CS 4460 Process Book

Team Name - Visualizers
Team Members
Anne Youn (ayoun8@gatech.edu) - Team Leader
Aaron Wang (awang653@gatech.edu)
Namrata Kondala (nkondala3@gatech.edu)
Kirito Machida (kmachida3@gatech.edu)

Team Agreement
- How will you communicate?
- We will use Discord and Messages to communicate for the project.
    - The general channel will be used for important updates and general communication, and the important links channel will be used for important links. The tasks channel will be used to keep track of tasks.
    - For calls, we will use the Discord voice channel.

- How/when will you meet?
- We will meet twice a week: once during our Wednesday lab and once virtually.
- We will meet in person/virtually on Mondays before class for thirty minutes.

- How will you collaborate on implementation?
- We will allocate specific tasks to each member and track and uphold deadlines together.
- We will try to partition based on relevant skills and try to have members work on one throughline.
- We will agree on which work to take and keep track of it using a google sheet and discord (the events feature and in a channel)

- How will you deal with non-performing members?
- We are establishing expectations at the beginning of the project and will hold all members to the team agreement.
- Three missed deadlines will result in a group discussion.
    - The team members and team leader will make sure to communicate with the non-performing members to understand why they are not completing their tasks and try to find a solution.
    - If someone feels they will miss a deadline, they are expected to notify the team ideally a few days in advance but anytime ahead of the last day of the deadline.
- If someone does not do their allocated work or becomes aggressive towards other members, we will discuss the situation with our professor.

- Deadlines
- Deadlines should be met. However, life happens, so if there is proper communication, then pushing some deadlines is okay.
- Deadlines should be set earlier than when we need that item by, so that way if there is a “life happens” situation, our project does not fall apart.
- Deadlines will be set together as a group, so there will be no surprises.

Project Proposal
Title: Exploring Brain Disease Causes and Impacts

Abstract (200-300 words):

The data from common brain diseases, such as Alzheimer's, Parkinson's, and migraines, will be used to analyze, compare, and communicate the symptoms, conditions, and treatments of these diseases. We will be able to draw conclusions from these metrics to see how the diseases are distributed across regions in the United States and worldwide. We will also be able to better understand different at-risk groups and the extent to which these conditions affect populations.The dataset is based on epidemiological, demographic, and socioeconomic data and provides three distinct factor types: disease prevalence, treatment access, and patient demographics. The complementary dataset breaks down geographic trends and socioeconomic disparities related to each condition. By creating visual comparisons of the dataset, we will explore how these diseases vary by geography and socioeconomic status, offering a clear view of correlations between environmental or social factors and disease impact. The goal of this project is to support evidence-based strategies for disease prevention and equitable healthcare resource allocation. We will achieve this goal through the use of visual comparisons of prevalence and demographic data provided by the dataset and comparing this with socioeconomic and geographic data to discover overall insights between both metrics.

Potential Data Sources:

- https://atlasofms.org/map/global/epidemiology/number-of-people-with-ms
-

Signatures

Anne Youn (ayoun8@gatech.edu) - Team Leader 

Aaron Wang (awang653@gatech.edu) 

Namrata Kondala (nkondala3@gatech.edu) 

Kirito Machida (kmachida3@gatech.edu) 

Map

TOPIC:

Types of neurological diseases/disorders and their region distribution

- Epilepsy- Anne
  - Overall:
    - https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(24)00038-3/fulltext
    - https://www.thelancet.com/action/showFullTableHTML?isHtml=true&tableId=tbl2&pii=S1474-4422%2824%2900038-3
    - https://www.who.int/news/item/14-03-2024-over-1-in-3-people-affected-by-neurological-conditions--the-leading-cause-of-illness-and-disability-worldwide
    - https://my.clevelandclinic.org/health/diseases/22934-brain-diseases
    - DALYs, prevalence, mortality: https://pmc.ncbi.nlm.nih.gov/articles/PMC7607495/
  - Epilepsy:
    - https://www.who.int/news-room/fact-sheets/detail/epilepsy
    - https://www.cdc.gov/epilepsy/data-research/facts-stats/index.html
- Alzheimer's - Namrata
  - https://pmc.ncbi.nlm.nih.gov/articles/PMC11066745/
  - Data banks:
    - https://www.cdc.gov/healthy-aging-data/data-portal/index.html
    - https://data.cdc.gov/browse?category=Healthy+Aging&sortBy=relevance&page=1&pageSize=20
    - https://www.parkinson.org/understanding-parkinsons/statistics/prevalence-incidence
    - https://www.radc.rush.edu/reports/demographicsReport.htm
- Parkinson's - Kirito
  - Data banks:
    - https://www.parkinson.org/understanding-parkinsons/statistics
    - https://pmc.ncbi.nlm.nih.gov/articles/PMC12035419/
    - https://pmc.ncbi.nlm.nih.gov/articles/PMC2865395/
    -
- Multiple Sclerosis (MS) - Aaron
  - https://www.nature.com/articles/s41597-025-05250-y (not quite sure what this has tbh, I see mostly figures rather than numbers)
  - https://atlasofms.org/map/global/epidemiology/number-of-people-with-ms
    - https://atlasofms.org/what-is-the-atlas-of-ms (scroll down to bottom for Excel worksheets)
  - https://www.nationalmssociety.org/
  - https://msda.emif-catalogue.eu/ (this website takes a while to load…)

1. Data Collection and Project Plan

I. **Who is your audience? Come up with at least three options and pick one target audience.**

- Older Generations (most susceptible population) - TARGET AUDIENCE
- General Public/Journalism Outlets
- Younger Generations (preventative population)
- Healthcare officials
- Caretakers
- Government Agencies/Policy Makers

II. **Describe your target audience in more detail. What do they know? What are their interests? What visualization literacy do they have? At what level of detail will you present information to them?**

- Our target audience is the older generations, but also the general public. Most elders do not know much about the brain and the diseases that can affect it. This age group is an ideal target audience since it is the age at which individuals should be informed about the risks of common brain diseases and possible preventative measures for the onset and further progression of the diseases. As most elderly individuals are interested in staying healthy, this furthers our notion. Elderly individuals also tend to be less proficient in using technology and are likely not used to reading into difficult visualizations, so we will present the information in an easy-to-understand manner to maximize impact.

III. **What questions about your data will be interesting for your audience? Come up with a list of interesting questions that your audience may have about your data. The more, the better, but your team should come up with at least 15 questions.**

1. General Questions:
    - 1.1. What are the most common brain diseases?
    - 1.2. Top 10 neurological disorders with the highest DALYs?
    - 1.3. Top 10 neurological disorders with the highest prevalence?
    - 1.4. Top 10 neurological disorders with the highest mortality?
    - 1.5. How vulnerable am I?
        - 1.5.1. Exploratory - age, gender, region inputs -> risk output
    - 1.6. What is the most effective preventative care?
    - 1.7. When are the onsets of [disease]?
    - 1.8. What are the most common symptoms of [disease]?
    - 1.9. What are the mortality rates for [disease]?
    - 1.10. What are the costs of treatments for [disease]?
    - 1.11. Does gender impact the risk of getting [disease]?

1.12. Does location impact the risk of getting [disease]?
1.13. Does age impact the risk of getting [disease]?
1.14. What are the most common barriers to getting proper treatment for [disease]?
1.15. What is the economic status of the most common countries for [disease]?
1.16. How quick is the progression of [disease]?
1.17. What are the most common early warning signs of [disease]?
1.18. What makes a region more susceptible to [disease]?
1.19. Which environmental factors affect the rates of [disease] the most?
1.20. At what point are symptoms a concern?

IV. What data do you have? Give a brief description of each attribute and its data type (categorical, ordinal, or quantitative) in your process book. It's OK if you are not sure about the data type for some attributes - you can simply describe them (e.g., geographic location).

- Geographic location
    - U.S. cities & states, world countries
- Name of Neurological Disease/Disorder: Categorical
    - The specific disease/disorder. The primary grouping variable.
- Absolute number of DALYs: Quantitative
    - DALYs: Disability-Adjusted Life Year = lost healthy years (years of life lost + years lived with disability)
    - Reported in millions of life years
- Prevalent cases: Quantitative
    - Prevalence: The number of existing cases in our population
    - Reported in millions of cases
- Mortality: Quantitative
    - Number of deaths from the disease/disorder
    - Reported in millions of deaths
- Mental Unwellness: ordinal
    - Survey respondents reported scales of mental unwellness symptoms and are given a relative score based on how (un)healthy they are
- Gender: Categorical
    - Shares the demographic gender of the Alzheimer's patient
- Race: Categorical
    - Shares the demographic race info of the Alzheimer's patient
- Age: Quantitative
    - Shares the demographic age info of the Alzheimer's patient

- Mean Age of Onset (Male & Female): Quantitative
    - Average age of when MS cases begin
- Percent of MS Incident/Prevalent Cases (Male & Female): Quantitative
    - Percent of cases that affected [male/female]

V. **Do you expect to do substantial data cleanup? How will data processing be implemented? Try to minimize the amount of cleanup you have to do by finding cleaned and ready-to-go data sources whenever possible.**
    - For some of the data and uniting data, we anticipate a fair degree of pruning and cleaning. We will want to have as much data for our visualizations as possible so pulling from multiple sources means normalizing them. Gathering as much data as possible also means finding large sets that contain a lot of information we may not need, which will require pruning.
    - CDC survey data is compressed to XPT files with technical names for all headers. Using Python (and libraries such as pandas), we were able to not only pull and convert the data but alter headers to be more readable and concatenate data sets like demographic information with diagnosed migraine, marital, and educational status.
    - Tableau may label some categories as incorrect data types due to cells being filled with "Not Available" or something similar, which will require parsing through the categories to change the data types manually.

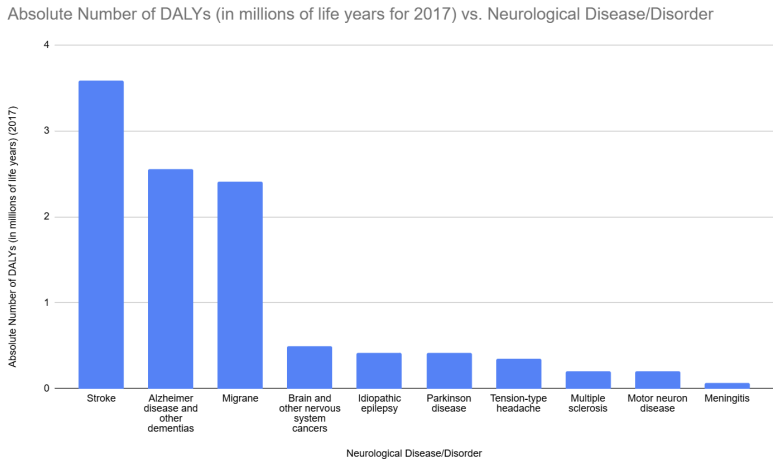
---

2. Data Exploration

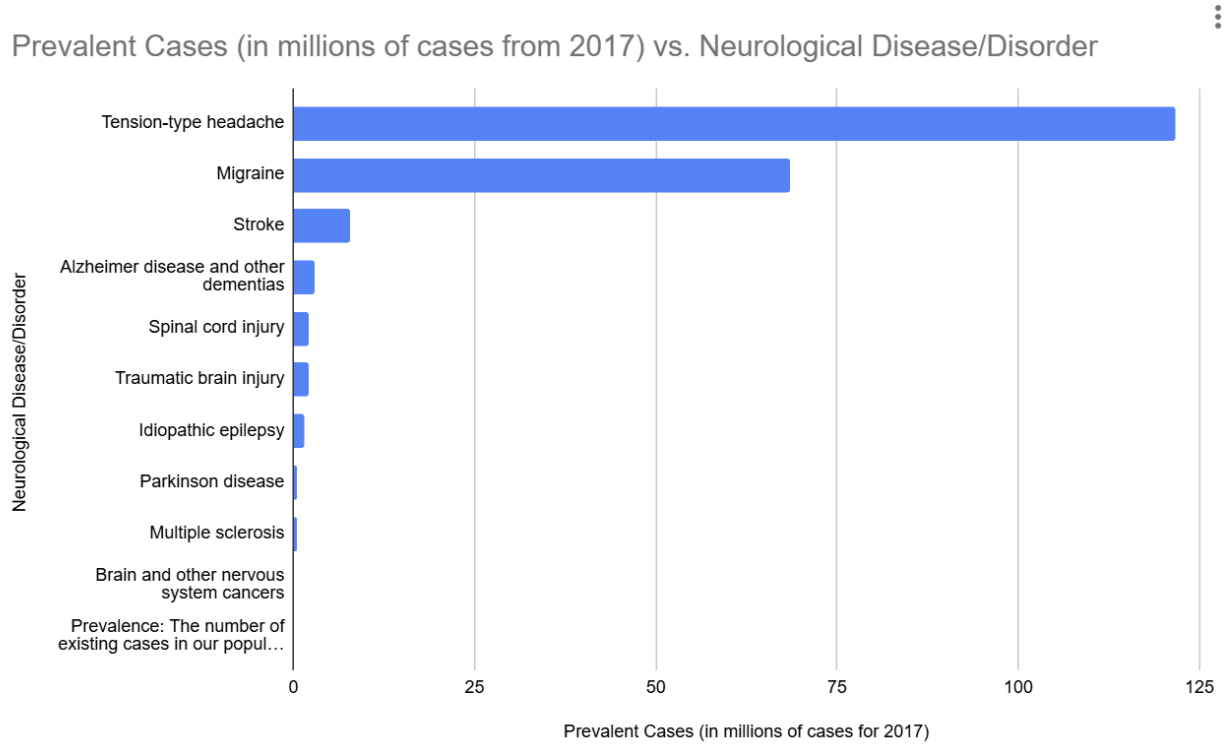
I. **Use Tableau/Excel/Python/R (or any visualization tool you like) to create preliminary visualizations to try to answer the proposed questions. If you realize some questions cannot be easily answered, this is totally fine. Just make a note in the process book and explain why.**
    - Since most of our questions are formatted to answer the same question for each of our different diseases (the ones that have "[disease]"), we decided to only put one visualization per question for each of these as a preliminary answer.
    - 1.2 ("How vulnerable am I?") cannot be easily answered because it is a broad question that needs more analysis from multiple data sources.

Anne:

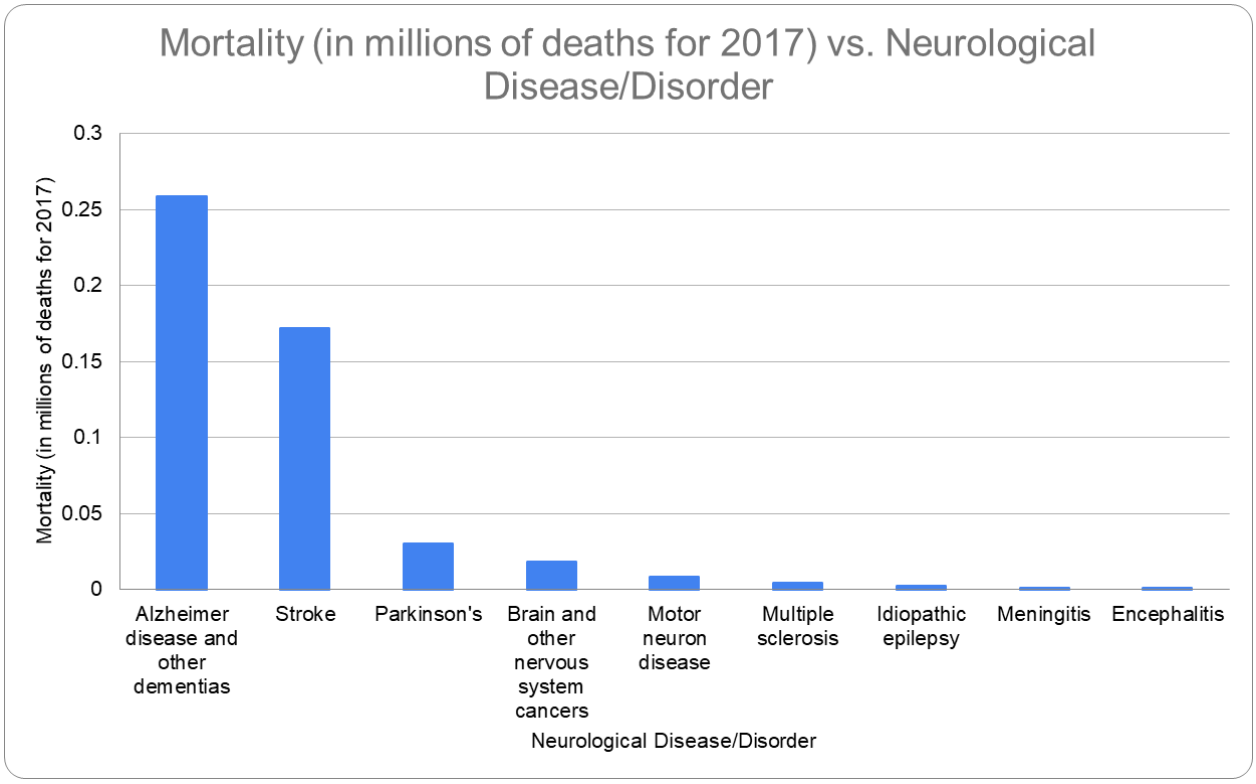
- Figure One: DALYs: Disability-Adjusted Life Year = lost healthy years (years of life lost + years lived with disability)



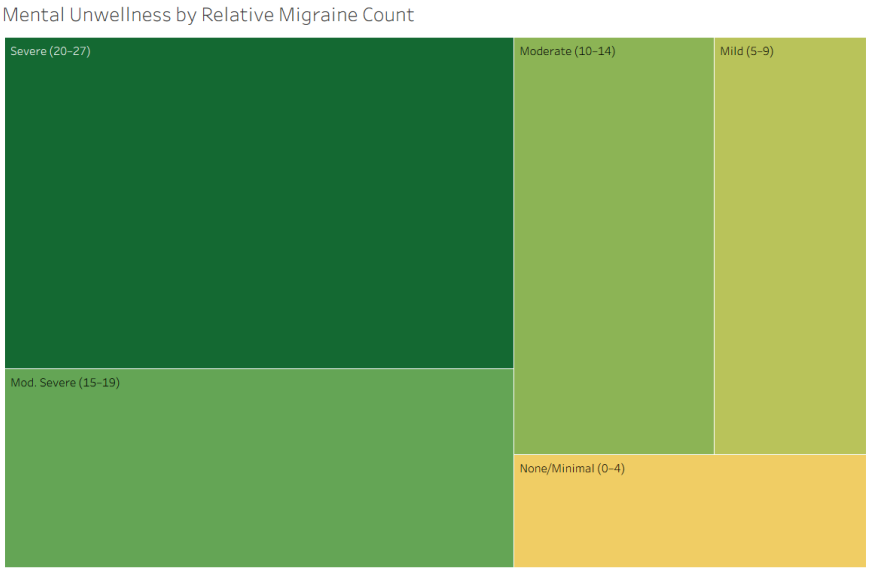
- Figure Two: Prevalence: The number of existing cases in our population



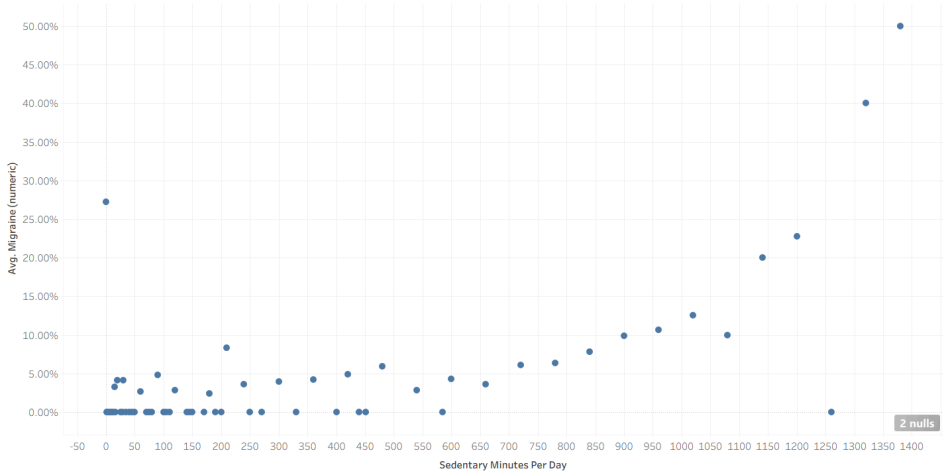
- Figure Three: Mortality



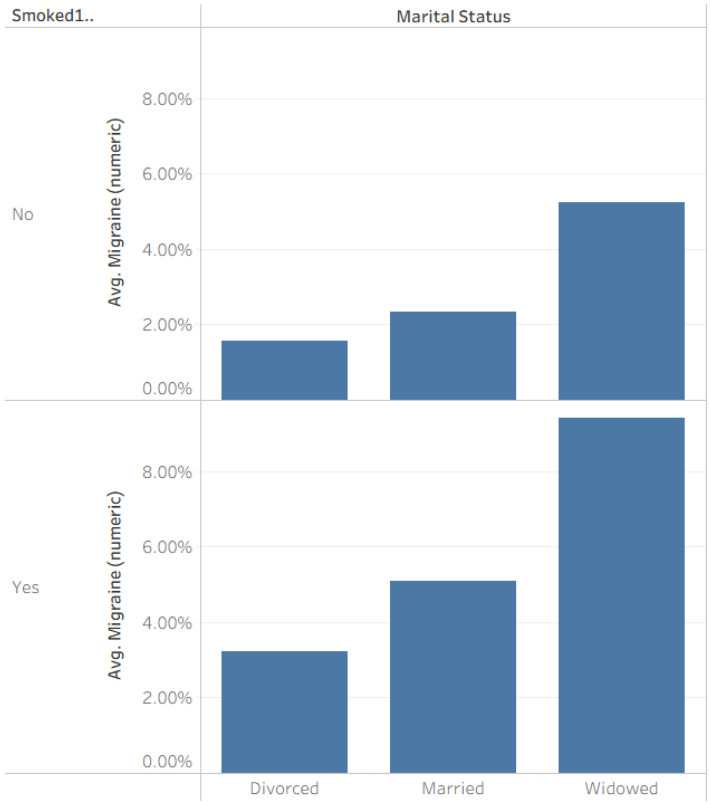
Kirito

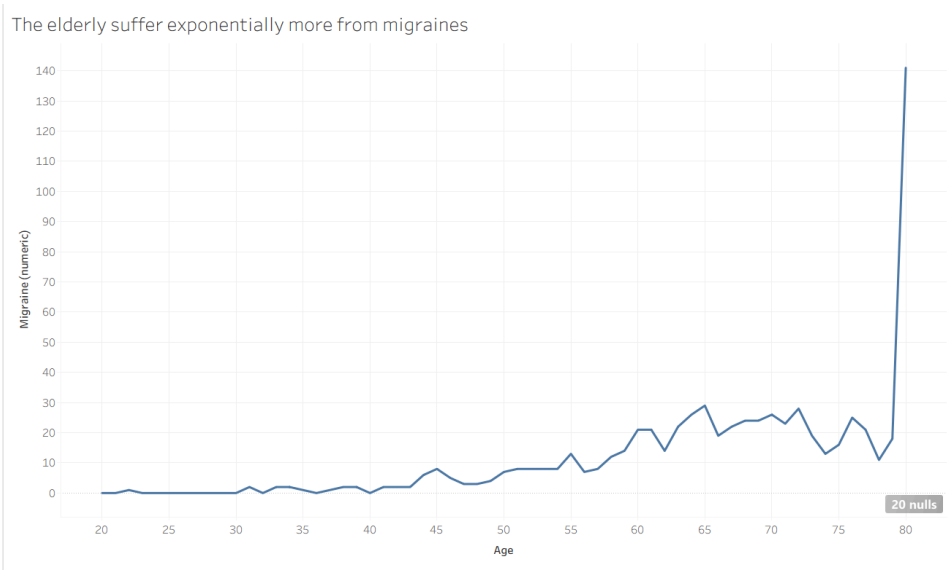


Sedentary lifestyle by percent who report migraines



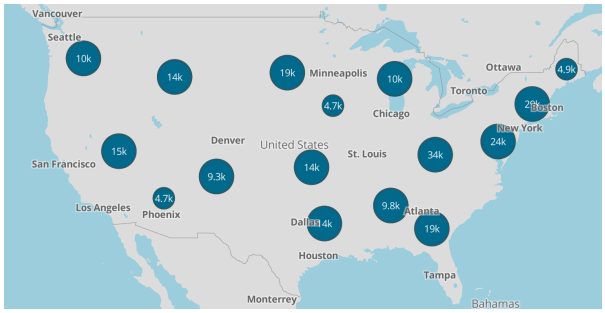
Those widowed most likely to suffer migraines, even more so if they smoke



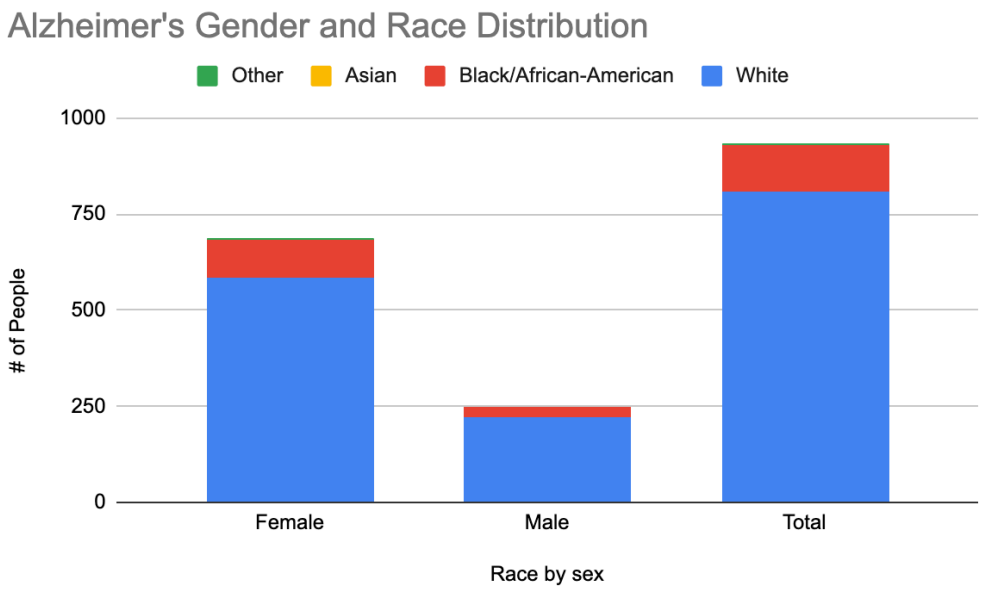


Namrata:

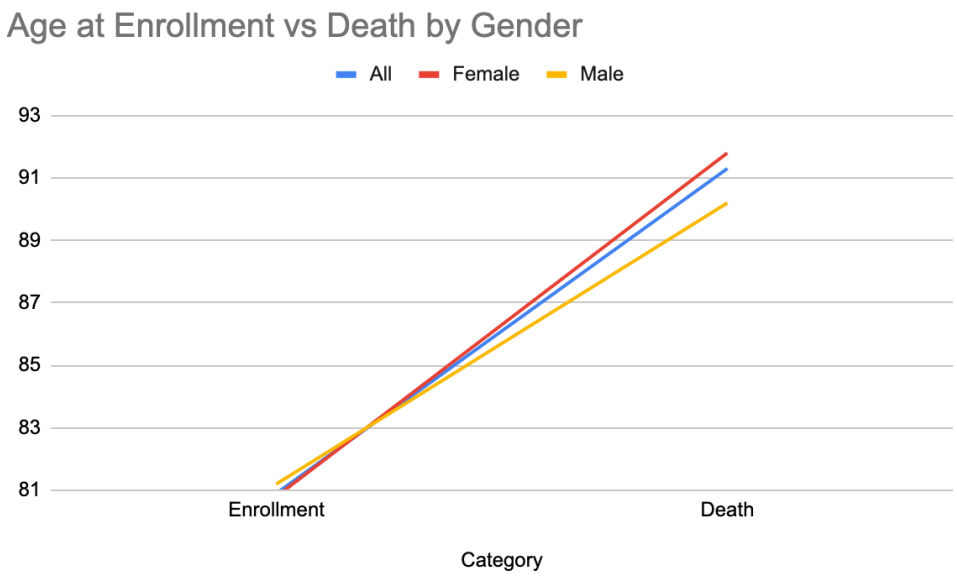
Alzheimer’s Case Distribution across the United States



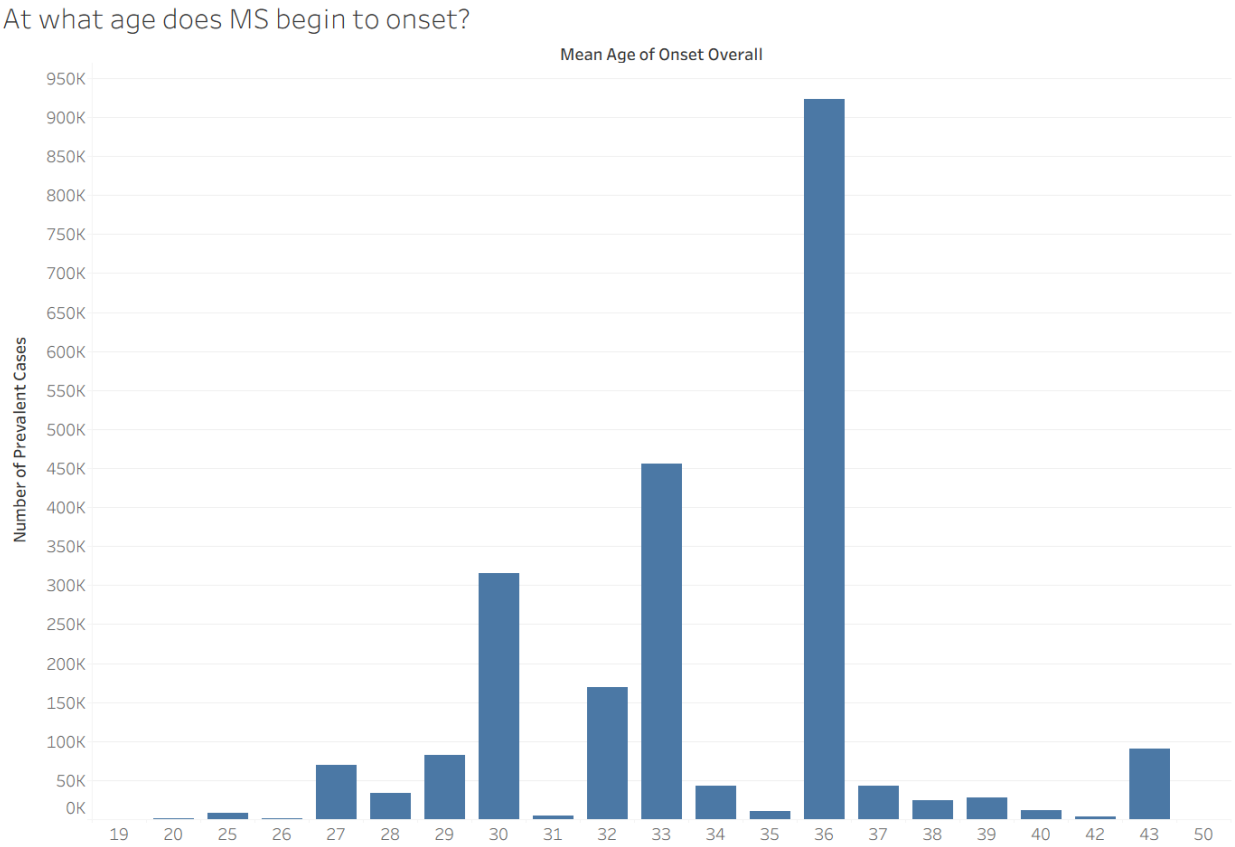
Alzheimer's Distribution over Sex and Race - Sample Survey Data



Alzheimer Patient Study - Age at Enrollment vs Death by Gender

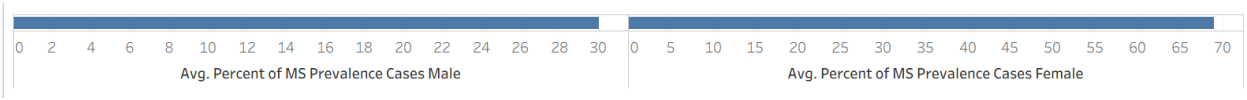


Aaron:



Number of Incident Cases per Country:

Male vs. Female MS Cases:



(Aaron was struggling to use Tableau due to conflicts with files and Tableau Public.)

II. Since these are preliminary visualizations, I used Excel and made simple ones. All of the 10 conditions listed are the top 10 for that category.
    a. Important findings:
        i. As seen in Figure 1, Stroke has the highest DALYs, which means it is the disease that results in the most amount of healthy years lost in the U.S.
        ii. Alzheimer's has a relatively low prevalence (Figure 2) but the highest DALYs and mortality (Figures 1 and 3), which shows how it is extremely lethal, disabling, and burdening.
            1. Tension-type headache, on the other hand, has a high prevalence (Figure 2), but has lower DALYs (Figure 1) and mortality rates (Figure 3), indicating that it is the most common but not deadly.
        iii. Diseases with low prevalence have relatively high DALYs, showing how burdensome diseases can be relative to how rare they are.
        iv. Alzheimer's/dementia diseases and stroke caused the most deaths, showing how mortality rate is concentrated in a few disorders.
    b. Kirito Findings

i. Mental illness has a moderate correlation with higher migraine diagnosis. This is especially poignant considering the number of people reporting "severe" mental unwellness is substantially smaller than those in any other category. It may be a symptom or a cause, potentially a feedback loop between them as well.
ii. There is a similar correlation with sedentary lifestyles to migraines as well as smoking.
iii. Widows suffered much more from migraines than other marital statuses but a keen detail is that smokers doubled the percent chance of getting migraines.
iv. The elderly are at much higher risk of reporting migraines.

c. Namrata Findings
i. The distribution of Alzheimer's cases throughout the United States is shown in the 'Alzheimer's Case Distribution across the United States' map. It shows a high number of cases around major cities and mostly along the East coast.
ii. The 'Alzheimer's Distribution over Sex and Race - Sample Survey Data' chart shows the distribution of the sample study of the Alzheimer's patients. The chart shows the distribution over gender and also the overall distribution by race as well. There is a higher number of females with Alzheimer's based on the study and majority white race.
iii. The 'Alzheimer Patient Study - Age at Enrollment vs Death by Gender' line graph shows the enrollment to death age of females and males of the study. We can see that the age of death for males is lower than that of the females though both females and males started the enrollment at very similar average ages.

d. Aaron Findings
i. MS typically begins to onset in a person's early 30s, although it can begin as early as in their 20s or as late as in their 40s.
ii. MS consistently affects women more than men; on average, prevalent cases are split 70-30 between women and men.
iii. MS seems to be particularly common in Europe, Sub-Saharan Africa, and Latin America from the map of "Number of Incident Cases per Country".

III. **Write down a list of important or interesting findings you found from the visualizations. The more, the better, but at least 3-4 per team member. Those will be important resources for you to build up your data story later. Record these findings as well as screenshots of your visualizations in your process book with each team member's name.**

- Anne: Done
- Aaron: Done
- Namrata: Done
- Kirito: Done
- All findings have been listed above.

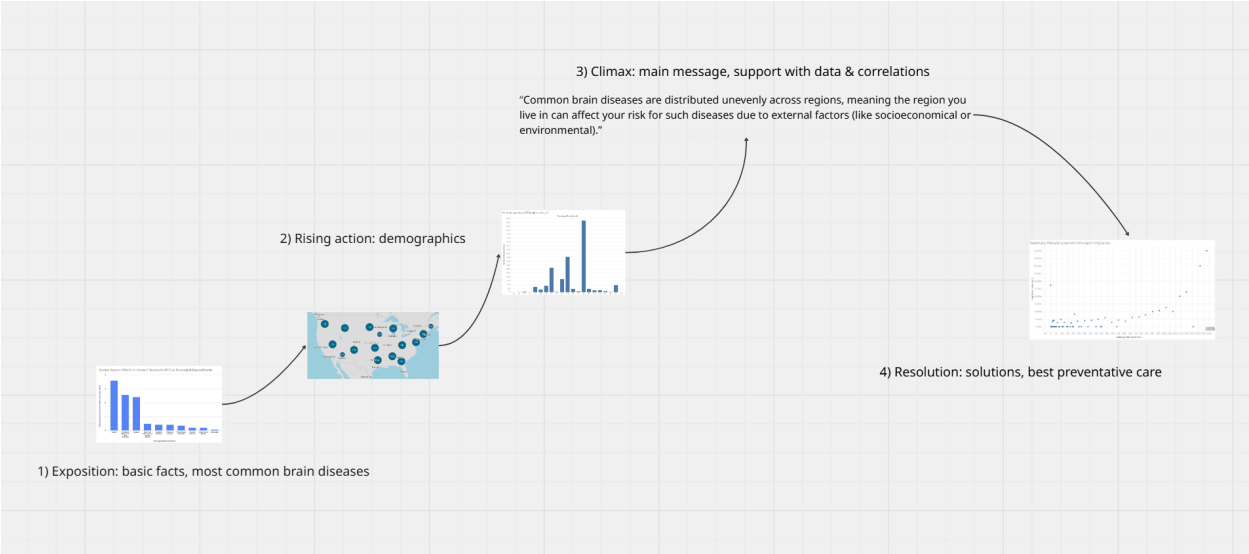
IV. In the process, you may want to ask questions that are not in your original list, and this is entirely fine, keep a record of them in your process book.

Storyboard

1. **Pick your main message:**
    1. **As a team, look at and discuss the list of insights you came up with. Pick one main insight that you think is most important or interesting to your audience. Formulate the insight as a message ("so what") in one sentence. You can choose multiple insights as long as they are coherent.**
    2. **Record your main message and the reasons that your group chose it in your process book.**
        1. Main message: Common brain diseases are distributed unevenly across regions, meaning the region you live in can affect your risk for such diseases due to external factors (like socioeconomical or environmental).
            a. Reasoning: Our group chose this message because it is the most relevant to our target audience: older individuals. Typically, the older generations want to understand their personal risk and typically assume that brain diseases primarily depend on aging. However, our data shows that a multitude of factors, such as geographical, socioeconomic, and environmental ones, play a critical role. By focusing on regional disparities, we can make the story relatable, engaging, and actionable. This insight ties together all of our datasets and provides a strong message for our story.
            b. Solution:
                i. Prevention: practice healthy life styles (healthy diets, regular physical exercise, social interactions, cognitive simulation).
                ii. Awareness: learn the early warning signs
                iii. Healthcare system: schedule screenings and discuss personal risk factors with doctors and specialists.

2. **Sketch your data storyboard. Following the four steps that you learned in the lecture. You will create a data storyboard to communicate your main message. You can use pen&paper, which is the preferred option. You can also use Miro with the following steps.**
    1. **You will work with your team to create the story arc. Please make sure that your story arc is lined up with the four elements of a data story (hook, rising insights, main message, and solution), similar to what we showed in the lecture.**
    2. **It is fine if you have multiple insights for each data story element, again as long as they are coherent.**
    3. **Take a screenshot of the final data arc (or a photograph of your papers) and put it in your process book.**

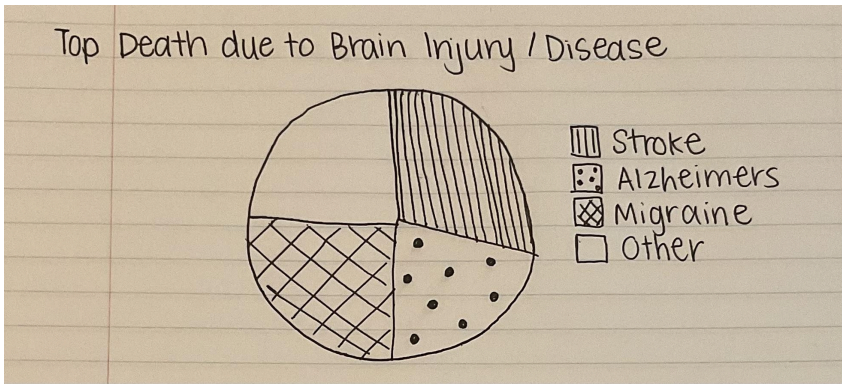
1. Main message: "Common brain diseases are distributed unevenly across regions, meaning the region you live in can affect your risk for such diseases due to external factors (like socioeconomical or environmental)."
2. Beginning: start with general visualizations and observations, such as "most common brain diseases by case number".
3. Rising insights: demographics; geographic distributions, race/ethnicity/gender differences, etc.
4. Empower audience: solutions; best preventative care - "You can reduce your risk for brain diseases through the following means."



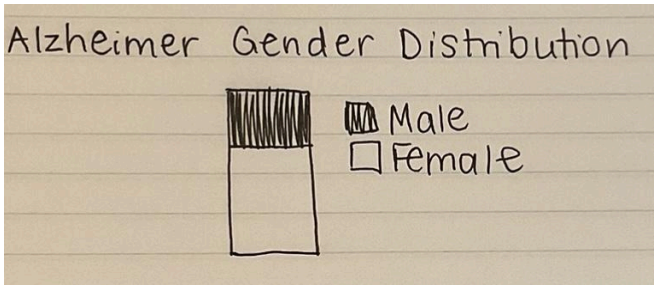
Sketch

Namrata's Sketches -

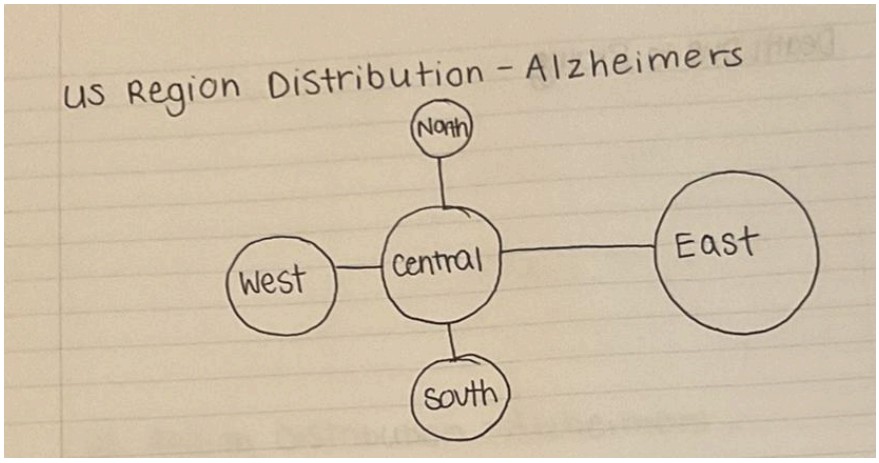
1. Message: This sketch is a pie graph that reflects the top deaths due to brain injury or disease. Through this pie graph, we are able to see the percent distribution of the injuries and disease that cause the top number of deaths related to the brain. Additionally, the difference in the labeling and patterns for each of the sections of the pie graph communicate the categorization of the different brain injuries and diseases.



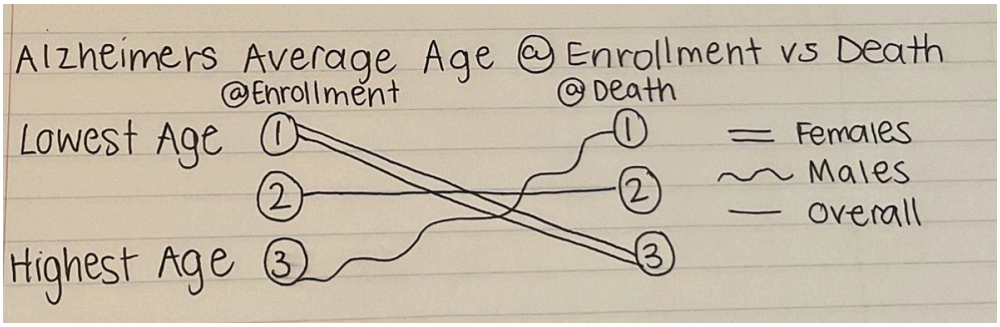
2. Message: The chart of the Alzheimer's Gender Distribution shows the distribution of Alzheimer's patients between females and males. As shown from the graph, Alzheimer's predominantly affects females. This shows the demographic effect as females more commonly are susceptible to having Alzheimer's.



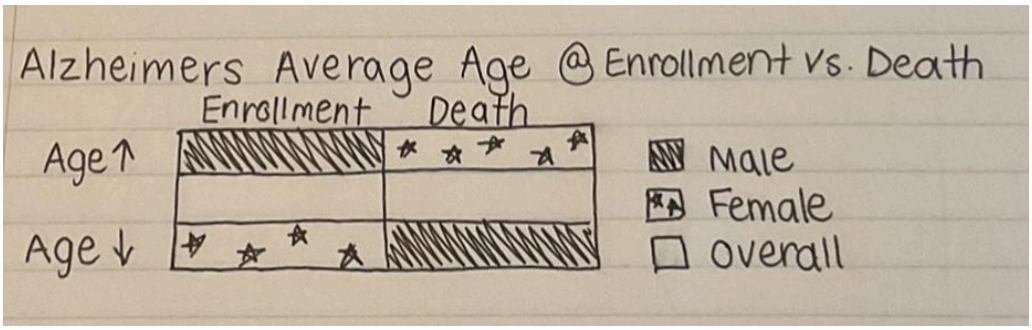
3. Message: The following sketch shows the distribution of Alzheimer's cases by size by region. This visually shows how Alzheimer's is more predominant in regions such as the East rather than regions with a less quantity of cases such as the North. The size of the East bubble is notably larger than the North bubble, meaning the East has a larger number of Alzheimer's patients than the North.



4. Message: The sketch below shows the average age at enrollment and death for females, males and overall. It shows the comparison of the age progressions between genders and the overall trend of the relative ages categorized by gender. We see that females initially have a lower age at enrollment but they have a higher age at death than the males.

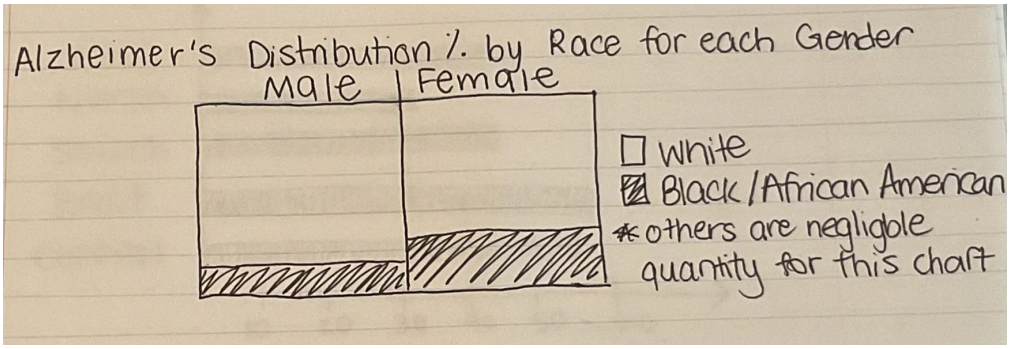


5. Message: The sketch below has a similar message to the previous sketch. We can overall understand the demographic impacts, specifically with gender and age. Females get Alzheimer's at a younger age than males but die at an overall later age. Males on the other hand get Alzheimer's at an older age and die at an earlier age compared to females.

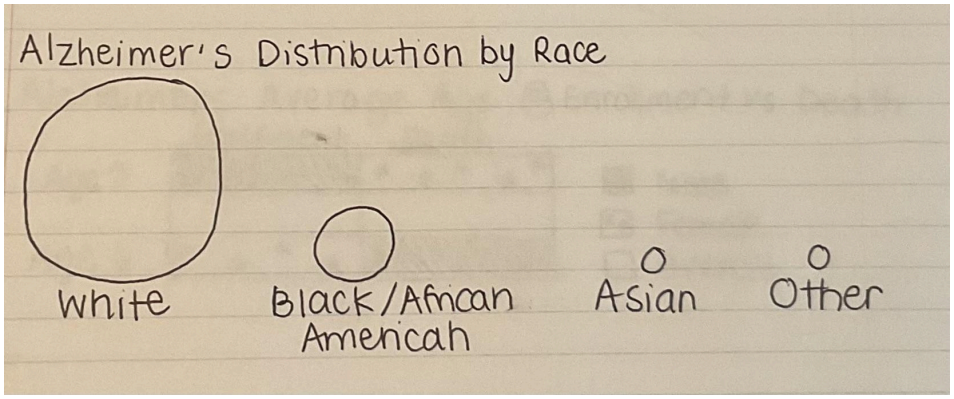


6. Message: The following chart shows Alzheimer's distribution % by Race for each gender. We understand demographic impacts on both gender and age combined. The White and

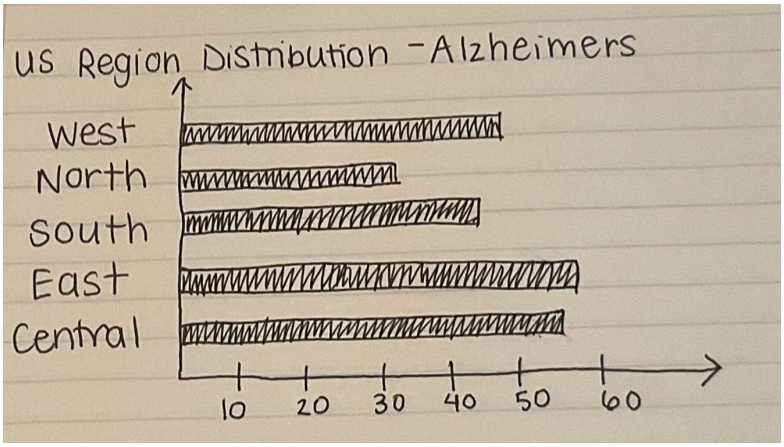
Black/African American races are the most notable for Alzheimer's with the White race having a very predominant percentage of cases. The difference in race % between gender is also comparable.



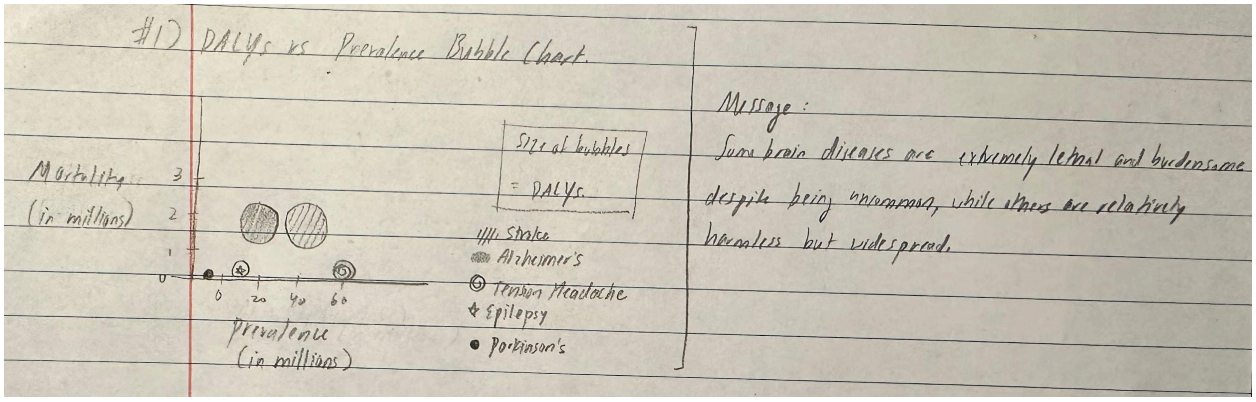
7. Message: The bubble graph below shows Alzheimer's Distribution by Race. The size of the bubbles for each race show the size of the cases of Alzheimer's that race has overall. The bubble for the White race is notably the largest with the Asian and Other races being almost negligible in size.

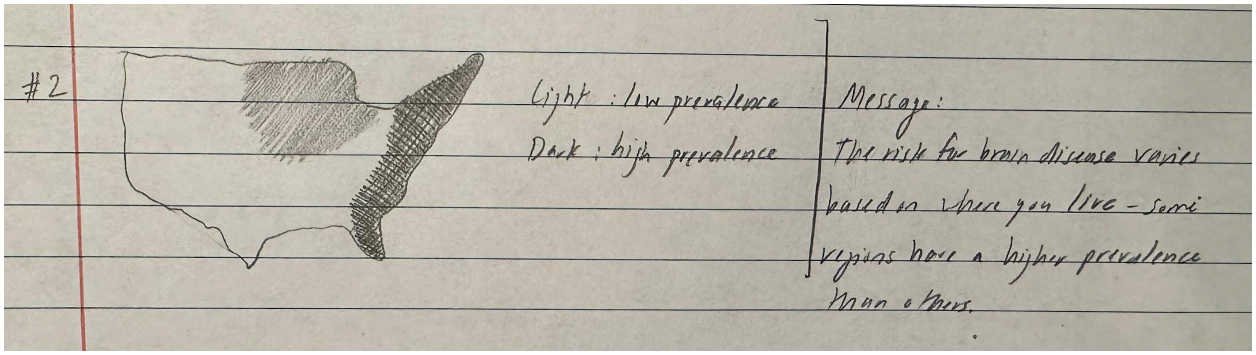


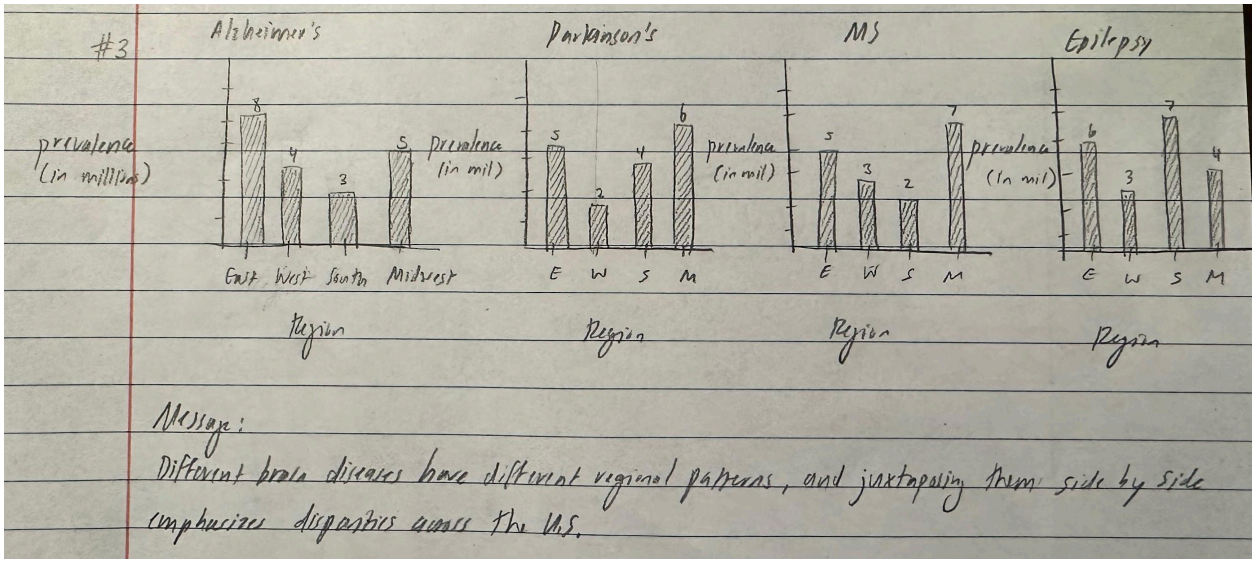
8. Message: The bar chart below shows the US Region Distribution for Alzheimer's. We can understand the geographic distribution of Alzheimer's patients within the regions in the US. The East has the largest quantity with the North having the least.

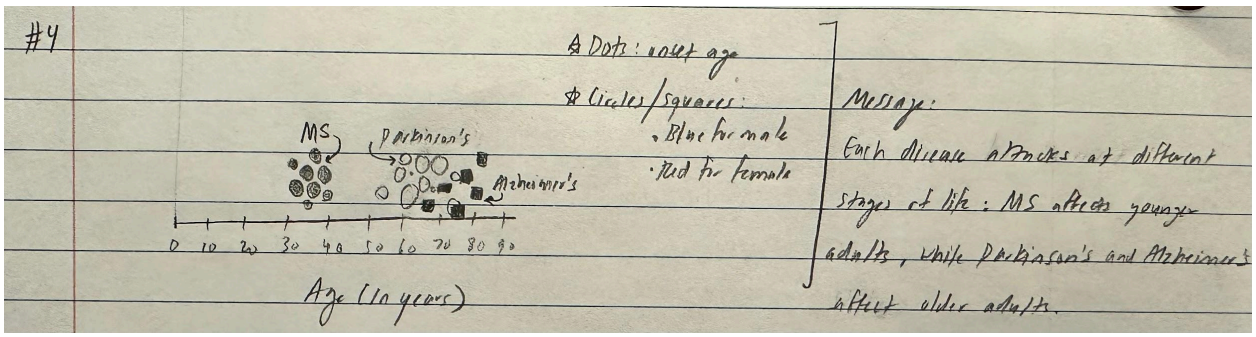


Anne's sketches:

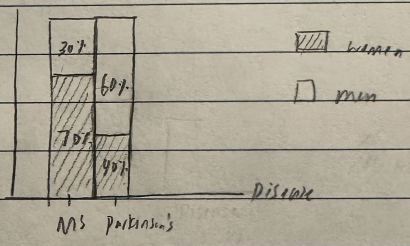








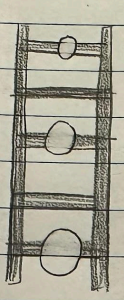
#5



Gender Distribution of MS vs Parkinson's

Message: MS disproportionally affects women, while parkinsons affects more men, showing clear gender differences in disease burden

#6

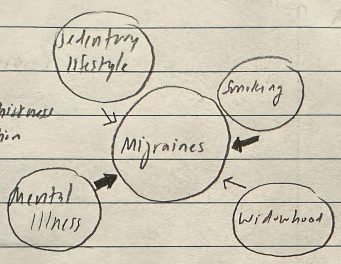


Size of dots: prevalence

Message: Lower income groups show higher disease prevalence, which suggests socioeconomic disparities in brain health.

Income vs. Disease Prevalence

#7



Message: Lifestyle and socioeconomic factors like smoking, a sedentary lifestyle, and mental health strongly increase migraine risk.

#8

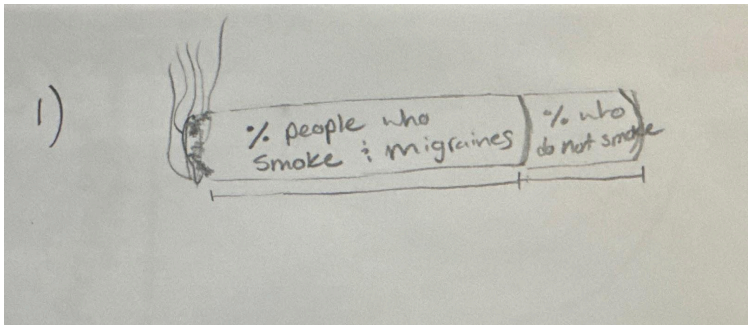
#1 Sketch	#2 Sketch
#5 Sketch	#4 Sketch

Brain Disease Burden Dashboard

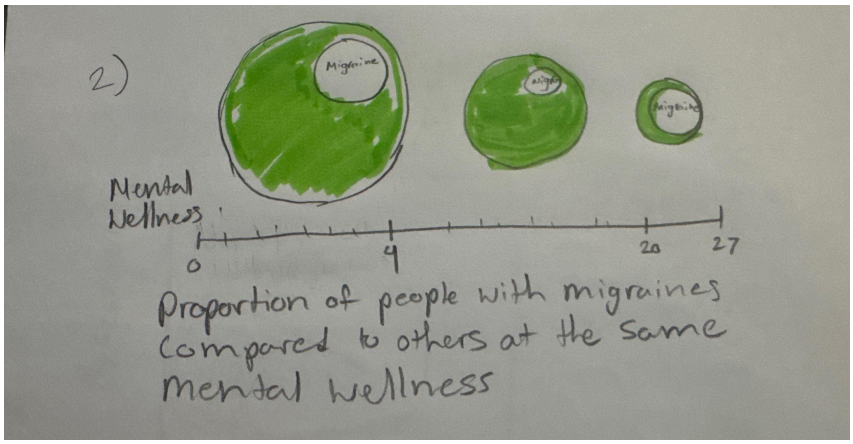
Message: A dashboard view combines mortality, geography, gender, and socioeconomic data to give a comprehensive view of brain disease disparities.

Kirito sketches:

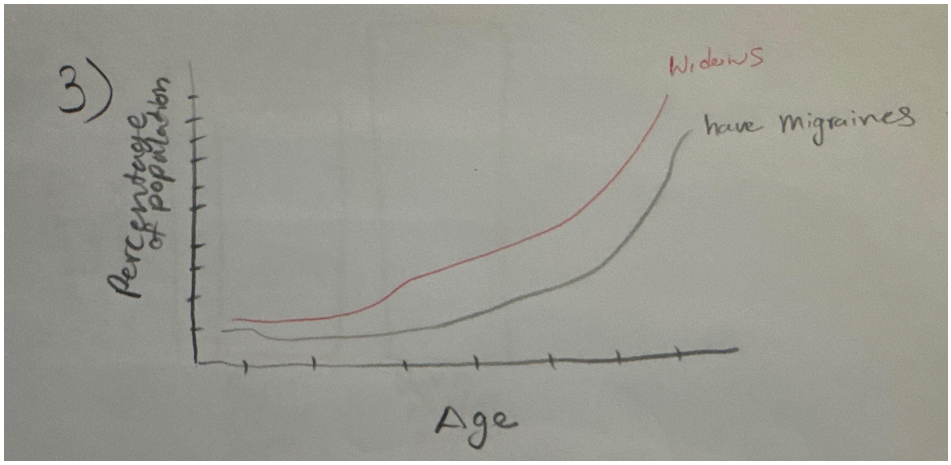
Of those who have migraines, many more of them smoke relative to the population as a whole



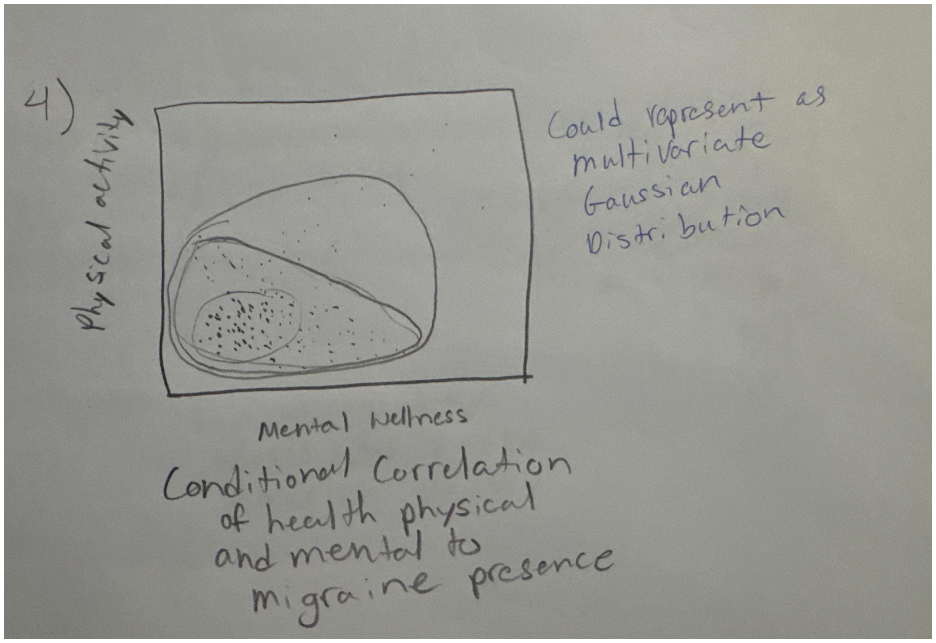
Similar to the smoking case, while considerably fewer people report significant mental unwellness, of those that do, a much larger ratio of them have migraines than those who reported moderately or very healthy



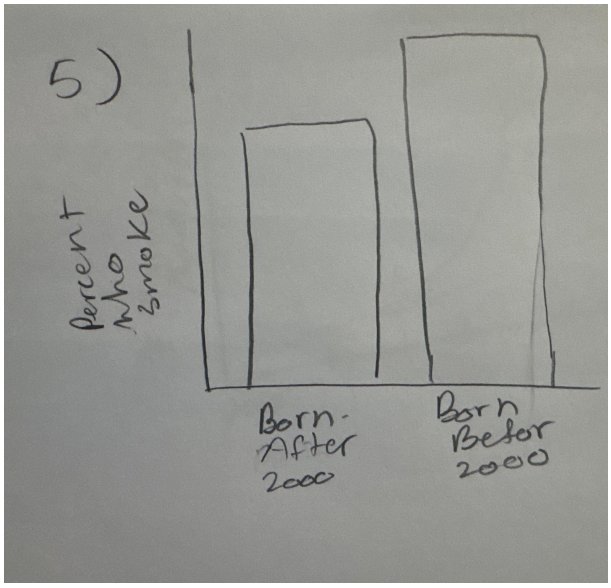
While more people who report migraines (relatively to their population size) are widows, that may just be because widows tend toward the older end of the age range since younger people are much less frequently widowed



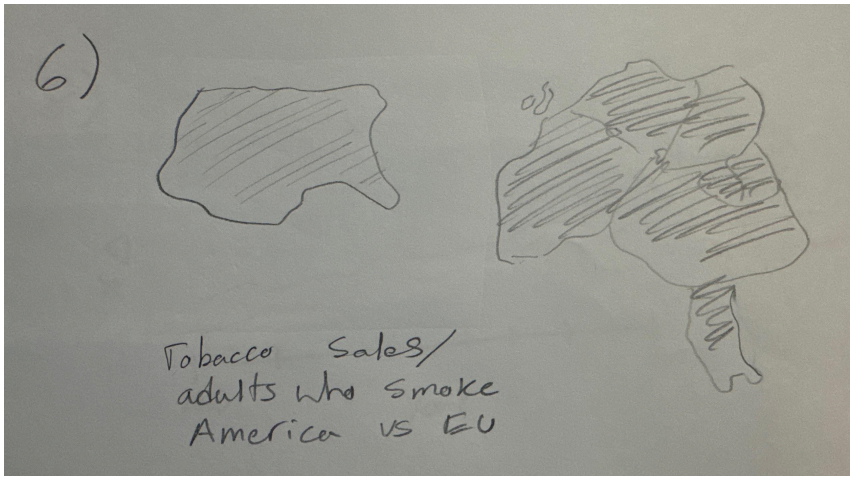
We can use multivariate distribution to make (maybe in 3D heat map) a Gaussian distribution to see how mental and physical health play into migraines. People who tend toward less healthy overall tend toward more migraines and we can find this conditional probability



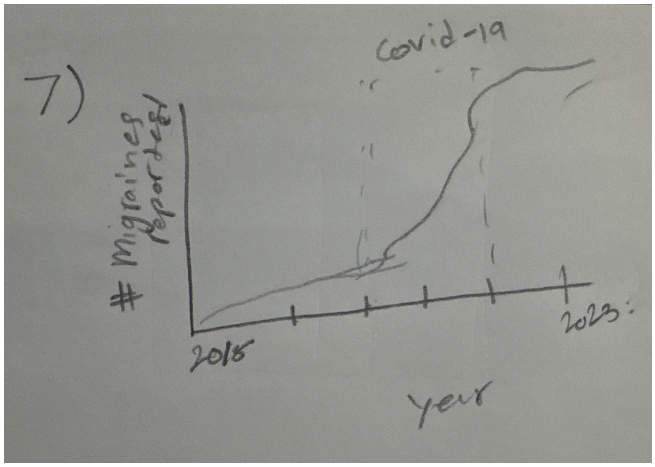
Smoking is much less prevalent among those born after the 90s and 2000s which contribute to a broad health benefit. This means healthier lungs and stronger immune systems that can fight stress and other migraine inducing events



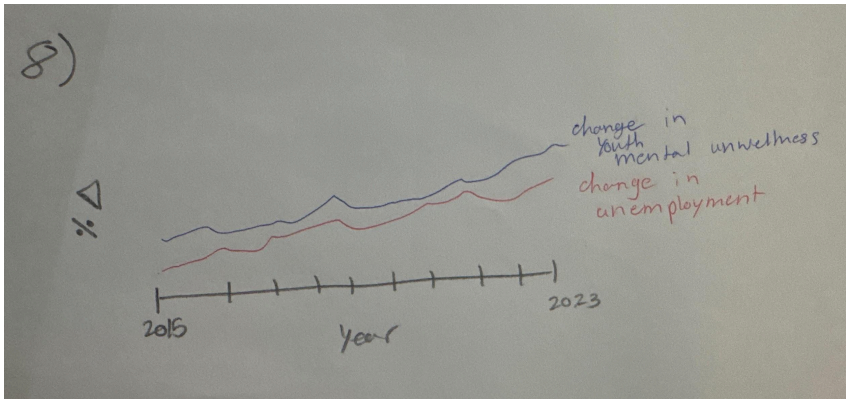
Does smoking really play a role when Europe smokes much more and (maybe) experiences equivalent levels of migraines?



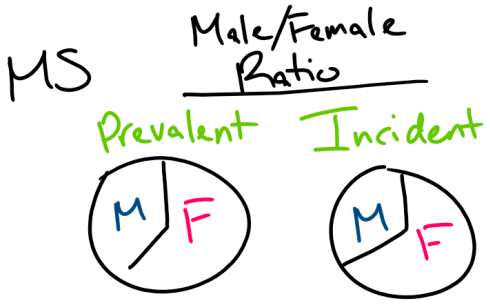
Covid had a large impact on a LOT of the data and people are still coming back down from those times. Adjustment was very hard and it has had lasting impacts. Youth may be stunted, leading to declining capability or health



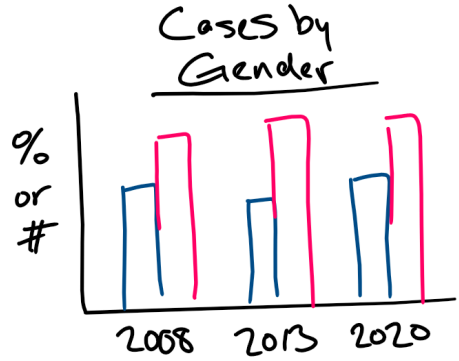
Youth who experience migraines may correlate with unemployment. As more people struggle to find their own, the increases in unemployment especially in youth and entry markets may cause an increase in migraines among those youth.



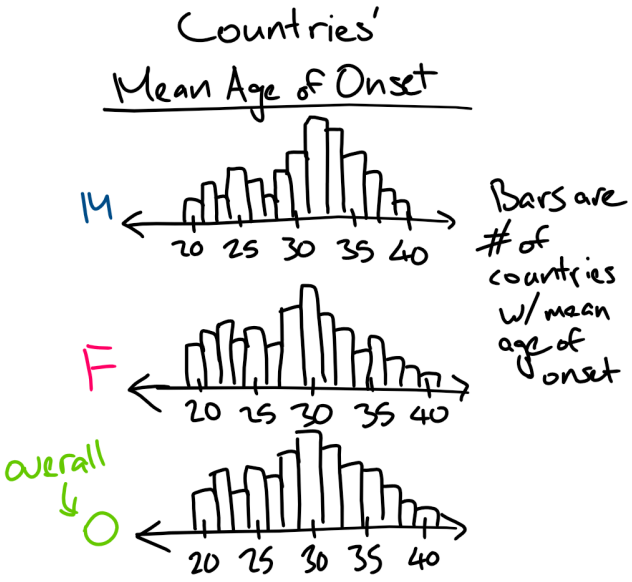
Aaron's sketches (multiple sclerosis):



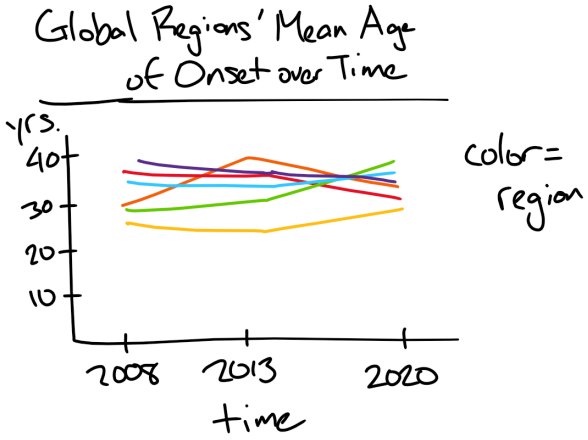
1. Message: In both prevalent and incident cases, the percentage of MS cases that are female is significantly greater than the percentage of MS cases that are male.



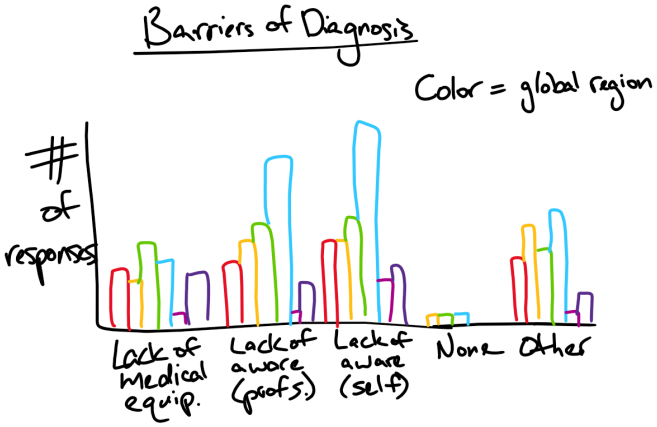
2. Message: The percentage of female cases of MS has been consistently greater than the percentage of male cases of MS over time/the ratio of female to male cases of MS has been consistently greater than 1.



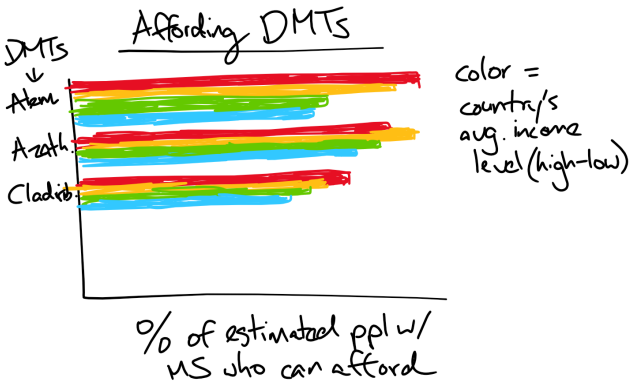
3. Message: The mean age of onset is consistently between 30-40 years of age on average, although it can start in as low as the early 20s or in as high as the early 40s.



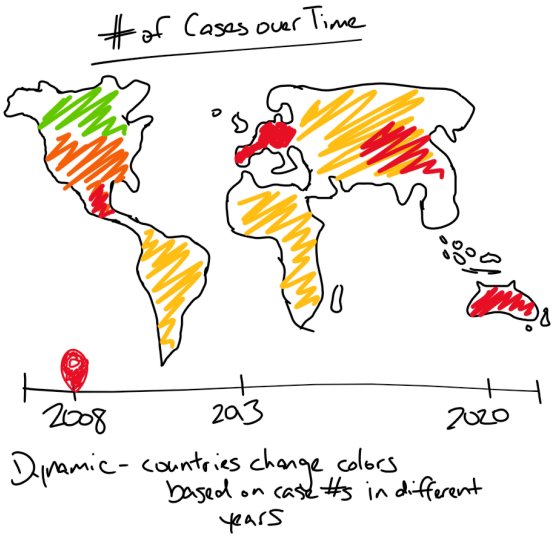
4. Message: The mean age of onset is consistently between 30-40 years of age on average across different regions of the world.



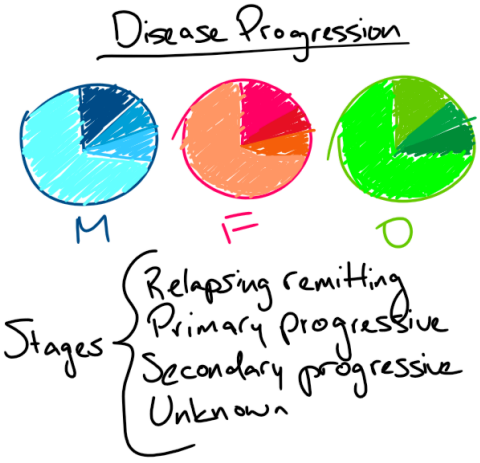
5. Message: There are various barriers to diagnosing MS early, some of which are more prominent than others depending on region, but the most common barrier to diagnosis is a lack of awareness of the symptoms of MS.



6. Message: Some DMTs (disease modifying therapies) are less affordable than others, which greatly affects the usage rate, depending on the level of income of the country.



7. Message: Some countries and/or regions of the world have higher prevalence (per 100,000) than others, indicating that there could be a higher risk for MS when living in those countries. (This may be influenced by other factors, such as population density and unrecorded data.)



8. Message: Most cases of multiple sclerosis fall into the relapsing-remitting category over the primary progressive and secondary progressive categories. On average, female cases tend to be of the relapsing-remitting type more commonly than male cases; conversely, male cases tend to be of either progressive type more than female cases.

Decide

We numbered the sketches with IDs from 1 to 32 in the order the sketches are given above; thus, Namrata's sketches are numbered 1-8, Anne's are numbered 9-16, Kirito's are numbered 17-24, and Aaron's are numbered 25-32.

Since each person gathered data on different diseases, most of our sketched visualizations pertained to the one disease that each person worked on. Therefore, we decided to name the message IDs by their topic and parameter(s) analyzed.

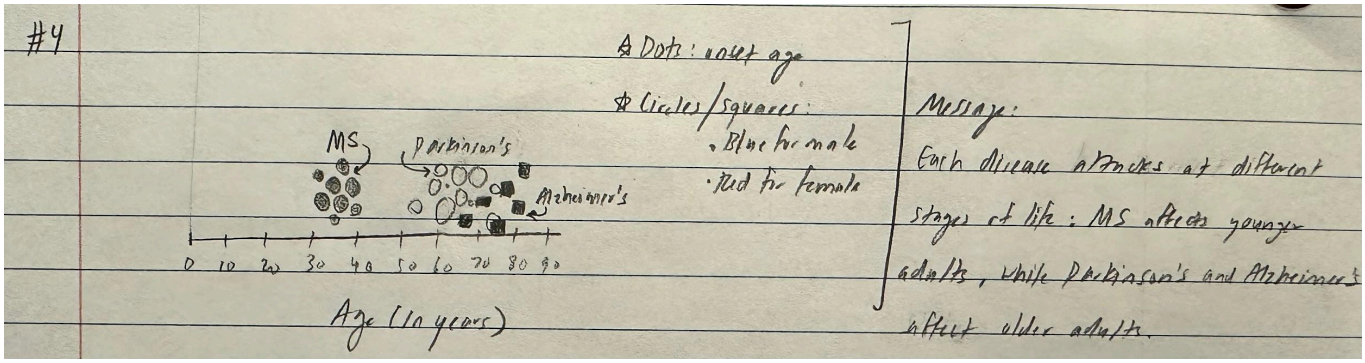
AFFINITY DIAGRAMMING TABLE:

ID	Message ID	Author	Voters
1	Brain	Namrata	
2	Alz	Namrata	
3	Alz	Namrata	Aaron, Namrata, Kirito
4	Alz, Age	Namrata	Namrata
5	Alz, Age	Namrata	
6	Alz, Demos	Namrata	Kirito
7	Alz, Demos	Namrata	
8	Alz, Demos	Namrata	
9	Brain	Anne	
10	Brain, Demos	Anne	Anne
11	Brain, Demos	Anne	
12	Brain, Age	Anne	Anne, Aaron
13	MS, Park	Anne	
14	Demos	Anne	
15	Lifestyle, Mig	Anne	
16	Demos	Anne	
17	Smoke, Mig	Kirito	Anne
18	Mental, Mig	Kirito	Anne, Aaron, Kirito
19	Demos, Mig	Kirito	Namrata
20	Mental, Mig, Lifestyle	Kirito	Aaron, Kirito
21	Demos, Mig	Kirito	
22	Demos, Smoke	Kirito	
23	Demos, Mig	Kirito	
24	Mental	Kirito	Namrata

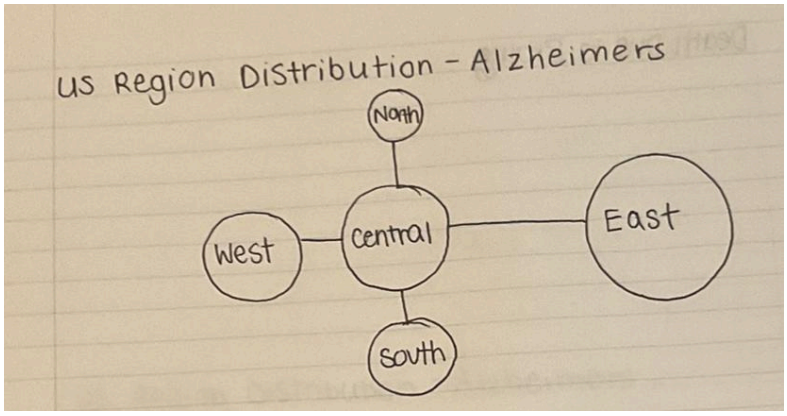
25	MS, Demos	Aaron	
26	MS, Demos	Aaron	
27	MS	Aaron	
28	MS	Aaron	
29	MS	Aaron	Kirito
30	MS, Demos	Aaron	
31	MS, Demos	Aaron	Anne, Aaron, Namrata
32	MS	Aaron	

CHOSEN SKETCHES:

12.

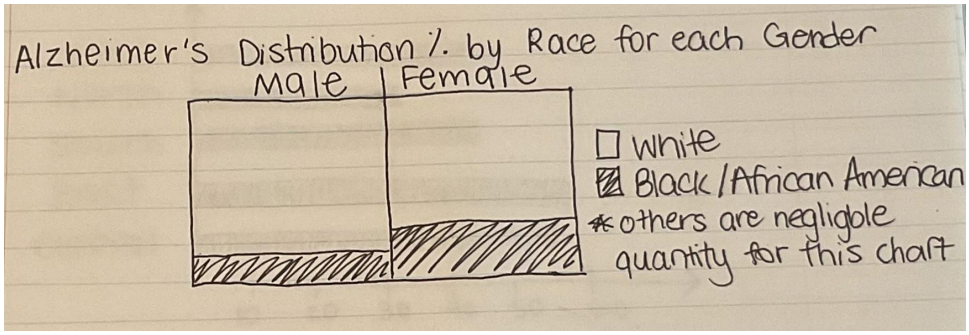


3. Message: The following sketch shows the distribution of Alzheimer's cases by size by region. This visually shows how Alzheimer's is more predominant in regions such as the East rather than regions with a less quantity of cases such as the North. The size of the East bubble is notably larger than the North bubble, meaning the East has a larger number of Alzheimer's patients than the North.

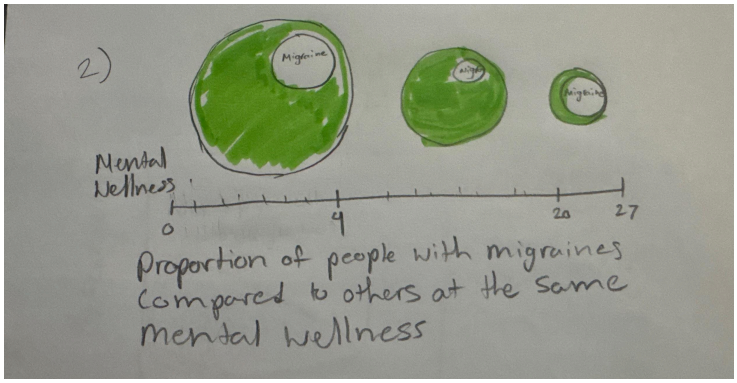


6. Message: The following chart shows Alzheimer's distribution % by Race for each gender. We understand demographic impacts on both gender and age combined. The White and

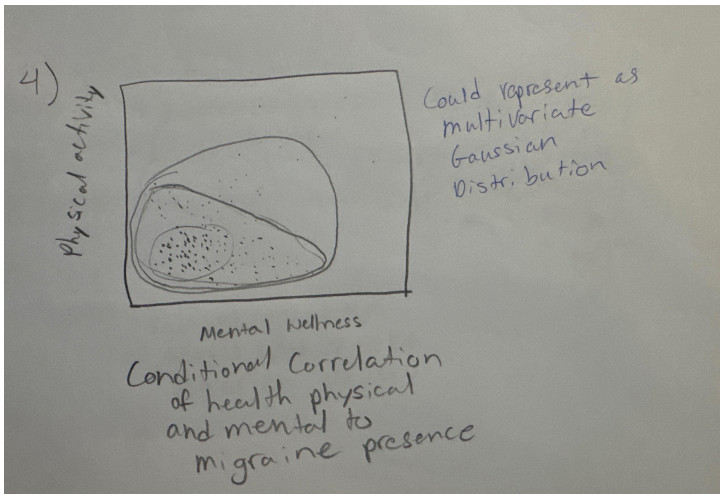
Black/African American races are the most notable for Alzheimer's with the White race having a very predominant percentage of cases. The difference in race % between gender is also comparable.

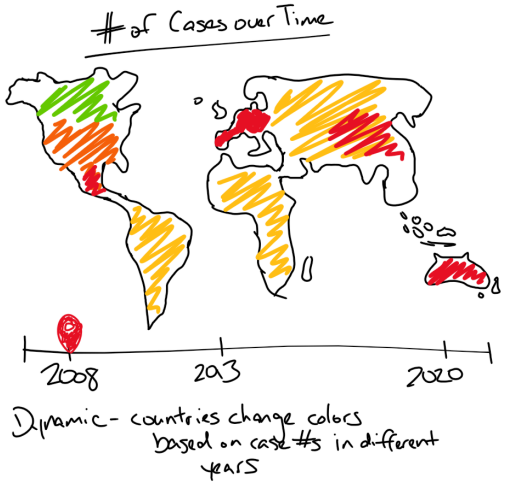


18. Similar to the smoking case, while considerably fewer people report significant mental unwellness, of those that do, a much larger ratio of them have migraines than those who reported moderately or very healthy



20. We can use multivariate distribution to make (maybe in 3D heat map) a Gaussian distribution to see how mental and physical health play into migraines. People who tend toward less healthy overall tend toward more migraines and we can find this conditional probability





31.

Message: Some countries and/or regions of the world have higher prevalence (per 100,000) than others, indicating that there could be a higher risk for MS when living in those countries. (This may be influenced by other factors, such as population density and unrecorded data.)

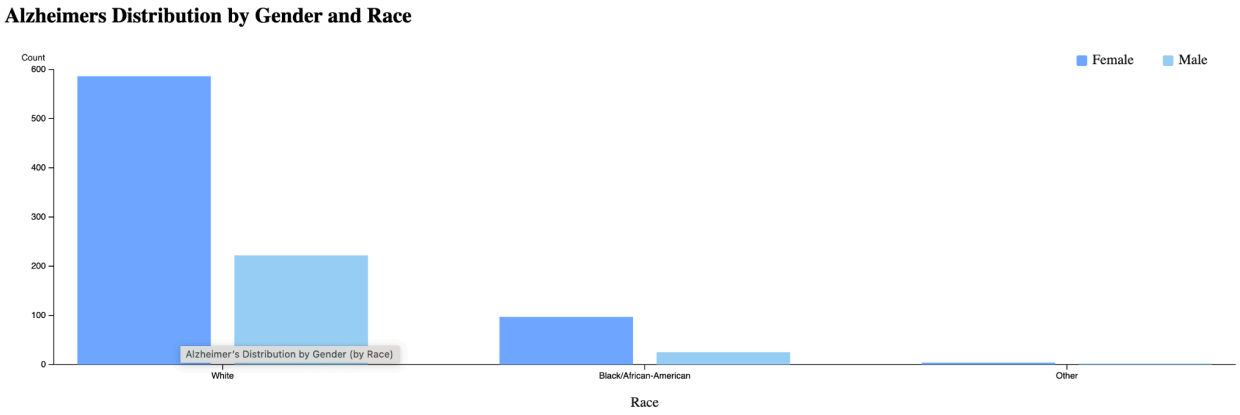
Summary/Rationale:

We ended up selecting six sketches to best represent our central message: brain diseases are shaped by demographic, behavioral, and environmental factors. We voted for sketches that build a comprehensive narrative, from vulnerability to disease-specific data. Sketch ID #12 introduces the overall idea of aging as a key risk factor. Sketch IDs #3 and #6 were chosen due to their clarity in showing demographic and regional disparities. Sketch ID #18 and #20 creatively build on the narrative by connecting mental health and lifestyle, providing prevention methods. Finally, Sketch ID #31 is a great global demographic comparison that reinforces the message of unequal disease distribution. All together, these visualizations are the best balance of demographic, behavioral, and medical perspectives. They will allow the audience to understand not just our data, but also the broader implications for their health and environment. This specific order was chosen to best weave our narrative together.

Prototype, Phase 1

Scraped & cleaned data link (multiple sheets):
https://docs.google.com/spreadsheets/d/1_dH_G4RwklM2oY6ZhJl0r0ICCFE_gtgAavBelNbiByQ/edit?usp=sharing

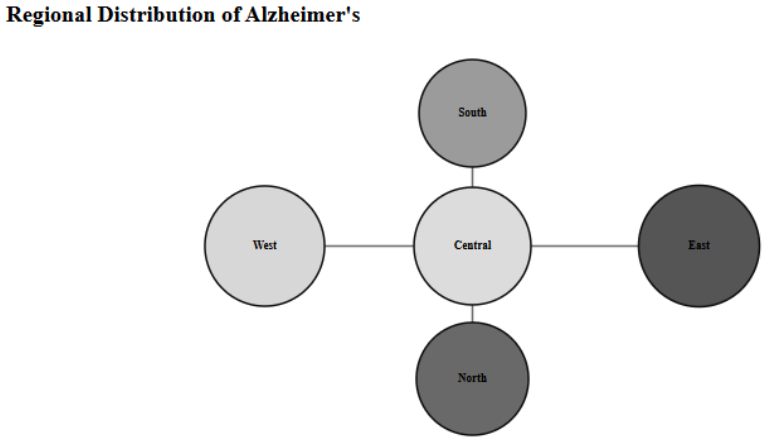
Visualization 1 implementation:



This bar chart shows the distribution of Alzheimers among Men and Women.

White
Female: 586

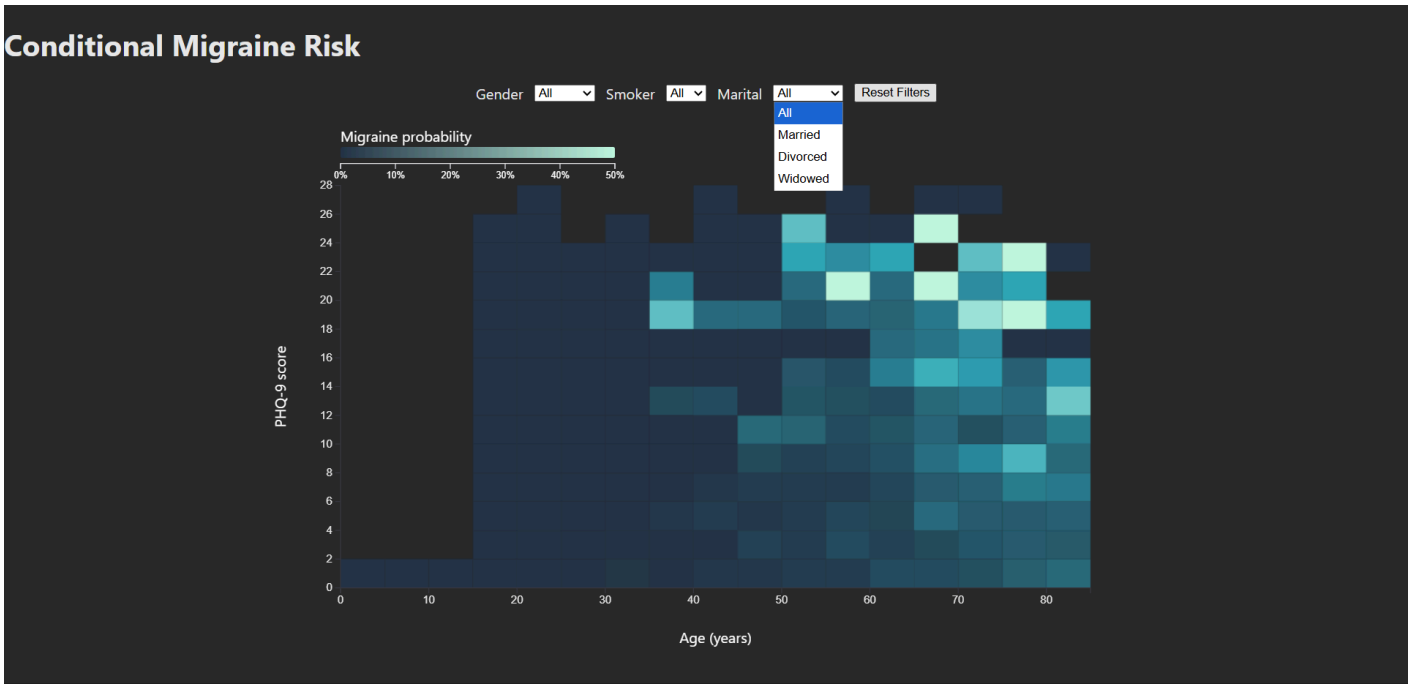
Visualization 2 implementation (partial):

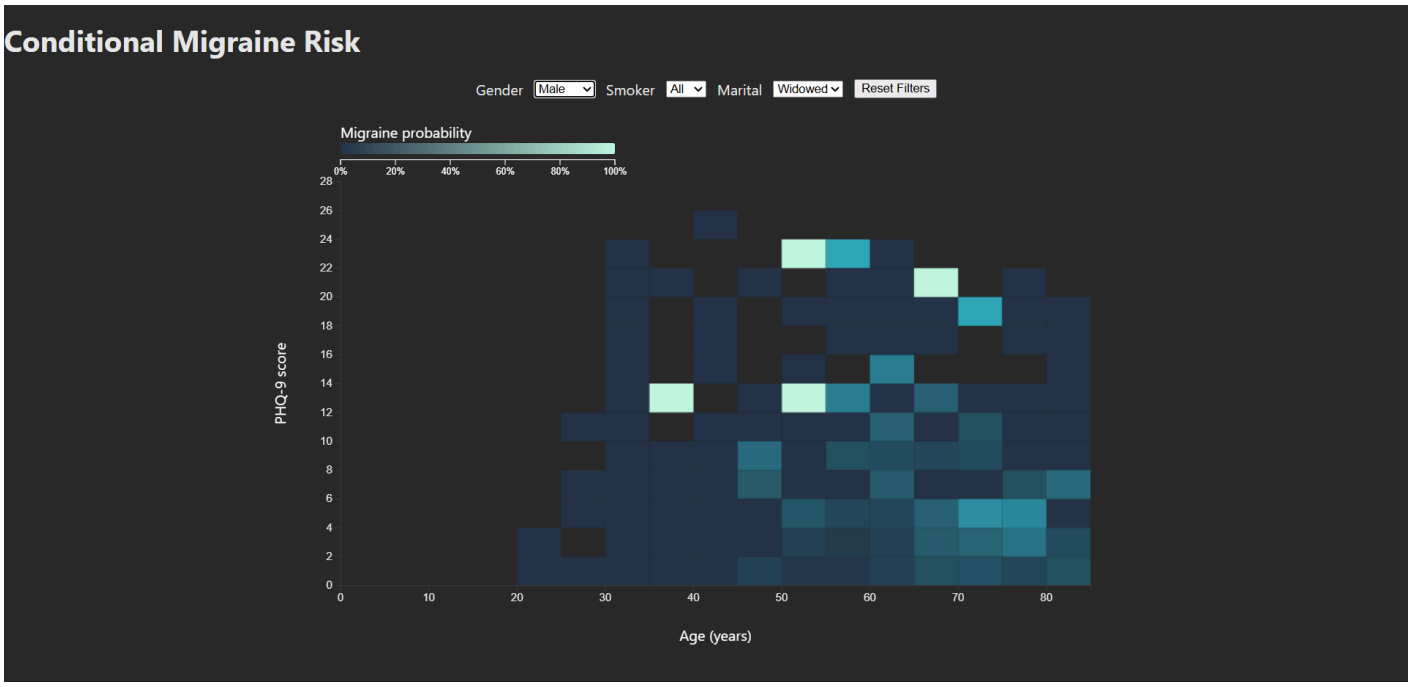


The chart shows the distribution of Alzheimers among US regions. Bubble area shows Alzheimer cases in thousands.

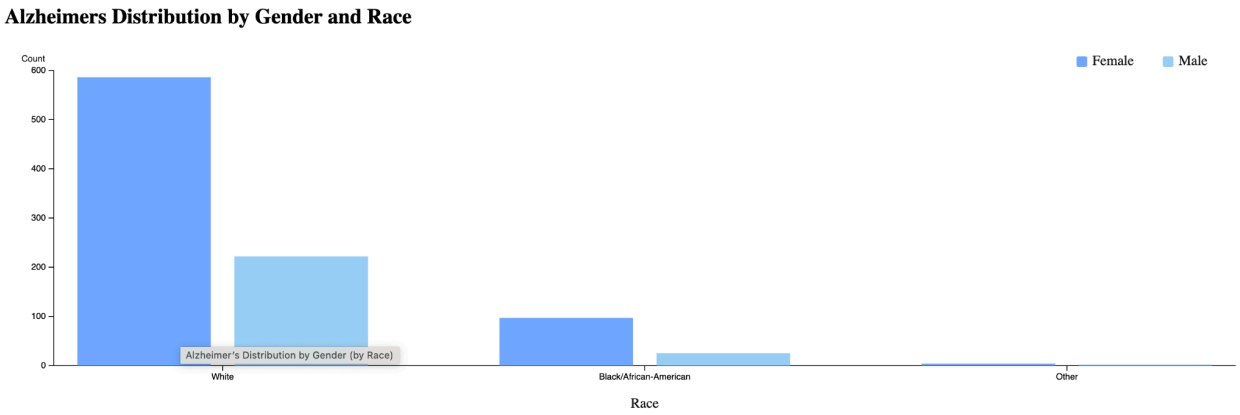
Prototype, Phase 2

Innovative visualization - Kirito



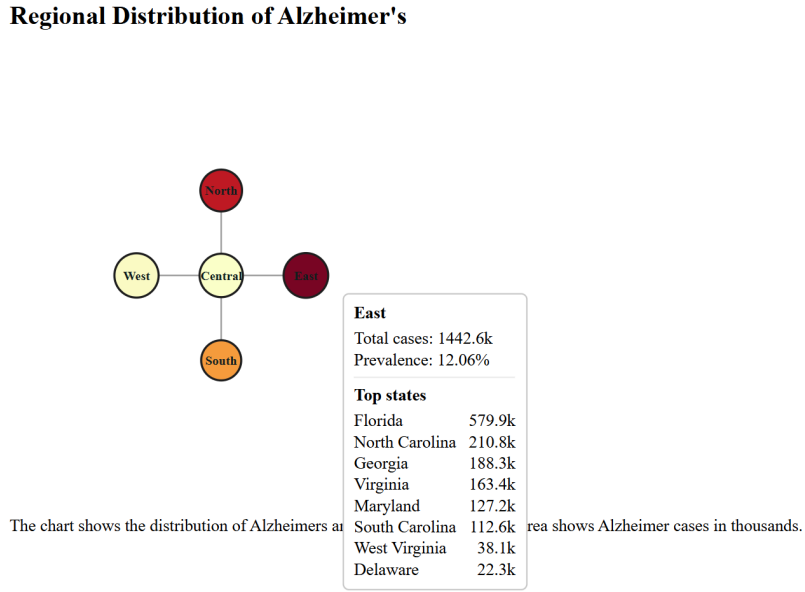


Alzheimer’s Visualization - Namrata



Some of the main interactions of this Alzheimer’s visualization is a tooltip that focuses on the exact bar the user would like to understand better. When the tooltip is over the bar, it shows the correlated data values under the graph for the user to better read the data and associate exact values and headers for what the bar is representing.

Alzheimer’s by Region Visualization - Anne

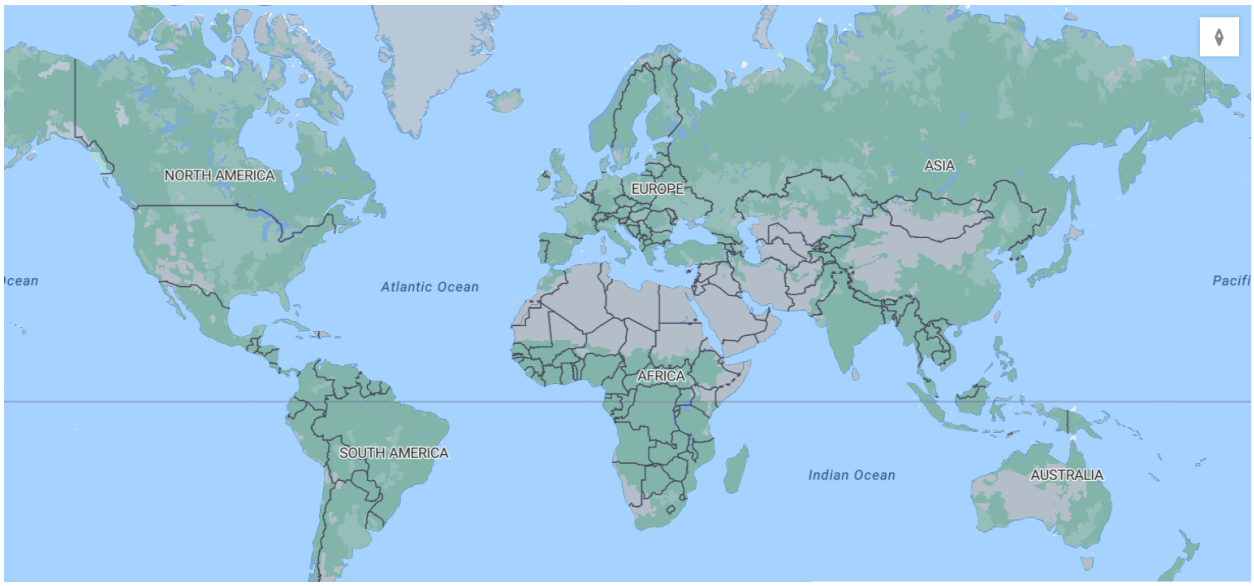


This visualization focuses on seeing the overall prominence of Alzheimer’s in different regions of the United States. By hovering over each circle, the overall statistics, individual states, and each of their numbers populate. It allows the user to elaborate on the abstract overview graph by zooming into the specific details of each region and their specific numbers. In addition, the colors show which area has more or less cases than the others.

MS by Country - Aaron

[Map of the world - likely 2D, Mercator map]

Example below:

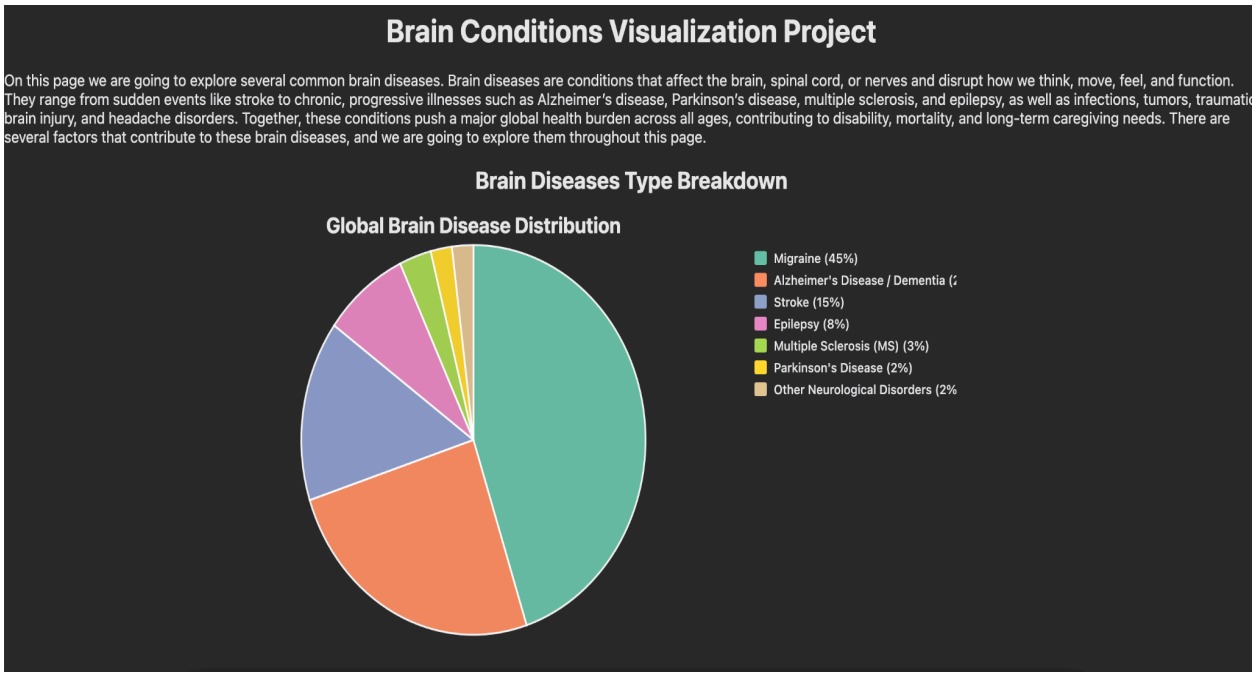


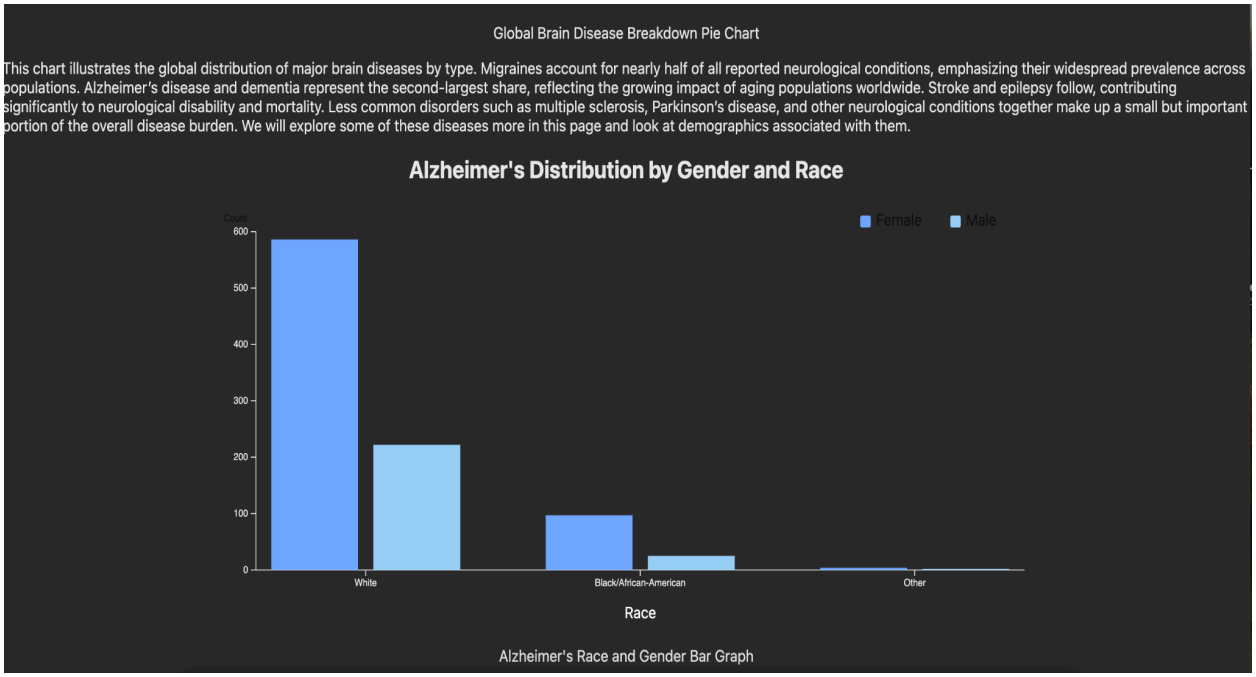
(Sourced from preview of https://github.com/johan/world.geo.json/blob/master/countries.geo.json - will probably use a .json file like this one)

This visualization, like the Alzheimer's by Region visualization, can show different statistics of multiple sclerosis by location; however, this visualization has a global scope, so the locations are categorized by countries and regions of the world rather than states and regions of the United States. The countries' data is represented through a color scale.

- The user can hover over certain countries (not all countries' data is available, which is indicated by color being "blank"/"transparent") to reveal a tooltip that shows the corresponding values. (A secondary layer may potentially be added to the map so that the global map only shows regions, and regions can be clicked on to zoom into them and show individual countries—with zooming back out available as a button to the side.)
- With datasets from multiple years (all data comes from The Atlas of MS), a new function is available: the user can click on icons on an axis below (?) the map to "move through time" from 2008 to 2013 to 2020. This allows the user to see how MS trends have changed over time.

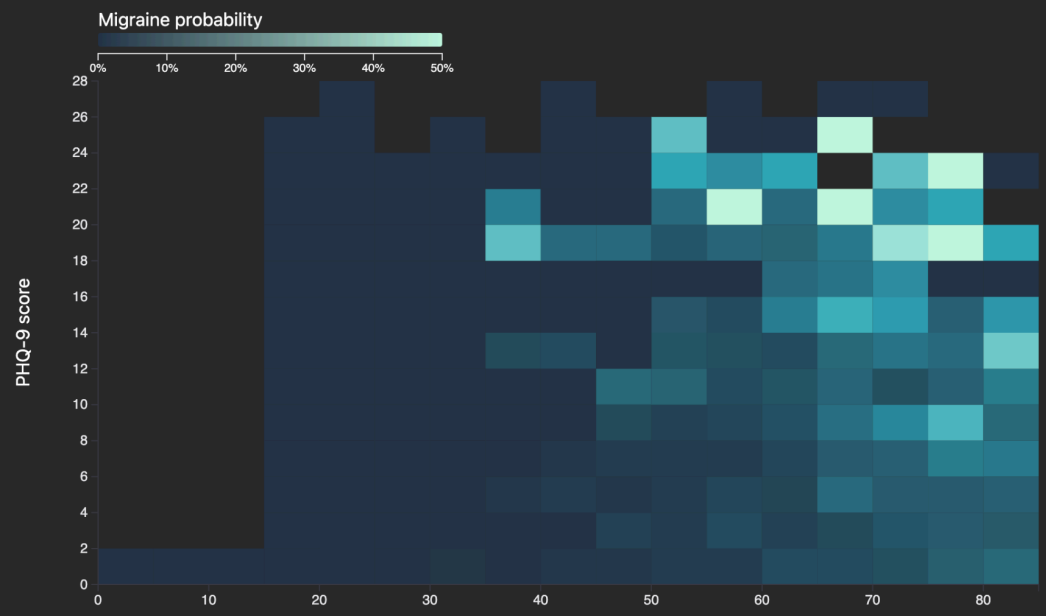
Prototype, Phase 3





Migraines Heatmap

Gender All Smoker All Marital All Reset Filters



Regional Distribution of Alzheimer's

The chart shows the distribution of Alzheimer's among US regions. Bubble area shows Alzheimer cases in thousands.

Placeholder text transitioning from US view of Alzheimer's to MS

MS Global Distribution

Placeholder text of MS

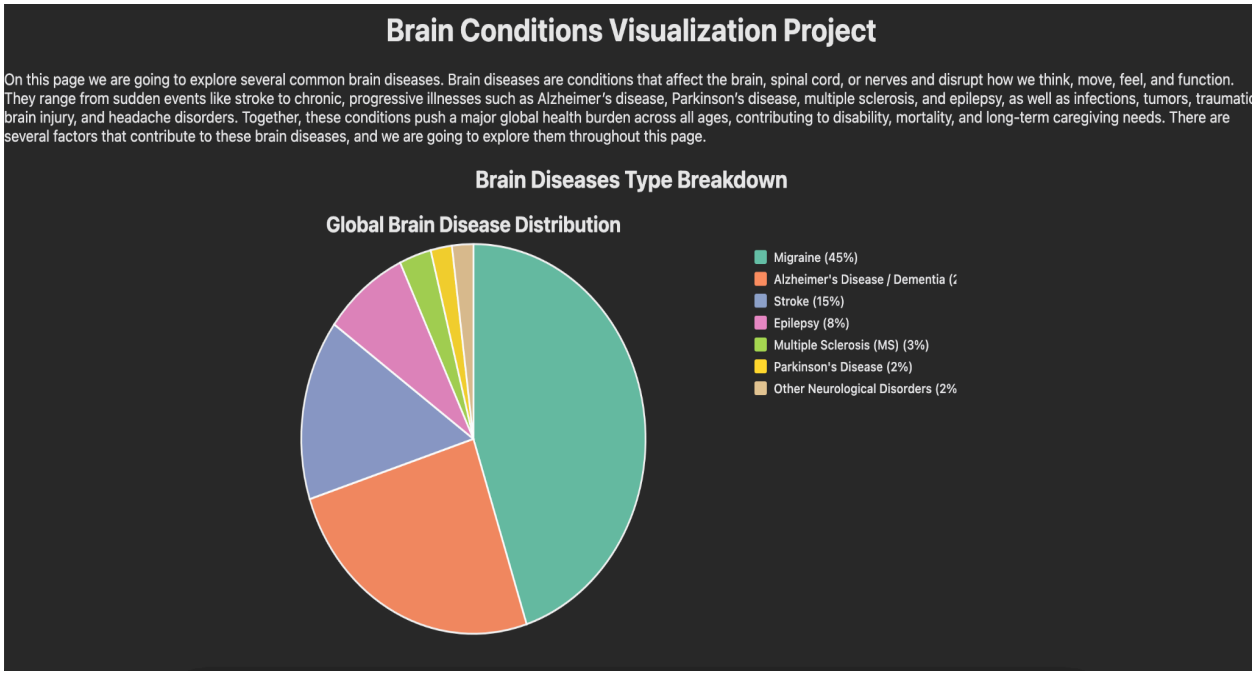
Placeholder text transitioning from MS to Migraines

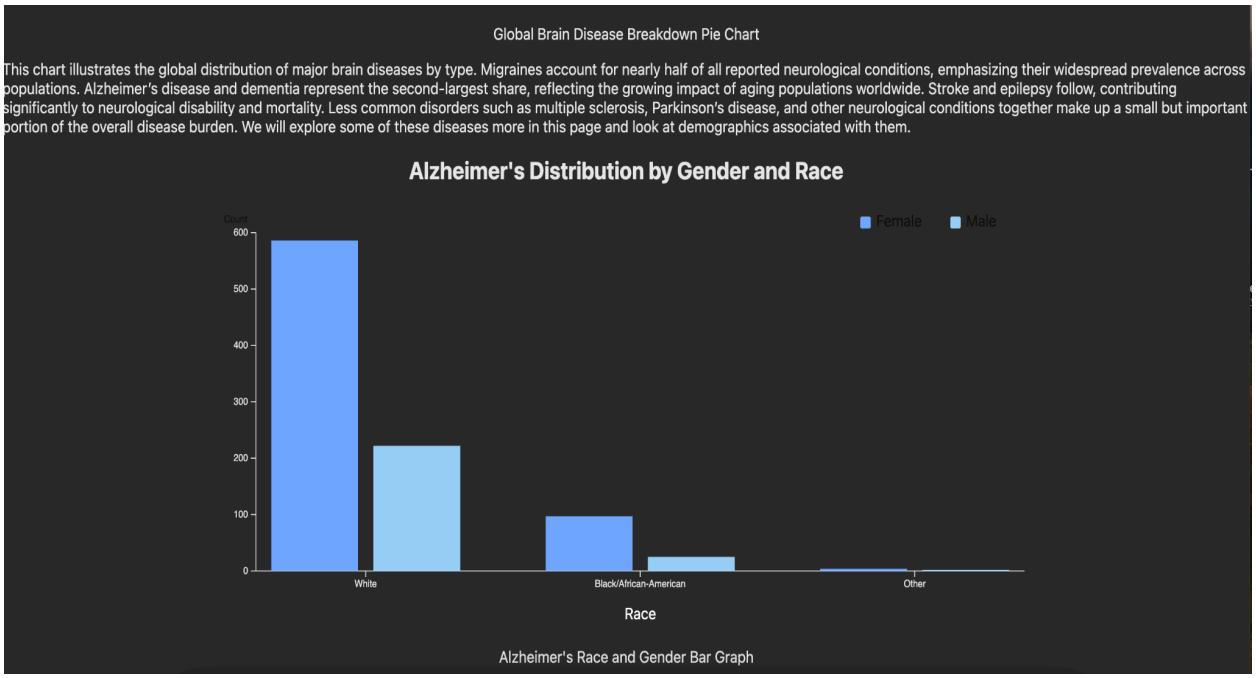
Placeholder text of migraines

Preventing Brain Diseases

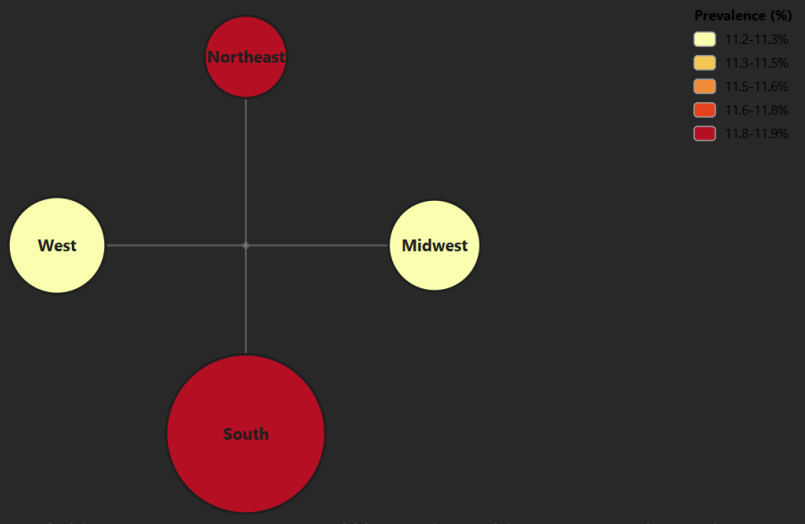
Steps you can take to prevent brain diseases: [insert cool things]

Prototype, Phase 4





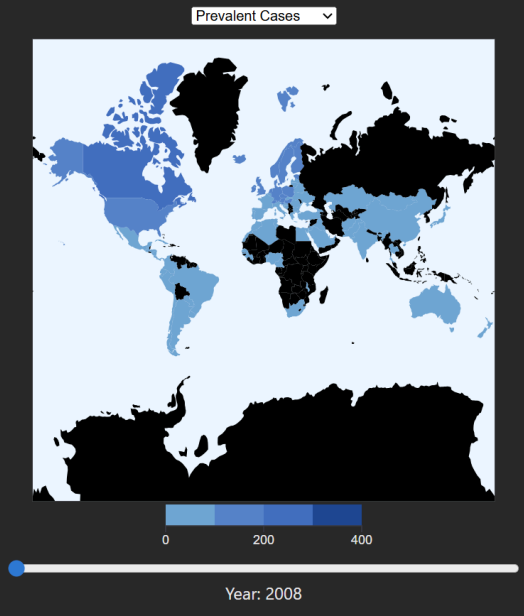
Regional Distribution of Alzheimer's



The chart shows the distribution of Alzheimer's among US regions. Bubble area shows Alzheimer cases in thousands.

This visualization highlights the regional distribution of Alzheimer's disease across the United States. The South shows the highest prevalence and number of cases, represented by the largest and darkest red bubble, suggesting both a higher rate and greater population impact. The Northeast also has relatively elevated prevalence, while the Midwest and West show lower rates, indicated by lighter yellow bubbles. The color gradient from yellow to red helps convey variations in Alzheimer's prevalence percentages among the regions. Overall, the chart emphasizes how geographic and demographic factors may influence Alzheimer's distribution patterns across the country.

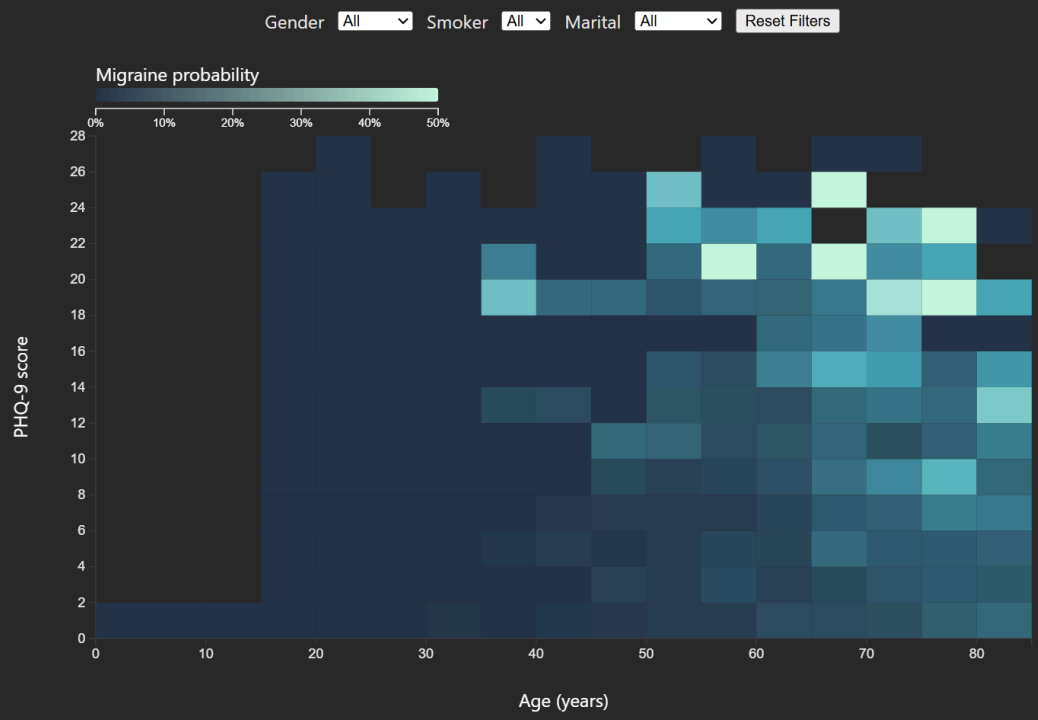
MS Global Distribution



This map shows the MS distribution around the world

This world map illustrates the global distribution of Multiple Sclerosis (MS), with color intensity representing varying prevalence rates across regions. Higher rates of MS are concentrated in North America, Northern Europe, and Australia, indicated by the darker shades of blue. Lower prevalence is seen in regions closer to the equator and in developing countries, such as parts of Africa, South America, and Southeast Asia. This geographical pattern suggests a potential link between latitude, genetics, and environmental factors in influencing MS risk. Overall, the map provides a clear visual comparison of how MS affects populations unevenly across different parts of the world.

Migraines Heatmap



For our last look, we invite you to explore how different lifestyle and demographic factors relate to the number of people experiencing migraines. The data covers U.S. citizens from 2017 to 2023 and includes information about marital status and smoking habits. The visualization compares these factors against age and mental health scores, helping reveal where migraine risk might be highest across different groups. The PHQ-9 (Patient Health Questionnaire-9) is a short, 9-question survey used to measure depression severity. Each question corresponds to one of the DSM criteria for major depressive disorder and asks how often a person has experienced certain symptoms over the past two weeks (like little interest or pleasure in doing things, trouble sleeping, or feeling tired). Each item is scored from 0 to 3 for "not at all," "several days," "more than half the days," or "nearly every day". Giving a total score from 0 to 27. 0–4: minimal or no depression 5–9: mild 10–14: moderate 15–19: moderately severe 20–27: severe

Preventing Brain Diseases

Based on the charts in our project, it's clear that demographic factors play a major role in shaping the prevalence of brain conditions. Regional and global patterns reveal that factors such as age distribution, lifestyle, genetics, and healthcare access significantly influence disease rates like Alzheimer's and Multiple Sclerosis. For instance, older populations in the South and Northeast U.S. show higher Alzheimer's rates, while MS prevalence is greater in higher-latitude countries with colder climates. These differences highlight how environmental exposure, diet, and socioeconomic conditions intersect with biological risk. Overall, understanding these demographic effects is essential for developing targeted prevention, awareness, and healthcare strategies that address the unique needs of each population.

Evaluation

Based on the results of your ‘think aloud’ study, what would you improve in your data story?

- Based on the results of our ‘think aloud’ study and the feedback from our two testers, our team is going to focus on visual improvements including text colors, graph colors, and font/sizing. Specifically, we need to fix various font colors so they are more visible (such as the legends).

Are there any additional insights and visualizations you would use? Would you amplify or change your message? Did your narrative work? Did the tester get your takeaways?

- There are not any additional visualizations we are planning to use for our data story; however, we are working on improving the final insight to represent the story better and provide a conclusion for our topic. We want to better amplify and convey our message to the audience. Our narrative for the introduction and our visualizations flowed well together; however, it was the conclusion that did not tie the topic and overall insight together well. The tester understood the insights from the visualizations, but felt that the conclusion could be better as they did not understand the concluding takeaway very easily.

Decide as a team which of these improvements you will implement and write down your decisions and why you made them in your process book as a numbered list.

Implement the intended changes and check them off your list (e.g., adding “done”). You can distribute the tasks among your team members. If you are unable to implement specific changes, please explain why and describe the expected results in your process book.

1. Fix the overall color scheme and make all of the text visible to make the website more visually appealing and aesthetically coherent. (done)
2. Change the color of the map viz so that the ocean is blue and not the land to make the map less confusing. (done)
3. Unify the color scheme to be matching (red). (done)
4. Get the heatmap to have a higher resolution because blocks of 5 years were less interesting (done)
5. Add the “Unmarried” option to the heatmap viz to have complete option set (done)
6. Add cursor effects and scroll effects to make the page more dynamic (skip - we need to fix meaningful errors before adding aesthetics)
7. Create more backgrounds and visually interesting features to make the website prettier (done)
8. Add more animations to the visualizations to increase cohesion, visual intrigue, and convey what is changing (done)
9. Add visualizations for more diseases like stroke or epilepsy to cover more ground (skip - too large a task with the time constraints, focused on improving the current visualizations instead)
10. Use buttons as opposed to a slider for the MS map viz to remove the confusion about continuity (we only have data sets from 3 separate years as opposed to multiple consecutive years) (done)
11. Have a more cohesive story put together through the conclusion since threads did not tie together well (done)
12. Have a conclusion (done)
13. Make the conclusion interesting (maybe a new viz or something else) since it was lackluster (done)

14. Have a more concrete “linked” section as it was not clear where that was (done)
15. Change the font colors to be more legible since reading was hard with dark text on dark background (done)
16. Change some of the wording like the PHQ score so that it is less taxing to read (done)
17. Add more tool tips to make reading and interacting make more sense (done)

Website Link:
http://localhost:63342/Final-project/index.html?_ijt=22at2jtersahg000386v7rb269&_ij_reload=RELOAD_ON_SAVE

GitHub Link:

https://github.com/ayoun8/Final-project.git

Demo Video Link:

https://youtu.be/tOSLhFC-7rk