RJ Duran
MAT259 Winter 2012
Data Visualization
Final Project

Introduction

The goal of this project is to visually explore and navigate the connections between words associated with the word “WATER” in book titles from the Seattle Public Library database from 2006 to 2011. By utilizing a partial FP-Tree algorithm to parse incoming book titles I am able to represent visual patterns in associated words. The data is represented as tree structures and volumetric pyramids in a polar plane within a navigable 3D space.

Query

```
select title from inraw where title like '%water%' and itemtype like '%bk%';
```

Explanation

This query examines the entire database “inraw” for book titles containing the word “water.”

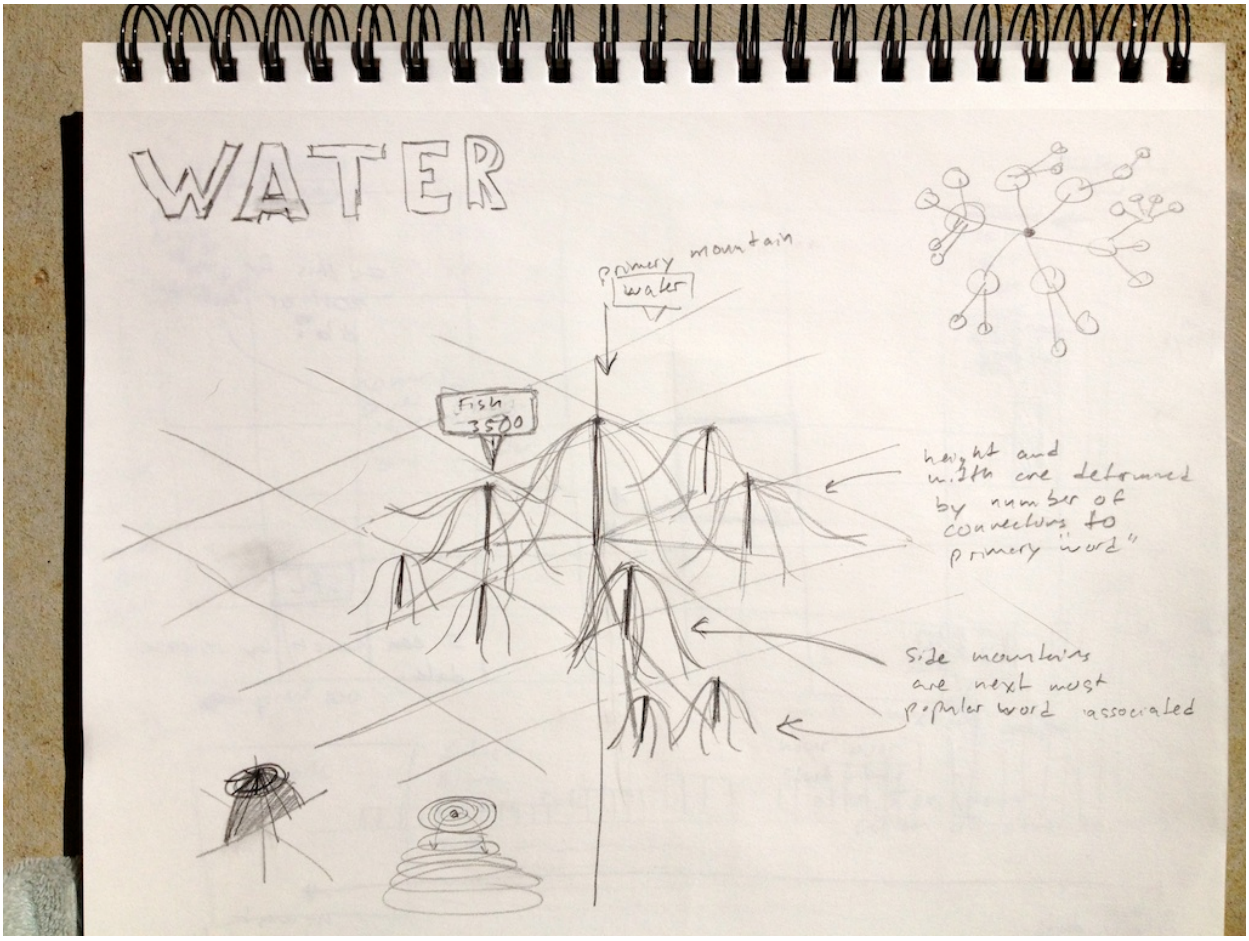
Query Request Time

In Sequel Pro the query takes about 3 or 4 mins to complete. After testing a small query where I looked at a specific day I expanded the search to the entire database. I used a custom Processing program to run the query and save the data into a text file for easy loading into my main visualization. The entire set of data can be seen in 20052011Data_WATER.txt.

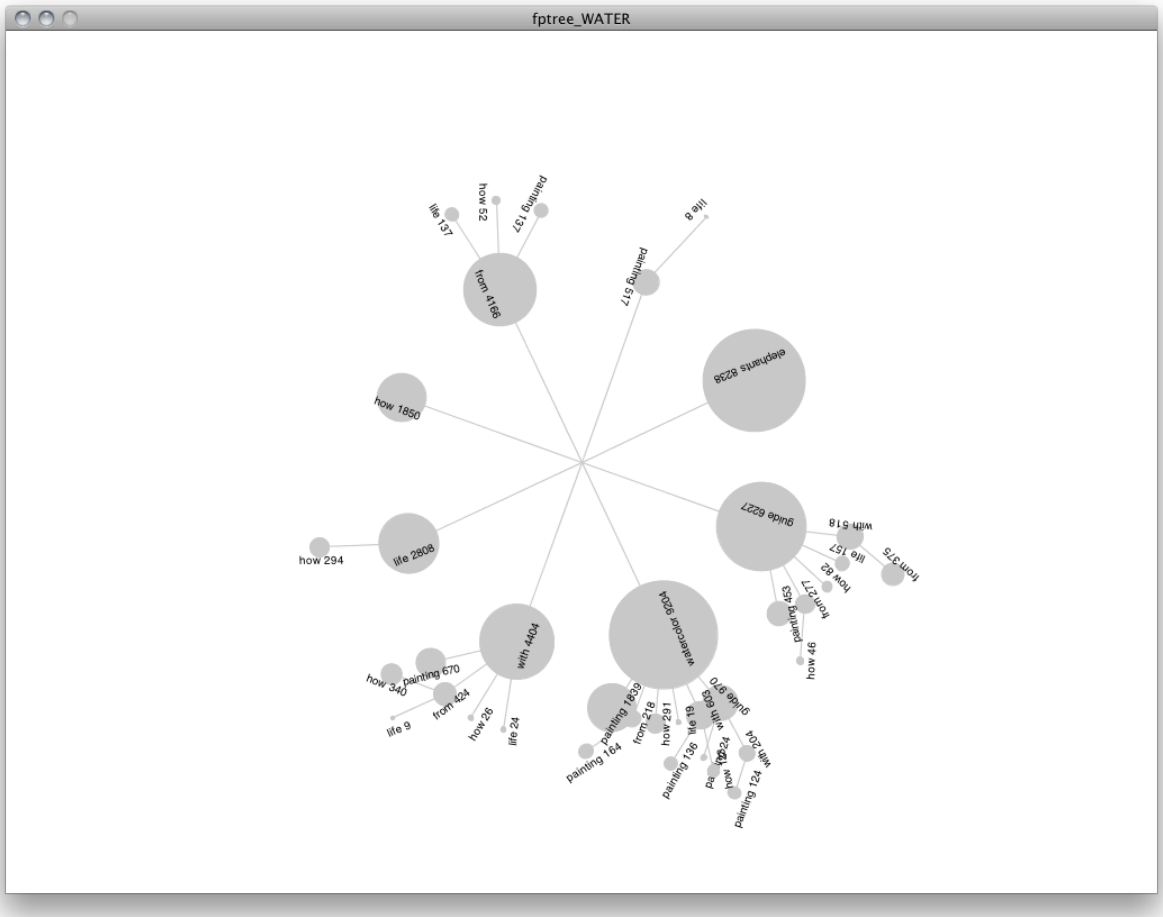
Process

The project started out as an extension of project #2, where I tried to determine how “green” Seattle Readers are. As I examined the broad topic of sustainability and environment I focused on one common topic for this project. By building a MySQL query for book titles containing the word “water” I was able to gather a rich set of data to represent.

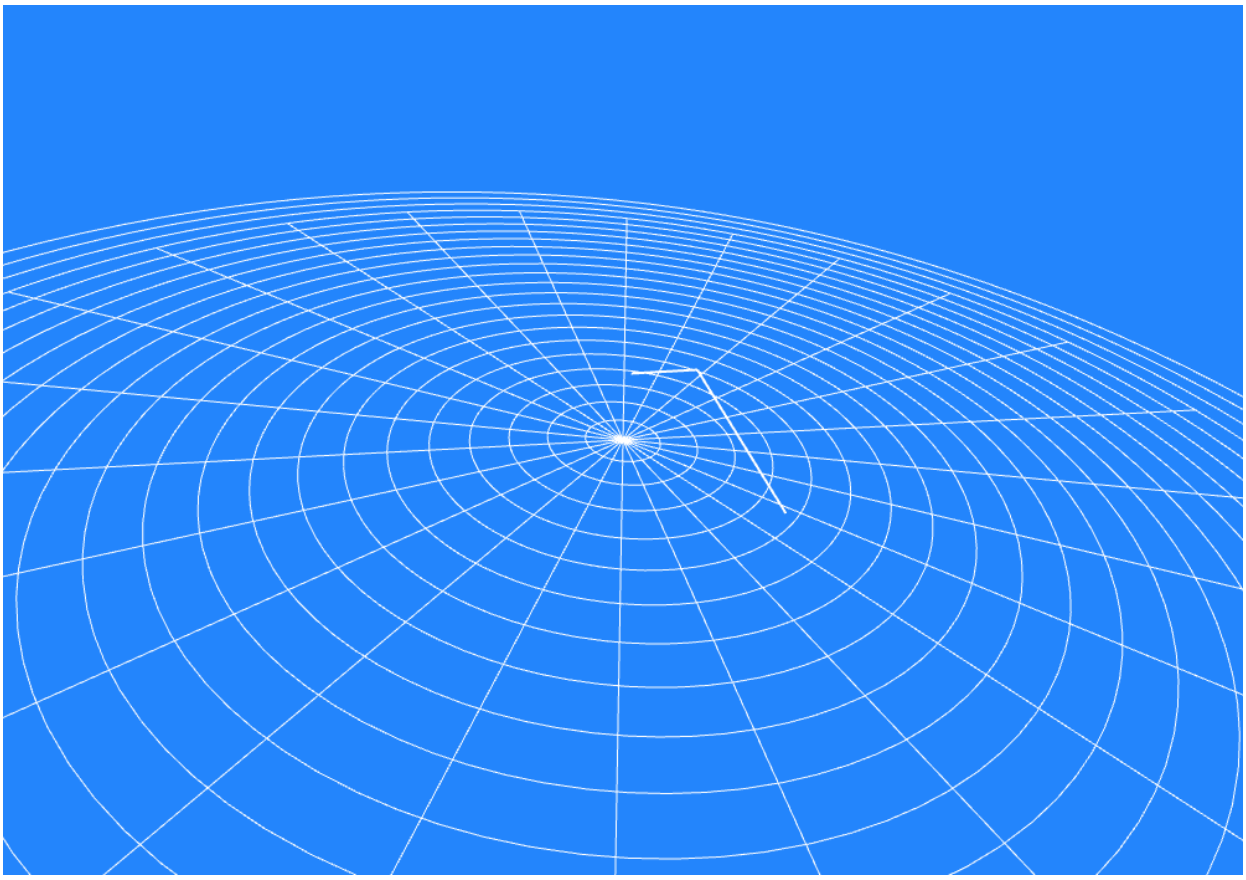
The process continued with a sketch of the initial idea. Initially I imagined representing the data as a topographical map with mountains representing the occurrence of words.

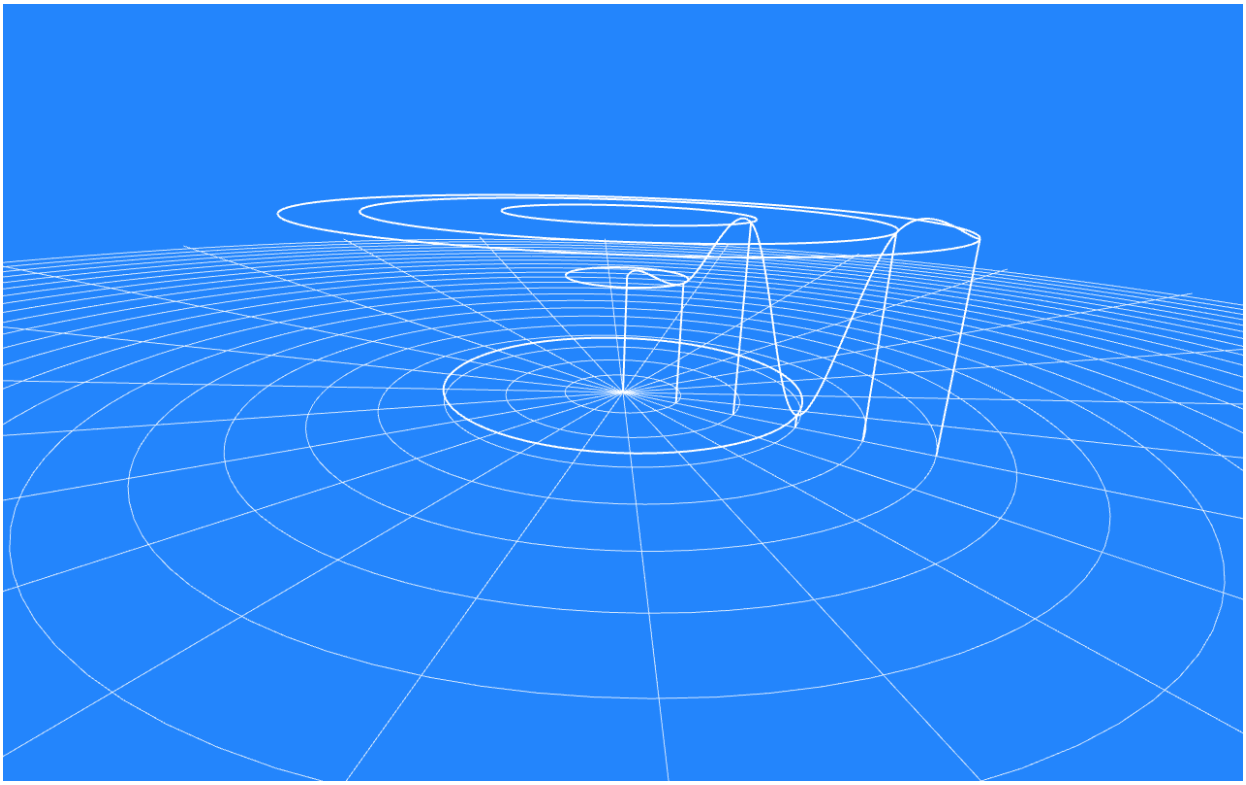


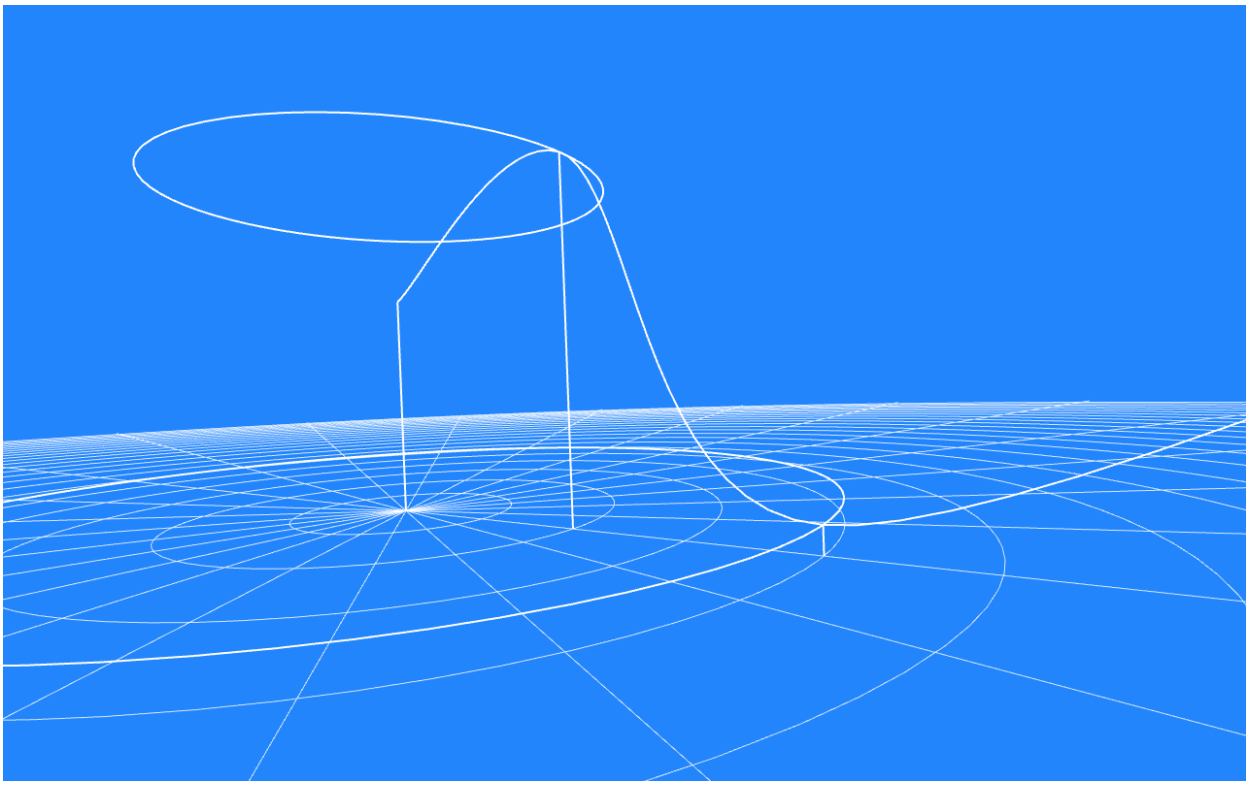
After I made the initial sketch I started prototyping the idea with Processing to see if I could represent it as I imagined. I ran a simple FP-Tree test to look at the data in 2D as seen below.

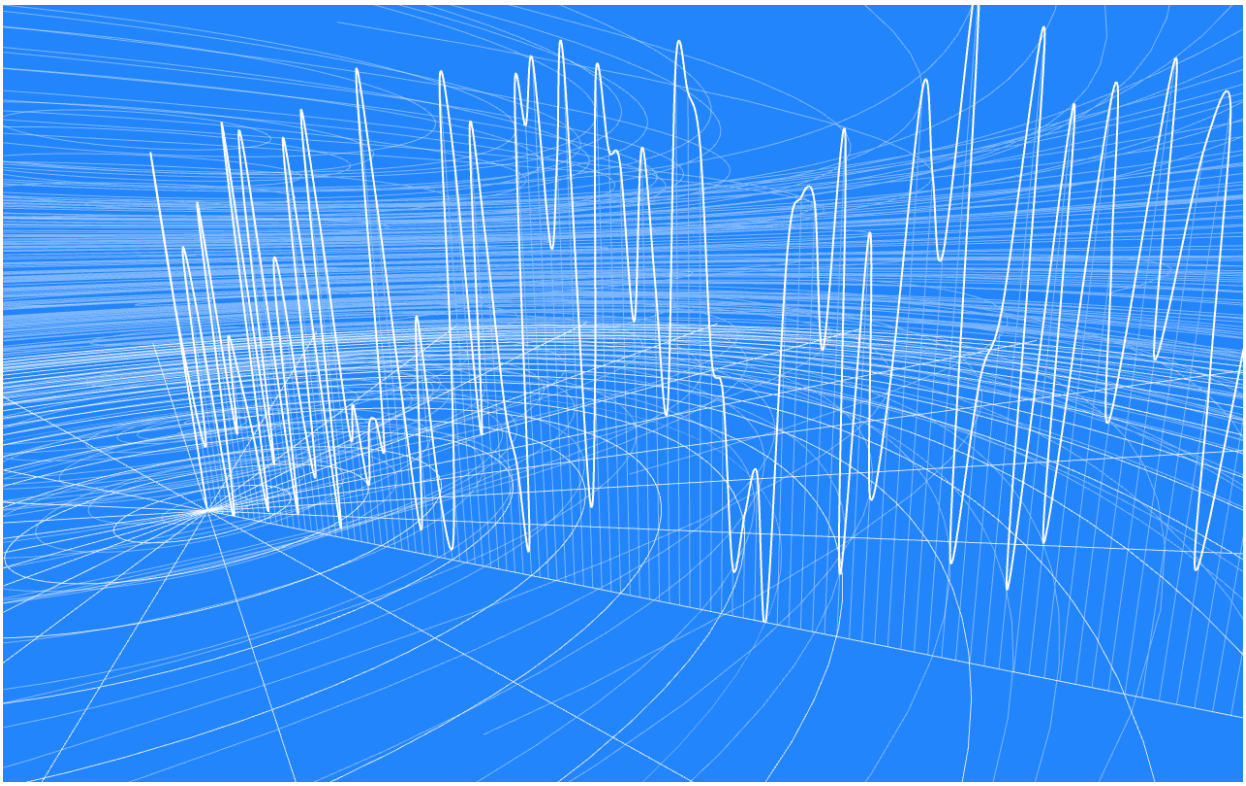


This led me to rethink the type of graph I wanted to place the data in relation to in 3D space. I built a simple polar graph to determine if this was a realistic direction. The images below show my first attempts at plotting a few data points within the space in relation to the polar graph.



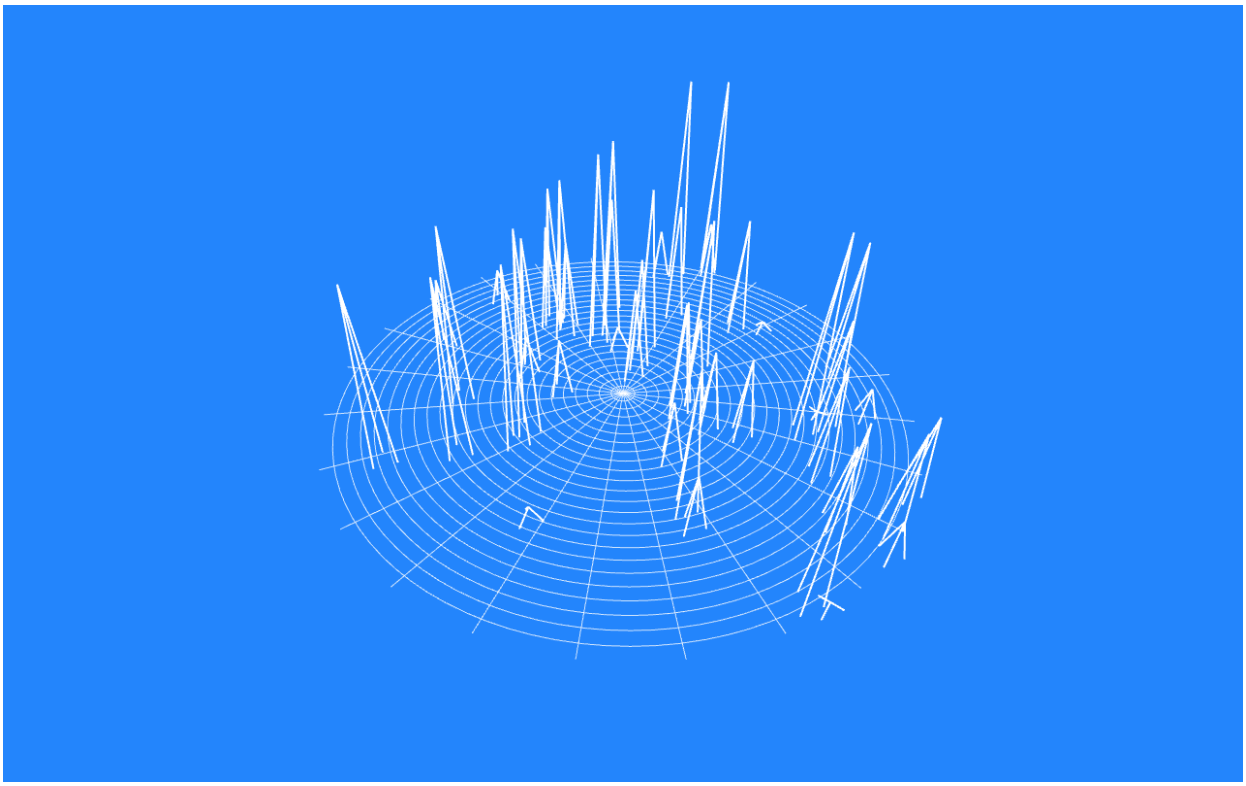




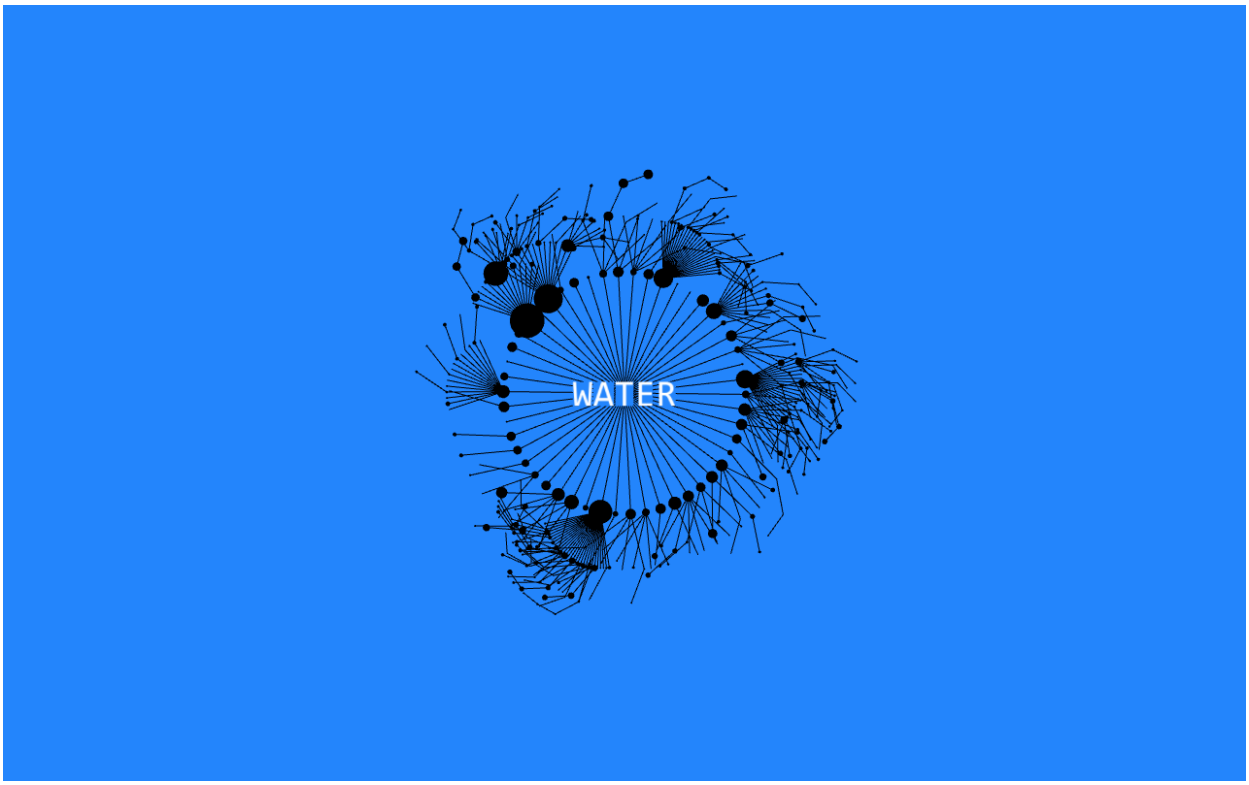


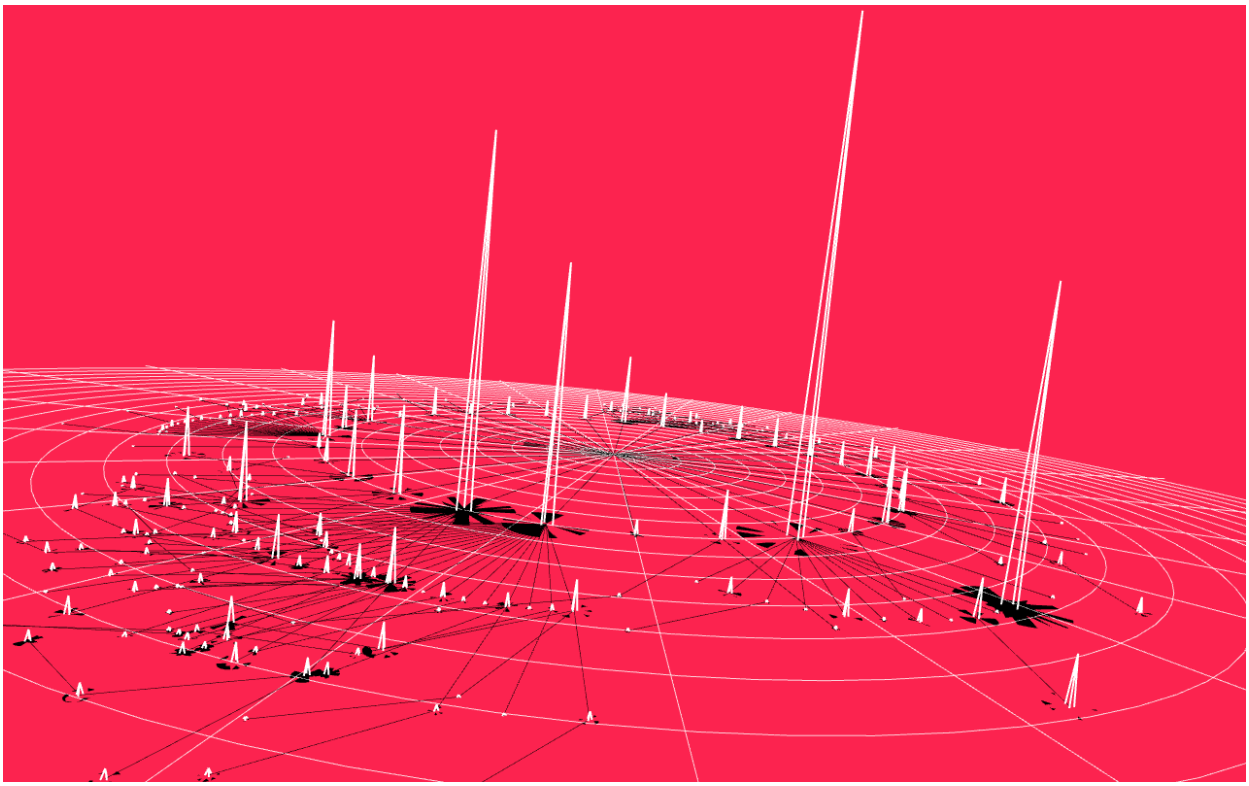
Once I was able to plot some lines within the space I decided to experiment with placing random pyramids on the graph. This was to help me determine what the actual data might look like

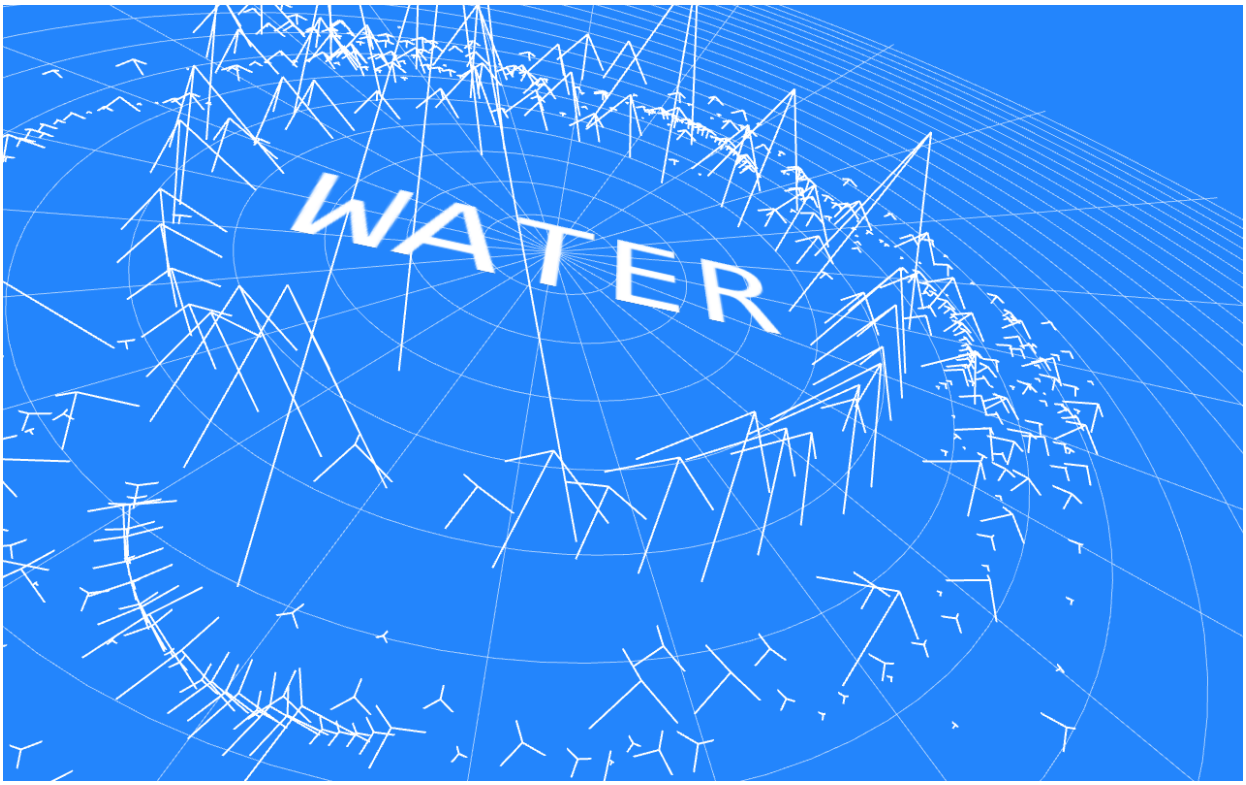
when plotted. The pyramids in the images below are placed at random locations from the center of the graph. They have random heights within a set range.

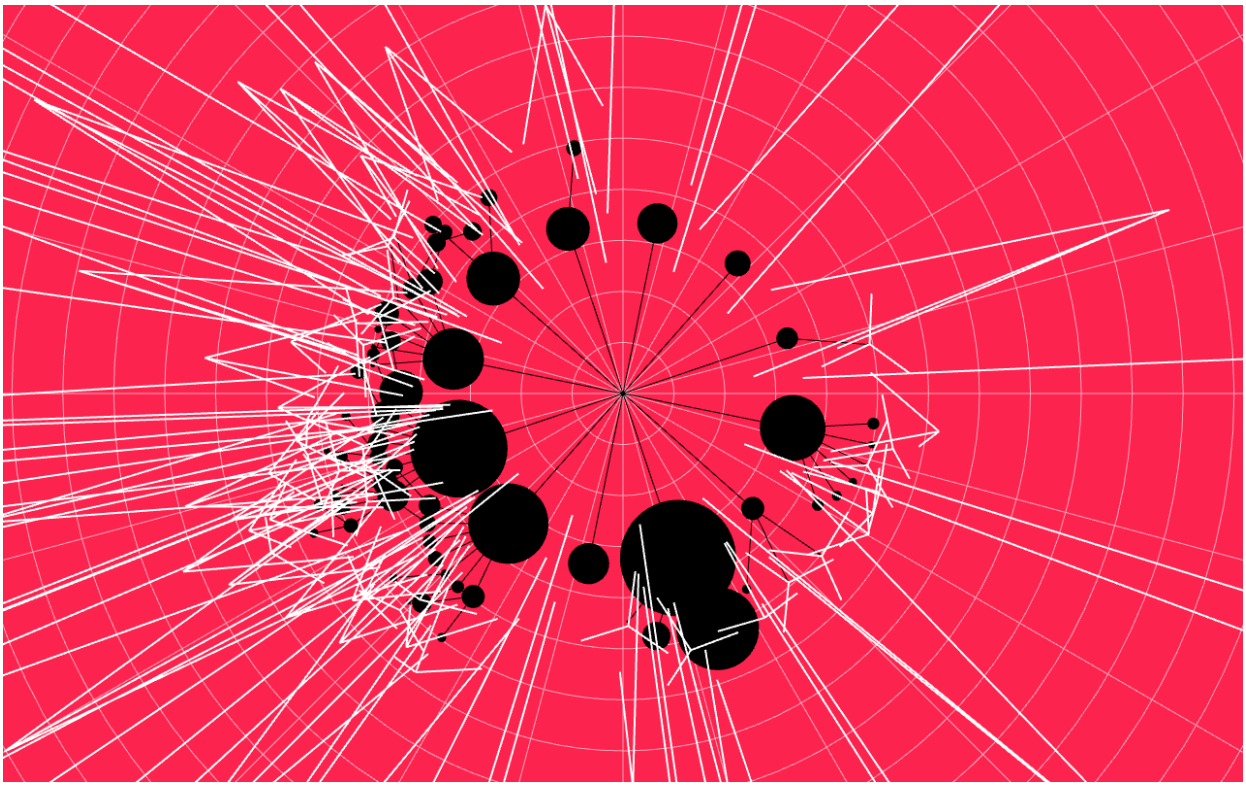


At this point I had the elements in place to plot data from the FP-Tree dataset. I started by plotting the tree structure itself on the graph. I experimented with color, pyramid width and scale, and positioning of the tree nodes. The images below show my first attempts at displaying the data this way. I experimented with hiding the lines and nodes and showing the pyramids.

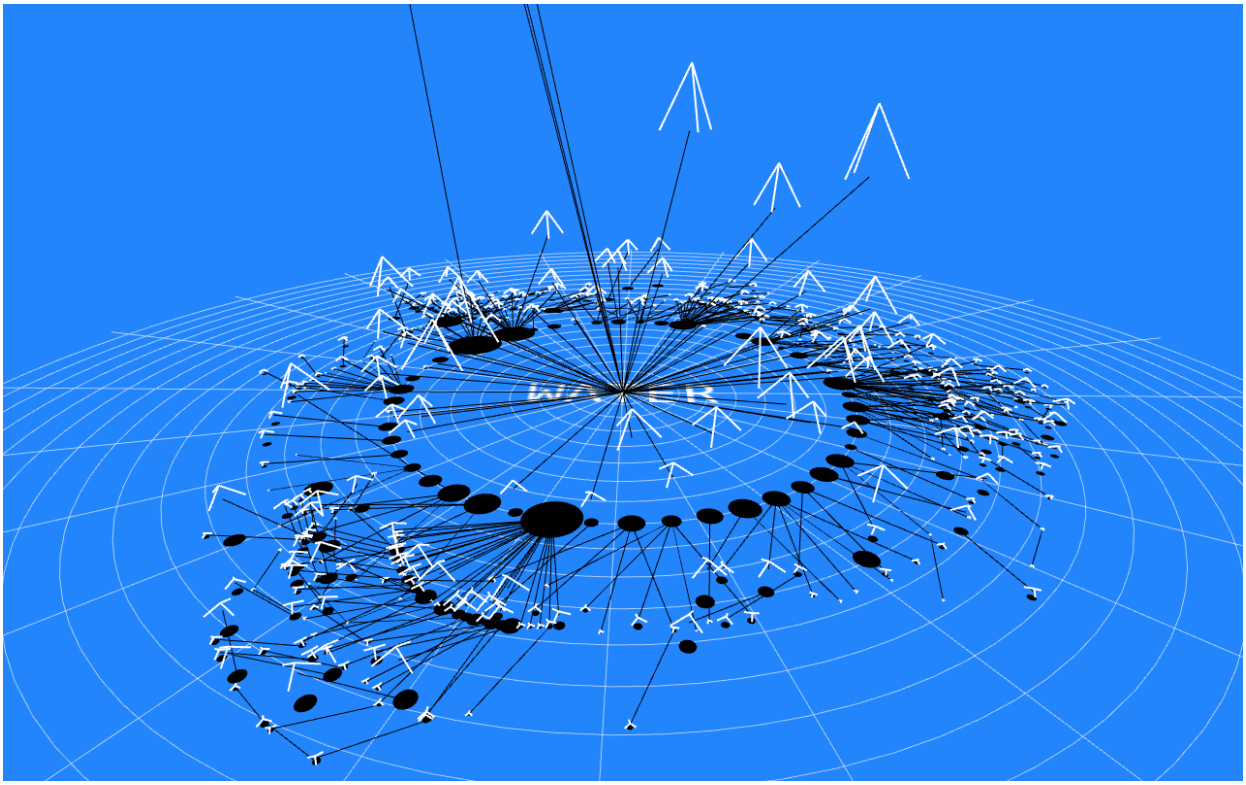

WATER



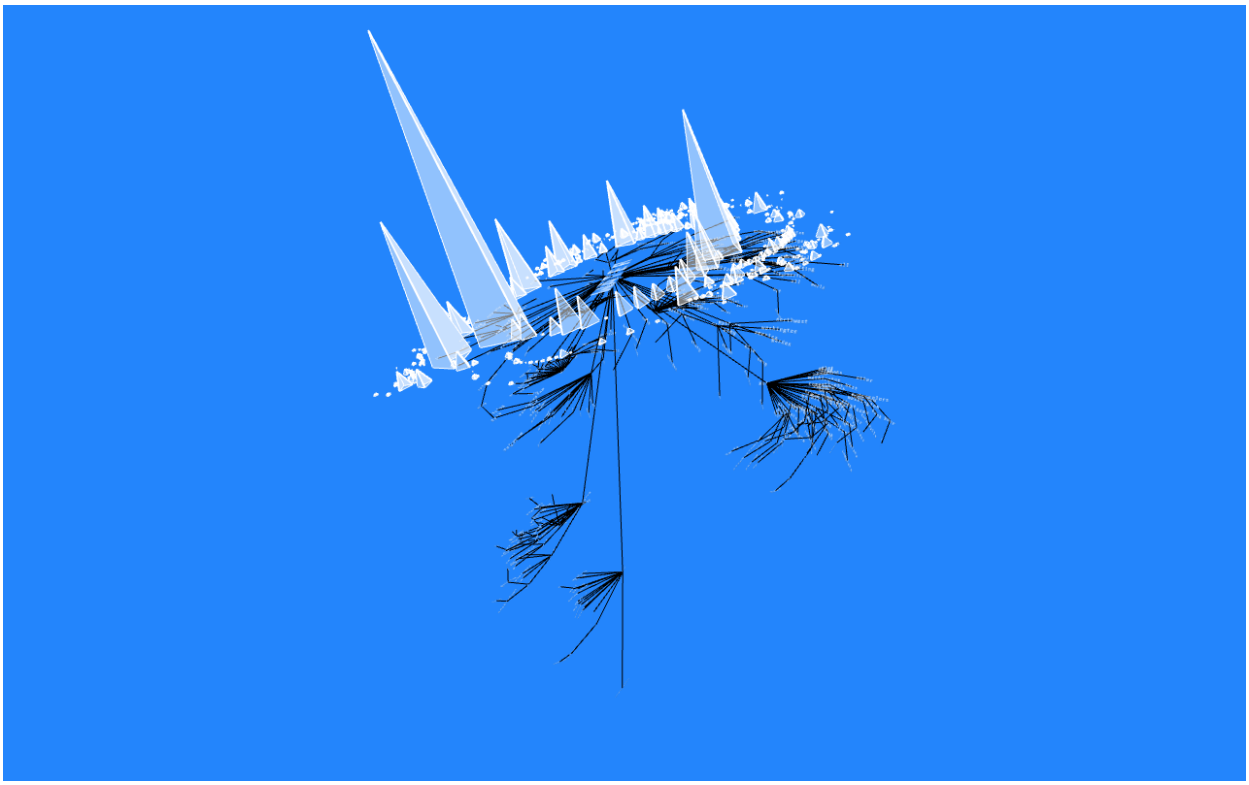


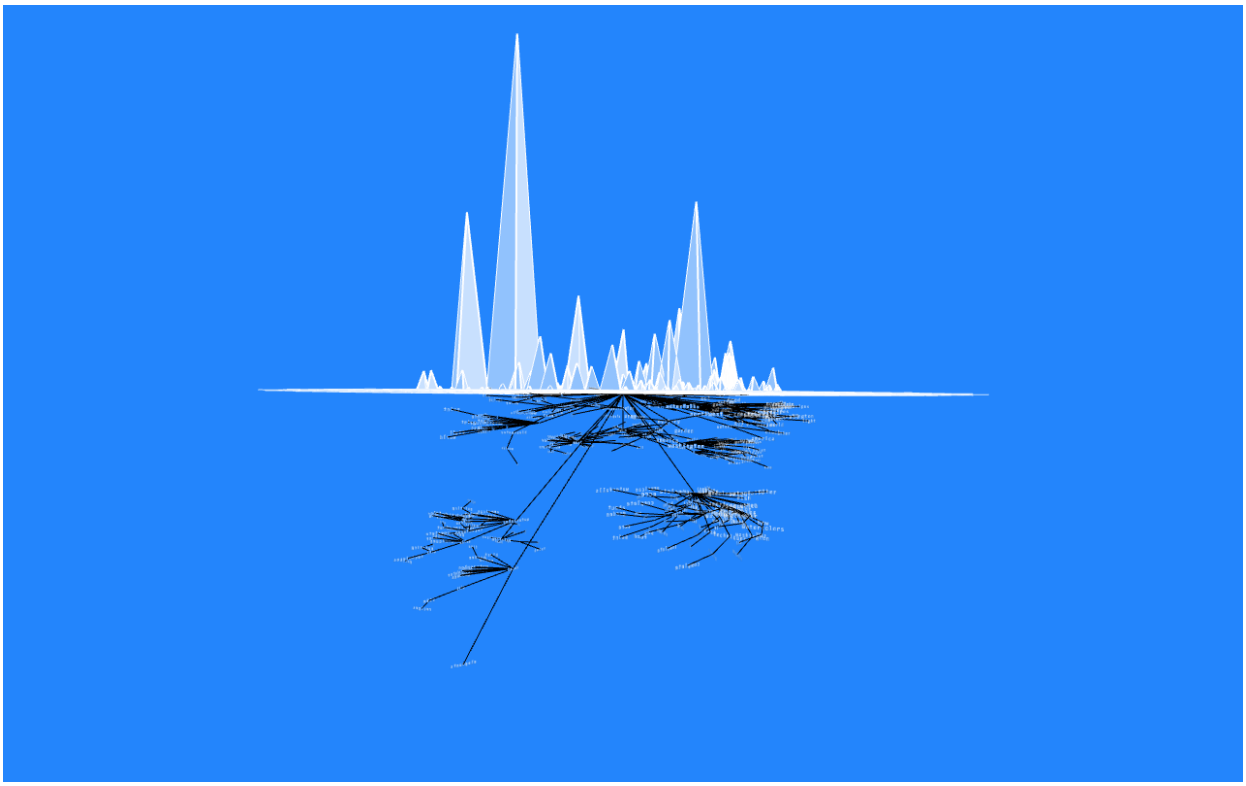


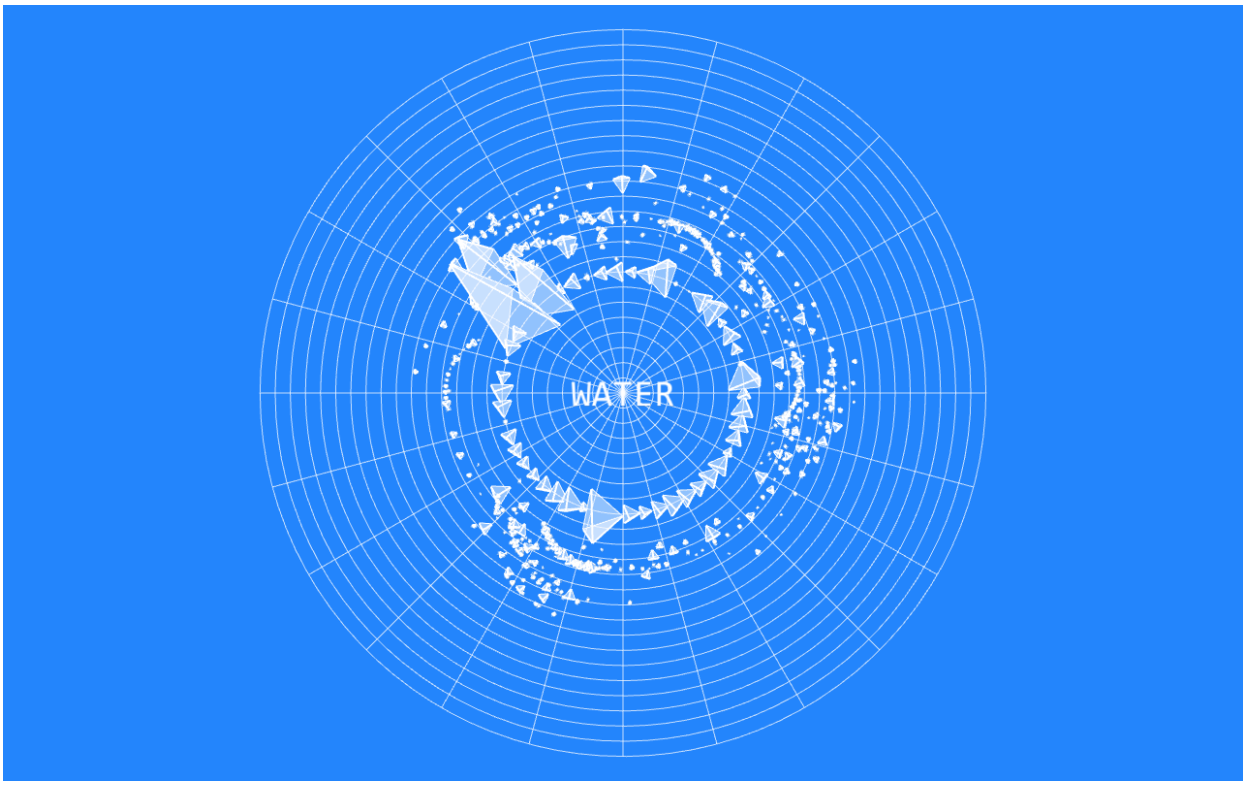
Following this attempt I wondered what the tree structure might look like in 3D. I experimented with projecting the tree above the graph and positioning the pyramids at the ends of each line segment.



The next step I took was to reverse the tree and represent it underneath the graph. Metaphorically, I saw representing the tree structure as a representation of roots in the ground. All living things have a root structure of some kind and they all need water to survive. I also played with the idea of connecting the mountains to the roots of everything through water. The pyramids represent the mountains and contain a volume of water just as the roots do. The images below start to bring out these ideas.

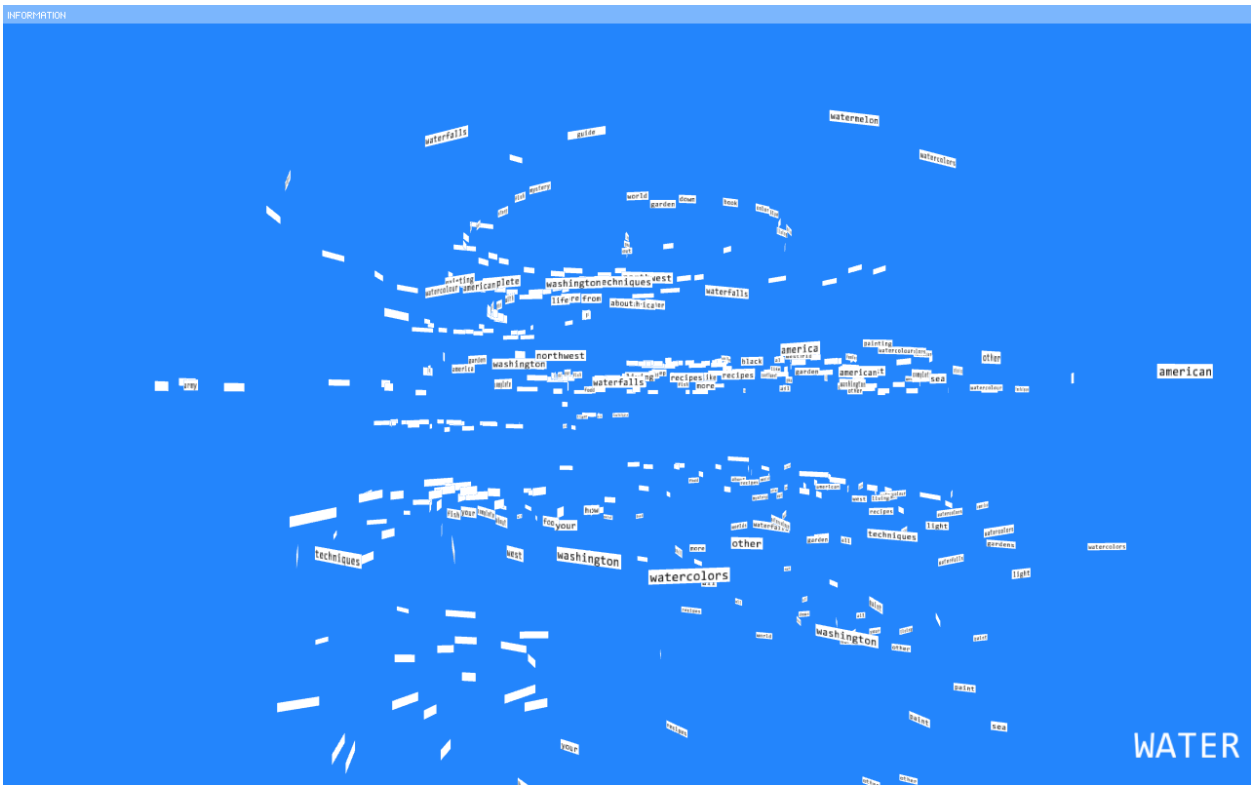


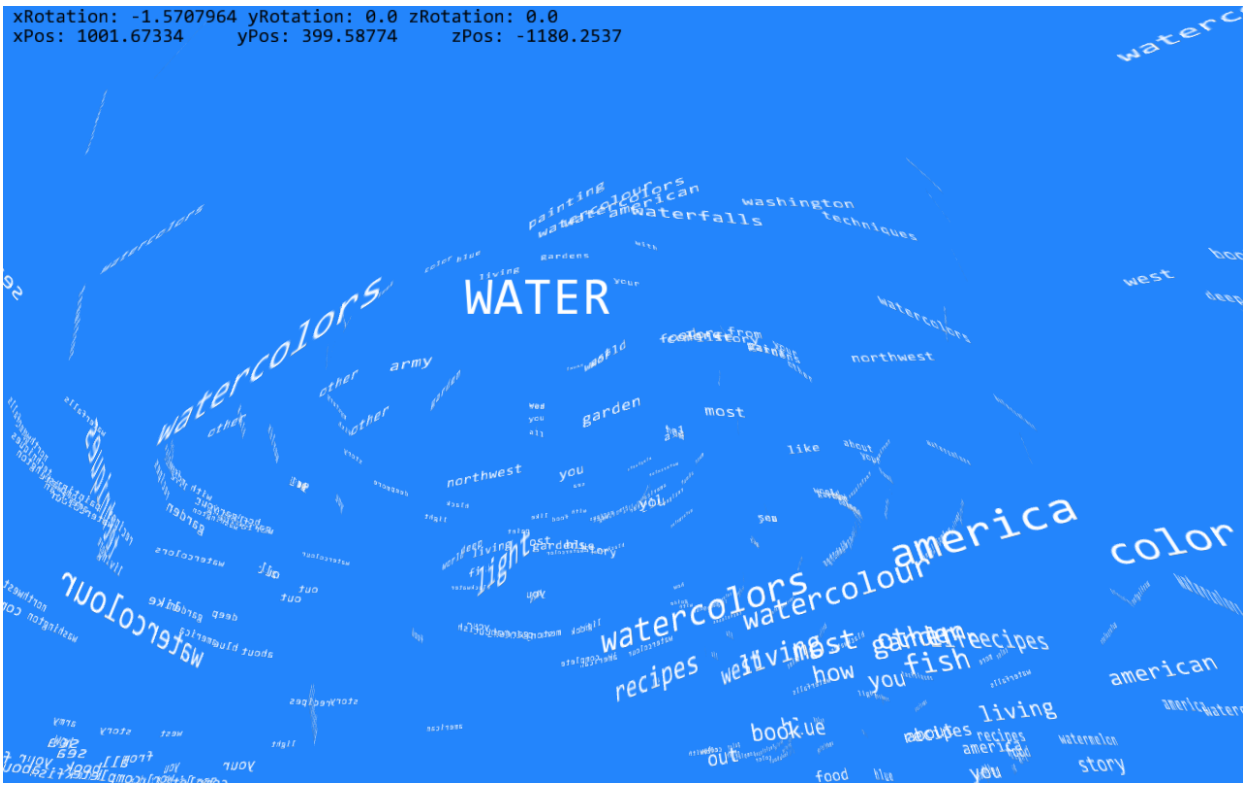




Once I had the data in place I started looking for ways to make the data interactive and navigable. I wanted to be able to move specific parts of the data set and see it in detail. For this I started to experiment with ControlP5. I added a small control bar to the top of my program to place elements. By moving the mouse over the area the panel drops down to show these controls.

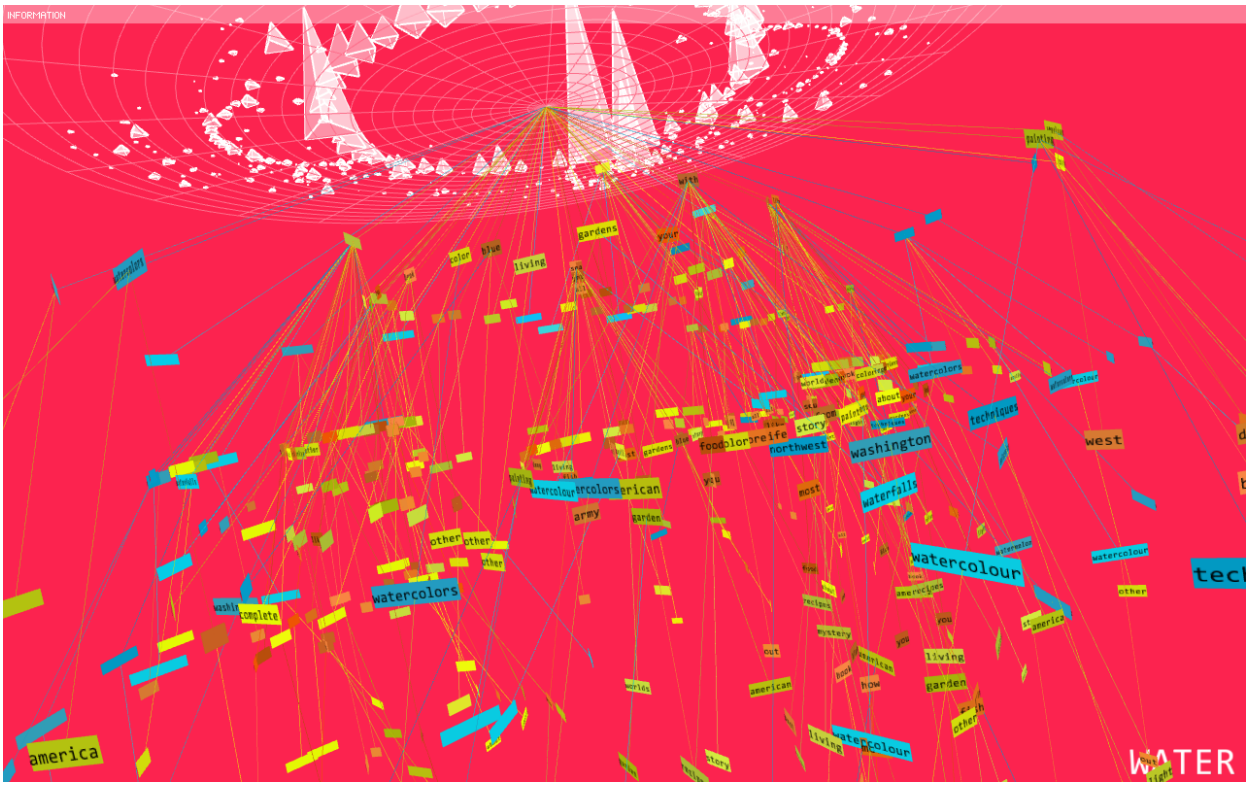
This is also where I started experimenting with representing the actual words associated with nodes in 3D. I applied a “tape like” look to the words with white backgrounds and black text. I hid the lines between the connecting nodes to bring out a interesting swirling visual effect. I added animation to the entire tree to rotate around the Y-axis (vertical).





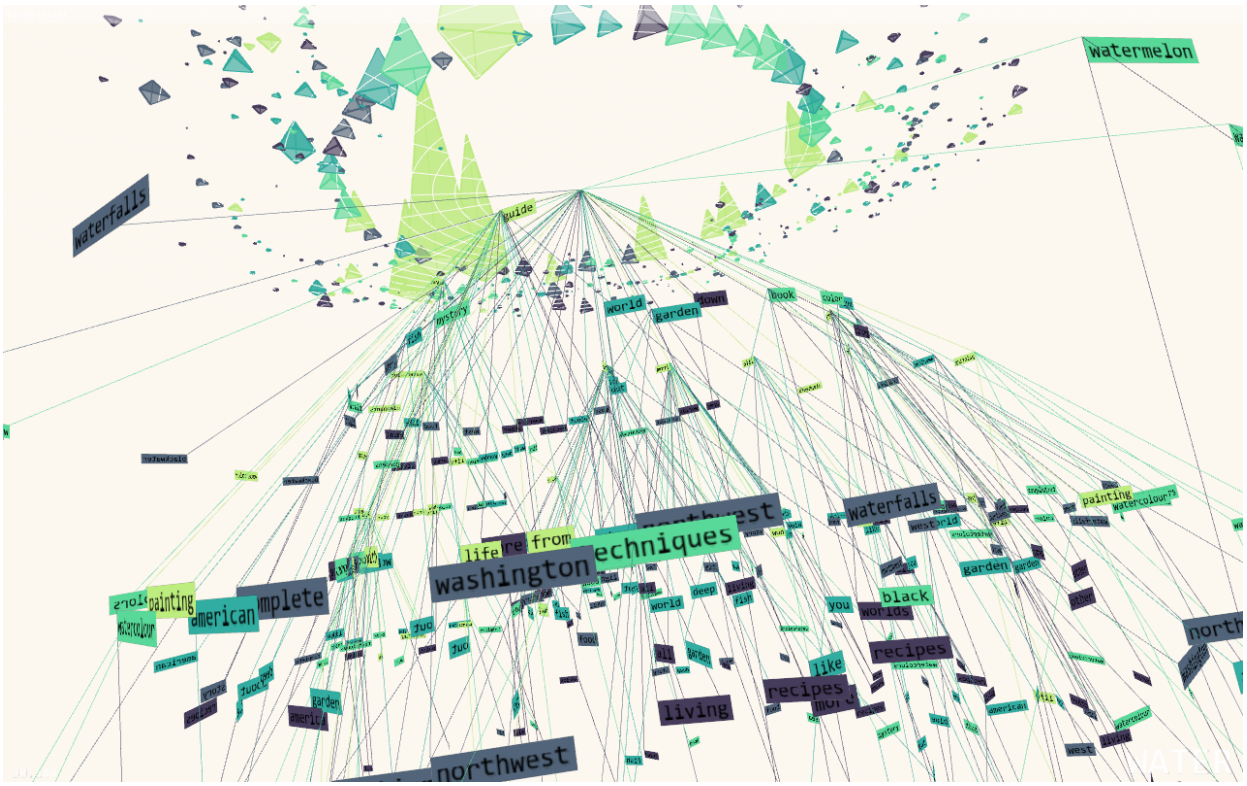
At this point I had a pretty good representation of data in both tree and pyramid forms. The only problem was in how to discern the difference between words at each node. To do this I

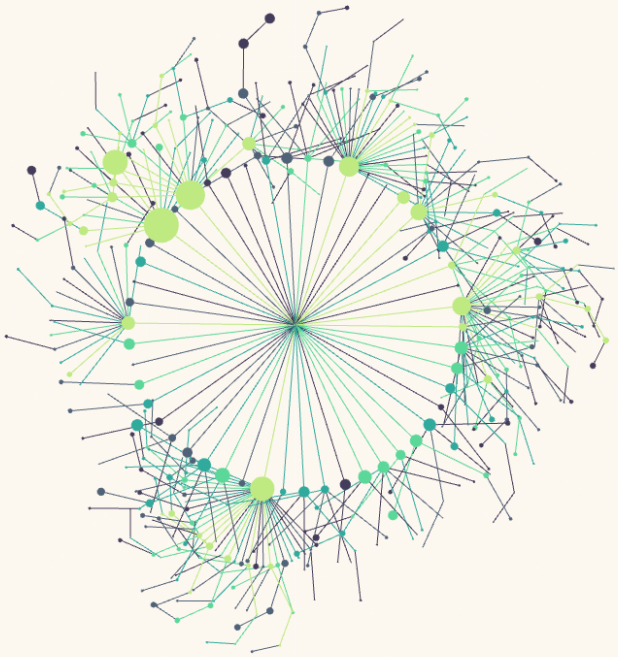
started experimenting with color. The image below shows words with associated colors chosen based on the length of each word. Longer words are represented as blue and shorter words are represented as oranges and yellows.



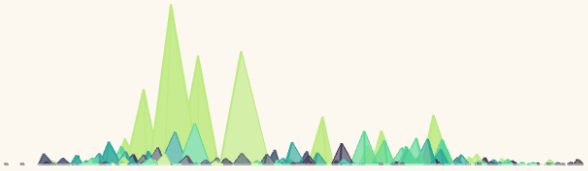
From here I continued to fine tune and optimize the program and color scheme. These are some other color experiments I tried.

world
garden
book
washington
techniques
american
waterfalls
life
from
about
northwest
washington
waterfalls
black
recipes
garden
painting
more
how
your
food
your

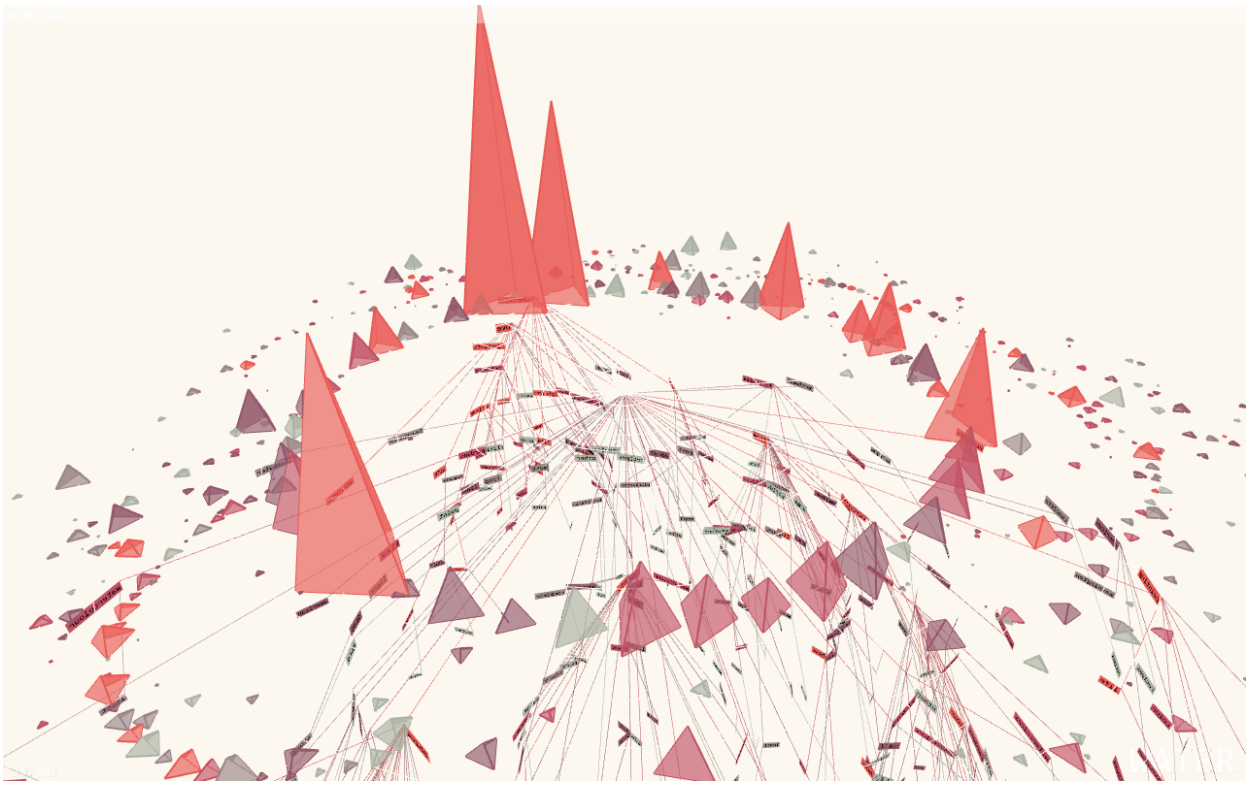

watermelon
waterfalls
guide
world
garden
book
waterfalls
painting
washington
techniques
life
from
american
black
recipes
living
northwest

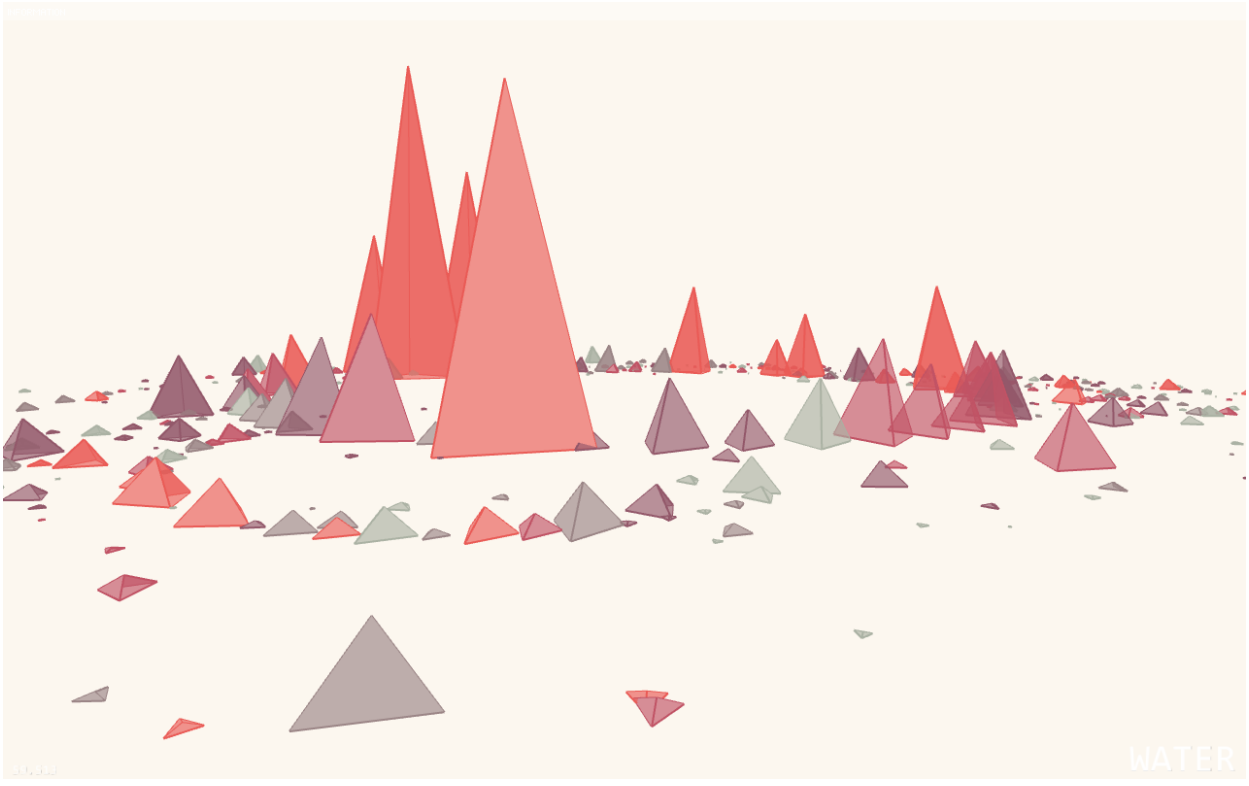


WATER



WATER

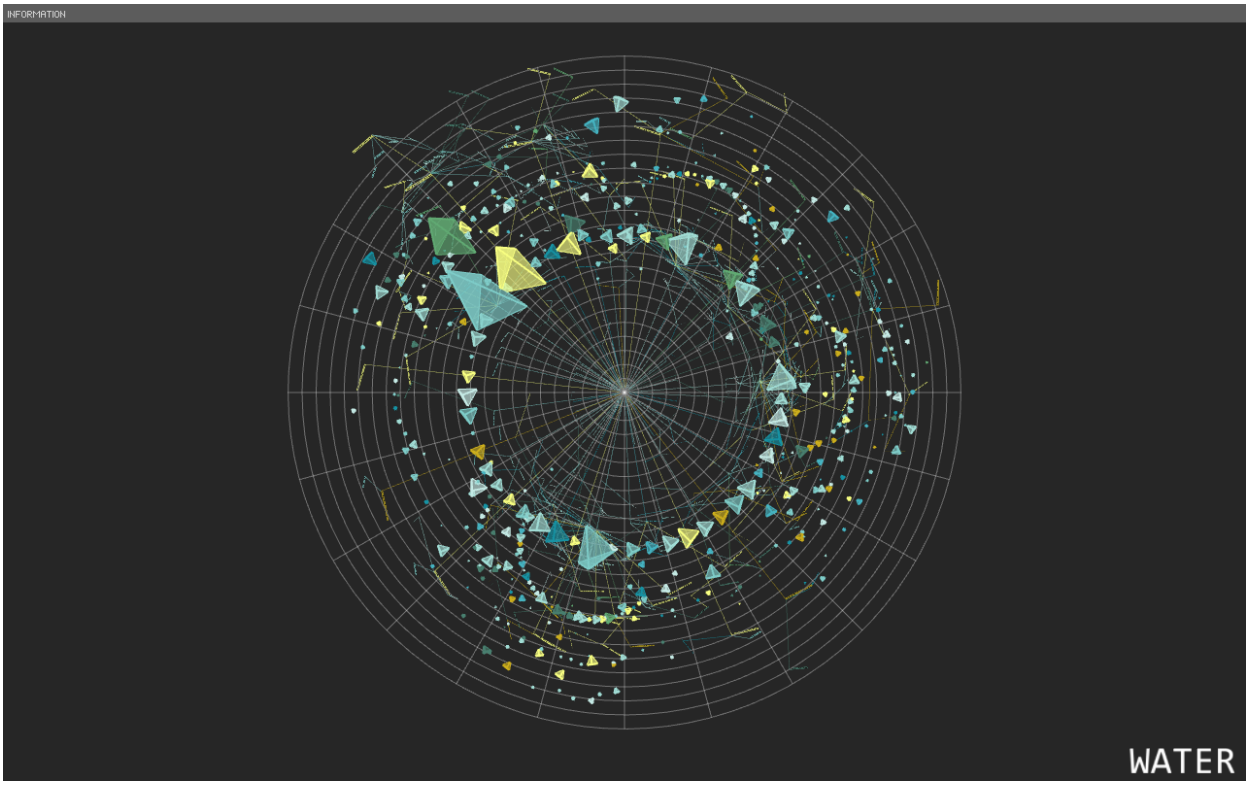


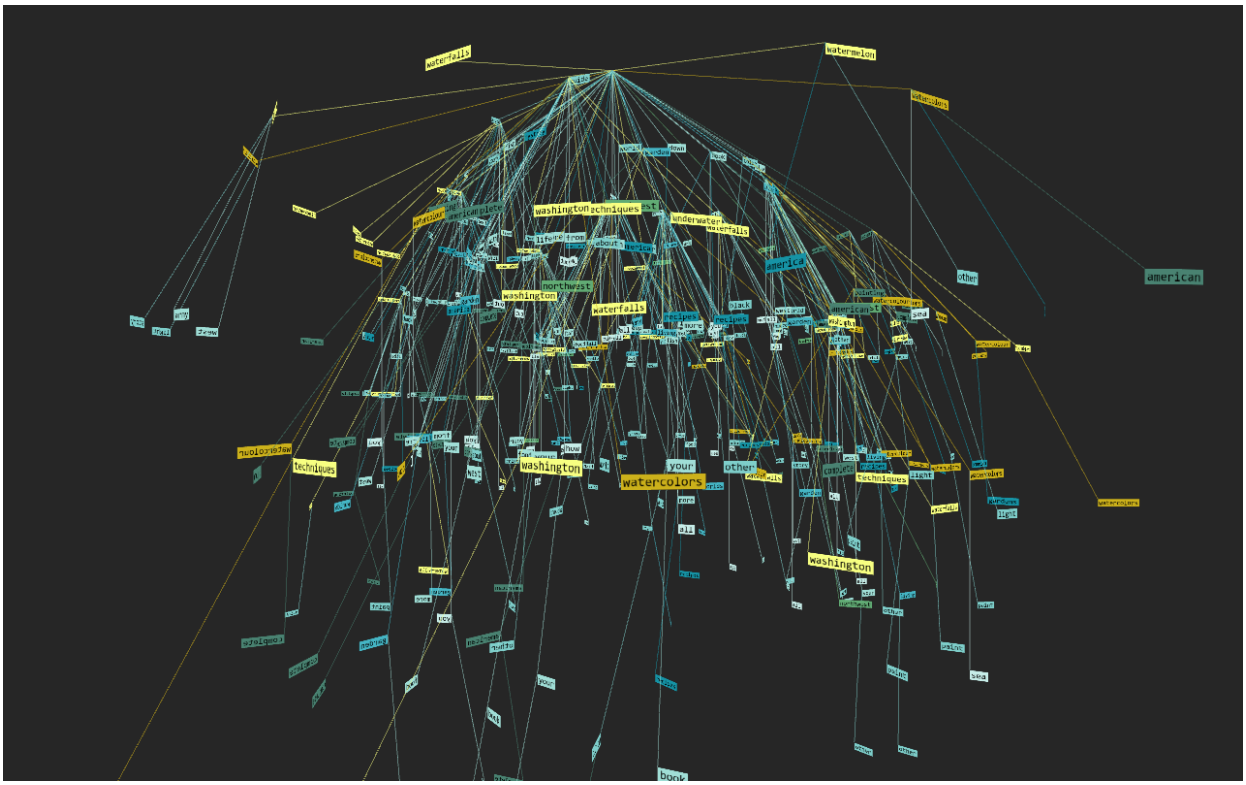


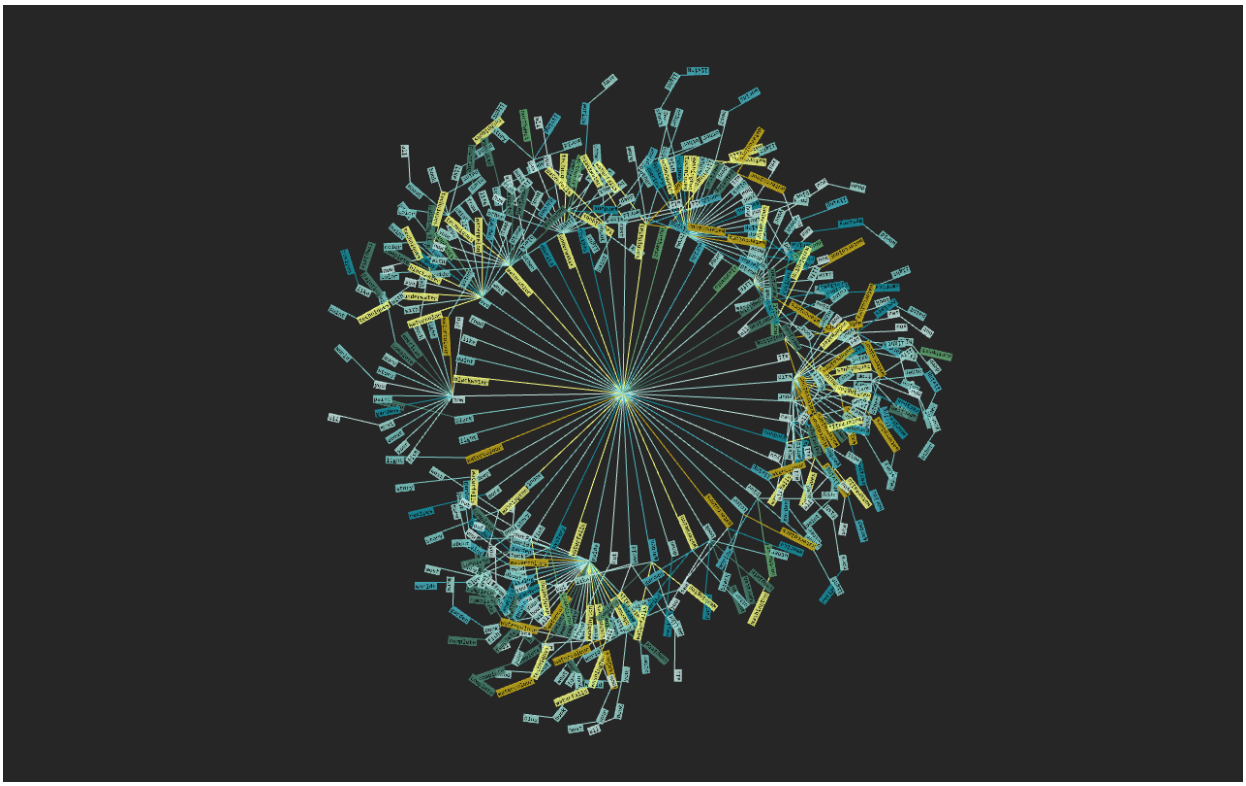
The final stage involved optimizing the frame rate, color selection, and cleanup. The final screenshots can be seen below. I decided to use a color palette of earthy tones consisting of

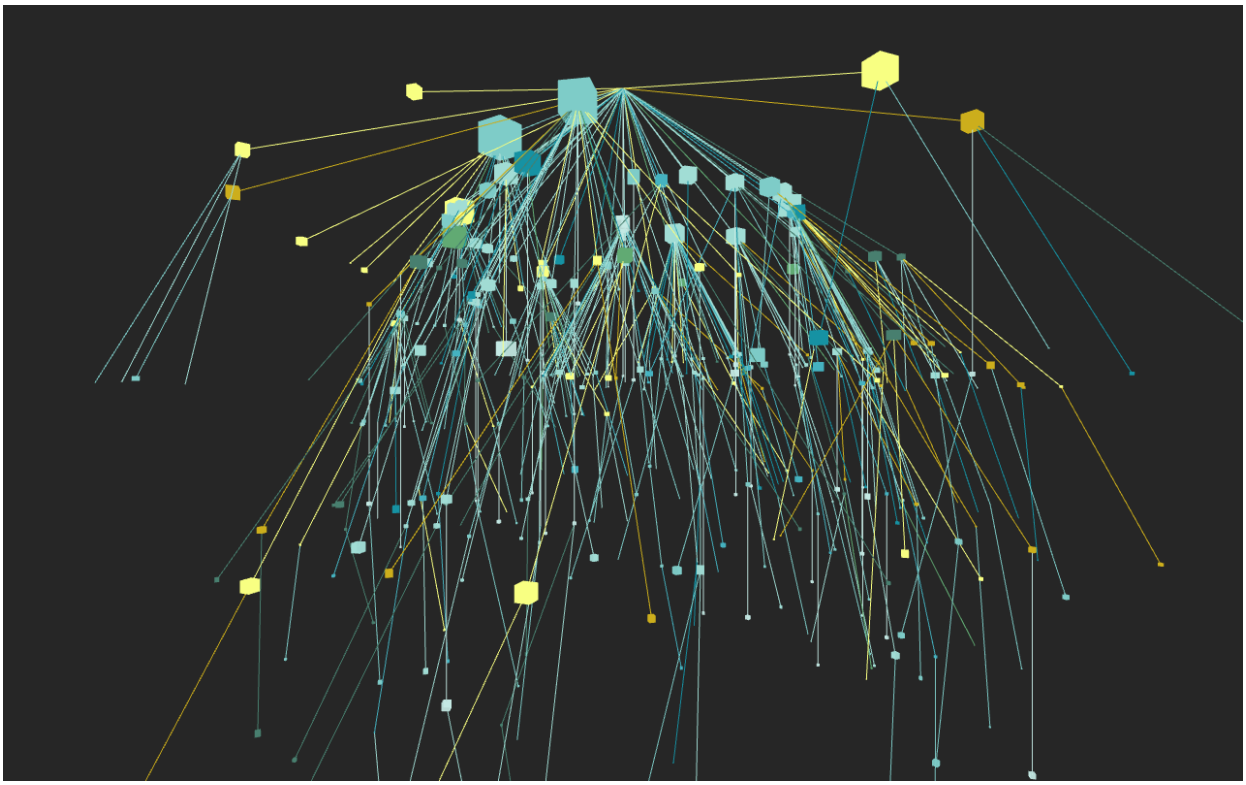
blues, greens, and yellows with a dark background to bring out the elements. I related color to the words parsed from the title by string length and applied it to both the pyramids, lines, and nodes.

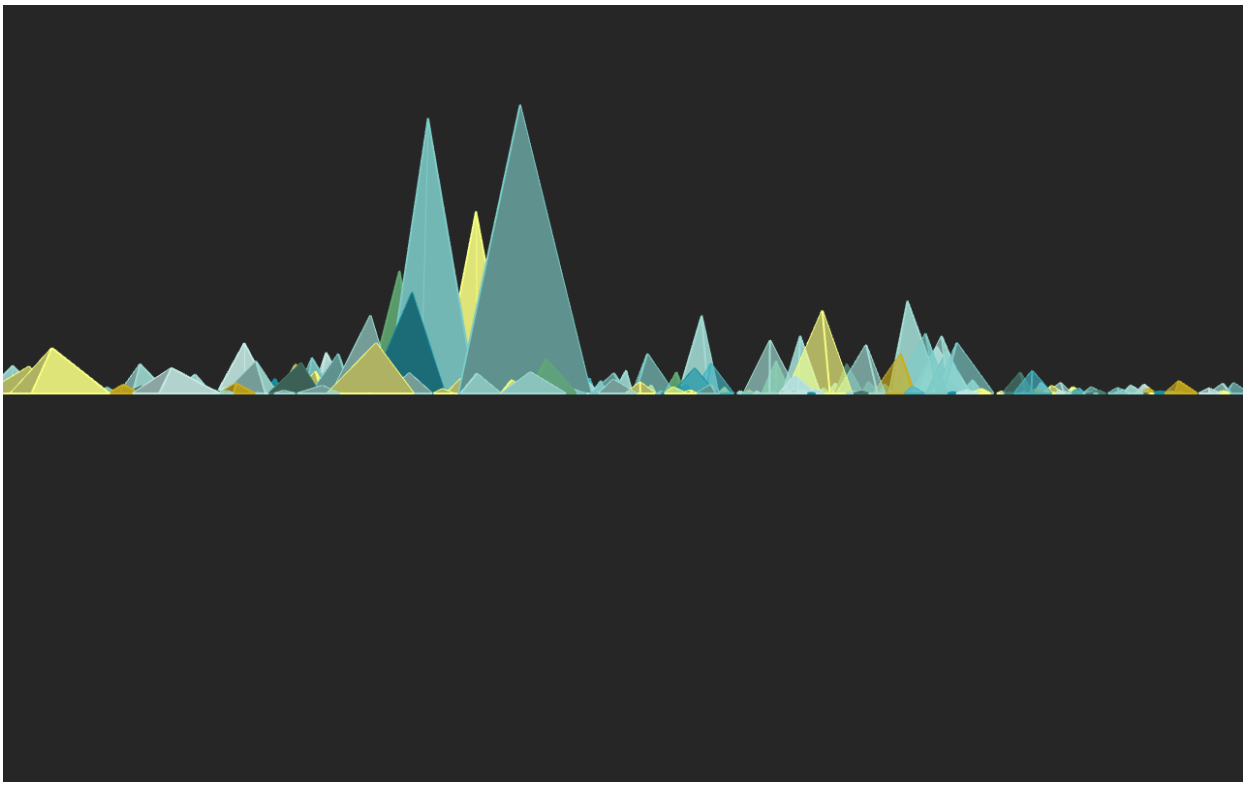
I assigned colors within the range of light blue hue's to deep green and yellow to short string lengths to longer string lengths. For example, shorter word lengths such as "army", "more", "with" were given a light blue color while long words such as "watercolor" or "elephants" were given a yellow color.

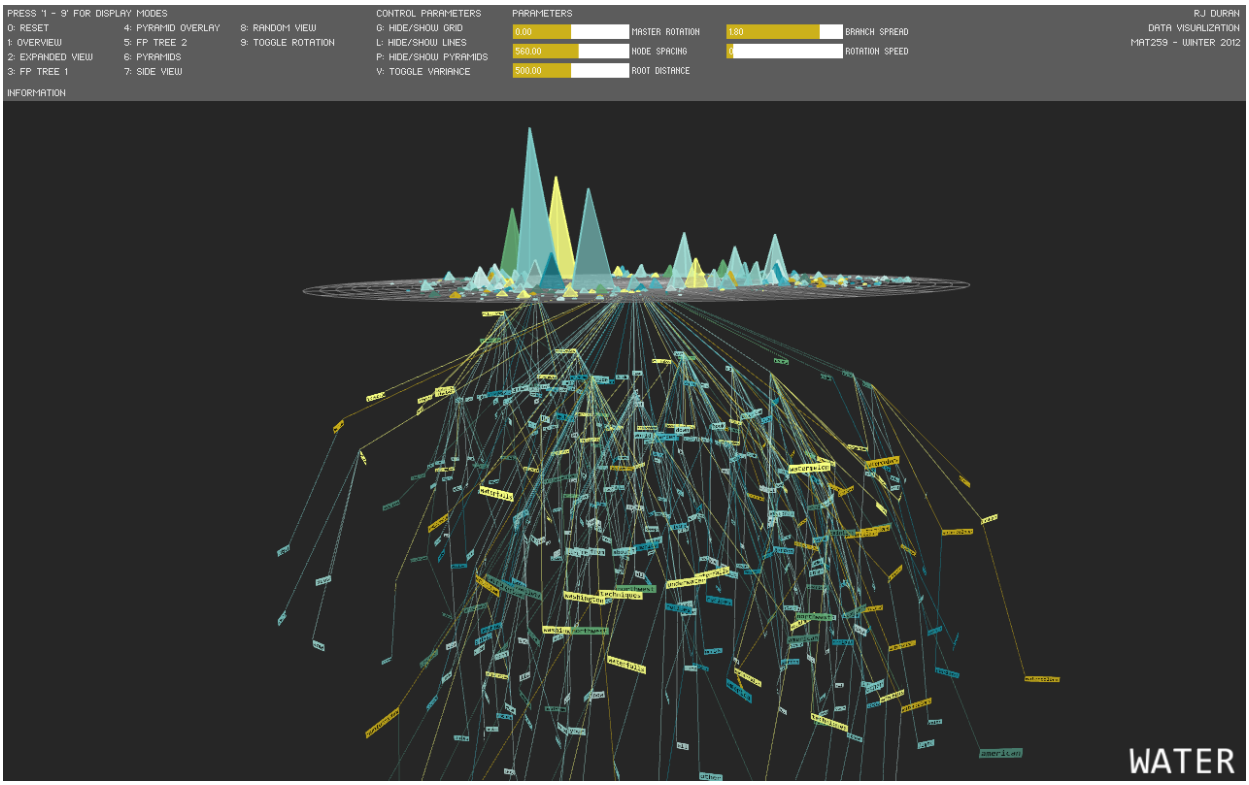


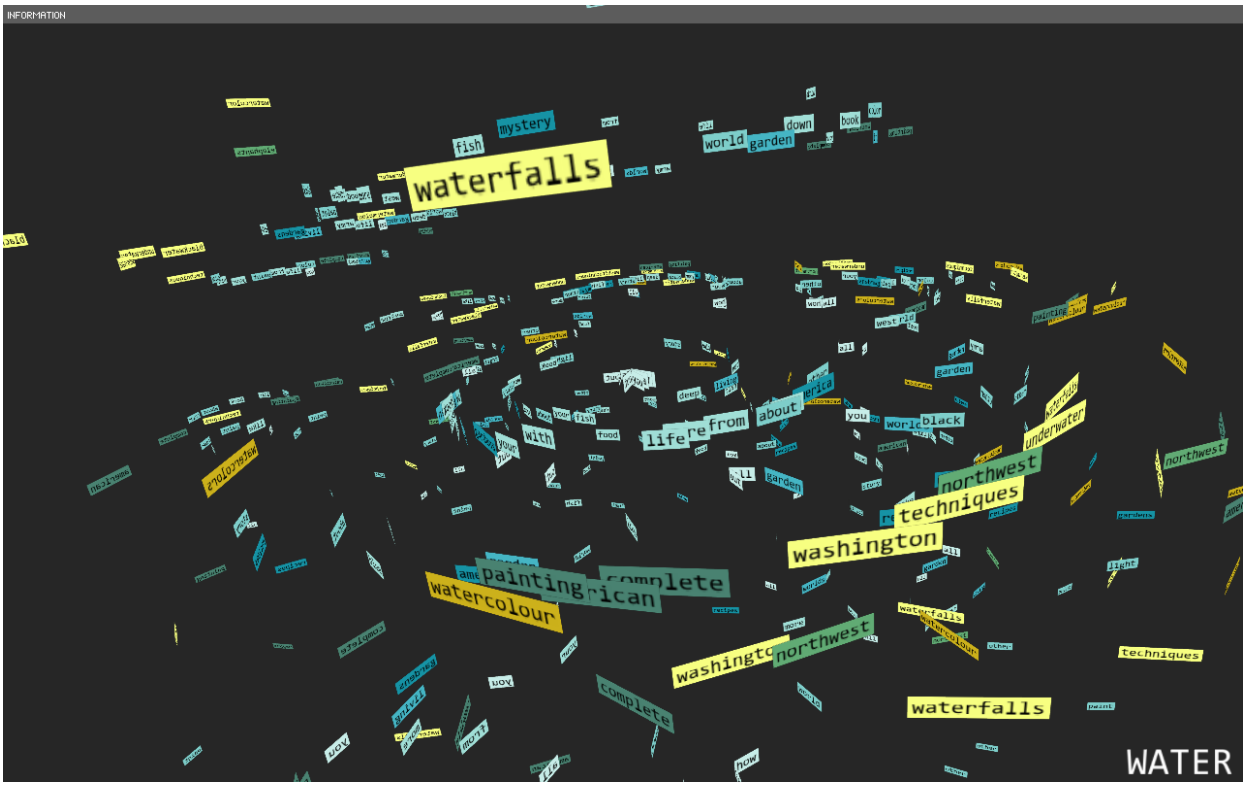

waterfalls
watermelon
washington
techniques
underwater
american
northwest
waterfalls
washington
techniques
washington
your
watercolors
other
other
techniques
washington
book











Analysis

Overall this project really pushed my designing abilities, which were very minimal at the

beginning. I explored methods for coloring data and representing layers of meaning within a 3D space through an interesting data set.

It also pushed my data filtering, searching, and sorting abilities to look for interesting ways of representing the data. From the data I was able to highlight the connections between water and the occurrence of each word. This string of terms shows the number of times a word happened to appear in a title.

novel:13018 watercolor:9204 elephants:8238 guide:7197 with:5729 from:5460 painting:4064 life:3162 underwater:3148 how:3104 your:2800 techniques:2714 color:2580 book:2437 watermelon:2416 watercolour:2412 watercolors:2366 sea:2328 mystery:2326 black:2302 gardens:2300 you:2267 fish:2198 garden:2118 world:2083 out:2061 like:2041 blue:1978 american:1969 deep:1966 about:1948 complete:1885 northwest:1763 most:1753 blackwater:1735 food:1733 washington:1652 story:1618 light:1615 west:1577 waterfalls:1569 america:1559 living:1507 worlds:1485 army:1473 more:1442 other:1441 all:1434 down:1428 recipes:1418 paint:1407

What this tells us is the most common word used with water is a title is “novel” followed by “watercolor.” This probably indicates that people were checking out mostly guide books, stories about things that have to do with water and watercolor art books. Water for Elephants was also a very popular book among readers.