Colors of the Desert

Lin Guan
Arizona State University
lguan9@asu.edu

Niharika Jain
Arizona State University
njain30@asu.edu

Mudit Verma
Arizona State University
muditverma@asu.edu

Parth Khopkar
Arizona State University
pkhopkar@asu.edu

1 INTRODUCTION

Notions of “beauty” affect where a landscape lies in a person’s scale of important biodiversity regions. [9] defends why deserts are not as “dull” or “boring” - a common misconception that many of us have. We focus on one such region of Arizona, USA - the Sonoran Desert. If you ask a person who has never been there to describe what it may be like, one is likely to hear a description of a barren wasteland which subdues any flowering plant that might dare to brave the scorching heat. The only flora you are likely to see, they would say are the prickly cacti that can adapt to the scant water available in the desert. They would be surprised to find that the Sonoran desert is in fact home to some of the most vivid flowers that you can come across. In this project, we have made an effort to showcase some of these flowers and make it easier for people to discover the beautiful flora of the southwest. Though many resources are available on the web for people to find more information about a particular plant, they are generally structured to serve as databases for scientists rather than user-friendly websites that one can browse for leisure. We take one such website, the SEINet data portal [6], and present some of the most interesting flowering plants using an intelligent, interactive visualization that indexes the flowers by color. We used data mining, an unsupervised learning technique, and an off-the-shelf text summarizer to populate different elements of this visualization.

We begin by discussing the motivation of this project in Section 2, followed by some of the related works which have similar goals as ours in Section 3. We visit the components of our visualization in Section 4 to overview the curation and scalability of the required data to populate the visualizations in Section 5. We then identify and answer some plausible research questions our project tackles in Section 6. Finally, we shed some light on our evaluation plan in Section 7 for this project and we present our conclusions in Section 8.

2 MOTIVATION

The motivation for this work is to allow lay users to experience the beauty of the desert. [30] talks about the theory of beauty from different standpoints, but in a general sense, does agree that vivid colors of spectra engender the notion of beauty to objects. This motivates the use of colors as a tool to attempt to bring a change in the perspective of humans about deserts from “dull” to “beautiful” vis-a-vis color. The work captures several interesting components of the meticulously catalogued flora in the SEINet data portal. The SEINet database is designed, for robust specimen management [17]. The project’s interface has primarily been developed for specimen digitization to aid local institutions and researchers. The plants are indexed by their scientific names and the descriptions are verbose, confusing, and overwhelming for lay users attempting to view the “beauty” of desert flora. The website, while a phenomenal endeavor to showcase the southwestern biodiversity by inventorying the myriad plants in Arizona and New Mexico, is not accessible to the common, curious mind. The particular research questions our work aims to solve is how to best structure and visualize well-documented scientific records in a straightforward and easily-understandable way to serve an audience of ordinary users who may not have sufficient background knowledge in environmental research or botany.

Our work depicts a visual representation of the diversity of desert flora, particularly in their color. To this end, we only investigate flowering plants; this is a small subset of the plants highlighted in the SEINet data portal. The visualization allows users to browse plants they may not recognize by name (especially scientific name). Our work juxtaposes the images of flowers so the plant data can be indexed by appearance. The user can then clearly compare and contrast the flowers, as well as see the overall diversity in the colors of Arizona plants.

3 RELATED WORK

The use of colors to elicit “beauty" is not a new concept. [12] posits that the concept of beauty is determined by the natural landscape as well as human judgment. Josh Eckert, discusses in his book "Psychology for Artists, Futurists, and Worldbuilders" [15] about how humans tend to like bright flowers over dull-colored ones which is supported by some Australian studies he cites. Interestingly, this is a whole field of research extensively utilized by animation companies to personify their characters. The key takeaway, however, is humans tend to get attracted by bright, vivid colors. Combined, these works emphasize the importance of pre-conceived notions about the use of a landscape for several activities, an important one being tourism. This motivates this project from a socio-economic point of view, to present deserts the way they are, places with colorful flora.

[7] is made for the region of Texas, USA. From a visualization point of view, it presents options to select plants by discrete, type (ornamentals, vines, etc.) and spread (width of a grown plant), but it has been aimed toward promoting water conservation. The plants in question are house plants rather than those found in the wild. [2] does a similar job, but is specific to large municipalities of Maricopa County, Arizona, USA. It lists various demonstration gardens and water conservation techniques but does not address the problem being discussed in this project.

[4] is an encyclopedia for gardeners. It does motivate the notion of beauty to the lay users, but the motivation is for ornamental purposes and has nothing to with gauging the user's perspective about deserts or landscapes. Finally, [5] gives a generic list of house plants and addresses the concept of beauty in similar ways to [4]. However, this tool is not specific to deserts and it provides short descriptions about house plants.

The related works we found missed out on the motivations that enable this project. Additionally, these tools are mostly for ornamental plants and so do not cover the natural beauty of a place. They lack features like plants spread across an area on the map, ability to visualize different colorful plants together, or referring users to other websites for an audience seeking technical/scientific details.

4 VISUALIZATION DESIGN

Our visualization design places emphasis on the colors of flowering plants. Instead of indexing plants by their scientific or even their common names, they are ordered by the dominant colors of their flowers. We identify the best images of a selection of plants' flowers from the SEINet dataset and arrange them in an intuitive rainbow spectrum. This ordering is aesthetically pleasing, and the choice of rainbow stems from people thinking about "vivid colors" as the seven distinct colors in the visible range of the light spectrum. Placing this visualization on a light background emphasizes the intensity of desert colors. The visualization is an image accordion of small rectangular strips corresponding to each plant, each image containing a blurred view of the flower. Placing this visualization on a light background emphasizes the intensity of desert colors. This component is interactive; when a user hovers over a blurred strip, it expands and displays the full, clear image of the flower. This interactivity simultaneously gives emphasis to one plant at a time while giving the user the controls to quickly navigate through the images to find one of interest. The user can then click on an expanded image to view two additional cards: a text card that displays the plant's scientific name and a summarized description from its SEINet listing, and the region on the Arizona map where the plant can be found.

Further, the map, like the image accordion, is also organized by color. Users can zoom in and out to see at a varying granularity of detail. When zoomed in fully, users will be able to see colored points on the graph corresponding to individual plants, where the color of a plant's point will be the dominant color of the flower for its thumbnail image. When the user zooms out further, the points will need to cluster together to form a larger point. The size of the cluster point represents how many individual plants are mapped to that region, and its overall color is found by finding the dominant color of all individual plant points.

The design choices of this visualization revolve around color. [35] stresses, "color helps us break camouflage. Some things differ visually from their surroundings." Here we are breaking the camouflage of the shades of green and brown which overpower the vivid colors of the desert flowers. We enhance flowering plants' colors and show their diverse range to tackle the misconceptions about the desert being a dull landscape. This visualization is both guided by and reinforces our belief that colors can be used as an effective tool for highlighting the beauty of a desert.

Rationale for Design Choices

We base our design choices on the difficulties we encountered in using the SEINet data portal. We are not well-versed in botany or environmental science, and so much of the webpage was not accessible to us. We were also inconvenienced by the organization of the database; the starting page is a long, textual representation of plants in a list, from which a user bounces from hyperlink to hyperlink to access information regarding a plant of interest. We prefer to provide a design focused on a graphical representation and with very little hyperlinks, but from which a user can still find relevant details.

To begin, our rationale to index our visualization by images rather than text is supported by studies in conceptual and perceptual memory [32]; images permit better recall and retention rates than text. Humans have a better memory for pictures than for words, and this is due to better ability of conceptual processing from viewing a picture than reading corresponding words. It is interesting to note, though, that there is not a large discrepancy in the ability of a viewer to process information from pictures with varying levels of detail. From this, we understand that a blurred image, while working to shift emphasis and a user's gaze somewhere else, is not a less descriptive method of visualization than a clear one.

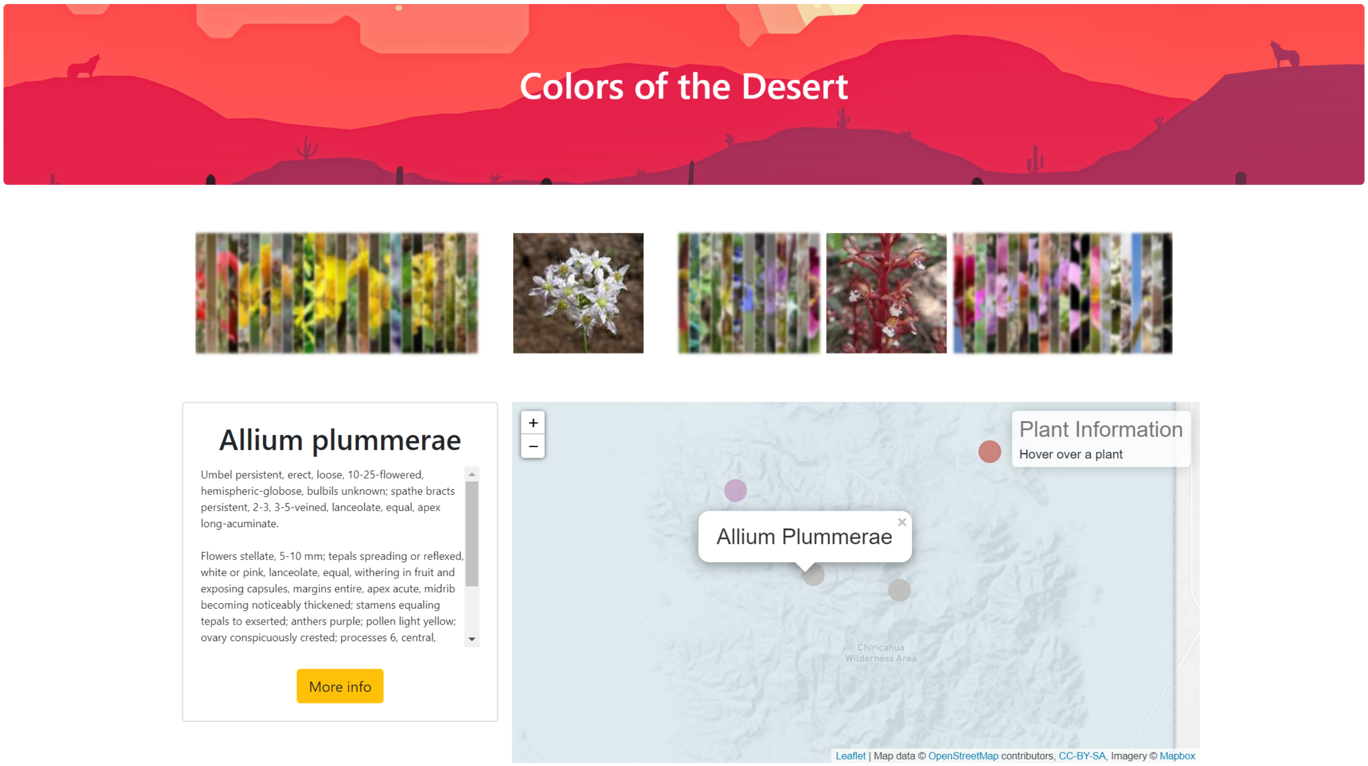


Figure 1: Screenshot of the visualization webpage

Our design follows the principle of the "Visual Information Seeking Mantra" [29]: "Overview first, zoom and filter, then details-on-demand." The image accordion visualization provides an overview of the plants in its initial view, with all images showing an abstract, fuzzy representation of their flowers. This overview manages to show 90 plants at once, but does not overwhelm the user; it condenses large amounts of information for the user to process quickly and naturally, with little effort. The zoom-and-filter represents how the user can browse through the images in the accordion; each image the user hovers over "zooms" in by expanding. The other images are not expanded, so this effect puts emphasis only on an image of interest, much in the same way filtering would do. Only one flower at a time is shown in its full size, and the surrounding images in the accordion are still blurred, so users' gazes are naturally kept where their cursors are. Finally, the user can access details on demand. Upon clicking on an image, two cards are shown offering information about the plant. There is one card containing textual information with the plant's scientific name and a short description. To the right of this card is an interactive map, so the user can visualize the location data of this plant and compare it to the others' locales. If the user is still interested in more information, we include one hyperlink to visit the SEINet listing for the plant. We also seek to make our visualization memorable. The authors in [10] perform a comprehensive study of visualization type on human memorability to propose a visualization taxonomy. Maps are the fifth-most memorable visualization type they study, after matrix, network, diagram, and area charts. Perhaps more interestingly, they find that the memorability of a visualization is directly proportional to the number of colors it has: the more colors, the more memorable. Visualizations with at least seven colors receive the highest memorability scores. We use these two facts to create our visualization, keeping color in mind. We still aim to avoid adding "chart junk," where the ink on the visualization does not relay any meaningful information about the data, so we opt for a minimalistic approach to introduce color and a map into our work.

We also carefully choose a design for the map to allow the user to intuitively compare and contrast plants by their locations. Unlike the proportional symbol maps in [11], which scales icons by the quantities they represent, we directly add a label to the icons to tell the number of plants they represent. This reduce the user's burden to understand the abstract "concepts" embedded in the map. We also add interactivity to this section. We follow the information-seeking mantra for our map and make it zoomable. Zoomable maps are a powerful tool to overlay information, so we can see how nearby plants are similar, as well as changing the level of detail to better contrast specific plants. We demand reality, but we understand the simplicity.

Finally, we follow established design principles for the card of text. From [19], we know that one of the easiest design mistakes to make in a user interface is to include *too much* text. We still opt to keep the textual description of plants in our visualization, but to minimize unnecessary text, we summarize the original SEINet listings' descriptions. We minimize the disruption to reading by left-aligning our text so the eye goes back to the same horizontal position at each new line. The description also uses a sans-serif font, which is more legible. We include a centered heading whose font size is significantly larger than the rest of the description text, so the user's eye is drawn to the most pertinent information about the flower.

Overall Interface

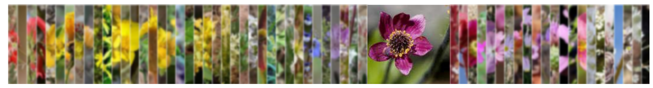
We used the Bootstrap framework [33] to design our interface in a manner that it is accessible across a wide variety of devices and screen types. The use of Bootstrap allowed us to develop the website faster and try out various design ideas. Fig 1 shows the layout organization. The banner at the top states the purpose - which is to highlight the "Colors of the Desert", which ironically, is placed over an image which only has shades of muddy red. This, in some sense, highlights the issue at hand, dealing with the misconception about desert having not many colors. A horizontal Image Accordion spans the top of the web page and allows interaction. This is an ideal position for the accordion as it immediately follows the heading. It appears centered at the top of the page. To adjust the content of the remaining web page there is consistent white-space padding on either side. The click-to-highlight feature of the accordion helps users grasp which image is being discussed by other visual elements at all times. Users can click an image and hover over other strips to compare two flower images at a time.

Immediately below the accordion is a plant description box. Beside it, a plant location map is present. The description box is scrollable owing to differing description text size and is accompanied by a "More Info" button that lands the interested users on the corresponding SEINet plant webpage. The map is interactive and by default locates the selected plant highlighting the location by the flower color on the map. Later sections discuss more on the interactive capabilities of these components.

Image Accordion

The accordion image browser (see fig. 2) is implemented using HTML and the D3.js library. A D3 function is used to read the JSON file containing details of the plants. This

(a) Initial view of the image accordion



(b) A single flower expanded

Figure 2: The image accordion

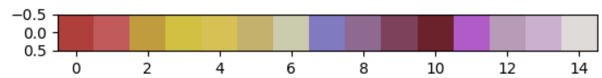


Figure 3: The final order of the 15 dominant color centroids

function generates the required hierarchy of HTML elements which when styled using CSS, give the desired effect of an accordion which expands the image you hover over (fig. 2b). When one of the flower images is clicked, the image stays expanded and another function is triggered which changes the data in the information card as well as focuses the map on the coordinates of the plant locations. This function is also responsible for removing the blur effect from the selected image.

The images are ordered to roughly follow a rainbow spectrum. There are two steps to this process. First, for each image, we find one value to represent the color of the flower [1]. We load each image as a Numpy array and use an unsupervised learning algorithm to cluster each pixel. We use k-means and hand-tune the number of clusters to best fit the flower image so that one of the dominant colors is similar to the visual appearance of the flower. This k-value is usually between 4-6. We attempt to automatically filter out dominant colors that do not correspond to the flower, such as blues of the sky, browns of the background, or greens of the stem and leaves. We verify this process, and in the cases where not all irrelevant colors can be filtered or where a flower's color is pruned, we manually choose the cluster centroid returned by k-means that matches the flower color. This process results in one color corresponding to the flower per image. We store the hex value of each of these colors with the name of the plant it describes.

The second step, once we have one hex value for each image, is to cluster the images to order them. Our goal is a rainbow spectrum, so we experiment with various numbers of clusters. A small cluster size is not sufficient because the number of flowers for each color in the rainbow are imbalanced; the majority of the flowers are yellow, so using a small k-value of around 4-5 returns suitable centroids in that region of the spectrum for yellow and red flowers, but never returns a blue centroid. We find that a centroid appears in the blue region of the spectrum with a k-value of at least 15. We manually arrange the resulting clusters to follow a spectrum, and allow the ordering of colors within a cluster to be random (fig. 3). We now have an ordering of all hex values; the final part of this step is to map these hex values back to their original flower images to obtain the final order of images used in the accordion.

Map

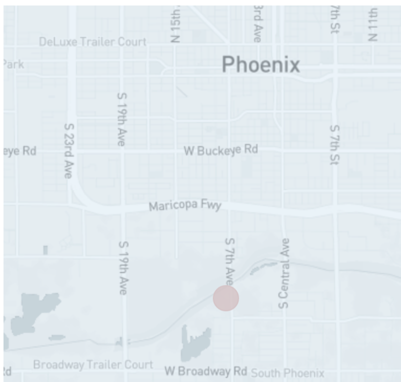


Figure 4: Color palette of the map

To implement the symbolic map, we use a light-weight open-source JavaScript library for interactive maps: *Leaflet* [21]. There are two main reasons why we choose *Leaflet*: 1. *Leaflet* is highly scalable and it has been shown to have the ability to handle over 50,000 markers at a time; 2.*Leaflet* is highly customizable, e.g. it accepts various map tiles from different map service providers. Here, we use the tile provider Mapbox [22] and choose Mapbox Light as the map theme (fig. 4) because it contains minimal geographic elements in the map and uses light colors with slight shifts in hue so that the users will not get distracted by the visual elements in the map while they are still able to recognize necessary geographic information of any highlighted location. Our selection of the map theme is inspired by two concepts data-ink [34] and visual complexity [23], which suggest that we should maintain a low visual complexity in our visualization by having a minimum number of irrelevant/decorative visual elements so that the user can perceive the message (i.e. recognize the marked location) at the first sight.

Considering that the symbols are only used to indicate the location and color of plants, we use markers of fixed size to avoid the ambiguity induced by various sizes since people often connect the size of symbols to different characteristics of plants like flower size or population size [16]. Nevertheless, there are still some potential issues of symbol map with fixed-sized symbols, one of which is the overlapping symbols

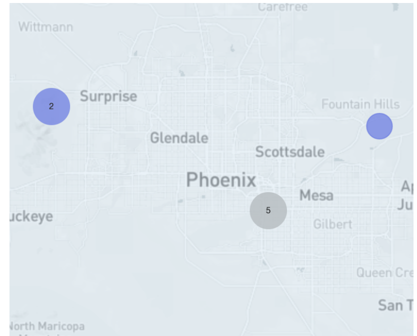


Figure 5: Example of cluster symbol

when the user zooms out, especially in those regions with high symbol density. Here, rather than using the reposition tricks, we create a multiscale symbol map [16], in which overlapping markers are coalesced into a blob when zooming out, and are de-clustered when zooming in. To imply the types of symbol (single plant or marker cluster), we use a slightly larger marker to represent the marker cluster. To be consistent with the fixed-size-symbol principle, we keep all the cluster symbols in the same size and use a number on the cluster symbol to tell the user how many plants are clustered together rather than using proportional size. Such a trick of using labels/annotations rather than modifying the color or size is inspired by the value-by-alpha trick, which visualizes one more dimension of information by varying alpha value without changing the existing colors or shapes in a thematic map [25].

Also, since the color of the symbol represents the dominant color of a plant, the color of a cluster symbol should also represent the dominant color of a group of plants accordingly. Thus, whenever we cluster markers together, we run k-mean clustering algorithm again to find the dominant color (fig. 5).

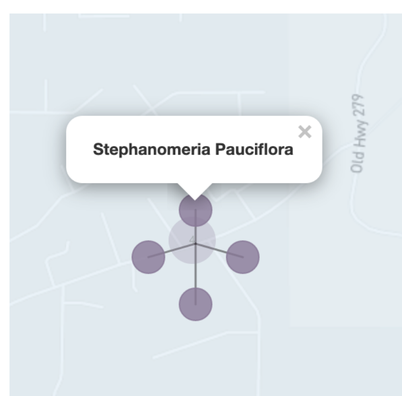


Figure 6: Example of spiderfying markers

Another notable technique we use in the map is spiderfying overlapping symbols. When plants are very close to each other or are at the same location, the corresponding symbols may still overlap when zooming in to the highest/deepest level. To deal with this extreme case, we apply a technique called spiderfying, which separates the overlapping symbols apart and uses a line to link them to their original position (fig. 6).

All the above-mentioned marker clustering functionalities are implemented with the open-source marker clustering plugin *Leaflet.markercluster* [20]

5 CURATING THE DATA

Different components bring together this visualization as explained in Section 4. All of these components take different types of data to populate the visualization we see, for instance, the map takes in the latitude and longitude information for each plant, the text description requires a summarized text, and the image accordion needs not only the flower images - but in the right order. To top it all, the visualization is cohesive, that is, the plant in concern binds what information is supplied to all of these components. Some of the challenges include being able to figure out the required plants (say, which has a flower), coming up with data collection methods that can scale to handle over 600,000 images of Arizona flora in the SEINet dataset, parsing numeric latitudes and longitudes from alphanumeric text descriptions embedded in a few of hundreds of images per plant and so on.

We have only used the data present on the SEINet portal for all purposes of this visualization, as it already has tremendous amounts of data and visualization that itself is a challenging task. The data is structured by a plant family. Each plant has a thumbnail image, a set of around hundred other images (which may be flower-bearing or dried and stored specimen images), a description box attached to it (there are several tabs within the description pane - also, there may be no description at all), and image description (browsing a particular image of the said plant gives more information like who took the image, location of the specimen and so on. The relevant information is the latitude and longitude values). Three different strategies had to be employed to obtain these three types of data. All the data was stored as JSON files. We optimized the data storage by storing links to images wherever possible, rather than downloading them.

We used Beautiful Soup [24] to obtain HTML source code for static webpages. Here we could exploit the structure of the webpage to extract the required information like the links to the images for each plant and their text descriptions. Proper error handling and certain naming conventions in HTML coding helped us build a feasible scraper that could be run to extract the 5800 plant directory. At this point, we have the plant thumbnail image, the raw text description, and image source links to each of the plant images.

Next, we modified the same scraper to traverse each image for each plant and attempted to find the latitude and longitude. A simple regex parsing was enough to extract floating points in an alphanumeric string. We further passed these coordinates to a custom function written to filter out coordinates that lie within the Arizona state boundary.

Now, we dealt with getting a text description that is concise and lucid for a lay user who intends to browse the website for general purposes, which could be solved by using a text summarizer tool. One of the ways was to implement a variant of Pointer Generator Networks [28], but given the need for the project - we found using off the shelf summarizers working enough for our use. We used [3, 8] for obtaining the text summaries, which were directly fed to our web page. Although these tools are very good, they do not provide respective APIs, hence we had to resort to using computer automation by Selenium [26] to obtain these summaries. It should be noted that we can, in effect, train a state-of-the-art text summarizer and make this process more time-efficient and self-reliant through automation. The processing of images for the image accordion is explained in Section 4.

Scalability

This project shows promise in scaling fully to the magnitude of SEINet, a huge dataset to navigate and parse. We have already collected plant data for all entries in Arizona from the data portal. The scalability constraints do not come from the limitation to mine the data, but rather from the limitation to visualize it. Since our visualization is indexed by color and shows an image for all the plants we display at once, we cannot show data to the full scale of the SEINet dataset without compromising legibility or our minimalistic approach, as our design stands. The process to select only flowering plants and find the dominant color for images involved manual effort, with human verification that some plants were irrelevant to our visualization of flowers and that our dominant-color-picking algorithm was filtering out background colors. To automate this process, we could use a commercial system for image segmentation to automatically filter out a mask of semantically irrelevant pixels in the plant images. We could also visualize our plants so the user does not see all flowers at once, but rather sees a subset corresponding to a chosen color of interest.

6 METHODOLOGY

The research questions that we pose are:

(1) What is the best way to visualize information from structured, scientifically cataloged data (such as those

contained in the SEINet portal) in a manner that is understandable by a lay user.
(2) Are visualizations capable of bringing a perspective shift for humans with prejudiced opinions?
(3) Can a visualization be treated as art?

The first question highlights an important issue: structure in data does not imply legibility. SEINet is an excellent tool for cataloging and digitizing plant specimens, but fails miserably at being accessible to lay users. The solution depends on the ability to abstract away complex scientific jargon. The first step we take to do this is to choose a different method of indexing the data. Our data structure revolves around the images of the plants, not their scientific names. This reflects a change in purpose and target audience from the SEINet dataset to our visualization: we expect the user to browse through plants by their appearance. Our target user does not have an environmental background, and may not even know the names of plants. The motivation for a user of our visualization is to explore and discover, not to refine already-honed knowledge. SEINet is built for researchers who have a background in botany. These users are using the portal with a set of plants already in mind and tap into the massive collection of data to learn more about their plants of interest. Having this audience in mind, our visualization has the main goal of not swamping the user with too much clutter or detail. Once users select an image of a flower, they will see more data for that plant. The information they find is compartmentalized into two parts: map and description. The sequence of the users' actions is controlled by our visualization, and the information is offered in a way so as not to overwhelm the user. The textual description is concise. The map uses a minimalist design, with a solid circular point marker for plants over a light overlay of major streets. Unlike the SEINet database, all these components are on the same frame of view; the user never has to click on a link to obtain a well-rounded idea of the plant. Appearance, location, and description are all displayed in cards on the same web page.

The effective visualization of this data involves effective mining. All plant names are retrieved from the data portal and used as a key. The values for these keys require images, descriptions, and coordinates. Each plant in the data portal corresponds to an average of 100 images, but we only display one. We manually select that image with the best depiction of a flower for each plant. Next, the description in the SEINet portal is verbose and difficult to sort through. We use the output of an automatic text summarizer in our description card. The location data for a plant is embedded in the description for individual images in the SEINet portal; we only choose to collect four of these locations, so there are not too many points on the map.

The organization and navigation of the visualization keep the audience in mind. One of the decisions we make is to only include 90 plants in our work, rather than creating a comprehensive catalog like SEINet's. We sample over 200 plants from the 6,000 featured in the data portal. From these 200, we hand-select only those plants which have images of flowers. This work does not display the colors of non-flowering plants because we aim to highlight the diversity of color in the spectrum, and most of the non-flowering flora are variants of green. Our image accordion does not show a textual representation (especially one which lists esoteric scientific names) of so many plants because it does not afford easy browsing. We expect our tool to be more popular because it will be easier for a lay user to find a plant of interest, by hovering through rectangular strips. Navigation is an important issue for a legible interface for lay users. We refrain from giving too much flexibility to the users because, after a certain point, too many choices confuses rather than offering additional functionality.

[18] posits that the goal of visualization is to aid our understanding of data by leveraging the human visual system's highly-tuned ability to see patterns, spot trends, and identify outliers. This aligns perfectly with our use of image accordion to display patterns in desert flora. The user can see trends to compare and contrast plants in their colors. The outliers are the colors that do not exist as frequently in the dataset; it only takes the user a quick glance to realize that there are many more yellow flowers in the desert than blue or purple.

Visualizations are also capable of shifting perspectives. Our work is aimed to change the minds of users who may believe that the desert is a dull, barren place. We do this simply by providing a depiction of the environment that may not match the user's idea of it; where an individual with a different opinion of the desert may imagine empty acres of just sand, rocks, tumbleweeds, and a sprinkling of cacti, we offer an illustration of a rainbow of flowers in Arizona. It only requires additional information for people to adjust their expectations of the world, and our visualization aims to tackle a specific prejudice lay people may have about the desert, in assuming its lack of color.

We present an evaluation plan in section 7 to provide an objective metric to further answer the second research question. Visualizations have played a key role in the digital era in shaping the ways of presenting the information. A well-created visualization can benefit the community as much as a poor visualization can hurt it. Hence, a subjective answer we believe, is yes data visualization can bring about a change in the way humans see the world. However, this project narrows this general task to specific prejudices one may have over issues that can have a socio-economic impact - perspective over the beauty of a landscape. Section 3 gives a brief into how beauty can affect tourism. It may also affect urbanization and gentrification, bringing in an influx of an affluent population and creating motivation for industrial improvement.

Finally, we explore that visualization can be an art, by creating a work that emphasizes and characterizes beauty. Previous sections (Section 2, 3) discuss the need of eliciting “beauty" of a place. Art, by definition, evokes appreciation and appeal from its audience. This is also the essence of this work: we highlight that deserts are diverse in appearance, worthy of appreciation, and full of appeal because they do not, in reality, match the caricature presented by the barren wasteland of a Wile E. Coyote cartoon. Art often functions as a commentary of the current world or status quo [13]. Big Data is now being used as a source for this art; for example, artist Nathalie Miebach converts weather data into musical scores [14]. The use of a work to create intrigue or spark curiosity is an affirmative answer that visualizations *can* be used as an art. Answering *how* they can be used as an art is not trivial; the charm of art is that there simply is no methodology to create it. In our visualization, we create art by invoking beauty and prompting users to consider a different point of view from their original biases.

7 EVALUATION PLAN

It is important that we evaluate whether our visualization captures the underlying phenomenon and helps the user understand it [31][27]. A fair evaluation would provide us with a scientific statement about the quality of the visualization rather than just personal opinion of the users. The evaluation of the visualization can be divided into following tasks:

(1) How well the visualization captures the underlying phenomenon
(2) Performance of users in performing the tasks associated with the visualization.
(3) Did the visualization make the users more informed about their opinions on deserts.

The phenomenon that we are trying to capture in our visualization is the gamut of colors of the flowers that bloom in the Sonoran desert and the information about the plants that they grow on. Based on this we would like to evaluate how well our users can navigate the different elements of our visualization while understanding the phenomenon that we intended to capture.

To this end we plan to use qualitative methods of evaluation such as in-situ observation where we present the users with the visualization and allow them to explore it without providing any further information. They would then be asked to fill out a questionnaire asking them to rate aspects of the visualization on a likert scale of 1 to 5, one being poor and five being best. They would then be explained the purpose of the visualization and asked to complete the second part of the questionnaire which they hadn’t previously seen. This part would have a few questions similar to the first and a text box to write general comments that the users have about the visualization. This first part would allow us to get the unbiased opinion of the users about the visualization which would serve as a baseline to compare against responses from the second part. This comparison would enable us to ascertain whether there are aspects of the visualization that are not intuitive. This feedback would allow us to make improvements to the design of the visualization in a direction that improves user friendliness. We can also pose questions which seeks a change in perspective of the users. A simple query asking them to list out what colors they feel are present in a desert before revealing the visualization, and the same question after exploring the tool should be enough to gauge whether we were able to create an impact in the way they look at deserts.

8 CONCLUSION AND FUTURE WORK

We have created a visualization that allows users to explore the beautiful flora of the Sonoran desert by looking at the colorful flowers that grow on the plants there. By using data mining and unsupervised learning techniques, we presented a visualization that enables a layman to access the rich, albeit overwhelming resources that the SEINet portal provides. By changing the way the user interacts with this data, our visualization is far more accessible and intuitive for lay users. For someone who is not familiar with or interested in the scientific taxonomy, our visualization offers a much better and guided experience similar to what one can expect at a botanical garden. Through this work, we were also able to explore the role that perception and appeal of colors plays in a visualization.

Our visualization not only converts data into information, but there are latent forces at play that it tries to affect. This visualization helps bring about a change in the user’s opinions that can prove pivotal in improving the economic situation of the region. The elements used in the web page are targeted for users with little interest in and knowledge about scientific factors but are visiting the page for the sole purpose of exploring the beauty of deserts.

We also discuss various rationales for our design implementation and the vast literature on various aspects of the project. Finally, we discussed the technical know-how of the work like the elements and libraries used, preparation and processing of data, and various intelligent facets of the work including data mining, text parsing, text summarization, unsupervised learning, etc.

A promising direction for future work is to expand the number of plants that we present in the visualization. Once we have a sufficient number of images, we could show multiple image accordions, each showing flowers of a single color ranging from light to dark shades. The evaluation plan can help us improve the design choices further and model our audience's preferences. We would also like to involve better user personalization through machine learning techniques.

REFERENCES

[1] 2017. Finding Dominant Colour on an Image. https://code.likeagirl.io/finding-dominant-colour-on-an-image-b4e075f98097

[2] 2020. AMWUA. https://www.amwua.org/

[3] 2020. AutoSummarizer. https://autosummarizer.com

[4] 2020. BHG. https://www.bhg.com/gardening

[5] 2020. houseplants. https://www.houseplantsexpert.com/a-z-list-of-house-plants.html

[6] 2020. SEINet Portal. http://swbiodiversity.org/seinet/

[7] 2020. TexasSmartScape. http://www.txsmartscape.com/plant-search

[8] 2020. TextSummarization. http://textsummarization.net/text-summarizer

[9] PAUL ADAM. 1987. Australian deserts defended. *Nature* 325, 6105 (1987), 570–570.

[10] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315.

[11] Alberto Cairo. 2016. *The truthful art: data, charts, and maps for communication.* New Riders.

[12] Terry C Daniel. 1976. *Measuring landscape esthetics: the scenic beauty estimation method.* Vol. 167. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range ....

[13] Josh Eckert. [n.d.]. Art in the age of ones and zeros: Turning big data into art.

[14] Josh Eckert. [n.d.]. Art Made of Storms.

[15] Josh Eckert. [n.d.]. Psychology for Artists, Futurists, Worldbuilders.

[16] Kenneth Field. 2020. Mapping coronavirus, responsibly. https://www.esri.com/arcgis-blog/products/product/mapping/mapping-coronavirus-responsibly/.

[17] Edward Gilbert, Corinna Gries, Nico Franz, Leslie R Landrum, and Thomas H Nash III. 2019. SEINet: A Centralized Specimen Resource Managed by a Distributed Network of Researchers. *Biodiversity Information Science and Standards* (2019).

[18] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. 2010. A Tour through the Visualization Zoo. *Commun. ACM* 53, 6 (2010), 59–67. http://idl.cs.washington.edu/papers/visualization-zoo

[19] Jeff Johnson. 2013. *Designing with the mind in mind: simple guide to understanding user interface design guidelines.* Elsevier.

[20] Leaflet. 2019. Leaflet/Leaflet.markercluster. https://github.com/Leaflet/Leaflet.markercluster

[21] Leaflet. 2020. Leaflet. https://github.com/Leaflet/Leaflet

[22] Mapbox. 2020. mapbox/mapbox-gl-js. https://github.com/mapbox/mapbox-gl-js

[23] Katharina Reinecke, Tom Yeh, Luke Miratrix, Rahmatri Mardiko, Yuechen Zhao, Jenny Liu, and Krzysztof Z Gajos. 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 2049–2058.

[24] Leonard Richardson. 2007. Beautiful soup documentation. *April* (2007).

[25] Robert E Roth, Andrew W Woodruff, and Zachary F Johnson. 2010. Value-by-alpha maps: An alternative technique to the cartogram. *The Cartographic Journal* 47, 2 (2010), 130–140.

[26] Sagar Shivaji Salunke. 2014. *Selenium Webdriver in Python: Learn with Examples* (1st ed.). CreateSpace Independent Publishing Platform, North Charleston, SC, USA.

[27] Beatriz Sousa Santos and Jean-Louis Dillenseger. 2005. Quality evaluation in medical visualization: some issues and a taxonomy of methods. In *SPIE Medical Imaging.*

[28] Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368* (2017).

[29] Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages.* IEEE, 336–343.

[30] Guy Sircello. 2015. *New theory of beauty.* Princeton University Press.

[31] Beatriz Sousa Santos and Paulo Dias. 2013. *Evaluation in visualization: some issues and best practices.* Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9017. 90170O. https://doi.org/10.1117/12.2038259

[32] Georg Stenberg. 2006. Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology* 18, 6 (2006), 813–847.

[33] Bootstrap team. 2020. Bootstrap. https://getbootstrap.com/

[34] Edward R Tufte. 2001. *The visual display of quantitative information.* Vol. 2. Graphics press Cheshire, CT.

[35] Colin Ware. 2012. Information Visualization: Perception for Design, Waltham, MA.