CPSC 583 Final Project Report

Topic: OSMI Mental Health in Tech

Nathan Cruz, 30030215

December 7, 2020

1. Introduction

This project is about a visualization of an OSMI survey in 2019 that covers mental health in the tech workplace. Its goal is to help with the discovery of various trends and correlations that can be found in the responses of the participants. The visualization takes the form of a sankey diagram that can be interacted with in order to filter out what responses a user is interested in. This report goes over the process of how the visualization was made and the various discoveries encountered during its creation. The resulting visualization overall is a simple yet effective and interesting way to visualize a survey dataset.

2. Data Description

2.1. Data Descriptions

Dataset Source: https://www.kaggle.com/osmihelp/osmi-mental-health-in-tech-survey-2019

I've chosen this dataset because it covers a field that I may possibly work in. Alongside that, I've always been interested in the idea of mental health and illness. It's important for people to not turn a blind eye on mental health as it could affect how they perform in the workplace. It will be interesting for me to see what people think about mental health issues. This dataset covers 82 different questions but I'll be focusing on a select few of them:

- Would you feel comfortable discussing a mental health issue with your coworkers?
- Do you *currently* have a mental health disorder?
- Have you observed or experienced an *unsupportive or badly handled response* to a mental health issue in your current or previous workplace?
- Have you observed or experienced a *supportive or well handled response* to a mental health issue in your current or previous workplace?
- Overall, how well do you think the tech industry supports employees with mental health issues?
- What is your age?
- What is your gender?

2.2. Pros and Cons of the Dataset

Pros

There seems to be a nice variety of data types to use in this dataset. There are also a fair amount of questions that seem to go hand in hand with each other which will help with comparisons.

Cons

A majority of the questions in this dataset are mainly "yes or no" answers. This may restrict the potential ways in visualizing the data available. Also by focusing on a select set of questions, there's a chance that I may miss some important relationships with certain questions. Some of the questions in this dataset can have multiple sub questions related to them. However, including all of them might not be necessary and some of them ask for a worded response that can't be measured easily

2.3. Dataset Decision

This dataset covers a large quantity of questions and it will be the only dataset I will be working with. Many questions in the dataset are similar to others with a single word separating them. I've tried to pick out the questions that I'm personally interested in seeing the response to in hopes of getting a better understanding on how people and companies feel about mental health.

3. Design Process

The design process for this project goes as follows: Firstly, subsets of the dataset had to be sketched in order to discover possible ways how the final visualization would look like. Secondly, visualization variations had to be created in order to explore what visualizations worked better than others when scaled up in dataset size. Lastly, one of the variations was chosen and interactions were implemented to help emphasize certain aspects of the dataset.

3.1. Sketch-able Data Subsets

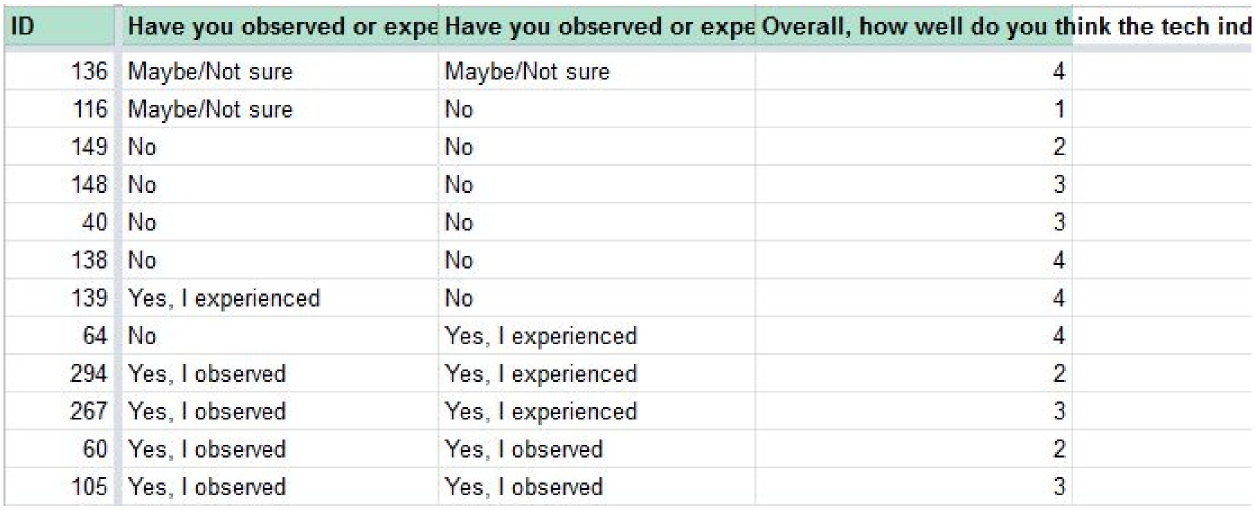
Each of these subsets are based around a certain question. Subset 1's question is wondering how one's overall opinion of the tech industry's outlook on mental health is affected by an observance of a response to a mental health issue whether it's good or bad. Subset 2's question is checking if those with a mental health disorder would be willing to discuss it with their coworkers and if there's any trends when it comes to the user's age. The tradeoff between and sketchability and subset representation was handled based on the types of data involved. Similar data types were grouped together in hopes of making a visualization that seems straight forward. Subset representation might vary as each subset's rows have been picked at random.

3.2. Design Direction

In the initial 10 sketches for both subsets, the idea was to cover as much variety as possible. Different means of representing the data was used such as using lines, grouping numbers or shapes, and node-links. Basic bar charts were generally avoided during these sketches as they seemed to lack the flexibility that the other representations had to offer. The variation sketches focused on a single initial sketch and played around with how it was shown. Some adjustments in the variations involved colors and shading while others focused on the arrangement of the data in terms of placement.

3.2.1. First Sketches & Variations (Subset 1)

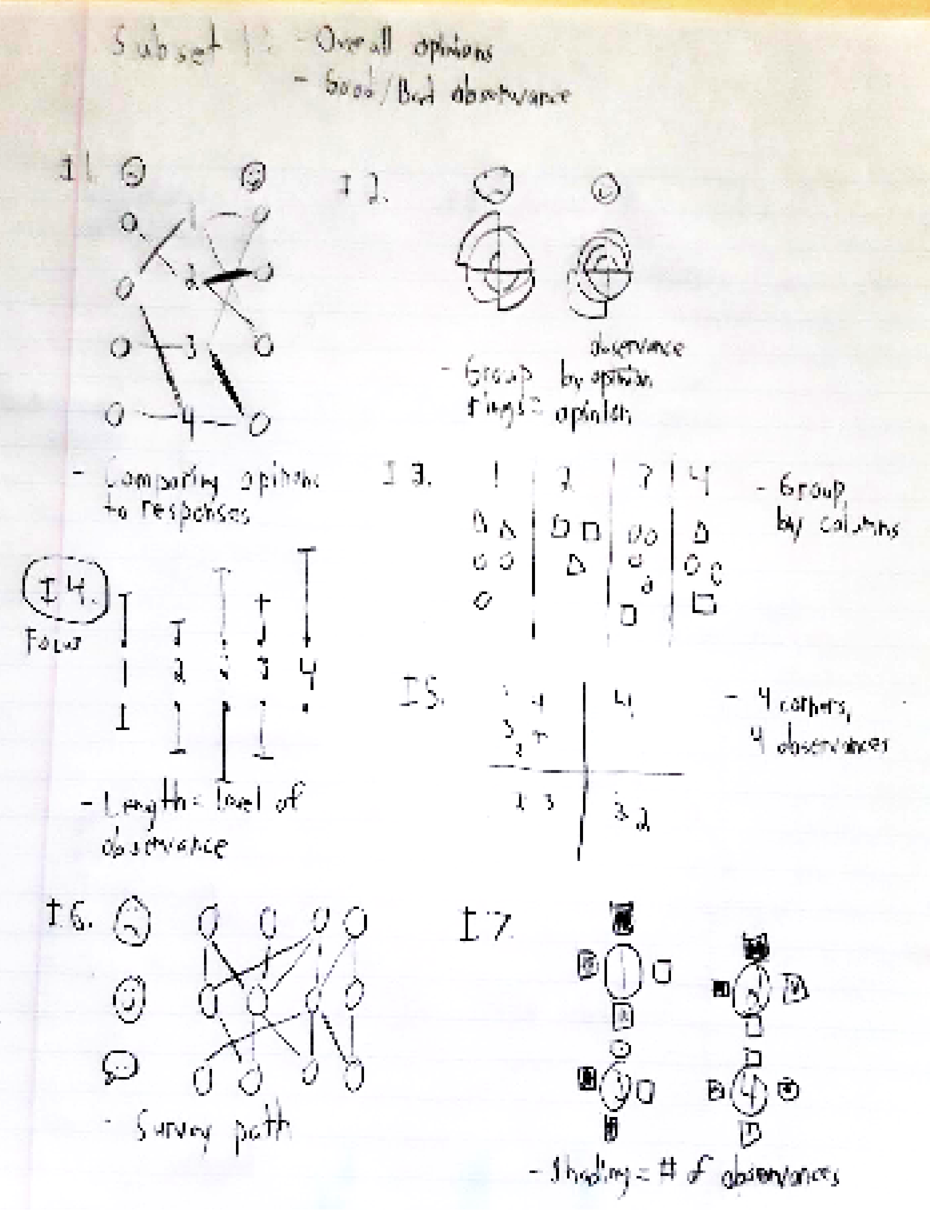
Subset 1

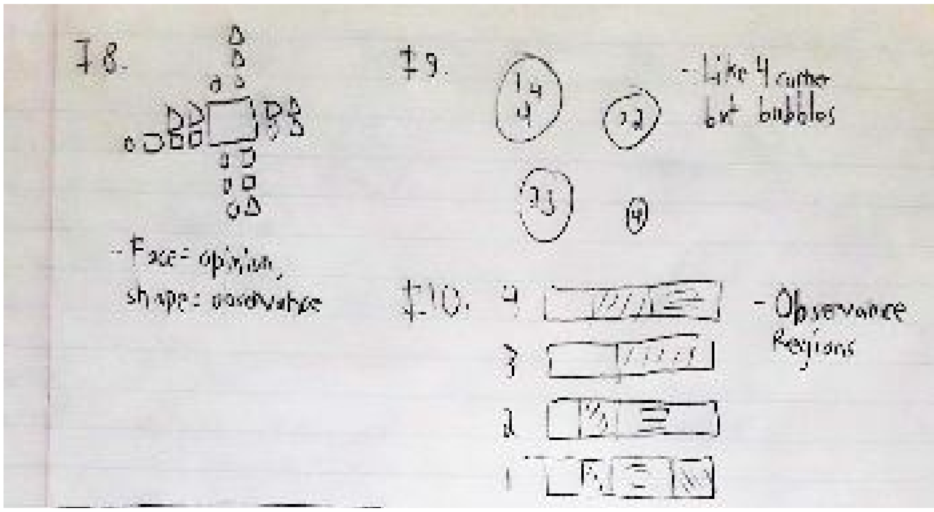


ID	Have you observed or expe	Have you observed or expe	Overall, how well do you think the tech ind
136	Maybe/Not sure	Maybe/Not sure	4
116	Maybe/Not sure	No	1
149	No	No	2
148	No	No	3
40	No	No	3
138	No	No	4
139	Yes, I experienced	No	4
64	No	Yes, I experienced	4
294	Yes, I observed	Yes, I experienced	2
267	Yes, I observed	Yes, I experienced	3
60	Yes, I observed	Yes, I observed	2
105	Yes, I observed	Yes, I observed	3

- Have you observed or experienced an *unsupportive or badly handled response* to a mental health issue in your current or previous workplace?
- Have you observed or experienced a *supportive or well handled response* to a mental health issue in your current or previous workplace?
- Overall, how well do you think the tech industry supports employees with mental health issues?

First 2 columns: Observance level, Observance type
Last column: Overall opinion


Subset 1: - Overall opinions
- Good/Bad observance
I1.
1
2
3
4
- Comparing opinions to responses
I2.
- Group by opinion
rings = observance opinion
I3.
1
2
3
4
- Group by columns
I4.
Four
1
2
3
4
- Length = level of observance
I5.
4 corners, 4 observances
I6.
- Survey path
I7.
1
4
- Shading = # of observances


I8.
- Face = opinion, shape = observance
I9.
- Like 4 corner but bubbles
I10.
4
3
2
1
- Observance Regions

Extra Sketch Notes

I2. Two rings that group good and bad observances. Each “quarter” is an observance level and each “layer” is based on the # of opinions that hold that observance level

I3. The opinions are split into columns with each shape representing an observance level. There would be 2 for good and bad observances.

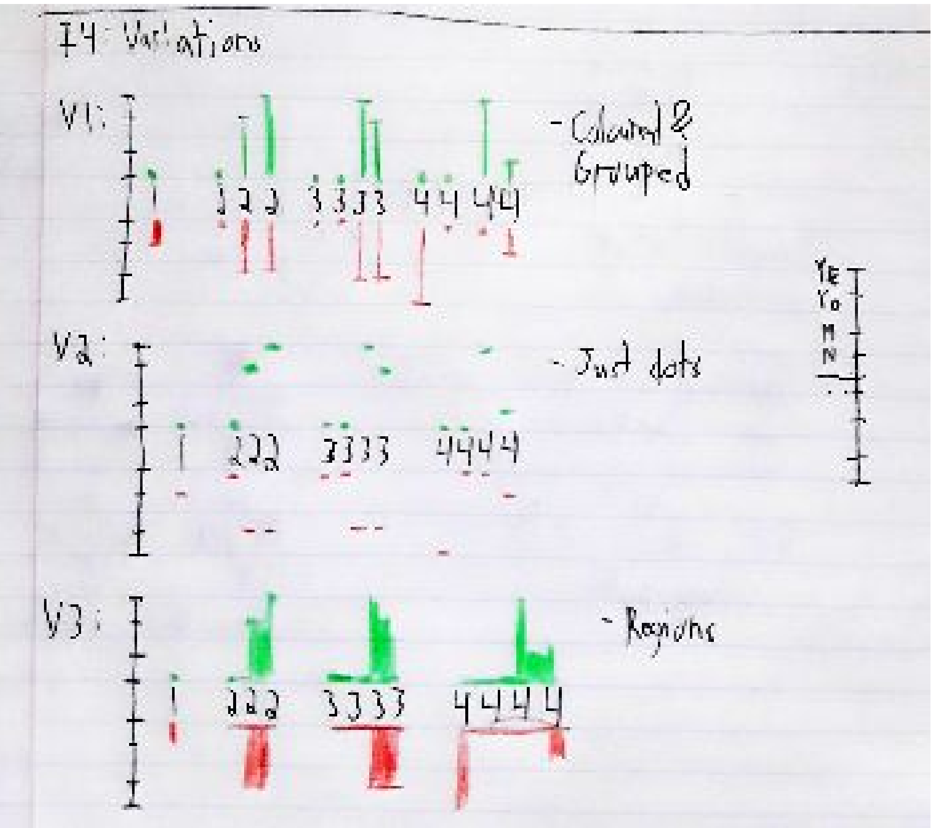
I4. The longer the line, the higher the observance level (No -> Maybe -> Yes, observed -> Yes, experienced)

I5. Like I3 but with a different layout. However the corners are based on the observance level instead of overall opinion

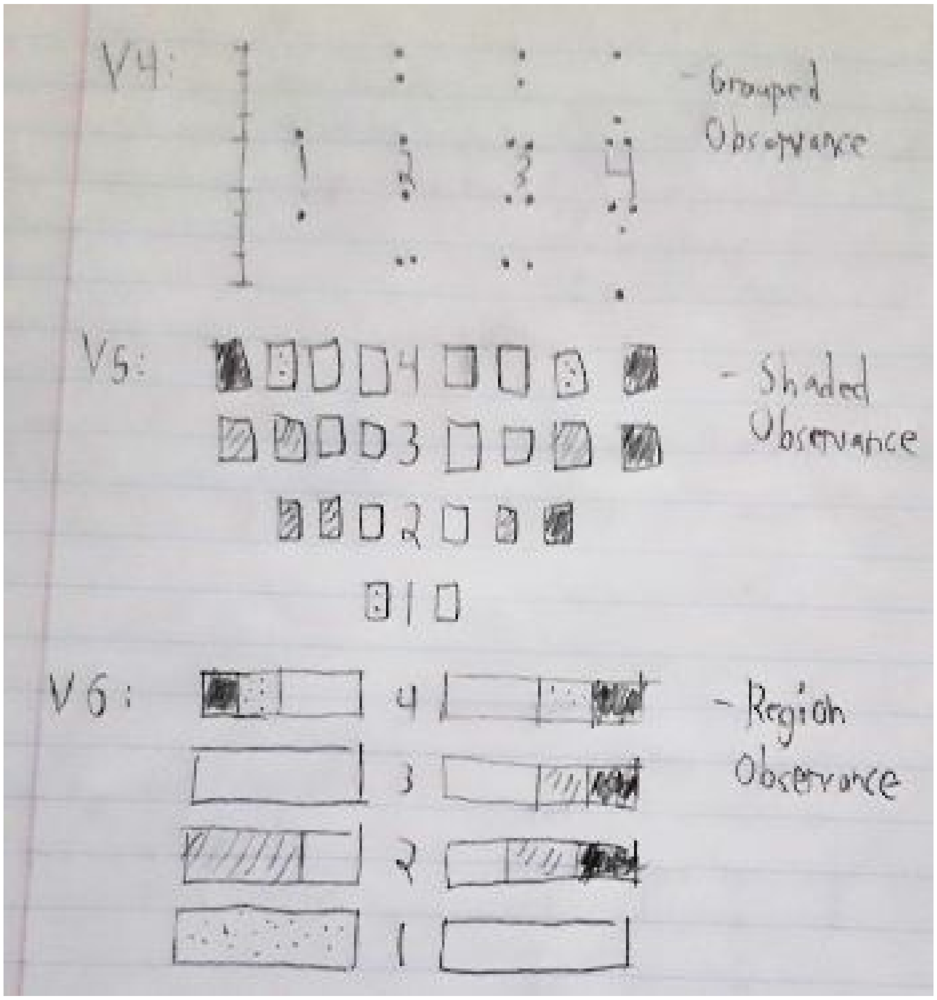
I7. Each block around an opinion represents an observance level. There would be 2 for good and bad observances.

I8. Two columns of shapes. One for each observance type

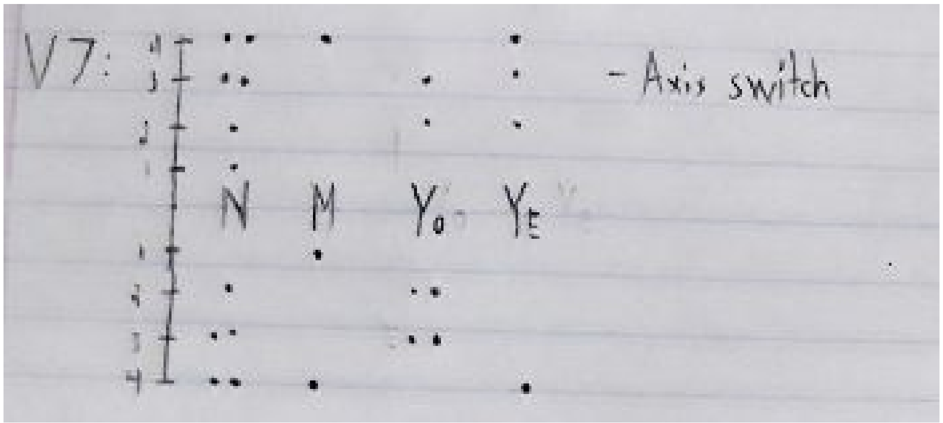
I10. Shading type = Observance level

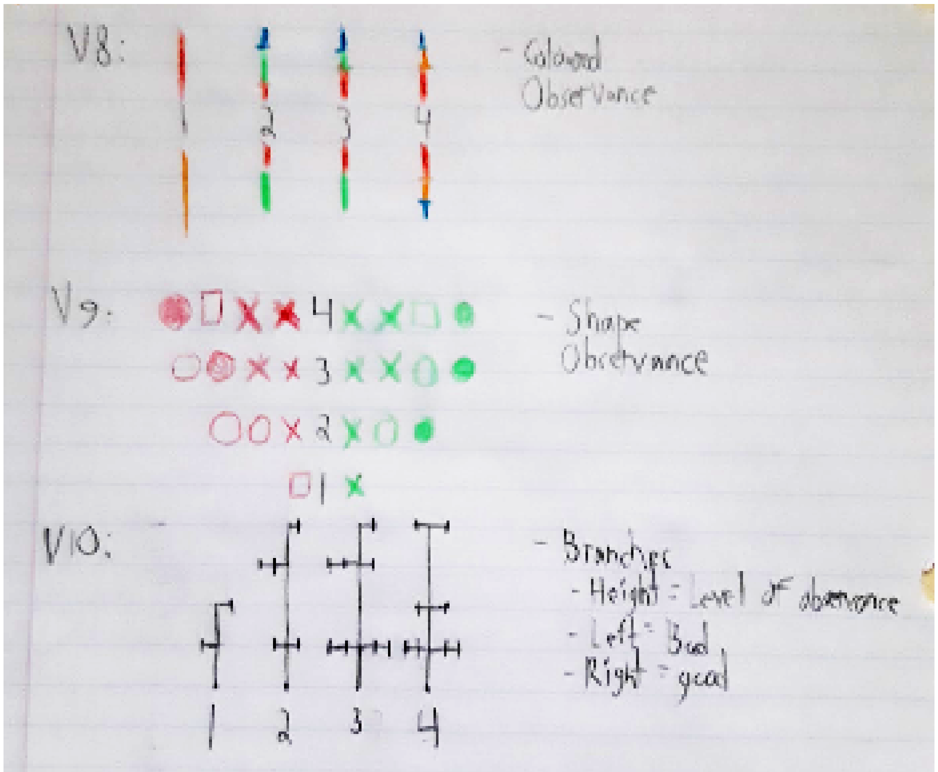


V1-4: Upper section is good responses, Lower section is bad responses



V5-6: The heavier the shading, the higher the observance level (No -> Maybe -> Yes, observed -> Yes, experienced)



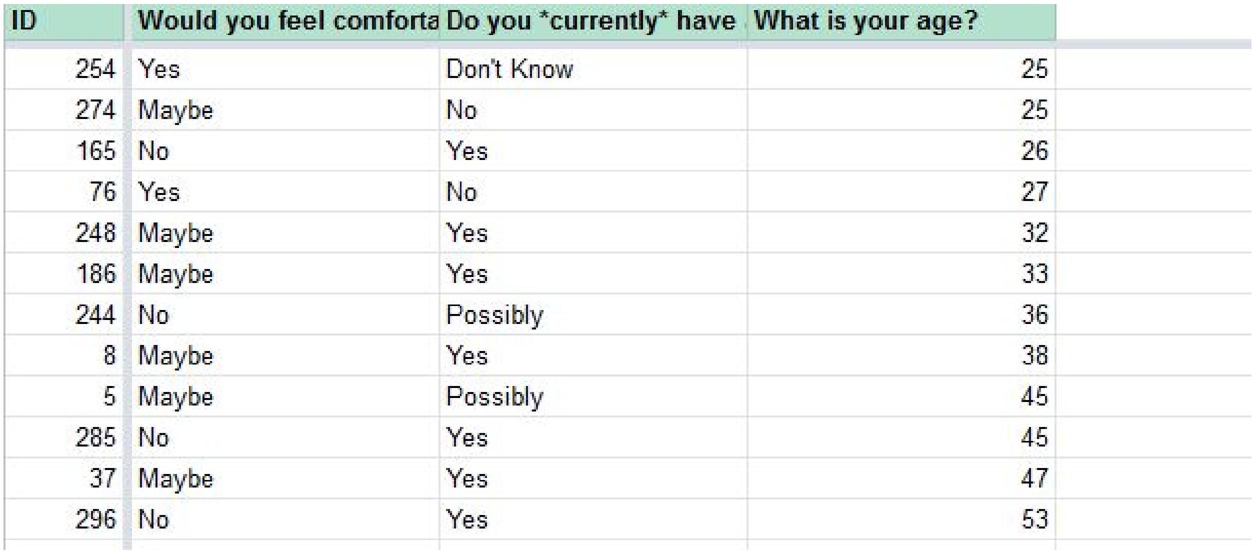


V8: Color is based on observance level
Red: No
Orange: Maybe
Green: Yes, observed
Blue: Yes, experienced

V9: Symbol is based on observance level
X: No
Square: Maybe
Empty Circle: Yes, observed
Solid Circle: Yes, experienced

3.2.2. First Sketches & Variations (Subset 2)

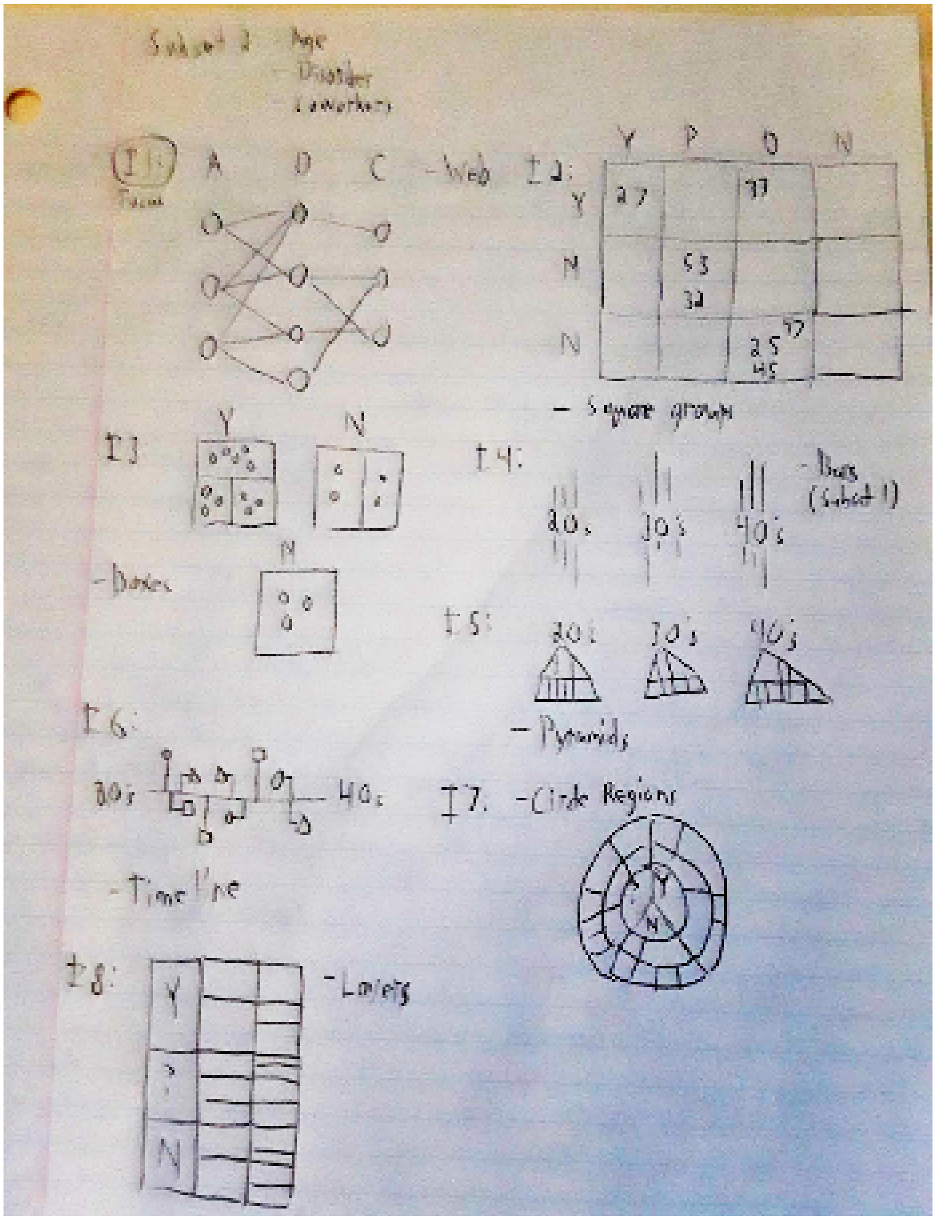
Subset 2

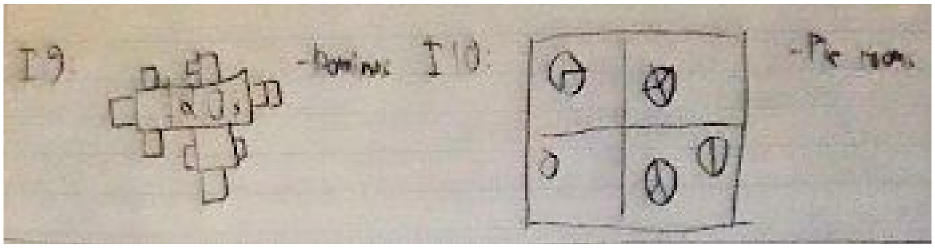


ID	Would you feel comforta	Do you *currently* have	What is your age?
254	Yes	Don't Know	25
274	Maybe	No	25
165	No	Yes	26
76	Yes	No	27
248	Maybe	Yes	32
186	Maybe	Yes	33
244	No	Possibly	36
8	Maybe	Yes	38
5	Maybe	Possibly	45
285	No	Yes	45
37	Maybe	Yes	47
296	No	Yes	53

- Would you feel comfortable discussing a mental health issue with your coworkers?
- Do you *currently* have a mental health disorder?
- What is your age?

Column 1 - Comfortability: **Y**es, **M**aybe, **N**o
Column 2 - Claimed mental disorder: **Y**es, **P**ossibly, **D**on’t Know, **N**o
Column 3 - Age: **20’s**: 20-29, **30’s**: 30-39, **40’s**: 40-49, **50’s**: 50-59





Extra Sketch Notes

I1. Age -> Claimed mental disorder -> Comfortability

I2. The grid is filled with the ages that hold those responses: Columns - Claimed mental disorder, Rows - Comfortability

I3. Boxes: Comfortability, Box sections: Claimed mental disorder, Dots: Age

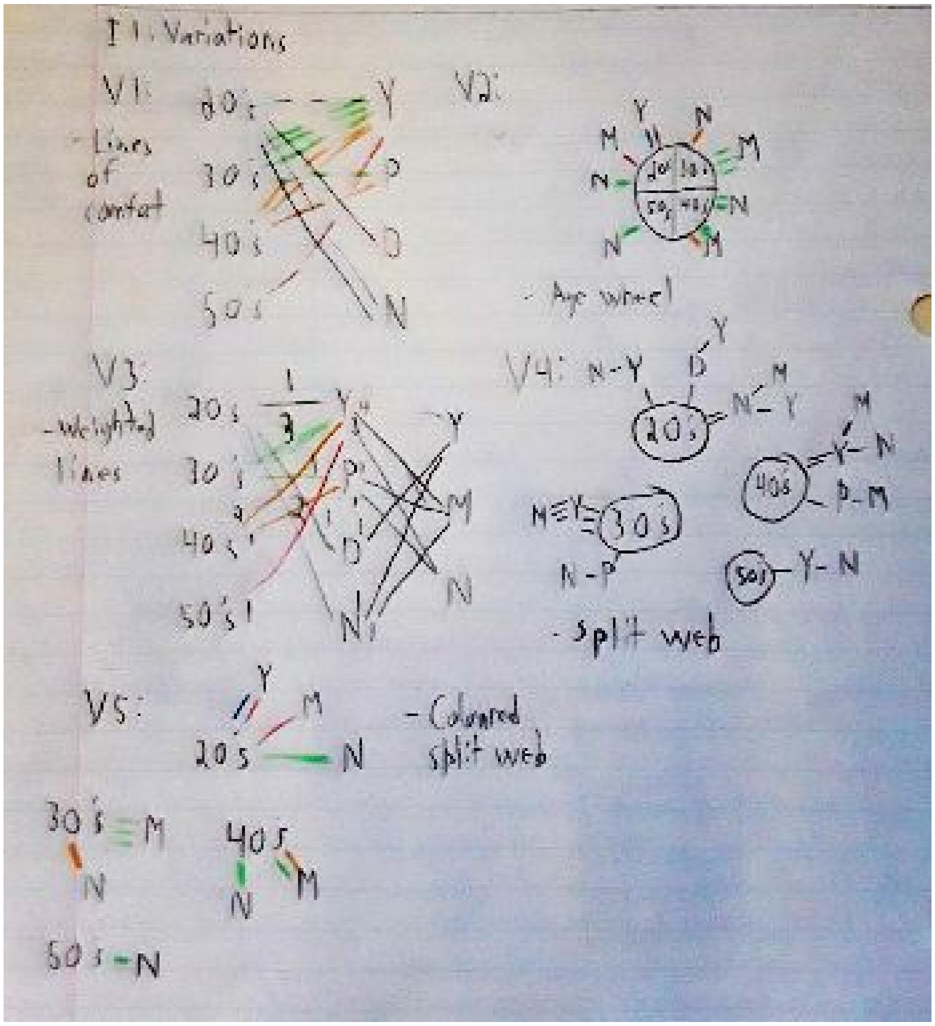
I5. Fans out into sections with the same ordering as I1

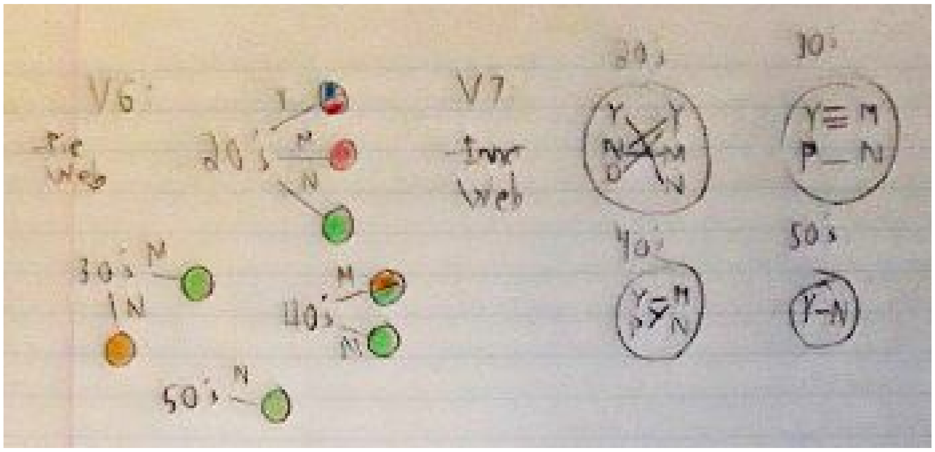
I6. The direction of the lines determines Comfortability. The shape determines Claimed mental disorder.

I7-8. Like I5 and I1 but the order is reversed. (Maybe = ?)

I9. Order is like I1. Similar to I6

I10. Similar to I3





V1, 3. Colors show which ages they come from

V1. More dashes = Less Comfortability

V3. “Sketch didn’t seem like a good representation”

V4-6. Order: Age -> Claimed mental disorder -> Comfortability

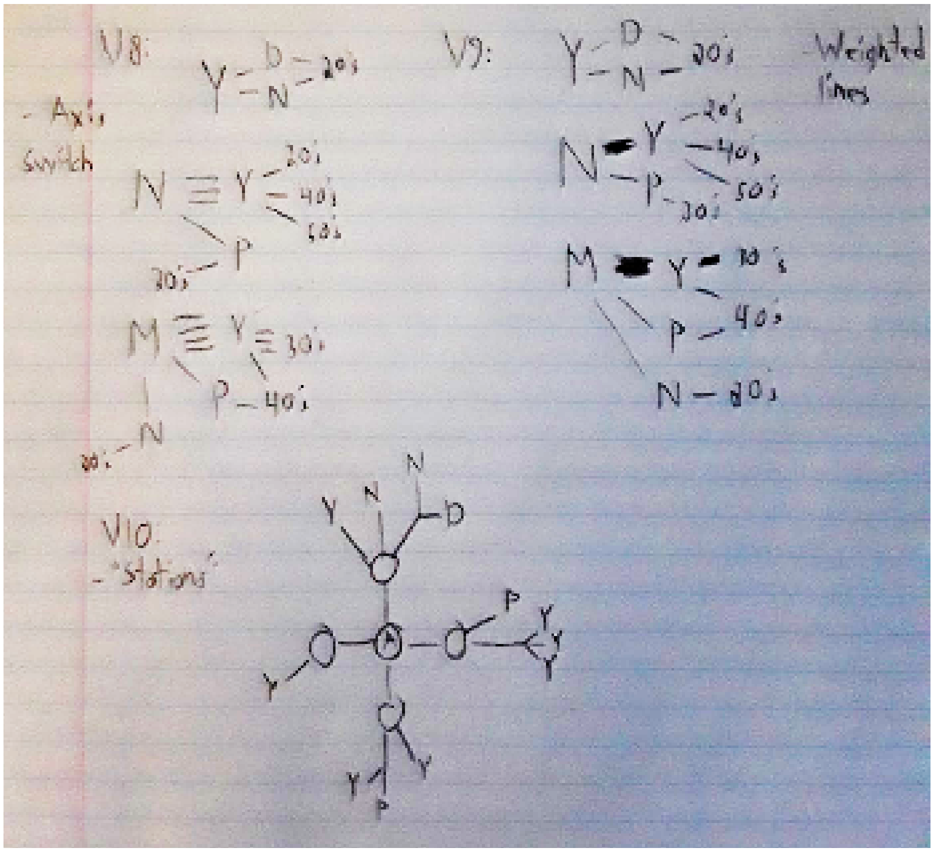
V5-6. Color = Claimed mental disorder

Blue = Don’t know

Red = No

Orange = Possibly

Green = Yes



V8-9. Order: Comfortability -> Claimed mental disorder -> Age

V9. Weight = # of responses

V10. Each circle is an age range. First branch is Comfortability. Second branch is Claimed mental disorder

3.3. Process

When it came to creating the data subsets, there were some issues when it came to selecting them. Going past 3 columns of data seemed hard to work with as the types of data would restrict what could be done if visuals. The nature of using a random set of data feels similar to that of a survey. However, a survey mainly relies on a very large group of data which means that the small data subset might not come to the same results as the entire survey. When it came to creating the sketches, trying to think of 10 different representations was a bit of a stretch as it felt like most options have been covered once 5-7 sketches have been made.

3.4. General Design Direction

The direction I want to take with my project is trying to illustrate the survey path the participants went through. The general goal of this project direction is to emphasize the “users find themselves” narrative pattern which sounds fitting for a survey dataset. By showing the question’s results in a similar manner to the survey, this could help make the data more relatable to the users as they get to see the other participants that had a shared response to them. The idea is to use a sort of “flow” style visualization that shows the various responses and questions that occurred during the survey.

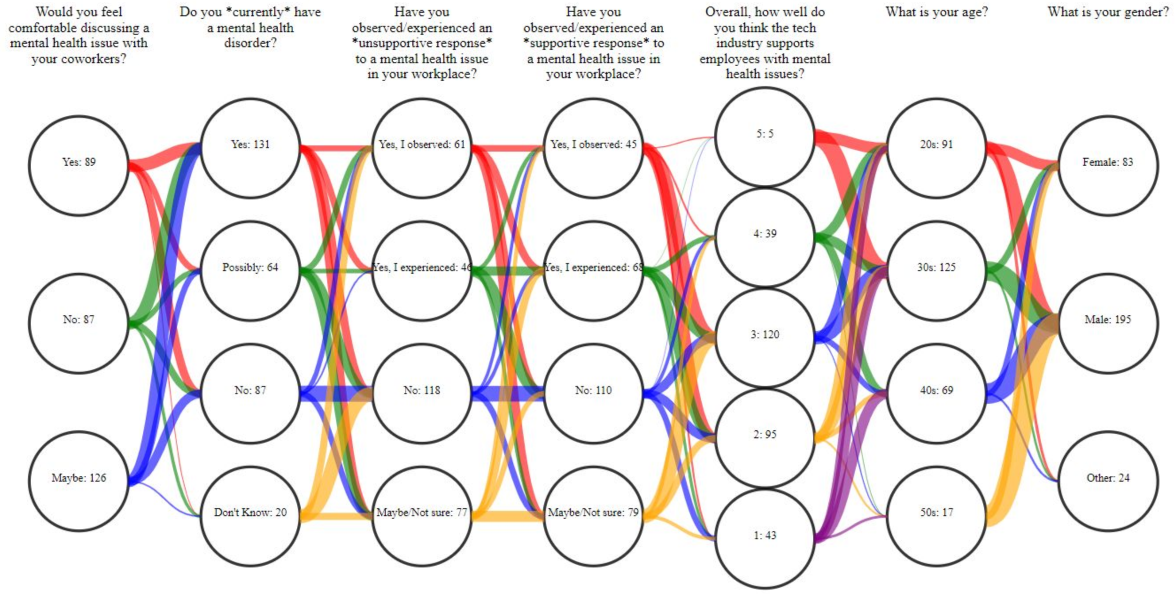
3.5. Prototyping Variations

General Design Concept

The general idea with these variations is to explore ways of showing the flow across the survey. By changing the way components appear visually on a surface level, I can find out what potentially works and what doesn’t when it comes to representing this dataset. Because a survey is a linear series of questions, the idea is to have the general attention of the user to flow in one direction and travel around from there as they see the variation of answers to each question in the survey.

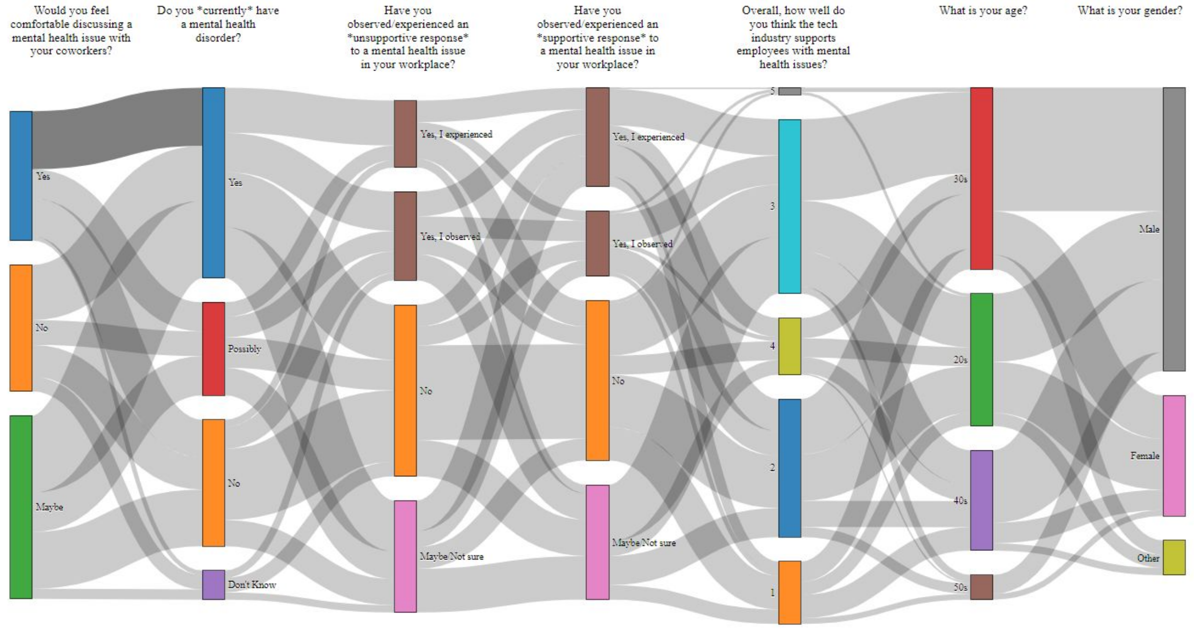
3.5.1. Variation A: Initial Idea

This first variation is an attempt at a direct interpretation of one of the sketches in the 10x10. During the creation of this variation, it was able to give some idea of what might not work in the final product. When representing flows, circles don't really work too well as nodes and to make sure the flows are clear, there needs to be a fair amount of space to visibly see everything.



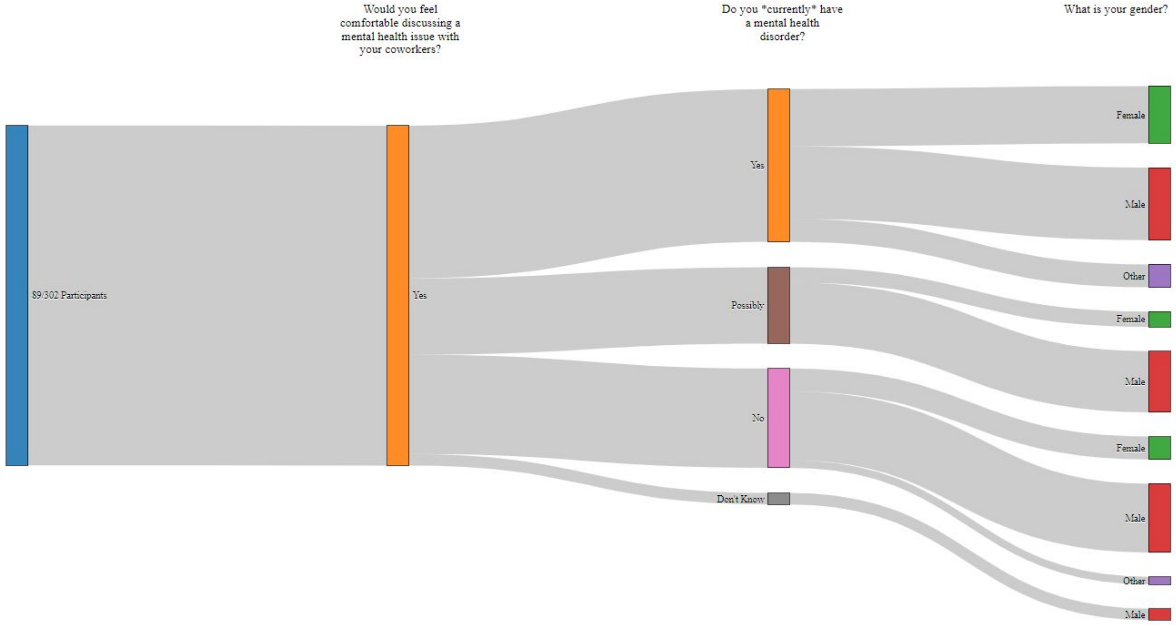
3.5.2. Variation B: Sankey Diagram

The linear feel of variation 1 feels like a good way to visualize the flow of a survey. To improve upon variation 1 would be to utilize a Sankey Diagram. By replacing the circles with rectangles, the visualization doesn't feel as tight or restrained which helps with visibility. One thing that this variation was able to explore that the first didn't execute well is node size as issues with space occur when trying to use circles as they have to expand in all directions.



3.5.3. Variation C: Dendrogram Snippet

A possible variation of the same step like variation 2. It is worth noting that this variation is just a snippet of data due to redundancy and the discovery of potential issues during creation. While the dendrogram may do better in showing the diversity of answers through the various decision paths, it might not look very good with this dataset. It might run into a similar problem with variation 1 where a lot of space is required to convey information that will inevitably be hard to read once all paths are implemented.



These variations can be viewed at this link in their respective folders:
https://pages.cpsc.ucalgary.ca/~nathan.cruz/CPSC%20583%20-%20Fall%202020/
(Variations B & C are based on this code:
https://bl.ocks.org/d3noob/d0212d9bdc0ad3d3e45b40d6d012e455)

3.6. Implementation Process

Through the process of implementing the representations and presentations of the dataset, several design ideas were affected. To start, most designs that I felt would be interesting would be scrapped as they're implementation is rather complex when converting their ideas into D3. This resulted in restricting my possible choices to more simplistic methods of representing that data. However, the use of more simplistic methods can make sure that the data that's being represented can come off as more understandable. Another point that affected what was implementable was the sense of scale. Like what was explored in variations 1 and 3, some design ideas do fine on a small scale like in the 10x10. Once more data is represented, sections that feel like they need space start to get drowned out by the larger sections of data. In a survey, most questions asked should play an equal part overall to get a good idea of what trends or correlations exist. In the end, this process of implementing representations and presentations helped me realize what aspect of my design ideas works better at larger scales and that having a rather simple representation isn't that bad of an issue as I perceive.

3.7. Final Static Design

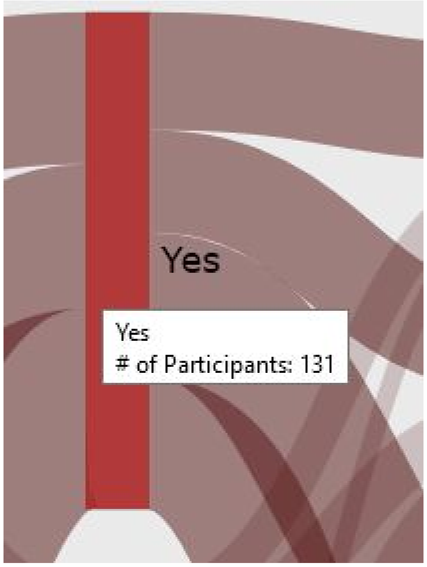
From feedback from my sketches, it would be an interesting idea to represent the path the participants took through the survey. While there isn't much variation during the prototyping phase, the main idea of representing the participant's survey path was kept in mind. Thus, the final implemented visualization will be based on Variation B. The use of rectangles over circles helps with visibility and its use of space is something that Variation C struggles to work with.

The main interaction the visualization will offer is the ability to filter out responses to see how many participants hold those responses. This interaction could be achieved by the use of selecting the rectangles, dropdown menus, or checklists. Some other interactions include filtering out entire questions or getting data surrounding a single question in the survey. This could be done through the use of extra buttons that are lined up across the question columns.

The main secondary information that can be placed in this visualization would be a summary of what the dataset is in order to give the user context of what they are seeing. This could take the form of a "information" button or a text box along the side of the visualization.

3.8. Prototyping Interaction(s)

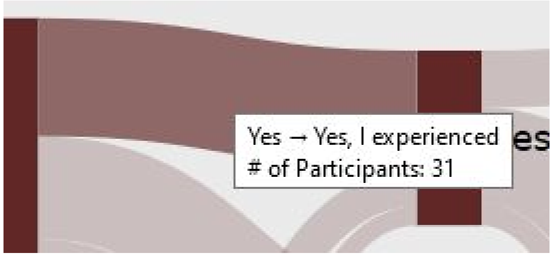
Hovers



Yes
# of Participants: 131

Nodes

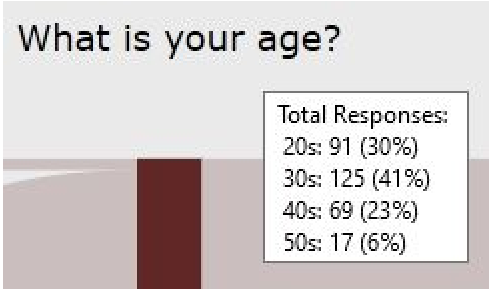
Hovering over a node highlights it and shows the amount of participants that hold that specific response. It will also highlight the paths/links that are related to it.



Yes → Yes, I experienced
# of Participants: 31

Paths/Links

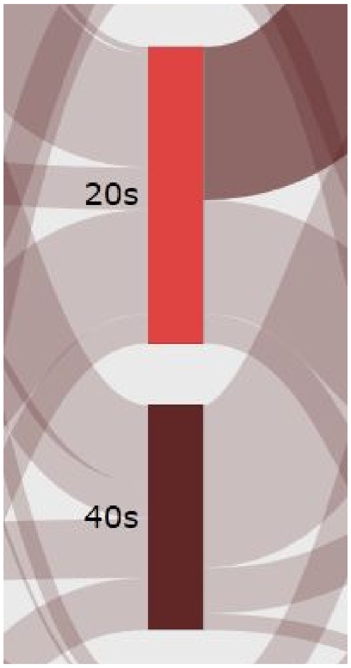
Hovering over a path/link highlights it and show the amount of participants that hold both the responses on each end of the path


What is your age?

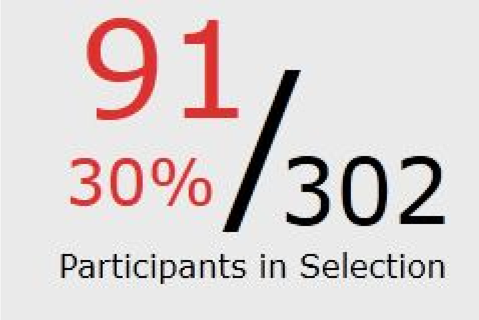
Total Responses:
20s: 91 (30%)
30s: 125 (41%)
40s: 69 (23%)
50s: 17 (6%)

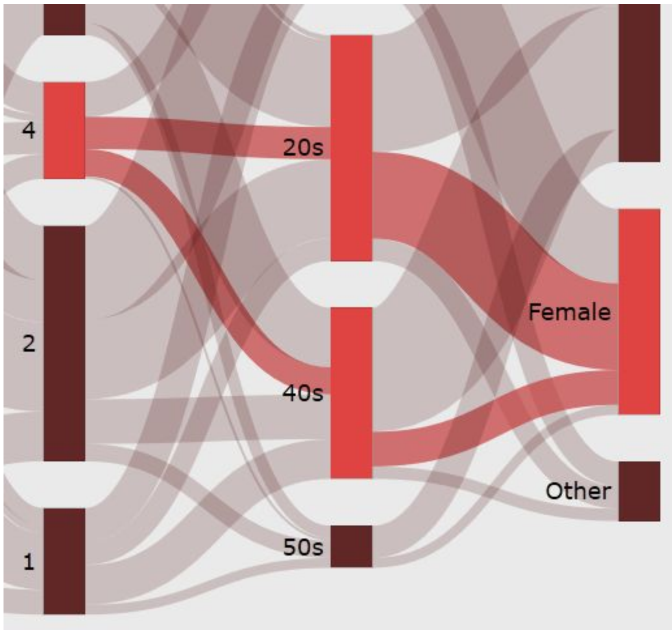
Questions
Hovering over a question shows the spread of the responses for that particular questions both as a number and percentage

Click (Main Interaction)



Nodes
Clicking on a node adds it into selection and updates the total number of participants that hold all the responses that are in selection. A selected node will stay highlighted regardless if it's being hovered over or not.
Clicking a selected node deselects it and reverts it back to be not highlighted. This will also update the total number of participants in selection as well.


91
30% / 302
Participants in Selection

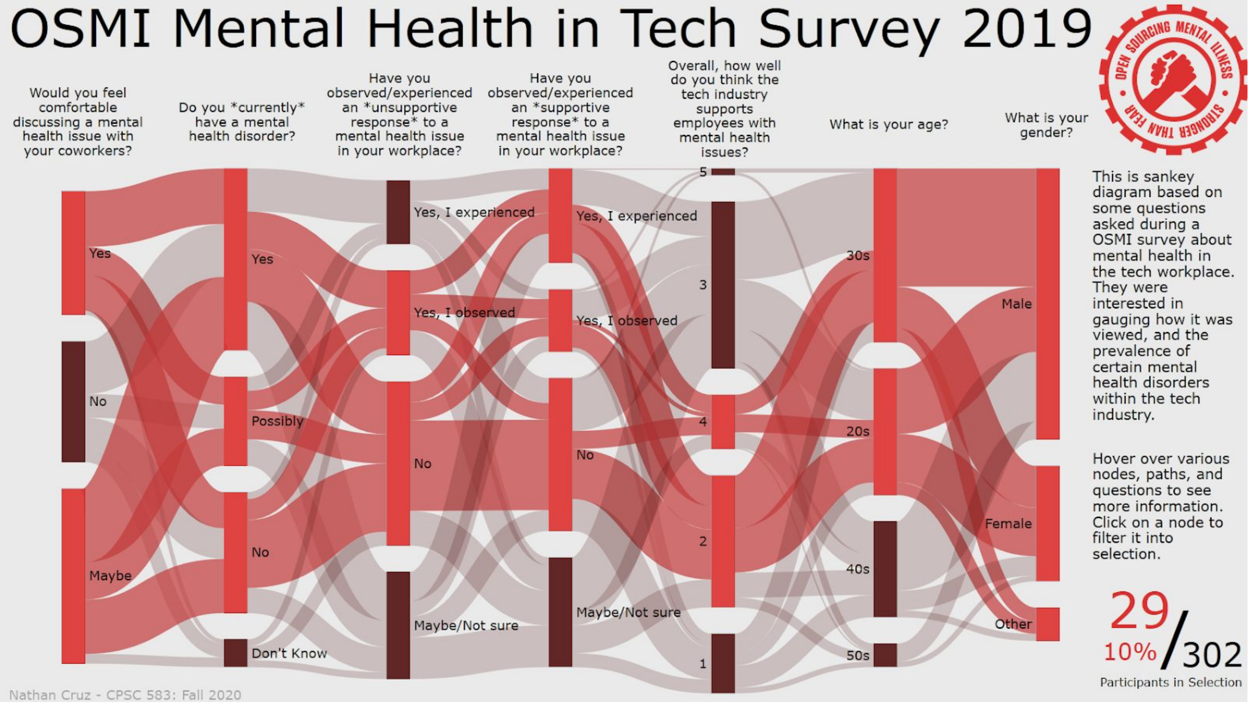


A Selection of Neighboring Nodes
If a selection of nodes are from neighboring questions, their paths/links will also be highlighted.

4. Final Implementation Process

This is the final implemented version of the visualization. Its interactions have been explored in the section before (3.8). It offers an overall view of all the responses in the survey shown as a sankey diagram. Users can select various nodes based on the response to the questions above them and can see the amount of participants that are in their selection both as a number and percentage. They can also hover over nodes and paths to get more information about the amount of participants who chose that response or hover over questions to see the spread of responses to that particular question. Should the user want to see the exact response of a participant in selection, they can find it in the “console” of the page (F12).

This could be used to help visualize trends or correlations between the responses of various questions or perhaps have the user go through the survey themselves and see who holds the same response that they do. Much like the purpose of the survey itself, users can get a better idea of how mental health is perceived in the tech workplace and discover how much people have different views of such a topic.



OSMI Mental Health in Tech Survey 2019

This is sankey diagram based on some questions asked during a OSMI survey about mental health in the tech workplace. They were interested in gauging how it was viewed, and the prevalence of certain mental health disorders within the tech industry.

Hover over various nodes, paths, and questions to see more information. Click on a node to filter it into selection.

Would you feel comfortable discussing a mental health issue with your coworkers?
Do you *currently* have a mental health disorder?
Have you observed/experienced an *unsupportive response* to a mental health issue in your workplace?
Have you observed/experienced an *supportive response* to a mental health issue in your workplace?
Overall, how well do you think the tech industry supports employees with mental health issues?
What is your age?
What is your gender?
Yes
No
Maybe
Yes
Possibly
No
Don't Know
Yes, I experienced
Yes, I observed
No
Maybe/Not sure
Yes, I experienced
Yes, I observed
No
Maybe/Not sure
5
3
4
2
1
30s
20s
40s
50s
Male
Female
Other
OPEN SOURCING MENTAL ILLNESS
STRONGER THAN FEAR
29
10%
302
Participants in Selection
Nathan Cruz - CPSC 583: Fall 2020

The final implementation can be viewed at this link: https://pages.cpsc.ucalgary.ca/~nathan.cruz/CPSC%20583%20-%20Fall%202020/Project%20Final/
(It is based on this code: https://bl.ocks.org/d3noob/d0212d9bdc0ad3d3e45b40d6d012e455)

4.1. Process Reflections

During the process of implementing the final visualization, what seemed to be minor design choices resulting in taking a fair amount of time and thought. The beginning of the implementation started with finding a color palette that worked with the visualization. In its initial stage, the background was of a dark blue and had the selected nodes and path be in orange to help emphasize the selection with the rest of the other nodes that were in grey. However later on, the color palette changed to a monochromatic red in order to go with the simple colored OSMI logo. The use of space was built to fit within the span of one screen. This was potentially inspired by visualizations that I've in the past that were rather confined to a small space but had a fair amount of information visualized.

Upon interacting with the visualization, it allowed the ability to compare and see correlations haven't been observed when directly looking at the data. One highlight of these interesting correlations include that only one participant holds all the most popular responses in each question. Another highlight would be that a participant who experienced a unsupported response to mental health in their workplace was still willing to claim that the tech industry does a good job at supporting those with mental health issues.

5. Discussion

Much was learned and discovered throughout the creation of this project. When it comes to datasets, surveys are an easy choice to pick. The data types that come from them are simple and straightforward to process which gives them flexibility in the most popular ways of visualizing data. However, the types of questions asked heavily affect what visualizations can be applied. As surveys tend to be a glimpse of a situation, it's hard to really utilize time aspects in their visualizations without having to look at a much larger dataset that spans across several years. Another problem that was clear in the early stages of the design process was the difficulty in trying to bridge together columns/rows of data. Questions in surveys overall make it hard to find interesting ways to overlap them as each part of an entry feels isolated from each other. This is unlike other datasets such as geographic ones which have a lot of variety of data types for a single point of data.

When it came to implementation, much that was desired to be put into the visualization had to be scrapped due to either a lack of knowledge or difficulty in implementation. The use of hovers felt overused due to the difficulty of implementing simple textboxes. The simple task of having a textbox that wraps around to fit its content is much harder to implement than it seems. Another part that had to be scrapped was the idea of letting the user go through the survey themselves in order to sort of "find themselves" in the group of participants. The implementation limits the ability to change the shapes of the nodes without having to affect other parts of the visualization. This could negatively impact the user's responses as the user would be able to see the larger response groups and may just choose them in order simply to "fit in" with the majority. Not having this aspect may hinder the intended interaction concepts of embodiment and experience but it can at least avoid a potential dishonesty problem that has been observed in a variety of psychology studies. Should more time be available or more experience with D3 be present, another interaction that would be built into the visualization would be the ability to compare and contrast two different selections. This is because some sources that look at survey data tend to compare two different response groups with each other which isn’t really present in this visualization.

6. Conclusion

Overall, despite my lack of experience with D3 and the limiting factors that the chosen dataset brought with it, I feel content with the final visualization. While the visualization seems simplistic in both design and implementation, it does help reveal trends and correlations that would have been hard to recognize in the original dataset. Although the dataset was hard to bridge questions together, I feel that the final visualization I chose felt like an appropriate choice outside the generic bar or pie charts that would have been the default for most survey datasets.