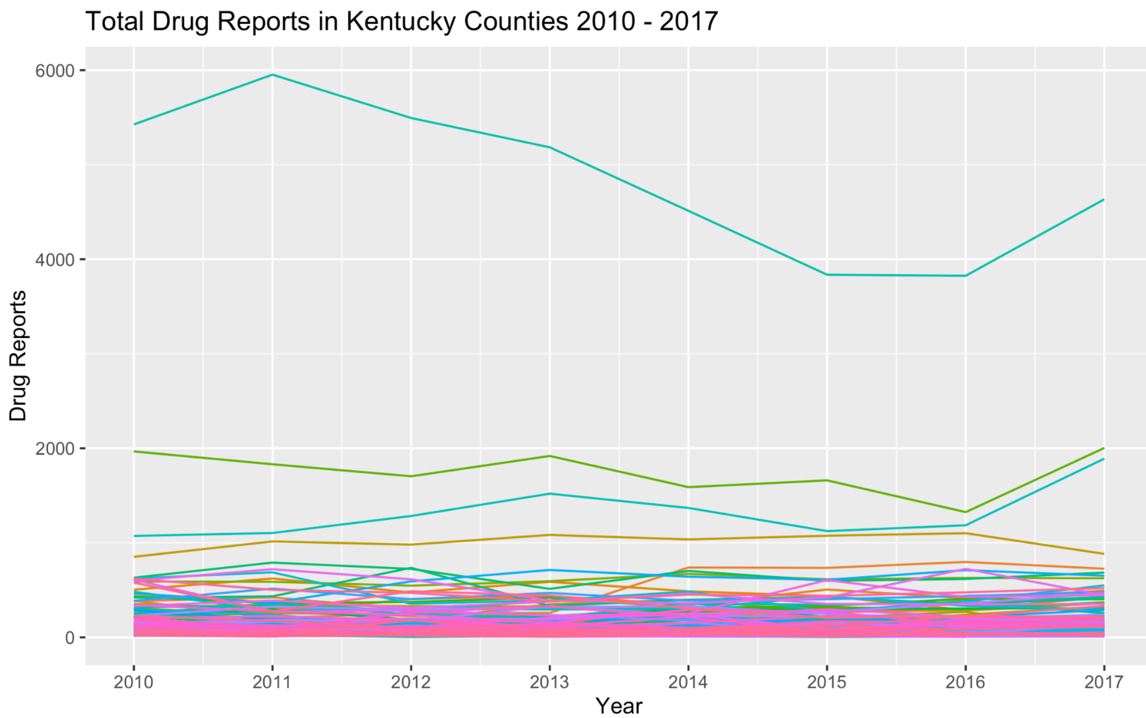
Before testing our model to find the epicenter of the opioid epidemic, we decided to analyze the given data in order to find trends or patterns that we can note. We began by creating a graph that displayed the total drug reports per year in each of the five different states. This graph was created with the data being recorded on a year-by-year basis:



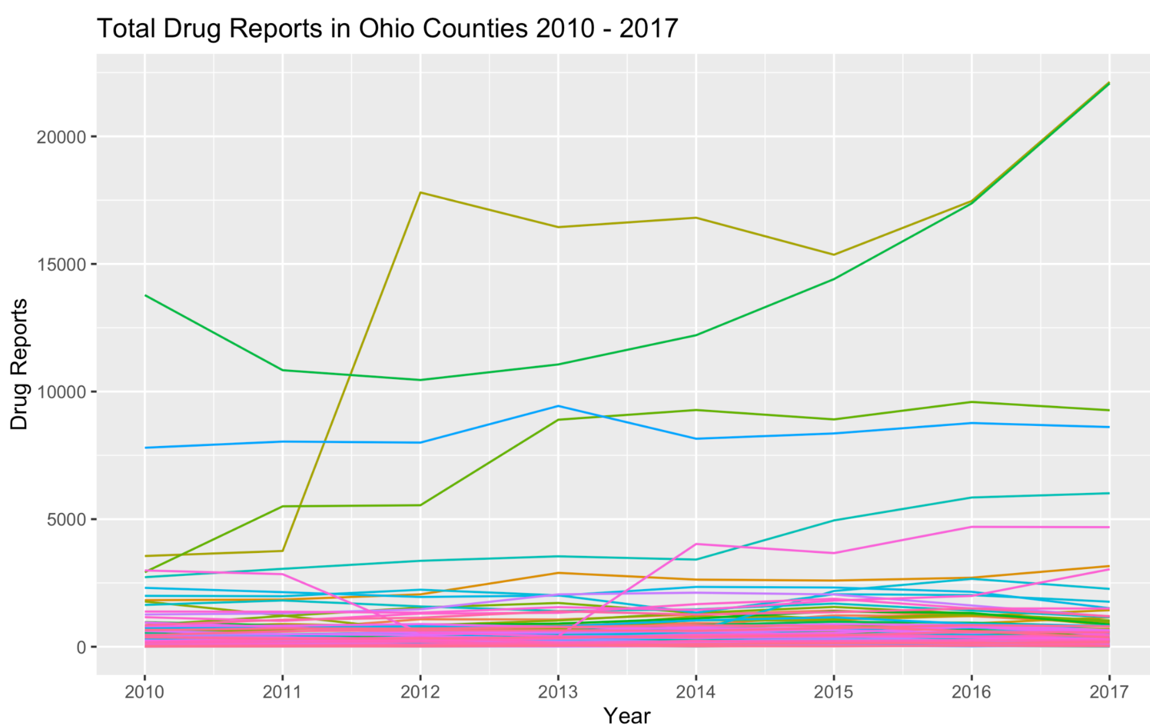
We can clearly see in this graph that Ohio has the greatest number of drug reports per year in the latter half of the 2010s. Additional observations are that Pennsylvania is steadily declining and Virginia had a peak around 2013, a drop around 2015, and is slowly starting to increase again. For Kentucky and West Virginia, the graph shows that the two states are plateauing; however, while Kentucky is showing a slight increase in 2017, West Virginia is maintaining a slow decline which started in 2014.

Analyzing Counties

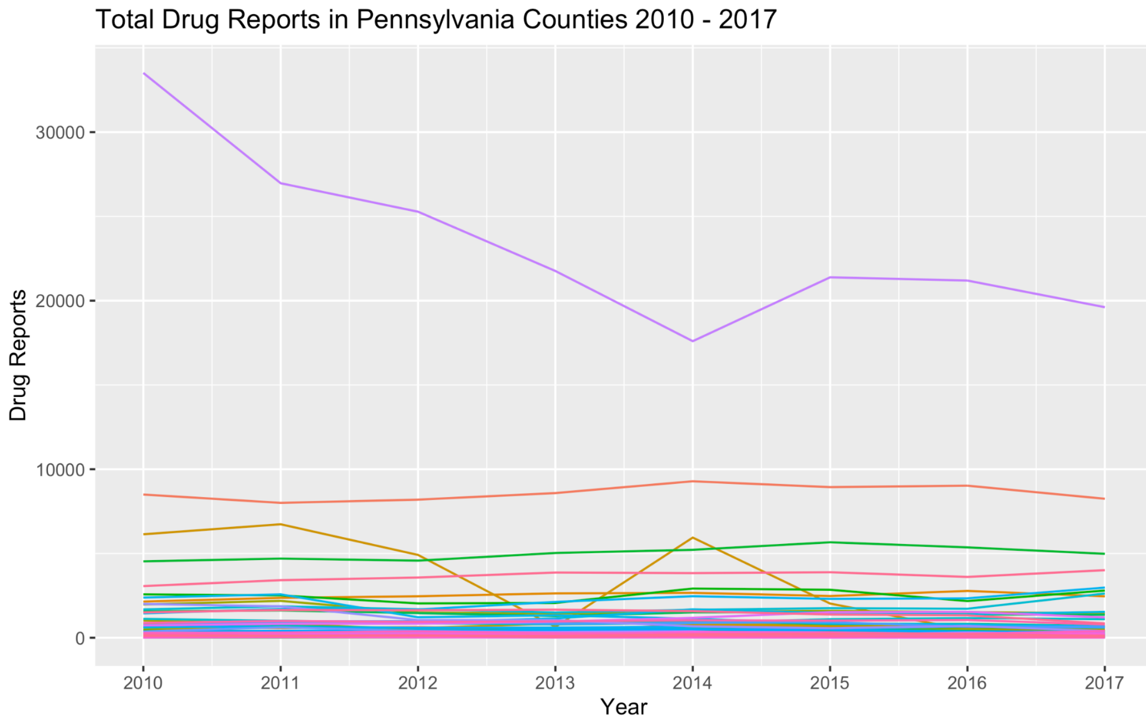
We made this decision because the larger counties not only have a significantly higher sample size, but the start of an opioid epidemic would most likely happen in a more populated region, rather than a smaller region. We analyzed the counties because we wanted to take a deeper look into the possible epicenters for the opioid epidemic.



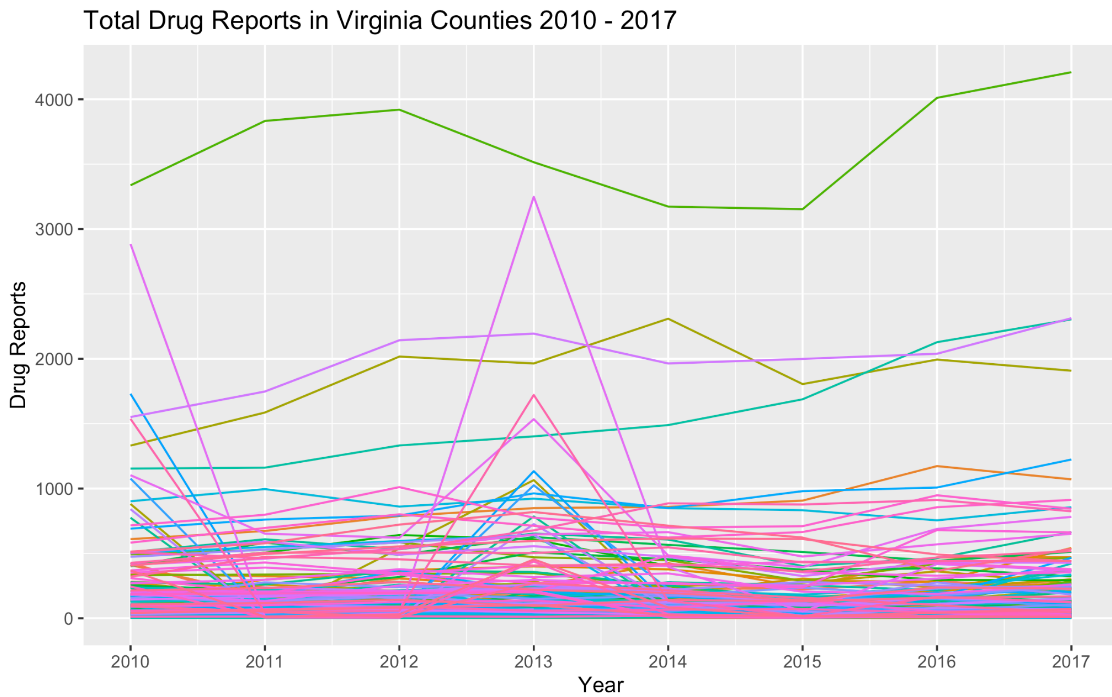
We first analyzed the counties in Kentucky. It is easy to observe that there exists four counties (Jefferson, Fayette, Kenton, Campbell) that are above the cluster of all the other counties, notably Jefferson (the highest line). These four counties are above what is called the "drug identification threshold", which is around 1,000 and is where the vast majority of counties in Kentucky lie. It is also glaring how far above Jefferson is compared to the other counties, which we may assume is due to the relatively large population.



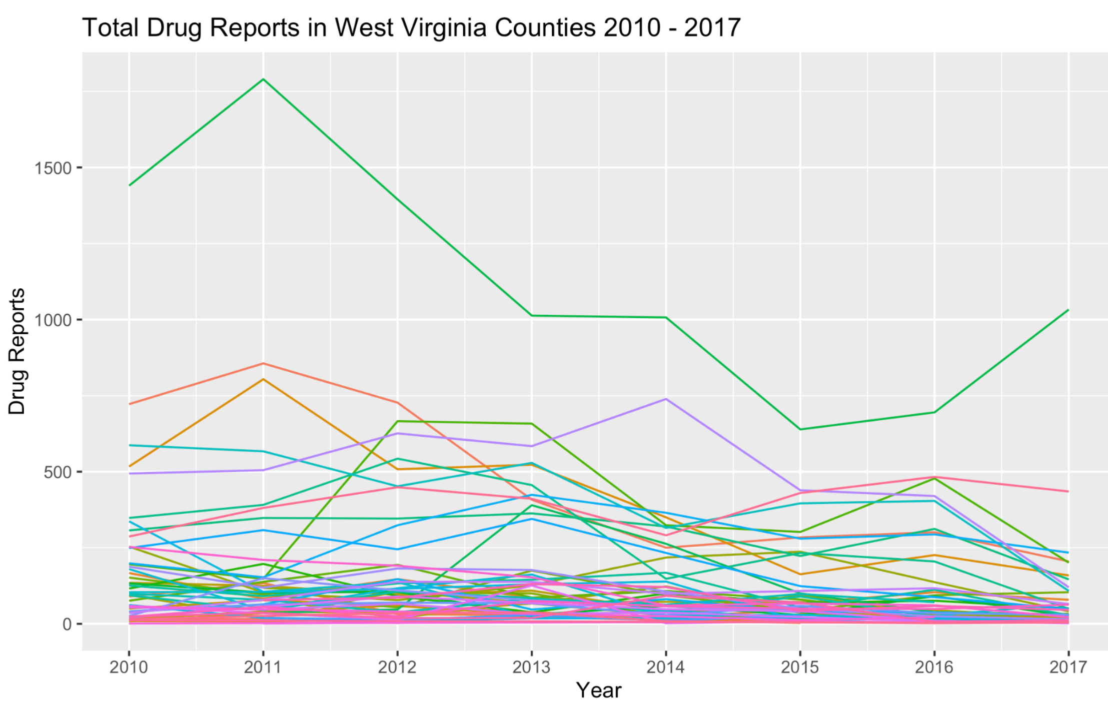
We then analyzed the counties in Ohio, which produced a slightly more intriguing graph than the previous. It is observable that there is a general increasing trend in drug reports for the counties that are well above the threshold. In particular, we notice that Cuyahoga (army green) and Hamilton (bright green) both have significant rises in the amount of drug reports since 2010. In comparison to the counties of Kentucky, we can see that there are more counties in Ohio that are well above the threshold, including counties such as Lake, Franklin, and Montgomery.



On the other hand, the number of drug reports in the counties of Pennsylvania generally seemed to stay consistent with the exception of Philadelphia. Similar to Jefferson in Kentucky, Philadelphia appears to be well above all other counties, which we may assume is due to the difference in total population. One intriguing thing to note is that the drug identification threshold seems thinner compared to that of Kentucky or Ohio, making it difficult to determine where the threshold lies.



Contrary to the patterns observed in the counties of Pennsylvania, the counties of Virginia seem to fluctuate. We can notice that there are several counties that begin with a high number of drug reports, but rapidly decrease in the span of just one year. One of these counties is in fact Richmond (magenta line that begins at around 3000), which rapidly increases from 2012. It appears to peak in 2013, and once again comes crashing down and hits the floor by 2014. However, the county with the highest number of drug reports, Fairfax, seems to have an increasing trend of drug reports overall over the span of 7 years.



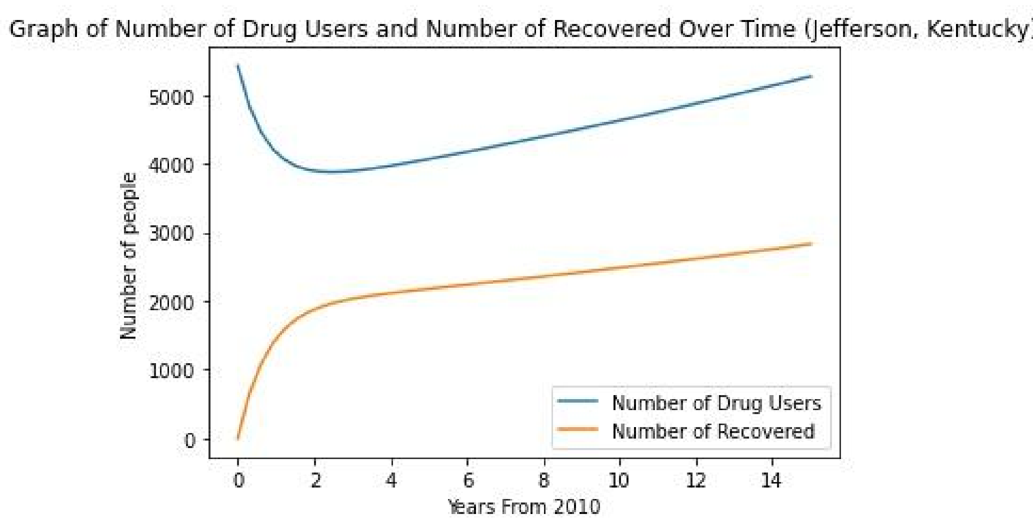
It is instantly recognizable that West Virginia has a much lower population per county compared to the other four states discussed above. Kanawha (top green line), starts off well above the other counties at about 1,500 reports, but sharply decreases from 2011 to 2015. All other counties that are above the vague threshold also appear to have a generally decreasing trend. It is also worth noting that the decrease in Kanawha's number of drug reports did not necessarily mean the increase in the number of drug reports in other counties.

<u>Determining the Epicenter</u>

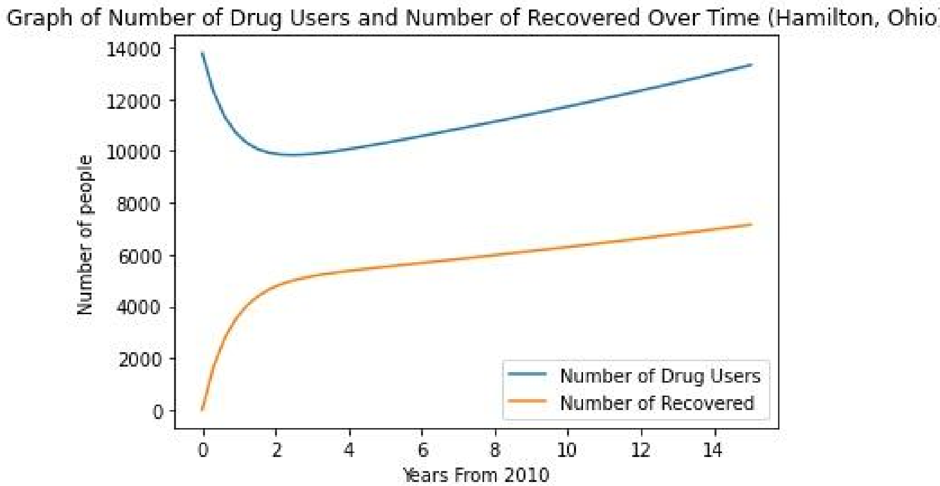
To find the epicenter, we had to determine what factors caused a place to most likely be the epicenter of the opioid epidemic. We eventually decided upon using the model on the five largest counties in each state and using the slope to determine if that county was where the epicenter is. We found this by graphing the data through the differential equations we made and then creating a line after the x-axis passes two to determine the slope. We felt that using the slope of our model was a better characteristic to determine the epicenter of each state than the number of drug reports. Using a number of reports is not an effective option because a region with a large number of drug reports could just have such a large number due sheerly off population size, whereas a region with a high rate of people becoming drug users shows that there was an epidemic happening in the area at that time. A high rate of people becoming "infected" - or drug users in this case- is a key characteristic in determining if there is an epidemic occurring or not.

After applying the data of the top five counties in each of the states into our model, the following graphs are the models of each county that we believe is the epicenter:

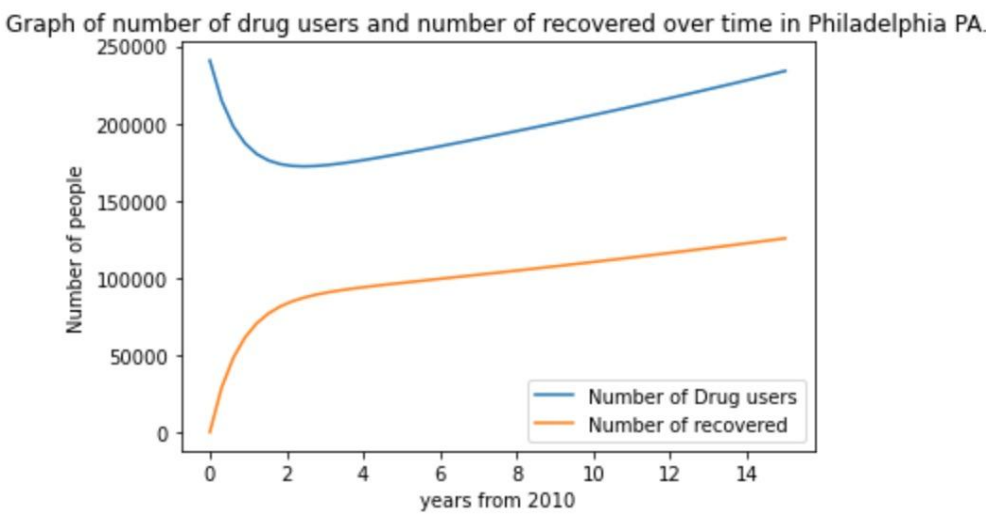
Kentucky:



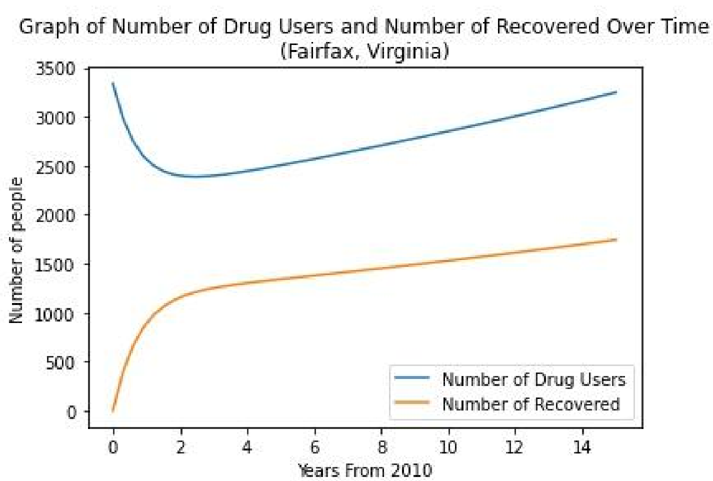
Ohio:



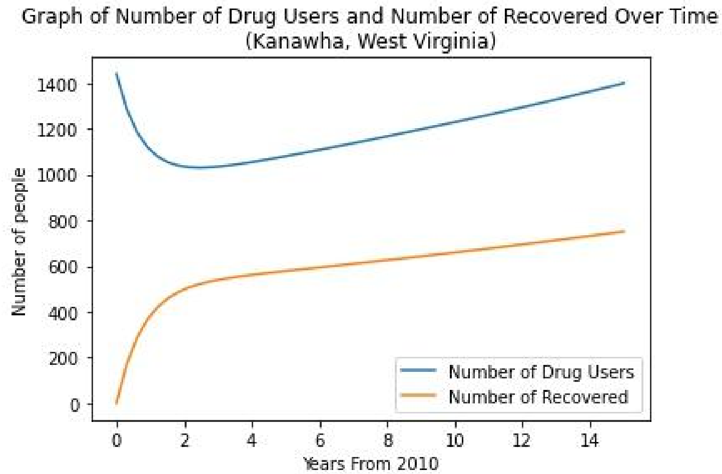
Pennsylvania:



Virginia:



West Virginia:



We also put the slopes of all the counties in the following dataframe:

	Number of people becoming drug users per year
oh1	73.449020
oh2	278.059150
oh3	60.341737
oh4	158.098289
oh5	55.484182
ky1	111.228261
ky2	40.629032
ky3	22.006051
ky4	17.395183
ky5	9.550576
pa1	689.721695
pa2	175.938825
pa3	126.908916
pa4	93.775275
pa5	63.443695
va1	68.861755
va2	31.047397
va3	31.935395
va4	27.415100
va5	23.803201
wv1	29.514816
wv2	14.786265
wv3	10.631144
wv4	10.136549
wv5	4.046191

Here, the indexing is where each number represents the ranking of size for the state (e.g., oh1 is the biggest county in Ohio, being Cuyahoga). We can see from this table that the counties with the highest slopes (rounded) for their state is:

- Hamilton, Ohio (278 new drug users per year)
- Jefferson, Kentucky (111 new drug users per year)
- Philadelphia, Pennsylvania (690 new drug users per year)
- Fairfax, Virginia (69 new drug users per year)
- Kanawha, West Virginia (30 new drug users per year)

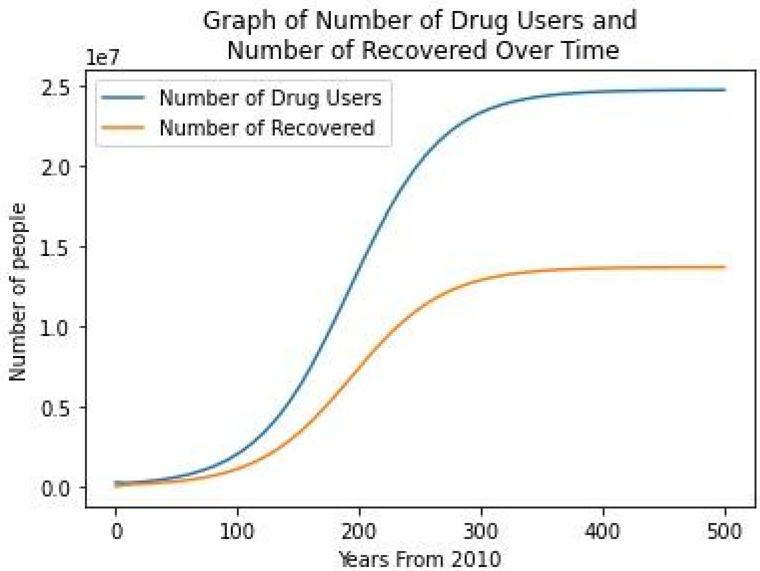
This is relatively accurate as these specific counties are more significant in comparison to the other counties in the graphs earlier shown of the total number of drug user reports every year per county for each state. An interesting observation is that Hamilton, Ohio was determined to be the epicenter even though Cuyahoga, Ohio was also very large in population. We determined that Hamilton was the epicenter because the rate that drug users were increasing in Hamilton was just significantly higher than the rate in Cuyahoga. Because the two counties had a similar total number of drug reports in the Ohio counties graph shown before, it can be argued that either one of these cities are the epicenter for Ohio.

Another big observation is how Philadelphia has such a high slope. This is interesting because the total number of cases is said to go down according to the counties in Pennsylvania graph earlier; however, according to our model, this is not the case. This may be an observation to note for further investigation later on.

We also determined that Jefferson was the epicenter of Kentucky. This is rather obvious because the slope for Jefferson is so much higher than the other counties. We can use the same logic for both Fairfax, Virginia and Kanawha, West Virginia. The total number of drug reports per year vs counties for each state graphs above support this data as Jefferson, Fairfax, and Kanawha are all the lines with the most drug reports compared to the other counties in the state.

Future Predictions

To see if this trend in the US continues, we decided to see if an asymptote would be reached in our model as well as where that asymptote would occur. We set the x-axis to go on for 500 years and the following results were obtained:



Here, we can clearly see that around the 350 years from 2010 mark the trend seems to reach an asymptote. This occurs when there are approximately 25,000,000 people who are drug users and 13,000,000 people who are sober. This means that according to our model, if this current trend of opioid use continues, then in about 350 years from now everyone will either be a drug user or had once been a drug user. Because of ever-changing policies and a general inability to predict something 300 years from now accurately, it is fair to say that this model is meant to be used for predicting events that occur in the near future (within 40 years from now) as that was the time frame that it was intended to predict.

Part 2 - Socio-economic Trends

To identify any possible correlations between the socio-economic background of the demographics in each state, we compared first computed what percentage of the population in 2010 and 2016 of each five states was:

1) A veteran
2) A high school graduate
3) Living in a single household

We then ranked states by these variables, and also ranked them by the percentage of drug users in the total population. We compared these ranks by using the following data tables:

For 2010:

	Kentucky	Ohio	Pennsylvania	Virginia	West Virginia
High School Diploma Percentage	38.300833	42.318182	44.097015	32.147015	44.147273
Single Household Percentage	26.559167	26.125000	27.386567	27.385075	28.372727
Civillian Veteran Percentage	9.851667	11.294318	11.904478	12.326866	11.801818
Opioid Drug Users Percentage	1.764648	1.559639	1.821264	1.393924	1.169970

In terms of rank:

	Kentucky	Ohio	Pennsylvania	Virginia	West Virginia
High School Diploma Percentage	4	3	2	5	1
Single Household Percentage	4	5	2	3	1
Civillian Veteran Percentage	5	4	2	1	3
Opioid Drug Users Percentage	2	3	1	4	5

For 2016:

	Kentucky	Ohio	Pennsylvania	Vrgnia	West Virginia
High school diploma percentage	9.539167	11.028409	12.064925	11.532836	11.489091
Single household percentage	38.258333	41.880682	32.031343	43.871642	43.800000
Civilian veteran percentage	26.579167	26.101136	27.318657	27.488060	28.567273
Opioid drug user percentage	0.622598	1.039507	0.417439	0.602738	0.304584

In terms of rank:

	Kentucky	Ohio	Pennsylvania	Virgnia	West Virginia
High school diploma percentage	5	4	1	2	3
Single household percentage	4	3	5	1	2
Civilian veteran percentage	4	5	3	2	1
Opioid drug user percentage	2	1	4	3	5

For both 2010 and 2016, we can observe that there is no noticeable correlation between the three socio-economic variables and the opioid drug user percentage. Assuming there is a correlation between them, a higher high school diploma percentage would suggest a lower ranking in opioid drug user percentage. Also, a higher rank in single household and civilian veteran percentages would suggest a higher rank in opioid drug user percentage. However, this does not seem to be the case. We can see that the rankings across all categories for both 2010 and 2016 are evenly scattered, in which we can state that our findings are inconclusive.

In conclusion, we were not able to find any correlation between opioid drug user percentages for all five states and veteran status, education status, and family size. We may attribute this lack of correlations to the methodology of our approach or simply, there is no actual realistic correlation.

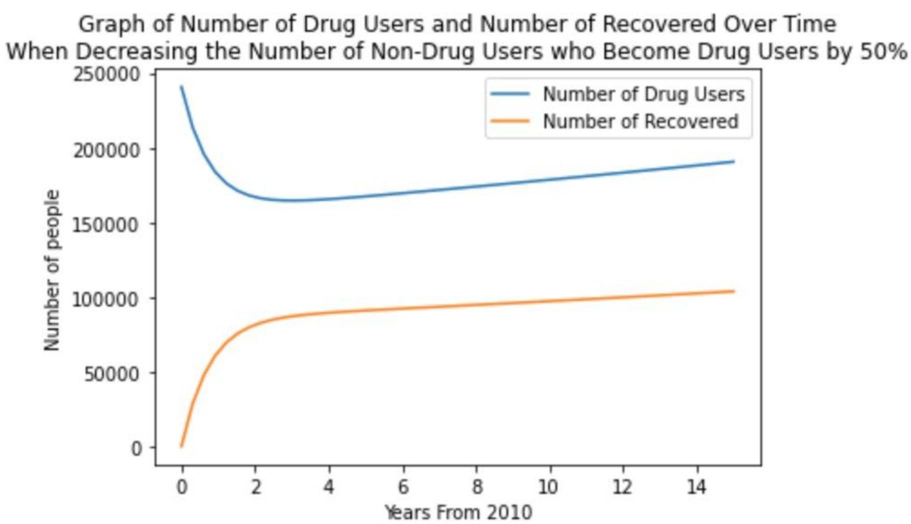
Part 3 - Strategy to Countering the Opioid Crisis

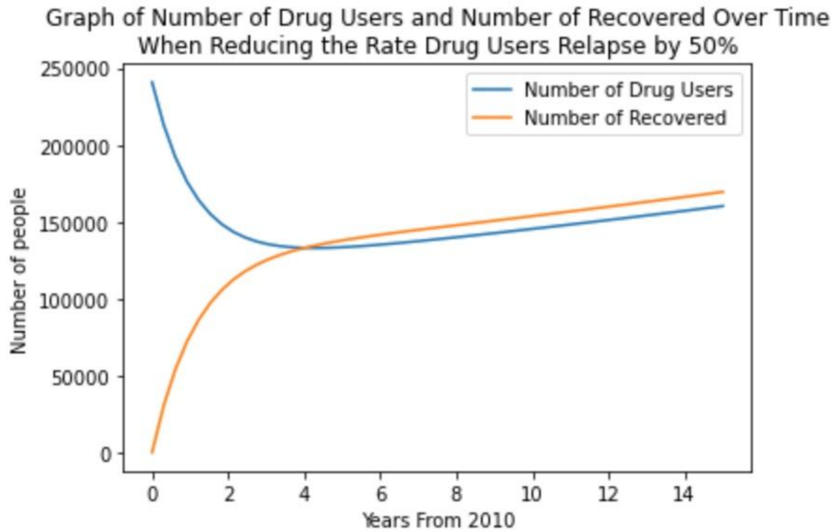
By altering the *r, v, p* variables for our model by 50%, it can be observed that altering the value of *p* had the greatest effect on the distribution of drug users and recovered users. To recall, the value of *p* represents the rate of change of abstinent drug users to going back to using opioids. In other words, it represents the rate of relapse. When decreasing *p* by 50%, we can observe that the number of recovered drug users eventually surpasses the number of drug users by the 4th year, which is the only graph in which this occurs. This demonstrates the influence that the rate of relapse has on the number of opioid users in a population.

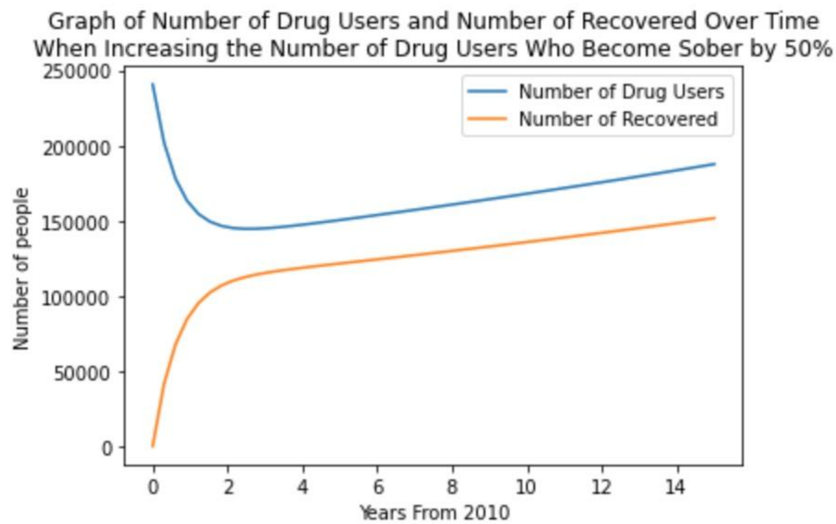
Therefore, we can state that the most effective strategy in combating the opioid crisis based on our model is that the government can increase funding for drug rehabilitation centers as well as making them more accessible to opioid abusers. This will decrease the probability that opioid users that have recovered to relapse. Furthermore, providing proper and long-term treatment for former drug users will also be able to strongly contribute to the decrease in the number of drug users.

Another notable observation is the effect of an increase in *v*, which is the rate of change of drug users to those who have recovered. Although it may not be as effective as the previous solution, raising awareness of the detrimental effects of opioid usage as well as facilitating the accessibility of resources such as drug hotlines may also be possible strategies to consider to counter the opioid crisis.

A different approach that we can take on combating the opioid crisis is focusing anti-drug efforts in specific locations that have strong influences on the neighboring counties. Our model is able to predict the county with the highest increase in opioid user population for each state, which we can assume is the epicenter of the opioid epidemic. The government can then target these specific counties in their efforts to counter the opioid crisis, which can make the battle against drugs much more time and cost efficient.







Based on the lack of socio-economic correlations with drug usage that we were able to find, we can also conclude that targeting specific locations is more likely to be successful than attempting to target certain demographics.

Strengths of our model is its accessibility due to the low number of parameters required. This makes the model easier to interact with, and also makes it easy to understand. There is also the fact that it is based on the well-known SIR model, which gives our model a level of reasonable credibility.

Weaknesses of our model include the limitations of such a simple system. The fact that the coefficients we use in our model to not change causes the predictive power of our model to not be as effective as possible. Furthermore, we assume that no individual can become truly abstinent, which creates a population model in which the amount of non-drug users can only continuously decrease. We also assume that the amount of individuals that are recovered is initially zero, which causes the graphs produced by our model to have an initial spike in the opposite direction as its end behavior.

Part 3 - Summary

For our project, we used a modified SIR model to model the spread of disease to predict the spread of synthetic opioid and heroin incidents. The adjustment that we made to the SIR model is we made it possible that the people that are recovered can actually become drug users again. Thus, the model being SIRI rather than SIR.

Using the NFLIS excel data, we first graphed the number of total drug reports of the five states over the time period 2010 ~ 2017. From this graph, we found out that Ohio had the most number of cases, followed by Pennsylvania, Virginia, Kentucky, and West Virginia. Then, we found the top five counties for each state that had the highest number of drug reports using the "TotalDrugReportsCounty" column.

The 3 ODEs of our model are $\frac{ds}{dt} = -r \cdot S(t) \cdot I(t)$ $\frac{dI}{dt} = r \cdot S(t) \cdot I(t) - v(t) + p \cdot R(t)$ $\frac{dR}{dt} = v \cdot I(t) - p \cdot R(t)$, where $S = Number\ of\ non-drug\ users,\ I = Number\ of\ drug\ users$

$R = Recovered\ drug\ users, r = rate\ of\ people\ going\ from\ non-drug\ users\ to\ drug\ users,$

$v = rate\ of\ people\ going\ from\ drug-users\ to\ recovered\ users,\ p = rate\ of\ people\ going\ from\ recovered\ back\ to\ drug\ users$

Using the total population of each state and the number of reported cases for the top five counties of each state in the year 2010, we created a model that showed the behavior of the number of drug users and number of recovered. Our hypothesis was that the counties that had the steepest slope would be the origin where opium and heroin uses spread out. From our model, the steepest slopes were Cuyahoga OH., Kanawha WV., Philadelphia PA., Fairfax VA., and Jefferson KY.

From our model, it was possible to conclude that the most likely places where the opioid and heroin use spread matched the graph of "Total number of drug reports of the top five counties", meaning that the county with the highest drug reports had the steepest slope in our model. Thus, allowing for a conclusion that our model is valid for use when assessing where the opioid and heroin use might have started.

Memo

Chief Administrator

DEA/NFLIS Database

June 10th, 2022

To Whom It May Concern,

After analyzing data about opioid use from Kentucky, Ohio, Pennsylvania, Virginia, and West Virginia, we have arrived at many observations and conclusions.

Through our data, we found that the epicenters for the opioid epidemic were located at: Jefferson, Kentucky; Hamilton, Ohio; Philadelphia, Pennsylvania; Fairfax, Virginia; and Kanawha, West Virginia. We determined this through finding the rate at which drug reports increase and those five stated counties possessed the highest rates out of their states. We chose this method of analysis because we felt that the rate of an epidemic was a bigger indicator of an epicenter than the sheer amount of reports.

Another observation that we made was that the education level (high school diploma or higher), single household status, and civilian veteran status did not cause a change in the percentage of drug use in each state. There is a chance that the methodology of our approach was not the best for this specific situation, but it is also very plausible that there is indeed no correlation between these socio-economic factors. Because of this ambiguity, more modifications to our method is required.

Along with these observations, we also came up with solutions to counter the opioid epidemic. Something important that we discovered was when we noticed through our model that reducing the rate of relapsing significantly impacted the number of drug reports per year, as well as the number of sober individuals per year. A solution that we suggest to accommodate this observation is to put more funding into rehabilitation centers in an effort to keep people from relapsing. Another finding is that increasing the rate that people become sober will help diminish the percentage of opioid drug users. A solution that we propose based on this result is to increase awareness of anti-drug efforts in regions with higher drug report rates predicted by our model. The government can spend more resources investing in these areas, rather than targeting locations based on the assumption that there are demographic correlations.

We hope you find our investigations useful in countering the opioid crisis. Thank you.

<u>Contributions</u>

In our group, everyone contributed very equally. When we built the model, we all worked off each other to come up with something. Whenever there were any demanding coding portions of this project, we split the work equally and put all the results together. We also split apart working on creating the report and slides.

References

1. Mathematical Association of America. The SIR Model for Spread of Disease - The Differential Equation Model, 2004.

https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model

2. Ashley Everything For Recovery. Drug Addiction Recovery Statistics, 2021.

https://www.ashleytreatment.org/drug-addiction-recovery-statistics/

3. National Library of Medicine. New Findings on Biological Factors Predicting Addiction Relapse Vulnerability, 2013.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3674771/