Exploring Increased Cancer Rates in Young Adults

PROCESS BOOK

John Ling Chen, Michael Dean Molenaar

December 6th, 2024

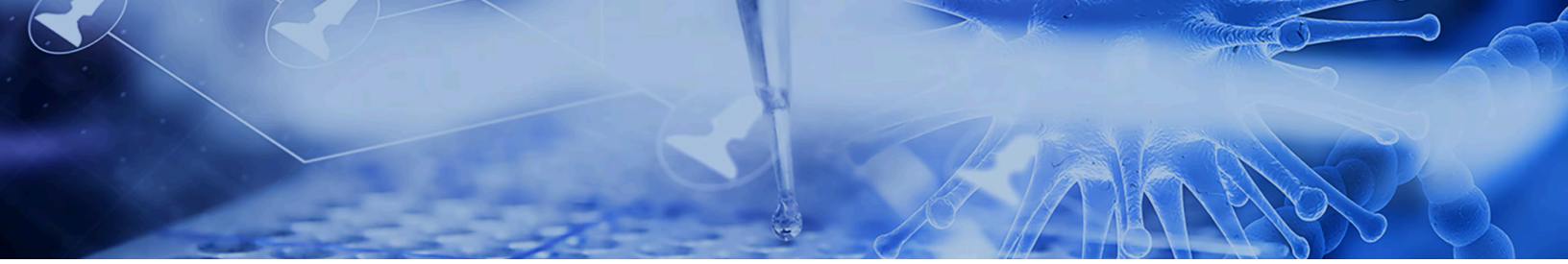
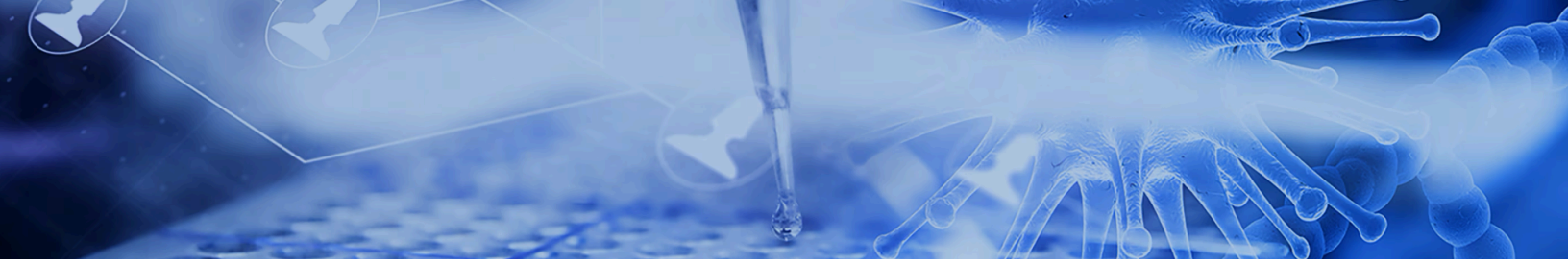


Table of Contents



Initial Meeting with TA regarding Project Proposal

On 9/19, we met with our TA regarding our proposal and her insights on improvements we can make toward the project. The largest recommendation she gave was to move away from using an anatomical visualization to one that is more focused on getting users to easily learn about different cancer rates. Her main fear was that using human anatomy would take away from the mission of our webpage and users may get lost exploring the anatomy rather than exploring cancers right away.

Because of this, we have decided to have our webpage start with specific cancers, which can immediately lead to data visuals of any changes in cancer incidence rates. This will take out the possible complication of the user having to explore human anatomy before finding cancers they may be interested in learning about more.

Regarding the specific data visualizations, she recommended we include nice interactions: showing growth in line charts, bar charts, or color matrixes. She also recommended adding in animation of dots showing any increase or decrease in the dots.

[entry by mmoneymike, 9/23]

John and I have drafted up some of the above recommendations, found in below sections:

Starting Cancers: **Design Evolution**

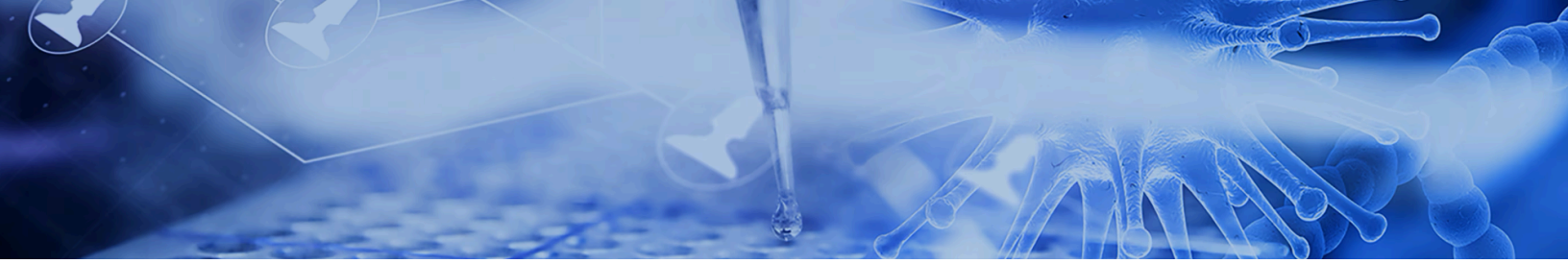
Data Visualization: **Exploratory Data Analysis**

[updated by mmoneymike, 10/12]

GIT Repository

Through GitHub Classroom, I have created our GIT repository, found here.

[entry by JCpennyChen, 9/13]



Data

> ***10/12*** *Today, John and I met to go over the first steps of the website. After creating our initial foundation for the main page, styling, js code, etc. we delegated two tasks: I am in charge of finding our main datasets, and John is in charge of creating our initial designs.*
>
> **[entry by mmoneymike, 10/12]**

Data Sourcing

Ultimately, we have chosen the **CDC Cancer Incidence 1999 - 2021** and **Cancer Mortality 1999 - 2021** datasets. These data have simple attributes and have already been cleaned up by the CDC. We also used the CDC filters for specific race, sex, and age subsets.

Data sources are found here:

> **Incidence:** https://wonder.cdc.gov/cancer-v2021.HTML
>
> **Mortality** https://wonder.cdc.gov/cancermort-v2021.HTML

[update by mmoneymike, 10/12]

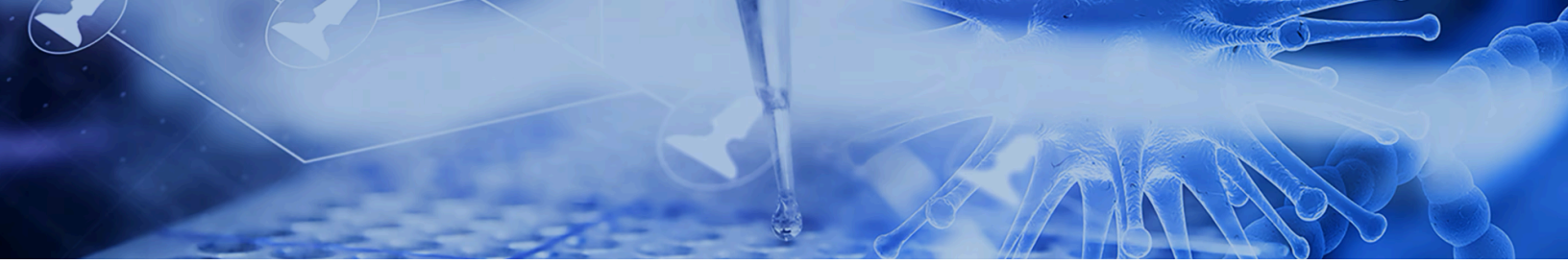
Data Conversion and Cleaning

File Conversion: Due to the CDC data sources download being .txt files, we have converted the files into more structured .csv files. To do so, we used PyCharm code to convert these.

Data Cleaning: This data from the CDC is incredibly concise and clean and already has all available counts for each leading cancer type. A few redundancies were removed, but other than that I believe we have our core datasets. Fortunately, this is a huge step for our graphical visualizations of cancer trends.

[entry by mmoneymike, 10/12]

I have added the Mortality datasets. Our final total is 10 cleaned datasets, which can be used to filter sex, race, and age groups within the young adult population.



[update by mmoneymike, 10/24]

After adding all the data visualizations we wanted using these datasets, we have began to look where to expand our webpage, which we believe will require more data.

[update by mmoneymike, 11/20]

As we have decided on adding a U.S. map of cancers by state, we are now pulling more data from the CDC, specifically the cancer incidence and mortality *AGE-ADJUSTED* rates by state. We believe we have used the 10 original datasets to the best of their capability and know this expansion will ultimately aid in our entire webpage design.

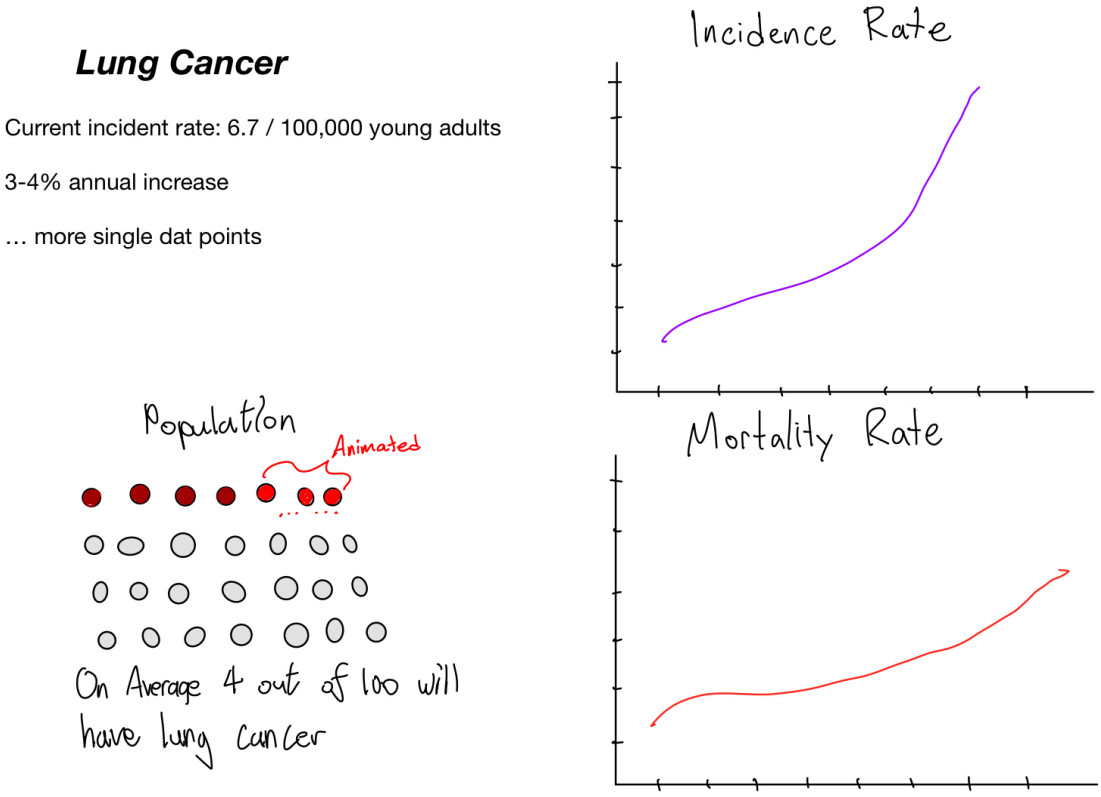
As a result we are expanding the data to 4 more datasets, with a final total of 14 datasets.

[update by mmoneymike, 11/21]

Exploratory Data Analysis

10/24 *Due to the specific data points we want to show, I believe we should show single data points and include graphs for further analysis. My idea of this is shown below of how we initially present the data after the user clicks to explore a specific cancer.*

[entry by mmoneymike, 10/24]



This chart implements the main core of our data visualizations, and recommendations from our TA during our Initial Meeting. Using the Leading Cancer Incidence and Mortality CSV files, we hope to create two interactive charts. Here, we can also add in interactions such as filters and hover abilities.

[update by mmoneymike, 10/24]

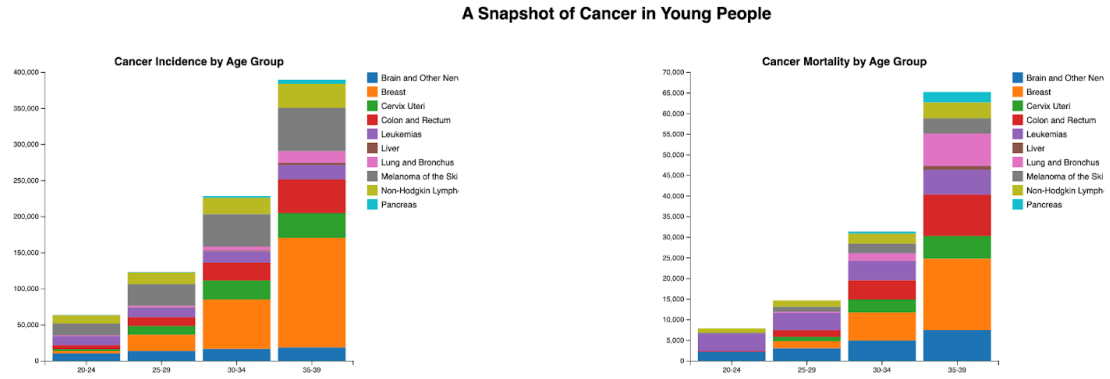
Data Visualization Implementation (Final Choices)

Cancer Incidence and Mortality Rates per 100,000 by Individual Cancer

This data visualization uses human-shaped icons to represent incidence and mortality rates per 100,000 people, making the information more relatable and visually engaging compared to traditional dot-based visualizations. This approach effectively conveys the magnitude of the rates in a way that is intuitive and easy to interpret, emphasizing the human impact behind the numbers. Unlike dots, which can feel abstract, the icons provide a clear, immediate connection to the population being represented.

[entry by mmoneymike, 11/19]

Cancer Incidence and Mortality Rates across Groups

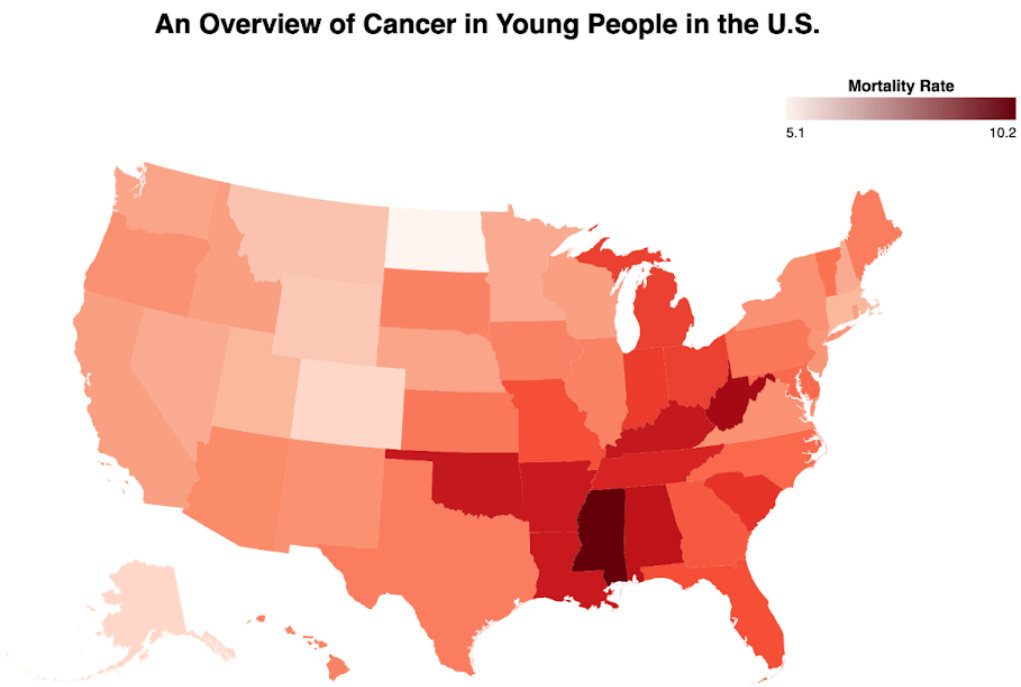


This visualization was originally designed as a stacked area chart to show the cumulative impact of different cancer types across age groups. However, we transitioned to stacked bar charts for clarity and precision. While the area chart effectively demonstrated

overall trends, it made it difficult to compare the contribution of individual cancer types within specific age ranges due to overlapping and blending areas.

The stacked bar chart resolves these issues by clearly delineating the individual cancer types within each age group. Each bar represents a specific age range, and the distinct color blocks make it easier to identify and compare the proportion of cancers like breast, leukemia, and brain cancer. This format provides a more accurate and intuitive view of how cancer incidence and mortality vary across demographics, ensuring the data is accessible and actionable for all users.

[entry by JCpennyChen, 11/21]



In tandem with the two stacked bar charts, we felt it would greatly help the narrative of our project to include another broad visualization. We decided to make a map depicting

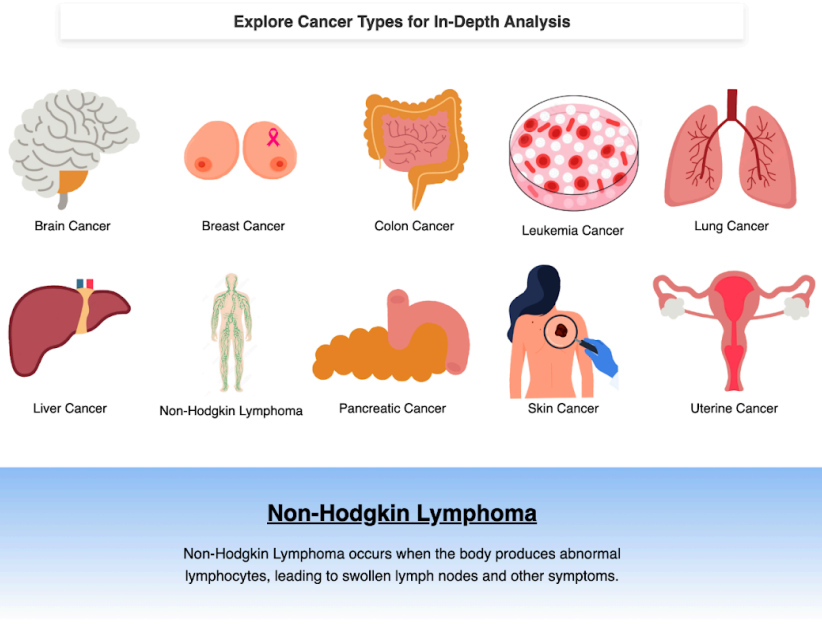
each state's mortality and incidence rate.

Furthering this visualization, we then decided to create a map for each individual cancer's incidence and mortality rate. This not only helps depict more information, but connects the individual cancer visualizations with the overall cancer rate section.

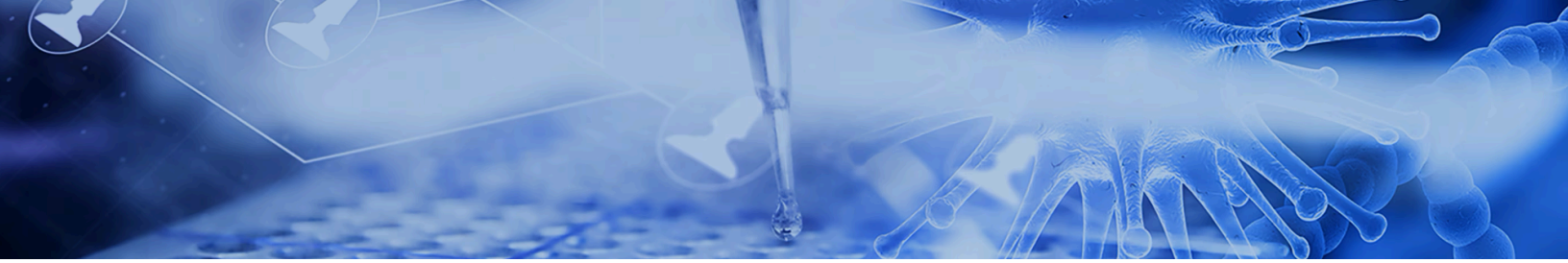
[entry by mmoneymike, 11/22]

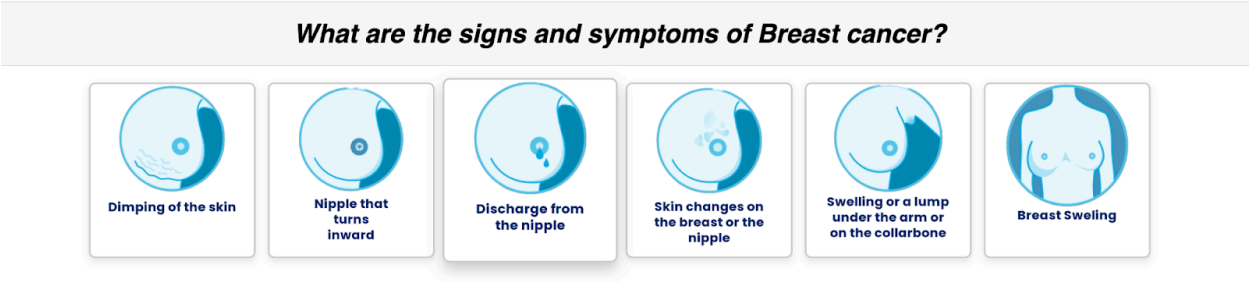
General Implementation

Individual Cancer Pages



A crucial goal of our visualization was to allow users to dive deeper into specific cancers they might be interested in—whether just common interest or they have genetic possibilities of developing these sooner than average.





While our incidence and mortality rates show sufficient data for understanding the current status of cancer rates among young people in the U.S., we felt it was incredibly important that we added in simple descriptives on how to spot certain cancers.

Through our research, we found that many young people run the risk of letting cancers go undetected for longer periods of time (see reference). In order to give more deliverables and takeaways for users, we decided to include generic visuals showing the signs and symptoms of each cancer.

We believe that if we didn't add this into our webpage, we would be leaving out a significant chunk of our mission of our project—thus we added them in.

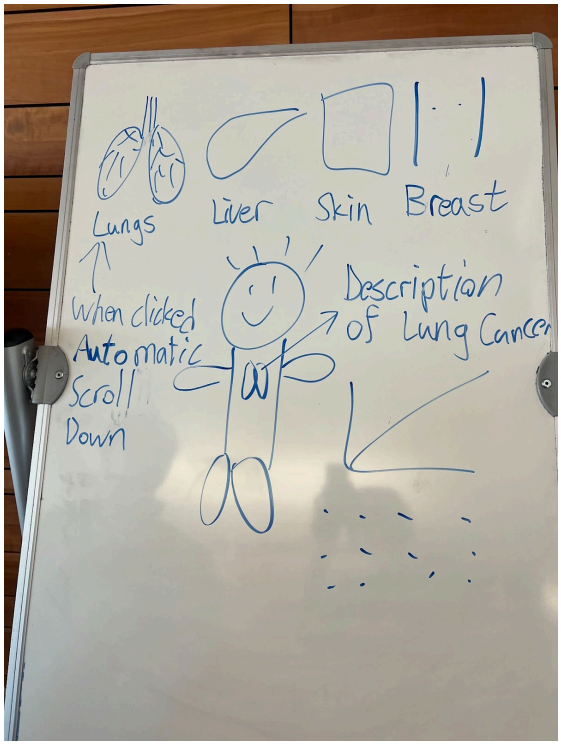
Reference:
https://www.cancer.org/cancer/types/cancer-in-young-adults/finding-cancer-in-young-adults.html

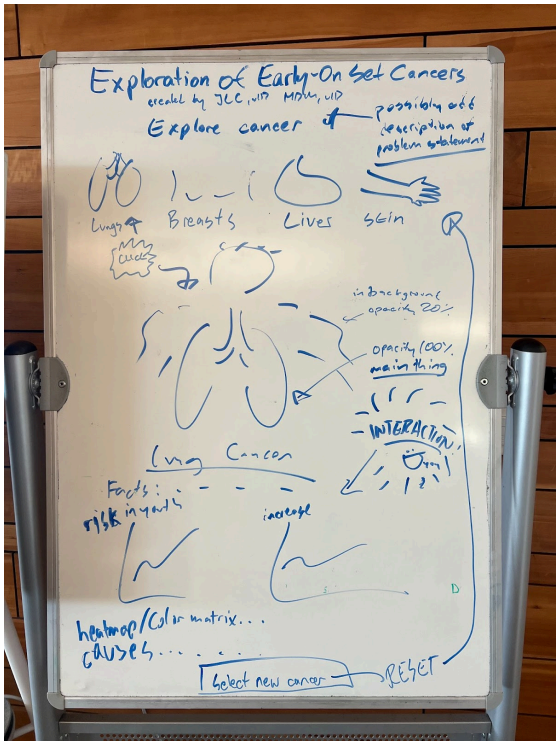
[entry by JCpennyChen, 12/1]

Design Evolution

First Screen User Sees of Webpage

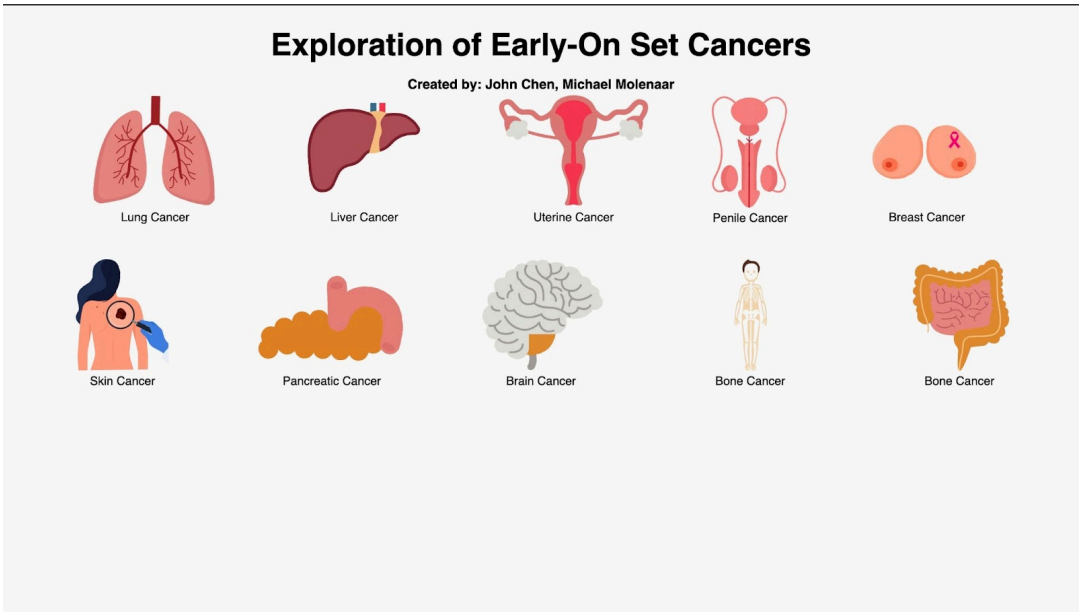
> ***10/12*** *As described in the notes above, I am in charge of the initial design implementation, that being how the user selects different types of cancer. During our meeting on 10/12, we made two, rough outlines for the beginning pieces of the website. This is be the start of the webpage.*
>
> **[entry by JCpennyChen, 10/12]**





The key idea in both of these designs is how the user begins interacting on our webpage. Without showing too much at once, we have decided to use buttons with images (as seen in both sketches) where users can then get an expanded analysis and visualization of the selected cancer type (as seen on the right sketch). We believe this helps mitigate showing way too much information and allows for fun, interaction where a user can stay on the webpage for a while exploring each cancer separately in its entirety.

[update by JCpennyChen, 10/12]

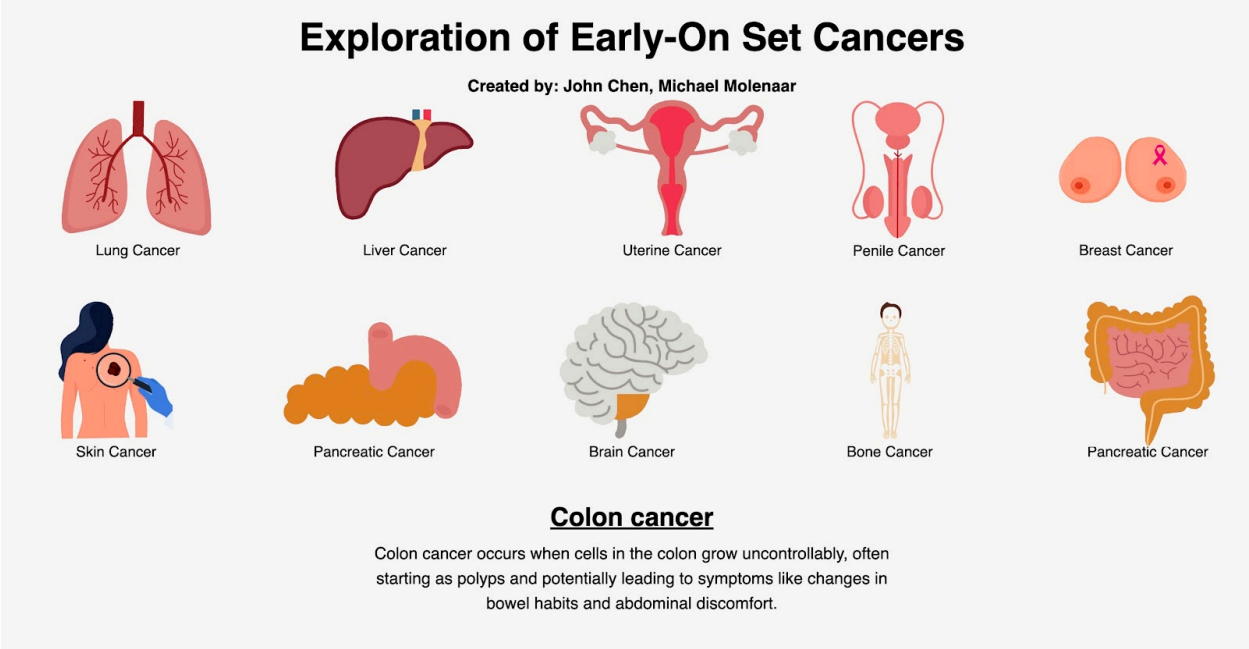


Below is the “first draft” version of the first screen, as explained above. I also decided to make images the buttons for users to go to. We initially planned to use a frame around the images to function as clickable buttons. However, we realized that this design didn’t align well with the overall style of our webpage. So thus, we opted for a subtle zoom-in effect on hover to signal that the images are interactive, creating a smoother, more cohesive experience.

[updated by JCpennyChen, 10/13]

__10/24__ Michael and I met again today, going over what the next steps are for the webpage after someone clicks on a specific cancer they want to explore. Michael made some sketches that I then updated, and have now begun the beginning of the individual cancer pages.

[entry by JCpennyChen, 10/24]



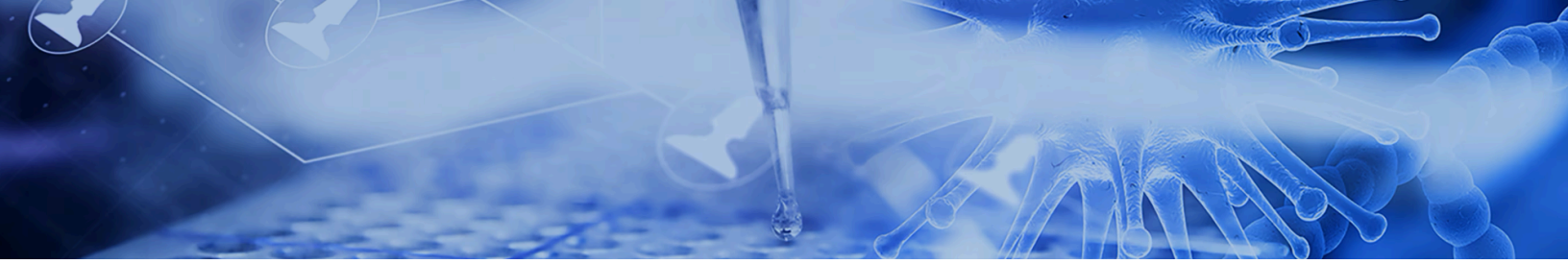
Upon clicking on a specific cancer button, I have added the following:

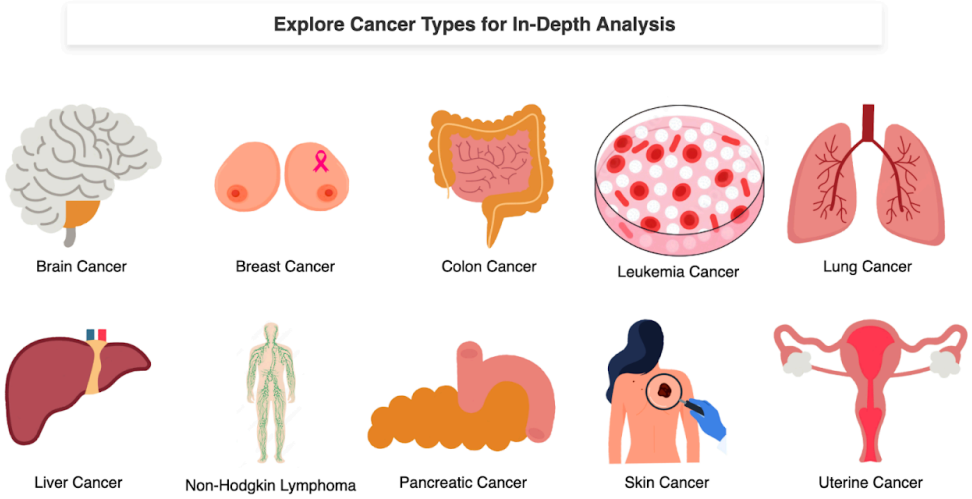
Title - ex: **Lung Cancer**

Description of Cancer - ex: Lung cancer is marked by uncontrolled cell growth..

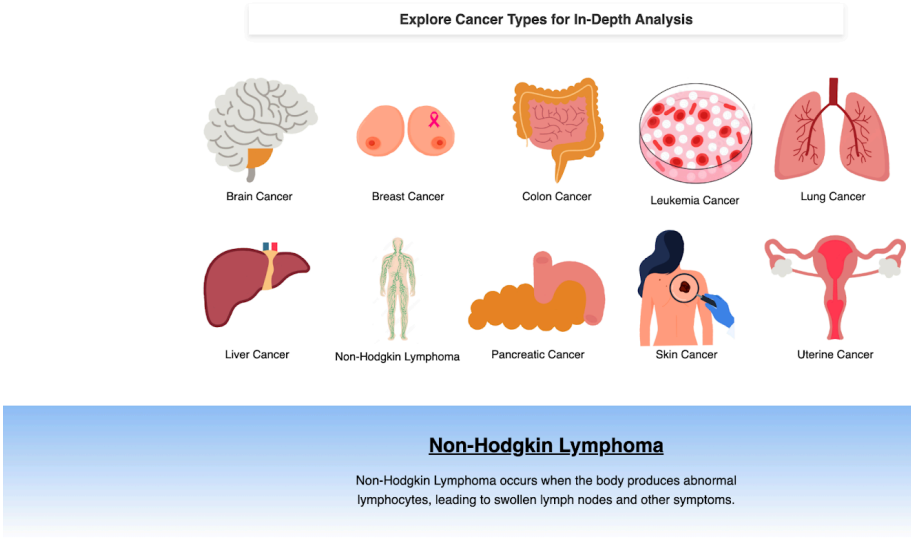
This is small introductory details for the individual cancer pages. With the conjunction of our graphs and other visuals, this can be the core of our visualization and where the user spends most of their time. We should start to think expanding our drafts and creating a fuller cancer page. However, I am prioritizing the introduction right now.

[update by JCpennyChen, 10/24]





After looking at the data more carefully, we decided to change the specific exploratory cancers - primarily including Leukemia Cancer, Non-Hodgkin Lymphoma, and Colon Cancer. This was due to their much higher prevalence than other cancers.

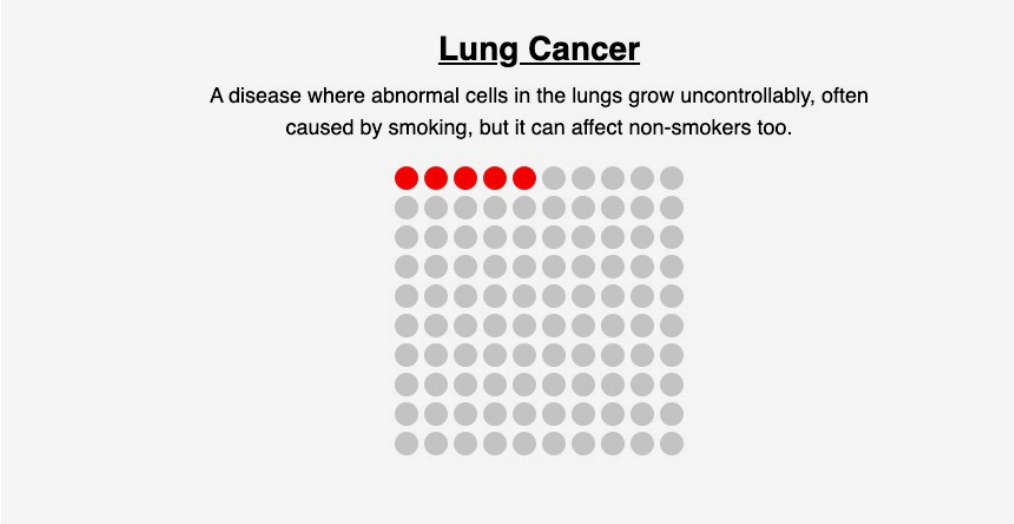


Also, as depicted underneath the blue, we kept the description of the specific cancer, with a change in background to help keep the distinction between the “menu” of cancer types and the currently selected cancer.

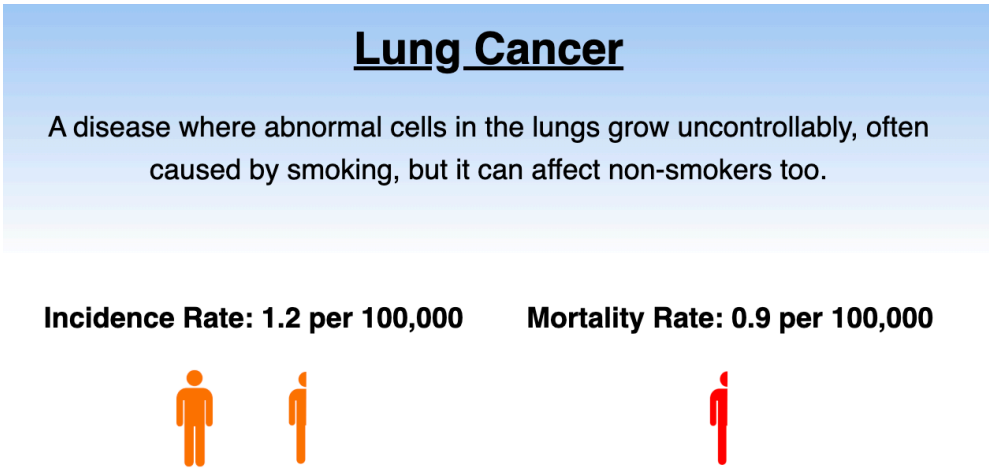
[updated by JCpennyChen, 11/17]

Visualizing Incidence and Mortality Rates

Here is the implemented look of a few of the features above. These features, specifically the text and dot chart, don't necessarily require any mapping to our data files, and thus are easier to implement right now. Below is the current implementation look:



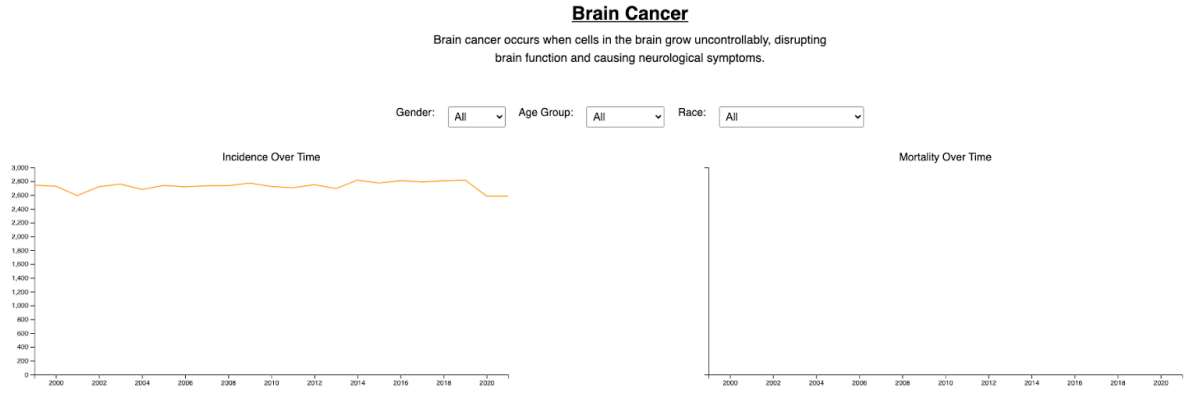
[entry by JCpennyChen, 10/25]



Above is the changed visualization. We switched to a much more simple and invoking design. Instead of representing the rates as dots, we felt it best to depict the rates as people, improving the overall visual design and the narrative of cancers in young people.

Second Screen (after Clicking on Cancer Image)

> ***11/06*** *Here is where we are planning on having our main visualizations for specific cancers. We want to be able to filter our data and have the charts react to show the filtered data.*

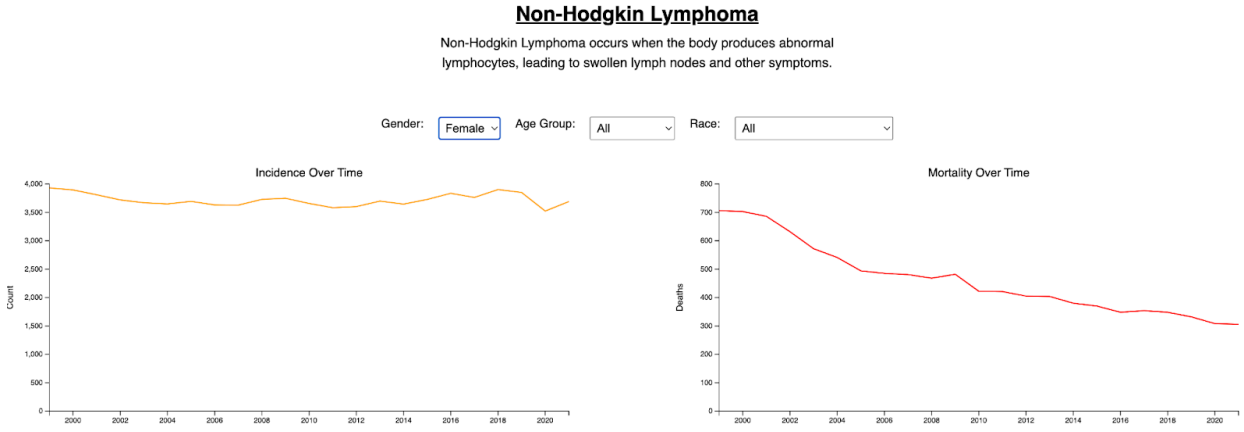


Once a user clicks on a specific cancer image, they are presented with a focused analysis screen. This page includes dynamic, interactive visualizations showcasing both incidence and mortality trends over time for the selected cancer type. The left chart depicts incidence trends, while the right chart highlights mortality, allowing users to explore data patterns effectively. *(Mortality is currently having errors.)*

Dropdown menus for filters (e.g., gender, age group, race) enhance interactivity, enabling users to customize the displayed data.

***Note: This should display the dot visualization as well before the two charts. We believe due to the errors in loading the Mortality dataset, the dot visualization is not appearing properly.

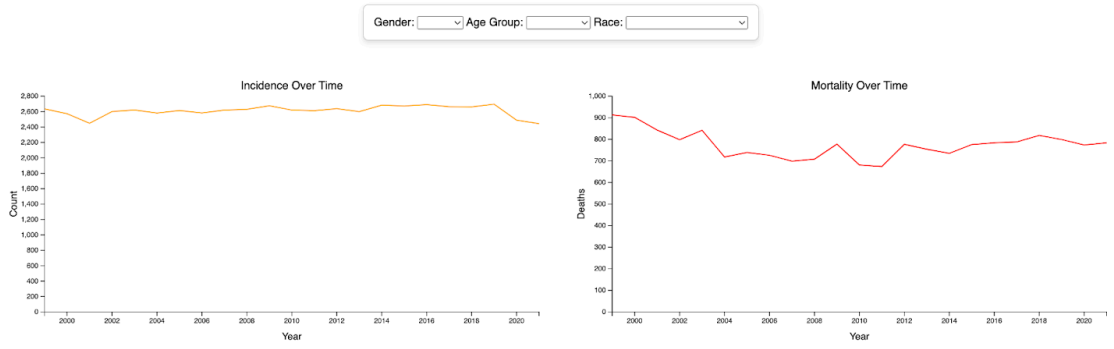
[entry by mmoneymike, 11/05]

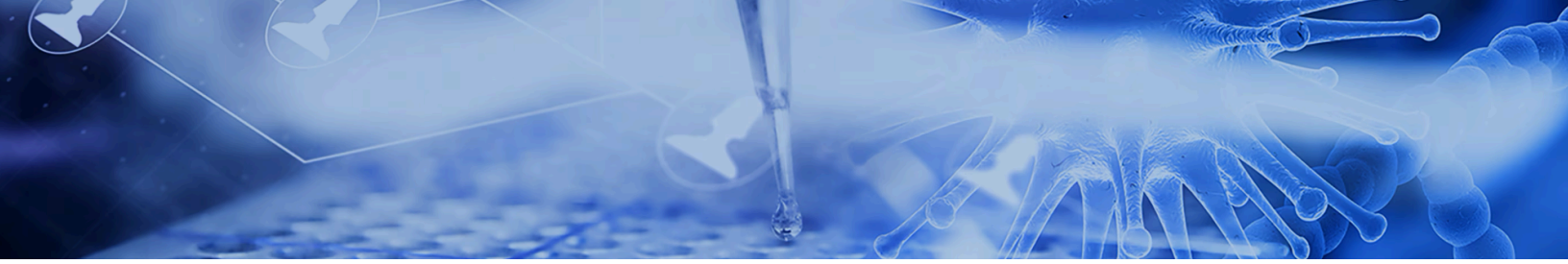


As seen in the chart here, we have been able to get most of the implementation for the graphs running. We are able to have interactions with the filters, which change the graphs' depictions (as seen that we have female selected).

We decided to incorporate these interactive filters for Gender, Age Group, and Race to provide users with a more personalized and detailed exploration of cancer data. By allowing users to dynamically filter the visualizations, we aimed to highlight trends and disparities across specific demographic groups, which are often critical in understanding cancer patterns. This functionality ensures the dashboard is not only informative but also adaptable to the unique questions and analyses users may have, making the tool more valuable for research, education, and decision-making purposes.

[update by mmoneymike, 11/19]

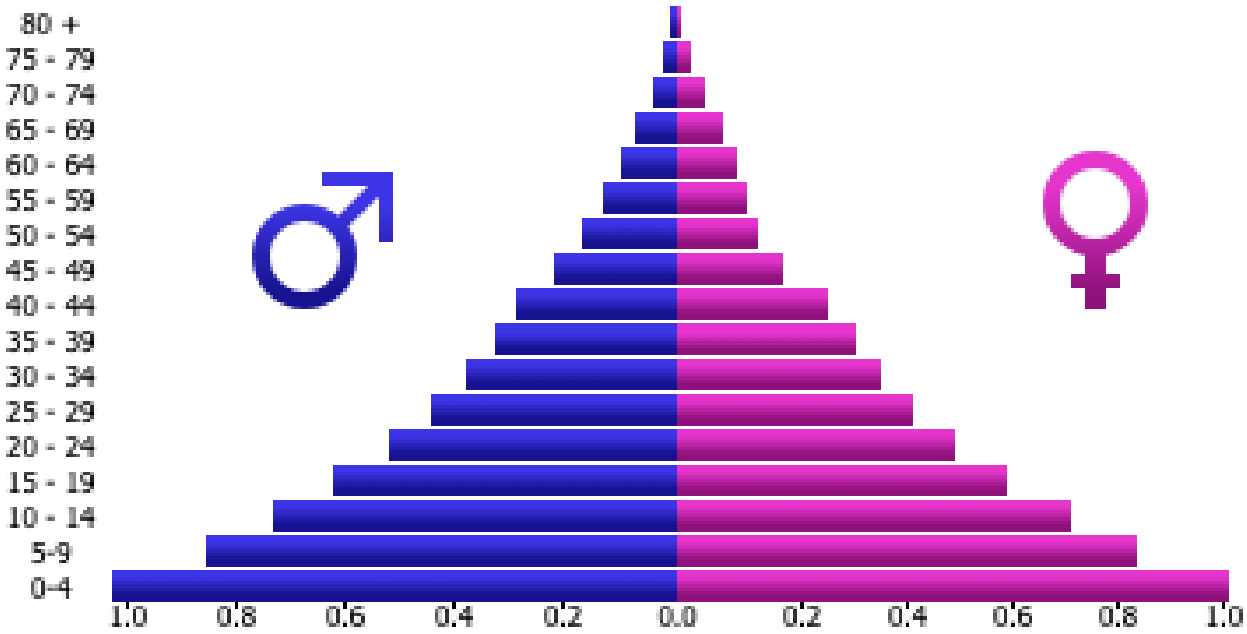




The goal was to enhance the user interface and user experience of the cancer visualization dashboard by improving interactivity, readability, and aesthetics. The updated dashboard provides clearer data insights and integrates both incidence and mortality across years in a cohesive design.

[update by mmoneymike, 11/21]

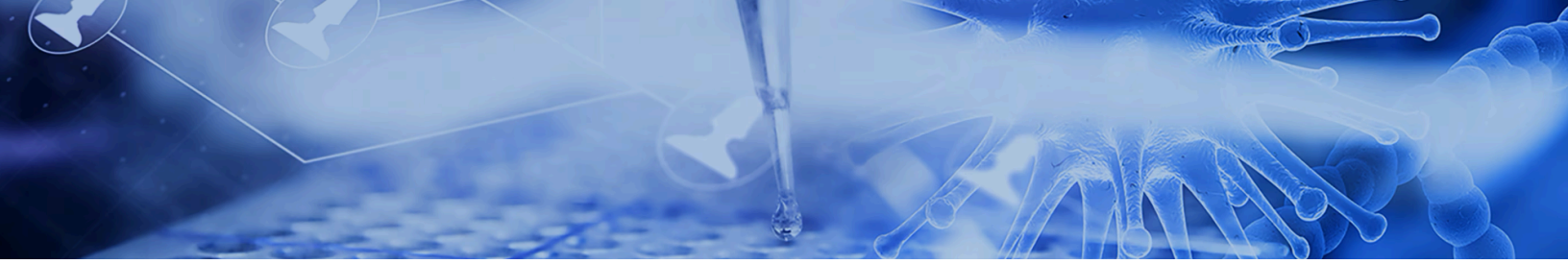
> ***11/20*** *We are currently looking to add new visualizations to enhance the detailed nuances of each cancer type.*
>
> **[entry by mmoneymike, 11/20]**



Example of Age Pyramid

I posited the idea of adding an age pyramid, showing comparison between each sex's specific rate of a specific cancer type, across all age groups. This is a new type of visualization from what we currently have implemented, and can be a nice addition.

[update by mmoneymike, 11/20]

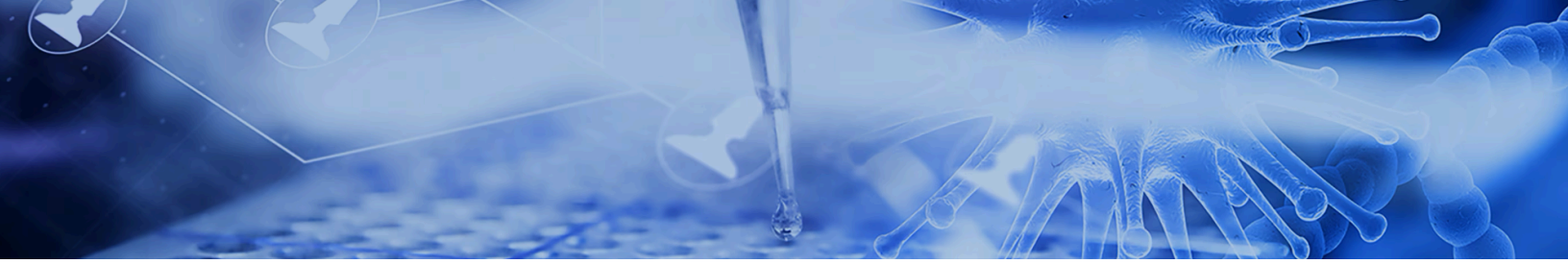


We are ultimately choosing to not include the age pyramid for the time being. Until we can further look at the implementation of it, we believe it doesn't provide a lot of data that is useful for our overall depiction of cancer data.

[update by mmoneymike, 11/22]

Upon further investigation, we have ultimately decided to not include the age pyramid for our final webpage edition. The data is insufficient for specific cancer types, and we feel that while the graphs don't explicitly show a direct comparison, users can still see the differences by gender and age through the line graphs. The addition of the age pyramid feels like it just adds already presented data and deters from the seamless of our data portrayal as a user progresses through our webpage.

[update by mmoneymike, 12/6]



General Code Updates

> ***10/24*** *With finally adding all of our necessary data today, I hope to make the main data structures that will load in our data. With this, there will no doubt be a need for some organization, which may allow us to do some more helper functions for calling the data.*
>
> **[entry by mmoneymike, 10/24]**

I have created our data loading functionality, and have created many helper methods to create specific charts for our visualizations. All csv files have keys with file mapping, and can now be used to begin our visualization.

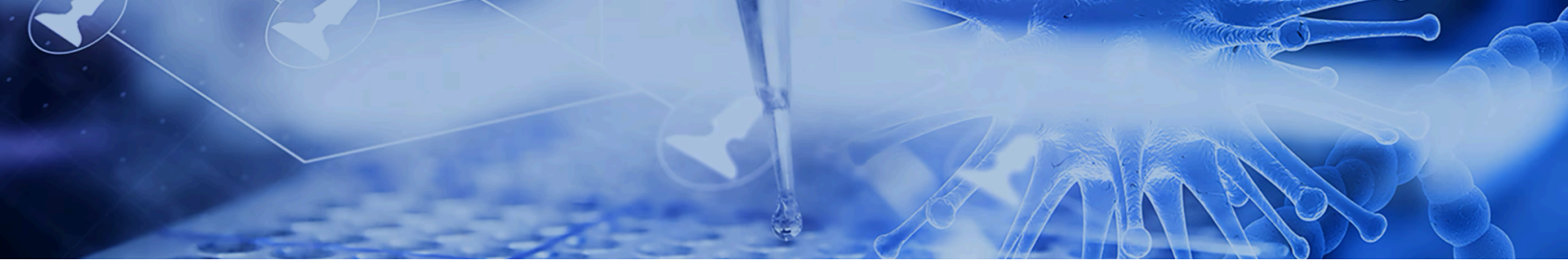
Next steps should now focus on how can implement the proposed first visualizations, specifically the interactive line charts of cancer incidence and mortality rates. I think we should focus on one specific page (i.e. Lung Cancer) and really incorporate as much readable, interesting and useful visuals. We currently have the three proposed, but we can do more research on different visualizations and techniques.

[update by mmoneymike, 10/25]

> ***11/01*** *We implemented a significant update to how interactions are handled for visualizing different cancers. Previously, the <**a**> tags in the HTML used inline **onclick** attributes to trigger visualizations directly. While functional, this approach cluttered the HTML and made it harder to maintain or extend functionality. Today, we transitioned all these interactions to JavaScript, using **addEventListener** for handling clicks.*
> **[entry by mmoneymike, 11/01]**

I created a centralized JavaScript object to store metadata for each cancer type, including its ID, display name, and description. Using **addEventListener**, we now dynamically attach click handlers to each link during the page load. This ensures our logic remains entirely in the JavaScript files, improving separation of concerns.

This update significantly cleans up the HTML, makes the codebase easier to scale, and supports future enhancements without requiring changes to the HTML structure.



Next steps will involve debugging existing issues caused by moving away from inline attributes, verifying all visualizations load as expected, and ensuring a smooth user experience. Additionally, this update sets the foundation for programmatically generating new cancer types or visualizations.

[update by JCpennyChen, 11/19]

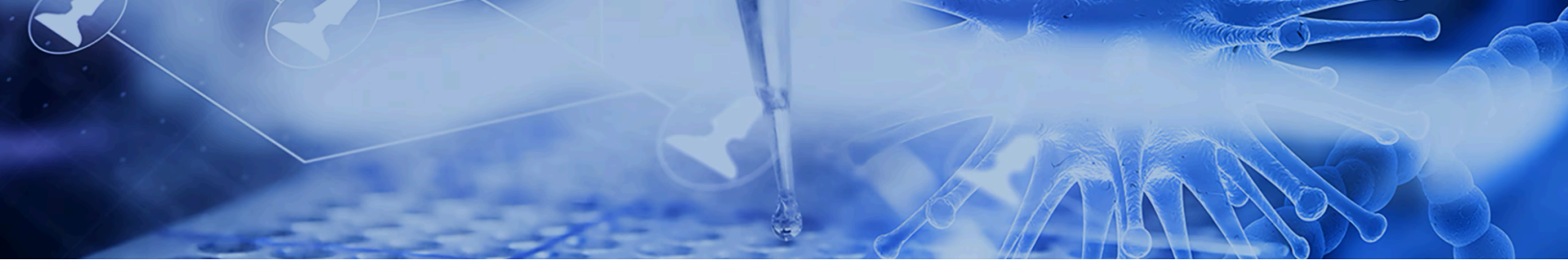
To improve the dot graph visualization, I implemented a detailed mapping system that bridges the interaction between the HTML structure and JavaScript functionality. Each cancer type in the visualization now has a corresponding mapping in JavaScript, linking user-friendly display names from the HTML interface to technical keys used in the data processing logic. This mapping ensures consistent communication between the UI elements and the underlying data.

The detailed mapping system implemented significantly improved the functionality and accuracy of the dot graph visualization. This new mapping system creates a solid foundation for the overall functionality between files, which we'll expand on with the rest of the graphs and visuals.

[update by mmoneymike, 11/19]

After looking at the dot visualizations, I realized that it didn't provide anything impactful if we had 4 dots colored and 999,996 left blank. So, I switched the functionality to be images of people, with only the incidence and mortality rate depicted. This felt like it helps with the storytelling, as showing 4 people "dead" from the mortality rate is a much more provoking data visual than dots.

[update by mmoneymike, 11/20]



Evaluation

What We Learned from the Data

By using our visualizations, we were able to see patterns and trends in cancer data that weren't immediately obvious in raw numbers. For instance, the stacked area charts helped show how certain cancers, like breast cancer, are more common in younger age groups, while others, like pancreatic cancer, tend to affect older populations. The U.S. map visualizations also made it easy to spot regional differences in cancer rates, which could point to disparities in healthcare access or environmental factors.

How We Answered Our Questions

Our visualizations allowed us to explore important questions about cancer. For example, we wanted to see if high incidence rates in certain areas were linked to high mortality rates, and the maps helped us identify those connections. The filters for age, gender, and race made it possible to focus on specific groups and get a better understanding of how cancer impacts people differently. This level of interactivity gave us clear and actionable insights.

How Well the Visualizations Work

Overall, we feel our visualizations do a great job of presenting complex data in a simple and interactive way. Features like synchronized updates between charts and the ability to explore specific cancer types make the tool user-friendly and informative. That said, there's room for improvement. For instance, some legends could be more clear, and the visualizations could be optimized better for mobile devices so that more people can easily use them.

How We Could Improve

To make this project even better, we could add more data, like survival rates or links to treatment success in certain regions. We could also include more filtering options, such as income levels or access to healthcare, to give a deeper understanding of the factors influencing cancer trends. Another big improvement would be adding predictive tools to show future trends, which could help researchers and policymakers. Lastly, we'd like to be able to present current research and ways people can get involved in cancer research.

[update by mmoneymike, 12/5]