Ranking of States Based on Combined Cancer Sites Prevalence Using TOPSIS Methodology

-Shrinidhi Rajesh

ABSTRACT

This study introduces a comprehensive data preprocessing and analysis framework applied to a dataset encompassing cancer incidence and mortality rates across diverse regions in the United States. The meticulous data cleaning process involved the selection of specific cancer sites and focused on data from all 50 states, resulting in a refined dataset that includes pertinent information such as age-adjusted rates. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method was employed to rank states based on multidimensional cancer age-adjusted rates. Three distinct weighing methods were applied separately for both incidence and mortality events. Sensitivity analyses were conducted on the weighing matrix, including an equal importance approach, a population proportion-based method, and a method based on weights derived from the Centers for Disease Control and Prevention (CDC) website. Notably, results indicated that CDC-based weights were the most suitable for this dataset. Within the TOPSIS analysis, the assigned ranks provide a clear picture of how each state stands in terms of the severity of both mortality and incidence rates for the specified cancers. If a state ranks highest, it's experiencing more profound challenges, whether in terms of the number of deaths or the occurrence of new cases, emphasizing the urgency for targeted public health interventions. The rankings were based on the combined age-adjusted rates for all nine cancer types, adjusted to facilitate fair comparisons between states or regions. The final TOPSIS rankings identified Washington as the state with the highest mortality rates and Wyoming with the lowest. Additionally, Alabama exhibited the highest cancer incidence rates, while Colorado had the lowest. The study concludes by using CDC-based weights for this specific dataset where the state's rank in the TOPSIS analysis serves as a comprehensive indicator of its relative standing in terms of mortality and incidence rates for the specified cancers. These findings contribute to a nuanced understanding of state-level variations in cancer incidence and mortality, underscoring the necessity for tailored approaches in public health interventions.

TABLE OF CONTENTS

Introduction
Methodology
Conclusion
Discussion
Limitations
Recommendation
References
Code Book

INTRODUCTION

Cancer, ranked as the second leading cause of death in the United States, poses a formidable health challenge that demands a meticulous understanding and targeted interventions. This study capitalizes on a comprehensive dataset curated from the Centers for Disease Control and Prevention (CDC), specifically the U.S. Cancer Statistics. This dataset amalgamates incidence data from the CDC's National Program of Cancer Registries and the National Cancer Institute's Surveillance, Epidemiology, and End Results Program, complemented by mortality data from the CDC's National Center for Health Statistics. Covering all regions in the US, this dataset incorporates pivotal variables such as age-adjusted rates, crude rates, population sizes, counts, gender-specific data, and racial attributes (Table 1).

Variables in the Dataset
AREA: This column represents the geographical area or region associated with the cancer data. It contains states, counties and regions in US.
AGE_ADJUSTED_CI_LOWER: This column contains the lower bound of the confidence interval for the age-adjusted cancer rates.
AGE_ADJUSTED_CI_UPPER: This column contains the upper bound of the confidence interval for the age-adjusted cancer rates.
AGE_ADJUSTED_RATE: This column represents the age-adjusted cancer rate, which is a measure that accounts for the age distribution of the population.
COUNT: This column contains the count of cancer cases or events for a specific area, year, and other relevant parameters.
EVENT_TYPE: This column specifies the type of cancer event, such as incidence or mortality.
POPULATION: This column represents the population size concerning which the counts are calculated.
RACE: This column indicate the racial or ethnic group associated with the cancer data. There are a total of 5 different categories mentioned in this dataset.
SEX: This column indicates the gender (male or female) associated with the cancer data.
SITE: This column specifies the anatomical site or type of cancer being considered. It represents different cancer types (e.g., lung, breast, colorectal).
YEAR: This column represents the year in which the cancer data was recorded or reported.
CRUDE_CI_LOWER: Similar to the age-adjusted confidence interval, this column contain the lower bound of the confidence interval for the crude (unadjusted) cancer rates.
CRUDE_CI_UPPER: This column contains the upper bound of the confidence interval for the crude cancer rates.
CRUDE_RATE: This column represents the crude (unadjusted) cancer rate, which is the raw rate without accounting for the age distribution of the population.

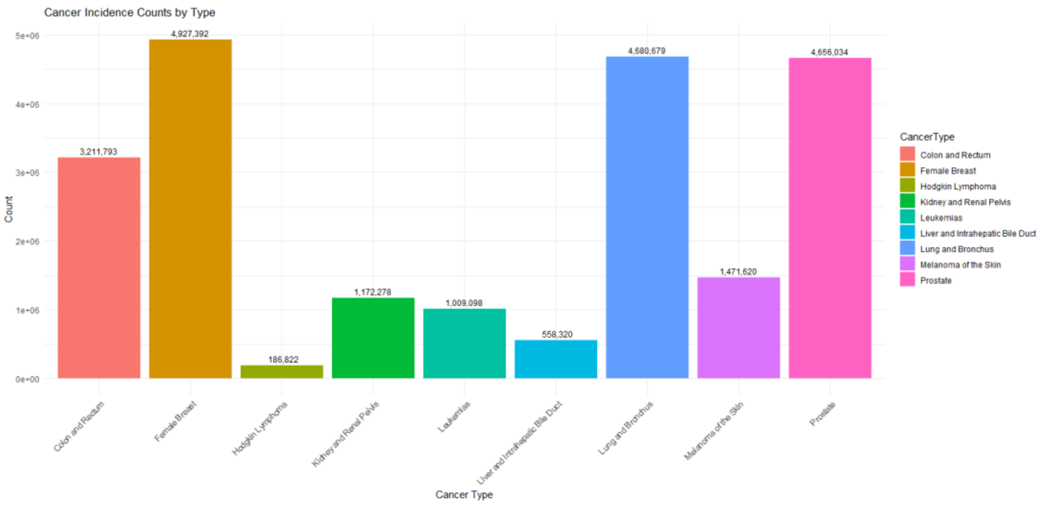
Table 1: Variables in the Dataset

The study introduces a robust framework involving data cleaning, the application of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method, and the exploration of various weighing strategies. The primary goal is to unveil state-level differences in cancer outcomes by ranking states based on their combined cancer age-adjusted rates. This approach provides insights into state performance concerning both mortality and incidence rates across specific cancer sites through TOPSIS analysis. Recognizing these state-level variations is crucial for informing public health interventions and optimizing resource allocation. The project's core objective is to pinpoint states with elevated mortality and incidence rates, enabling the public health sector to address these unique cancer types and their prevalence. It's pivotal to note that this ranking is anchored in age-adjusted rates, a consideration vital for accurately reflecting the relative burden of cancer, especially given higher rates in older populations. This choice is paramount, as age-adjusted rates provide a normalized metric that accounts for both counts and population variations across different age groups. By homogenizing the impact of age, the analysis ensures a fair and accurate comparison between states, essential for deriving meaningful conclusions. The study's insights aim to guide targeted interventions, acknowledging the distinct challenges and needs of each state in the battle against cancer. The dataset underwent multiple filtrations to ensure relevance, including the selection of specific cancer sites known for their substantial impact on public health. Subsequently, the dataset was refined to encompass only the 50 states, streamlining the scope for a more focused analysis. Noteworthy challenges arose in handling missing data in the USCS incidence dataset, particularly in years with data gaps for specific states like Mississippi and South Dakota. Addressing these gaps was crucial, and my approach utilized Multiple Imputation by Chained Equations (MICE) for its robustness in imputing missing values, ensuring the dataset's integrity. Redundancy was managed by focusing on specific gender and racial categories, simplifying the dataset for a more straightforward analysis. The collaborative efforts highlighted by Kohler et al. showcase major entities, including the American Cancer Society, CDC, National Cancer Institute, and North American Association of Central Cancer Registries. This collaboration emphasizes the annual production of updated national cancer statistics, ensuring a robust foundation for understanding cancer trends and informing public health strategies in the United States. The referenced article delves into the specific focus on breast cancer incidence by molecular subtype, explains its variations over the years and is based on all the regions in the United States. This study broadens this perspective by examining multiple cancer sites. Further articles by Senkomogo et al, provide insights into the global burden of cancer, with a focus on low- and middle-income countries (LMICs), and outline the CDC's role in supporting workforce capacity development through Field Epidemiology Training Programs (FETPs). The success of pilot testing the curriculum and its impact on increasing cancer knowledge among trainees underscore the tangible outcomes of the CDC's technical assistance. Our study, while focusing on state-level disparities in cancer outcomes, aligns with the global perspective by offering a comprehensive analysis that draws parallels with a world-level approach. By examining a multitude of cancer sites and considering the diverse factors influencing state-level variations, our research contributes to the broader understanding of cancer incidence and mortality. This approach adds a better understanding of the diverse landscape of cancer outcomes, taking into account various forms of the disease and their unique implications.

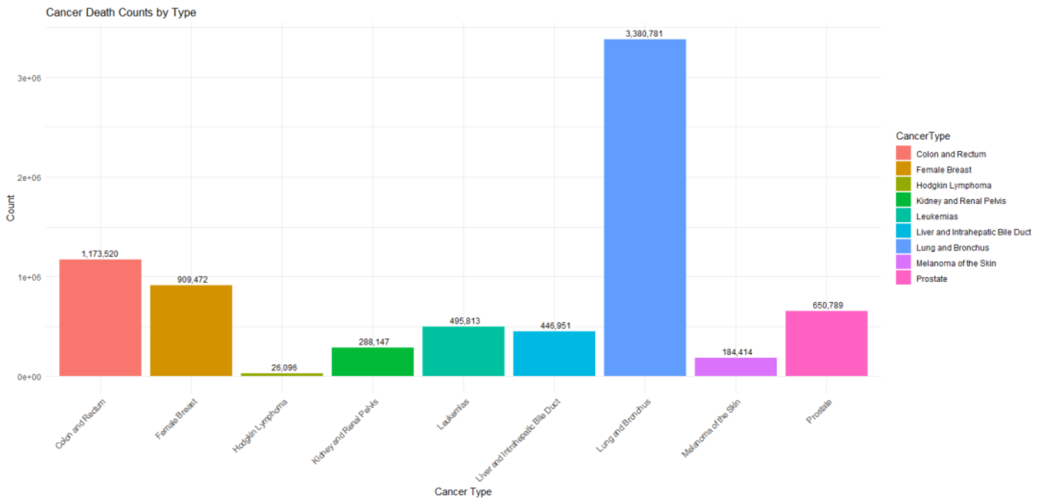
The purpose of this study is to comprehensively analyze cancer incidence and mortality rates across the United States, utilizing a robust data preprocessing and analysis framework. The primary objectives are to employ the Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) method to rank states based on multidimensional cancer age-adjusted rates. This ranking provides a comprehensive indicator of each state's relative standing in terms of severity for both mortality and incidence rates across specific cancer sites. Ensure precision in the analysis by implementing meticulous data-cleaning processes. This involves selecting specific cancer sites, focusing on data from all 50 states, and addressing missing values using Multiple Imputation by Chained Equations (MICE). Focus on age-adjusted rates as a key metric for fair state comparisons, considering the age distribution in each state's population. This adjustment aims to provide a more accurate representation of the relative burden of cancer. Finally, contributes valuable insights to guide interventions by identifying states with higher mortality and incidence rates. Strategies can then be developed to address the prevalence of cancers in highly ranked states.

METHODOLOGY

The study focuses on a cancer dataset obtained from the CDC by loading the dataset, containing information on cancer incidence and mortality rates across diverse regions in the United States. The dataset is split into two distinct datasets based on the Event Type, focusing on cancer incidence and mortality events separately. Similar data processing steps are applied to both datasets to ensure consistency in the analytical approach. This project focuses on both data preprocessing and raking the dataset based on states using the TOPSIS approach. A systematic approach is employed to clean the data, focusing on the following key steps. Specific cancer sites are identified as crucial for public health analysis, narrowing down the scope of the study. The dataset is refined to include only the 50 states, ensuring a targeted state-level analysis. Initial exploratory data analysis is conducted to understand the structure and features of the dataset. Table 2 and 3 contains overall information on the Rates spread against all the states in the USA from the year 1999-2012 for the 9 different cancer sites. Plots 1 and 2 depict their overall counts based on each cancer type.



Plot 1: Incidence Counts by Cancer Types



Plot 2: Mortality Counts by Cancer Types

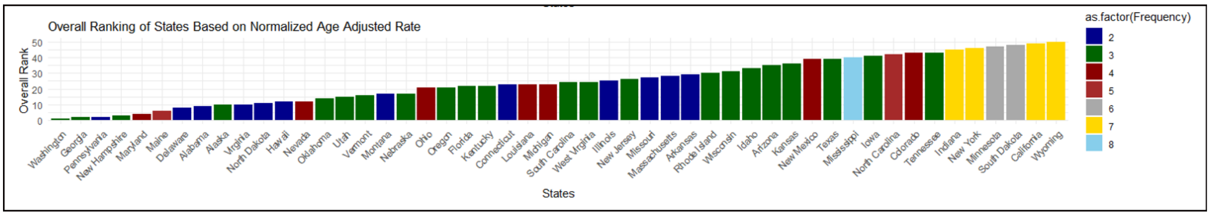
For both the split incidence and mortality dataset, only the age-adjusted rate column is retained for analysis, addressing the need for a normalized metric for fair state comparisons, as it accounts for the underlying variations over the ages. Redundancy is minimized by considering only the combined "Male and Female" gender category and the "All Races" category for race analysis. Multiple Imputation by Chained Equations (MICE) is applied to impute missing values, ensuring the integrity of the dataset. Anomalies, such as the combined "2008-2012" year category, are identified and rectified. Thus, all the other rates and intervals were omitted for this analysis. The cleaned data is pivoted into a wide format, creating a multidimensional dataset with columns representing 9 cancer site's age-adjusted rates, with the states and year columns(Plot 3, Plot 4). Redundant entries in the incidence dataset are removed to ensure a streamlined and accurate dataset. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method is employed for multi-dimensional state ranking based on combined cancer incidence and mortality rates (Table 4). For each of the two event types, the dataset is normalized by dividing each data point by the square root of the sum of squares on a column-wise approach. Three weighing methods are applied for sensitivity analysis: Equal Importance Method, Colum Wise Proportion-Based Method, and based Weights Method obtained in Table 1 and Table 2. The normalized matrix is weighted based on each of the selected weighing methods separately. Positive and negative ideals are determined for each column to calculate relative closeness. The relative closeness is calculated for each state, considering both positive and negative ideals. States are ranked based on relative closeness in descending order to obtain the states in the order of their combined prevalence of the chosen event i.e. mortality or incidence. The ranking is performed using each of the three weights. Overall state rankings are determined based on the frequency of ranks for each state across the years. If Alabama is ranked 1 5 times and other states are ranked 1 less than 5 times over 14 years, then Alabama becomes ranked 1. Also, the method was curated in such a way that if Alabama has the highest frequency even for Rank 20, it wouldn't be ranked 20, cause once a state is ranked then it gets removed from the ranking process. Comparative analysis is conducted to assess the robustness of different weighing methods (Plot 7, Plot 8). The study focuses on CDC based approach for this dataset, considering the nuances of age-adjusted rates. Additionally, the equal weights approach doesn't cause much of a change in the ranking compared to when no weights are used. Also, it should be kept in mind that we are using multiple features based on different cancer sites, and no one cancer is better or worse than the other. They are all caused to the human body, except for their different stages. However, this dataset focuses on cancer sites as a whole and not their stages. Thus, another way to weigh is by considering their adjusted rate proportion to this particular population of the dataset. Which seems to be a valid weighing method. Final rankings identify states with the highest mortality and incidence rates, contributing to a nuanced understanding of state-level variations in cancer outcomes (Plot 5, Plot 6). The methodology employed in this study enhances the understanding of state-level variations in cancer outcomes, providing valuable insights for targeted public health interventions and resource allocation strategies. In summary, the methodology integrates data cleaning, TOPSIS ranking, and sensitivity analyses, ensuring a comprehensive and robust approach to unraveling state-level patterns in cancer incidence and mortality rates. The study's findings are driven by a systematic and transparent analytical process, contributing to evidence-based decision-making in public health.

CONCLUSION

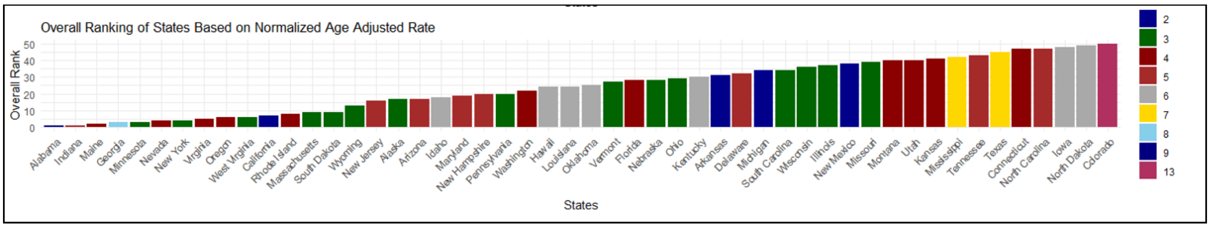
This study focuses on the intricate landscape of cancer incidence and mortality rates across the United States, employing a strict framework from data preprocessing to multi-dimensional state ranking. The dataset, sourced from the Centers for Disease Control and Prevention (CDC), underwent a rigorous cleaning process, focusing on specific cancer sites and refining the analysis to the 50 states. The Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) method was instrumental in revealing state-level variations in cancer rates. The choice of age-adjusted rates emerged as pivotal, providing a normalized metric for fair state comparisons that consider variations in population age distribution. Sensitivity analyses on weighing methods underscored the significance of CDC-based weights, aligning with the complexities of age-adjusted rates.

The state rankings yielded compelling results, with Washington standing out for the highest incidence rates and Wyoming exhibiting the lowest(Table 6). Alabama took the lead in cancer mortality rates, while Colorado has the lowest mortality rate of all cancer sites(Table 5). These findings contribute to the understanding of state-level disparities in cancer outcomes, emphasizing the urgency for tailored public health interventions. It is becoming thoughtful of the fact that each cancer site might have a higher prevalence in one particular state but very low in the other. This can be due to many more external factors affecting the cells to cause cancer. Further analysis of why a particular state is affected more, can help us come up with causes, which when controlled can help the cancer rate preventions, rather than curing after getting affected. The overall ranking is solely based on the state's frequency in the ranking.

Plot 11: Overall Ranking For The Incidence Rate



Plot 12: Overall Ranking For The Mortality Rate



DISCUSSION

This study unravels valuable insights into the complex landscape of cancer outcomes, providing a foundation for informed public health interventions. The state rankings emphasize the need for state-specific strategies. The weighing sensitivity analysis highlights the importance of methodological choices, urging researchers to align methodologies with the unique attributes of their datasets. However, it is also important to follow the weights based on a standardized approach. Since the standardized method wasn't applicable, this way of weighing aligns well with the dataset. The findings prompt further exploration into the determinants of state-level differences, considering socio-economic factors, healthcare infrastructure, and cultural influences. Collaborative efforts between states can facilitate knowledge exchange and the implementation of successful interventions. The identification of specific states with elevated cancer rates opens avenues for targeted research and resource allocation.

LIMITATIONS

Despite the rigor in methodology, this study is not without limitations. While the use of age-adjusted rates is essential for fair comparisons, this might obscure specific age-related patterns. One such example is that cancer rates tend to be higher in older age groups. The weighting sensitivity analysis, though comprehensive, relies on certain assumptions and is subject to variations based on the weighing method. Additionally, the dataset's reliance on CDC-based weights is specific to this dataset and may not be universally applicable.

Missing data, although addressed through Multiple Imputation by Chained Equations (MICE), introduces an element of imputation uncertainty. The study's focus on specific cancer sites might not encompass the entirety of cancer-related challenges, potentially overlooking emerging or less common cancer types.
Lastly, TOPSIS is a multidimensional method, the states are ranked based on all the 9 cancer sites combined rates for both incidence and mortality. Thus, this analysis might not indicate the rate of each 9 cancer sites in the US states. If individual cancer sites need to be measured based on their rates in a particular state, other approaches need to be performed. Additionally, each state's overall ranking is purely based on the frequency of the overall prevalence of the cancer sites. Thus, this ranking needs an in depth monitoring and analysis by including other socio economic factors to come up with other metric, rather than just based on the frequency to determine where they stand.

RECOMMENDATIONS

Ongoing monitoring of cancer outcomes at the state level is crucial. Longitudinal studies can unveil temporal trends and aid in the identification of emerging patterns. States identified with higher cancer mortality and incidence rates should implement targeted interventions. These interventions could encompass screening programs, public awareness campaigns, and improvements in healthcare access. Further, ranking is very well needed for each cancer type, which makes the practitioners more focused on that particular cancer type in that particular region. States with similar challenges can collaborate to share best practices and innovative solutions. Collaborative efforts can enhance the effectiveness of interventions and resource utilization. Further research should be into the socioeconomic factors influencing cancer outcomes. Understanding these determinants can guide the development of interventions that address root causes. As new datasets become available, periodic updates to weighing methodologies should be considered. This ensures that the analysis remains aligned with the evolving landscape of cancer data. Additionally, a proposed weighing method based on each cancer site needs to be devised. It is very challenging as the count and the population keeps varying. However,, a proposed weighing method based on all the data collected so far can help do multi-dimensional ranking with much higher perfection. This can result in cluster identifications if extended. Additionally, it is important to combine all the years together to understand which state has the highest prevalence. In this analysis it is ranked by a frequency approach. But it is necessary to come up with a metrics combining other socio economic factors to be able to rank them. This is important and avoids a lot of confusions when we want to compare the prevalence of each diseases seperatelely over the years and then rank them based on their aggregation over the years. This helps in identifying the trends and patterns over the year, instead of just taking the summation of all years together and then ranking their overall. In conclusion, this study provides a comprehensive snapshot of state-level differences in cancer outcomes, setting the stage for evidence-based public health interventions. Acknowledging the limitations, future research endeavors can build upon these findings, refining methodologies and expanding the scope to deepen our understanding of the complex dynamics of cancer incidence and mortality across diverse regions in the United States.

TABLES AND PLOTS

Step 1: Construct normalized decision matrix

$$r_{ij} = x_{ij} / \sqrt{(x^2_{ij})} \quad \text{for} \quad i = 1, \ldots, m; \; j = 1, \ldots, n \quad (1)$$

where x_{ij} and r_{ij} are original and normalized score of decision matrix , respectively

Step 2: Construct the weighted normalized decision matrix

$$v_{ij} = w_j \, r_{ij} \qquad (2)$$

where w_j is the weight for j criterion

Step 3: Determine the positive ideal and negative ideal solutions.

$$A^* = \{ v_1^*, \ldots, v_n^* \}, \quad (3) \quad \text{Positive ideal solution}$$

where $v_j^* = \{ \max (v_{ij}) \text{ if } j \in J ; \; \min (v_{ij}) \text{ if } j \in J' \}$

$$A' = \{ v_1', \ldots, v_n' \}, \quad (4) \quad \text{Negative ideal solution}$$

where $v' = \{ \min (v_{ij}) \text{ if } j \in J ; \; \max (v_{ij}) \text{ if } j \in J' \}$

Step 4: Calculate the separation measures for each alternative.

The separation from positive ideal alternative is:

$S_i^* = [\; (v_j^* - v_{ij})^2]^{½} \; i = 1, \ldots, m(5)$

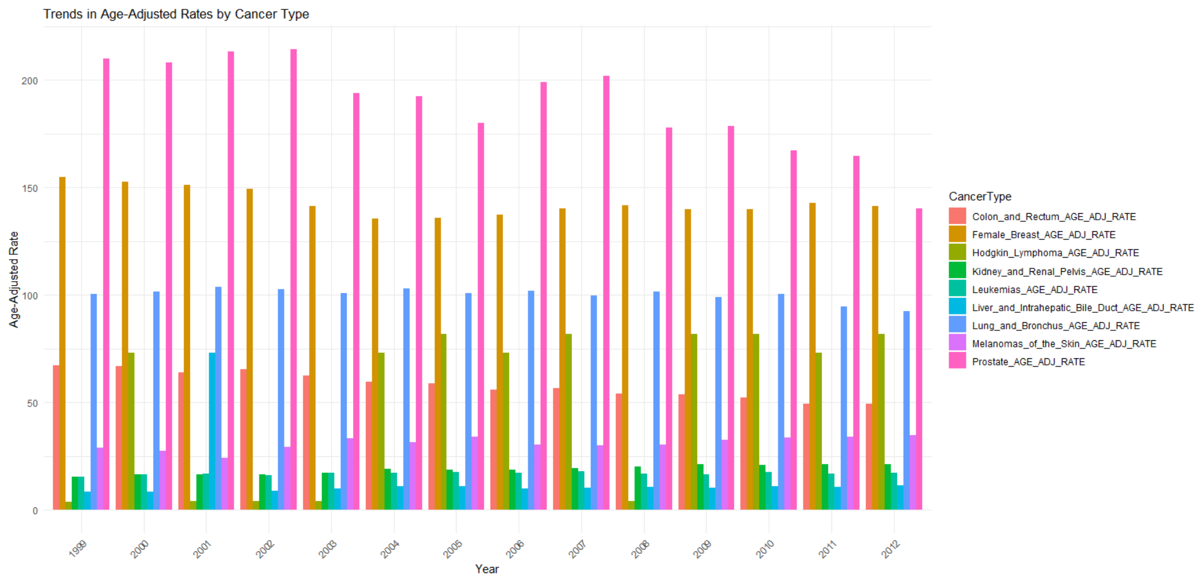
Similarly, the separation from the negative ideal alternative is:

$S'_i = [\; (v_j' - v_{ij})^2]^{½} \quad i = 1, \ldots, m \; (6)$

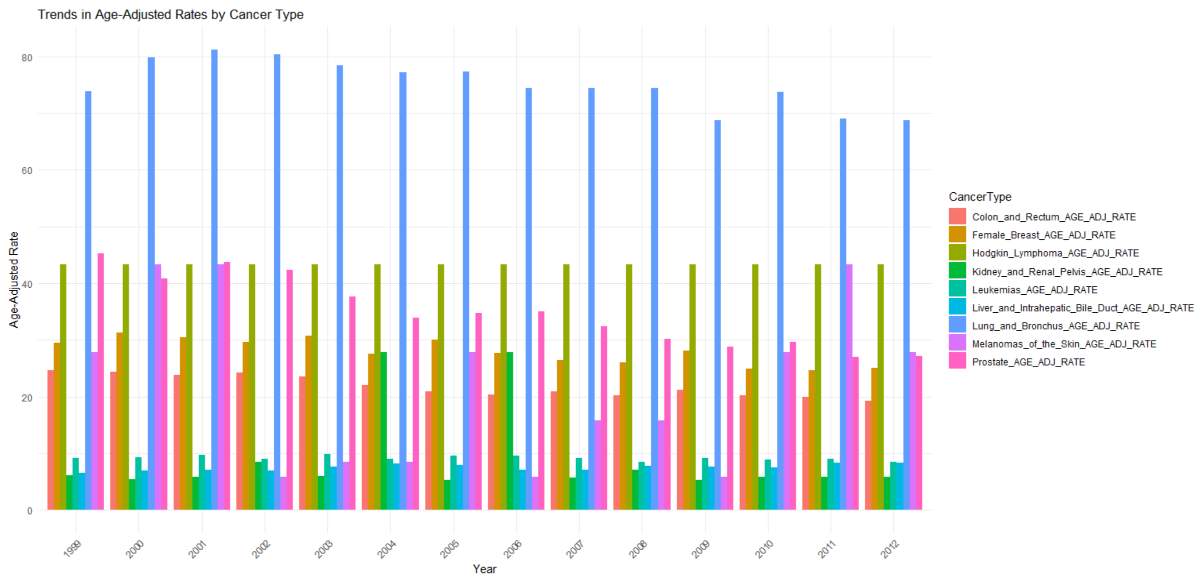
Step 5: Calculate the relative closeness to the ideal solution C_i^*

$$C_i^* = S'_i / (S_i^* + S'_i) \, , \quad (7) \quad 0 \quad Ci^* \quad 1$$

Table 4: TOPSIS Methodology



Plot 3: Incidence Age-Adjusted Rate Trends over the years

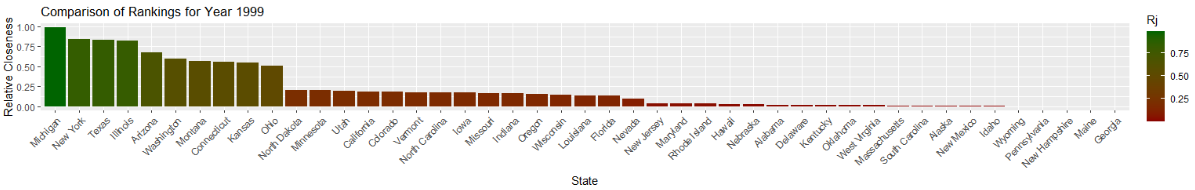


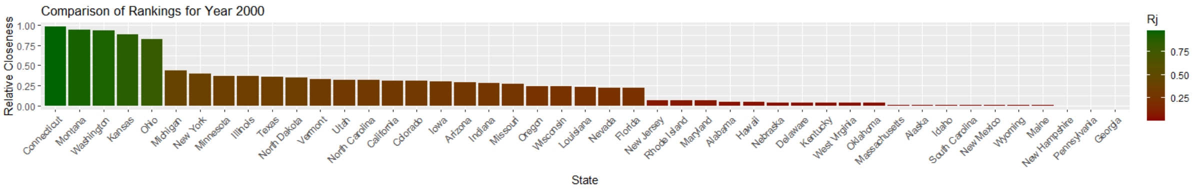
Plot 4: Mortality Age-Adjusted Rate Trends over the years

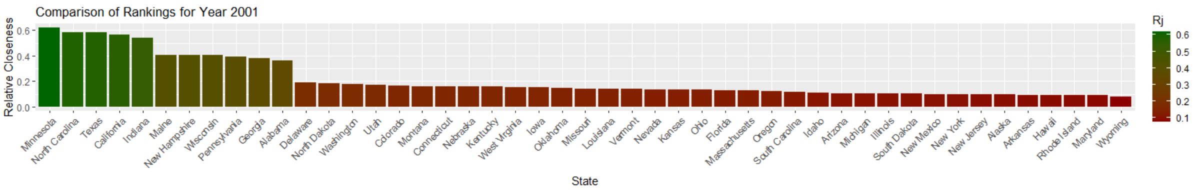
Mortality				
CANCERS	CRUDE_RATE	NORMALIZED_CRUDE_RATE	AGE_ADJ_RATE	NORMALIZED_AGE_ADJ_RATE
Lung and Bronchus	53	0.381569474	51.2	0.372634643
Colon and Rectum	18.2	0.131029518	17.6	0.128093159
Female Breast	27.2	0.195824334	23.8	0.173216885
Prostate	20	0.143988481	25	0.181950509
Leukemias	7.4	0.053275738	7.2	0.052401747
Liver and Intrahepatic Bile Duct	5.7	0.041036717	5.4	0.03930131
Kidney and Renal Pelvis	4.2	0.030237581	4.1	0.029839884
Melanoma of the Skin	2.8	0.020158387	2.7	0.019650655
Hodgkin Lymphoma	0.4	0.00287977	0.4	0.002911208

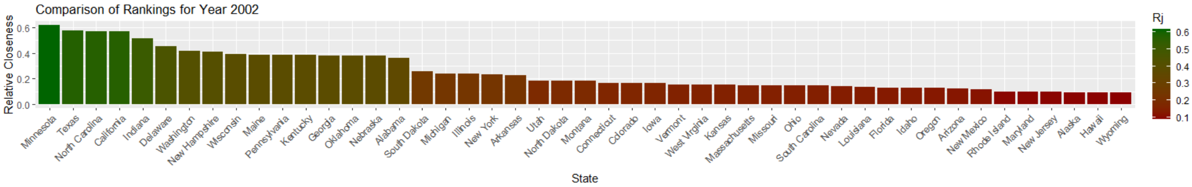
Incidence				
CANCERS	CRUDE RATE	NORMALIZED CRUDE RATE	AGE_ADJ_RATE	NORMALIZED_ADJ_RATE
Prostate	147.3	0.316706085	153	0.339020607
Female Breast	138.6	0.29800043	125.7	0.278528695
Lung and Bronchus	70.2	0.150935283	67.5	0.149567915
Colon and Rectum	50.1	0.10771877	48.2	0.10680257
Melanoma of the Skin	19.4	0.04171146	18.8	0.041657434
Kidney and Renal Pelvis	15.7	0.033756181	15.1	0.033458897
Leukemias	14.2	0.030531069	13.8	0.030578329
Liver and Intrahepatic Bile Duct	6.8	0.014620512	6.4	0.014181254
Hodgkin Lymphoma	2.8	0.006020211	2.8	0.006204299

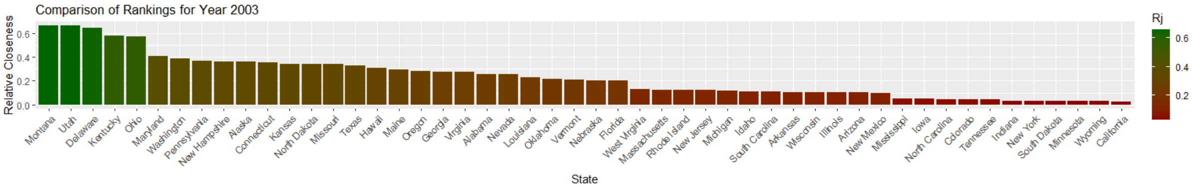
Table 2 and Table 3: The overall Rates of Each Cancer Sites

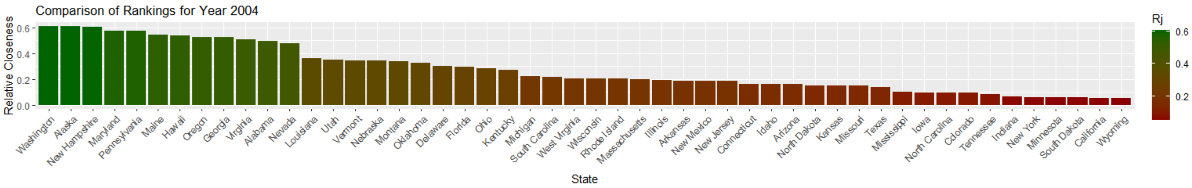


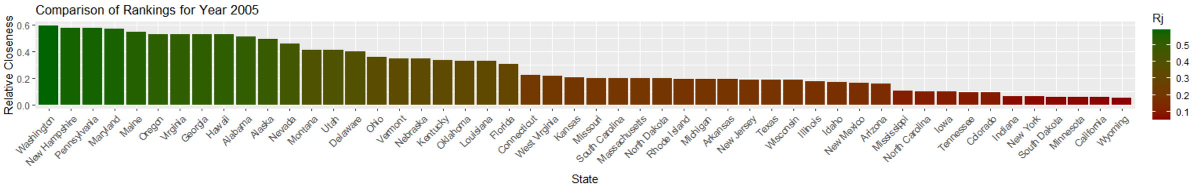


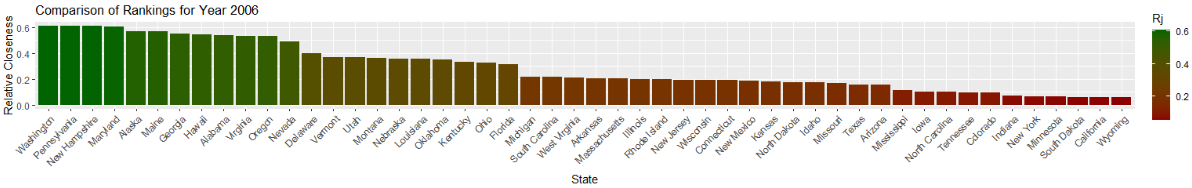


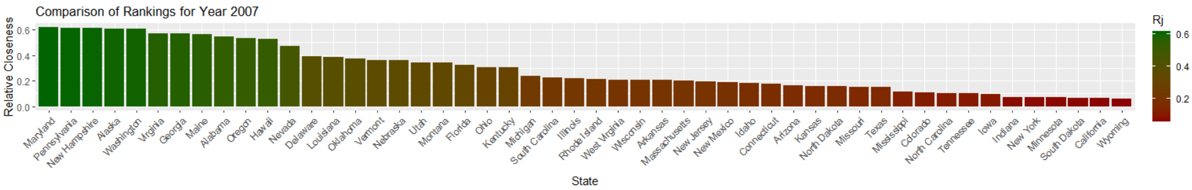

Comparison of Rankings for Year 2002
Relative Closeness
0.0 0.2 0.4 0.6
Minnesota Texas North Carolina California Indiana Delaware Washington New Hampshire Wisconsin Maine Pennsylvania Kentucky Georgia Oklahoma Nebraska Alabama South Dakota Michigan Illinois New York Arkansas Utah North Dakota Montana Connecticut Colorado Iowa Vermont West Virginia Kansas Massachusetts Missouri Ohio South Carolina Nevada Louisiana Florida Idaho Oregon Arizona New Mexico Rhode Island Maryland New Jersey Alaska Hawaii Wyoming
State
Rj
0.6 0.5 0.4 0.3 0.2 0.1

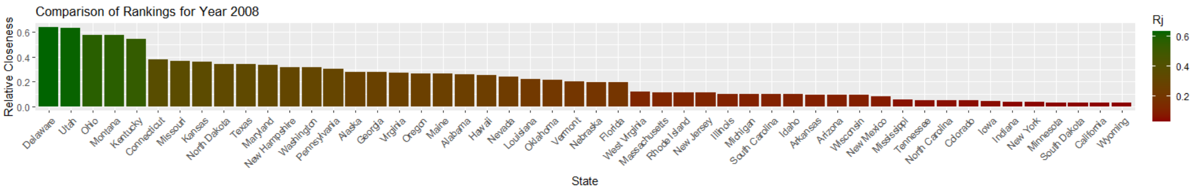

Comparison of Rankings for Year 2003
Relative Closeness
0.0 0.2 0.4 0.6
Montana Utah Delaware Kentucky Ohio Maryland Washington Pennsylvania New Hampshire Alaska Connecticut Kansas North Dakota Missouri Texas Hawaii Maine Oregon Georgia Virginia Alabama Nevada Louisiana Oklahoma Vermont Nebraska Florida West Virginia Massachusetts Rhode Island New Jersey Michigan Idaho South Carolina Arkansas Wisconsin Illinois Arizona New Mexico Mississippi Iowa North Carolina Colorado Tennessee Indiana New York South Dakota Minnesota Wyoming California
State
Rj
0.6 0.4 0.2

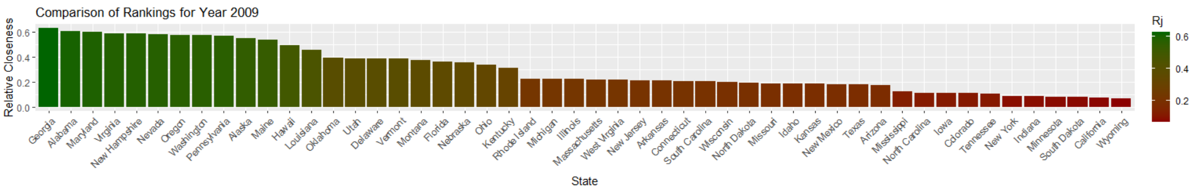

Comparison of Rankings for Year 2004
Relative Closeness
0.0 0.2 0.4 0.6
Washington Alaska New Hampshire Maryland Pennsylvania Maine Hawaii Oregon Georgia Virginia Alabama Nevada Louisiana Utah Vermont Nebraska Montana Oklahoma Delaware Florida Ohio Kentucky Michigan South Carolina West Virginia Wisconsin Rhode Island Massachusetts Illinois Arkansas New Mexico New Jersey Connecticut Idaho Arizona North Dakota Kansas Missouri Texas Mississippi Iowa North Carolina Colorado Tennessee Indiana New York Minnesota South Dakota California Wyoming
State
Rj
0.6 0.4 0.2

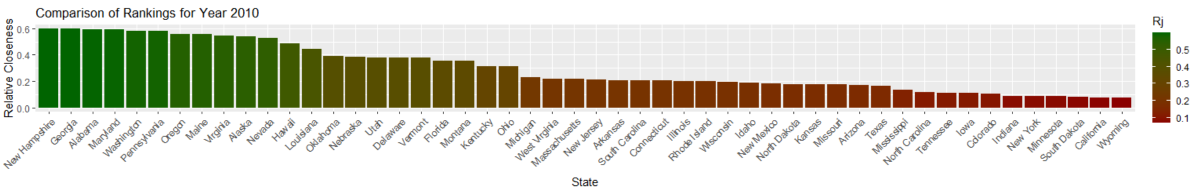

Comparison of Rankings for Year 2005
Relative Closeness
0.0 0.2 0.4 0.6
Washington New Hampshire Pennsylvania Maryland Maine Oregon Virginia Georgia Hawaii Alabama Alaska Nevada Montana Utah Delaware Ohio Vermont Nebraska Kentucky Oklahoma Louisiana Florida Connecticut West Virginia Kansas Missouri South Carolina Massachusetts North Dakota Rhode Island Michigan Arkansas New Jersey Texas Wisconsin Illinois Idaho New Mexico Arizona Mississippi North Carolina Iowa Tennessee Colorado Indiana New York South Dakota Minnesota California Wyoming
State
Rj
0.5 0.4 0.3 0.2 0.1

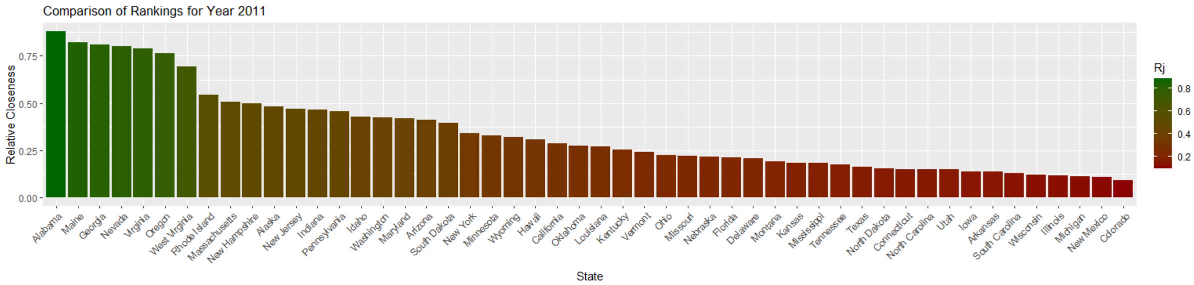

Comparison of Rankings for Year 2006
Relative Closeness
0.0 0.2 0.4 0.6
Washington Pennsylvania New Hampshire Maryland Alaska Maine Georgia Hawaii Alabama Virginia Oregon Nevada Delaware Vermont Utah Montana Nebraska Louisiana Oklahoma Kentucky Ohio Florida Michigan South Carolina West Virginia Arkansas Massachusetts Illinois Rhode Island New Jersey Wisconsin Connecticut New Mexico Kansas North Dakota Idaho Missouri Texas Arizona Mississippi Iowa North Carolina Tennessee Colorado Indiana New York Minnesota South Dakota California Wyoming
State
Rj
0.6 0.4 0.2

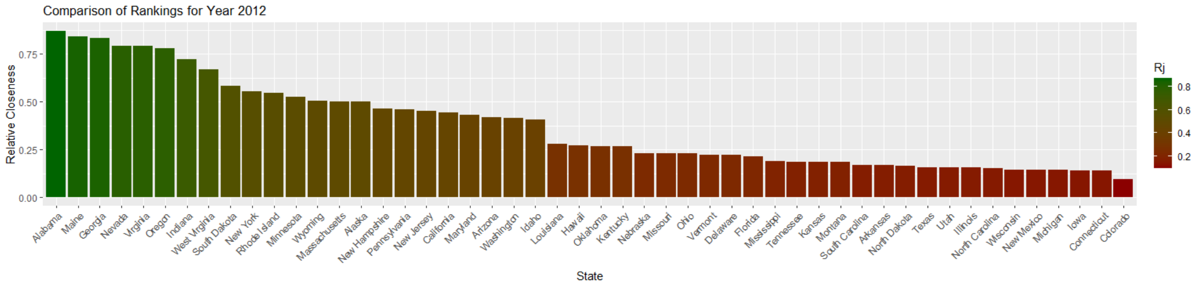

Comparison of Rankings for Year 2007
Relative Closeness
0.0 0.2 0.4 0.6
Maryland Pennsylvania New Hampshire Alaska Washington Virginia Georgia Maine Alabama Oregon Hawaii Nevada Delaware Louisiana Oklahoma Vermont Nebraska Utah Montana Florida Ohio Kentucky Michigan South Carolina Illinois Rhode Island West Virginia Wisconsin Arkansas Massachusetts New Jersey New Mexico Idaho Connecticut Arizona Kansas North Dakota Missouri Texas Mississippi Colorado North Carolina Tennessee Iowa Indiana New York Minnesota South Dakota California Wyoming
State
Rj
0.6 0.4 0.2


Comparison of Rankings for Year 2008
Relative Closeness
0.0 0.2 0.4 0.6
Delaware Utah Ohio Montana Kentucky Connecticut Missouri Kansas North Dakota Texas Maryland New Hampshire Washington Pennsylvania Alaska Georgia Virginia Oregon Maine Alabama Hawaii Nevada Louisiana Oklahoma Vermont Nebraska Florida West Virginia Massachusetts Rhode Island New Jersey Illinois Michigan South Carolina Idaho Arkansas Arizona Wisconsin New Mexico Mississippi Tennessee North Carolina Colorado Iowa Indiana New York Minnesota South Dakota California Wyoming
State
Rj
0.6 0.4 0.2

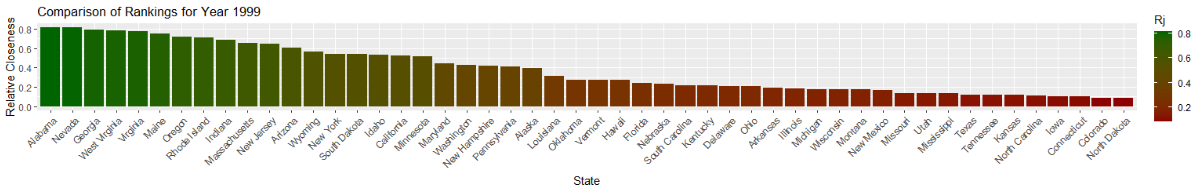

Comparison of Rankings for Year 2009
Relative Closeness
0.0 0.2 0.4 0.6
Georgia Alabama Maryland Virginia New Hampshire Nevada Oregon Washington Pennsylvania Alaska Maine Hawaii Louisiana Oklahoma Utah Delaware Vermont Montana Florida Nebraska Ohio Kentucky Rhode Island Michigan Illinois Massachusetts West Virginia New Jersey Arkansas Connecticut South Carolina Wisconsin North Dakota Missouri Idaho Kansas New Mexico Texas Arizona Mississippi North Carolina Iowa Colorado Tennessee New York Indiana Minnesota South Dakota California Wyoming
State
Rj
0.6 0.4 0.2

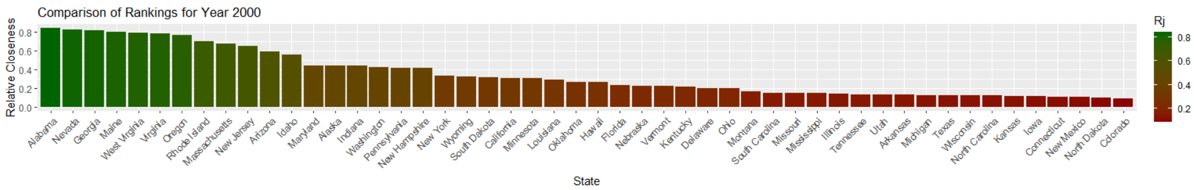

Comparison of Rankings for Year 2010
Relative Closeness
0.0 0.2 0.4 0.6
New Hampshire Georgia Alabama Maryland Washington Pennsylvania Oregon Maine Virginia Alaska Nevada Hawaii Louisiana Oklahoma Nebraska Utah Delaware Vermont Florida Montana Kentucky Ohio Michigan West Virginia Massachusetts New Jersey Arkansas South Carolina Connecticut Illinois Rhode Island Wisconsin Idaho New Mexico North Dakota Kansas Missouri Arizona Texas Mississippi North Carolina Tennessee Iowa Colorado Indiana New York Minnesota South Dakota California Wyoming
State
Rj
0.5 0.4 0.3 0.2 0.1

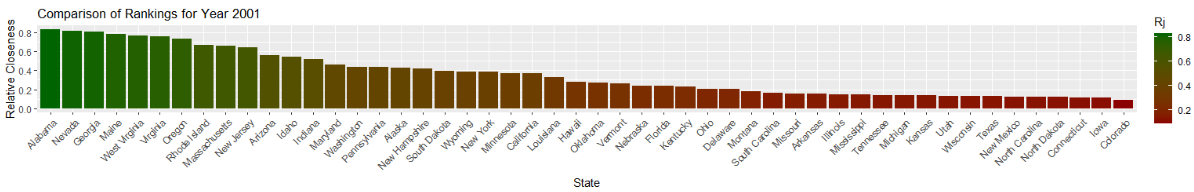


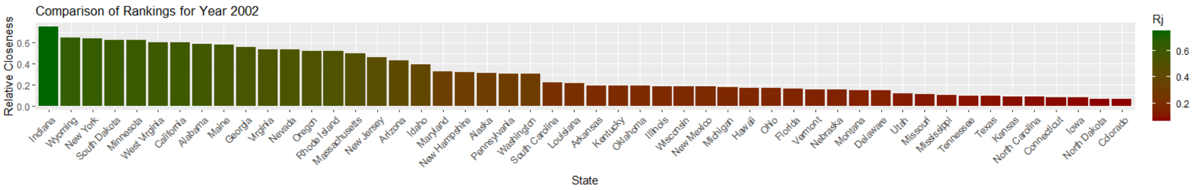


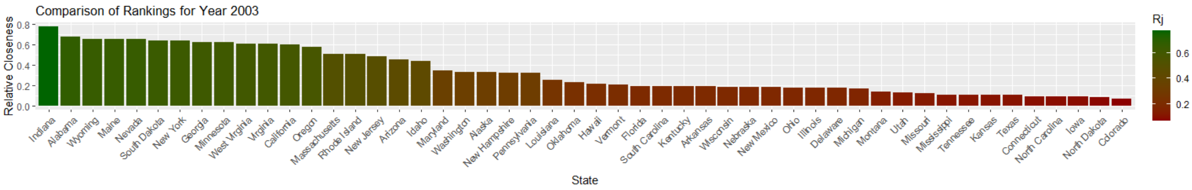
Plot 5: Incidence Age Adjusted Rate Ranking of the States Based on Years

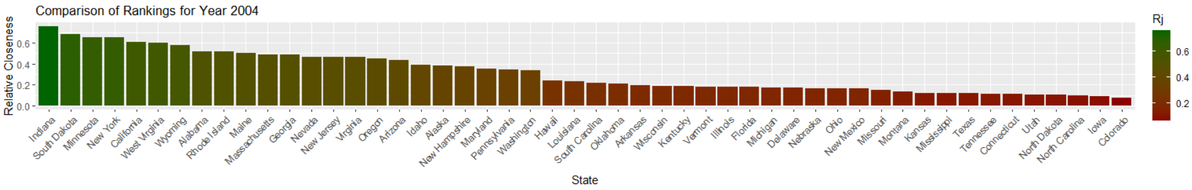


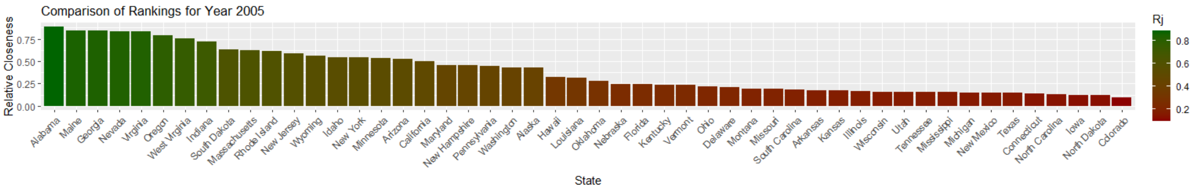


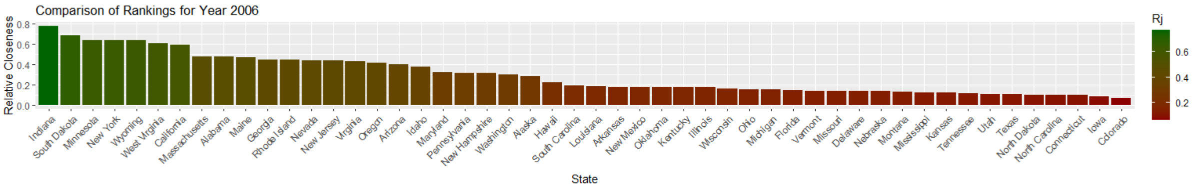


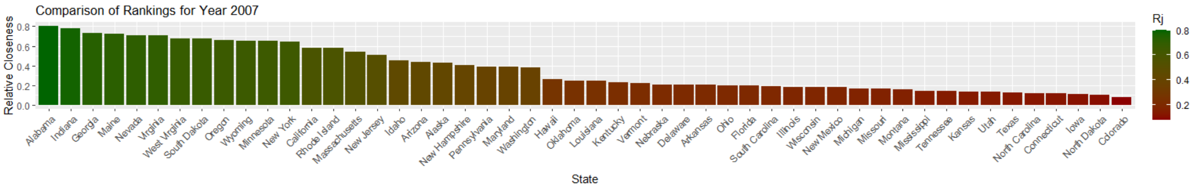

Comparison of Rankings for Year 2002
Relative Closeness
State
Rj

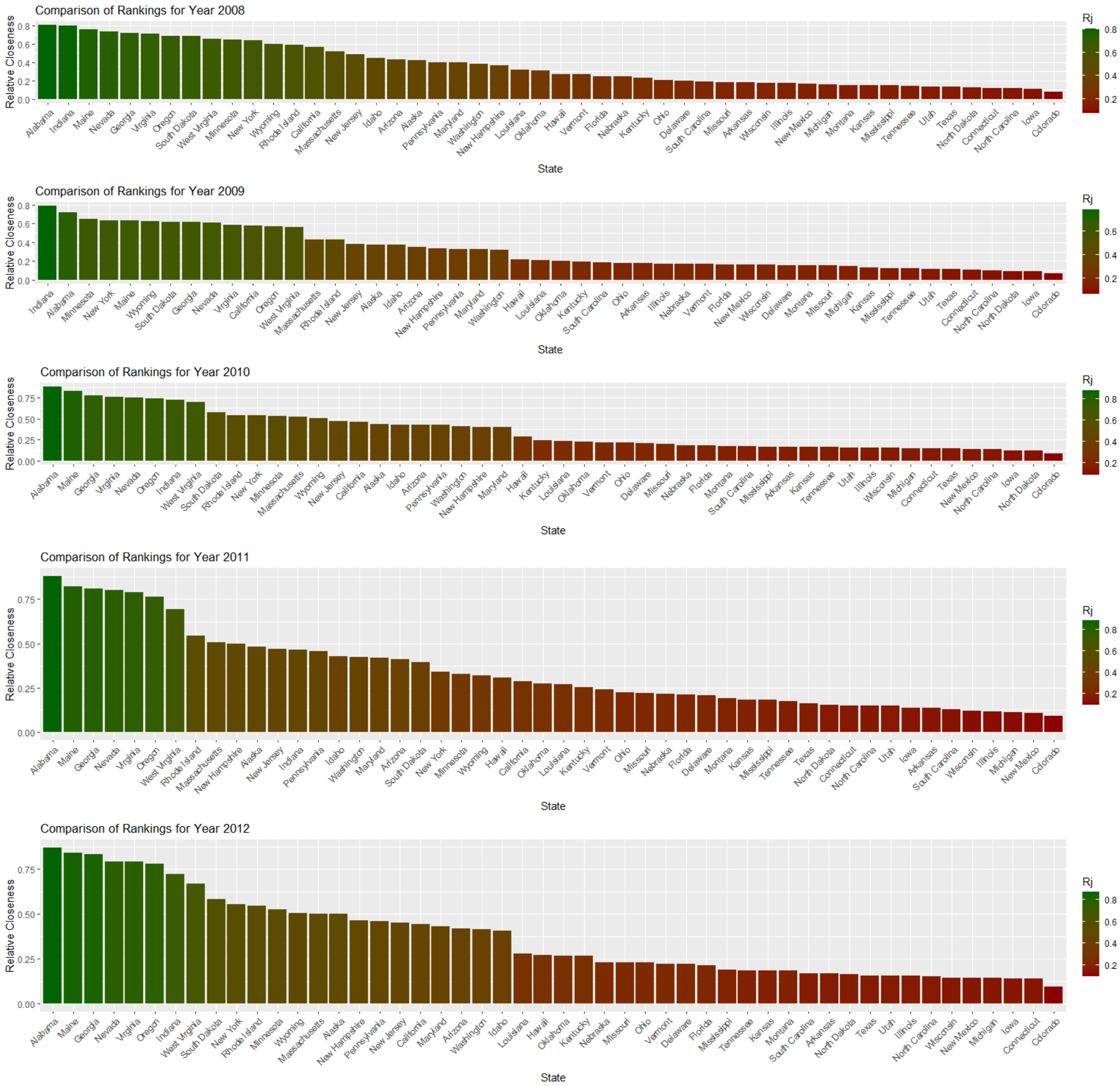

Comparison of Rankings for Year 2003
Relative Closeness
State
Rj


Comparison of Rankings for Year 2004
Relative Closeness
State
Rj


Comparison of Rankings for Year 2005
Relative Closeness
State
Rj


Comparison of Rankings for Year 2006
Relative Closeness
State
Rj


Comparison of Rankings for Year 2007
Relative Closeness
State
Rj



Plot 6: Mortality Age Adjusted Rate Ranking of the States Based on Years

Incidence Rank														
STATES	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Alabama	31	29	11	16	21	11	10	9	9	20	2	3	4	1
Alaska	38	37	42	45	10	2	11	5	4	15	10	10	18	14
Arizona	5	18	35	40	38	35	39	39	35	37	39	38	42	43
Arkansas	NA	NA	43	21	35	30	32	26	29	36	29	27	40	37
California	14	15	4	4	50	49	49	49	49	49	49	49	48	48
Colorado	15	16	16	26	43	43	44	44	41	43	43	44	46	44
Connecticut	8	1	18	25	11	33	23	32	34	6	30	29	23	25
Delaware	32	32	12	6	3	19	15	13	13	1	16	17	8	8
Florida	24	25	30	37	27	20	22	22	20	27	19	19	25	22
Georgia	45	45	10	13	19	9	8	7	7	16	1	2	2	2
Hawaii	29	30	44	46	16	7	9	8	11	21	12	12	21	10
Idaho	40	38	34	38	33	34	37	36	33	35	35	33	28	28
Illinois	4	9	37	19	37	29	36	28	25	32	25	30	33	34
Indiana	20	19	5	5	45	45	45	45	45	45	46	45	49	49
Iowa	18	17	22	27	41	41	42	41	44	44	42	43	39	40
Kansas	9	4	28	30	12	37	25	34	36	8	36	36	11	23
Kentucky	33	33	20	12	4	22	19	20	22	5	22	21	15	13
Louisiana	23	23	25	36	23	13	21	18	14	23	13	13	7	7
Maine	44	42	6	10	17	6	5	6	8	19	11	8	6	6
Maryland	27	28	46	43	6	4	4	4	1	11	3	4	14	12
Massachusetts	36	36	31	31	29	28	28	27	30	29	26	25	30	32
Michigan	1	6	36	18	32	23	31	23	23	33	24	23	35	35
Minnesota	12	8	1	1	48	47	48	47	47	47	47	47	44	46
Mississippi	NA	NA	NA	NA	40	40	40	40	40	40	40	40	38	36
Missouri	19	20	24	32	14	38	26	37	38	7	34	37	27	27
Montana	7	2	17	24	1	17	13	16	19	4	18	20	10	18
Nebraska	30	31	19	15	26	16	18	17	17	26	20	15	20	17
Nevada	25	24	27	35	22	12	12	12	12	22	6	11	NA	NA

New Hampshire	43	43	7	8	9	3	2	3	3	12	5	1	1	4
New Jersey	26	26	41	44	31	32	33	30	31	31	28	26	12	11
New Mexico	39	40	39	41	39	31	38	33	32	39	37	34	36	41
New York	2	7	40	20	46	46	46	46	46	46	45	46	32	33
North Carolina	17	14	2	3	42	42	41	42	42	42	41	41	45	47
North Dakota	11	11	13	23	13	36	29	35	37	9	33	35	41	39
Ohio	10	5	29	33	5	21	16	21	21	3	21	22	26	26
Oklahoma	34	35	23	14	24	18	20	19	15	24	14	14	16	19
Oregon	21	21	32	39	18	8	6	11	10	18	7	7	19	21
Pennsylvania	42	44	9	11	8	5	3	2	2	14	9	6	3	5
Rhode Island	28	27	45	42	30	27	30	29	26	30	23	31	17	9
South Carolina	37	39	33	34	34	24	27	24	24	34	31	28	29	29
South Dakota	NA	NA	38	17	47	48	47	48	48	48	48	48	34	30
Tennessee	NA	NA	NA	NA	44	44	43	43	43	41	44	42	47	45
Texas	3	10	3	2	15	39	34	38	39	10	38	39	43	42
Utah	13	13	15	22	2	14	14	15	18	2	15	16	9	24
Vermont	16	12	26	28	25	15	17	14	16	25	17	18	13	16
Virginia	NA	NA	NA	NA	20	10	7	10	6	17	4	9	22	20
Washington	6	3	14	7	7	1	1	1	5	13	8	5	5	3
West Virginia	35	34	21	29	28	25	24	25	27	28	27	24	24	15
Wisconsin	22	22	8	9	36	26	35	31	28	38	32	32	31	31
Wyoming	41	41	47	47	49	50	50	50	50	50	50	50	37	38

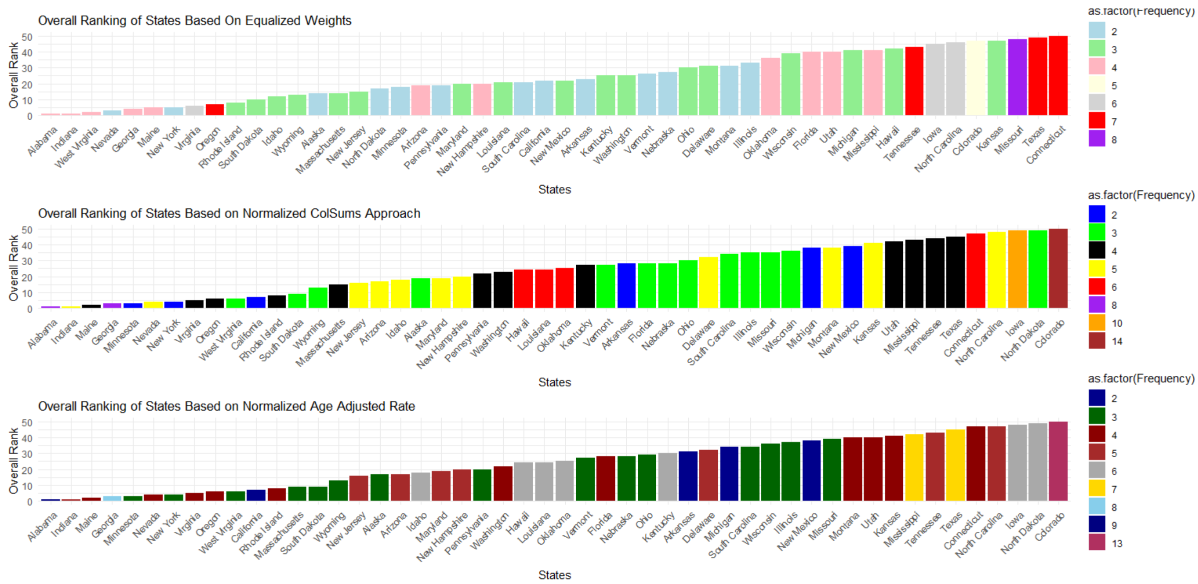
Table 5: Incidence Rate Ranking

Mortality Rank														
STATES	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Alabama	**1**	**1**	**1**	**8**	**2**	**8**	**1**	**9**	**1**	**1**	**2**	**1**	**1**	**1**
Alaska	**23**	**14**	**17**	**21**	**21**	**19**	**23**	**23**	**19**	**19**	**17**	**17**	**11**	**15**
Arizona	**12**	**11**	**11**	**17**	**17**	**17**	**17**	**17**	**18**	**18**	**19**	**19**	**18**	**21**

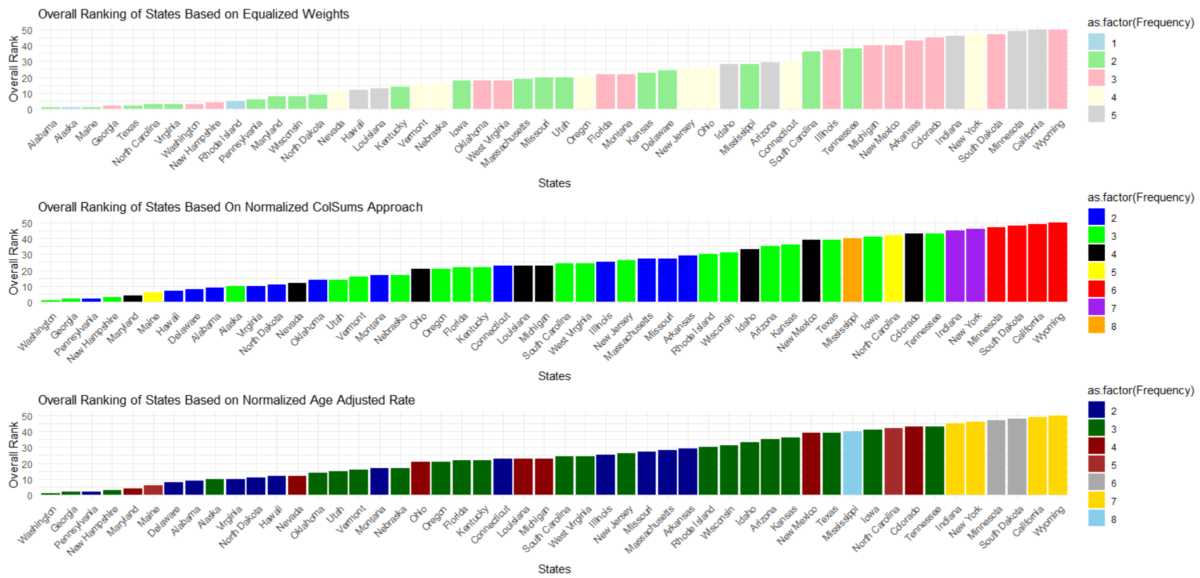
Arkansas	34	40	36	26	31	28	36	27	31	35	30	37	44	39
California	17	22	23	7	12	5	18	7	13	14	11	16	24	19
Colorado	49	50	50	50	50	50	50	50	50	50	50	50	50	50
Connecticut	48	47	48	47	46	45	46	48	47	47	46	44	40	49
Delaware	32	31	32	39	37	35	32	38	30	32	37	30	33	32
Florida	28	27	29	35	28	33	28	35	33	28	34	33	32	33
Georgia	3	3	3	10	8	12	3	11	3	5	8	3	3	3
Hawaii	27	26	25	33	26	24	24	24	24	26	24	24	23	25
Idaho	16	12	12	18	18	18	14	18	17	17	18	18	15	23
Illinois	35	37	37	29	36	32	38	31	35	37	31	41	47	43
Indiana	9	15	13	1	1	1	8	1	2	2	1	7	13	7
Iowa	47	46	49	48	48	49	48	49	48	49	49	48	43	48
Kansas	45	45	41	45	44	41	37	42	43	41	41	38	35	36
Kentucky	31	30	30	27	30	30	29	30	27	30	27	25	27	27
Louisiana	24	24	24	25	24	25	25	26	26	24	25	26	26	24
Maine	6	4	4	9	4	10	2	10	4	3	5	2	2	2
Maryland	19	13	14	19	19	21	19	19	22	21	22	23	17	20
xMassachusetts	10	9	9	15	14	11	10	8	15	15	14	13	9	14
Michigan	36	41	40	32	38	34	43	34	38	39	40	43	48	47
Minnesota	18	23	22	5	9	3	16	3	11	10	3	12	21	12
Mississippi	42	36	38	42	42	42	42	41	41	42	42	36	36	34
Missouri	40	35	35	41	41	39	34	37	39	34	39	31	30	29
Montana	38	33	33	38	39	40	33	40	40	40	38	34	34	37
Nebraska	29	28	28	37	33	36	27	39	29	29	32	32	31	28
Nevada	2	2	2	12	5	13	4	13	5	4	9	5	4	4
New Hampshire	21	18	18	20	22	20	20	21	20	23	20	22	10	16
New Jersey	11	10	10	16	16	14	12	14	16	16	16	15	12	18
New Mexico	39	48	45	31	34	38	44	28	37	38	35	46	49	46
New York	14	19	21	3	7	4	15	4	12	11	4	11	20	10
North Carolina	46	44	46	46	47	48	47	47	46	48	47	47	41	44
North Dakota	50	49	47	49	49	47	49	46	49	46	48	49	39	40
Ohio	33	32	31	34	35	37	31	33	32	31	29	29	29	30
Oklahoma	25	25	26	28	25	27	26	29	25	25	26	27	25	26
Oregon	7	7	7	13	13	16	6	16	9	7	12	6	6	6
Pennsylvania	22	17	16	22	23	22	21	20	21	20	21	20	14	17

Rhode Island	8	8	8	14	15	9	11	12	14	13	15	10	8	11
South Carolina	30	34	34	24	29	26	35	25	34	33	28	35	45	38
South Dakota	15	21	19	4	6	2	9	2	8	8	7	9	19	9
Tennessee	44	38	39	43	43	44	41	43	42	43	43	39	37	35
Texas	43	42	44	44	45	43	45	45	45	45	45	45	38	41
Utah	41	39	42	40	40	46	40	44	44	44	44	40	42	42
Vermont	26	29	27	36	27	31	30	36	28	27	33	28	28	31
Virginia	5	6	6	11	11	15	5	15	6	6	10	4	5	5
Washington	20	16	15	23	20	23	22	22	23	22	23	21	16	22
West Virginia	4	5	5	6	10	6	7	6	7	9	13	8	7	8
Wisconsin	37	43	43	30	32	29	39	32	36	36	36	42	46	45
Wyoming	13	20	20	2	3	7	13	5	10	12	6	14	22	13

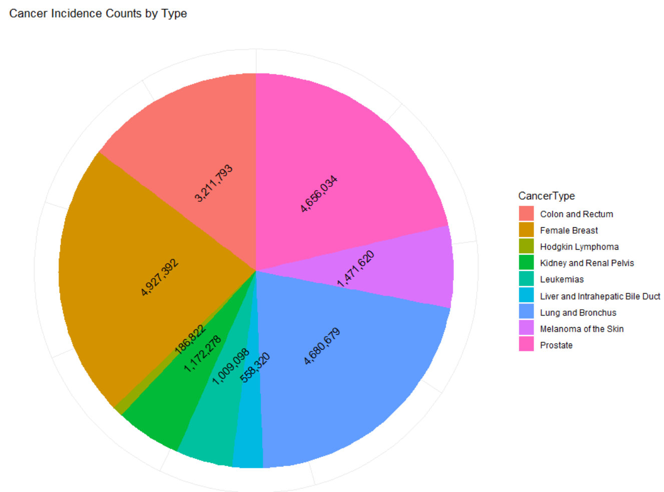
Table 6: Mortality Rate Ranking



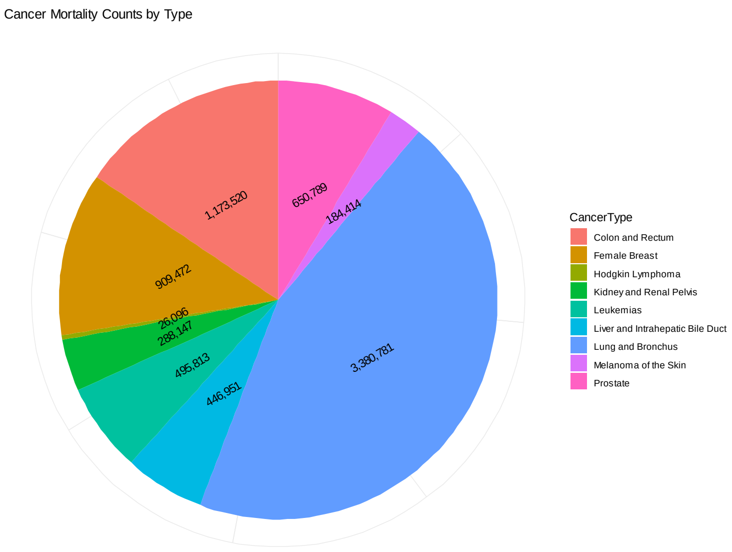
Plot 7: Incidence rate ranking based on three weighing methods



Plot 8: Mortality rate ranking based on three weighing methods



Plot 9: Incidence Cancer Rate Proportion



Plot 10: Mortality Cancer Rate Proportion

References

1. CDC Cancer Statistics - USCS Trends:
   - Link: [CDC USCS Trends](https://gis.cdc.gov/Cancer/USCS/#/Trends/)
2. PubMed - Article on Cancer:
   - Title: "CDC Activities to Enhance Training in Cancer Prevention and Control in Field Epidemiology Training Programs in Low- and Middle-Income Countries"
   - Link: [PubMed Article](https://pubmed.ncbi.nlm.nih.gov/30084697/)
3. PubMed - Another Article on Cancer:
   - Title: "Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State"
   - Link: [PubMed Article](https://pubmed.ncbi.nlm.nih.gov/25825511/)
4. SEER Cancer Statistics:
   - Link: [SEER Cancer Statistics](https://seer.cancer.gov/about/)
5. CDC Wonder - Cancer Statistics:
   - Link: [CDC Wonder - Cancer Statistics](https://wonder.cdc.gov/cancer.html)
6. Purdue University - Canine Cancer Database:
   - Title: "Purdue Scientists Join in Launch of Cloud-based Canine Cancer Database to Benefit Humans and Their Best Friendse"
   - Link: [Purdue University Article](https://vet.purdue.edu/news/pvr-a20-purdue-scientists-join-in-launch-of-cloud-based-canine-cancer-database-to-benefit-humans-and-their-best-friends.php)