THE EFFECTIVENESS OF DIFFERENT ATTRIBUTES USED IN WORD CLOUDS

JULES CAZAUBIEL, ELAINE CHEN, JOHVANNI PEREZ, NICK TOURILOTT
(jcazaubiel@wpi.edu, echen2@wpi.edu, jperez@wpi.edu, njtourtillott@wpi.edu)

FINAL PROJECT CS 573 (SPRING 2022)

Overview and Motivation

Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

Word clouds are a visualization strategy used to represent data that can be pulled from articles, websites, or any kind of input data. The purpose of this study is to analyze which elements of a word cloud contribute to its effectiveness. Summarization techniques used in data visualization vary quite severely depending on the audience, the kind of data, and the method in which they are reading the information. We found that word clouds were an interesting method of visualization because of their representation of data for textual information, which is a form of information that is not as commonly looked into or represented. Word clouds can be created in a variety of ways, and due to different methods and attributes, their effectiveness can differ. In a day and age where less is more, word clouds are able to be seen as interactive and entertaining, while still representing a large amount of information that is able to be consumed in a short period of time.

Based on previous studies, we are able to look into completed experiments and test different elements. The attributes we looked into were bar chart frequencies, horizontal occurrences, vertical occurrences, downhill occurrences, uphill occurrences, varied colors, and grouped varieties. After doing research about previous studies completed across various age demographics, we thought that observing undergraduate college students would be the most interesting and accessible to survey. Because all the college students we would be testing are WPI students, we expected that they would be able to understand a scientific paper, and be even more interested in one relating to COVID-19. Additionally, a college student demographic would allow us to better gauge how levels of education might play a role in which attributes are the most effective versus the least.

Participants were expected to provide their demographics, so that we could better understand how gender roles, majors, and language processing levels may impact the interpretation of reading a word cloud. For example, some majors may be more inclined to read scientific documentation and research papers more frequently, allowing them to easily summarize and identify the important elements in a word cloud. However, other majors might be reading less complex or different types of papers, and want to utilize different attributes to create the most effective word cloud. On the other hand, English being an individual's native language might also impact their ability on how they wish to interpret the information of a word cloud, and they want to provide a more visually intuitive interpretation like bar frequency.

Word clouds are a method that are variable enough to be used across all age groups, easy to understand, and quick to decipher. However, they are a controversial concept in the data

visualization community, and it is worthwhile to explore methods for visualizing textual content as opposed to numerical information. After finding a variety of studies that look into testing the effectiveness of word clouds across different types of audiences, it provides a backing that they are able to be easily understood, and have an impact on audiences in some regard. Other studies look into methods that make word clouds more or less comprehensive, and our study is planning on examining these methods to draw our own conclusions. We are collecting quantitative data that helped us conclude on the most effective traits of a word cloud and identify why they are helpful. With the rise of data visualization growing rapidly, it is important to establish the most effective and fastest methods of portraying all types of information, such as the text that is portrayed in word clouds.

Related Work

Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.

In order to learn the effectiveness in different attributes viewed in previous studies, we closely followed and took inspiration from Hearst's *An Evaluation of Semantically Grouped Word Cloud Designs* done in Northwestern University. This study looked at a large variety of different word clouds and helped us identify the most effective ways to visualize our results. One of the most frequency and effective ways that Heart presents their results is by utilizing filter fields and side by side comparisons. Some examples of how they utilize this are shown below.

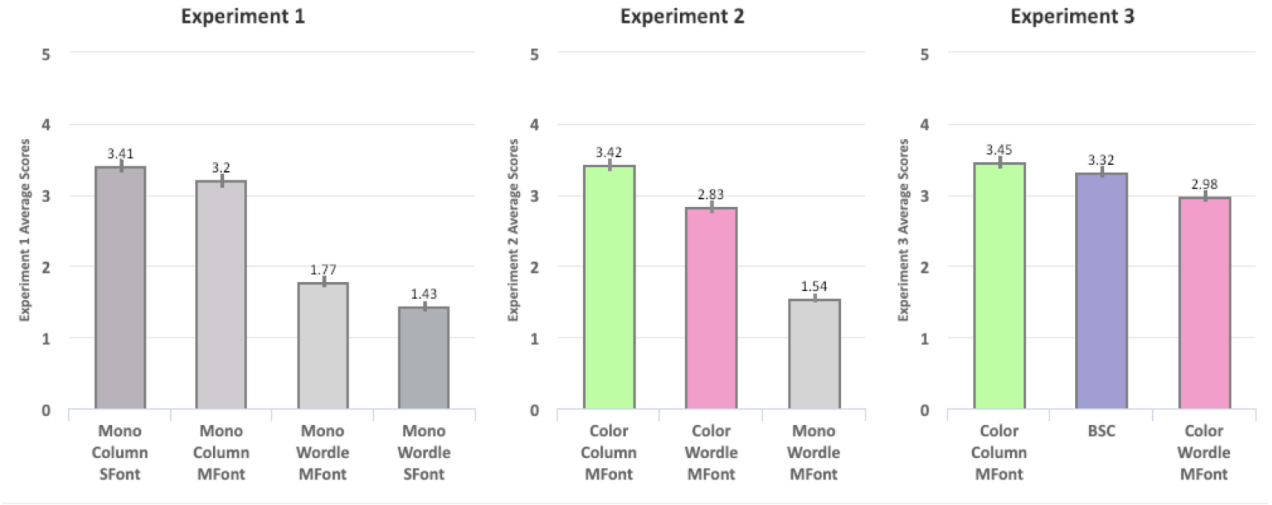


Fig. 4. Average scores and standard errors for the conditions of Experiments 1, 2, and 3. Bar colors indicate shared conditions across experiments.

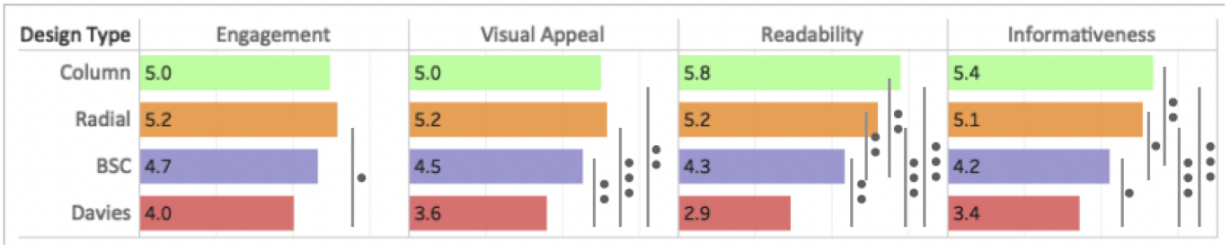


Fig. 9. Experiment 4 average ratings, (7 pt. scale, 1=poor, 7=excellent). Bars indicate pairwise significant differences; 1 dot corresponds to $p < 0.05$, 2 dots to $p < 0.01$, and 3 dots to $p < 0.001$.

This method allows the user to very easily identify the differences between the various experiments/variables being compared, and allows them to better grasp the results of the highest impact. Utilizing these methods, we decided to follow suit and create a visualization with a filter field as well that would allow the user to select which text they were reading and to observe the respective results.

Questions

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

- What attributes are the most effective amongst college students reading word clouds?
- Is there a correlation between English proficiency levels and the effectiveness of word clouds?
- Which attributes contribute to the effectiveness of word clouds in relation to English proficiency and English being a native language?
- Does major and educational background play a role in how textual information is able to be understood and preferred?
- Is there an attribute that is universally preferred amongst demographics?

These questions evolved over the course of the project because we were able to realize during the project that we had selected similar types of articles that were related to a biological field. Because of this selection of articles, there was an inherent bias to biological/life sciences majors. If we had done papers about different majors, it would have been interesting to see how these results might differ. New questions that we considered in the course of our analysis was trying to identify if there was a universal trait that was perceived more effectively across the board, rather than for a specific demographic.

Data

Source, scraping method, cleanup, etc.

SURVEY RESULTS:

major	gender	language	Horizontal_1	Bar_1	Vertical_1	Downhill_1	Normal_1	Color_1	Uphill_1	Horizontal_2	Bar_2	Vertical_2	Downhill_2	Normal_2	Color_2	Uphill_
Life Sciences	Woman	Spanish	2	1	3	3	5	3	5	3	1	3	3	4	4	5
Life Sciences	Man	French	2	2	1	2	5	3	2	2	2	1	2	5	4	2
Life Sciences	Man	English	3	1	2	4	5	5	4	3	1	2	3	4	4	3
Math/Physics	Nonbinary	English	3	2	4	3	5	5	3	4	2	2	3	4	5	5
Life Sciences	Woman	French	1	2	1	1	4	3	1	1	2	1	1	4	2	1
Business/Humanities	Woman	English	4	1	2	3	5	5	3	2	2	1	3	5	5	4
Computer Science	Woman	Other	3	3	1	2	4	4	2	2	3	1	2	5	4	2
Engineering	Man	English	4	1	2	3	5	5	4	3	2	2	2	5	5	3
Life Sciences	Woman	English	2	1	1	2	5	5	3	3	2	2	3	5	4	3
Business/Humanities	Man	Other	3	2	1	2	5	5	2	3	2	1	2	5	5	2
Computer Science	Man	English	3	2	2	3	4	5	3	2	3	2	4	5	4	3
Computer Science	Woman	English	2	3	4	1	5	2	1	5	1	3	5	4	1	3
Computer Science	Man	Other	2	4	2	2	5	4	2	3	4	2	2	5	5	3
Math/Physics	Nonbinary	Portugese	2	2	4	3	3	1	5	1	1	3	1	1	2	4
Engineering	Man	English	3	2	5	3	4	2	1	3	2	5	5	4	3	2
Business/Humanities	Woman	English	3	2	2	5	4	3	3	5	3	3	4	3	3	1
Engineering	Woman	English	2	3	1	2	4	4	2	2	3	1	2	5	4	3
Life Sciences	Man	Spanish	3	2	4	4	2	4	4	2	2	2	4	5	4	4
Engineering	Woman	Portugese	3	3	1	2	4	4	2	1	2	1	2	5	4	3
Computer Science	Man	English	3	4	4	2	3	1	1	5	4	3	2	2	1	3
Business/Humanities	Woman	French	3	3	3	2	5	2	1	4	1	1	3	2	2	3
Life Sciences	Man	English	1	2	1	1	5	4	1	1	2	1	1	5	4	1

Exploratory Data Analysis

What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

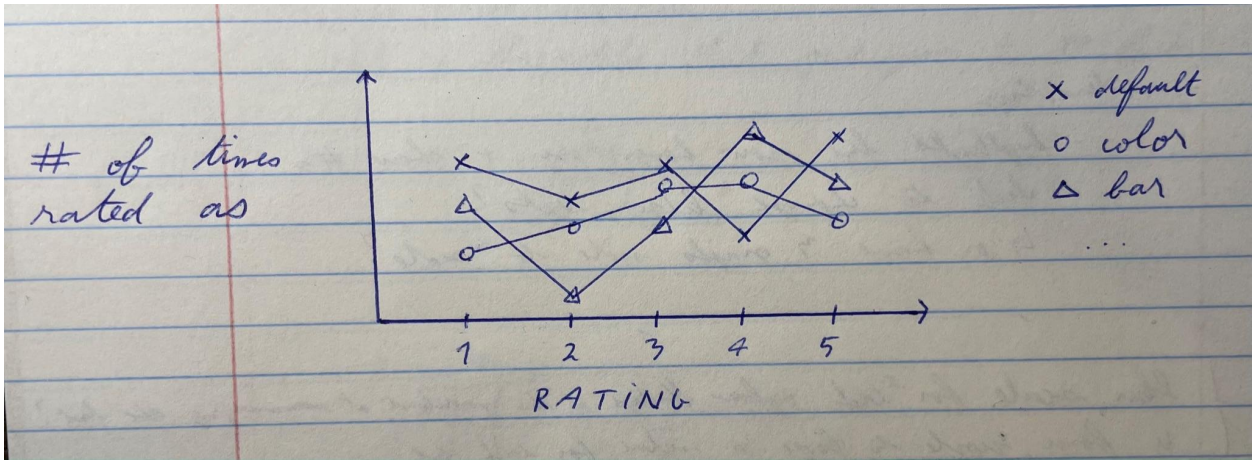
Initially, we had confusions about how to gather our information and the best way to portray our data. We originally tried to decide between two different methods, which we eventually decided to scrap and to prioritize having participants ranking all of the word clouds, which gives us more data to work and visualize. We were initially trying to collect data in the methods very loosely illustrated below, where colors represented the different attributes being tested.

However, after trying to identify the best way in which we can use to represent our information, we decided that having the participant rank each word cloud how effective they find it is more effective. This is because we would then be able to identify every participant's preferences, and also identify if there are any trends or similarities among each other or certain attributes.

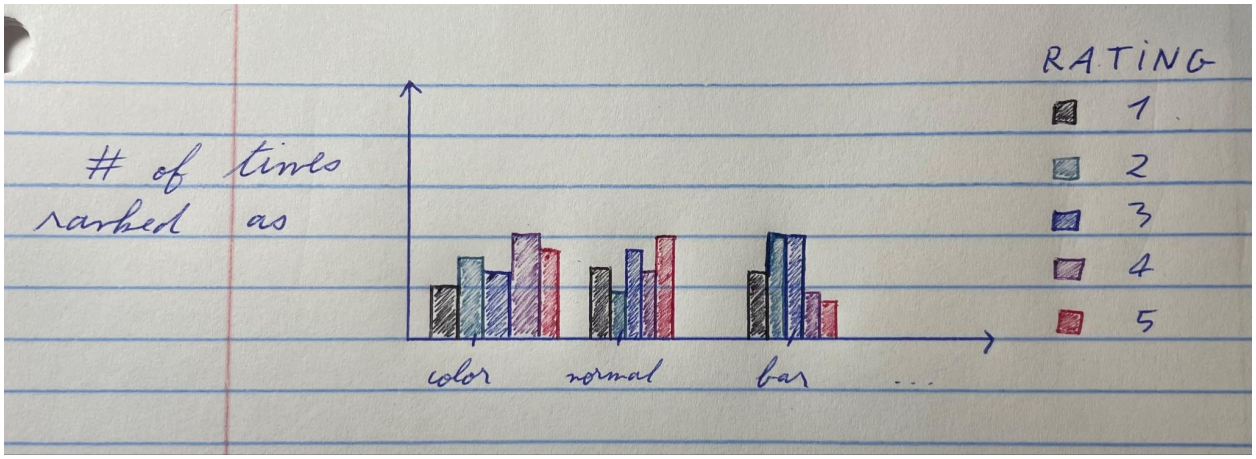
Design Evolution

What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

The different visualizations we considered included creating a scatter plot to show the frequencies of certain attributes, and how often individuals were selecting a “5” for a specific attribute and for a specific paper. This would have been done by then having different paper’s data be represented using different colors and plotting each data result.



However, we thought that this data would be best represented using a bar chart, and that utilizing filtering dashboards would make it easier to draw these comparisons than a scatter plot. This is because a bar chart is able to be easily compared side by side with each other like how our visualizations are done and utilizing the filter helps the reader test and see how these results might have differed depending on the reading, or if it did. We did not deviate from the proposal's initial purpose or work, but we deviated slightly how we were going to collect our data.

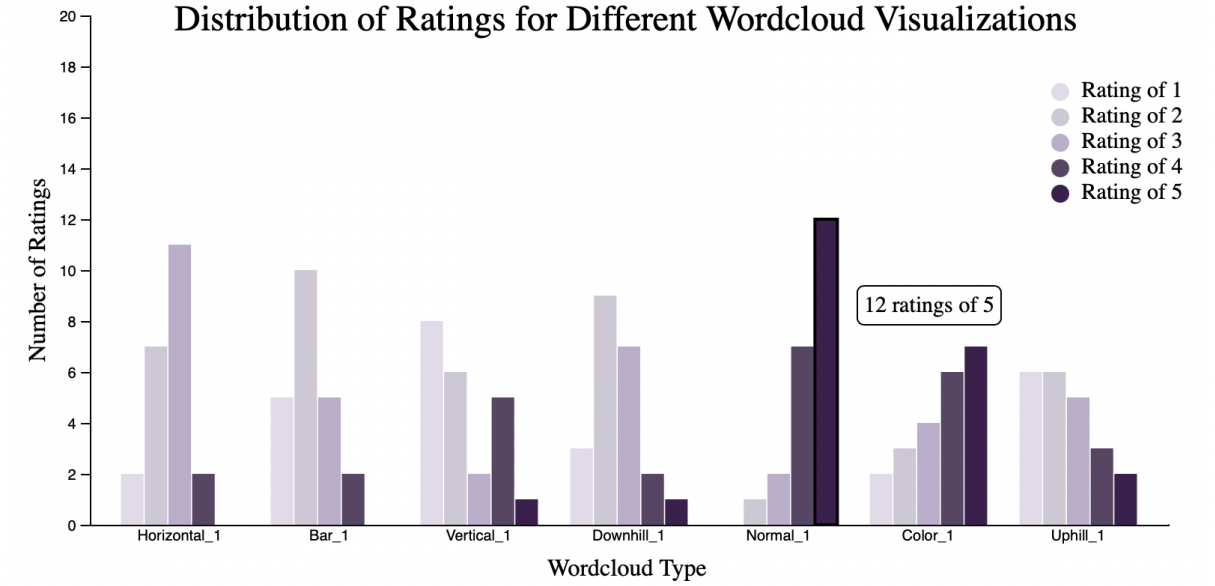


Implementation

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements

Our implementation of our data visualization includes a bar chart with a filter dashboard that allows users to select the results associated with each reading they read. This filter is created to ensure that the reader is not overwhelmed by the results and see how results might have differed or stayed the same depending on the reading. Our implementation also shows the results and frequencies of rankings of each attribute. This was done to prove and showcase how different

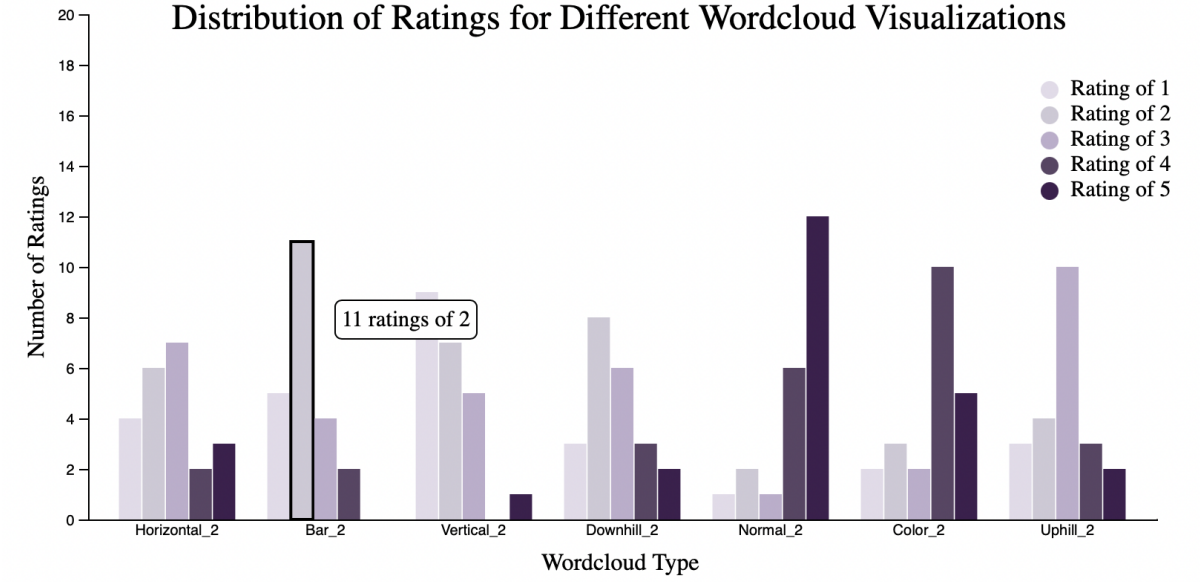
attributes compared against each other and allows the reader to see which attribute was the most preferred for each reading. We also utilized the different colors to represent each attribute's scoring.



Text filter:

1

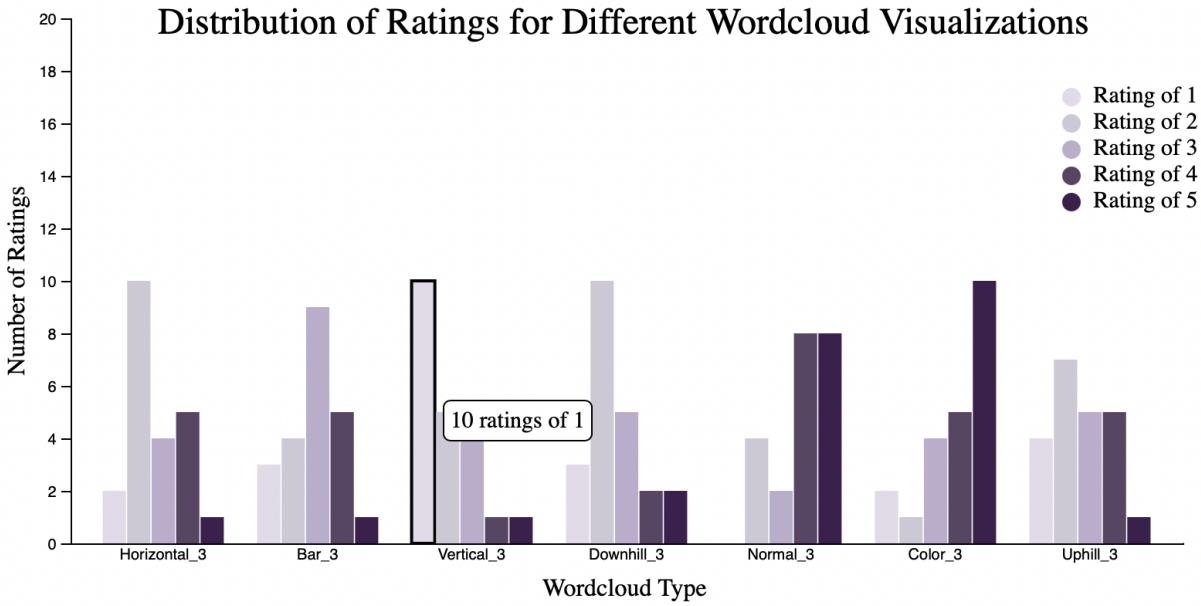
Results from Reading 1 with Hover



Text filter:

2

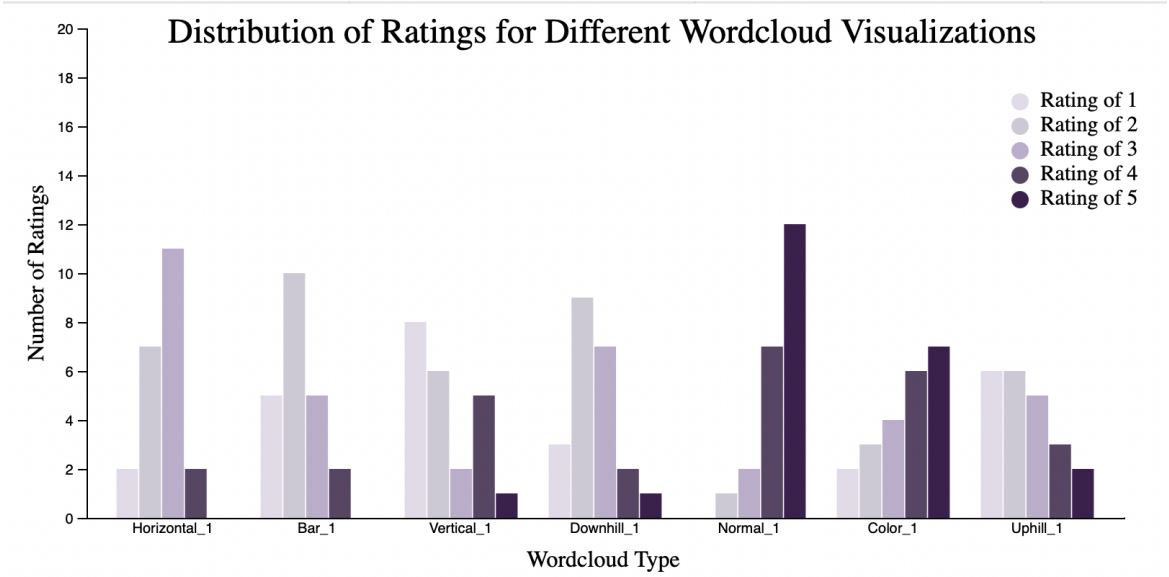
Results from Reading 2 with Hover



Text filter:

3

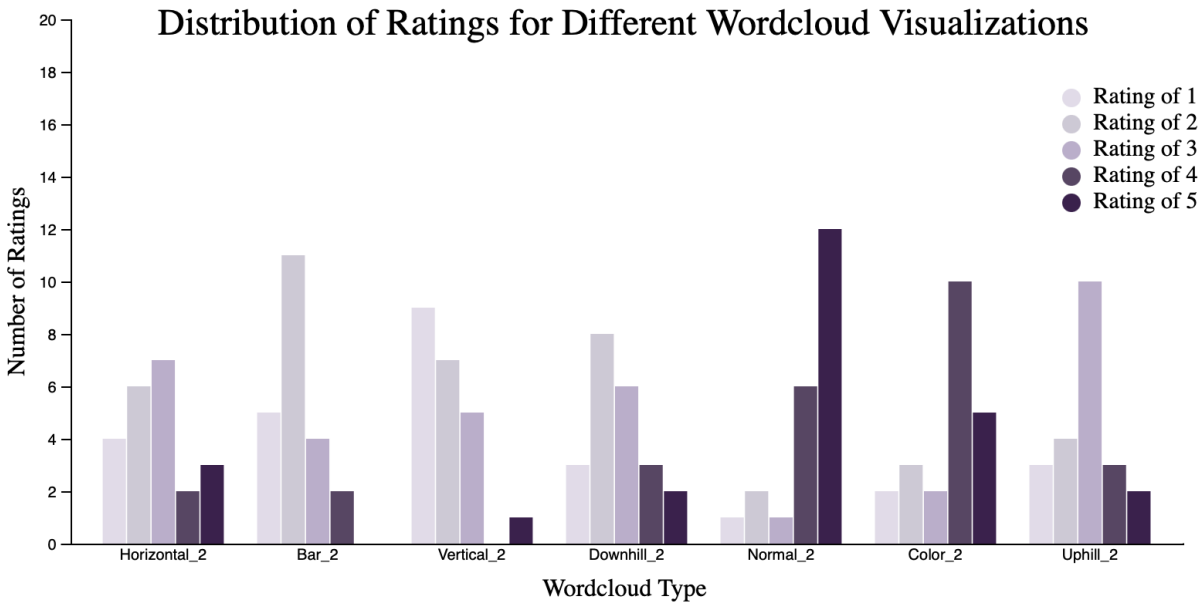
Results from Reading 3 with Hover



Text filter:

1

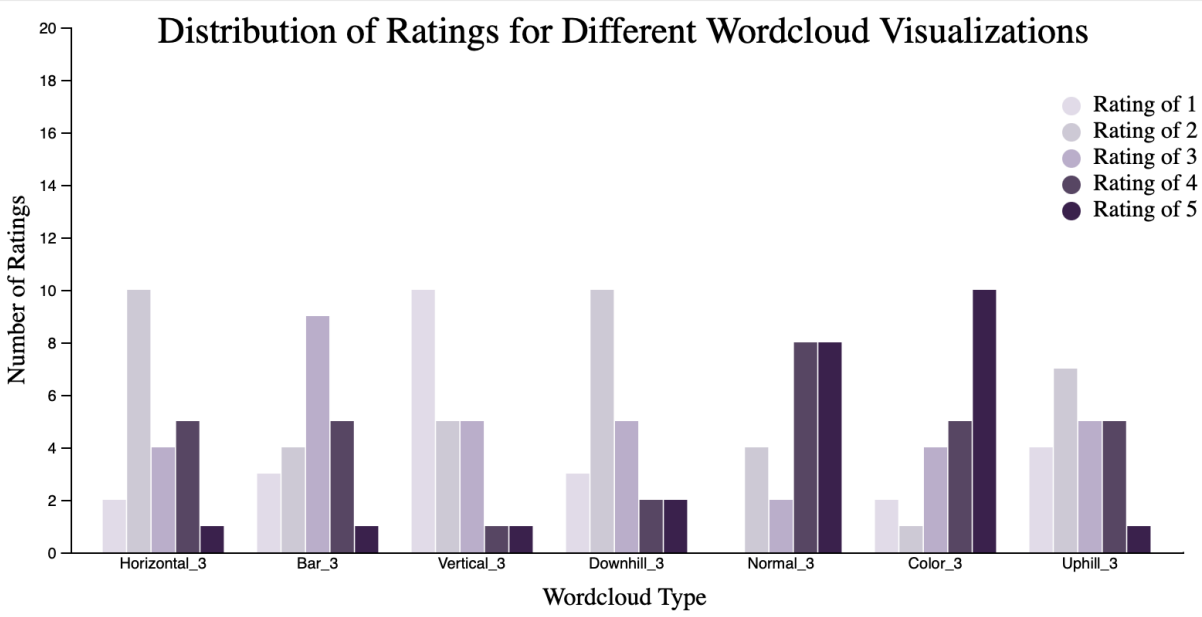
Results for Reading 1



Text filter:

2

Results from Reading 2



Text filter:

3

Results from Reading 3

Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

Across the three different readings, it appears that color was the most preferred attribute amongst the word clouds. While the most favored attribute differed depending on the word clouds, we found that color is universally preferred depending on the readings. However, depending on the readings, the most favored attribute appears to have been the normal attribute, which has one color and a variety of sizing in text. We also found that those who have English as their second language dislike the bar method, which is the opposite of what we had assumed. Those who were non-native speakers appeared to like the normal word cloud, where the word cloud is one color and text sizes would vary. This was an interesting finding, considering we had initially though that varied color might be easier for non native readers to find important parts to. We also found that relating to majors, there was a strong correlation between engineering and life sciences to prefer multi colored visualizations, in comparison to the other methods regarding positioning. This was also an interesting observation because it allows us to understand which attributes might be more effective for different educational backgrounds. Our visualization works well in its ability to easily read and understand the information being presented. It is able to easily change its results based on the reading selected, and the results from the rankings are clearly indicated using a cohesive color scheme. To further improve our visualization, if we had more time and resources, we would want to try to be able to isolate certain rankings and when selected, have the majors who selected that ranking to appear. This way, it would be easier to identify what demographics selected each ranking. This interaction would make the visualization tell more of a story than a sole representation of data.