Visualization of Climbing Data

Authors

Torin McDonald (MS Candidate Computer Graphics and Visualization)
Tianlong Zhang (MS Candidate Electrical and Computer Engineering)
Pranav Rajan (Third Year Candidate BS Computer Science)

Table of Contents

Project Proposal

Basic Info

Title: Climbing Policy Visualization

Members:
Torin McDonald, u0940253 (torin@sci.utah.edu)
Pranav Rajan, u1136324 (u1136324@utah.edu)
Tianlong Zhang, u1209611 (tianlong.zhang@utah.edu)

Github: https://github.com/tdmcdonald/dataviscourse-pr-climbersurveyvis

Background and Motivation

For our project, we chose to visualize data from a survey conducted by the Salt Lake Climber's Alliance (SLCA). The SLCA is what is known as a Local Climbing Organization (LCO), a non profit designed to organize, educate, and advocate for climbing communities and climbing access. From a public policy perspective, the climbing community is unique in that it is largely self governed through social norms, culture, and clear in-group identities. The purpose of the SLCA's survey is to not only profile the climbing community in the greater Salt Lake area and their interactions with the SLCA, but to explore how this community views specific climbing related issues, from allowing dogs at climbing crags to bolting ethics. Because of the explosion in the popularity of climbing, along with an increase in the number and diversity of climbers who are challenging traditional climbing norms, it is more important than ever to understand how the community views climbing specific questions in order to direct future policy and land management. Developing a visual framework to explore the relationships between certain demographics of climbers and their view on climbing policy can help organizations like the SLCA plan for the future. In addition, this type of data has the potential for exploring interesting visualizations for multidimensional categorical data.

Project Objectives

We want our project to explore two different components of the data. The first objective is to give a broader overview of the data and how answers are distributed. The second objective is to design a visualization that allows the exploration of correlations in the answers. For example, the type of question that we hope our visualization will answer would be along the lines of how do climbers who are predominantly trad climbers view

re-bolting ethics versus climbers who are predominantly sport climbers? Being able to quickly answer questions like this one through a visual interface can help the SLCA reach out to specific groups in the climbing community to try and come to a resolution on a specific issue, or develop policies for specific climbing areas.

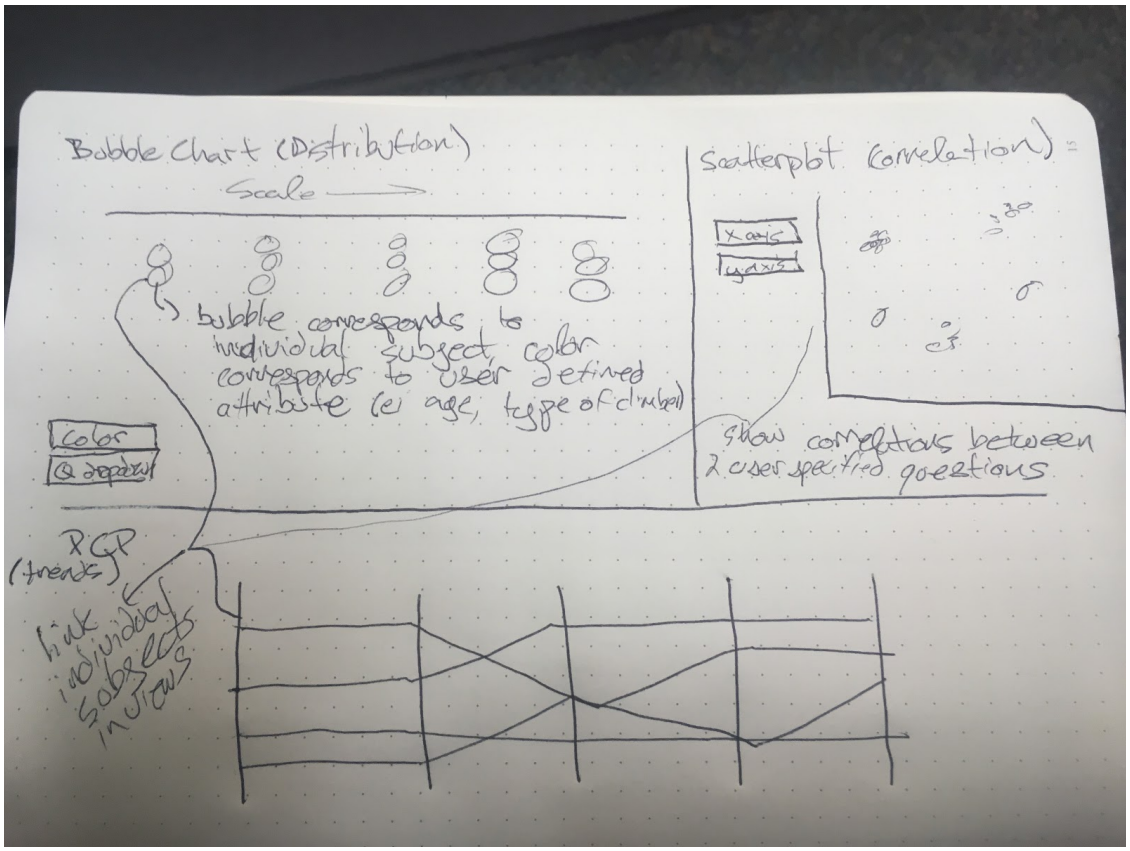
Data and Data Processing

We obtained our data from David Carter, an assistant professor in the Public Affairs Program in the Dept of Political Science at the University of Utah. Along with the SLCA board, he designed the survey and gave a summary of the results in [1]. He has given us access to the raw data from the survey for this project. He has also requested that the raw data not be distributed beyond our group and the teaching staff for the class. The data will require some preprocessing. Because the questions have different formats, we will have to come up with a data structure for each format. We plan on doing the preprocessing in python and outputting a JSON file.

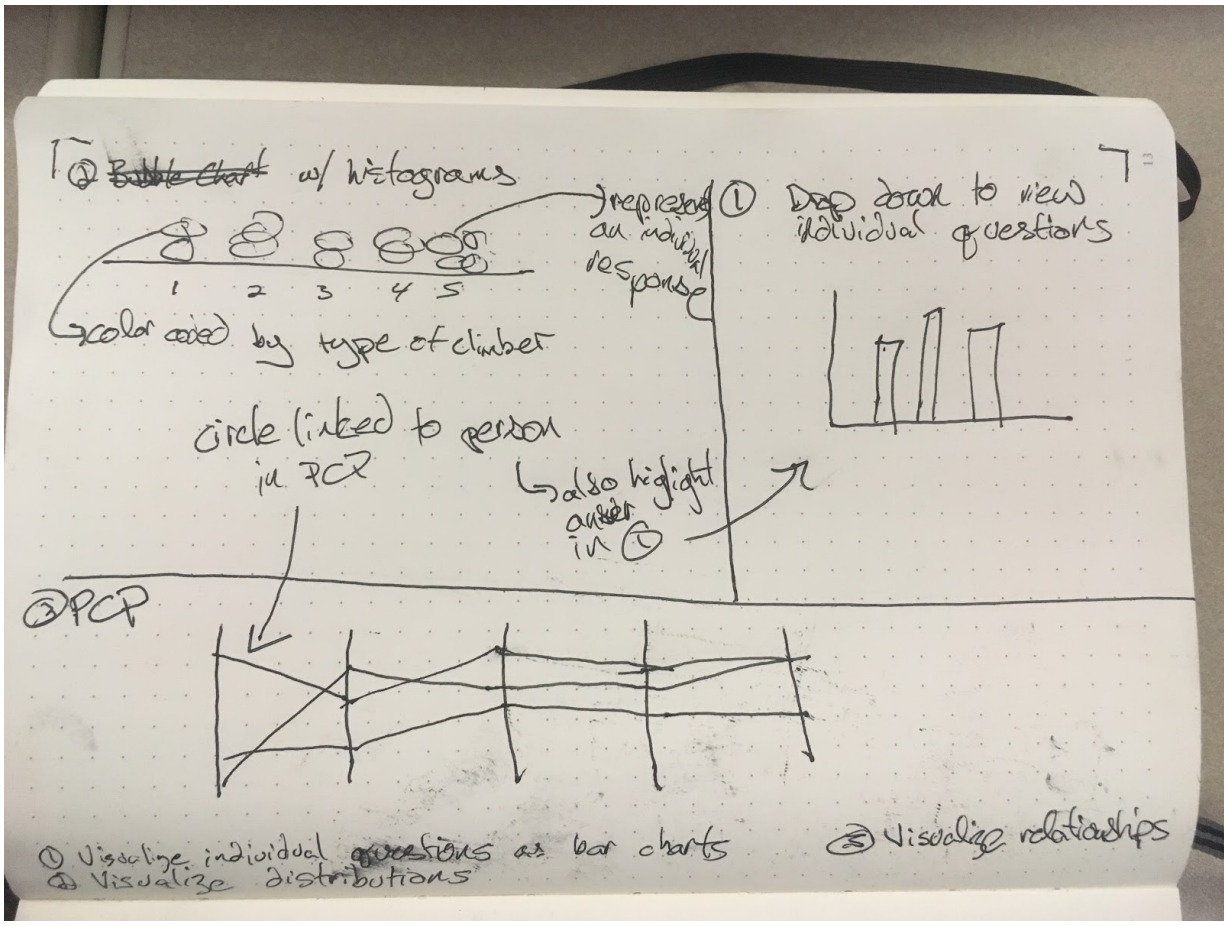
Visualization Design

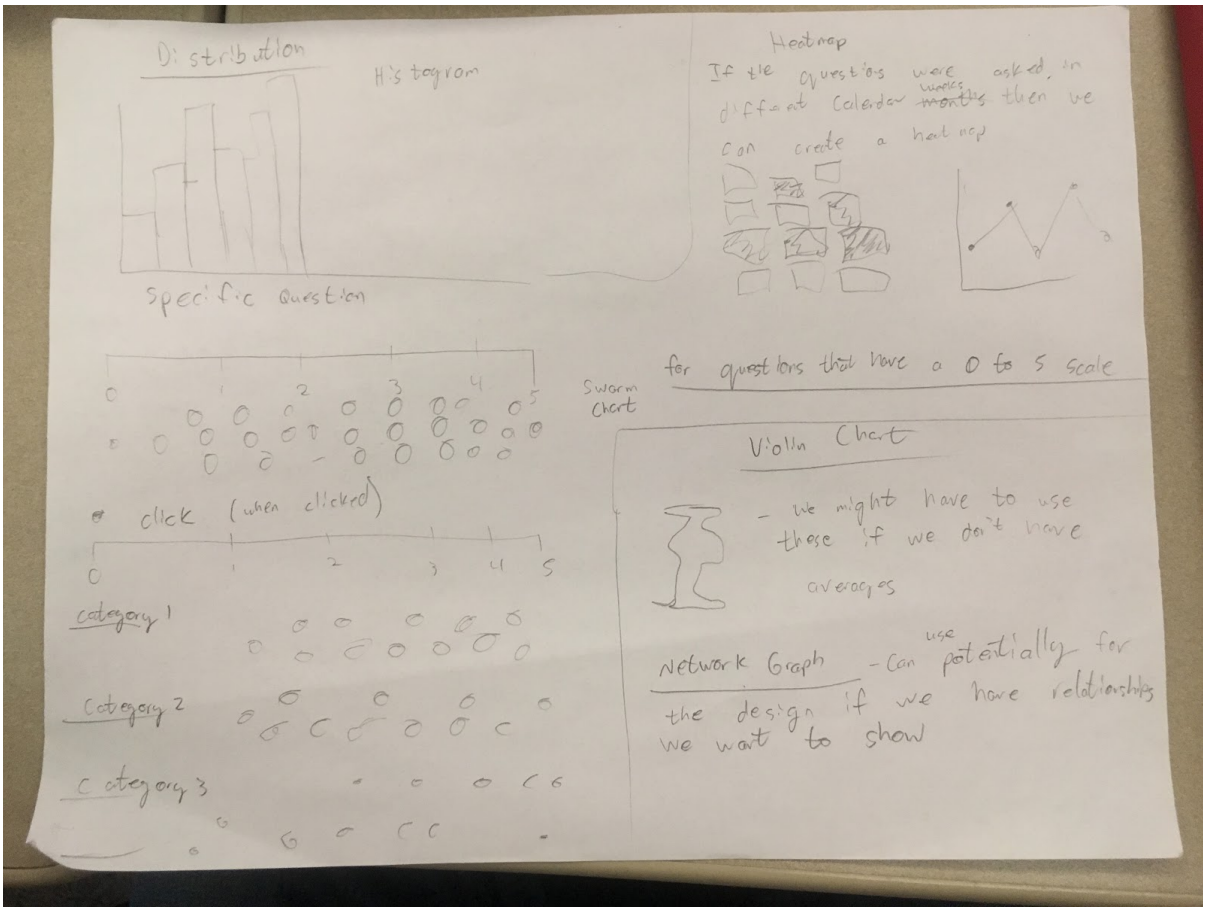
For our visualization, we want to combine three different views: A bubble chart that shows the distribution of data for specific answers to questions, a parallel coordinates plot to explore relationships between answers to different survey questions based on the subjects who took the survey, and a scatter plot that also shows the relationships between answers between two user specified questions to explore correlations and how answers cluster. The views will be linked together for further exploration in relationships in the data. For example, the user will be able to hover or click on a circle in the bubble chart, and then the line corresponding to the circle's subject in the parallel coordinates plot will be highlighted.

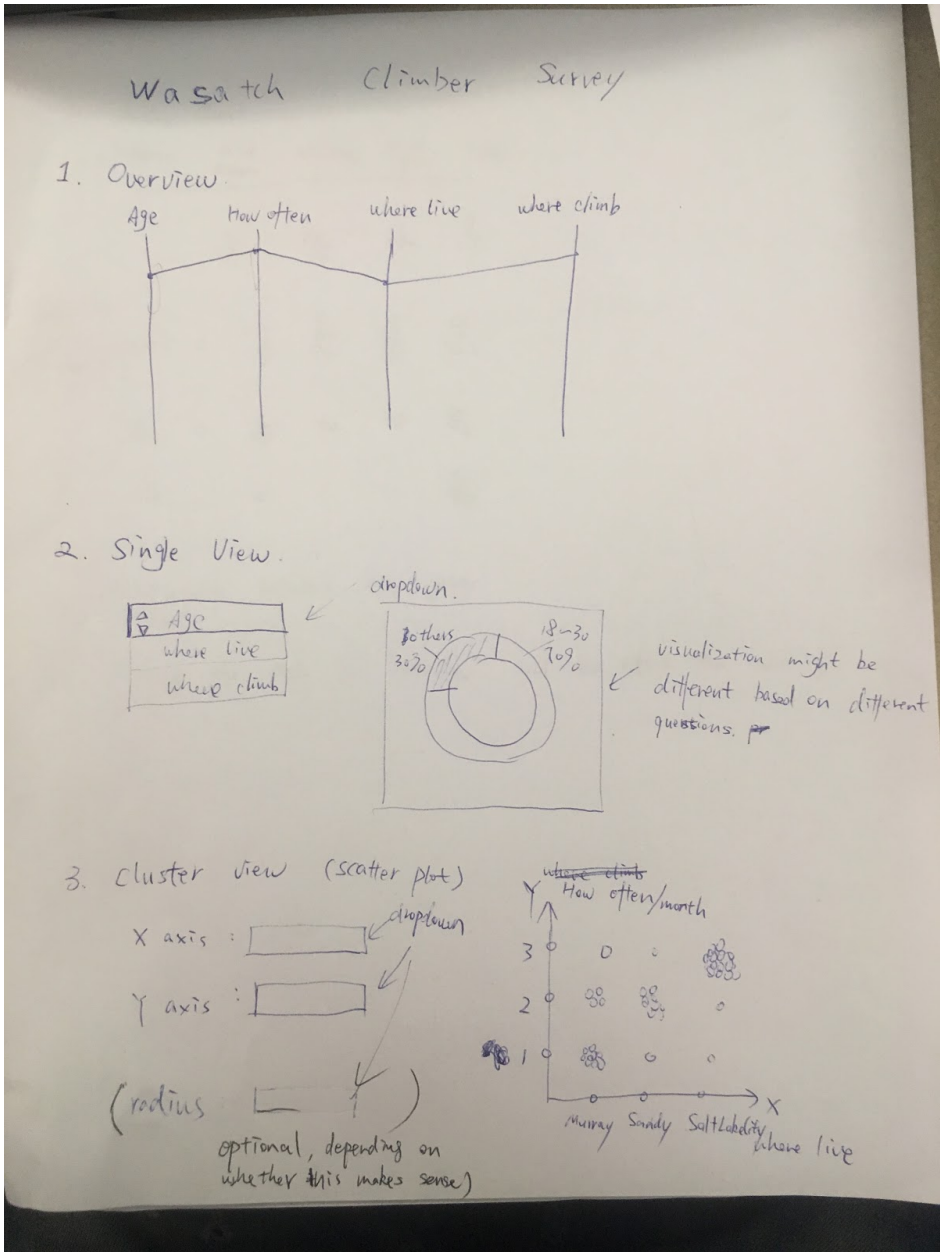
Final Visualization Design

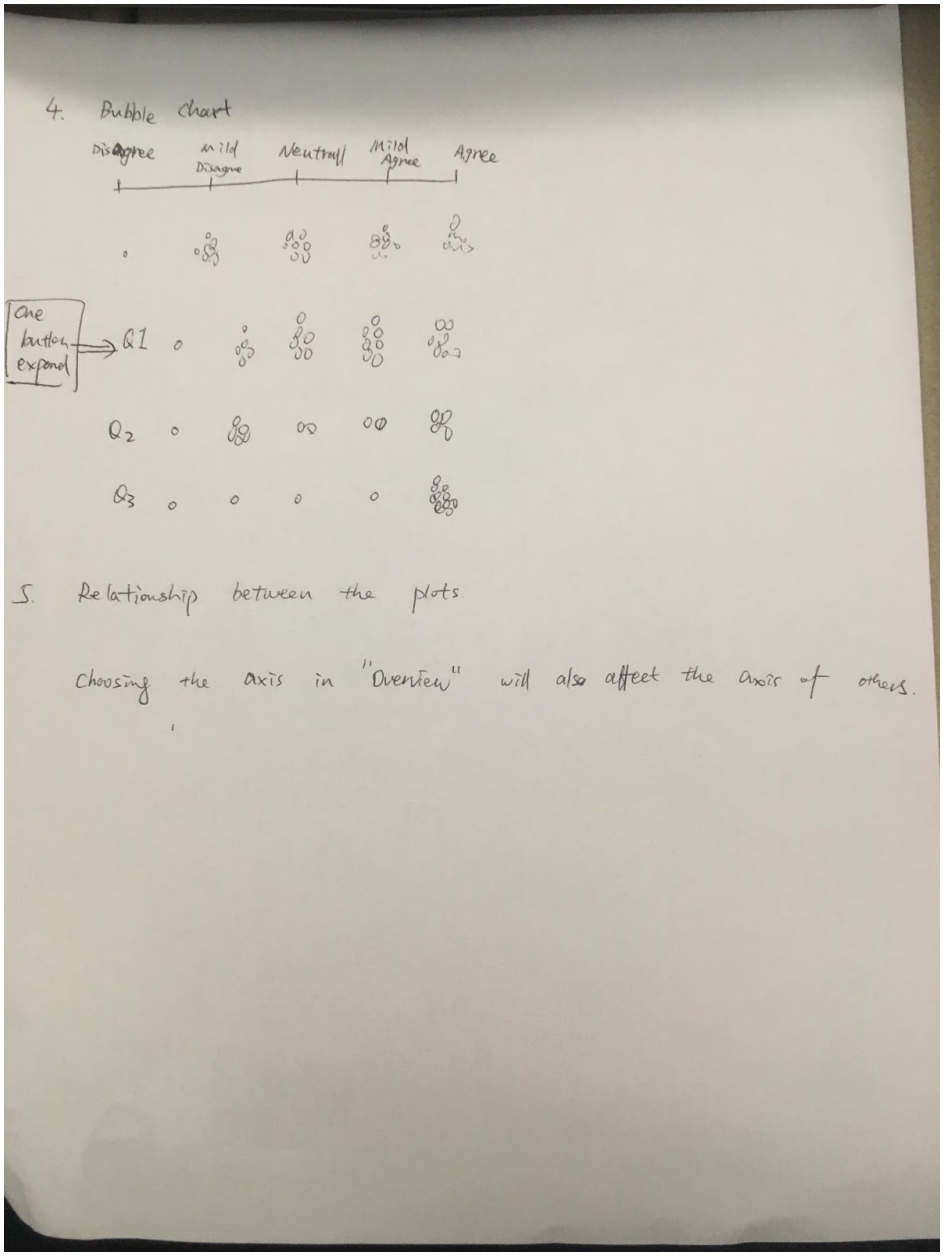


Alternative Design Prototypes




Distribution
Histogram
specific Question
Heatmap
If the questions were asked in diffent Calendar months then we can create a heat map
Swarm Chart
for questions that have a 0 to 5 scale
click (when clicked)
category 1
Category 2
Category 3
Violin Chart
- we might have to use these if we don't have averages
Network Graph
- Can use potentially for the design if we have relationships we want to show


Wasatch Climber Survey
1. Overview
Age
How often
where live
where climb
2. Single View.
dropdown.
Age
where live
where climb
Bothers 30%
18~30 70%
visualization might be different based on different questions.
3. cluster view (scatter plot)
X axis :
Y axis :
dropdown
(radius
optional, depending on whether this makes sense)
where climb
How often/month
3
2
1
X
Murray Sandy SaltLakeCity
where live



Must-Have Features

- Parallel Coordinates Plot
- Scatter plot with user defined axises
- Bubble chart to show distributions of answers
- Links between the views to help explore specific subjects answers or groups of subjects answers

Optional Features

- A fourth view for showing the answers to a specific question
- Brushing for linking the views together
- A view that shows some type of dimensionality reduction or clustering for the entire dataset.

Project Schedule

Nov 1: Finish all data preprocessing and design data structures (group)
Nov 8: Submit project milestone (group)
Nov 15: Full draft of views: PCP (Torin), Scatterplot (Tian), bubble chart (Pranav)
Nov 22: Link views and stylize (group)
Nov 27: Submit project

References

1. Carter, David P. 2019. “2019 Wasatch climber survey summary report.” Salt Lake Climbers Alliance: Salt Lake City, Utah.

Peer Evaluation

Peer review session notes:
During the class we had the peer review session with **‘Uber Moves’** group. Their members are Dayanidhi Tandra, Tianyu Li, Mia Ngo.

First half of the class they reviewed our project, and their comments are the following:
Dayanidhi Tandra:

1. One possible way to handle ‘yes/no’ question in beeswarm plot can be: set yes to 1 and no to 2, corresponding to 1-5 scale questions.
2. When do the beeswarm plot, maybe it is a good idea to add another table chart next to the beeswarm plot, just like hw6
3. How to encode color for this project?
   Answer: we decide to use color for some demographic properties like age and gender, and these properties will be present throughout the whole project.

Mia Ngo:

1. What do X axis and Y axis represent for the scatter plot? Why don’t you guys use bar chart?
   Answer: A. axes will be chosen by the users (user-driven), and those are (some of) the questions answers. Therefore the axes are not fixed. B. we did not choose bar chart because we want to show the cluster of the data, so we can see the relationships between two

questions, and also the demographic properties. Also some questions are answered in discrete answers rather than sequential numbers, so bar chart is not a good idea for this data structure.

2. What's your plan for storytelling?

   Answer: we plan to do an extreme case button just like hw6, but this is optional since we also wants to find the relationship between questions which means we currently do not have any idea of what is the extreme case. Also we want the project to be user-driven so the users can find whatever information they want.

Tianyu Li:

1. What is your plan for the data structure since this is not a sequential dataset?

   Answer: we would like to add all of the attributes to one data point. But we will treat the demographic properties as a basic attribute that will be encoded as color for every plot, and then use other attributes as the axes or properties for all of the plots.

TA Proposal Feedback (Jen Rogers, PhD Candidate SCI Institute)

This looks good! Really good background and motivation. You have a good overview of the domain space and the motivation is clear and concise. Your first objective is a little vague. What do you mean by broader overview? What attributes are you using in the data to plot the distributions or are they distributed for each question? Will this overview change when a user selects a different question or attribute?I think the questions are great. I would define 3-5 example questions and use those consistently through the design process. The background and motivation section has 2 questions - whether climbers that self identify if distinct groups have any difference in opinion regarding bolting ethics or tolerance of dogs at the crag than other groups. I think this is great - I am interested in seeing if you could go through the questionnaire and define a set of 3-5 climbing community concerns. Bolting ethics, dogs at the crag, opinion on stick clips, involvement in adopt a crag, gear stashing, chopping bolts, use of wag bags (not sure what you have in the questions so I'm rambling) would be options. The more you can define specific use case questions the easier it is to build a tool for them. For
the data - do you have demographic information separate from the questionnaire? The design sketches look good! I would encourage you to use a visual hierarchy and make one view the dominant one with secondary visualizations.

Implementation

Data Processing

The survey that the data was collected from contained thirty three questions. In coordination with Professor David Carter, our team decided which questions would yield meaningful results for the SLCA and climbing policy. As a group, we went through the survey questions and

selected questions that we thought could be used as a characteristic of a type of climber and which questions could be used for visualizing aspects of the data based on the climber type. The following questions were selected for defining climbers for the visualization:

Climber Definition Questions

1) Question 1 (Climbing Experience)
2) Question 2 (How often do you climb)
3) Question 3 (Favorite Climbing Location)
4) Question 5 (Climbing Gym)
5) Question 6 (How did you learn to climb)
6) Question 7 (Type of Climbing)
7) Question 8 (Climbing Motivation)
8) Question 23 (Age Group)

Climbing Visualization Questions

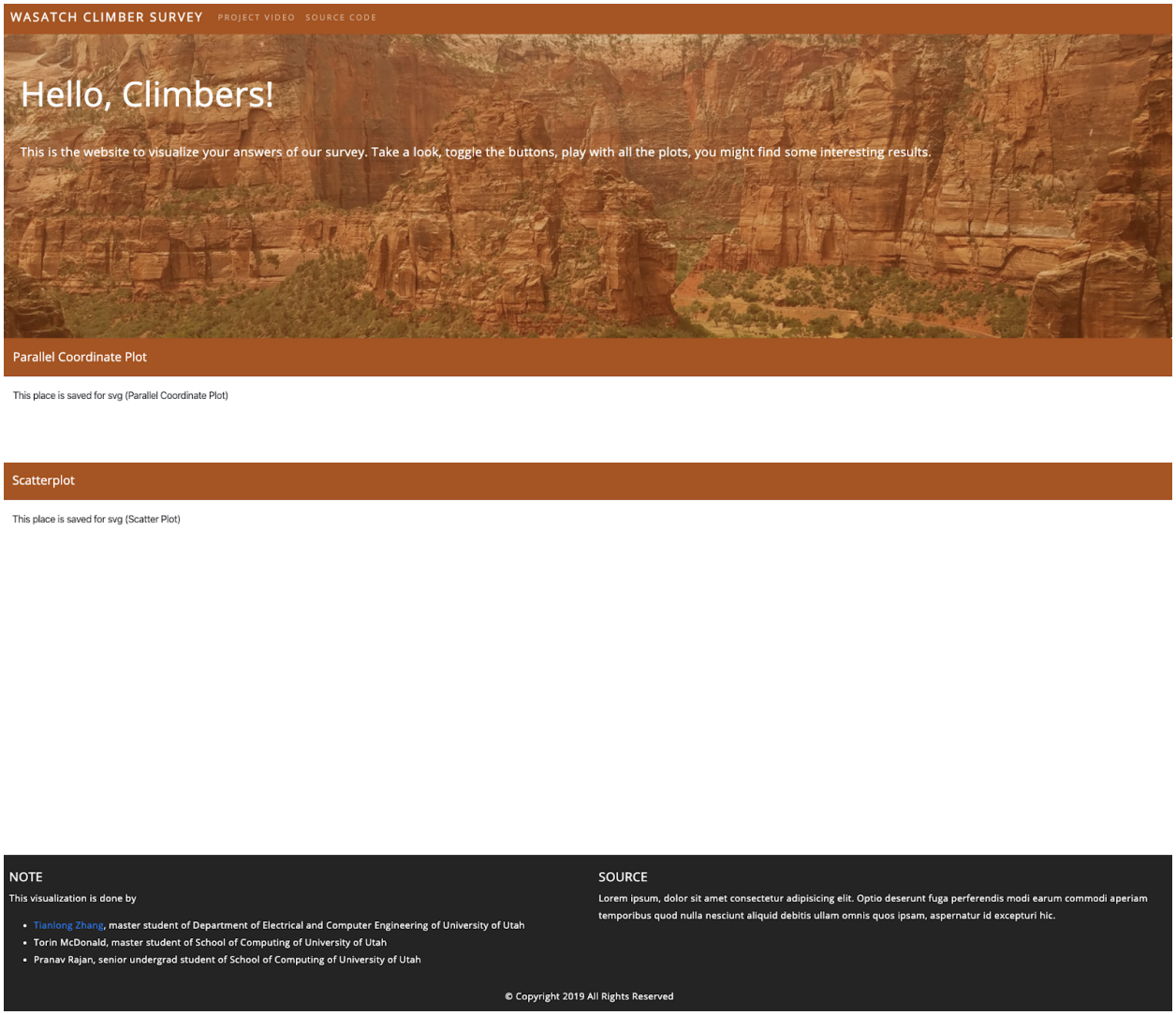
1) Questions 12a and 12b (Climbing Policy Opinions)
2) Questions 16a and 16b (Climbing Impact Concerns)
3) Question 17 (Climbing Behavior)
4) Question 18 (Toilets in the outdoors)
5) Questions 21a and 21b (Contributions toward Climbing and the Outdoors)

Data Structures

For structuring the data for the different views we created objects made up of lists based on the responses of users. For each property that defined for a climber we assigned a list that would store the responses of the user. For some of the questions, users were allowed to select multiple options and to account for this we used lists.

Website Design

The whole idea is to use Bootstrap to make the layout for our website. We choose Bootstrap version of 4.3.1 to do the implementation.
Our group decides on the style, color and fonts etc.., and Tianlong Zhang does the layout implementation.
By November 8 (milestone submission deadline), our layout for the website looks like this:



The structure is based on 4 parts: navbar, jumbotron, cards, footer.

For navbar, we use two links button to link to the project video (for future final submission) and our source code.

For jumbotron, the image is obtained from unsplash.com, they have awesome pictures but free to download. In order to fit the style for the whole page, we add an overlay layer on top of the image.

For the cards, our idea is to show all three plots by default, and users can not toggle whether they what to hide any cards. Since A, we want them to see all three plots, and B, there might be no need to collapse and toggle the plots.

For the footer, we have our team members information and the source information.

Credits: https://www.saltlakeclimbers.org/ and SLCA. Our color and style choice is based on their website, since we are just computer science student, we focus more on the implementation and visualization of the data.

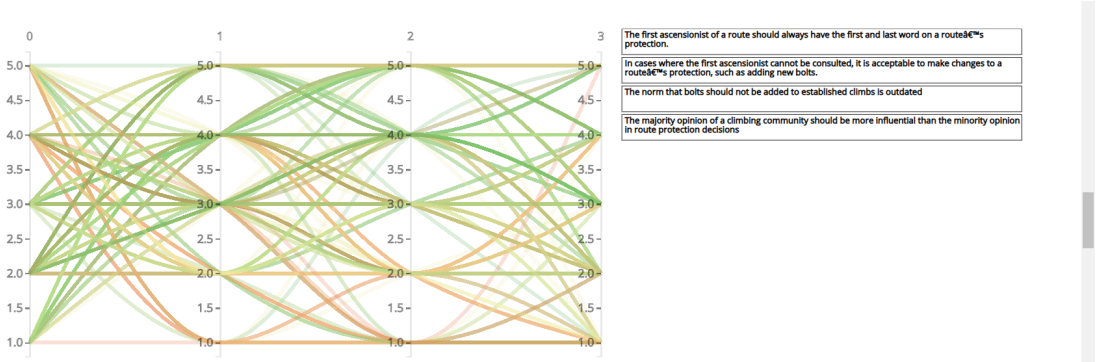
TA Milestone Feedback (Jen Rogers, PhD Candidate SCI Institute)

The website is looking good!
I think it is good that you included the feedback from the proposal and the peer session, but you should have some kind of description on what pieces of that you are considering in your design and how you are changing the design (if you are changing it).

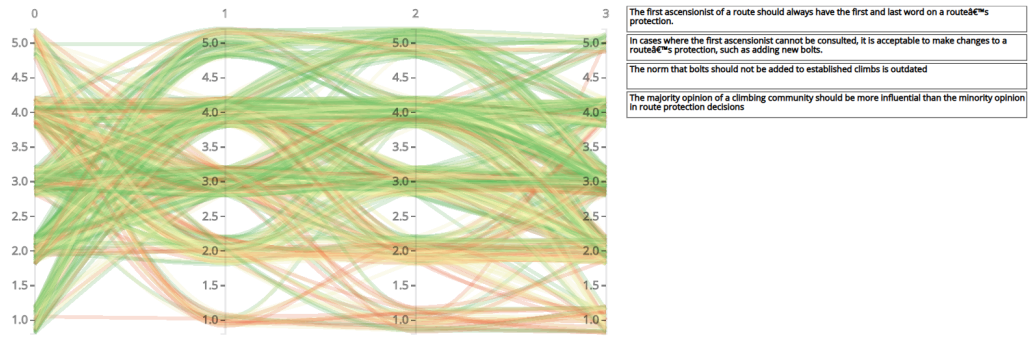
After meeting with you guys, I think the parallel coordinate plot is a good candidate for the main visualization. The scatter plot and distribution view are good supplementary views. Think of 2 or three questions you have about the data and use one for highlighting in the application on load. This gives the user a starting point to dive into the data and counts as a storytelling aspect.

Parallel Coordinates Plot Visualization (Torin McDonald)

The parallel coordinate plot (PCP) part of our visualization is designed to give the user insight into trends or relationships in how the survey responders responded to a sequence of questions. There were a few design considerations that had to be made in order to make the PCP usable. First of all, because all of the data generally falls in between 1 and 5 or 1 and 4, the initial PCP view looked like:

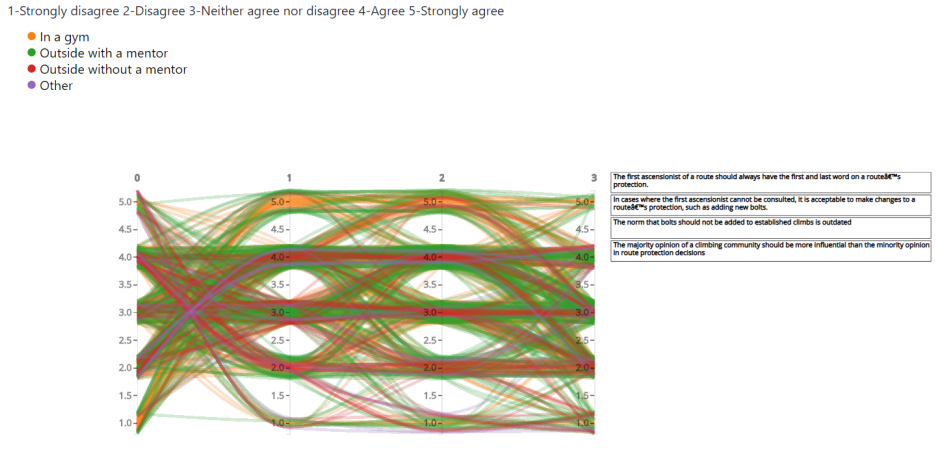


To overcome this problem, I took two different approaches. In order to give spread to the data, I initially adjusted the values to randomly fall within .3 of the actual value. This accomplished spreading the data out in the pcp. To improve this spread and avoid problems inherent in randomly changing the data, I adjusted the values to be stacked evenly within .3 of the value. Furthermore, I added a curve to the paths in order to better visualize the ‘flow’ of responses. The result was this:

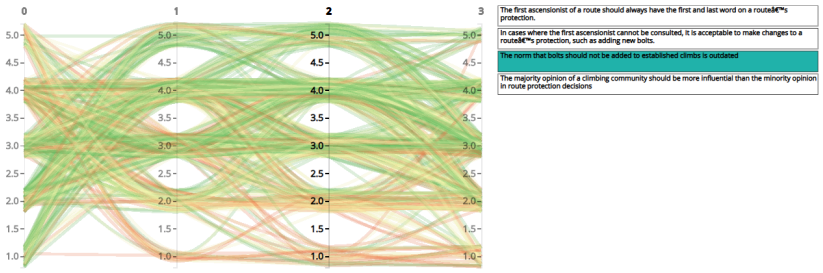


In addition, the user has the option of either color coding the PCP according to whether the respondent was predominantly a sport or trad climber (above), or by the value of the response (below).

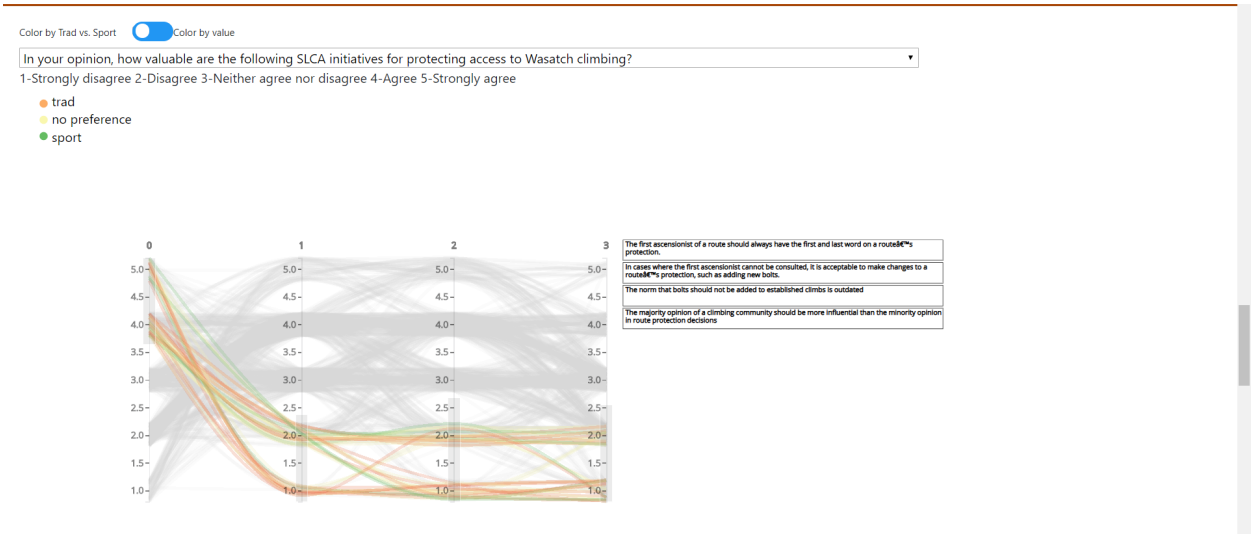
The user can further filter the PCP by selecting histograms in the control panel. In aggregate mode, the user can select individual bars, and in category mode the user can select an entire histogram. In this mode, the PCP is color coded by the bins in the histogram (below, for how a climber learned to climb)



The user can hover over either the axes or the text boxes to highlight the axis and corresponding question.



Finally, the user can brush each of the axes to further narrow down the visualization specific response values.



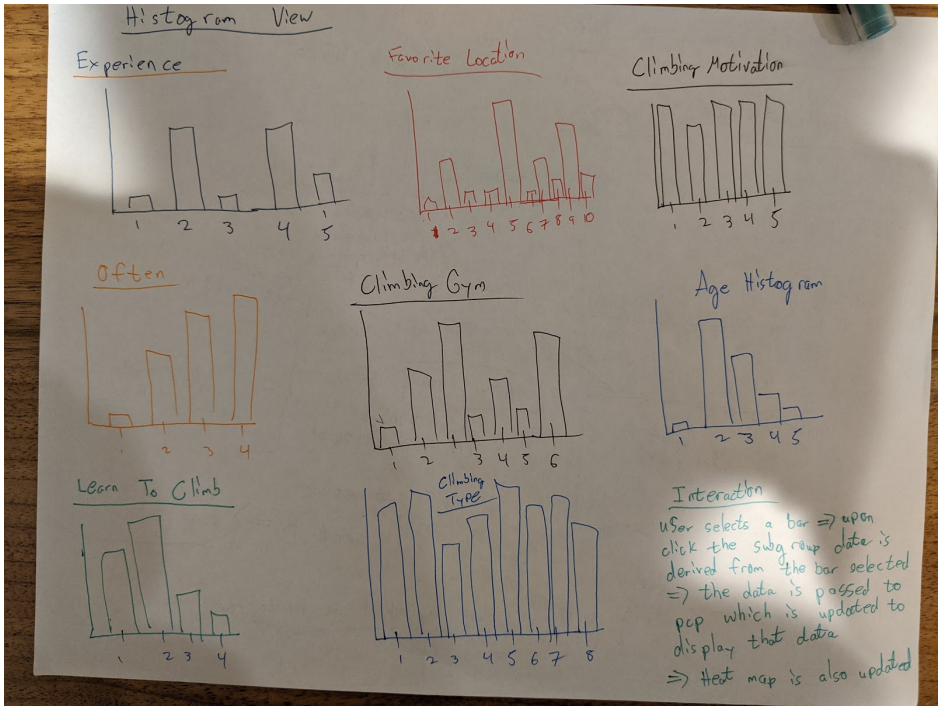
Here, we can see that respondents who typically have more traditional views towards bolting practices are predominantly trad climbers.

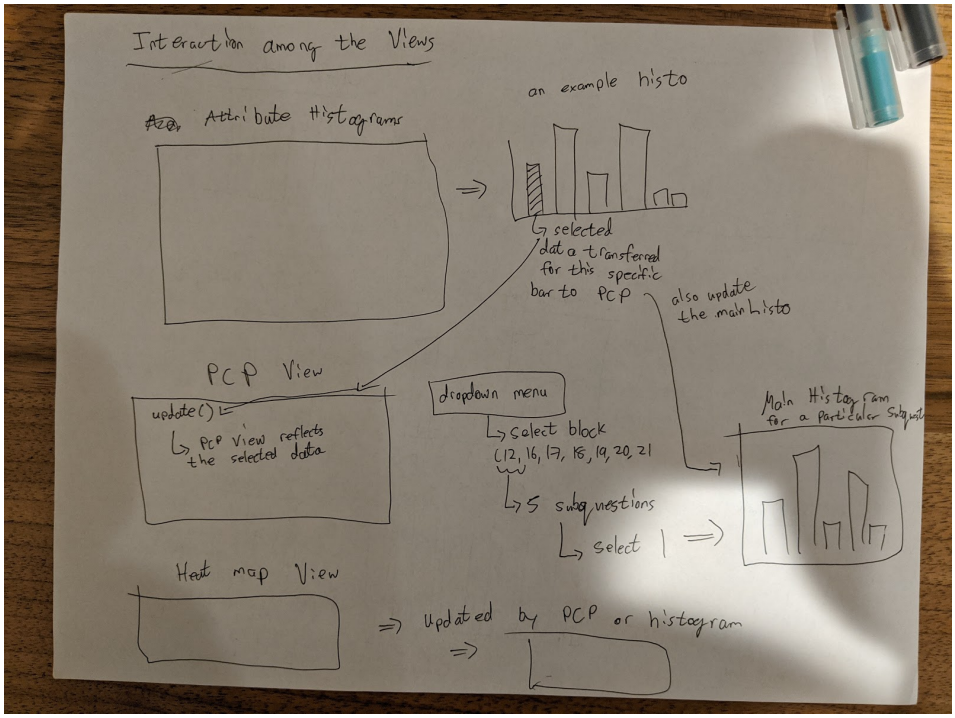
Histogram Visualization (Pranav Rajan)

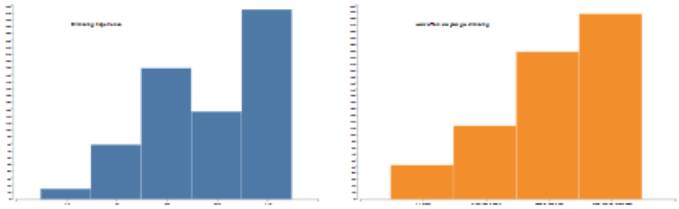
For this view, the visualization was redesigned from a drop down scattergram and scatter plot histogram to eight histograms and one main histogram. While designing the data structures and figuring out how to visualize different response distributions, the swarm chart caused issues with visualizing the different distributions for particular climber attributes. One problem that was not solved with the swarm approach was how to handle multiple responses because there were some questions where a person was allowed to select multiple choices. With the swarm approach there would have to be a lot more labels and information on the page to distinguish that a person selected a certain number of responses for a multiple response question such as favorite climbing locations. With the histogram approach, the data structures were designed

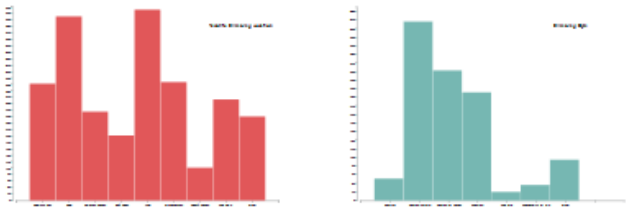
such that each bin of the different histograms corresponded to all the results for a particular question selection. For the eight different histograms, each histogram analyzes a different attribute question from the climbing data and bins and aggregates the total responses for a particular answer. Users can then interact with the different histogram bars to see different answers for a particular subgroup that belongs to a particular attribute category. The main histogram visualizes how people responded to the climbing visualization questions as defined in the data structures section. Blocking the questions into larger overall categories, users can select a particular subquestion that belongs to a category and view a histogram of the responses for that particular answer. For the climbing motivation and the climbing type histograms, all the participants of the survey answered for all types so there was not an interesting distribution to visualize. I took the data for these two questions and decided to visualize the responses that were the most important climbing type that climbers selected and the motivations that were very important and extremely important. For visualizing the total values, and the full name of a bin label, a tooltip was used to display the information.

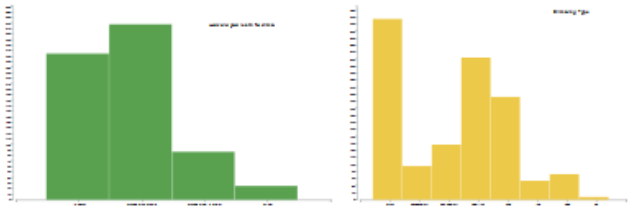
Redesign Images:

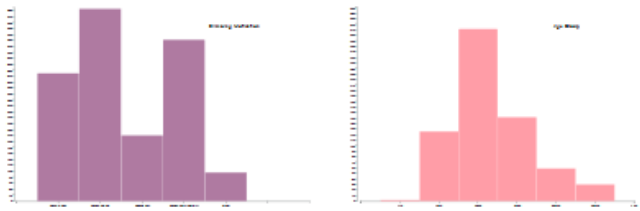


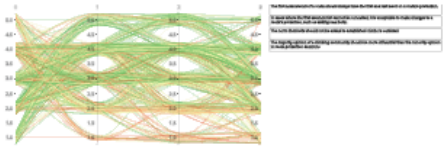

Interaction among the Views
Attribute Histograms
an example histo
selected data transferred for this specific bar to PCP
also update the main histo
PCP View
update()
PCP view reflects the selected data
dropdown menu
select block
(12, 16, 17, 18, 19, 20, 21
5 subquestions
select
Main Histogram for a particular Subquest
Heat map View
Updated by PCP or histogram

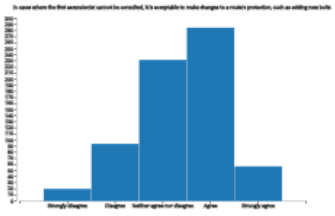








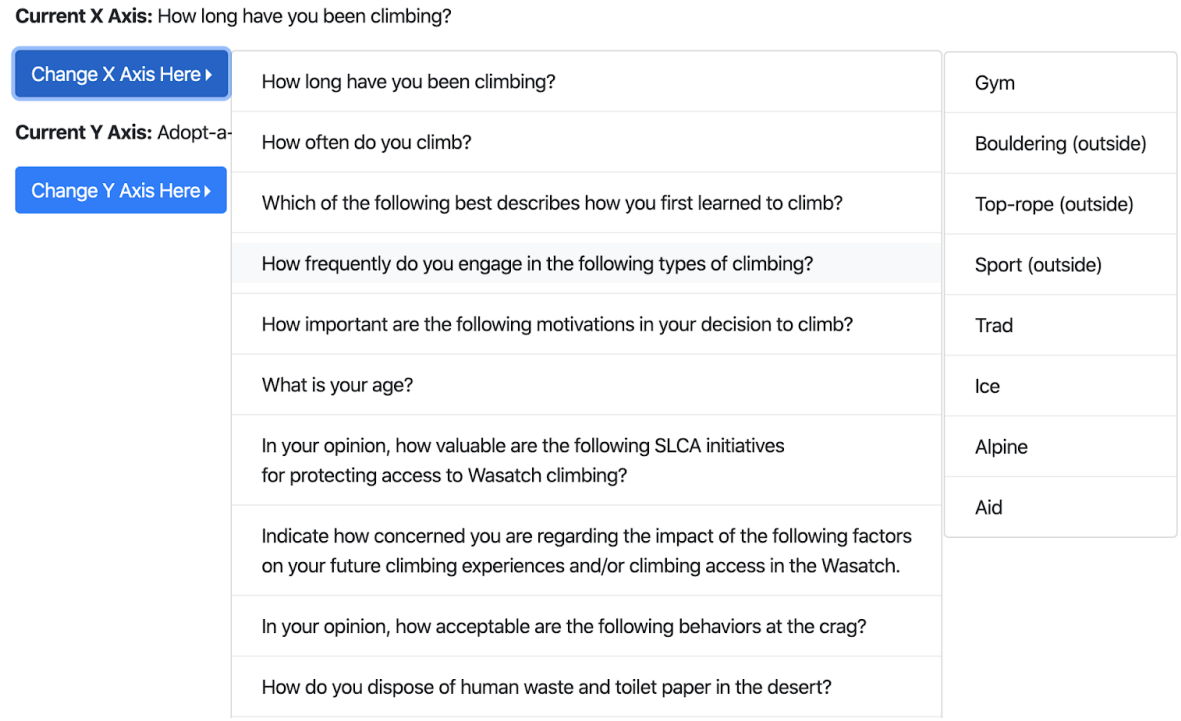




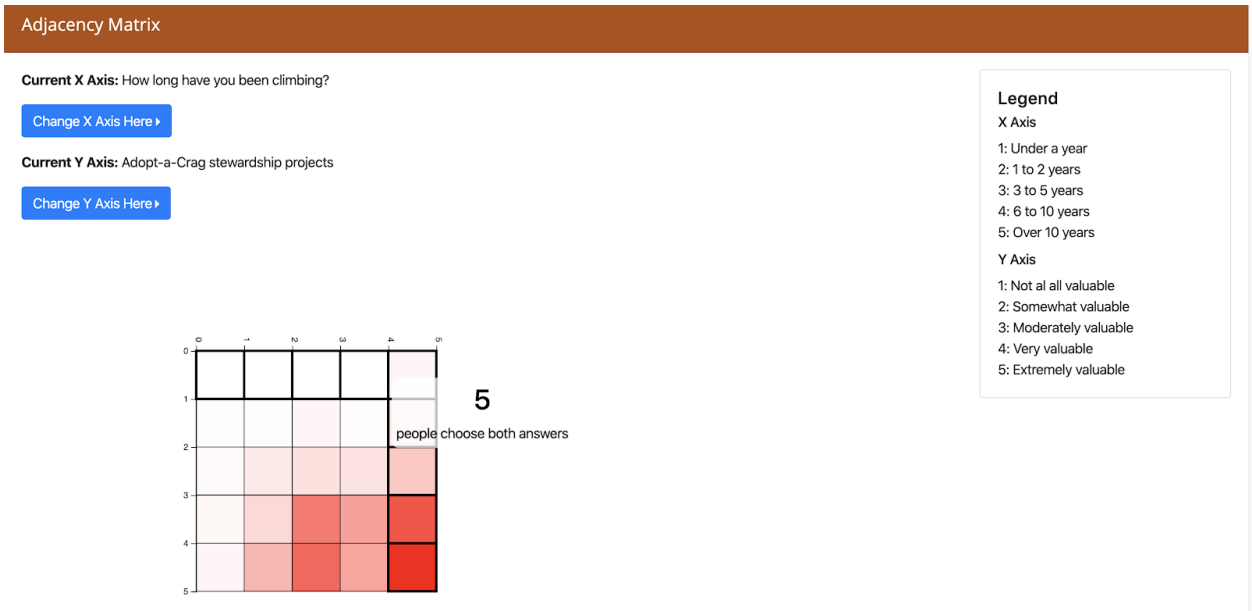
Heat Map Visualization (Tianlong Zhang)

The original idea for this part is actually the scatter plot, similar idea like homework 4. But after discussing with teammates and TA, I decided to do the Heat Map/Adjacency Matrix. The reason to choose this view is because it can not only show the relationship between 2 questions, like

scatter plot, but also avoid the drawbacks of scatter plot, since our data (questionnaire) is not like traditional sequential data. Our data has different format for different questions (some questions have subquestions, shown as picture below).



These problems can be well-handled by the adjacency matrix with heat map color representation. Shown as below:



In order to show the actual number of heat map, I did tooltips when hovering over the blocks, and for users convenience, we add the click function to show the actual answers that are respectively to that block. Also for user to read, we add the legend and text for the axes.

Overall Visualization Design

This website's layout is done by Bootstrap 4, this framework will help our layout looks better and additional feature is that our website will be compatible to laptop, pc, tablet, phone screen. Also it is compatible when you shrink or enlarge your browser.
The overall visualization has three main views: 1) The histogram view which contains eight different histograms of the different climber attributes and a main histogram for viewing subquestion distributions, 2) The Parallel Coordinates Plot which visualizes the responses for trad climbers, sport climbers, and all other climbers 3) The adjacency matrix view is for viewing the relationship of two questions. The idea for coloring is like a heat map, which is based on how many people choose those two selected questions. On the top left, you can change the questions by clicking on the dropdown buttons, after choosing the question for X or Y axis, you can see changes on 4 parts, the text of the axis, the matrix organization, the axes of the matrix and the legend. Additional features are 2 things, when you hover over a block in the matrix, a tooltip of that block will show up, describing how many people are choosing those 2 answers. Also if you click on the block, you will see the black borders on the blocks that will be easier for readers to know which answers are respectively to that block. The eight histogram view functions as the control panel for the visualization. Users can select different bars which upon being clicked, update the pcp view to see the distributions for the sub questions. The main histogram is linked to the PCP chart, where viewers can select a sub-question from the menu and the main histogram visualizes the distribution for that specific sub question.