

Project Report:

Cancer Incidence and

Socioeconomic

Analysis

Analysis Done by:

Basanth Kumar Varaganti

Email ID: bv8946@rit.edu

1. Project Overview

The primary objective of this project was to analyse cancer incidence and mortality across the United States, with a specific focus on identifying socioeconomic disparities. This analysis aims to support Rochester Regional Health (RRH) in understanding how factors like poverty and income levels impact cancer outcomes across different U.S. regions (East, West, North, South). The dashboard was designed to provide actionable insights that RRH can use to target healthcare interventions effectively.

2. Data Sources and Key Variables

The dataset used for this analysis includes comprehensive data across U.S. zip codes, covering the following critical variables:

- **Cancer Incidence Rate:** Number of new cancer cases within a specified region.
- **Mortality Ratio:** Proportion of cancer cases that result in death, highlighting the severity of outcomes.
- **Median Income (Med Income):** A measure of economic status within each region, providing insight into the overall wealth and access to resources.
- **Poverty Percent:** The percentage of the population living in poverty, indicating socioeconomic challenges.

Regional Classification of U.S. States for Analysis

To analyse cancer incidence and mortality more effectively, the United States was divided into four primary regions: **East, North, South, and West**. Each region comprises specific states, allowing for a clear comparison of socioeconomic and health outcomes.

Region	Subregion	States Included
East	Northeast	Maine (ME), New Hampshire (NH), Vermont (VT), Massachusetts (MA), Rhode Island (RI), Connecticut (CT), New York (NY), New Jersey (NJ), Pennsylvania (PA)
	Southeast	Delaware (DE), Maryland (MD), Virginia (VA), West Virginia (WV), North Carolina (NC), South Carolina (SC), Georgia (GA), Florida (FL), District of Columbia (DC)
North	Midwest (North-Central)	North Dakota (ND), South Dakota (SD), Nebraska (NE), Kansas (KS), Minnesota (MN), Iowa (IA), Missouri (MO), Wisconsin (WI), Illinois (IL), Indiana (IN), Michigan (MI), Ohio (OH)
South	South-Central	Kentucky (KY), Tennessee (TN), Mississippi (MS), Alabama (AL), Oklahoma (OK), Texas (TX), Arkansas (AR), Louisiana (LA)
West	Pacific	Washington (WA), Oregon (OR), California (CA), Alaska (AK), Hawaii (HI)
	Mountain	Idaho (ID), Montana (MT), Wyoming (WY), Nevada (NV), Utah (UT), Colorado (CO), Arizona (AZ), New Mexico (NM)

This classification allows for nuanced insights into regional trends in cancer incidence, mortality, and socioeconomic factors like income and poverty.

3. Methodology

This section outlines the steps taken to clean, organize, and visualize the data, allowing for a comprehensive analysis of cancer incidence and mortality rates across different socioeconomic backgrounds and geographic regions.

Data Cleaning and Preparation

- **Initial Data Examination:** The dataset was examined for missing values, inconsistencies, and outliers. Any anomalies were addressed to ensure data accuracy and reliability.
- **Categorization by Region:** To facilitate a regional comparison, U.S. states were categorized into four primary regions—**East, West, North, and South**. This classification allows for targeted regional insights and highlights geographical disparities in cancer outcomes.
- **Income Group Creation:** Based on the median income values, the data was further segmented into income groups (e.g., low, middle, high income). This grouping helps in understanding the relationship between socioeconomic status and cancer metrics, such as incidence and mortality rates.

Grouping by Region

- The states were systematically grouped into the four regions (East, West, North, and South) to analyse cancer incidence and mortality through a regional lens. This grouping allows us to uncover unique regional patterns and provides context for how cancer outcomes differ across the United States.

Visualization Approach

A selection of visualization techniques was employed to explore and highlight relationships among **incidence rate, mortality ratio, poverty percent, and median income**:

- **Heatmaps:** These visualizations identify areas with high poverty and high mortality rates, making it easy to spot socioeconomically disadvantaged regions that may require healthcare interventions.
- **Bar Charts:** Bar charts were used to compare metrics across regions and income groups, providing a straightforward view of disparities in cancer outcomes.
- **Stacked Bar Charts:** Stacked bars were implemented to show the composition of various factors (e.g., income levels within regions) and how they contribute to overall cancer metrics.
- **Interactive Filters:** Filters, including income group and region, allow users to explore specific subsets of data. This interactivity helps users dive deeper into particular demographic segments or geographic areas, enabling a more tailored analysis.

This approach allowed for an in-depth examination of cancer outcomes across different socioeconomic contexts, with interactive elements that support user-driven exploration.

4. Key Findings and Insights

Highest Cancer Incidence Areas

- Analysis reveals that the **Northeast and South regions** show the highest cancer incidence rates across the United States.

- **Northeast:** States like **New York, Pennsylvania, and New Jersey** exhibit elevated cancer incidence rates. This may be due to a combination of urban pollution, lifestyle factors, and higher population density, which often correlates with increased cancer cases.
- **Southern States:** States including **Kentucky, Mississippi, and Alabama** have notably high cancer incidence. This could be influenced by lifestyle factors prevalent in the region, such as smoking and dietary habits, as well as a lower availability of preventive healthcare.

States with the Highest Mortality Ratios

High Mortality Ratio States

The **mortality ratio**, indicating the lethality of cancer cases, is particularly high in **California (CA), Arizona (AZ), Florida (FL), and Nevada (NV)**. These states have higher proportions of cancer deaths relative to incidence rates, which may point to challenges in healthcare access and socioeconomic barriers impacting patient outcomes.

Top Mortality States

States like **California, Arizona, Florida, and Nevada** exhibit some of the highest cancer mortality ratios in the country. These states have varied demographic and socioeconomic factors that may limit access to timely and effective cancer care.

Contributing Factors

- **Healthcare Access:** In certain areas within these states, healthcare facilities may be sparse, especially in rural or underserved urban areas,

leading to delayed diagnoses and limited access to advanced treatment options.

- **Socioeconomic Barriers:** Economic challenges and high poverty levels within specific communities can hinder access to healthcare services. Individuals may face obstacles such as the inability to afford treatments, lack of transportation to specialized care facilities, or insufficient insurance coverage, all contributing to higher mortality rates.

Socioeconomic Outcomes and Regional Cancer Mortality Hotspots

Poverty's Impact on Cancer Mortality by Region

- **South Region:** States in the South, exhibit a strong correlation between high poverty levels and high cancer mortality. Poverty restricts access to treatment and preventive healthcare, creating mortality hotspots, especially in rural, low-income areas. While **Texas (TX)** has a relatively higher median income within the region, there are still significant disparities across Southern states.
- **Northeast Region:** Certain states like **Pennsylvania** and **West Virginia** show pockets of poverty that correlate with elevated cancer mortality. In these areas, urban poverty and environmental factors (such as industrial pollution) contribute to higher cancer rates. **New Jersey (NJ)**, however, has a comparatively higher median income within the East, which can influence better healthcare outcomes.

Income Disparities and Healthcare Access

- **Income and Cancer Outcomes:** There is a noticeable trend across all regions where lower-income populations experience higher cancer mortality rates. This disparity is particularly severe in the South, where

poverty and mortality rates are both elevated. Higher-income states within regions, like **New Jersey (East)**, **California (West)**, **Minnesota (North)**, and **Texas (South)**, may show relatively better cancer outcomes due to improved healthcare access.

- **Healthcare Infrastructure:** Regions with well-developed healthcare infrastructures, such as parts of the **West Coast** (e.g., **California** and **Washington**), tend to have lower mortality ratios. This is likely due to better access to healthcare facilities and preventive programs. In contrast, the South and some areas in the Midwest with limited healthcare networks face higher mortality rates, with rural and underserved areas lacking sufficient resources for timely cancer care and prevention.

4. Targeted Interventions Based on Regional Findings

- **High Priority Areas for Intervention:**
 - The **South** (particularly in Mississippi, Alabama, and Kentucky) should be a focus for Rochester Regional Health if considering expanding services. Interventions like mobile cancer screening units, community-based education, and financial support for treatment could reduce mortality in high-poverty rural areas.
 - **Urban Low-Income Regions in the Northeast:** Targeting high-incidence urban areas with tailored healthcare outreach can improve outcomes. RRH might focus on preventive care in densely populated cities, particularly in lower-income neighbourhoods where environmental risk factors are also prevalent.

5. Recommendations for RRH

- **Prioritize High-Poverty Areas:** Consider focusing healthcare resources on regions with high poverty and mortality rates, especially in the South.

- **Increase Early Detection Efforts:** Implement screening programs in low-income areas to catch cancer cases early, which can significantly reduce mortality.
- **Collaborate with Local Organizations:** Partner with community organizations to address socioeconomic barriers to healthcare, such as transportation or insurance coverage.
- **Monitor Changes Over Time:** Track changes in cancer outcomes as socioeconomic factors evolve, allowing for adaptive interventions based on updated data.

6. Challenges and Limitations

- **Data Completeness:** Limited data in some areas may impact the accuracy of results, especially in less populated regions.
- **Age and Demographic Factors:** While the analysis considered income and poverty, further demographic details (such as age and ethnicity) could provide additional insights.
- **Non-Linear Relationships:** Socioeconomic impacts on health can be complex and influenced by multiple interacting factors, which were not fully explored in this analysis.

7. Future Directions

To build on this analysis, future studies could:

- **Incorporate More Demographic Variables:** Adding age, ethnicity, and education level would provide a more detailed view of cancer disparities.
- **Analyse Healthcare Access Factors:** Including data on healthcare facilities, insurance coverage, and preventive care availability could give a clearer picture of the barriers affecting cancer outcomes.

- **Longitudinal Analysis:** Tracking data over a more extended period could help identify long-term trends in cancer outcomes relative to socioeconomic changes.

8. Conclusion

This analysis has highlighted significant socioeconomic disparities in cancer outcomes across the U.S. By understanding these relationships, Rochester Regional Health can better tailor its healthcare interventions and resource allocation. Addressing poverty-related healthcare gaps in targeted regions could lead to improved outcomes and reduced mortality rates, making a measurable impact in high-risk communities.