A Visual Analysis of recent Cancer Incidences in the United States, including Causal Factor Research Trends

Debra Abrams, Adam J. Johs, and Matthew Lowery

Abstract—Cancer is one of the leading causes of death in the United States. Studying trends in the incidence of cancer may provide insights into the cause(s) of the disease. This report utilizes a set of information visualization tools, specifically, a country map, treemap, and two co-citation networks, to analyze both United States cancer incidence rates and causal factor research trends. An examination of the country map indicated that of the 51 geographic areas in the United States, Connecticut has experienced some of the highest overall cancer incidence rates. The treemap demonstrated that colon and rectum cancer, or colorectal cancer, has had the highest incidence rate of all the cancer types included in this analysis, predominately in the United States' Black and American Indian or Alaska Native populations. The first co-citation network, comprised of 521 nodes, indicated that colorectal cancer is also one of the chief areas of United States causal factor research. The same network suggested that mitochondrial disease may be an emerging area of cancer causal factor research. The second (487 node) co-citation network showed that much of the research being conducted on the leading causes of colorectal cancer in the United States has centered on the use of colonoscopies as a method to prevent colorectal cancer.

Index Terms—Cancer, colorectal, visualization, country map, treemap, co-citation, network

1 INTRODUCTION

According to the World Health Organization (WHO), at least one-third of cancer cases are preventable [1]. The implications of successful intervention are pervasive, ranging from increased personal quality and longevity of life to national financial relief. For example, it is predicted that approximately 580,350 individuals in the United States alone will die from cancer, making it the second most common cause of death [2]. Furthermore, the American Cancer Society reported that in 2008, the cost associated with cancer was $201.5 billion [2], a number expected to increase in future years. Identifying the factors associated with cancer development is essential in reducing the incidence and devastating effects of this disease.

One method of identifying the etiology of cancer is to review the data for the incidence of the disease. Regions where incidence is high or increasing may be subject to malevolent environmental or behavioral factors, while the opposite scenarios may be present in regions with low or decreasing rates of cancer incidence. Such trends can suggest subsequent modes of study. For example, Vieira, Webster, Weinberg, and Aschengrau studied the incidence of breast cancer in upper Cape Cod, Massachusetts using spatial-temporal data. Their results identified a region near the Massachusetts Military Reservation with a statistically significant increased risk for developing breast cancer during a specific time frame [3]. Such results provide a starting-point for a more in-depth analysis of causative factors, which may then have broader significance.

There are many methods in addition to spatial-temporal studies in which to analyze trends in cancer. For example, one way to obtain a global view of past and current progress in cancer research is with co-citation analysis. Cluster analysis of topics in literature can identify not only studies that have already been conducted, but also possibly reveal previously undiscovered relationships among potential causative factors or prevention strategies. In this paper, visualizations of the incidences of all types of cancer in the United States over a ten-year period were generated. As a complimentary analysis, a co-citation analysis of the literature regarding potential causes of cancer was performed.

2 TOOLS

Two different IBM Many Eyes visualization tools were utilized to analyze differing data types, namely the tools (1) *Country Map* and (2) *Treemap* [4]. The country map provides a geographical view of the data in question, while the treemap shows the relationships among hierarchal data sets. The power of the two tools is that they are interactive, thus allowing for more advanced analysis.

In addition to Many Eyes, two other tools were utilized to facilitate this analysis: Thomson Reuters' Web of Science® [5] and Chen's CiteSpace application [6]. The Web of Science® is a unique database portal/exploration and retrieval tool that

- *Debra Adams is with Drexel University. E-mail: dja64@drexel.edu.*
- *Adam J. Johs is with Drexel University. E-mail: ajj37@drexel.edu; adam.johs@yahoo.com.*
- *Matthew Lowery is with Drexel University. E-mail: ml958@drexel.edu.*

enables users to access scholarly works contained in thousands of the foremost academic journals. The ability to locate and extract bibliographic data from the Web of Science®, along with its overall compatibility with CiteSpace, resulted in its use as the sole collection source for the bibliographic data visualized in this report. CiteSpace was chosen as the citation visualization tool because the authors have learned from previous efforts how effective it can be at yielding novel insights into the research patterns of scientific fields. As Chen, CiteSpace's creator explains, "CiteSpace is a freely available Java application for visualizing and analyzing trends and patterns in scientific literature [that] focuses on finding critical points in the development of a field or a domain, especially intellectual turning points and pivotal points" [7].

3 METHODS

3.1 Many Eyes

The analytical tools available from IBM's Many Eyes website were selected for the creation of two visualizations on the United States cancer incidence rates by cancer type across the demographic category of race and ethnicity. To this end, the selection of visualization types was based upon the taxonomy of the data and the ability to show the relationships among the data elements from multiple points of view. For this analysis, the taxonomy type for the data is considered to be multi-dimensional. The general methodology employed for each of the Many Eyes visualizations is as follows:

1. Download of data from data source
2. Reconfiguration of source data for creation of a data set
3. Selection and creation of a particular visualization type
4. Interaction and adjustment with the visualization to create an instance that offers meaningful insight

The creation of a data set and two visualizations is described in specific detail in the following subsections.

3.1.1 Many Eyes: The Data Set – 'Project D Data Set'

Shneiderman, a pioneer in the information visualization field, introduced a type by task taxonomy (TTT) for which creators of information visualization systems can reference when determining the appropriate taxonomy to apply to a given case study [8]. The seven types of data discussed in the TTT are: 1-dimensional, 2-dimensional, 3-dimensional, temporal, multi-dimensional, tree, and network [8]. Once the data type is classified into a unit of measure, a system that quantifies relevant data can be constructed, which in turn provides the basis for an interactive graphical image system that facilitates the execution of several tasks, namely "overview first, zoom and filter, then details-on-demand" [8].

A single data set, entitled 'Project D Data Set,' was created from the source data to support and facilitate the two Many Eyes visualizations. The data elements were then configured into a hierarchal structure as, from top to bottom: geographic area or state, cancer type or physical site of the cancer, and race and ethnicity.

Three distinct units of measure associated with the above multi-dimensional data type were aligned within the hierarchal structure. The specific units of measure were: state population, state incidence rate (per 100,000 age-adjusted to the 2000 U.S. standard population), and incidence rate by cancer type and demographic of race and ethnicity (per 100,000 age-adjusted to the 2000 U.S. standard population).

State population is the total population for each state of the United States for the year 2009 and includes Washington the District of Columbia (D.C.) for a total of 51 data elements [9]. Note that in some instances, Many Eyes provides the value of the state population by the thousand (K). State incidence rate is the rate for all cancer types and demographics in that state, per 100,000 age-adjusted to the 2000 U.S. standard population. Note that the state incidence rate includes other cancer types and demographic categories besides those selected for this case study [9]. Incidence rate is the specific incidence rate per 100,000 age-adjusted to the 2000 U.S. standard population for each state based upon cancer type and the demographic of race and ethnicity [9].

The CDC data pertaining to incidence rates have three notable caveats. First, some rates are suppressed if fewer than 16 cases were reported in the specific category (geographic area, race and ethnicity). Second, some rates are suppressed at the state's or metropolitan area's request. Third, Hispanic origin is not mutually exclusive from race categories (White, Black, etc.).

3.1.2 Many Eyes: Visualizations

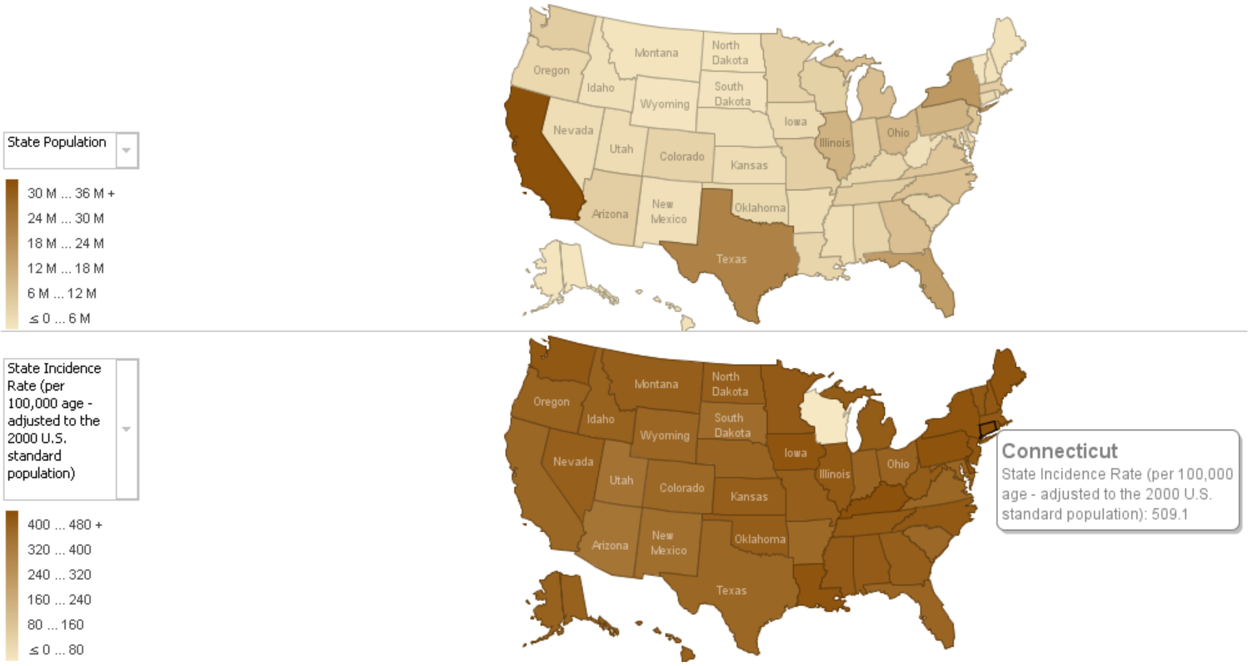
According to Shneiderman, some of the essential tasks that an information visualization should be capable of performing are overview (first), zoom, filter, details-on-demand, relate, history, and extract [8]. In order to achieve those capabilities for this analysis, two different visualization types were selected: a country map for overview first, and a treemap for zoom & filter, details-on-demand, and relation. Each visualization type is explained in further detail in the following subsections.

3.1.2.1 Overview Visualization – 'Project D Country Map USA'

The first Many Eyes visualization created, 'Project D Country Map USA,' was chosen in order to present a high-level overview of geographical cancer incidence rates in the United States. This viewpoint also provided a broad starting point to explore the subject area for further detail and analysis.

The country map visualization is composed of two maps that display the continental United States along with Alaska and Hawaii, allowing for a juxtaposed comparison of state population with the state incidence rate. This visualization includes zoom capabilities and details-on-demand and permits the data to be explored in further detail with the supporting visualization, 'Project D Treemap.'

Figure 1 shows a static image of 'Project D Country Map USA.' The image displays the zoom and details-on-demand functionality as depicted by the pop-up window containing the value of 509.1 for the state incidence rate for all cancer types and demographics for the state of Connecticut, the highest state incidence rate of all 51 geographic areas.



Connecticut
State Incidence Rate (per 100,000 age - adjusted to the 2000 U.S. standard population): 509.1

Fig. 1. Project D Country Map USA.

3.1.2.2 Zoom & Filter, Details-on-Demand, and Relation Visualization – 'Project D Treemap'

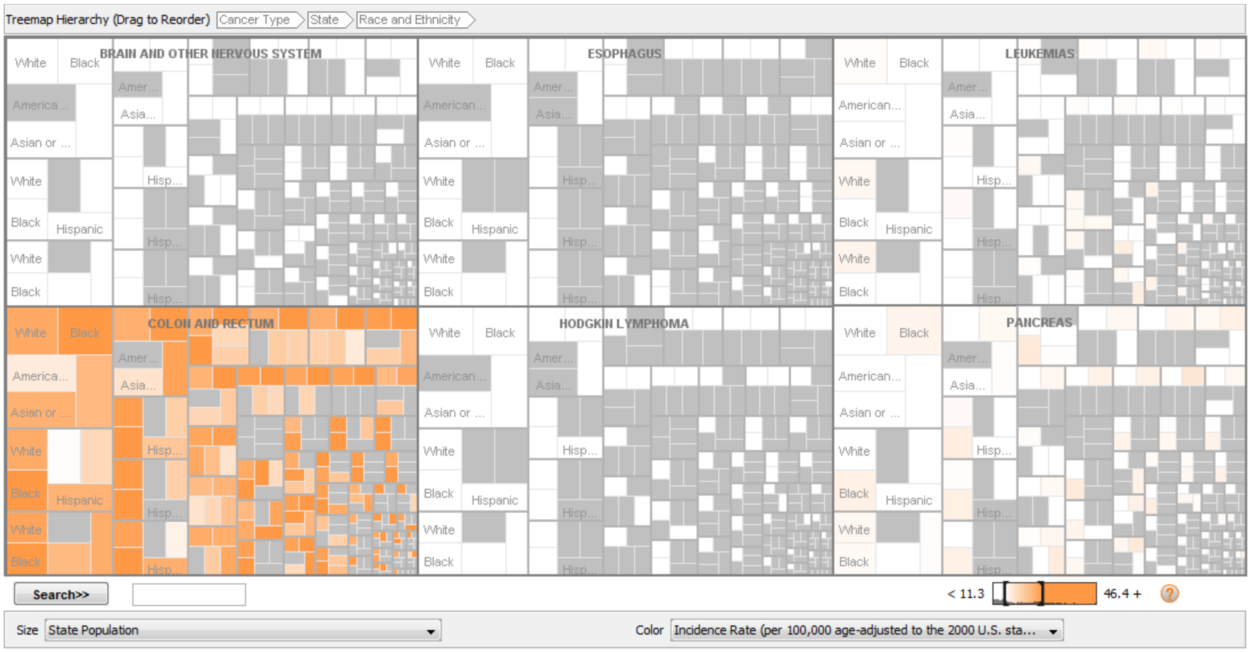
The second Many Eyes visualization was created in support of the first, allowing the overview presented in the 'Project D Country Map USA' to be drilled down in further detail. A treemap, titled 'Project D Treemap,' was created to provide multiple perspectives on the relationships among cancer type, geography, and race and ethnicity.

The analytical capabilities of the treemap lent itself to interactive user analyses of the data. A hierarchal structure was laid out to view the relationships between the data elements. One approach was to view the incidence rates by Cancer Type, State, and Race and Ethnicity. The size of the area of the state depicted in the visualization is proportional to the state population.

Two different static images of the 'Project D Treemap' are provided below for reference. Figure 2 displays the hierarchy by Cancer Type, State, and Race and Ethnicity. Figure 3 shows the same data with a differing hierarchal structure of Race and Ethnicity, Cancer Type, and State. Furthermore, the Race and Ethnicity Incidence Rate by Cancer Type is illustrated by the use of the intensity of color, in this case dark orange being relatively high and white being relatively low. Another note to the color scale is the use of gray to depict incidence rates that do not meet the minimum threshold or were not reported by the state in question.

To further provide visual discrimination among incidence rates, settings made to the color filter adjustment function of the treemap were employed. The range below 11.3 is indicated with opaque white, while the range above 46.4 is indicated by the deepest shade of orange. These set points were established based upon the statistical data provided by the CDC. The mean for those states reporting an incidence rate above the minimum threshold for Colon and Rectum Cancer Type was calculated to be 30.56, with a standard deviation based upon the population of 15.9. Therefore, one standard deviation upward from 30.56 is approximately 46.5 and as such 46.4 was chosen to be the bottom of the upper range. The settings for the color filtering adjustment are located in the lower right-hand portion of the visualization.

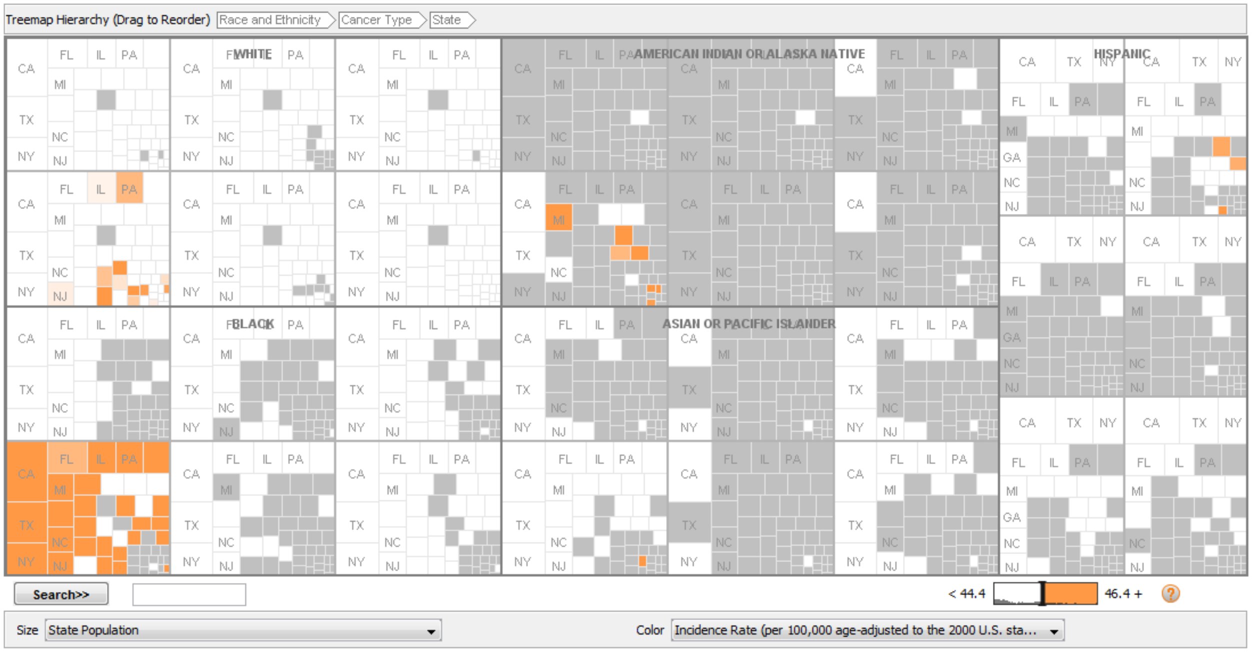
A different hierarchal structure is depicted in Figure 3, with the order now being Race and Ethnicity, Cancer Type, and State. Discriminating adjustments made to the color filter were set to < 44.4 and 46.4 + to accentuate significant outlying data greater than one standard deviation above the mean incidence rate for the Colon and Rectum Cancer Type by Race and Ethnicity.


Treemap Hierarchy (Drag to Reorder) Cancer Type > State > Race and Ethnicity

Size: State Population

Color: Incidence Rate (per 100,000 age-adjusted to the 2000 U.S. sta...

Fig. 2. Project D Treemap - Display of the relatively high Colon and Rectum Incidence Rates compared to other cancer types.


Treemap Hierarchy (Drag to Reorder) Race and Ethnicity > Cancer Type > State

Size: State Population

Color: Incidence Rate (per 100,000 age-adjusted to the 2000 U.S. sta...

Fig. 3. Project D Treemap - Display of the relatively high Black and American Indian or Alaska Native Incidence Rates of the Colon and Rectum cancer type compared to other races and ethnicities.

3.2 CiteSpace: Bibliographic Data Collection

To analyze the potential trends and patterns in the literature discussing the leading causes of cancer in the United States, a collection of bibliographic data was obtained from the Web of Science® for visualization in CiteSpace. The Web of Science® was searched for all of the articles that examined the leading causes of cancer in the United States. The initial search returned a total of 2,175 records. These results were filtered to only include ‘articles’ as the document type, which led to a total of 1,632 papers being returned. This ultimately led to the authors obtaining a data set comprised of 1,632 bibliographic records that spanned the years 1991-2013. The search configuration and filtering/refinement options that were used are detailed below:

> Topic=(leading cancer causes in United States) OR Topic=(leading cancer causes in U.S.) OR Topic=(leading cancer causes in US)
> Refined by: Document Types=(ARTICLE)
> Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI.

A preliminary examination of the information collected for this project (to be discussed in subsequent sections) indicated that additional, low-level data centering on colorectal cancer would need to be acquired to further facilitate the authors’ analysis. As a result, a second collection of bibliographic data was obtained from the Web of Science®. For this data set the authors searched for the entire collection of records that examined the leading causes of colorectal cancer in the United States. Because of the inclusive nature of the term ‘colorectal,’ various phrase combinations were included in the search configuration to ensure no relevant research went overlooked. The initial search returned a total of 1,019 records. Like the first bibliographic data set, these results were refined to only include ‘articles’ as the document type. This refinement led to a final data set of 817 records spanning the years 1991-2013 that was extracted for visualization in CiteSpace. Below is the search configuration that was utilized:

> Topic=(causes of colorectal cancer in United States) OR Topic=(causes of colon cancer in United States) OR Topic=(causes of rectal cancer in United States) OR Topic=(causes of rectum cancer in United States) OR Topic=(causes of colorectal cancer in U.S.) OR Topic=(causes of colon cancer in U.S.) OR Topic=(causes of rectal cancer in U.S.) OR Topic=(causes of rectum cancer in U.S.) OR Topic=(causes of colorectal cancer in US) OR Topic=(causes of colon cancer in US) OR Topic=(causes of rectal cancer in US) OR Topic=(causes of rectum cancer in US)
> Refined by: Document Types=(ARTICLE)
> Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI.

3.3 CiteSpace: Co-Citation Network Generation

Once the bibliographic data sets were retrieved and extracted from the Web of Science®, they were imported into CiteSpace for visualization. In particular, the visualization generated for the literature on the leading causes of cancer in the United States (the high-level co-citation research network of this analysis) was created with CiteSpace using the following settings:

1. Time Slicing: The time interval was configured to range from 1991-2013 to reflect the publication dates of the earliest and most recent articles in the data set. Consequently, the length of each individual time slice was set to two years. This resulted in a network comprised of 11 two-year time slices
2. Term Source: Titles, Abstracts, Author Keywords (DE), and Keywords Plus (ID) were the term source components selected for the network
3. Term Type: Not specified
4. Node Types: The type of node contained in the visualization was configured to be the cited references of the records contained in the data set
5. Links: The default settings for strength and scope of links were utilized: Cosine and Within Slices, respectively
6. Top N per slice: The top 50 most cited or occurred items from each slice were selected
7. Pruning: No pruning options used
8. Visualization: The visualization was set to display a static cluster view and merged network

The following configuration was utilized to generate the low-level co-citation research network from the records retrieved on the leading causes of colorectal cancer in the US:

1. Time Slicing: Like the high-level network visualization, the time interval was also configured to range from 1991-2013 to reflect the earliest and most recent records obtained in the data set. Accordingly, the length of each individual time slice was also set to two years, resulting in a network consisting of 11 two-year time slices
2. Term Source: Titles, Abstracts, Author Keywords (DE), and Keywords Plus (ID) were also selected as the term source components
3. Term Type: None specified
4. Node Types: To ensure that a progressive, accurate analysis could be accomplished the node type of the low-level network was also set to be the cited references of the data set articles
5. Links: The default settings for strength and scope of links were utilized: Cosine and Within Slices, respectively
6. Top N per slice: The top 50 most cited or occurred items from each slice of the network were also chosen
7. Pruning: No pruning options used
8. Visualization: The visualization settings were also set to display a static cluster view and merged network

Once each network was generated the authors displayed citation bursts, clustered the nodes, and labeled the clusters with title terms. To ensure the most interesting information was captured and conveyed, the authors labeled the clusters in the high-level network by log-likelihood ratio (LLR) and the clusters in the low-level network by tf*idf. Once these steps were completed the authors then analyzed each network from multiple viewpoints in an attempt to acquire insight into the trends, patterns, and possible directions of the research being conducted on the causal factors of U.S. cancer incidences, namely cases of colorectal cancer.

4 RESULTS

4.1 High-Level Co-Citation Network of the Leading Causes of U.S. Cancer Cases

A 521 node co-citation network visualization was generated for the high-level data set obtained for the leading causes of cancer in the United States. A total of 43 co-citation clusters were identified in the network. The largest cluster, ID #19 colorectal cancer, was comprised of 110 nodes. The research of Jemal et al. [10], the most highly cited article in the network, contained a total of 40 citation counts and was located in cluster ID #19. A timeline view depicting the high-level co-citation network is displayed below (Figure 4):

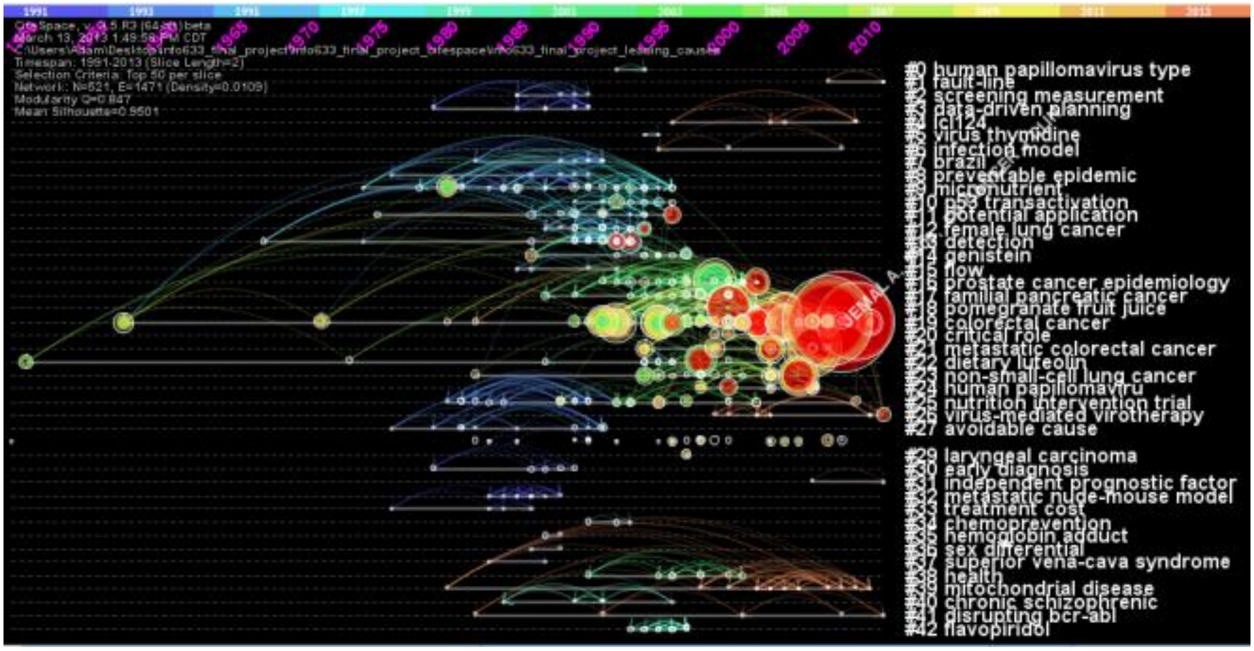


Fig. 4. The timeline view of a 521 node co-citation network visualization generated from the literature collected on the leading causes of cancer in the United States. Cluster labels are configured uniformly for optimal readability, citation bursts are depicted in red, and the node of the most highly cited article in the network, the work by Jemal et al. [10], is labeled.

4.2 Low-Level Co-Citation Network of the Leading Causes of U.S. Colorectal Cancer Cases

For the low-level colorectal dataset, a 487 node co-citation network visualization was generated. In this network a total of 8 co-citation clusters were identified. Cluster ID #7, rectal cancer, was the densest cluster in the network and included the majority of the network's overall nodes, 465 to be precise. The paper with the highest citation count, a study by Winawer et al. [11], was located in the rectal cancer cluster and cited a total of 37 times. A timeline view displaying the low-level network is provided below (Figure 5):

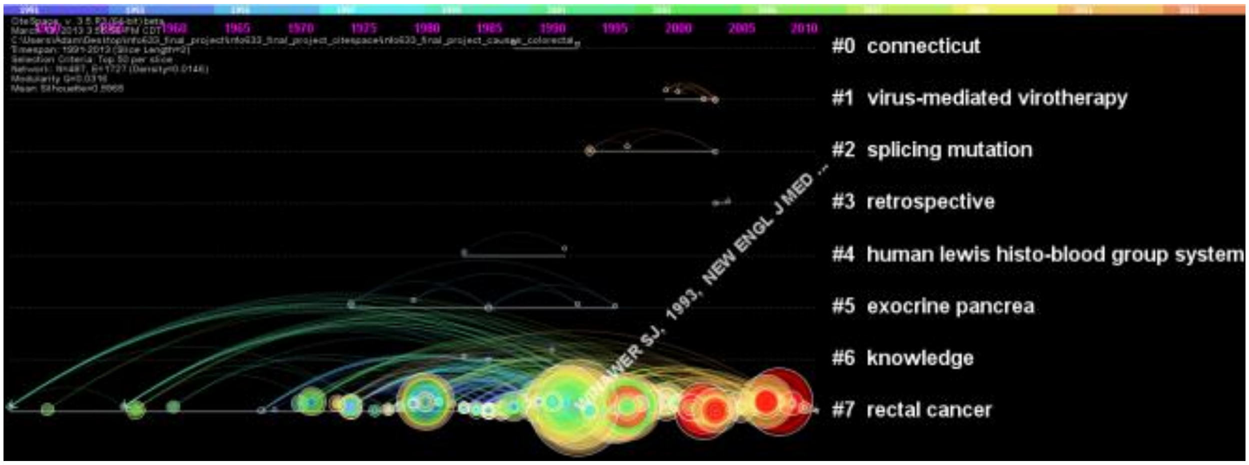


Fig. 5. The timeline view of a 487 node co-citation network visualization generated from the literature collected on the leading causes of colorectal cancer in the United States. Cluster labels are sized uniformly for optimum readability, citation bursts are portrayed in red, and the node of the article with the highest citation count in the network, the research of Winawer et al. [11], is labeled.

4.3 Combined Analysis of the Country and Treemap Visualizations

Using a combination of the 'Project D Country Map USA' visualization for overview and interactive analysis by way of the supporting 'Project D Treemap' visualization, an image of where the highest incidence rates have occurred emerged. Across the nation, colon and rectum cancer incidence rates were the highest amongst the six cancer types explored in this case study. Furthermore, the Black and American Indian or Alaska Native populations had the highest incidence rates of colon and rectum cancer.

5 DISCUSSION

The ultimate goal of information visualization is to facilitate user insight. Insight has been defined as "unexpected discoveries, a deepened understanding, a new way of thinking, eureka-like experiences, and other intellectual break-throughs" [12]. In the context of this analysis, the generated visualizations were designed with the hope of elucidating our understanding of cancer.

Cancer statistics have been reported exhaustively over the years, though only recently has the data collection become somewhat standardized. The data can be presented in myriad ways, including by gender, race, ethnicity, geographic location, disease location, and any permutation of these variables. The article by Jemal et al. on cancer statistics in 2009 is inundated with tables of numbers based on the variables listed above [10]. The data itself is comprehensive, but often too dense for the reader to immediately understand their significance. For example, Table 3 shows the death rates for all cancer types for a five-year period (2001-2005) as well as estimated death rates by state. This table consists of 13 columns and 52 rows of statistics, from which it is time consuming for the reader to try to make sense. Our approach in communicating similar data was to use a country map and a treemap.

An advantage of using a treemap for this type of data is that it allows for comparison of values among groups. For example, cancer type, race and ethnicity, and state can be visualized simultaneously, demonstrating that the incidence of colorectal cancer is highest in the Black population and in the American Indian or Alaska Native population. A detailed visual analysis indicated that for the top 20 highest incidence rates for the colon and rectum cancer type, 14 belonged to the Black population, 5 belonged to the American Indian or Alaska Native population, and 1 was attributed to the Hispanic ethnicity group. The highest incidence rate was found to be in the American Indian or Alaska Native group from the state of Alaska with a value of 96.0.

Vieira et al. also used geographic maps in the portrayal of their data. They initially showed a local map, without cancer data, of the area and its relation to the entire state. Subsequent maps showed different aspects of the data, but it was easy to understand their relevance and relation to each other when displayed on the geographic background [3].

Literature analyses can also be used to assess the state of knowledge in cancer research. The co-citation analyses of articles relevant to the leading causes of cancer yielded interesting results. An initial review of the treemap and the high-level network visualization demonstrated that colorectal cancer was a prominent type of cancer occurring in the United States and a chief area of casual factor research. While it makes intuitive sense that research efforts would focus on the cancers that affect the most individuals, the authors wanted to explore this topic in greater depth in an attempt to uncover further information about colorectal cancer.

To this end, a low-level analysis of the citations related to the leading causes of colorectal cancer in the United States was performed. While there is mention of colorectal cancer periodically before 1960, the number of articles relating to this type of cancer increased around 1990, and markedly rose higher in 2005 and after. In fact, in 1993, Winawer et al. reported in the *New England Journal of Medicine* that not only removing colonic polyps via colonoscopy resulted in a "lower-than-expected incidence of colorectal cancer," but also confirmed the pathophysiology of the progression of the disease [11]. The significance of these findings cannot be overstated as the results support the use of colonoscopies as a method to prevent colorectal cancer. The impact on public health of such a highly effective and minimally invasive procedure to potentially reduce the incidence and burden of colorectal cancer is substantial. Their findings also describe how colon cancer is a progressive disease, beginning as benign growths or polyps that have the potential to become cancerous over time. This multi-stage paradigm has served as a model for how cancer develops, suggesting why this paper might be highly cited in a search of cancer causes.

The prominent article in the latter timeframe is the paper by Jemal et al. from 2009. As mentioned previously, this article is a comprehensive report of cancer statistics, which logically would serve as a reference for subsequent research [10].

A relatively sizeable cluster that was not connected to other clusters in the high-level co-citation network was cluster ID #39, 'mitochondrial disease.' As mitochondrial disease typically affects cell energy production as opposed to regulating cell proliferation, it was not immediately clear why this topic would appear in this result set. A timeline analysis demonstrated that the first mention of mitochondrial disease in relation to cancer causation was in the early 1980s. It does not appear to be mentioned again until the mid to late 1990s, and then again in the 2000s, giving the impression that the present understanding of the role of mitochondria in cancer is either limited or being dismissed as an influential factor. In 2002, Carew and Huang published an article reviewing the state of knowledge of mitochondrial disease and cancer. They explain that mutations in mitochondrial DNA have been found in cancer cells and provide a biological mechanism rationale, but a proven role of such mutations in clinical disease is lacking [13]. This may in fact be an emerging topic and could offer a means for novel therapeutic intervention.

6 CONCLUSION

The type of visualization used to display data is essential to promoting insight. Colorectal cancer can be preventable when discovered at an early state, making screening strategies an essential part of public health. Two factors come to mind when considering prevention measures on a population scale. The first is education, as populations with high rates of colon cancer should be targets of public health education. The second is access, because at-risk individuals may not be seeking colonoscopies due to lack of health insurance or access to medical care. Information visualization can help identify these at-risk groups and track overall progress made in overcoming these barriers. Country maps and treemaps clearly illustrated which populations were at highest risk for colorectal cancer, as well as this risk compared to other cancers.

Co-citation analysis showed that most research in colorectal cancer causation has occurred beginning in the 1990s, with the most influential articles being those published in the last 15 years. Our understanding of cancer causes and incidence has increased greatly as a result of this analysis, suggesting that the visualizations presented here did in fact foster insight.

ACKNOWLEDGMENTS

The authors are grateful to IBM Research and the IBM Cognos software group, Thomson Reuters, and Dr. Chaomei Chen for their work on the Many Eyes visualization experiment, the Web of Science$^{®}$, and CiteSpace, respectively. For without their efforts this analysis could not have been accomplished.

REFERENCES

[1] World Health Organization. (n.d.). Cancer prevention. Retrieved from http://www.who.int/cancer/prevention/en/

[2] American Cancer Society. (2013). Economic impact of cancer. Retrieved from http://www.cancer.org/cancer/cancerbasics/economic-impact-of-cancer

[3] Vieira, V. M., Webster, T. F., Weinberg, J. M., & Aschengrau, A. (2008). Spatial-temporal analysis of breast cancer in upper Cape Cod, Massachusetts. *International Journal of Health Geographics, 7*(46). doi:10.1186/1476-072X-7-46

[4] IBM. (n.d.). Many eyes. Retrieved from http://www-958.ibm.com/software/data/cognos/manyeyes/

[5] Thomson Reuters. (2013). Web of Science$^{®}$. Retrieved from http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/

[6] Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359-377. doi:10.1002/asi.20317

[7] Chen, C. (2013). CiteSpace: Visualizing patterns and trends in scientific literature. Retrieved from http://cluster.cis.drexel.edu/~cchen/citespace/

[8] Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. Proceedings from *IEEE Symposium on Visual Languages*, 336-343. Los Alamitos, CA: IEEE Computer Society Press.

[9] Centers for Disease Control and Prevention. (2013). *United States cancer statistics: 2009 incidence and mortality*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Retrieved from http://www.cdc.gov/cancer/dcpc/data/index.htm

[10] Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., & Thun, M. J. (2009). Cancer statistics, 2009. *CA Cancer J Clin 2009, 59*, 225-249.

[11] Winawer, S. J., Zauber, A. G., Ho, M. N., O'Brien, M. J., Gottlieb, L. S., Sternberg, S. S., . . . National Polyp Study Workgroup. (1993). Prevention of colorectal cancer by colonoscopic polypectomy. *The New England Journal of Medicine, 329*(27), 1977-1981.

[12] Chen, C. (2010). Information visualization. *Wiley Interdisciplinary Review:*

Computational Statistics, *2*(4), 387-403. doi:10.1002/wics.89

[13] Carew, J. S., & Huang, P. (2002). Mitochondrial defects in cancer. *Molecular Cancer*, *1*(9).doi:10.1186/1476-4598-1-9