RIT | Saunders College of Business

White paper on factors relating to cancer deaths in the US

November 6, 2022

Arpita Sheth

Neha Awasthi

Sanjeev Hirudayaraj

RIT | Saunders College of Business

1. INTRODUCTION

The United States has experienced an increase in Cancer rates, and it continues to be the second most common cause of death in the US after heart disease. A total of 1.9 million new cancer cases and 0.6 million deaths from cancer are expected to occur in the US in 2022, which is about 1,670 deaths a day. Though there is substantial progress in reducing the overall mortality rate due to cancer, significant socio-geographic disparity across the Unites States persists.

In our current investigation, we explore various factors relating to cancer incidence and death in the US, considering information at the county level including median income, cancer incidence rate, and mortality rates.

2. SCOPE OF ANALYSIS

We have considered the economic factors, population, and trends sourced from publicly available data from the US census website[1] and state cancer profiles[2].

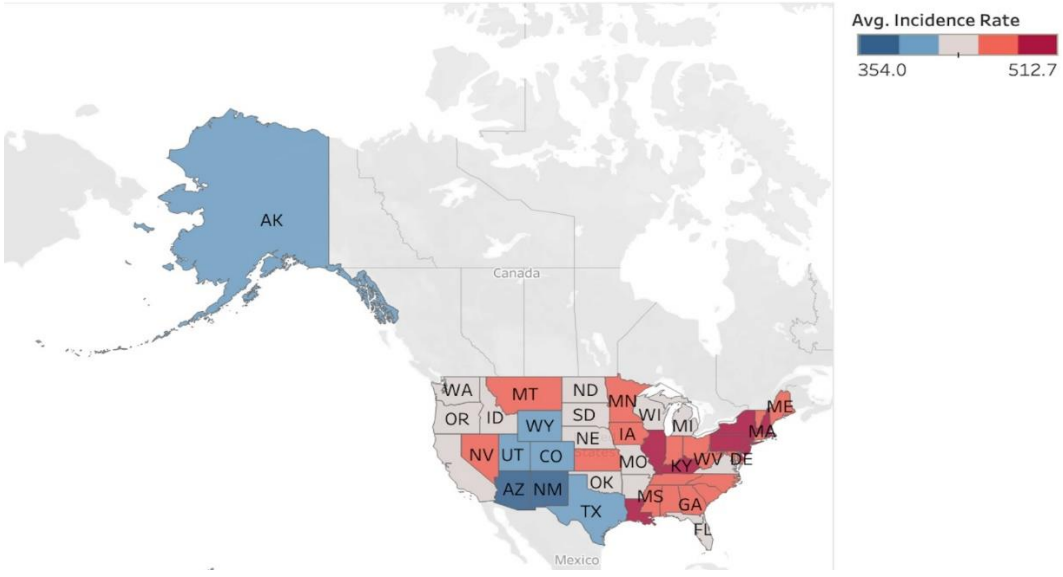
Our study includes an analysis of the data mentioned above to explore the impact of the economic disparity in cancer incidence and death rates across the United States.

Additionally, we have also tried to explore associations among various factors present in our data.

3. DETAILED ANALYSIS

3.1. Regions in the United States most prone to Cancer:

We have tried to find the regions most prone to cancer based on the average incidence rate per state. The following geographic heat map demonstrates the cancer incident hot spot to be in the Southern belt.



[1] https://www.census.gov/popest/data/counties/totals/2015/index.html

[2] http://statecancerprofiles.cancer.gov/incidencerates/index.php?stateFIPS=51&cancer=071&race=00&sex=0&age=001&type=incd#results

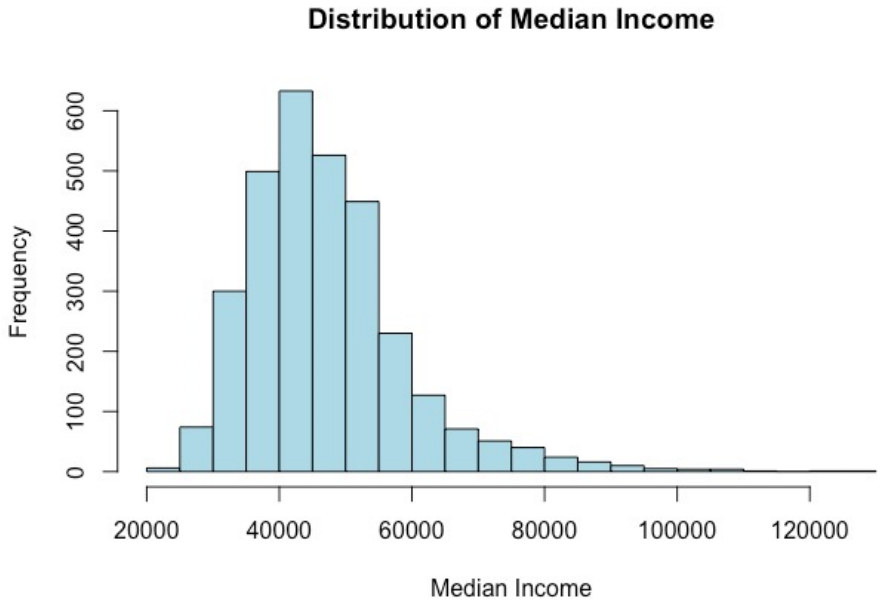
We have identified the following top five states as most prone to cancer:

Sr no	State	Average Incidence rate
1	Kentucky	513
2	Delaware	502
3	New York	498
4	New Jersey	495
5	New Hampshire	486

In the current finding, we have not considered the recent trends variable in our analysis of regions most prone to cancer due to missing information.

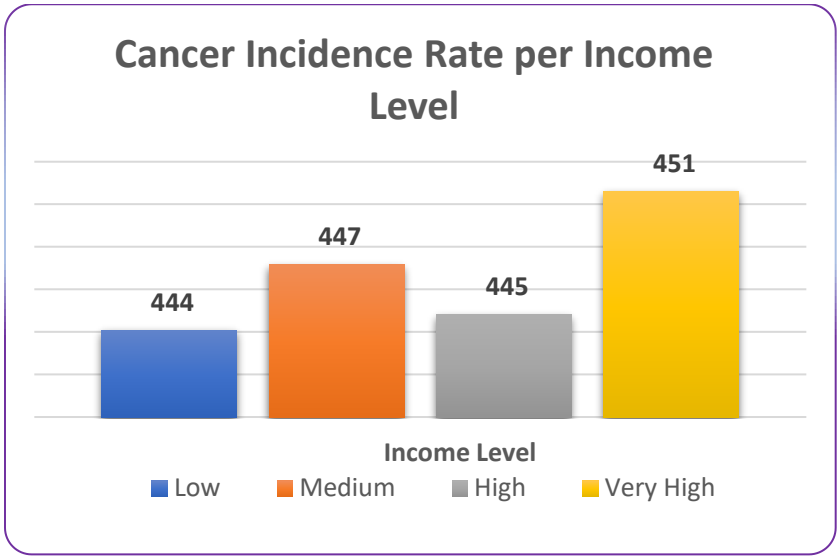
3.2. Cancer Incidence Rate in relation to Income level:

In the current finding, we have tried to identify associations between cancer incidence rates and median income across 3072 counties in the United States.
To categorize median income into four categories, we have tried to understand the distribution of median income.



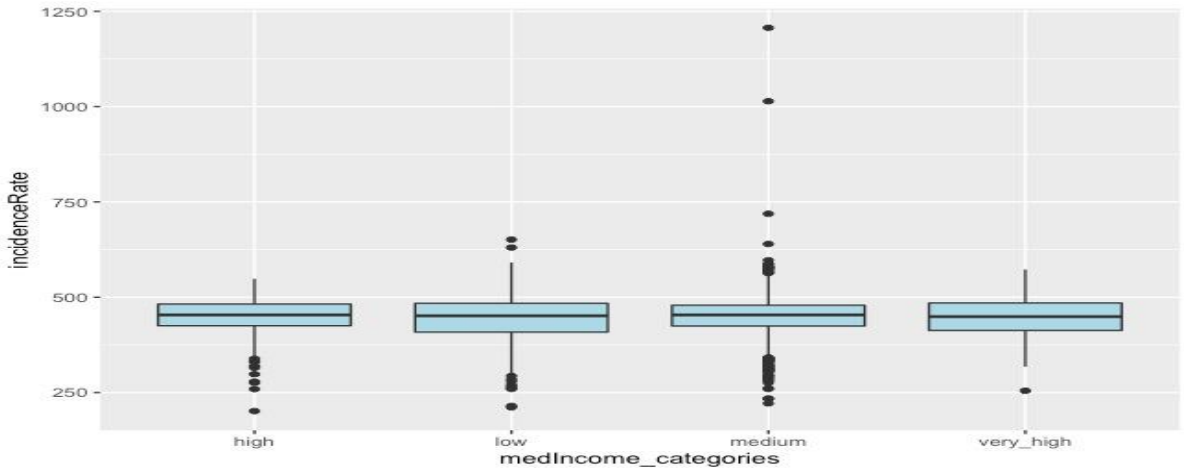
Based on the above histogram, we have categorized the median income into the following ranges:

Sr No	Income PA	Income Category
1	<=$ 39,999	Low
2	$ 40,000 -$ 59,999	Medium
3	$60,000- $ 79,999	High
4	>$ 80,000	Very High



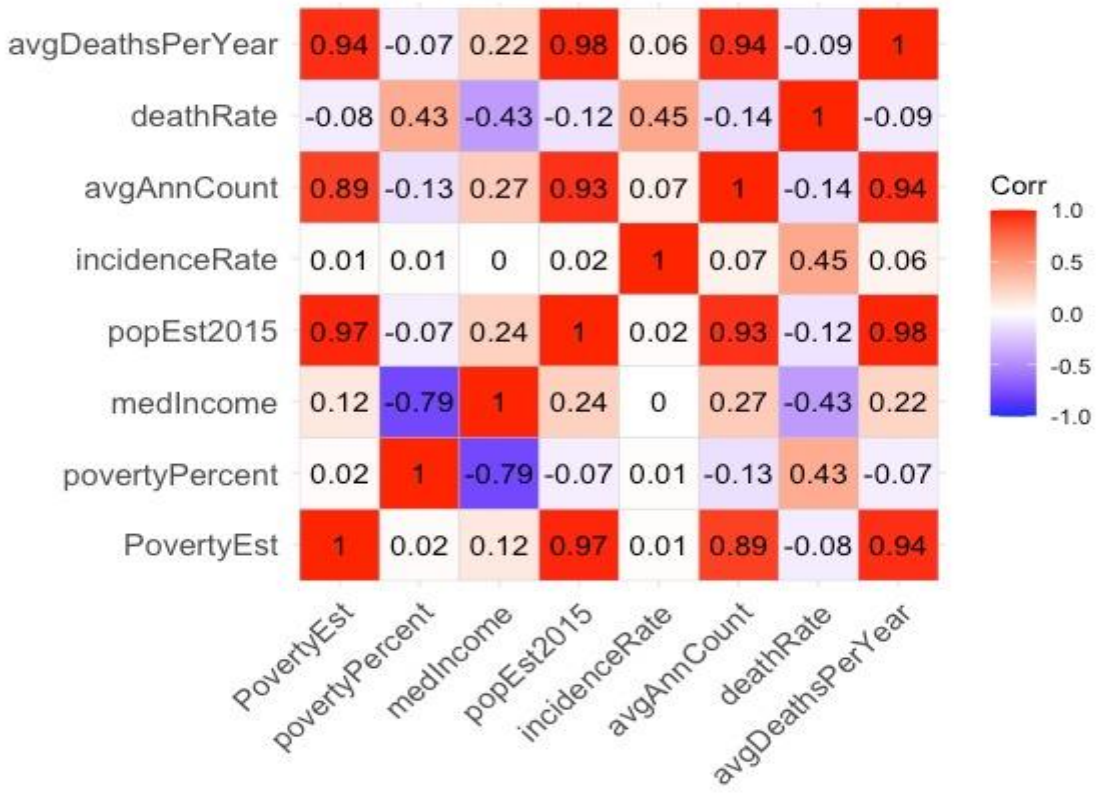
The average incident rate was 444 in counties with low income compared to 451 in counties with very high income. Hence, we can conclude that there is no direct relationship with the median income.

Additionally, we also tried to visualize the distribution of incidence rates across all the median income categories:



3.3. Analysis of the relationship between Incidence Rates and other factors:

We have tried to find meaningful relationships between various factors relating to cancer in the US using the following correlation matrix:



We can infer the following:

- Cancer incidence rate doesn't have any significant linear relationship with other numeric factors except for the cancer death rate which can be explained by the dependence of the death rate on the incidence rate.
- Cancer death rate, however, has a clear positive relationship with the poverty percentage and a clear negative relationship with the median county income by the same magnitude.
- There is also a very strong relationship between poverty estimate and average deaths per year due to cancer which shows that although poverty might not affect your chances of getting cancer significantly, it plays a significant role in your dying from it.
- Although the median county income and poverty estimate for a given time doesn't show any significant relationship with the cancer incidence rate, the historical average incidence rate per county shows a significant relationship with the poverty estimate over time.

- Interestingly, there is almost no relationship between the poverty estimate and the poverty percent although the poverty percent is derived from the poverty estimate.

3.4. Regression Analysis:

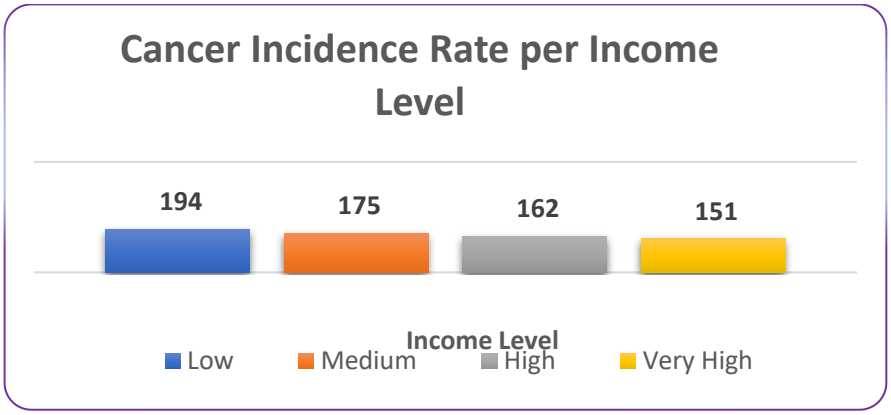
We built a linear regression model for cancer incidence rate. Some major insights derived from the model are:

- All the coefficients are significant signaling that a linear relationship between the incidence rate and the predictive factors.
- Historical average incidence rate per county has a positive impact on cancer incidence rate but not by a significant magnitude.
- Population estimate negatively affects the incidence rate but by a very insignificant amount.
- Recent trend in incidence rate is a significant predictor of incidence rate. If the recent trend is stable, it positively impacts the incidence rate by 11.3 units when compared to the falling recent trend. The rising recent trend positively impacts the incidence rate by 33.8 units as compared to the falling recent trend.
- We can see that missing values in recent trends negatively affect the cancer incidence rate. However, since we have clubbed together three kinds of missing values, the exact impact cannot be interpreted easily due to assumptions.
- While the model is significant and stable, it is not a good and accurate model to predict cancer incidence rate corroborating the results of the correlation analysis of incidence rate that shows almost no linear relationship with other factors.
- We might benefit from introducing non-linearity in the model or using a non-linear regression model if we want a useful predictive model for cancer incidence rate.

4. ADDITIONAL INSIGHT:

Cancer death rate and income level:

We have tried to find associations between the cancer death rate and income level across the counties. The cancer death rates showed an inverse relationship with income levels.



This socioeconomic disparity in cancer deaths can be attributed to soaring costs of cancer diagnosis and treatment, and better healthcare systems accessible by the affluent section of society.

5. LIMITATIONS

- Based on the limited availability of data, our analysis is limited to the year 2015.
- In our regression analysis, we have not considered the five-year cancer incidence trend as a significant factor due to the unavailability of complete information.
- We have been provided with a poverty estimate for the year 2014, however, the population estimate is provided for the year 2015. For the purpose of our analysis, we have ignored the differences in years.
- Behavioral and structural factors leading to cancer are beyond the scope of the analysis.
- The analysis does not aim to recommend any prevention or cure for any type of cancer.