Investigation on Relationship between Power Plants and Cancer in the U.S.

Introduction

Cancer has always been a hot topic in the United States. People always looking for sources that cause cancer. Some people claims that the radiation from nuclear power plants will cause citizens living around them to have cancer. Also, some study tells that the pollution from burning coal will cause people to get lung cancer more effectively. Based on the data about number of cases from each county in the United States and the locations of all power plants, this project tends to find some relationships between power plants and cancer.

Data Collection

The first dataset is from the National Institutes of Health, which is called USA_CancerRates_All_ByCountry.csv. It is the cancer incidence rate report by county in United States from 2012-2016. It contains the adjusted cancer cases per 100,000 people of most of the counties in the United States. The second dataset called US_Dataset_PowerPlants_Locations_Nature_County.csv, which contains all the power plants in the United States and their primary source. This dataset is gained from U.S. Energy Information Administration.

Methods

The first part is cleaning the dataset. This project used Python to combine the two datasets based on the latitude and longitude. Then for the analysis part, Tableau is great in map visualization and combinations of two kinds of data, such as showing both cancer cases and locations of certain primary source of power plants. Therefore, the following map will be shown using Tableau.

Analysis

Overall Cancer Cases Exploratory Analysis

The following map shows the overall heatmap for adjusted cancer cases per 100,000 people from the dataset. The map shows that mid United States has relatively high adjusted cases than other areas.

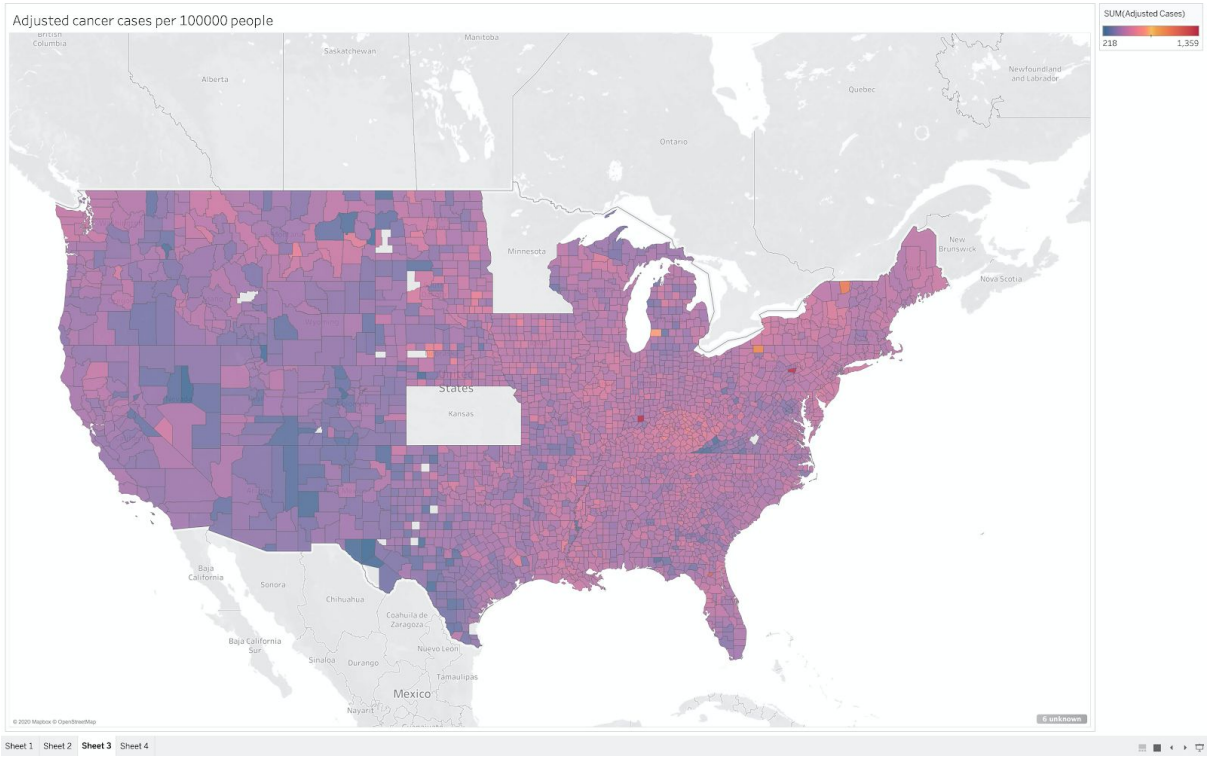

fig.1 heat map of adjust cancer cases

And using the bar chart, I tend to find the top 20 counties with highest adjusted cancer cases, as shown in figure 2.

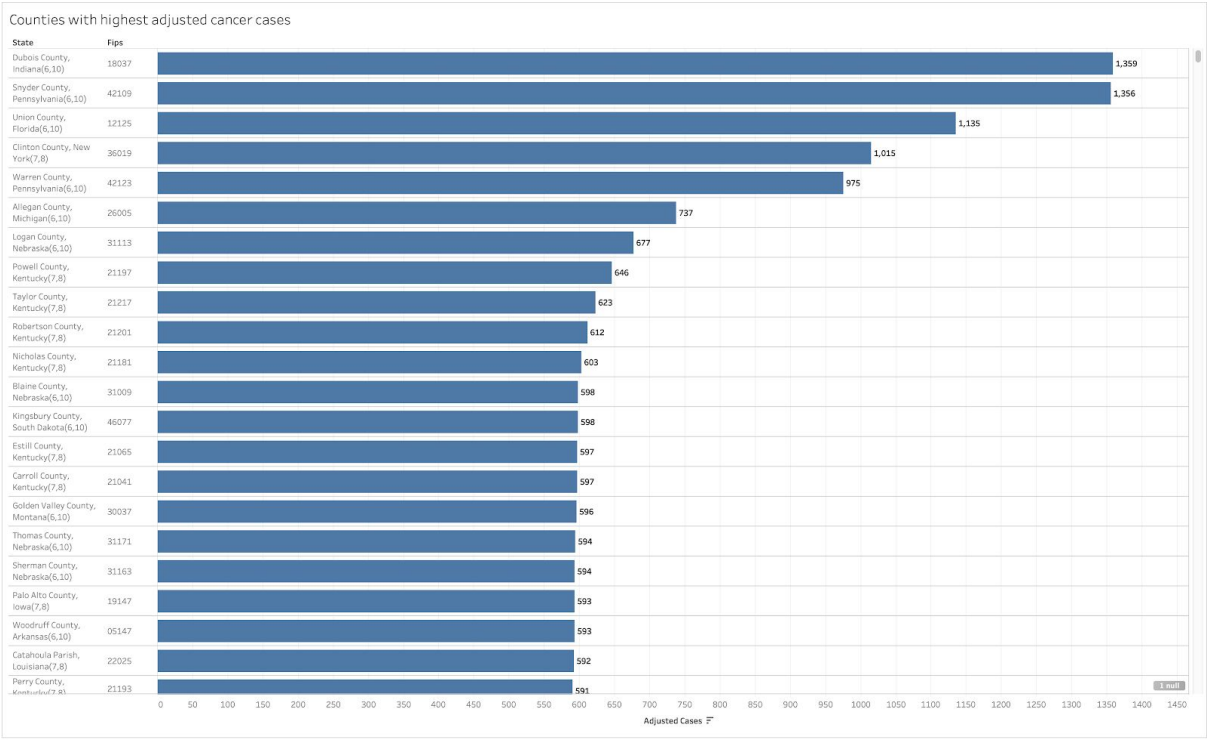

fig.2 bar chart of counties with highest adjusted cancer cases

In the graph, the top 20 counties contain 6 counties in Kentucky, 4 counties in Nebraska, 2 counties from Pennsylvania, 1 county each in Indiana, Florida, New York, Michigan, South Dakota, Montana, Arkansas and Iowa.

Are Power Plants Related to High Cancer Cases?

Based on the given datasets, the first part of the analysis will contain exploratory analysis on the relationship between power plants and cancer cases based on state. The above analysis tells that Kentucky and Nebraska contain about 10 in top 20 highest adjusted cancer cases in United States. The following graphs will show the locations of the power plants and what those power plants are.

The first graph shows all the locations of power plants in the United States. The number of power plants in an area is based on the population of that area. West coast and east coast are loaded with power plants while there are not so many power plants in midwest.

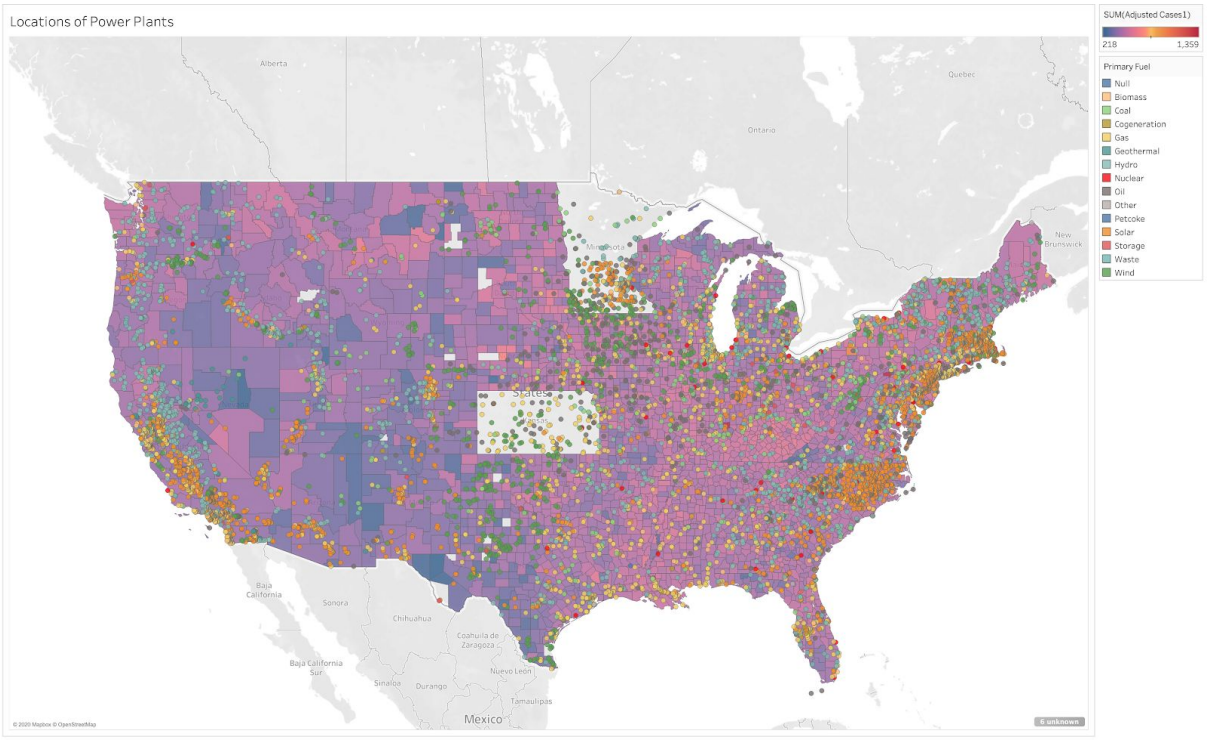

fig. 3 all locations of power plants

The following graph tells the locations of power plants and the adjusted cases of cancer in Kentucky, which is the state containing most top 20 counties of adjusted cancer cases.

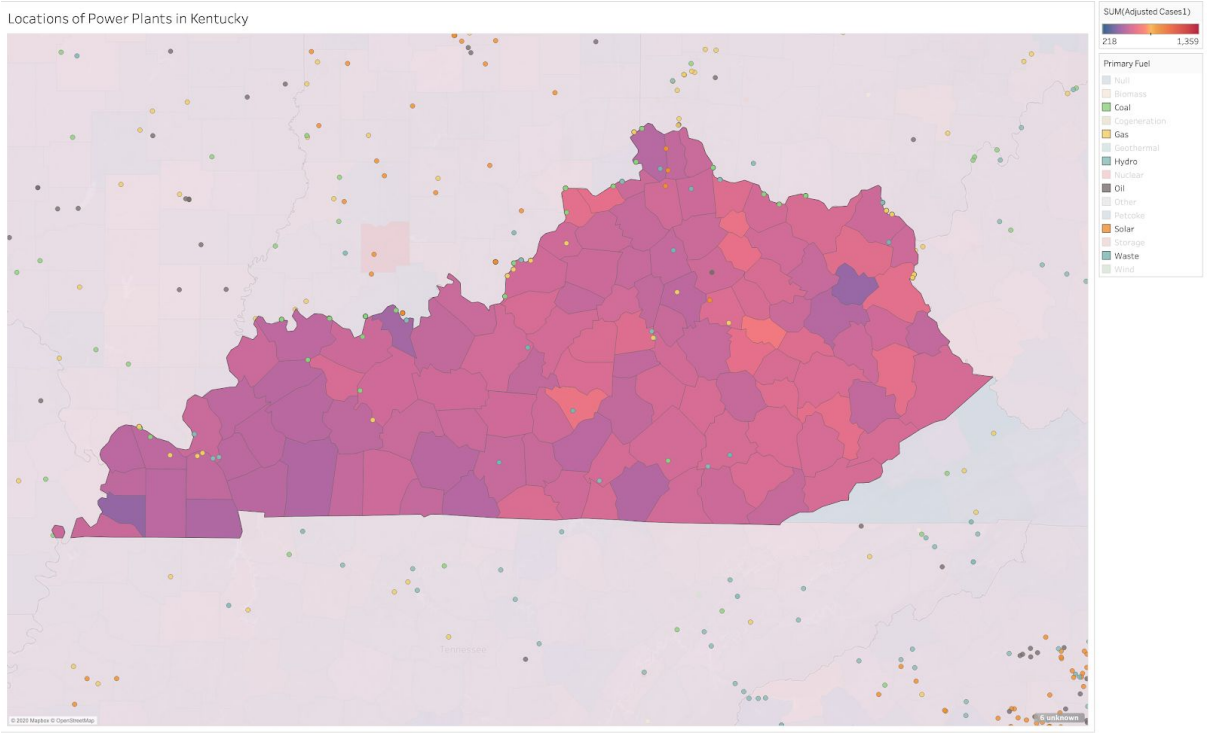

fig. 4 all locations of power plants in Kentucky

The state of Kentucky only contains 6 types of power plants, which are coal, gas, hydro, oil, solar and waste. And most of the power plants are located on the border line. There are no strong evidence that power plants are related to high adjusted cancer cases in Kentucky.

Coal Power Plants

According to Centers of Disease Control and Prevention, lung and bronchus cancer is the third highest rates of new cancer in United States. Also, based on the new study from Havard T.H. Chan School of Public Health, “The more a country relies on coal-fired power plants to generate energy, the greater the lung cancer risk is among its citizens”. Therefore, the following graph gives all the coal-fired power plants in the United States.

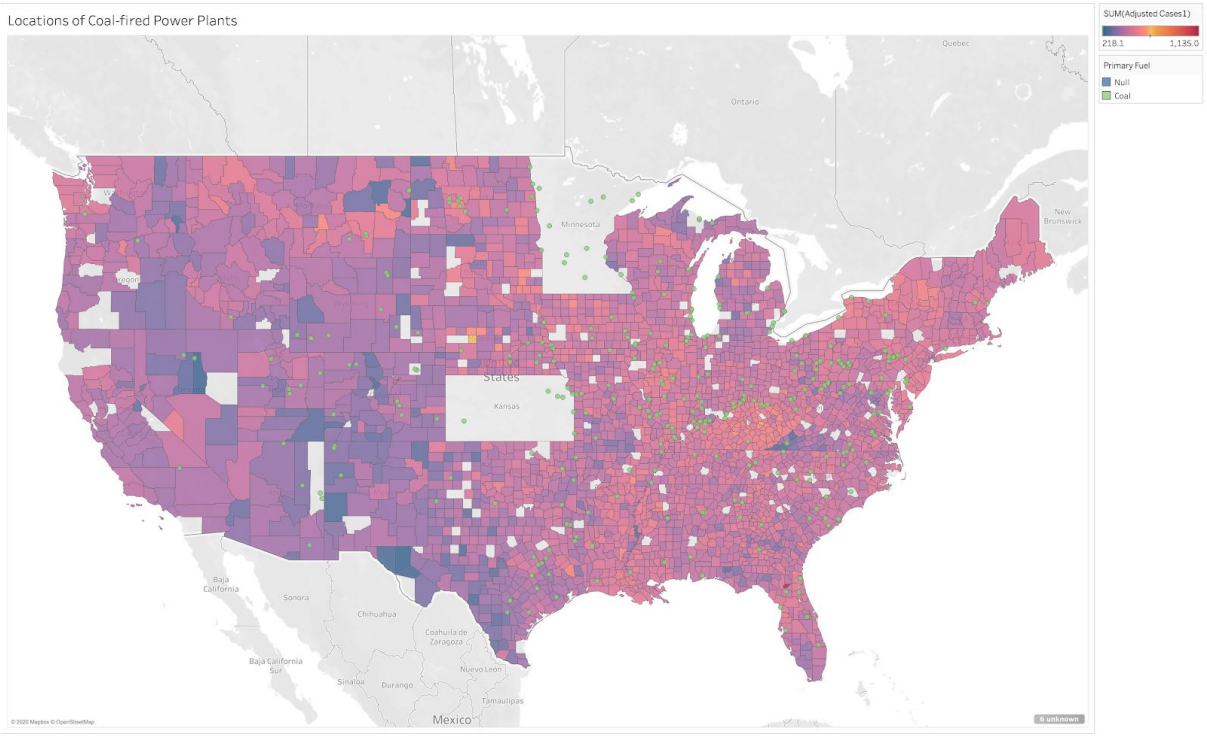

fig. 5 all coal power plants in U.S.

As shown in the map, most green points which are the locations of coal-fired power plants are located in the mid and mid-eastern of United States, which are also the areas with high adjusted cancer cases. Therefore, from the given map, there can be some positive correlation between numbers of coal-fired power plants and cancer cases.

Nuclear Power Plants

The following map shows the locations of nuclear power plants in the United States. There are not many nuclear power plants in the United States compared to other sources and most of the plants are located in the east side. And there are not nuclear plant located in Kentucky, and 1 nuclear power plant locate on the border of Nebraska. There's no clear evidence like coal power plants above to show the relationship between nuclear plant and cancer cases.

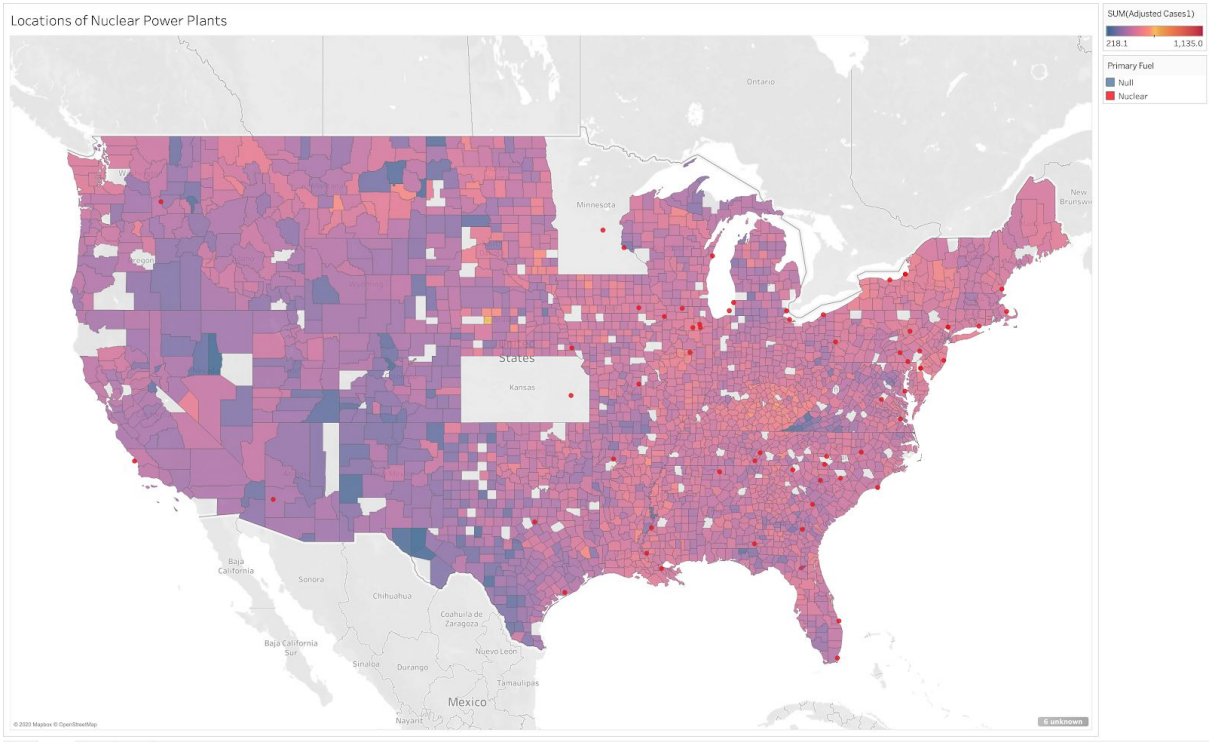

fig.6 all nuclear power plants in U.S.

Results

From the analysis above, the project shows that being the state with 10 out of 20 top adjusted cancer cases in the United states, there's no much relationship between the locations of power plants and the number of cancer cases. Also, there's no evidence to prove that nuclear power plants will increase the number of cancer cases of that area. However, from the locations of coal-fired power plants and the study about the coal burning and lung cancer, there seems to have a relationship between coal-fired power plants and higher numbers in cancer cases.

References

https://gis.cdc.gov/cancer/USCS/DataViz.html

https://news.harvard.edu/gazette/story/newsplus/reliance-on-coal-linked-with-lung-cancer-incidence/