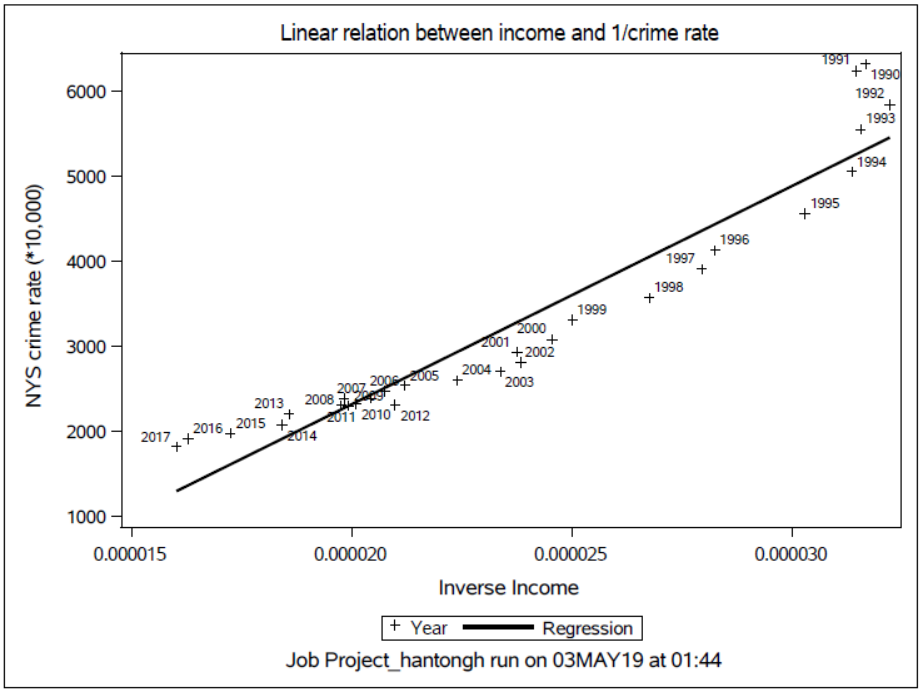
Project Write-up for BIOS 669: Relationship between crime rate and median household income and chemical dependence treatment admission of New York State

Hantong Hu

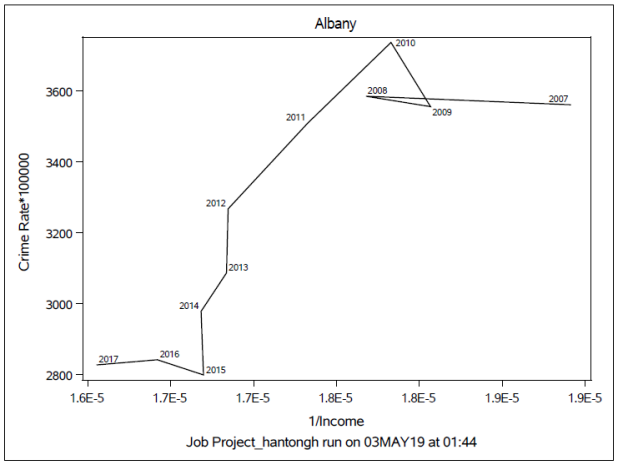
This project analyzed the relationship between crime rate and median household income in New York State (state level) from 1990 to 2017, and introduced chemical dependence treatment admission to analyze the relationship in county level from 2007 to 2017.

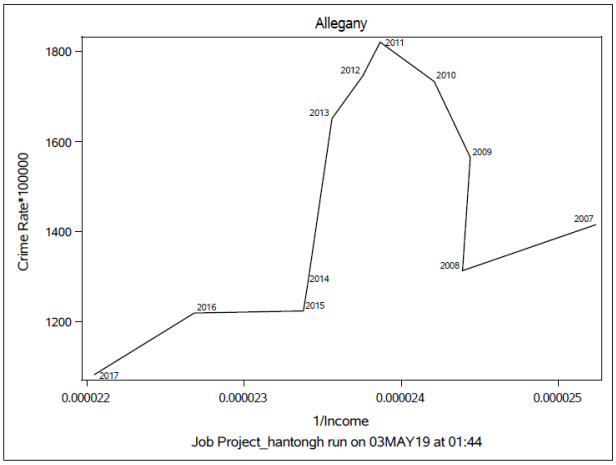
First, the program imported data sets of county-level crime rate from 1990 to 2017, chemical dependence treatment from 2007 to 2017, and New York State median income from 1990 to 2017. State crime rate is calculated by the sum of crime cases divided by state population and inflated by 100,000 to be consistent with the original data set. Then, several line charts are plotted (P15-P18) and it indicates that there seem to be a linear relationship between crime rate and 1/income (P18).



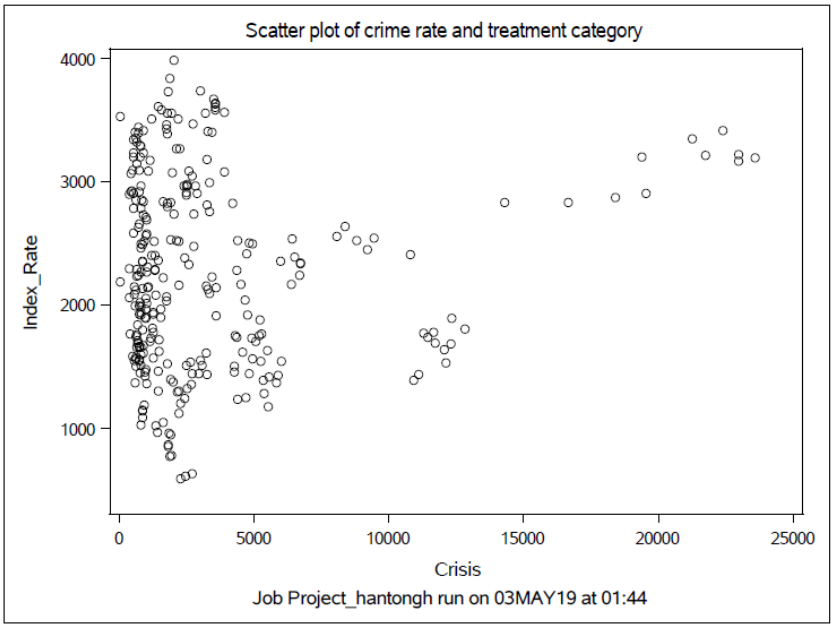
In addition, it is obvious that the linear relationship is somehow violated starting from 2007. Thus county level analysis is conducted.

County-level data is scattered in several data sets, so to import median household income for each county, a macro is created using techniques such as metadata (to find data set name), regular expression (to find the needed column), SQL etc. Then another macro is created to plot the line chart for crime rate vs 1/income for each county (P23-82). Examples of 2 counties are shown below. Note that some counties (there are 62 counties in N.Y.) are deleted because they have no data for income or crime rate starting from 2007.





These non-linear line charts show that on the county level, crime rate and median income has almost no linear relationship. Thus we introduce the chemical dependence treatment admission to see if this variable affects crime rate within each county. First, I consider that different categories may affect crime rate differently, so scatter plots are made for each treatment category (P83-85). Example of the crisis category is shown below.



Since the scatter plot didn't indicate too much linear relationship between total admission per category and crime rate, so there is no reason to analyze categories. Also there are too many missing values if analyzing each program category as a predictor, so I will use only total admissions per county per year. A simple random selection is conducted to split the data set into a training set and a test set. A linear regression model is finally fitted for the training set as:

$$\text{crime rate} = \beta_0 + \beta_1 * \frac{1}{median\ income} + \beta_2 * total\ admission + e$$

and we get the result of

$$\text{crime rate} = 1994.52 + 220.16 * \frac{1}{median\ income} + 250.73 * total\ admission + e$$

Since the variables are standardized, the model can be interpreted as, if a county has the mean median income and mean total admission for treatment program among all counties, the crime rate is expected to be $1994.52/100000 = 0.020$, and the effects of median income and total admission on crime rate are similar, with total admission having slightly more impact (Statistics on P89).

When fitting this model to the test set, the MSE is very high, indicating that though there is certain relationship between crime rate and median household income and treatment program admission, the model is not very good, so this will just be a reference for calculating or predicting crime rate in counties within New York State.

Item specification:

Program folder includes the SAS program and SAS log.

Output folder includes the PDF output of the program.

Data folder includes all data set imported into the program. Websites of the data sets are:

- https://catalog.data.gov/dataset/index-violent-property-and-firearm-rates-by-county-beginning-1990
- https://catalog.data.gov/dataset/chemical-dependence-treatment-program-admissions-beginning-2007/resource/ecb2186e-54ae-425e-bcea-0926c16268be
- https://www.statista.com/statistics/205974/median-household-income-in-new-york/
- https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
    - Data sets obtained by guided search