By performing data analysis, we determine several goals and construct corresponding models to reflect the hidden characteristics of the spread of the opioid crisis.

First, we conduct a series of data preprocessing work, including abnormal value processing, default value processing and data normalization. We then reclassify all the socio-economic factors included into 4 categories to analyze the data more efficiently. In order to consider the spread between adjacent counties, we also sort the data based on geographic location and adjacent relations.

Next, through the application of multiple linear regression model, we estimate the coefficients of various kinds of socio-economic factors, which reflect the relationship between elements and the number of drug reports. We discover that 38 factors have the most significant influence among all the possibilities. By comparing these coefficients, we find out the top 8 factors with the greatest effects on the condition of drug abuse.

Then, we introduce the index, comprehensive socio-economic property(CSE), to reflect the propensity for people in a county to use synthetic opioid or heroin. Based on the hierarchical cluster analysis(HCA), we firstly reduce the number of variables along with the consideration of common consensus. The HCA is a comprehensive index integrate multiple aspects, including family condition, education level and cultural background. Ranking CSE, we find that the opioid crisis in Ohio is the most severe among the five states.

Finally, we revise the traditional Cellular Automata Model to simulate the spread of opioid crisis among counties. In our CA model, we shed the constraint by using arrays to store the neighbors, which implies that we can have theoretically infinite neighbors. We represent a county by a cell, and any two cells are adjacent in our CA model if and only if the two counties they respectively represent are geographically adjacent. Rules applied in our CA model, including prediction rules and backstepping rules, are devised to either forecast the trend in the future or estimate the situation in the past. The simulation of our CA model shows that Allegheny(PA), Hamilton(OH) and Philadelphia(PA) may be the places where the opioid crisis first started, while Delaware(PA) and Philadelphia(PA) might face a severe opioid crisis in the future.