CSC316 Project Process Book

Week 3: Team and topic

Project Topic: Drugs for Neurological/Psychotropic Conditions

- **Project Title:** Conditions of the brain, treatments, and connections to each other.
- **Project Abstract**
    - The brain is an organ with complex anatomy and various neurotransmitter systems that enable humans to carry out important functions. Disruptions on brain structures and pathways can lead to clinical conditions such as Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), Dementia, and other neurological conditions. Conditions do not always appear in isolation, and it is quite common for certain disorders to have comorbidities– in a worldwide survey of individuals with Major Depressive Disorder (MDD), around 45% of respondents also reported history with at least one anxiety disorder [source]. These interplays between conditions might make some people wonder what's the relationship between conditions in the first place.

      One reason why comorbidities may happen is due to sharing the same brain structures. Therefore, this project seeks to explore the relationship between drugs and the conditions they're used to treat, both label and off-label, and explore potential relationships between conditions based on the medications used to treat them and their Mechanism of Action (MOA). The goal of the visualization(s) is not to seek any discoveries (as none of us have expertise), but rather to serve as a way to inform the general yet curious audience on different disorders of the brain, and how they might play off from each other.

      The main dataset that delineates brain disorders, the drugs used to treat them,

and the mechanism of action (explanation of MOA and receptors) was found in this link: https://www.sciencedirect.com/science/article/abs/pii/S0010482524006206 . When it comes times to building a more cohesive narrative (such as explaining the percentage of population with a certain condition, or drug usage), data can be collected from government datasets (such as statscan, example https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310046501) or health organizations (such as the CDC, https://www.cdc.gov/nchs/fastats/depression.htm).

Week 4: Team Agreement, Detailed Project Plan, and Data

1. Team Agreement

Communication

- Primary Channel: WhatsApp
    - All team-related updates and announcements will be posted in the groupchat.
    - Ideal response time: within 24 hours
- Urgent Matters
    - Direct mentions on WhatsApp
    - If there's no response within a few hours, someone will text/call.
- Weekly Meetings
    - Every Wednesday at 2 pm, via Zoom (link)
    - Each member gives a brief status update on tasks.

Work Distribution & Accountability

- **Role-Specific Responsibilities**
    - Team Lead (Jennifer): Project timeline, planning + deliverables tracking, design brainstorming, contact with course coordinators, development.
    - Data Specialist (Tianyu): Data collection, cleaning, documentation, development.
    - Visualization & Development Specialist (Yizhi): Design brainstorming, video presentation preparation, development.
- **Shared Responsibilities:** Contribution to discussions and documentation, cross-role assistance when needed, updates on assigned tasks. If unexpected delays happen, send a message to the group chat ASAP.
- **Accountability Measures:** Weekly check-in @ team meetings.

Version Control & Documentation

- **Code**
    - Short yet descriptive variable names. Provide comments if necessary.
    - JSDoc comments/descriptions for each function. Write inline comments for code blocks that aren't intuitive, and important steps of the code.
- **GitHub Repository**
    - Protected main branch, pull requests require at least one reviewer.
    - Each team member will have its own branch from main.

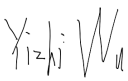
Signatures

- Members**:**
    - Jennifer A. Fong Li (Team Lead / Coordinator)
    - Tianyu Fan (Data Specialist)
    - Yizhi Wu (Video & Development Specialist)

- Signed:

Jennifer A. Fong Li:  Feb 5, 2025

Tianyu Fan:  Feb 5, 2025

Yizhi Wu:  Feb 5, 2025

---

2. Project Details

Basic Information

- **Project Title:** Conditions of the brain, treatments, and connections in between.
- **Team Name:** Chartbusters
- **Members**:
    - Jennifer A. Fong Li (j.fongli@mail.utoronto.ca)
    - Tianyu Fan (tianyu.fan@mail.utoronto.ca)
    - Yizhi Wu (yizhi.wu@mail.utoronto.ca)

Background & Motivation

The human brain is incredibly complex. Disruptions in its pathways can lead to various conditions such as Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), Dementia, etc. These conditions often share overlapping symptoms or comorbidities (for example, about 45% of individuals with MDD report at least one anxiety disorder in their lifetime [source]). Another point of interest is how certain medications can exacerbate symptoms for other conditions (for example, meds for depression can cause manic episodes in Bipolar disorder, perhaps even with hallucinations or delusions [source]).

Our team is interested in exploring how psychiatric and neurological medications connect different diagnoses, perhaps including on-label vs. off-label uses, and shared mechanisms of action.

Related Work

- Network visualizations (such as those made by Gephi!)
- Inspiration if we decide on including a geographical map: https://www.reuters.com/graphics/world-coronavirus-tracker-and-maps/vaccination-rollout-and-access/

Audience and Questions

- **Audience**: The target audience is the general public. More specifically, those that are curious about disorders of the brain, but lack a strong science background.
- **Primary Questions**: Which drugs are used for multiple diseases? Is there any relationship between the diseases treated by the same drug?
- **Overall goal**: Create a visual, user-friendly representation of the relationships among brain conditions and treatments.

Data & Cleanup Procedures

- **Current Datasets**
    - Dataset of diseases, drugs, and targets (csv) (Main dataset): https://www.sciencedirect.com/science/article/abs/pii/S0010482524006206 .
    - Dataset of mental health around the world (csv): https://www.kaggle.com/datasets/imtkaggleteam/mental-health
- **Current Issues with the main dataset:**
    - Redundant or inconsistent drug name fields (e.g., Drug_Name_x, Drug_Name_y)
    - N/A values in Disease_Name, might need to replace N/A from the description in Labeled_Indications, drug synonyms might need parsing.
    - Mechanisms of Action (MOA) are string descriptions, needs extraction/tagging
- **Cleanup Approach:**
    - Use Python (Pandas) scripts to:
        - Merge redundant drug name columns
        - Standardize disease/condition names
        - Parse drug synonyms into a structured format
        - Extract MOA details into separate columns or key-value pairs

Project Timeline

- **Week 4: Team Agreement and Detailed Project Plan, Data**

    - **Meeting**: Finalize team agreement, detailed project plan, and discuss cleaning data.
    - **Submission**: .pdf of team agreement and detailed project plan.
    - **DUE**: February 6, 2025 @11:59pm
- **Week 5: Data, Map**
    - **Meeting**: Discuss storyboard. Initial visualization sketches and user flow design.
    - **Submission**: Cleaned datasets.
    - **DUE**: February 13, 2025 @11:59pm (Quick check-in on Feb 10)
- **Week 6: Sketches, Decide & Storyboard**
    - **Submission**: Submit sketches and declare final decision.
    - **DUE**: February 20, 2025 @11:59pm (Quick check-in on Feb 17)
- **Week 7: Prototype V1**
    - "The overall structure and the content should be clear. We will ask you to hand in your code in its current state."
    - **Submission**: Project in its current state.
    - **DUE**: March 6, 2025 @11:59PM (Quick check-in on Mar 1 and Mar 3)
- **Week 8: Prototype V2 (95% done)**
    - **Submission**: updated project.
    - **DUE**: Mar 13, 2025 @11:59pm (Quick check-in on Mar 10)
- **Week 9: Prototype and Test**
    - **Submission**: Think Aloud Study Results
    - **DUE**: Mar 20, 2025 @11:59pm
- **Weeks 10–11: Incorporate user feedback, wrap up, group assessment.**
    - **Quick check in**: Mar 24, 2025 and Mar 27, 2025
    - **Submission:** Final project.
    - **DUE**: Monday, Mar 31
- **Week 12: Watch party.**
    - Apr 2, 2025 (During lecture)

Project Plan Overleaf Link: https://www.overleaf.com/project/660fe89269d639baf73104ab

Week 5: Data, Map

Week 5 Instructions

Group Discussion

1. Who is your audience? Come up with **at least three** options and pick one target audience.
    a. **General Public**
    b. High school students (?)
    c. Data analyst (???)
    d. Patients and/or family of patients
2. 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?
    a. Our target audience will be the **General Public**. They likely know the names of some conditions, and some basic information they heard on the news or social media. Maybe even some medications and unofficial ways to treat common diseases (such as treating Seasonal Affective Disorder with Vitamin D supplements).
    b. The audience might want to understand brain conditions, how they are connected, and how some disorders often occur together and learn about the drugs used to treat these disorders, including on-label and some off-label uses. Using data visualizations to understand complex medical relationships more.
    c. The visualization literacy of the general public can vary across the board. Therefore, the information will be simplified and we will remove some complicated technical jargon. The level of detail to be presented will have to be fairly straightforward (not getting into the nitty-gritty details, such as neuroanatomy), and there will have to be some basic introductory definitions (For example, defining what an “agonist" is). We will design Interactive elements to attract users reading the information and engage more with the content (for example, a tooltip that explains what an “agonist” is).

3. 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 ten questions**.
    a. What diseases can each drug treat?
    b. What are some common comorbidities?
    c. Why might medications be effective at treating the conditions they are prescribed for? (What are the medications' targets?)
    d. What are some medications that are prescribed off-label?
    e. Which drugs can treat a particular disorder? Conversely, which disorders can a single drug treat?
    f. What medications might trigger symptoms for other conditions? (e.g., antidepressants triggering manic episodes in bipolar disorder)
    g. Do drugs that target the same receptor treat similar diseases?
    h. Which receptors act as "hubs" connecting many drugs and diseases?
    i. Which diseases have the most drugs available?
    j. Are there any drugs used for multiple diseases?
    k. What drug actions are the most common?
4. What data do you have? Look at it in Excel or Google spreadsheet and briefly describe each attribute and its data type (categorical, ordinal, or quantitative) in your process book. It's OK if you are unsure about the data type for some attributes - you can simply describe them (e.g., geographic location).
   The datasets can be found in this gdrive folder: datasets
   The **"general_drug_data.csv"** is the dataset that contains our main data. In this file, we have:
    a. drug_name **categorical(nominal)**: the name of the drug.
    b. inn_name **categorical(nominal)**: the INN (International Nonproprietary Name) name of the drug.
    c. brand_names **categorical(nominal)**: the brand names of the drug.
    d. list_of_conditions **categorical(nominal):** description of the intended treatment of the drug labeled by FDA.
    e. target_name **categorical(nominal)**: the name of a drug's MOA intended target..

f. type_of_action **categorical(nominal)**: action of the drug on the target.

The "**off-label_uses.csv**" file has the off-label uses for a drug. Since there aren't that many records of off-label usage, we decided to keep it in a different table. Here we have:

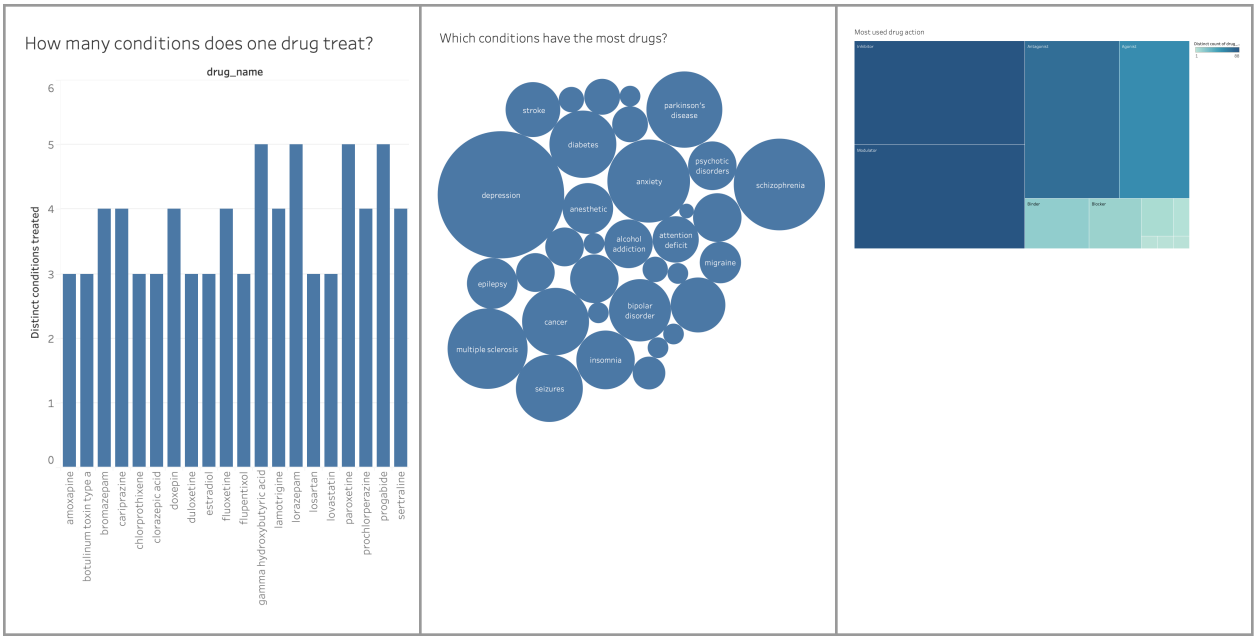
g. drug_name **categorical(nominal)**: the name of the drug.
h. offlabel_uses **categorical(nominal)**: description of the drug usage that is not labeled by FDA.

There is also a dataset called "**drugs_categories.csv**". The original dataset classified the drugs in "categories" based on the conditions they treat. So, this cleaned dataset was made to group the drugs into a general area of what they generally treat while keeping it separate from list_of_conditions in the main dataset as conditions may overlap (ex. "depression" and the "depressive disorders" category).

i. drug_name **categorical(nominal)**: the name of the drug.
j. category **categorical(nominal):** category the drug belongs to.

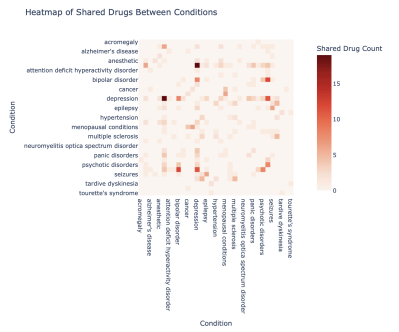
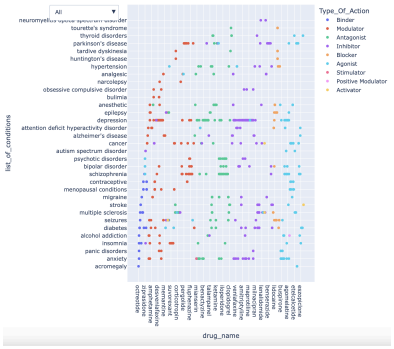
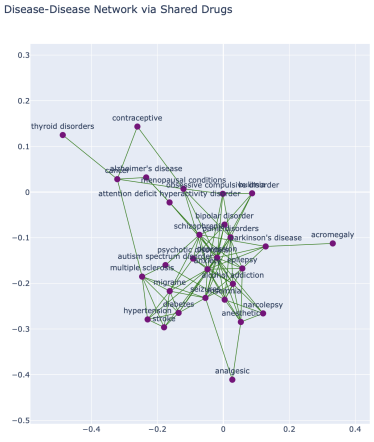
Initial visualizations per team member:

- Jennifer's Tableau Work

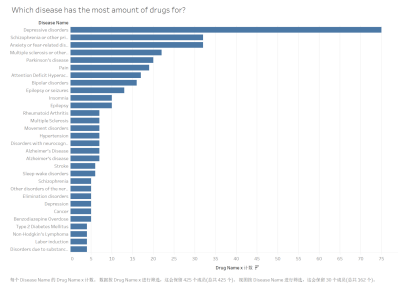
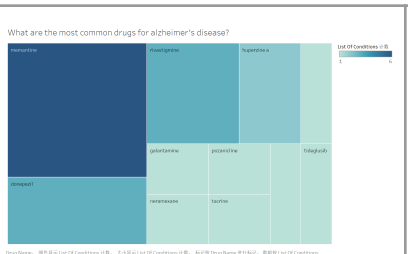
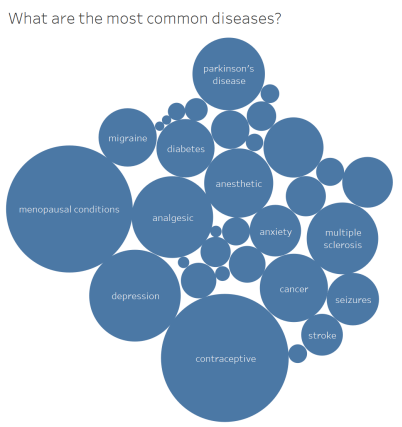


The question this visualization answers is “How many conditions does one drug treat?” This is similar to question “a.) What diseases can each drug treat?”. The main difference is that the tableau question focuses on the total count of conditions, while the original question seems to ask the list of conditions– not the count. For a concise view, the tableau question might be better, but for actual comparisons the original question.	The question this visualization answers is “Which conditions have the most drugs?”. This is the most similar to question “i.) Which diseases have the most drugs available?”. There isn’t much difference between the two questions, other than the original one seems to ask for a “top x diseases” versus a general view across the board.	The question this visualization answers is “What drug actions are the most common”. This is the same as “g.) What drug actions are the most common?”. Since there is no difference, no question is better than the other. This is the same question because any other question about the detail of how drugs affect their targets might be out of scope for the target audience. So, for general curiosity, this question is enough.

- Tianyu’s Tableau Work

		
Shared Drugs between conditions: how many conditions share the same drug, and how many drugs can treat the same condition?	Drug vs Conditions: For each drug of a given type, what diseases can they treat?	Disease-to-Disease network: We connect drugs if they can cure the same diseases. Then we form a network of drugs.

- Yizhi's Tableau Work

		
This graph answers the question that is mentioned in the 10 questions as question i: Which diseases have the most drugs available? It sticks to the original question because the result has referential value and is interesting.	This graph answers the question: What are the common drugs for Alzheimer? It is not in the original question, but as prevalent as the disease is, it is definitely true that the general public cares about the cure of the disease.	The graph is a reduced version of question b: What are some common comorbidities? The reason for the reduction is there is not enough data to know whether diseases happen together as a comorbidity.

Week 6: Sketches, Decide & Storyboard

[Week 6 Instructions](#)

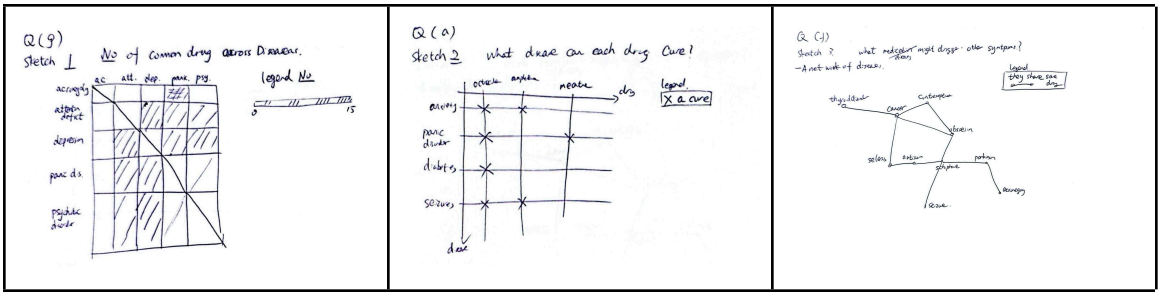
Sketches Step

Instructions

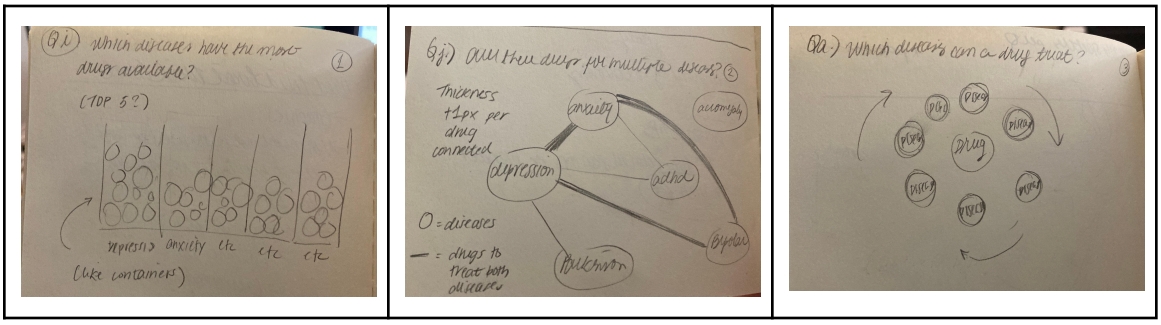
- Five (5) different sketches per team member.
- Use pen and paper.
- Label each sketch with the question(s) it is answering.
- For each sketch, create a legend that describes the visual encodings of your data.
- Upload them to your process book

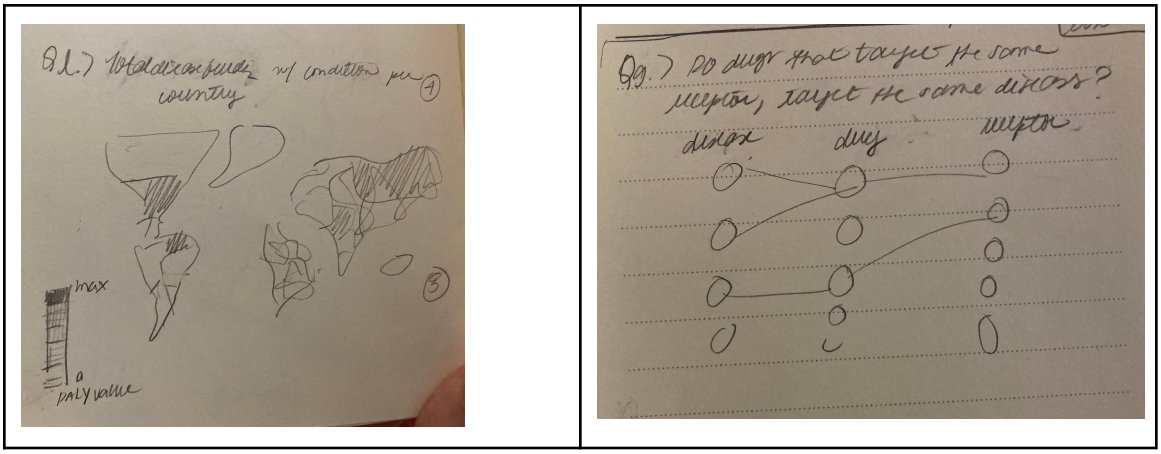
Photos sketches

- Tianyu:
  1, 2, 3, 4, 5

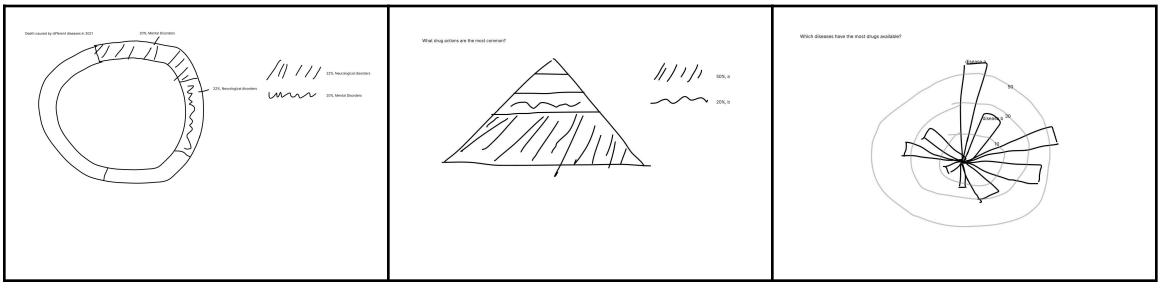


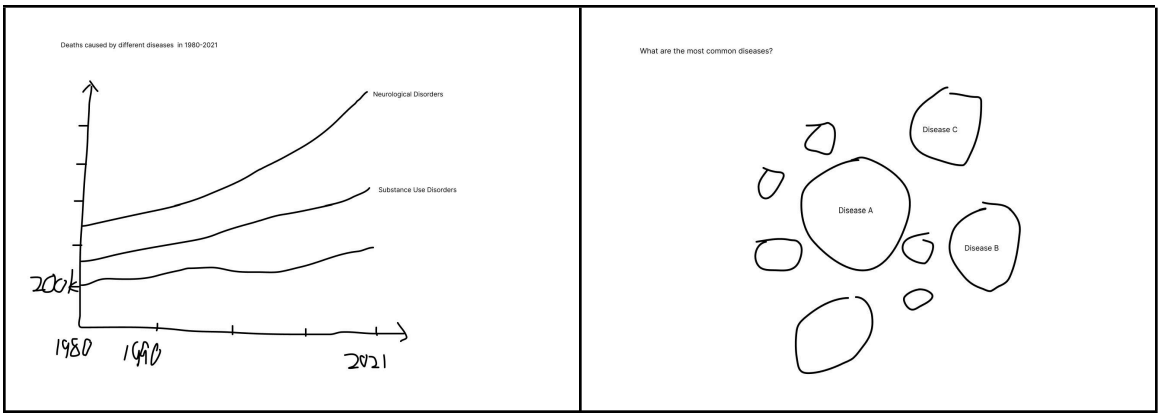
- Jennifer:
  6, 7, 8, 9, 10





- Yizhi:
  11, 12, 13, 14, 15





Decide Step

Decide which sketches to implement. Pick between 4-6 sketches.

Instructions

- Create a table where each question has a question ID, each sketch has a sketch ID, and author name, and number of votes.
    - If there are duplicate visualizations, combine them into one row.
- If sketches need more explanations, ask the author to explain their ideas in more detail.
- Vote.
    - Everyone gets 5 votes.
    - Multiple votes are allowed.
    - Votes on your own sketches are allowed.
- Select 4-6 sketches, copy and paste them and the questions that they answer.
    - Arrange them logically- from relevant to least relevant.
- Add a one-paragraph explanation at the end summarizing your decisions and rationale for choosing the sketches you plan to implement.

Affinity Diagramming wk06-affinity-diagram

Sketch Decision Meeting Documentation wk06-sketches-meeting-minutes

Summary: During the 45-minute meeting, we uploaded our visualizations, further explained on any visualizations anyone was confused about, voted, and came up with the final visualizations: 1, 3, 4, 7. Afterwards, we discussed some insights and the main message of the project.

Storyboarding Step

1. **Pick your main message**

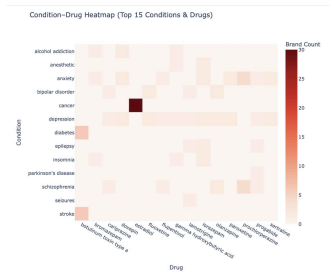
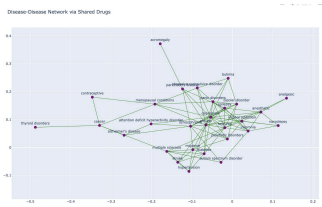
Insights per team member:

- Jennifer

1. **Insight**: Depression is a condition with most treatments available. **So what?**: Perhaps depression is one of the most treated conditions due to how common it is, or	2. **Insight:** Inhibitors are one of the most common mechanisms of action. **So what?**: Perhaps the reason why inhibitors are popular is because it inhibits the reuptake of	3. **Insight:** For depression, antagonist drugs are the second most common after Modulators. **So what?:** The first line of defense for depression are SSRIs, which are

perhaps it is a common comorbidity with other conditions.	neurotransmitters– therefore, potentially bringing the brain's chemicals to a level where the condition is attenuated.	inhibitors. However, this data contains more antagonists and modulators. This shows that the preferable action is not always the most common one to come across.

- Tianyu

		
1. **Hook:** Do you know that many diseases share the same drugs, even though they seem unrelated? **Insight:** Many Diseases share the same drugs. It is far much more common than we think. **Solution:** Optimizing treatment strategies based on drug-sharing patterns.	2. **Hook:** How can we use the fact that some drugs could be used to treat multiple diseases? **Insight:** Many diseases share common drugs. This maybe due to them having the target inhibitor for the drugs **Solution:** Utilize AI-driven analysis to identify more cross-disease drug repurposing opportunities based on the existing data.	3. **Hook:** Each disease has many drugs to cure. Each drug also has many brands. **Insight:** So it is important to compare the effectiveness and safety of different drugs.

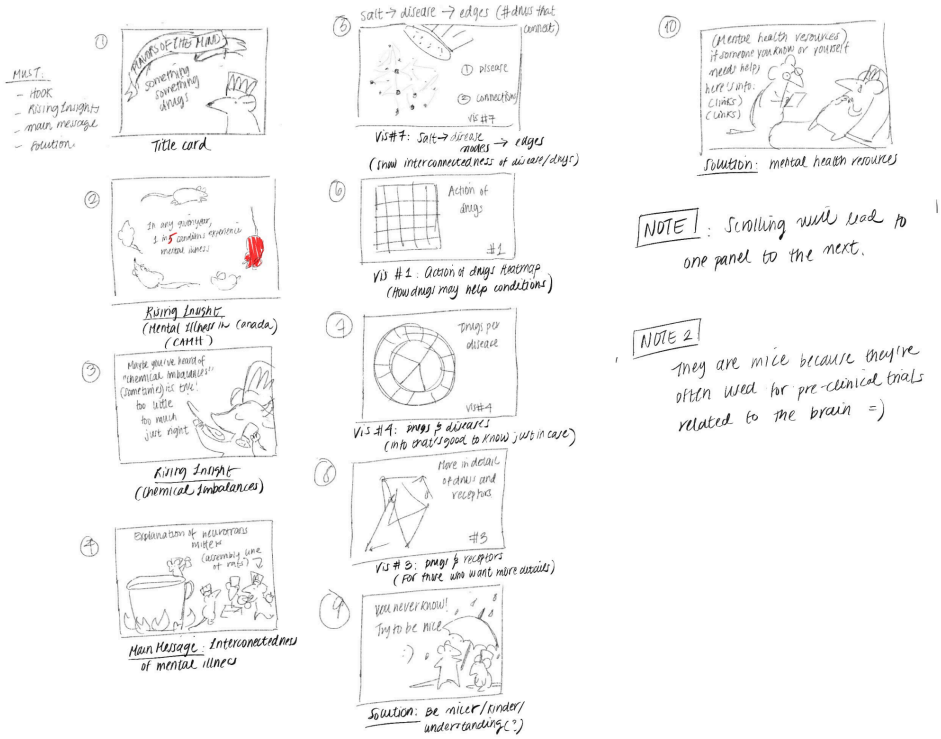
- Yizhi

1. **Insight**: There are some drugs that treat multiple diseases. **Message**: Perhaps the diseases that are treated by the same drugs have something in common.	2. **Insight**: Deaths caused by neurological diseases keep increasing over the past 20 years. **Message**: One possible reason for the deaths is the diagnosis of such diseases are getting more prevalent over the years.	3. **Insight**: Mental Disorders result in significantly more DALYs(Disability-Adjusted Life Years) than neurological diseases. **Message**: Mental disorders have more impact on matters of life quality other than deaths.

Main message: Conditions of the brain are intricately interconnected.

2. **Sketch your data storyboard**
   Ensure the storyboard has: hook, rising insights, main message, and solution.

   Final Storyboard: Storyboard.jpg



Week 7: Prototype V1

Week 7 Instructions: CSC316-Final-Project-Week-7-Instructions_2025CSC 316.docx

Requirements:

- At least two D3 visualizations already partly implemented (including data loading and the basic vis, filtering does not have to work yet)
- Detailed drafts for 2-3 more visualizations.
- Rough webpage design and structure has to be done and implemented (placeholders for visualizations, text, and images allowed)
- The first design of an innovative view.
- Interactions (e.g., filtering, brushing, etc.) have to be designed (at least in a textual description and some sketches)

Information:

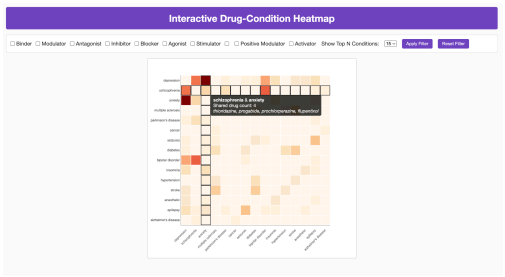
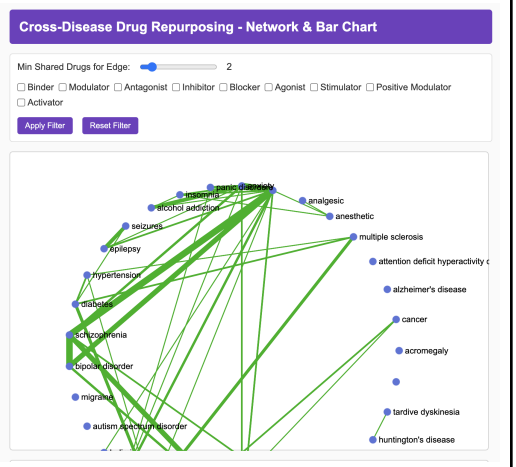
- Name of students who worked on V1 submission:
    - Jennifer (Web Page design & structure, data cleaning, new visualization sketches, process book upkeep)
    - Tianyu (Data cleaning, 2 visualizations)
    - Yizhi (1 visualization)

- Data scraping and data cleaning:
    - Link to data folder: https://drive.google.com/drive/folders/1lRPqgJ1PWAzY59kt_M2-nbsJ-N4Rfls6?usp=sharing

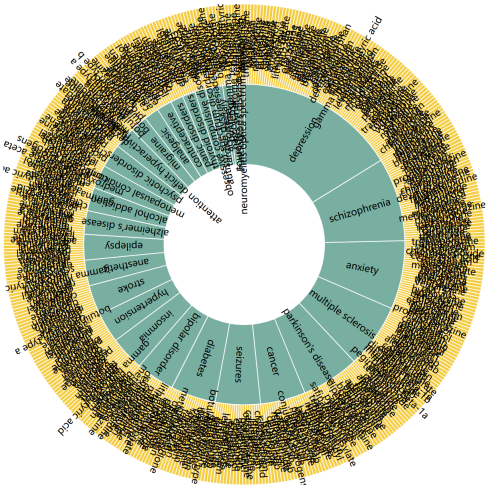
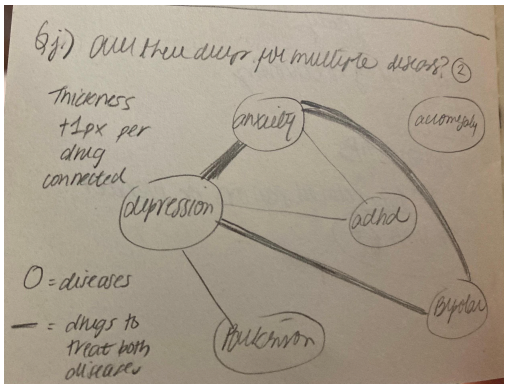
- Partially implemented visualizations:
    - Tianyu: https://1drv.ms/u/s!AtwqGMtC2L3yg7l-3lMHjf4c4V5LtQ?e=7PKkpa
    - Yizhi: Visualization 3.zip

- Interaction:

- Tianyu

	
Interaction: Tooltips showing common drugs and numbers Filters for drug type, TOP N filters	**Interaction**: Filters for minimal edges to add to the graph FIlters for types of drugs Tool tips showing diseases that connects themz~!D

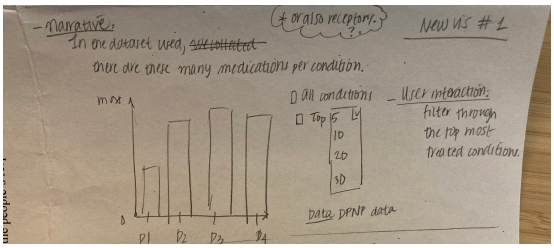
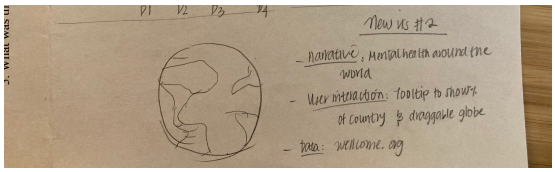
- Yizhi

	
Interaction: Highlight drug cures for specific conditions when mouse hovering **TBD:** Improve readability	**TBD:** To be implemented next week Possible interaction: - Hovering over an edge emphasizes the two nodes at

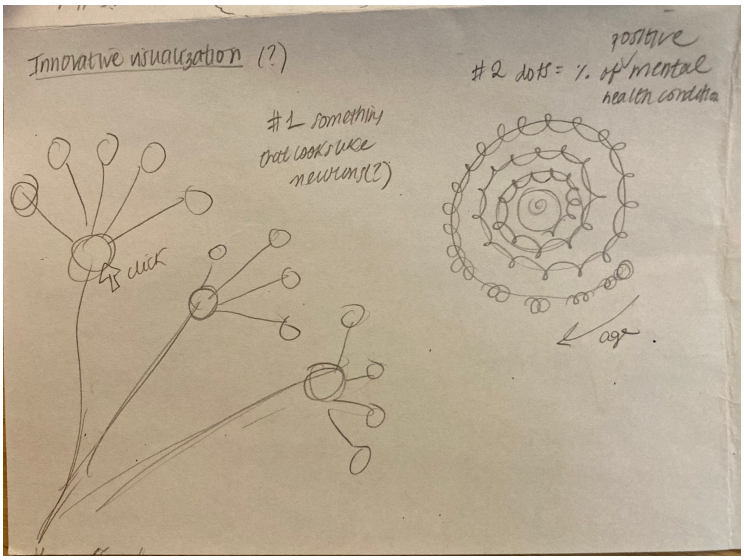
	each end.

- Rough webpage design and structure:
  - Jennifer: link to structure work
  - Recording: V1-Webpage-structure-recording.mov

- New visualizations:

	
Details: ● Simple bar chart. ● Question: how many treatments are available per condition. ○ Alternative Q: How many receptors are targeted by a drug's medication? ● Categorized by color (hue?). ● Interaction: ○ User can filter to show all, top 10, top 5, etc.	Details: ● GeoJSON draggable globe, like the one from a previous lab. ● Question: % reports of depression and anxiety in the world. ● Each country is colored based on its percentage value (from light (least report) to dark (most reports). ● Interaction: ○ Users can drag the globe to look at the % in other countries. ○ Hovering over a country shows the percentage.

- Innovative view design:

- 

  - **Explanation of #1:**
    - User clicks a node, then the node produces other nodes related to it.
    - It's an interesting idea, but not sure what question it could answer :(
  - **Explanation of #2**:
    - Based on this dataset: https://www.kaggle.com/datasets/anthonytherrien/depression-dataset/data
    - A spiral that contains dots representing the age of a respondent. The size of the dot represents the % of reports with some sort of history of mental illness.
    - User's can change the 'mental illness history' to other metrics such as substance abuse/family history of mental illness/etc.

Link to all project files/documents:

https://drive.google.com/drive/folders/1CZhttHOp_n1lL0-Kge1lH1CSIlWmU13U?usp=sharing

---

Week 8: Prototype V2

Nothing to write– just submit the code.

Week 9: Test

Think aloud test:
Gdocs link: Think-Aloud-Study.docx .
The completed study template will be attached in the second page as well.

Questions:

1. Based on the results of your 'think aloud' study, what would you improve in your data story?
    - I think maybe some more explanation about the mechanisms would be helpful. The tester seemed to be a bit confused by it.
2. 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?
    - The tester seemed to get the idea of mental health being important, which is pretty accurate. Specifically though, the goal was to show how the brain is interconnected so maybe some tweaks in the narrative would be good.
3. 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.
    - From testing, these are the things to fix:
        i. Bigger font
        ii. Fix scrolling function sensitivity
        iii. Better readability in the donut chart
4. 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.
    - Currently working on them! The TODOs for Tianyu and Yizhi can be found here: TODO: VIS FIXES. For Jennifer, the TODOs are in the Github repository.

Week 10-11: Wrap Up & Submission & Peer Assessment

...

Week 12: Watch Party, Best Project Prizes