Save the Bees: Data Visualization Project on Bee Colony Population throughout the US

Angela Rubalcava
College of Arts and Sciences
University of San Francisco
San Francisco, California 94117
Email: amrubalcava2@dons.usfca.edu

***Abstract*—Bee colony populations across the US have been rapidly decreasing and affected by a multitude of stressors. Our ecosystem is not prepared for a world without bees which means we must spring into action and begin counteracting what is killing our bee population. In order to bring more attention to this issue, I created five different visualizations that portray bee colony populations in the US and some of the issues bee colonies are facing. The five visualizations consist of a US heat map, a parallel coordinate chart, two line charts, and a bee swarm. The goal of these visualizations was to best communicate information to the average viewer as accurately and straightforward as possible, which I was able to do after a few attempts.**

I. Introduction

This project will be able to communicate the rate at which bees are dying and what is contributing to their demise through easy-to-read charts and novel ways to engage the reader/viewer to care more about not just bees, but the well-being of our planet. The project objectives were to provide:

- overview of the bee colony population of every state in the US.
- insight as to what variables/stressors have caused bee colonies to die.
- comparison between bee colony populations across different regions of the US
- comparison between bee colony population loss and bee colonies added

II. Related Work

In the process of deciding what project to choose, I read a few articles on the topic of bee population issues and the potential for drastic change in our ecosystems if we do not attempt to rebuild bee population numbers and protect them. Despite colony loss being a huge problem for the future of our ecosystem, not enough action has been taken [12]. The articles I read are responsible for my choice to pursue this topic and attempt to create a visualization aid to alert people of the issue at hand. We often see statistics being shown instead of images or visualizations to allow the reader/user to truly imagine the scale of bee loss. I believe that visualizations will be far more engaging and the visual aspect will be hard to ignore. You can forget percentages [6] but can you forget the steady increase in bee colony loss in a line chart? I credit these articles to be the motivation for this project and for specifically choosing this topic.

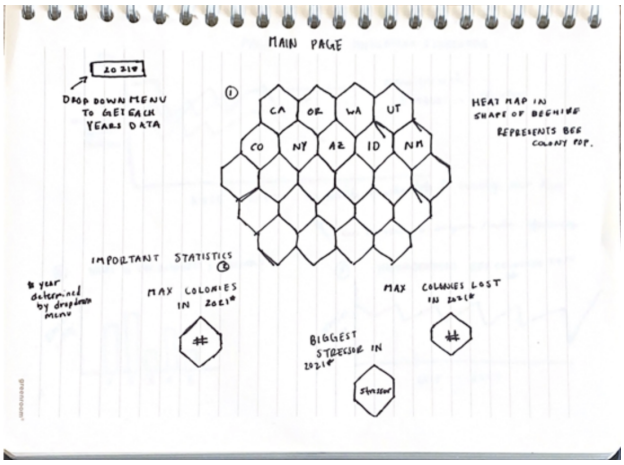
In order to ensure I understood the data, I looked into some articles, research papers, and books to provide context. I was able to find a resource that highlights bee colony loss due to stressors that includes pests and pathogens as well as looking into what other environmental, agricultural, and socioeconomic changes may be responsible [8]. This article was instrumental in my decision to include a chart that highlights the stressors often found in bee colonies which led to the creation of my bee swarm chart. I also wanted to increase my general knowledge of bees and the way they operate. This led me to a book titled *Bumble Bees of the Western United States* [4]. I was able to read excerpts of this book to familiarize myself more with bees. They are fascinating creatures that are integral to our ecosystem and deserve to be protected. The reading really allowed for myself to gain more background knowledge on western bumble bees and their changing role/impact due to climate change.

I used used d3.js [1] to create all my visualizations, albeit different versions at times, which is a very useful JavaScript library. One of the websites I visited to further my understanding of bee colony changes throughout the US inspired my heat map for bee colony population [9]. They have a great US Map that highlights Colony loss throughout the US. This gave me a start point for creating my heat map, but I also used a similar example of a US hex heat map [7] to get started. I also relied heavily on previously created visualizations on D3 as reference points for my own visualizations. For my parallel coordinate chart, the D3.js graph gallery had a great example [3]that allowed me to incorporate the hover effect and color code for the different regions. I also used many of the principles taught in *Fundamentals of Data Visualization*, especially for use of color [10] in the parallel coordinate chart. I originally wanted to emphasize each state, but coloring each state with its own color would be confusing and far too much for a user to have to remember. For the bee swarm I created, I used Mike Bostock's [2]example as a reference point and then expanded from there to customize it to my needs. When creating all of my visualizations, I attempted to remind myself the importance of proportional ink [11] in visualizations but especially a large project like this. I wanted to ensure I did not overwhelm the user with too many components in my visualizations. I also added captions to the web page to ensure their was non-data ink that provided context and necessary information for each

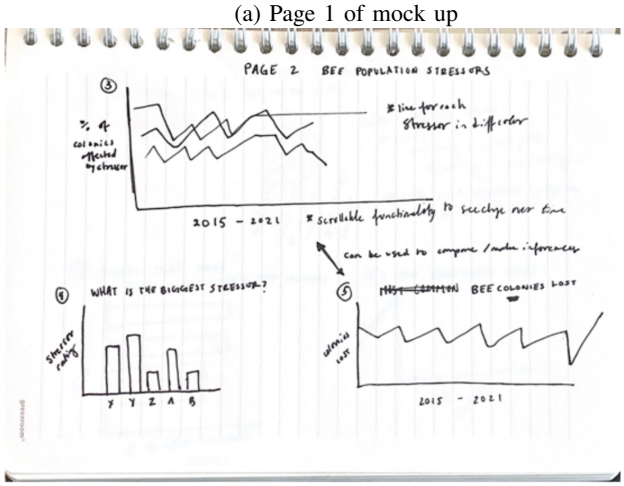
visualization.

III. Approach

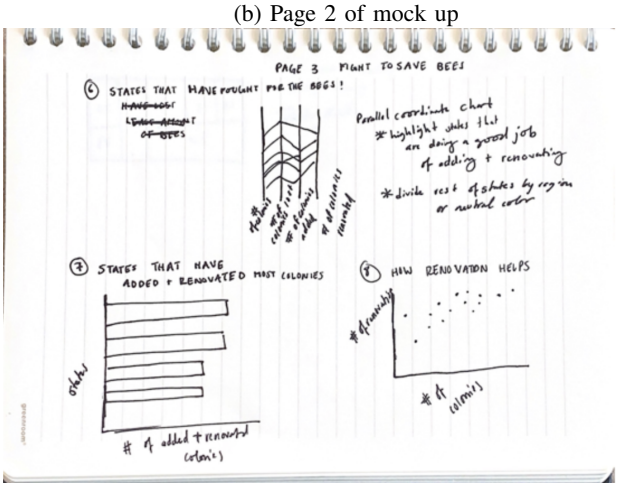
When I first began the project, I had planned to create 8 different visualizations. In Figure 1 you can see a mock up of what I had originally planned.



(a) Page 1 of mock up



(b) Page 2 of mock up



(c) Page 3 of mock up

Fig. 1: Mock up of original visualization plan

I believe I may have been too ambitious with the amount of visualizations I had planned. Eight separate visualizations would have been a lot of work but it would have also been difficult to justify 9 different visualizations. Since the data I used only contained so much information, to create eight different and meaningful visualizations seemed almost impossible. Not only that, when I first began the project I did not sift through the data enough and conclude how it would appear when using the visualization techniques I had planned. For example, I had planned to use bar charts to show the biggest stressors bee colonies faced and the top states that had renovated and added colonies. In figure 2 and 3, you can see the attempts to create these bar charts.

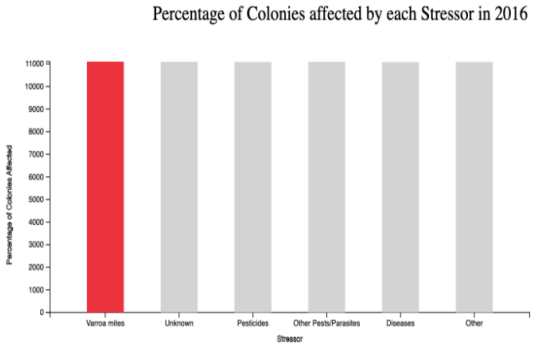


Fig. 2: Bar chart created to portray the biggest stressors bees face

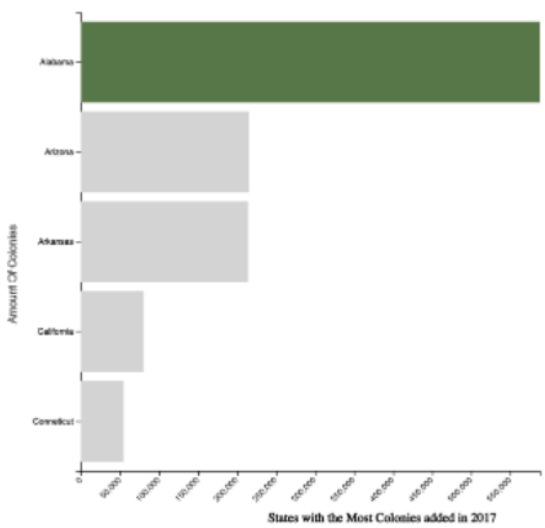


Fig. 3: Bar chart created to portray the top 5 states that had renovated and added bee colonies.

As you can see, the bar chart in figure 2 hardly pointed out a difference, even with the bar with the highest value being highlighted. Due to this, I decided to scrap this idea and instead focus on creating a chart that would better utilize the data to show both the different stressors bees face, but also how

common and effective they were. This led me to create a bee swarm that was able to show the different stressors as well as how effective they were. Figure 3 is an image of the bar chart I had created to show the top 5 states that had both added and renovated the most amount of bee colonies. While this information may have been interesting, was a bar chart the most effective way to do this? Highlighting the only top 5 states felt too exclusionary and did not provide more insight to users. It was also somewhat redundant as users could use the parallel coordinate chart I had created to compare these different values, albeit it would be comparison by region rather than state. These failures led me to reexamine almost all my charts and decide whether their inclusion was meaningful and added value to the overall project. This led me to eliminate my originally planned bar charts, scatter plot, and statistic display as they were repetitive with the information they disclosed.

The pieces I had to implement to execute my new approach was to truly look over the data I had and decide which techniques would best fit my data. From there, I would edit the data to isolate only what I needed. This led to much higher quality visualizations and much more informative ones as well. When implementing this approach, I found that the way I had edited and read through the data was extremely inefficient. I was using Google Sheets to parse through all the data and edit what was necessary. Had I used a different software or method, I may have been able to save a lot of more time and effort that could have been put toward the actual visualizations.

IV. Results

The main goal of the project was to ensure that the visualizations I created aligned with the project objectives. The project objectives I had set out were to provide:

- (1) overview of the bee colony population of every state in the US.
- (2) insight as to what variables/stressors have caused bee colonies to die.
- (3) compare bee colony population across different regions of the US
- (4) compare bee colony population loss against bee colonies added

I believe I was able to meet these objectives with the visualizations I created.

The heat map seen in figure 5-6 meets the first objective as users are able to compare the bee colony population across states in the US. The map contains hovering functions that allows for the user to hover over a state and the exact value will be displayed. The user can also use the legend to determine the amount of bee colonies present in that state. There are also two buttons: a back button and drop down button. The back button allows the user to return to the home page if they are finished viewing the visualization. The drop down button allows for users to select a different year and view the heat map for that year instead.

The parallel coordinate chart seen in figure 7-9 fulfills objective 3 as users are able to see the number of bee colonies, number of bee colonies lost, number of bee colonies added,

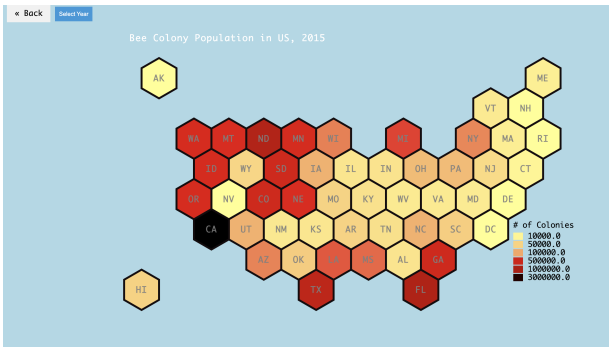


Fig. 4: This heat map of the US is for the bee colony populations for the year 2015. The visualization contains a legend [5], a back button, and a select year drop down button. The legend allows for the reader to compare the value the given state has to the legend. You can also hover over the state and it will show the exact value. The back button takes you back to the main home page and the drop down menu allows you to select a year from 2015-2021 and view the heat map for that year.

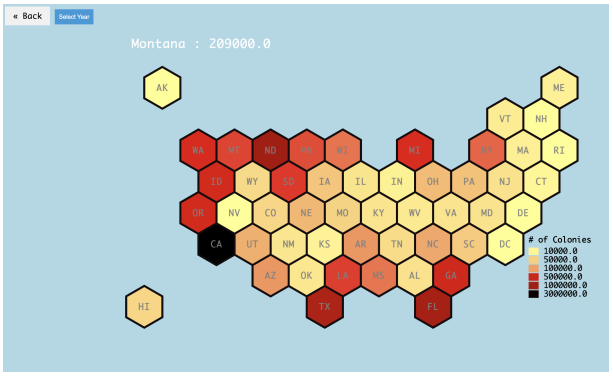


Fig. 5: In this figure, you can see the hover effect in use. We are currently hovering over the state of Montana. This causes for the name of the state to be shown as well as the bee colony population for that state.

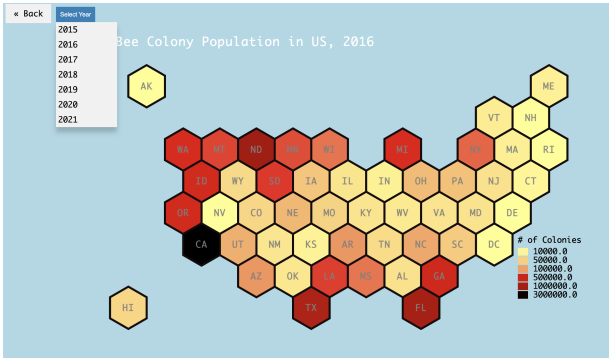


Fig. 6: In this image, you can see the drop down menu functionality. When you click the drop down button, you can select a year you would like to view and you will be take to the heat map for that year.

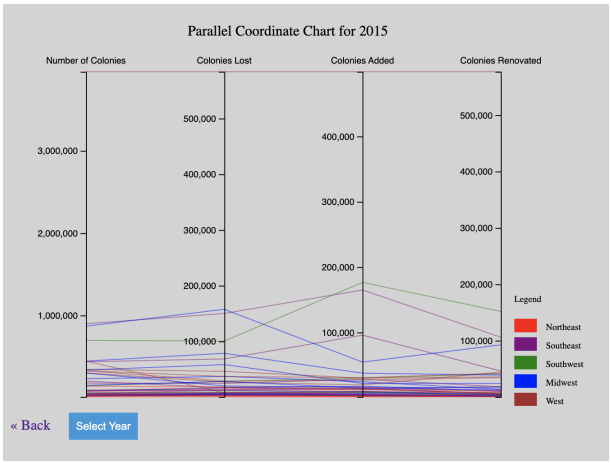


Fig. 7: This parallel coordinate chart for the year 2015 shows the number of bee colonies, number of bee colonies lost, number of bee colonies added, and number of bee colonies renovated. Each line represents a state and the lines are color coded to represent the Northwest, Southeast, Southwest, Midwest, and West regions of the US. There is a legend [5] that specifies the color to the specific region. You can hover over a line and it will cause only the lines that pertain to the same region to be visible. There is also a back button to return to the home page.

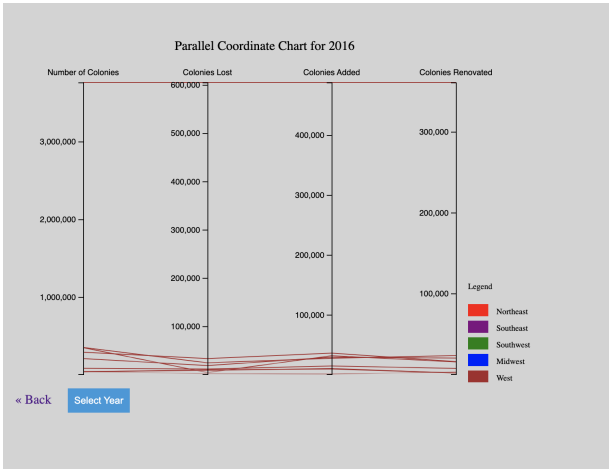


Fig. 8: In this figure, you can see the hover effect in use. We are currently hovering over a line that pertains to the Western region of the US. This caused for only the lines that belong to the Western region of the US to be visible. This allows for easy comparison between states in the same region. Once you stop hovering over a line, all regions will appear again.

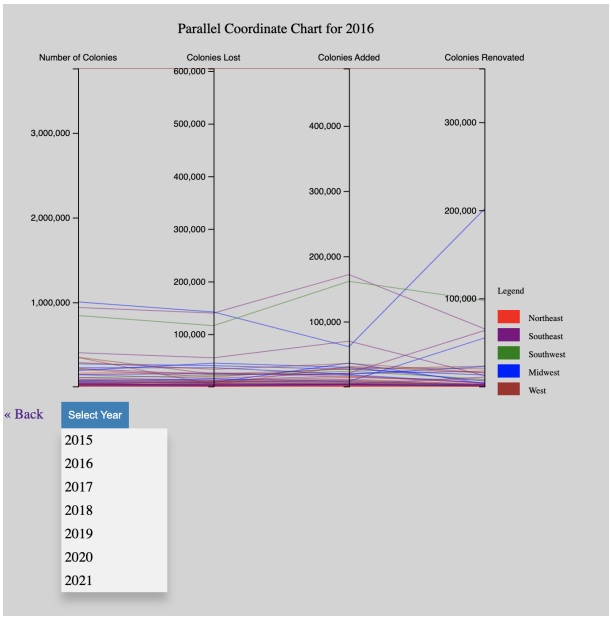


Fig. 9: This figure shows the drop down menu in use. You can click on the ”Select Year” button which allows for the user to see the years and select from 2015-2021. After selecting a year, the user will be take to the parallel coordinate chart for that year.

and number of bee colonies renovated. The separation of the lines by regions allows for users to make easier comparisons between regions. The hovering effect makes it easier to do this as well as you can isolate specific regions you would like to see. It also gives insight beyond bee colony population as you can also compare bee colony loss, colony additions, and colonies renovated. The back button allows for users to return to the home page. The drop down button allows for the user to select a year from 2015-2021 and view the visualization for the selected year.

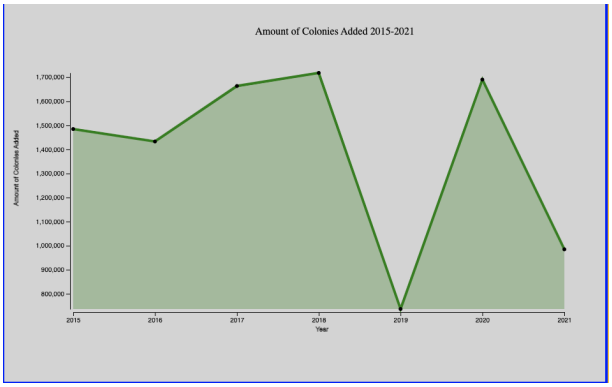


Fig. 10: This line chart maps the change in bee colonies added throughout the entire US. The line chart is shaded to emphasize the difference in change year to year. The user can hover over the point for a given year to see the exact value.

In figure 10-11, you can see that part of objective 4 is met.

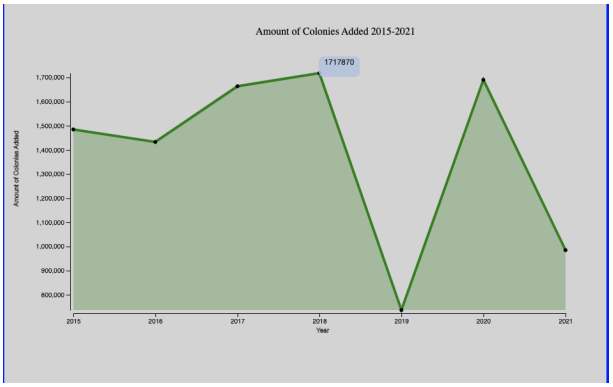


Fig. 11: In this figure, you can see the hover effect in use. We are hovering over the point for the year 2018 and the value is shown as a tool tip.

A line chart is used to show the change in amount of bee colonies added to the US throughout the years 2015-2021. The bottom of the line chart is shaded to emphasize the change and bring attention to both the increase and decrease throughout the years. The hover effect allows for the visualization to be more effective as the user can hover over the dot for a specific year and receive the exact value.

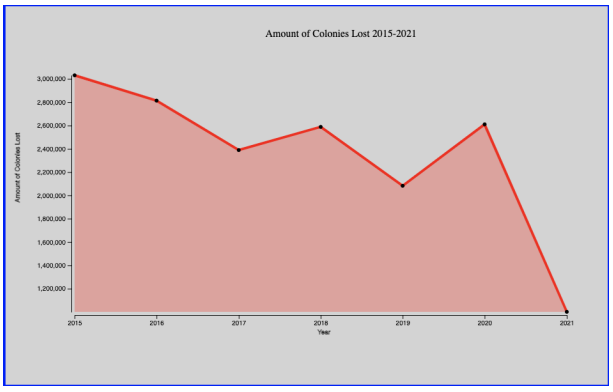


Fig. 12: This line chart maps the change in bee colonies lost throughout the entire US. The line chart is shaded to emphasize the difference in change year to year. The user can hover over the point for a given year to see the exact value.

In figure 12-13, you can see that the rest of objective 4 is met. A line chart is used to show the change in amount of bee colonies lost to the US throughout the years 2015-2021. The bottom of the line chart is shaded to emphasize the change and bring attention to both the increase and decrease throughout the years. The hover effect allows for the visualization to be more effective as the user can hover over the dot for a specific year and receive the exact value. Both the line charts that present colonies added and lost allow for the user to compare the change in both throughout the years. The user can ask themselves if the amount of colonies added truly compares to the amount of colonies lost every year.

In figure 14-16, objective 2 is met as it provides users insight

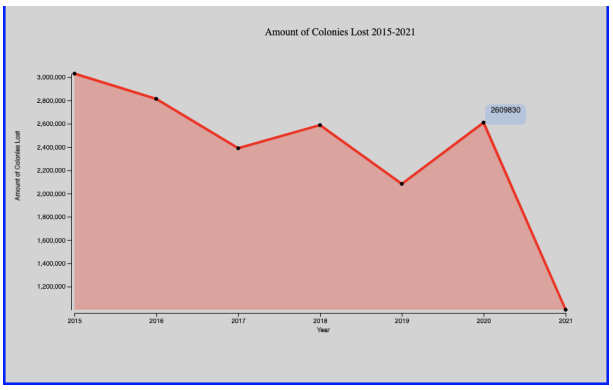


Fig. 13: In this figure, you can see the hover effect in use. We are hovering over the point for the year 2020 and the value is shown as a tool tip.

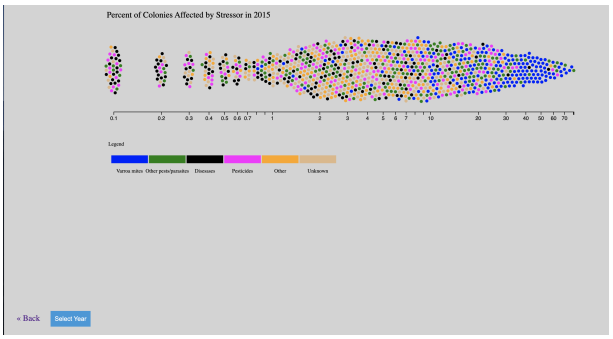


Fig. 14: This figure consists of a bee swarm that plots each stressor found in bee colonies in the specified year. Each point has a specific color associated with the type of stressor it is. The visualization has a hover function in which hovering over a specific plotted point will show the value and highlight the point you are on red. This visualization contains a legend [5], a back button, and a drop down menu button. The legend consists of the 6 different type of stressors found and the color associated with it. The back button takes the user back to the main page. The drop down menu allows for the user to select a specific year from 2015-2021 and view the bee swarm for the selected year.

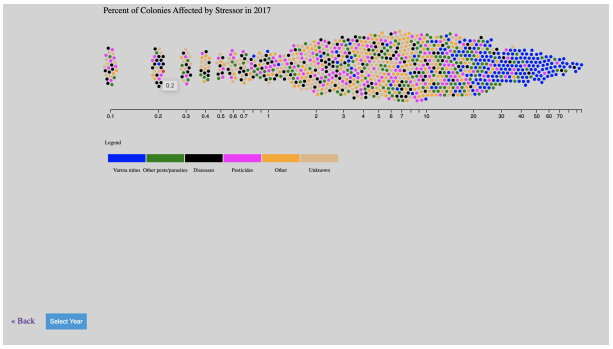


Fig. 15: In this figure, you can see the hover effect in use. We are currently hovering over a plotted point and the value for that point is now visible. The point is also highlighted red to separate it from the rest of the points.

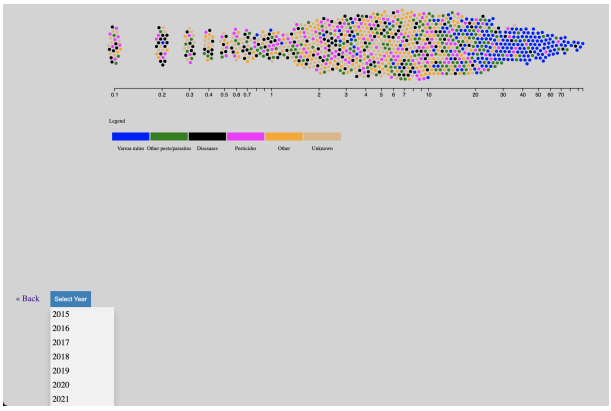


Fig. 16: In this image, you can see the drop down menu functionality. When you click the drop down button, you can select a year you would like to view and you will be take to the bee swarm for that year.

on the different stressors found in bee colonies as well as how effective they were compared to each other. This visualization consists of a bee swarm with a hover functionality, legend, back button, and drop down menu. The hover functionality allows for users to hover over a plotted point and view the exact value of that point. Hovering over the point also causes the point to turn red which allows the user to differentiate it from the other plotted points. The legend contains the six different stressors bee colonies face and the color associated with each stressor. The back button allows for users to return to the home page. The drop down menu allows for users to select a year from 2015-2021 and view the bee swarm for that given year.

I measured success by ensuring that the visualizations I created were clear, concise, and straightforward. I wanted to produce visualizations that the average user could understand and come to their own conclusions regarding the state of bee colonies in the US. I ensured the visualizations were not misleading by receiving feedback on my visualizations and fixing anything that may have been confusing. I also wanted to ensure the visualizations I created were meaningful. I did not want to create charts that were weak in what they were attempting to portray. This led me to downsize and focus on quality charts and useful interaction. Each chart has an interaction that allows for users to gather more specific information to truly have all the data. Providing as much information as possible while creating appealing charts was essential to this project and it is what I based my success on.

V. Discussion

I found the approach I took to be promising but delayed. When I first began the project, I wasted a lot of time on ineffective visualizations. Once I examined the data I had, I chose visualization techniques that would best fit with the data I had and what I was attempting to portray. After selecting a technique and sifting the data to only contain what was necessary for that visualization, I had to decide how I could make the visualization interactive for the user. I wanted the user to be able to interact with the visualization and gather more information from that interaction. From here, I added any other necessary buttons or information necessary such as back buttons, drop down buttons, and legends. The goal was to create a valuable and informative visualization that anyone could understand and gather information from.

An approach that may have been more effective would have to do with the way I edited and sifted data. I would have liked to have used a method that automated the process of extracting data I actually needed from the CSV files I used. I did it all by hand essentially through Google Sheets which was a slow and repetitive process. I would have also liked to possibly found a way to find a more effective way to show the data from 2015-2021 without using a drop down menu to see the different charts for the specified year. I believe the eight different charts for each year may be a little redundant and an ineffective way to view change. I am not sure what approach may have been better but I believe if I had come to my original approach quicker, I would have found a better method of managing multiple charts for multiple years.

I learned a lot from doing this project. I had previously not created such a step-by-step project that included multiple releases, reviews, feedback, and preparation. When I first began the project, I was somewhat naive with the process and did not fully weigh the data I had against the techniques I had originally chosen. Our first feedback session grounded me and caused me to realize that I was prioritizing quantity over quality. I was creating ineffective charts that did not communicate the data well. I had to adapt quickly and develop new visualizations using different techniques, some I had never used before. I also learned the steps to fully finishing a project, from the prototype to the final user manual. I gained valuable experience fully creating and developing an in-depth project on my own.

If I could change my approach I would have definitely prioritized by data a lot more. I would have examined the numbers a lot closer and found a more efficient method to process all the data. I then would have chosen visualization techniques based on the data rather than simply choosing just any technique. This would have granted me more time to create additional visualizations and added more interactions to my visualizations I had already created.

VI. Future Work

I would have liked to add and change a few things in my visualizations and web page for the project. I would have liked to add a hover tool tip to the parallel coordinate chart that would allow users to hover over a line and see the specific state the line belongs to. This way the user could also compare states and not just regions. I also would add an animation technique to both the colonies added and colonies lost line charts. The line for the chart would have rendered as the user scrolled to the line charts or possibly clicked a button. This would have made the changes seen throughout the years more effective as the user could see it rendered in real time. Lastly, I

attempted to embed the charts into the index.html page for the project website but was unable to do so. I would have wanted to do this so I could possibly add a scrolling element to the website. I hope to pursue these changes in the future and to continue improving this project.

REFERENCES

[1] Mike Bostock. Data-driven documents.
[2] Mike Bostock. Beeswarm, May 2021.
[3] Yan Holtz.
[4] Jonathan Koch, James Strange, and Paul Williams. *Bumble Bees of the Western United States*. U.S. Forest Service, 2012.
[5] Susie Lu. D3 svg legend (v4).
[6] Oliver Milman. Us beekeepers lost 40% of honeybee colonies over past year, survey findsoliv, Jun 2019.
[7] Sundar Singh. D3js: Us hex map, May 2019.
[8] Kristine M. Smith, Elizabeth H. Loh, Melinda K. Rostal, Carlos M. Zambrana-Torrelio, Luciana Mendiola, and Peter Daszak. Pathogens, pests, and economics: Drivers of honey bee colony declines and losses - ecohealth, Feb 2014.
[9] The Bee Informed Team. United states honey bee colony losses 2020-2021: Preliminary results, Jun 2021.
[10] Claus O. Wilke. Fundamentals of data visualization.
[11] Claus O. Wilke. Fundamentals of data visualization.
[12] Josh Woods. Us beekeepers continue to report high colony loss rates, no clear progression toward improvement, Jun 2021.