Identifying Priority Survey Sites for Early-Season Milkweed Conservation

Technical Documentation

A Capstone Project submitted in partial satisfaction of the requirements for the degree
of Master of Environmental Data Science
for the
Bren School of Environmental Science & Management

By

Amanda Herbst
Sam Muir
Anna Ramji
Melissa Widas

Committee in charge:
Dr. Ruth Oliver
Dr. Carmen Galaz García

MAY 2024

Identifying Priority Survey Sites for Early-Season Milkweed Conservation

As developers of this Capstone Project documentation, we archive this documentation on the Bren School's website such that the results of our research are available for all to read. Our signatures on the document signify our joint responsibility to fulfill the archiving standards set by the Bren School of Environmental Science & Management.

____ Amanda Herbst

____ Sam Muir

____ Anna Ramji

____ Melissa Widas

The Bren School of Environmental Science & Management produces professionals with unrivaled training in environmental science and management who will devote their unique skills to the diagnosis, assessment, mitigation, prevention, and remedy of the environmental problems of today and the future. A guiding principle of the School is that the analysis of environmental problems requires quantitative training in more than one discipline and an awareness of the physical, biological, social, political, and economic consequences that arise from scientific or technological decisions. The Capstone Project is required of all students in the Master of Environmental Data Science (MEDS) Program. The project is a six-month-long activity in which small groups of students contribute to data science practices, products or analyses that address a challenge or need related to a specific environmental issue. This MEDS Capstone Project Technical Documentation is authored by MEDS students and has been reviewed and approved by:

____

MEDS 411B Instructor: Dr. Carmen Galaz García

____

Faculty Advisor: Dr. Ruth Oliver

____

Date

Abstract

Monarch butterfly (*Danaus plexippus*) populations are experiencing decline due to habitat degradation and climate threats. In 2024, this iconic species became considered for protection under the Endangered Species Act. Critical to the persistence of this species of cultural and ecological importance is milkweed (*Asclepias spp.*), which monarchs rely on as an essential resource for food and reproduction. Motivated by concerns over the loss of monarch habitat, the Santa Barbara Botanic Garden (SBBG) and local collaborators are working to identify potential restoration areas in the Los Padres National Forest (LPNF). However, the vast size and complex terrain of the LPNF pose a logistical challenge to surveying. Therefore, there is a need for a tool to aid in selecting sites to prioritize. Our team created a priority index for surveying milkweed within the LPNF by predicting the habitat suitability for four early-season milkweed species using Maximum Entropy (MaxEnt) species distribution modeling and creating a novel survey site accessibility index. We identified high-priority sites for the team to survey based on high predicted suitability for each species of early-season milkweed and high physical accessibility. To communicate these results, we created an interactive web dashboard, which the SBBG team will use for field planning to support the ongoing monitoring of milkweed populations for monarch habitat restoration efforts within the LPNF.

Executive Summary

Western monarch butterfly (*Danaus plexippus*) populations have been heavily declining over the past 50 years, largely due to the loss of their overwintering and breeding habitat, the majority of such being milkweed (*Asclepias spp*.) (Pelton et al. 2019). The monarch butterfly is currently under consideration for protection under the Endangered Species Act. Monarch butterflies are essential pollinators and are the only butterfly species to complete a two-way migration like many bird species do. The Los Padres National Forest (LPNF) is a crucial stop in the western monarch's migration as it is home to four species of early-season milkweed native to the area: *A. californica*, *A. vestita*, *A. eriocarpa*, and *A. erosa*, making the forest an important reproductive site. Monarch butterflies lay their eggs on milkweed leaves, which serve as a larval food source and are crucial in the monarch life cycle. Unfortunately, milkweed populations are also facing pressing threats such as habitat loss, drought, and commercial pesticide use (Bahmani et al., 2018; Boyle et. al., 2019; Los Padres Forest Watch, 2013; Pleasants & Oberhauser, 2013).

The Santa Barbara Botanic Garden (SBBG), located in the Mission Canyon area of Santa Barbara, was established almost one hundred years ago and operates under their mission "to conserve native plants and habitats for the health and well-being of people and our planet" (Santa Barbara Botanic Garden, n.d.). Motivated by concerns over the loss of essential monarch habitat in the LPNF, the SBBG and local collaborators are working to restore milkweed habitat and identify potential restoration areas in the forest. The SBBG has completed the first of a two-year survey, in which the research team aims to revisit up to 1,000 historic locations of populations of native early-season milkweed. These initial surveys are being conducted with the intent to build an understanding of the current distribution of milkweed throughout the LPNF and to collect seed and tissue samples to prepare for future restoration efforts. The second year of surveys is scheduled for this summer, 2024, but planning the locations of surveying proves challenging, as the team cannot feasibly visit every part of the forest, since it spans a total of 1.75 million acres and stretches almost 220 miles from north to south (USDA Forest Service National Website, n.d.). Therefore, the SBBG team needed a tool to help them strategically plan what locations they should visit in the LPNF.

This capstone project has been built to facilitate informed survey efforts by identifying high-priority survey locations throughout the LPNF and presenting all components in a user-friendly interactive web dashboard. The SBBG provided the capstone team with the locations of milkweed found during their 2023 field surveys. The first step in the project used this provided data, along with environmental data collected from publicly accessible sources, to predict and map milkweed habitat suitability in the LPNF via Maximum Entropy (MaxEnt) species distribution modeling (SDM) for each species of early-season milkweed. The second step consisted of creating a novel survey site accessibility index that identified how physically accessible locations in the LPNF were to the SBBG field survey team. This was based on five factors that were determined from multiple discussions with the client: distance from roads, distance from trails, vegetation density, slope, and land ownership. Lastly, a survey site priority index was calculated by combining the habitat suitability maps and the survey site accessibility index. A location that is "high-priority" is predicted to be both highly suitable for the specific milkweed species and highly accessible for the SBBG field survey team.

To provide the SBBG team with a deliverable they can easily use to visualize the results of the capstone project, an interactive web dashboard was developed that houses the habitat suitability maps, survey site accessibility index, and survey site priority indices. This dashboard allows the SBBG team to compare the results for each milkweed species and download a table of high-priority locations for their upcoming surveys. Once the 2024 field surveys are completed, functionality has been built into the code for the SBBG team to add the new milkweed occurrence data and update both the habitat suitability models and the survey site priority index. The results of this project are an important step in the SBBG's long-term

conservation efforts that will benefit future early-season milkweed and, consequently, monarch butterfly populations.

Table of Contents

1. Problem Statement

Western Monarch butterfly (*Danaus plexippus*) populations have been declining for decades due to climate and habitat threats. Over the last few years, the monarch butterfly has come under consideration for protection under the federal Endangered Species Act. Under this designation, their essential reproductive habitat, early-season milkweed (*Asclepias spp.*) would also be protected. The four species of early-season milkweed (*A. californica*, *A. vestita*, *A. eriocarpa*, and *A. erosa*) found throughout the Los Padres National Forest (LPNF) are also threatened by climate change and habitat loss. Motivated by the population decline of both the monarch butterfly and its early-season milkweed habitat, the Santa Barbara Botanic Garden (SBBG) and local collaborators are working to protect the milkweed habitat and identify potential restoration areas throughout the LPNF. An important step in this conservation process is identifying and understanding where milkweed is currently located in the forest and collecting milkweed seeds in anticipation of future restoration efforts. The SBBG has completed the first of a two-year field survey, in which the research team aims to visit up to 1,000 historic locations of the four species of early-season milkweed.

However, this is a large-scale, resource-intensive task, as the LPNF covers a total of 1.75 million acres and stretches almost 220 miles from north to south with complex terrain (USDA Forest Service National Website, n.d.). The SBBG cannot survey the entire forest and wants to prioritize visiting locations where milkweed is likely present, so the survey team needs a tool to help direct their surveys within the LPNF. This project aimed to identify locations within the LPNF where milkweed is likely to be located and it is highly physically accessible for the SBBG team to visit.

2. Specific Objectives

To achieve this goal, we have identified the following objectives:

1. Identify high-priority milkweed survey sites, based on habitat suitability and physical accessibility, throughout the Los Padres National Forest to direct Santa Barbara Botanic Garden summer 2024 field surveys.
2. Deliver all components of Objective 1 in the form of an interactive web dashboard to facilitate field survey planning and allow for future updates as additional data is collected.

3. Summary of Solution Design

This project was developed in the programming language R, in the integrated development environment, RStudio. For all packages and versions used, see Appendix I.

To identify high-priority sites for the SBBG to visit in their upcoming surveys of early-season milkweed in the LPNF, we determined locations that were both highly suitable for each species of milkweed and highly physically accessible to enable the garden team to visit many locations efficiently. To find locations in the forest that are highly suitable for milkweed, we created a **Habitat Suitability Map** using a maximum entropy (MaxEnt) species distribution modeling (SDM) approach (see Section 3.2). To identify how physically accessible each location in the LPNF is, we developed a novel **Survey Site Accessibility Index**. This was created based on factors that the client deemed important in measuring physical accessibility (see Section 3.3). Finally, to identify high-priority survey site locations, the outputs of the habitat suitability maps and the survey site accessibility index were combined to create a **Survey Site Priority Index** (see Section 3.4). All of this information was incorporated into an **Interactive Web**

Dashboard as a tool to aid the SBBG staff in survey planning. See Sections 3.5 and 6.3 for a more thorough walkthrough of the dashboard components and methodology.

3.1 Data

One of the key datasets used in this project was the milkweed survey data, collected in the summer of 2023 by the SBBG team in the southern region of the LPNF. The "milkweed_sp" (indicating milkweed species) and "geometry" columns were used to extract milkweed occurrence points for the species distribution modeling (see Section 3.2). Other key datasets were publicly available environmental data: bioclimatic variables, canopy cover data as a proxy for bare ground, and solar radiation data that we calculated using the slope and aspect data from a digital elevation model (DEM). These environmental variables were used in the species distribution models to create the Habitat Suitability Maps, and the slope and canopy cover as a proxy for vegetation density were additionally used in the development of the Survey Site Accessibility Index. Shapefiles of trails, roads, and land ownership status in the LPNF were also used to create the Survey Site Accessibility Index. A template raster was created to standardize raster dimensions and was used as a blueprint to reproject and resample all of the layers that built the Survey Site Accessibility Index in order to perform zonal calculations.

Two sources were used for the trail and road data: Los Padres Forest Watch and USGS. The Los Padres Forest Watch data was used to calculate the distance from trails and distance from roads in the southern region of the LPNF to maintain consistency since it is the same dataset that the SBBG team has been using to plan their surveys. The Forest Watch data did not include the northern region of the LPNF, so USGS data was used for that region. The raw Los Padres Forest Watch data and all data used to create intermediate and final outputs and products for this project were archived on Dryad (see Section 7).

Table 1. Data Summary

Data	**Source**	**Use**
SBBG Milkweed Survey Locations (Polygon Data, 2023 Survey): milkweed species, presence/absence, location, number of plants (.shp)	Dr. Sarah Cusser (SBBG) via Google Drive	Habitat Suitability Map, Interactive Web Dashboard
Bioclim (19 variables)**:** annual trends, seasonality, extreme/limiting environmental factors based on monthly temperature and rainfall data (.tif)	**{wallace}**'s **envs_worldclim()** function, which utilizes **{raster}**'s **getData()** function (see Appendix I)	Habitat Suitability Map, Survey Site Priority Index, Interactive Web Dashboard
California Multi-Source Land Ownership: classification of public and private land in the LPNF and surrounding areas (.shp)	California Department of Forestry and Fire Protection; California State Geoportal, hosted on CAL FIRE Portal (via gis.data.ca.gov)	Survey Site Accessibility Index, Survey Site Priority Index
Canopy Cover: horizontal cover fraction occupied by tree canopies (.tif)	California Forest Observatory, selecting "Canopy Cover" from available datasets to	Habitat Suitability Map, Survey Site Accessibility Index, Survey Site Priority Index, Interactive Web

	download	Dashboard
Digital Elevation Model (DEM): topographic surface of the earth and flattened water surfaces (.tif)	USGS's The National Map (TNM) download interface; Elevation Products (3DEP) 1 arc-second DEM	Habitat Suitability Map, Survey Site Accessibility Index, Survey Site Priority Index, Interactive Web Dashboard
Los Padres National Forest (LPNF) Boundary: boundaries of the northern and southern regions of the LPNF (.gdb)	California State Parks GIS Data (via parks.ca.gov) ESRI geodatabase	Habitat Suitability Map, Survey Site Accessibility Index, Interactive Web Dashboard
LPNF 2023 Trails and Roads: trails and roads in the southern region of the LPNF (.shp)	Los Padres Forest Watch via ArcGIS	Survey Site Accessibility Index, Survey Site Priority Index, Interactive Web Dashboard
USGS 2024 Trails and Roads: trails and different types of roads in California (.shp), used data from within the northern region of the LPNF	National Transportation Data (NTD) California Shapefile, courtesy of the U.S. Geological Survey	Survey Site Accessibility Index, Survey Site Priority Index, Interactive Web Dashboard

3.2 Habitat Suitability Map

To identify suitable habitat for each of the four species of early-season milkweed in the LPNF, we used a Maximum Entropy (MaxEnt) species distribution modeling (SDM) approach (Figure 1). MaxEnt SDM works to maximize entropy, where the best estimate of a probability distribution is the one with the largest entropy while fitting the constraints imposed by the environmental variables at the occurrence points (Figure 1). MaxEnt modeling uses presence data, not taking into account absences, so MaxEnt uses background points, which are randomly selected locations within the study area to represent the available environmental conditions.

Since point presence data is needed to create an SDM, milkweed species occurrence polygon data provided by the SBBG was converted to points. After a conversation with the client, it was decided that the polygon's size is proportional to the number of plants within the polygon's boundary, so the polygons were converted to points using **sf::st_cast()**. This method converts each outer convex point on the polygon outline as a point. This was performed to maximize the data points for the SDM and was appropriate because the border of the polygon represents the farthest extent at which a particular species was identified in the survey, meaning that the area is the full range of where that species was observed at that survey location.

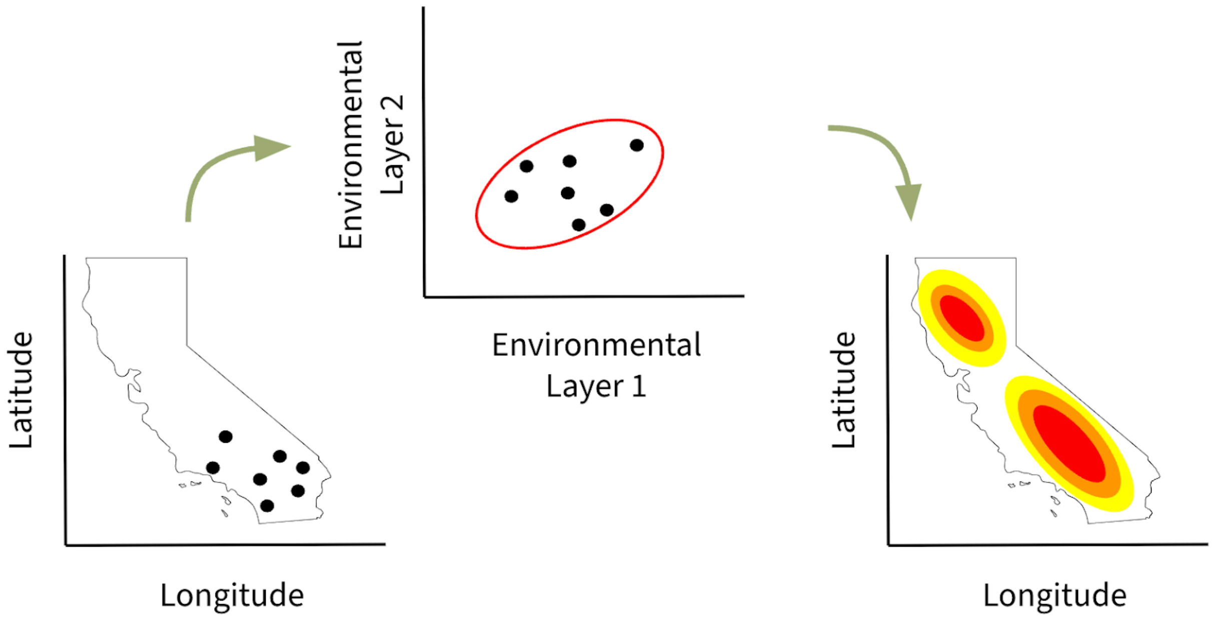

Figure 1. MaxEnt species distribution modeling: This approach estimates habitat suitability based on occurrence points based on their association with key environmental factors and projects this into environmental space.

The geographic result of a MaxEnt model is a continuous probability map (heat map) where each grid cell has been given the relative suitability for the species on a scale of 0 to 1. The higher values can be interpreted as a more suitable habitat, where the model predicts a higher probability of finding a specific species in that location based on the similarity in environmental conditions.

Milkweed occurrence points in the southern region of the forest and the environmental data layers were input into the model. The model was trained on occurrence data in the southern region and was then projected to the northern region to create a map of predicted milkweed habitat suitability across both regions of the LPNF. Details and methodology on model selection can be found in Section 5.2.1. The resulting heatmap shows the habitat's predicted suitability in each 1km x 1km grid cell (Figure 2).

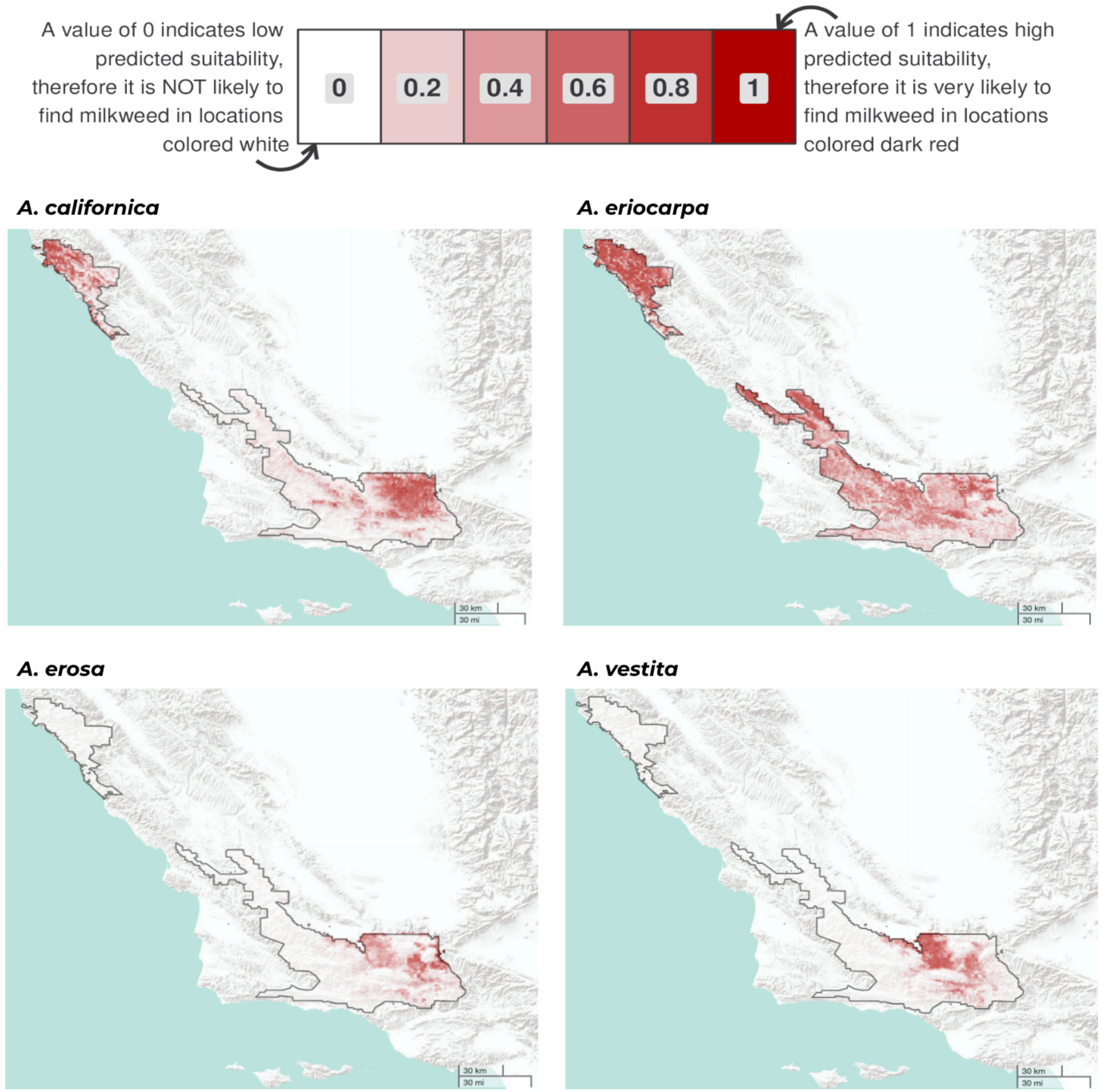

Figure 2. Habitat suitability, mapped within the boundary of the LPNF, for the four milkweed species. Dark red indicates the highest suitability (1) and white indicates the lowest suitability (0).

3.3 Survey Site Accessibility Index

A novel index for survey site accessibility was calculated based on distance from roads, distance from trails, slope, canopy cover (as a proxy for vegetation density), and land ownership status. These factors were determined to be most important for physical accessibility based on conversations with the client. See Section 3.1 for data sources and processing methodology. The distance from roads and distance from trails were calculated by extracting the centroid of each raster cell as a point using **raster::rasterToPoints()** and then calculating the distance from each point to each road/trail using **sf::st_distance()**. Then, the minimum value, which represents the distance to the nearest road or trail, was selected. Finally, the data was converted back to a raster using **stars::st_rasterize** and **sf::rast()**. Once processed, the slope, vegetation density, distance from roads, and distance from trails rasters were all cropped to the LPNF boundary using **terra::mask()**. Lastly, these layers were rescaled to a scale of 0 (least accessible) to 1 (most accessible) by a created function, **rescale_raster()**, to streamline the rescaling process. This function utilizes **terra::minmax()** to pull out the minimum and maximum values of the given raster, then rescales each raster cell with this calculation:

$$rescaledCell = \frac{rasterCell - minimumValue}{maximumValue - minimumValue}$$

so that the minimum value of the raster becomes 0 and the maximum value of the raster becomes 1. After this function is applied, the minimum values correlate to the highest accessibility, so a final calculation was performed to ensure that a value of 1 indicated the most accessible location:

$$1 - rescaledCell$$

The land ownership layer was converted to a mask containing only two values where a value of 1 indicates public land and a value of 0 indicates private land since the SBBG team cannot survey on privately owned land. This ensures that when all layers are combined by multiplication, any private lands remain as a 0, or inaccessible. A final map of survey site accessibility throughout the LPNF on a scale of 0 to 1 was then created as the product of the rescaled layers of slope, vegetation density, distance from roads, distance from trails, and the mask of land ownership (Figure 3).

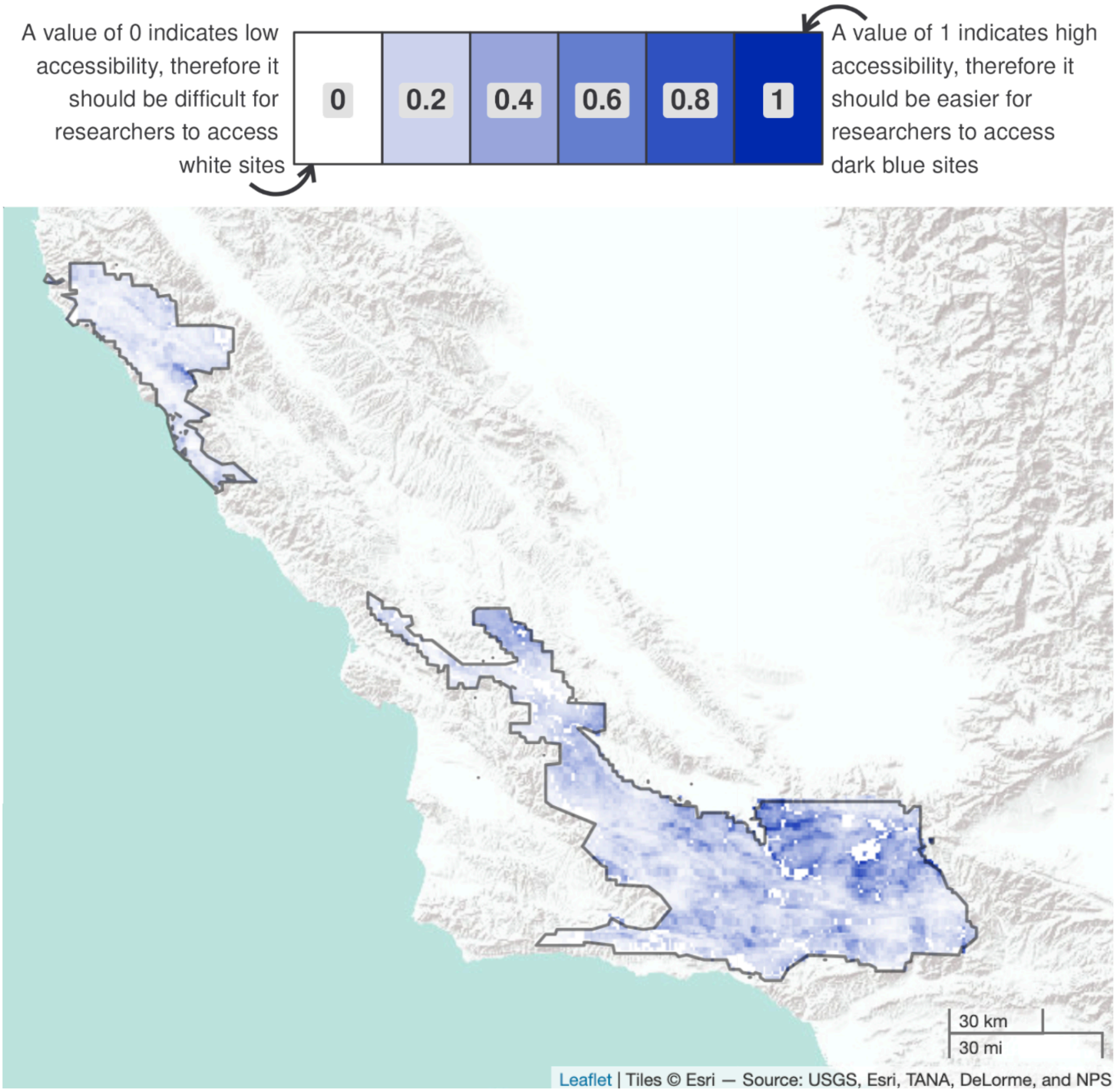

Figure 3. Survey Site Accessibility Index within the LPNF. The darkest shade of blue indicates the highest level of survey site accessibility (1), while white represents the lowest survey site accessibility (0).

3.4 Survey Site Priority Index

High-priority survey sites were identified as locations with high predicted milkweed habitat suitability and high physical survey site accessibility. Habitat suitability and survey site accessibility, both on a scale from 0 (least suitable and least accessible) to 1 (most suitable and most accessible), were multiplied together to calculate a priority score. The survey site priority index for each species was then rescaled using **rescale_raster()** such that the highest relative priority became a value of 1 and the lowest relative priority remained a value of 0 (Figure 4). Thus, the highest priority survey sites are the areas that have the values closest to 1. The centroid coordinates and priority score of each raster cell were compiled in a data table with each row containing the raster cell coordinates, relative priority score for each milkweed species, accessibility score, and status indicating if the SBBG team has previously visited the site.

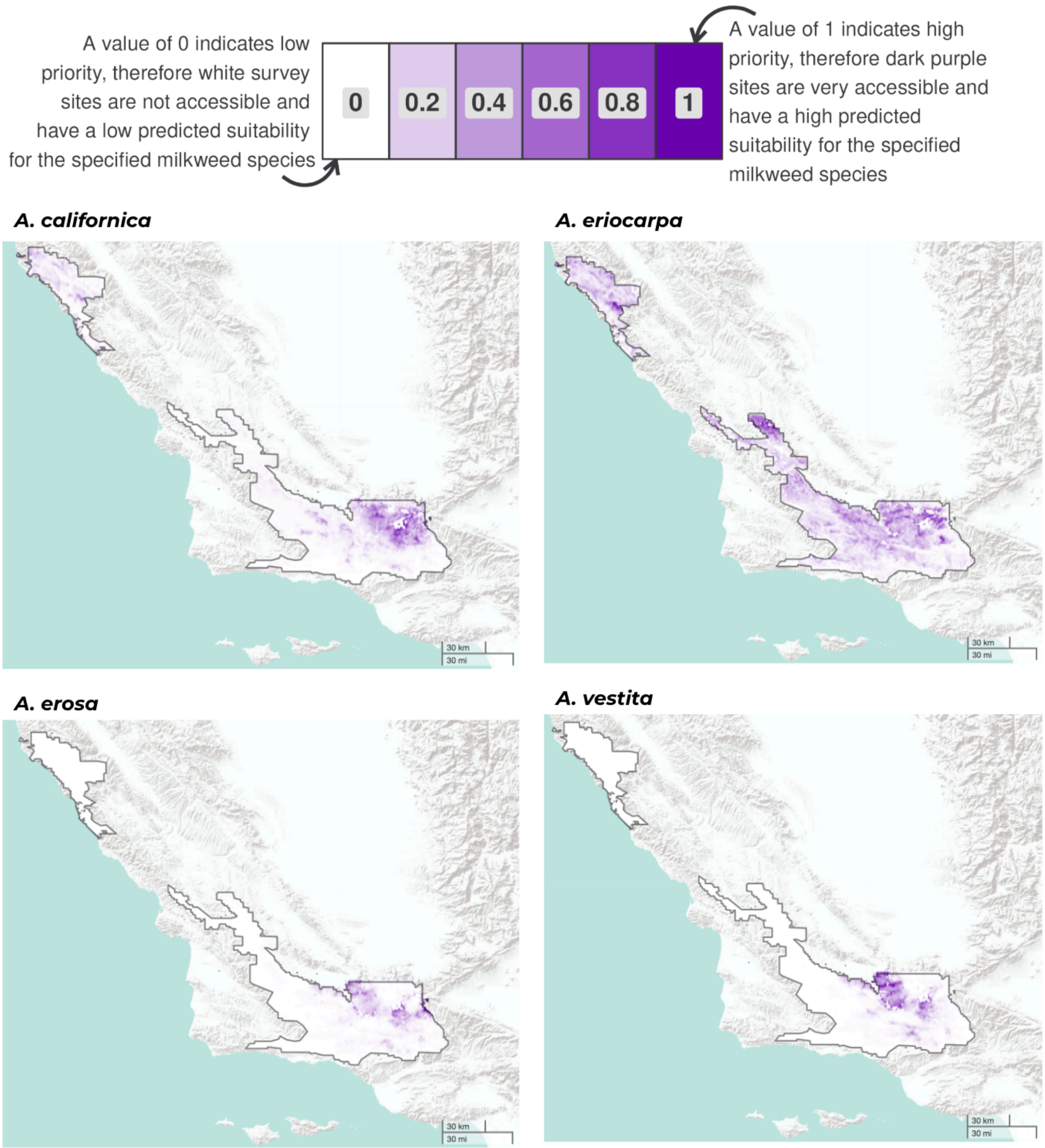

Figure 4. Survey site priority, mapped within the LPNF boundary for all four milkweed species. The darkest shade of purple indicates the highest level of survey site priority (1), while white represents the lowest survey site priority (0). This color palette was chosen to serve as a visual indication of the combination of habitat suitability (red palette) and survey site accessibility (blue palette).

3.5 Interactive Web Dashboard

An interactive web dashboard was developed in R via the Shiny package to communicate the project results and provide an interface to aid the SBBG team in planning field surveys. The interactive web dashboard consists of background information about the project, milkweed habitat suitability maps, a survey site accessibility map of the LPNF, and a map and table of high-priority survey sites. The interactive web dashboard's source code is publicly available on GitHub. For more detailed information on the interactive web dashboard please see Section 6.3 Navigating the interactive web dashboard.

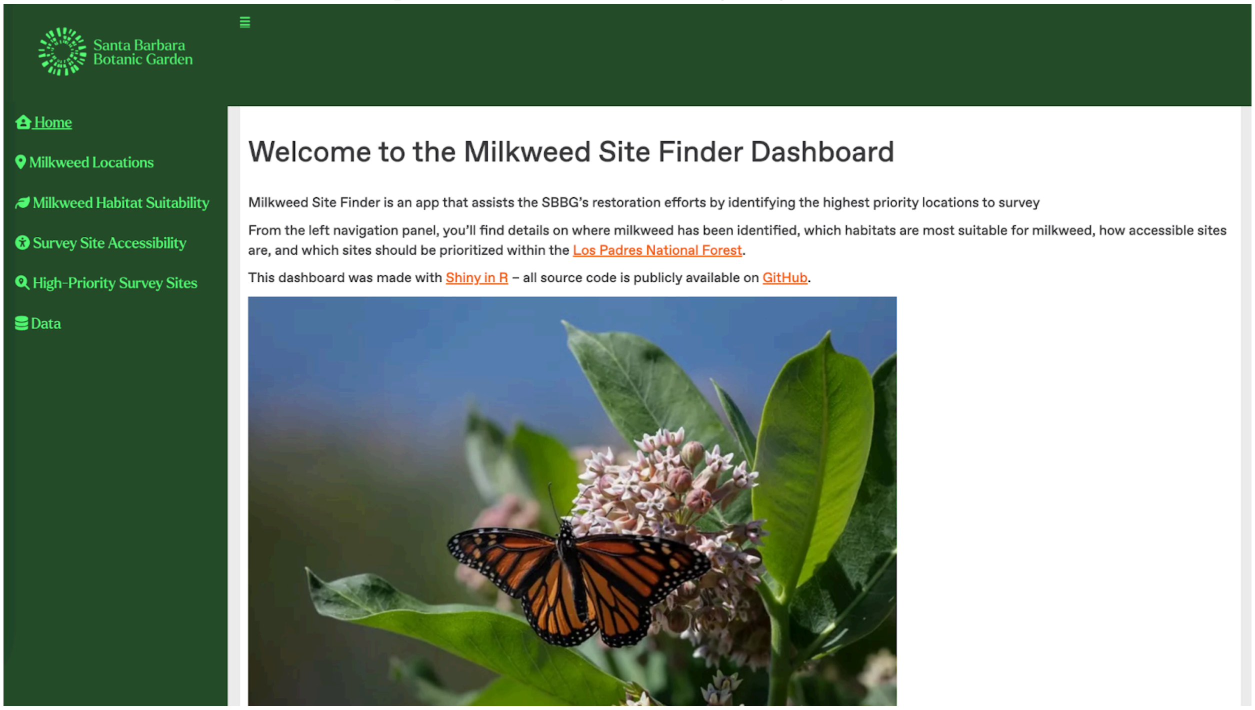

Figure 5. Homepage of the interactive web dashboard created to showcase the results of this project.

4. Products and Deliverables

4.1 Interactive Web Dashboard

The interactive web dashboard "Milkweed Site Finder" was built using the Shiny package in R and is intended to be used for planning field surveys for the SBBG team. The interactive web dashboard is comprised of the following information:

- **Home:** Brief overview of the project and a data collection disclaimer (Figure 5)
- **Milkweed Locations:** Data visualization of the locations where the SBBG team has surveyed early-season milkweed (Figure 6)

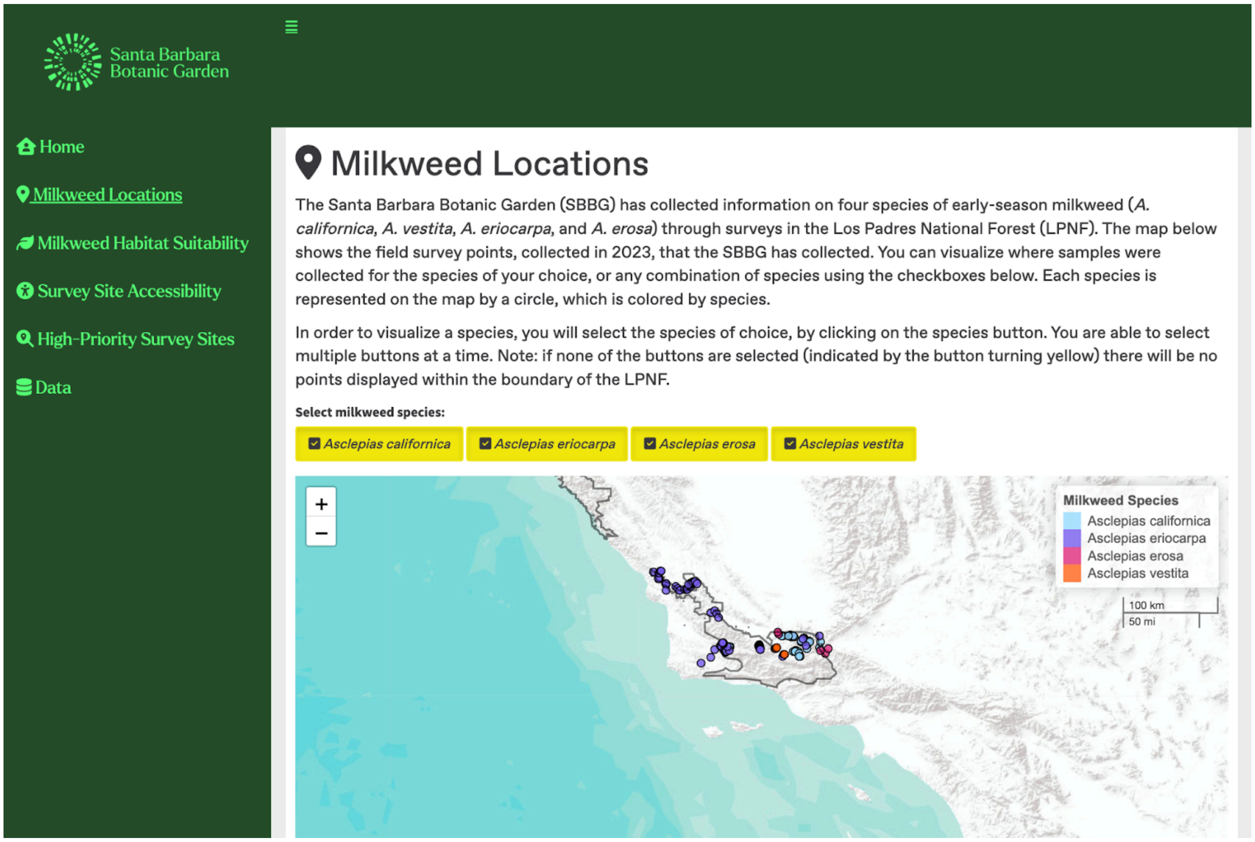

Figure 6. The Milkweed Locations tab of the interactive web dashboard showing the sites where milkweed was previously found by the SBBG team.

- **Milkweed Habitat Suitability:** Data visualization of the habitat suitability for the four species of early-season milkweed (Figure 7)

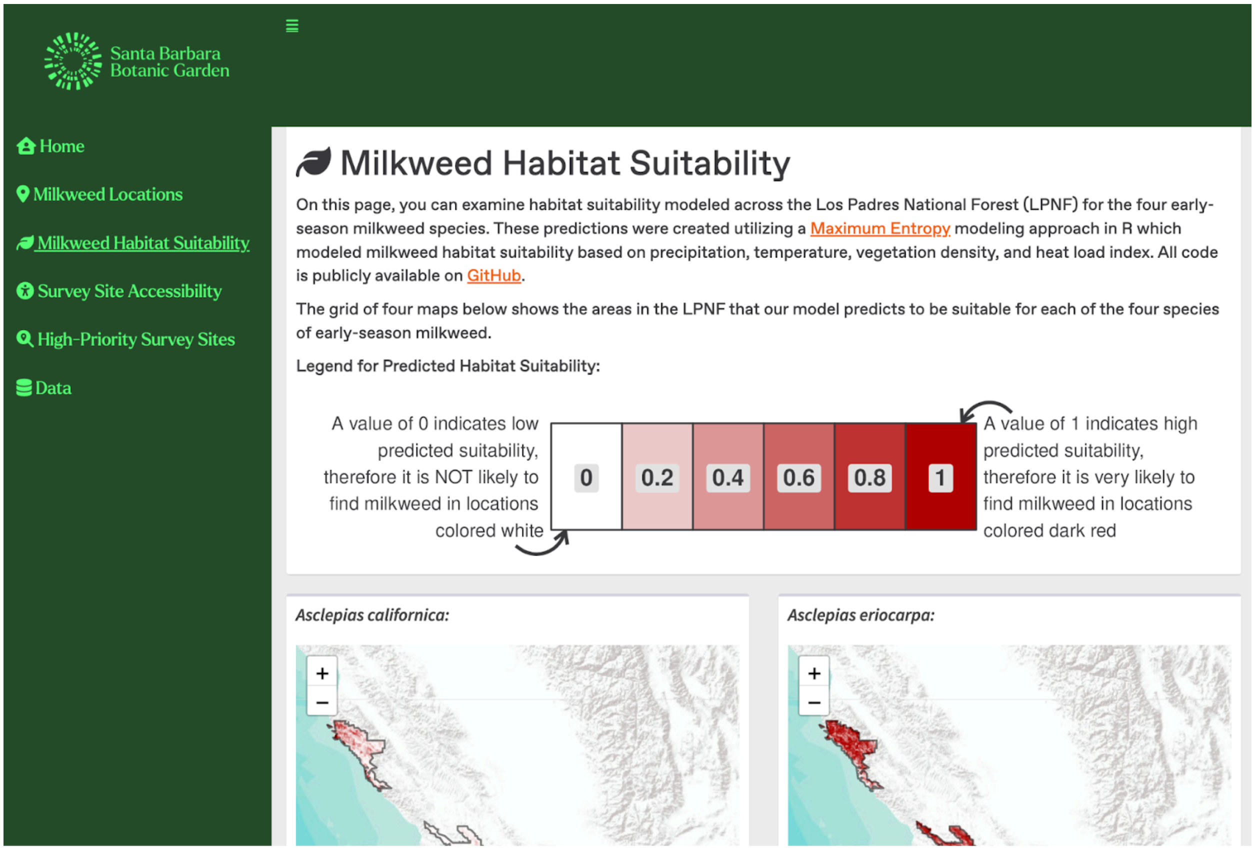

Figure 7. The Milkweed Habitat Suitability tab of the interactive web dashboard showing the SDM results.

- **Survey Site Accessibility:** Data visualization of the survey site accessibility map for the LPNF (Figure 8)

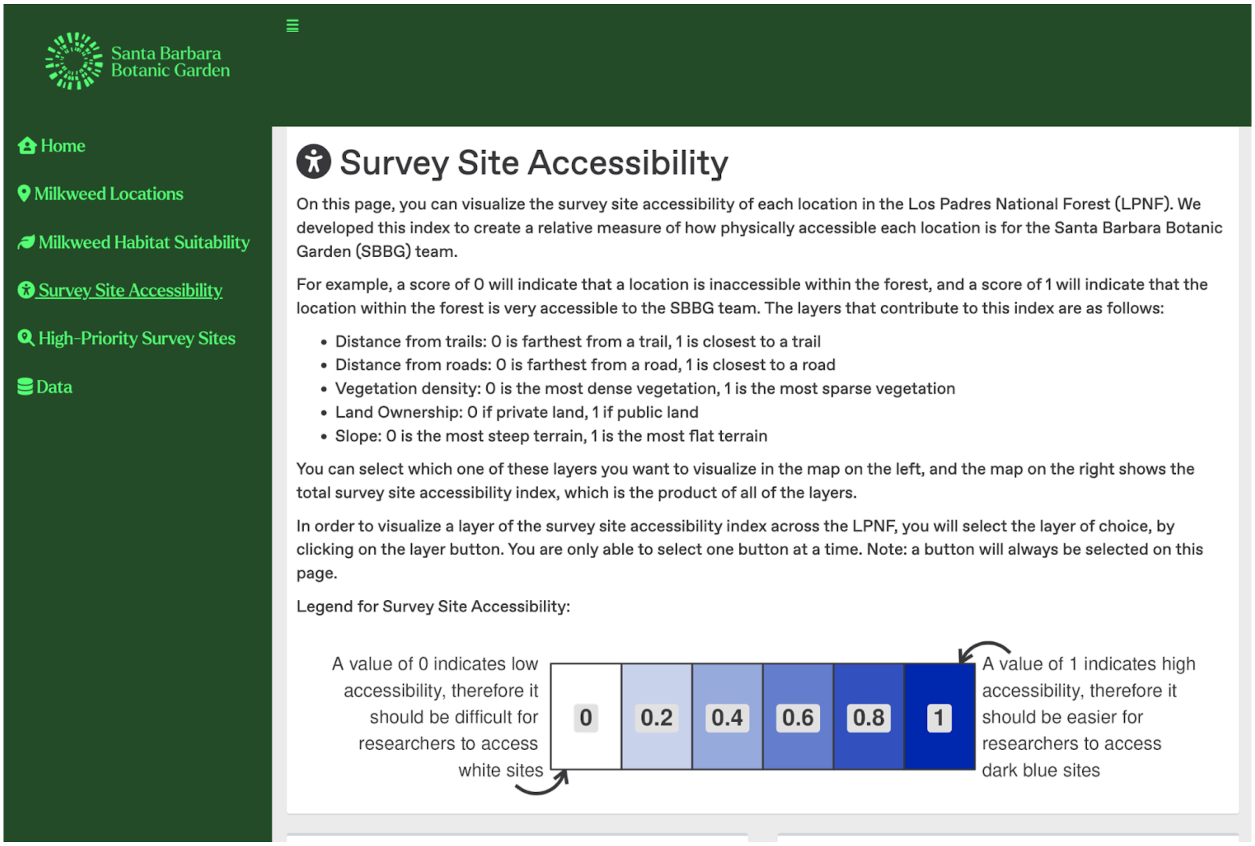

Figure 8. The Survey Site Accessibility tab of the interactive web dashboard showing the survey site accessibility maps.

- **High-Priority Survey Sites:** Data visualization of the high-priority survey sites with the capability to download a list of sites ranked by the priority index score (Figure 9)

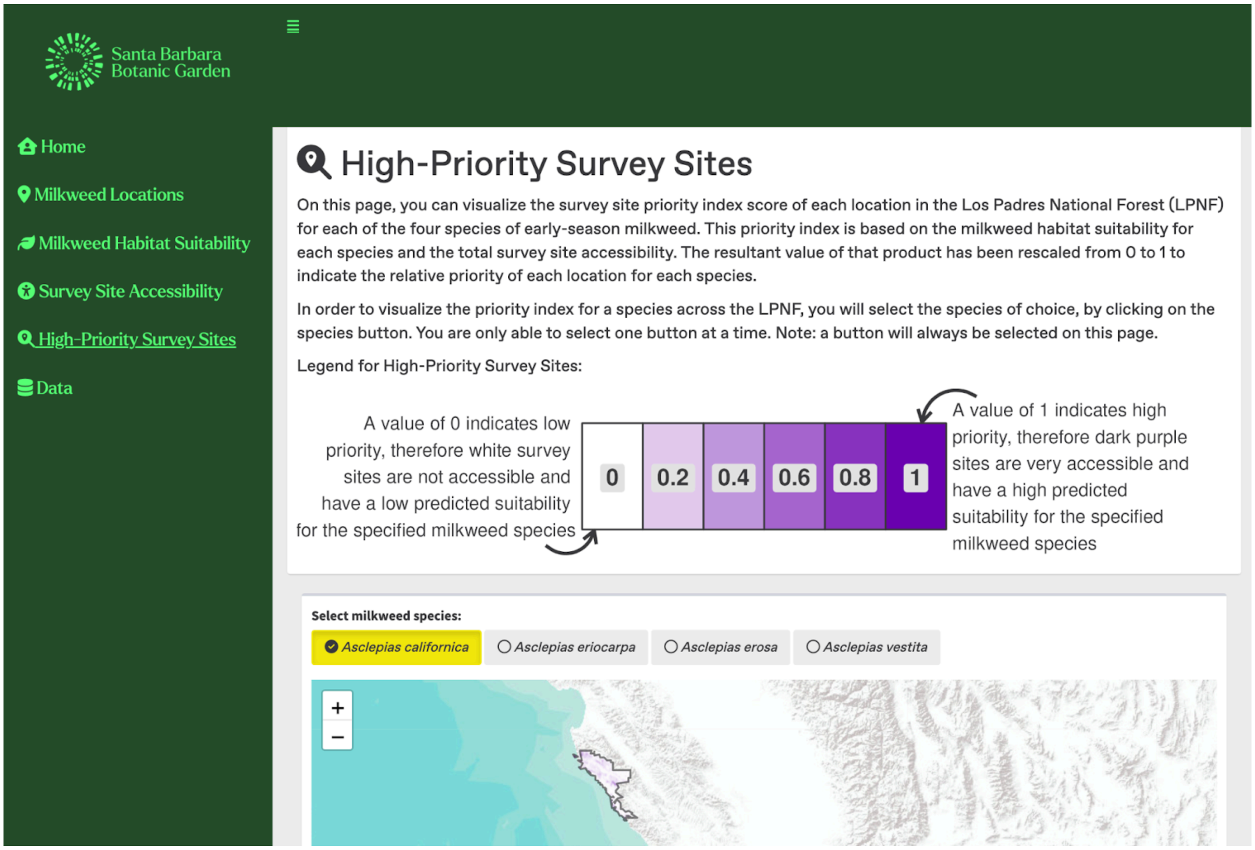

Figure 9. The High-Priority Survey Sites tab of the interactive web dashboard showing the resulting high-priority survey site maps.

- **Data:** Description of data sources and data access used to develop the application and its components (Figure 10)

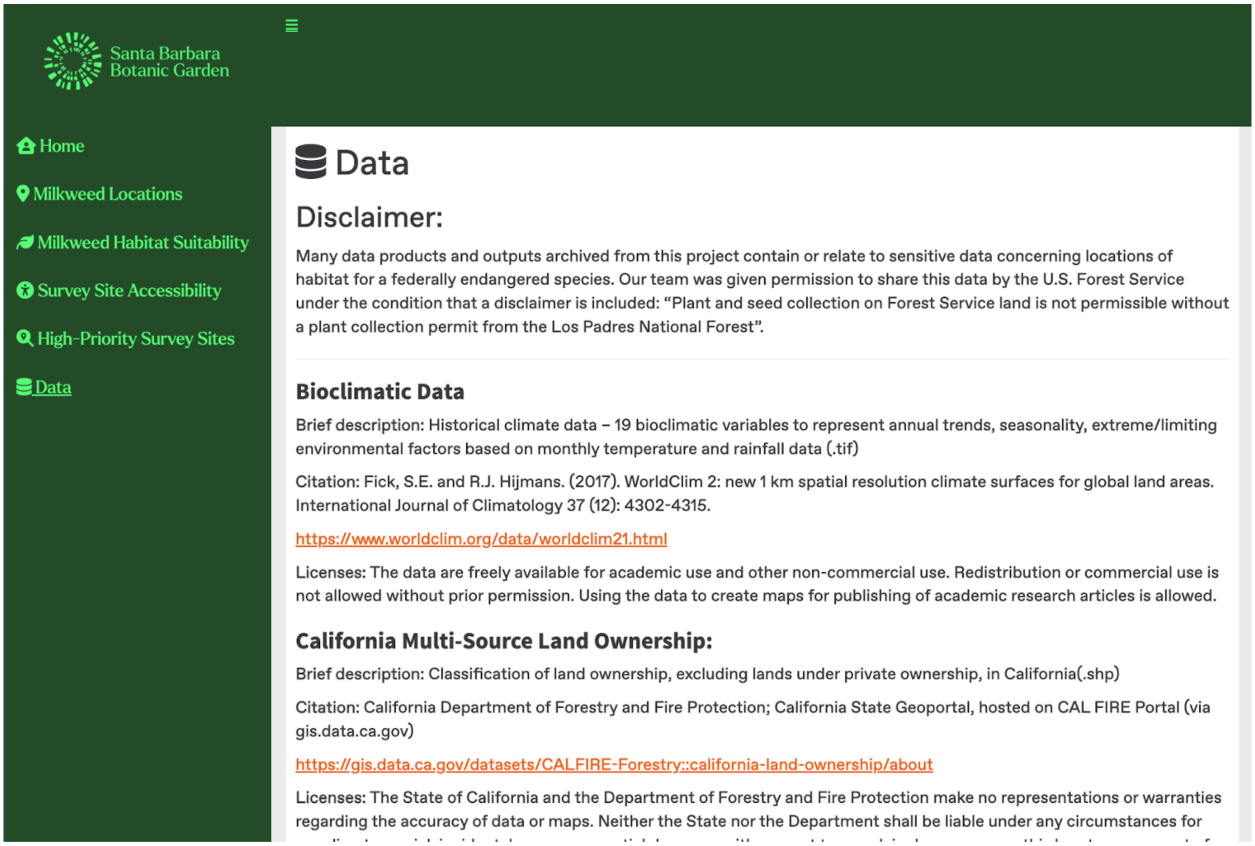

Figure 10. The Data tab of the interactive web dashboard detailing information on the data sources.

4.2 GitHub Organization

All components of this project are stored in the public GitHub organization. This GitHub organization contains a README.md file for each of its repositories and one for the overall organization. The repositories housed by this organization are an interactive web dashboard repository and a modeling repository.

4.3 Data Compilation

All data used in this project were stored in a Box folder owned by the SBBG. The Box folder contains all raw data, intermediary data, and final outputs used and produced throughout this project, as it is imperative that the client can build this project. The intermediate data and final outputs are also archived in Dryad (Section 7).

When archiving and sharing our data products, we aimed to follow the FAIR principles of being Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). Additionally, the architecture of our code allows for easy integration of new SBBG survey data. For more information on future data integration, see Section 6.2.2.

5. Summary of Testing

5.1 Unit Testing for Data Preparation

To create the necessary raster stacks and conduct zonal calculations with all raster data, the designated criteria were defined:

- Coordinate Reference System: EPSG 4326
- Resolution: 0.0083 x 0.0083 degrees per pixel
- Extent: xmin = -122.0, xmax = -117.5014, ymin = 33.6, ymax = 36.9034
- Origin: (0.0017, 0.0016)

A template raster was created with the above metadata and values of zero, which was used to both reproject environmental layer rasters and verify all rasters meet the criteria. Unit testing has been built in to manually verify this information by printing a question and then the output of a logical equation. For example:

```
# canopy cover crs check
print("Does canopy_rescaled have the correct CRS?")
st_crs(canopy_rescaled) == st_crs(temp_rast)

>  Does canopy_rescaled have the correct CRS?
> TRUE
```

5.2 Habitat Suitability Model Selection

5.2.1 Model Tuning and Selection with {ENMeval}

To prepare our data for evaluation, we used block partitioning ($k = 4$), where species occurrences were spatially partitioned into equally sized groups. This method is suitable for many sample sizes and is commonly used in species distribution modeling (Fourcade et al., 2018, Radosavljevic & Anderson, 2014). This is a form of *k*-fold cross-validation, in which the model was built sequentially on all but one fold, and model performance was evaluated based on the remaining fold. After this modeling was complete, the model performance statistics were averaged across iterations, and final models were built using all of the data (Johnson et al., 2022).

For this modeling, various feature classes were used to model the relationships between species occurrence and environmental conditions. Linear features model relatively simple, linear relationships. For example, if a species prefers higher temperatures, a linear feature might show a consistent increase in the probability of presence as temperature rises. Quadratic features model non-linear relationships by including squared terms. For example, if a species thrives at moderate temperatures but is less likely to be present at very low or very high temperatures, this relationship forms a parabolic shape. Hinge features create piecewise linear splines to model abrupt changes in species presence. For example, a species might only present above a temperature threshold (Johnson et al., 2022). For the milkweed suitability models, Linear and Linear-Quadratic feature class settings were used. Hinge was not used since the relationships that we are modeling are relatively simple.

Regularization multipliers were used to penalize model complexity, where higher values indicate smoother, less complex models. This is used to retain only the variables with the greatest predictive contribution in the model (Johnson et al., 2022), thus performing feature selection. Regularization multipliers from 0.5 to 4 by increments of 0.5 were used, resulting in 16 candidate models (L, LQ, x1 each per incremental increase) for each of the four species of early-season milkweed.

From the 16 candidate models, we selected the one with the highest Area Under the Receiver Operator Characteristic Curve (AUC/ROC). The AUC is the total area under the plotted ROC line, which is the true positive rate plotted against the false positive rate, where a straight 45º line represents a 50:50 false positive to true positive ratio. This gives us an idea of model performance, as a straight line is no better than guessing species presence and we want to maximize the true positive rate, thus maximizing the area under the curve.

5.3 Interactive Web Dashboard Testing

5.3.1 Manual Testing

The primary mode of testing with the interactive web dashboard was manual testing. Manual testing was done by creating a code change with an expected return (ie. updating a raster file being used), and running the application locally. The interactive web dashboard was then checked to ensure it visually reflected the change made. Manual testing was performed after each change made in the interactive web dashboard code.

5.3.2 User Testing

To ensure the interactive web dashboard functionality and positive user experience, we tested the dashboard with the Santa Barbara Botanic Garden team. During the testing, users used the interactive web dashboard tabs to select different species, accessibility index layers, and provided feedback on areas that required additional development. The feedback gathered was used to enhance and update the design and development of the interactive web dashboard.

5.3.3 Deployment / Hosting

The interactive web dashboard is hosted through shinyapps.io and will be used by the client in this state to aid in the planning of upcoming surveys. The client has indicated that their tech team will ultimately be hosting and deploying the dashboard through their website. The SBBG will use this web dashboard as an example of how scientists use tools to aid in conservation efforts.

6. User Documentation

6.1 Overview of data folder and GitHub repositories file structures

This section contains references to the appropriate appendices that detail the file structures of the shared data folder we will be transferring to our client to facilitate their understanding of the file organization and how to find any component of the final product. It also contains the file structures for two public GitHub repositories created for this project: `milkweed-mod` and `milkweed-site-finder`.

6.1.1 Shared data folder

For a full overview of the file structure of the shared Box folder delivered to the SBBG, see Appendix II.

6.1.2 GitHub Repository: Habitat Suitability Modeling, Survey Site Accessibility & Survey Site Priority

For a full overview of the file structure of the `milkweed-mod` GitHub repository, see Appendix III. This file structure details where key elements are located within the repository, as well as where the client can find the walkthrough notebooks detailed in Section 6.2.1.

6.1.3 GitHub Repository: Interactive Web Dashboard

For a full overview of the file structure of the `milkweed-site-finder` GitHub repository, see Appendix IV. This file structure details where key elements are located within our interactive web dashboard repository, as well as the walkthrough notebook detailed in Section 6.2.2.

6.2 Updating Models and Outputs

6.2.1 Updates within model notebooks

The SBBG team has been provided full access to the GitHub organization so they can incorporate new milkweed survey data and make changes as needed. Details on how to update habitat suitability models with new data, add or remove roads and/or trails by name or open/closed status to update the Survey Site Accessibility Index, and consequently update the Survey Site Priority Index, are found in the Quarto Markdown notebook `updating_milkweed-mod.qmd`.

6.2.2 Updates within the Interactive Web Dashboard

To make updates to outputs displayed on the interactive web dashboard, please use the following process:

1. Run the notebooks with code changes that will create new outputs, following instructions laid out in `updating_milkweed-mod.qmd`. These notebooks include:
    - `A_californica_sdm.qmd`
    - `A_eriocarpa_sdm.qmd`
    - `A_erosa_sdm.qmd`
    - `A_vestita_sdm.qmd`
    - `create_accessibility_index.qmd`
    - `max_suitability_sdm.qmd`
    - `priority_sites.qmd`
    - `priority_sites_table.qmd`
    - `rescale_all_layers.qmd`

2. Push all updates and changes to the GitHub `milkweed-mod` repository.

Note: all new outputs created will be saved within the correct folders automatically. No new files should be created.

3. In RStudio within the `milkweed-site-finder` repository open `global.R` and re-run code

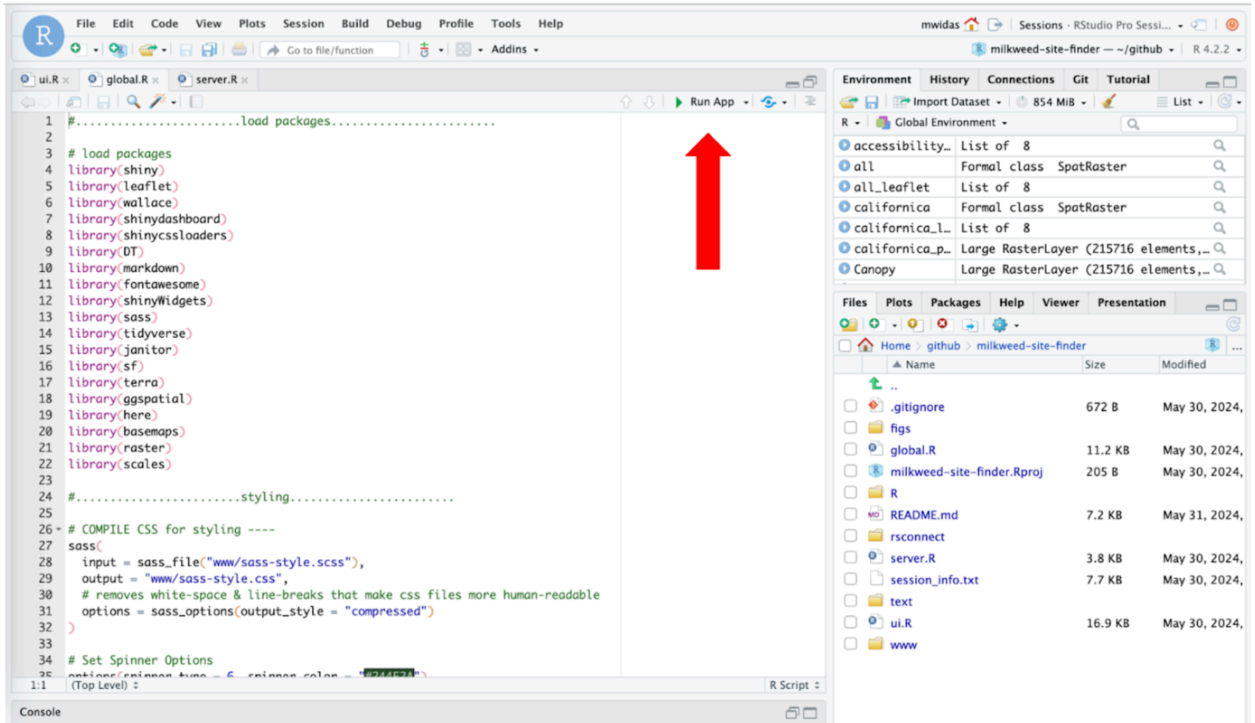

Figure 11. Image of milkweed-site-finder repository open in RStudio. The global.R file is open and in the upper right corner, there is a red arrow indicating where the 'Run App' button is to launch the web dashboard locally.

Note: No names or file paths should have changed during this data update, therefore all code should execute with the updates

Note: If a file name was changed during the creation or the replacement of data folders, the file path will need to be changed in the `global.R` where that specific file is called.

4. All data is loaded through a URL linking to where the file is saved in GitHub, therefore the interactive web dashboard when run locally will pull in updated files.
5. If new data is being added to the interactive web dashboard the data should be read in from a GitHub URL. Navigate to GitHub and copy the URL for the file, which will be in the following format:

   https://github.com/MEDS-SBBG-milkweed/milkweed-mod/blob/main/outputs/dashboard/californica_sdm.tif

   You will edit the URL to match this format:

   https://raw.githubusercontent.com/MEDS-SBBG-milkweed/milkweed-mod/main/outputs/dashboard/californica_sdm.tif

   Note: the two main changes are the section directly after "https://" and removing "/blob" from the URL.

6. To complete the update process the interactive web dashboard will need to be re-deployed through shinyapps.io. For more information on deployment please see Section 6.4.

6.3 Utilizing the Dashboard for Survey Planning

The main goal of the interactive web dashboard is to facilitate the planning of milkweed field surveys in the LPNF. The dashboard is deployed through a Shiny account created in the client's name so they will have full access and ownership. The dashboard was built to allow the SBBG team to visualize each component of the project and, most importantly, download a dataset of high-priority survey sites to visit.

6.3.1 Exploring Visualizations

To view locations where the SBBG team has found milkweed in previous surveys, navigate to the "Milkweed Locations"' tab and select any combination of milkweed species to explore on a zoomable map (Figure 6). To view and explore the outputs of the milkweed habitat suitability models for each species, navigate to the "Milkweed Habitat Suitability" tab (Figure 7). If interested in viewing where milkweed habitat is predicted to be suitable in the LPNF regardless of species, scroll to the bottom of the tab to view an interactive map built from the maximum suitability in each location for all of the four early-season milkweed species. To view the survey site accessibility index and the individual layers used to build it, navigate to the "Survey Site Accessibility" tab (Figure 8). Lastly, to view the survey site priority index for each milkweed species, navigate to the "High-Priority Survey Sites" tab and select the species of interest (Figure 9). At the top of each page, users can find more details on the contents of each tab and how to explore them.

6.3.2 Downloading High-Priority Survey Sites

The bottom of the "High-Priority Survey Sites" tab contains a reactive data table where each row corresponds to a raster cell in the LPNF and the columns correspond to the latitude and longitude, priority score for each milkweed species, accessibility score, and survey status indicating whether the SBBG team has visited the location previously (Figure 12). To filter the dataset, click on the "Search" bar in the top right (1) and type in a term. For example, if a user wanted to only look at locations previously visited by the SBBG team, they could type "Visited" in the search bar, and only rows containing "Visited" would populate in the table. To order the dataset based on a specific column, click on the up or down arrows next to the column name to sort in either ascending or descending order (2). Once the dataset is filtered and ordered to the user's liking, they can download it as a .csv file to their device by clicking on the "CSV" button in the top left (3).

	Longitude	Latitude	A. californica Priority	A. eriocarpa Priority	A. erosa Priority	A. vestita Priority	Accessibility Score	Survey Status
1	-121.82155	36.40125	0.341	0.494	0.000	0.000	0.256	not visited
2	-121.81325	36.40125	0.261	0.400	0.000	0.000	0.208	not visited
3	-121.80495	36.40125	0.091	0.157	0.000	0.000	0.105	not visited
4	-121.79665	36.40125	0.058	0.120	0.000	0.000	0.103	not visited
5	-121.78835	36.40125	0.065	0.153	0.000	0.000	0.096	not visited
6	-121.78005	36.40125	0.095	0.253	0.000	0.000	0.129	not visited
7	-121.77175	36.40125	0.000	0.000	0.000	0.000	0.000	not visited
8	-121.76345	36.40125	0.124	0.247	0.000	0.000	0.123	not visited

Figure 12. Downloadable data table located on the "High-Priority Survey Sites" tab on the dashboard.

6.4 App Deployment and Maintenance

Note: Section 6.4 is intended for Santa Barbara Botanic Garden Staff.

To deploy the interactive web dashboard, please use the following process:

7. Once any updates have been made within the `milkweed-site-finder` repository locally or updated models have been made in the `milkweed-mod` repository locally, make sure to save all changes in RStudio and push the changes to GitHub via either the RStudio Git GUI or through the command line.

Note: Once all changes have been saved in the local and upstream environments of the tool, we will start the deployment process:

8. Go to shinyapps.io and login. It is recommended to create your account and log in using GitHub. To utilize shinyapps.io, you first need to link your account with RStudio on your computer. Follow the instructions on shinyapps.io when you first create your account to install the **{rsconnect}** package and authorize your account. You can use the following image as a reference:

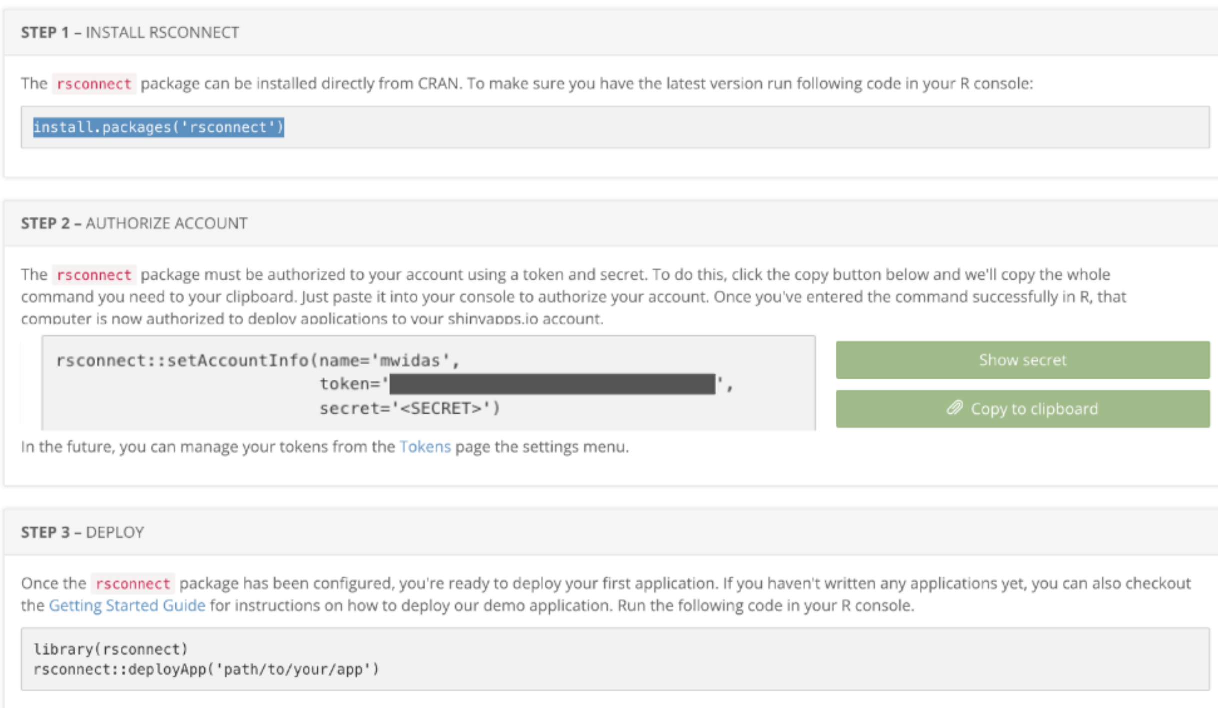

Figure 13. Instructions for configuring a shinyapps.io account for web dashboard deployment and maintenance. An example account is shown above.

9. Once your account has been created please follow the instructions under authorize account.

   Tip: It will be helpful to select "show secret" and then copy all of the code in the chunk.

```
# execute in the Command Line in RStudio to connect shinyapps.io
rsconnect::setAccountInfo(name = ‘account_name’,
                          token = ‘<token>’,
                          secret = ‘<secret>’)
```

10. You have now configured your environment to be connected to the shinyapps.io environment. The next step will be to deploy the application. You will follow the instructions under Step 3 - DEPLOY, as shown in Figure 13. To deploy this specific application please run the following code, updating the file path to reflect where it is stored locally:

```
# example of dashboard-specific directions
library(rsconnect)
# filepath to app directory
rsconnect::deployApp("file-path/milkweed-site-finder")
```

Tip: If the name of the app-directory name changes, this code will need to be updated to reflect the new naming convention.

11. Once deployed, a browser will open to your application. The URL will take the form: `https://username.shinyapps.io/your_app_directory_name`. You will also now see an updated `/rsconnect` folder within the app’s directory – this is generated when an application bundle is successfully deployed and information on the deployed content (i.e. the name, title, server address, account, URL, and time).
12. You will then save the updates that have been made to the app within the `/rsconnect` folder locally to save all updates
13. Once you have saved all changes in RStudio locally, you will push all of these changes to GitHub via either the RStudio GUI terminal or through the command line.

    Tip: It is important to save all the updates from the deployment within GitHub to ensure that the application deploys and is hosted successfully.

14. The application is now updated successfully and fully redeployed.

6.5 Deployment Troubleshooting

It is not uncommon for applications to have failed deployments. Our recommendation is to check the Log tab on shinyapps.io if you encounter an error with a deployment (Figure 14).

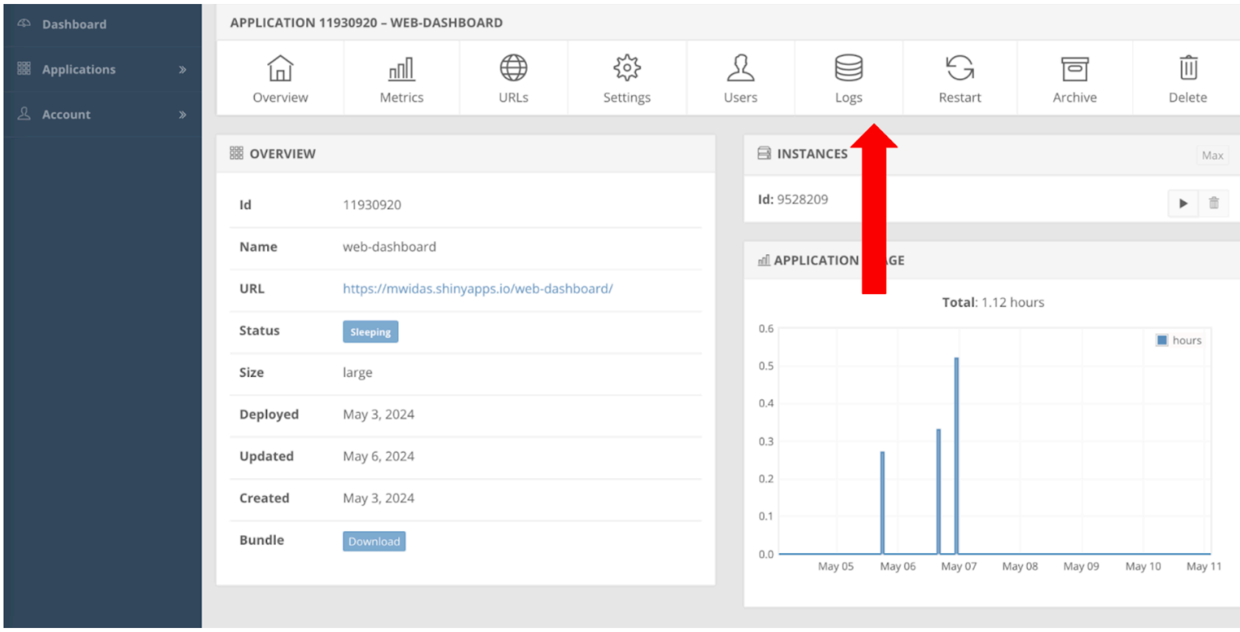

Figure 14. A red arrow indicating where the shinyapps.io Log tab can be found for reference when troubleshooting web dashboard deployment and maintenance.

- Problem 1: The application deployed successfully but a graphic is replaced with a white box and a red error message.

The next step is to navigate to the shinyapps.io website that is hosting the application and look at the log tab. The logs tab will present a record of actions taken during the deployment. You will be looking for any Warning: error message text. The text should be able to direct what next steps need to be taken.

- Problem 2: Could not find function "function-name"

To address this type of error message we recommend combing through your app-directory and making sure that any functions or variables are present. Once you have found the function and have ensured that it is being called into the app with the appropriate nomenclature you should be able to re-run and deploy it successfully.

Note: During updates, it can be easy to lose documents or move files while forgetting to update a specific file path

- Problem 3: Error: package 'package-name' version # was found, but >= different version # is required by 'package-name'

To address this type of error message we recommend upgrading the packages to later versions. In this case, this will require running install.packages("package-name") locally. Once you have installed the updated package and have ensured that it is running in the app locally you should be able to re-run and deploy it successfully.

- Problem 4: "Disconnected from server" message in the browser.

To address this type of error message we recommend visiting The Shiny Troubleshooting Guide. There could be a number of code errors that cause this error message to show including code dependencies, library package errors, incorrect file paths for loading data, and a few others.

7. Archive Access

Data archiving and access. As described in Section 4.3, all data, model, and index products were delivered to the client in a shared Box folder. The data used in our model development and interactive web dashboard are cited in Section 3.1. A formal Dryad archive of our intermediate data and final outputs was developed along with a comprehensive README.md detailing the data sources and archived products. See Appendix V for a comprehensive list of all archived data products. The following list is a high-level summary of the archived data, all of which have been cropped to the LPNF boundary:

- Bioclimatic data
- Canopy cover data
- DEM data, including aspect, eastness, northness, and slope.
- Land ownership data
- 2023 SBBG milkweed survey data as points (see Section 3.1, 3.2 for methodology) and centroids
- A raster stack of environmental layer data used in the SDM
- Data used to create the novel survey site accessibility index, including rasters of calculated distance from trails and distance from roads and the template raster (see Section 3.1 for details)
- Trail and road data
- Survey site accessibility index outputs, including each rescaled layer that defines the index
- Survey site priority index rasters for each species
- SDM outputs for each species and a raster of the maximum suitability across all species for each location in the LPNF
- Raw data from the "2023 Regional Trails and Roads" data created by the Los Padres Forest Watch

The environmental layer data we used for our models is publicly accessible in its raw formats, as is all of the survey site accessibility data with the exception of the LPNF trail and road data made by the Los Padres Forest Watch. Publicly accessible raw data was not included in the Dryad archive. All of these data sources are also documented in the model development GitHub repository's README.md, with clear sources and links following the UCSB RDS README.md guidelines. The Dryad archive is also linked in the README.md of our GitHub organization.

The trail and road data for the southern region of the LPNF used in this project is currently only accessible through ArcGIS, which is proprietary software. To ensure reproducibility for our client and anyone who wants to use our model, an R-compatible download of this shapefile was included in both the shared Box folder and our Dryad data archive.

With the permission of our client and associated parties, the cleaned SBBG milkweed survey data, all intermediate data products, and final outputs of our project were made publicly accessible via Dryad, accompanied by a disclaimer stating: "Plant and seed collection on Forest Service land is not permissible without a plant collection permit from the Los Padres National Forest."

Code. All scripts and notebooks developed for this project are publicly accessible through the GitHub repositories `milkweed-mod` and `milkweed-site-finder` within the GitHub organization `MEDS-SBBG-milkweed`. Because these repositories are publicly accessible, the scripts were not included in our Dryad archive. The coding environment which includes a list of all packages and versions used in each repository can be found in the `session_info.txt` file at the highest level of organization in each repository.

Intellectual property and re-use. The LPNF trail and road data that we downloaded from ArcGIS were made by the Los Padres Forest Watch and currently have no associated license beyond being considered "public access," according to the FeatureServer's ItemInfo. The USGS data used to identify trail and road data in the northern region of the LPNF resides in the Public Domain and may be used without restriction. Due to the publicly accessible nature of most of the data used and with the meticulous archive of our data products, researchers beyond the SBBG will be able to use our data cleaning and model scripts to reproduce our results or for developing other related projects. Both GitHub repositories are under MIT Licenses.

8. References

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A. , Vilela, B. & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography, 38: 541-545. URL https://onlinelibrary.wiley.com/doi/10.1111/ecog.01132.

Allaire, J., Horner J. & Xie, Y. (2023). _markdown: Render Markdown with 'commonmark'_. R package version 1.5, <https://CRAN.R-project.org/package=markdown>.

Atkins, A., Allen, T., Wickham, H., McPherson, J., & Allaire, J. (2024). _rsconnect: Deploy Docs, Apps, and APIs to 'Posit Connect', 'shinyapps.io', and 'RPubs'_. R package version 1.2.2, <https://CRAN.R-project.org/package=rsconnect>.

Bahmani, M., Naghdi, R. & Kartoolinejad, D. (2018). Milkweed seedlings tolerance against water stress: Comparison of inoculations with Rhizophagus irregularis and Pseudomonas putida. Environmental Technology & Innovation 10:111–121.

Boyle, J. H., Dalgleish, H. J. & Puzey, J. R. (2019). Monarch butterfly and milkweed declines substantially predate the use of genetically modified crops. Proceedings of the National Academy of Sciences 116:3006–3011.

California Department of Forestry and Fire Protection. (n.d.). California land ownership. [Data set]. California State Geoportal. https://gis.data.ca.gov/datasets/CALFIRE-Forestry::california-land-ownership/about

California Forest Observatory. (2020). A statewide tree-level forest monitoring system. [Data set]. Salo Sciences, Inc. https://forestobservatory.com

Chang, W. & Ribeiro, B. B. (2021). _shinydashboard: Create Dashboards with 'Shiny'_. R package version 0.7.2, <https://CRAN.R-project.org/package=shinydashboard>.

Chang, W., Cheng J., Allaire J., Sievert C., Schloerke B., Xie Y., Allen J., McPherson J., Dipert A., & Borges, B. (2022). _shiny: Web Application Framework for R_. R package version 1.7.4, <https://CRAN.R-project.org/package=shiny>.

Cheng, J., Mastny T., Iannone R., Schloerke B., & Sievert, C. (2023). _sass: Syntactically Awesome Style Sheets ('Sass')_. R package version 0.4.7, <https://CRAN.R-project.org/package=sass>.

Cheng, J., Schloerke B., Karambelkar B., & Xie Y. (2023). _leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library_. R package version 2.2.0, <https://CRAN.R-project.org/package=leaflet>.

Cheng, J., Sievert C, Schloerke B, Chang W, Xie Y, Allen J (2023). _htmltools: Tools for HTML_. R package version 0.5.6.1, <https://CRAN.R-project.org/package=htmltools>.

Csárdi, G., Hester, J., Wickham, H., Chang, W., Morgan M., & Tenenbaum, D. (2023). _remotes: R Package Installation from Remote Repositories, Including 'GitHub'_. R package version 2.4.2.1, <https://CRAN.R-project.org/package=remotes>.

Dunnington, D. (2023). _ggspatial: Spatial Data Framework for ggplot2_. R package version 1.1.9, <https://CRAN.R-project.org/package=ggspatial>.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17: 43-57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. [Data set]. International Journal of Climatology, 37(12), 4302-4315. https://www.worldclim.org/data/worldclim21.html

Firke, S. (2023). _janitor: Simple Tools for Examining and Cleaning Dirty Data_. R package version 2.2.0, <https://CRAN.R-project.org/package=janitor>.

Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography, 27(2), 245–256. https://doi.org/10.1111/geb.12684

GBIF.org. *Asclepias californica* Greene in GBIF Secretariat (2023). GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2024-01-28.

Hijmans, R. (2023). _raster: Geographic Data Analysis and Modeling_. R package version 3.6-26, <https://CRAN.R-project.org/package=raster>.

Hijmans, R. (2023). _terra: Spatial Data Analysis_. R package version 1.7-55, <https://CRAN.R-project.org/package=terra>.

Hijmans, R.J., Phillips, S., Leathwick, J., & J. Elith. (2022). _dismo: Species Distribution Modeling_. R package version 1.3-9, <https://CRAN.R-project.org/package=dismo>.

Iannone, R. (2023). _fontawesome: Easily Work with 'Font Awesome' Icons_. R package version 0.5.2, <https://CRAN.R-project.org/package=fontawesome>.

Iannone, R., Cheng, J., Schloerke, B., Hughes, E., Lauer, A. & Seo, J. (2023). _gt: Easily Create Presentation-Ready Display Tables_. R package version 0.9.0, <https://CRAN.R-project.org/package=gt>.

Johnson, Bethany A., Gonzalo E. Pinilla-Buitrago, Andrea Paz, Jamie M. Kass, Sarah I. Meenan, Robert P. Anderson. (2022). Wallace Ecological Modeling Application v2.0 Vignette. updated 2023-03-11. https://wallaceecomod.github.io/wallace/articles/tutorial-v2.html#partition-occurrences

Kass, J.M., Pinilla-Buitrago, G.E., Paz, A., Johnson, B.A., Grisales-Betancur, V., Meenan, S.I., Attali, D., Broennimann, O., Galante, P.J., Maitner, B.S., Owens, H., Varela, S., Aiello-Lammens, M.E., Merow, C., Blair, M.E., & R.P. Anderson. (2023). "wallace 2: a shiny app for modeling species niches and distributions redesigned to facilitate expansion via module contributions." Ecography, e06547. https://doi.org/10.1111/ecog.06547.

Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A., Soley-Guardia, M., & R.P. Anderson. (2021). "ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions." Methods in Ecology and Evolution, 12(9), 1602-1608. https://doi.org/10.1111/2041-210X.13628.

Los Padres Forest Watch. (2013). Monarch Butterfly.

Los Padres Forest Watch. (2023). 2023_Regional_Trails_and_Roads (FeatureServer). [Data set].

Los Padres National Forest - About the Forest. USDA Forest Service National Website. (n.d.). https://www.fs.usda.gov/main/lpnf/about-forest#:~:text=Los%20Padres%20National%20Forest%20encompasses,of%20two%20separate%20land%20divisions. Accessed 17 Jan 2024.

Müller, K. (2020). _here: A Simpler Way to Find Your Files_. R package version 1.0.1, <https://CRAN.R-project.org/package=here>.

Owens, H., Barve, V., & Chamberlain, S. (2023). _spocc: Interface to Species Occurrence Data Sources_. R package version 1.2.1, <https://CRAN.R-project.org/package=spocc>.

Pebesma, E., & Bivand, R. (2023). Spatial Data Science: With Applications in R. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

Pelton, E. M., C. B. Schultz, S. J. Jepsen, S. H. Black, & E. E. Crone. (2019). Western Monarch Population Plummets: Status, Probable Causes, and Recommended Conservation Actions. Frontiers in Ecology and Evolution 7.

Perrier, V., Meyer, F., & Granjon, D. (2024). _shinyWidgets: Custom Inputs Widgets for Shiny_. R package version 0.8.1, https://CRAN.R-project.org/package=shinyWidgets.

Pleasants, J. M., & Oberhauser, K. S. (2013). Milkweed loss in agricultural fields because of herbicide use: effect on the monarch butterfly population: Herbicide use and monarch butterflies. Insect Conservation and Diversity 6:135–144.

Phillips, S.J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. In Proceedings of the Twenty-First International Conference on Machine Learning, pages 655-662. PDF

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of biogeography, 41(4), 629-643.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sali, A., & Attali, D. (2020). _shinycssloaders: Add Loading Animations to a 'shiny' Output While It's Recalculating_. R package version 1.0.0, https://CRAN.R-project.org/package=shinycssloaders.

Santa Barbara Botanic Garden. (n.d.). *Our Mission*. https://sbbotanicgarden.org/about/mission-vision.

Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida.

State and Federally Listed Endangered and Threatened Animals of California. (2024). State of California: Natural Resources Agency, Department of Fish and Wildlife, Biogeographic Data Branch, California Natural Diversity Database (CNDDB).

United States Department of Agriculture Forest Service. (2024). Administrative forest boundaries [Data set]. https://data.fs.usda.gov/geodata/edw/datasets.php.

U.S. Geological Survey. (2019-2021). USGS 3D Elevation Program Digital Elevation Model [Data set]. https://apps.nationalmap.gov/downloader/.

U.S. Geological Survey. (2024). USGS National Transportation Dataset (NTD) for California [Data set]. https://www.sciencebase.gov/catalog/item/5f6345ee82ce38aaa238c9df.

Wickham, H. (2011). “testthat: Get Started with Testing.” The R Journal, 3, 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). “Welcome to the tidyverse.” Journal of Open Source Software, 4(43), 1686. doi:10.21105/joss.01686.

Wickham, H., Hester, J., & Bryan, J. (2023). _readr: Read Rectangular Text Data_. R package version 2.1.4, <https://CRAN.R-project.org/package=readr>.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.3, <https://CRAN.R-project.org/package=dplyr>.

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). Chapter 5: Data Tidying. In *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (2e). O’Reilly. Retrieved May 9, 2024, from https://r4ds.hadley.nz/data-tidy.

Wickham, H., Vaughan, D., & Girlich, M. (2023). _tidyr: Tidy Messy Data_. R package version 1.3.0, <https://CRAN.R-project.org/package=tidyr>.

Wickham, H. & Seidel, D. (2022). _scales: Scale Functions for Visualization_. R package version 1.2.1, <https://CRAN.R-project.org/package=scales>.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding

Principles for scientific data management and stewardship. Sci Data 3, 160018 https://doi.org/10.1038/sdata.2016.18
Xie, Y., Cheng, J., & Tan, X. (2023). _DT: A Wrapper of the JavaScript Library 'DataTables'_. R package version 0.27, <https://CRAN.R-project.org/package=DT >.

9. Appendices

Appendix I: Summary of Software & Tools

R Package(s) & Version	Citation(s)	Use
tidyverse (2.0.0): including **dplyr** (1.1.3), **tidyr** (1.3.0), **readr** (2.1.4). **janitor** (2.2.0) **here** (1.0.1)	Wickam et al., 2019 Wickam et al., 2023 Wickam et al., 2023 Firke, S., 2023 Müller, K., 2020	Cleaning, wrangling, tidying data, reproducible file paths
raster (3.6.26) **sf** (1.0.14) **terra** (1.7.55) **stars** (0.6-4)	Hijmans, R., 2023 Pebesma, E., 2018 Hijmans, R., 2016 Pebesma & Bivand, 2023	Geospatial data processing and manipulation, reading in Bioclim data
dismo (1.3.9) **wallace** (2.1.1) **spocc** (1.2.1) **spThin** (0.2.0)	Hijman et al., 2022 Kass et al., 2023 Owens et al., 2023 Aiello-Lammens et al., 2015	Species Distribution Modeling
ggplot2 (3.4.4) **leaflet** (2.2.0) **plotly** (4.10.2) **ggspatial** (1.1.9) **gt** (0.9.0)	Wickam H., 2016 Cheng et al., 2023 Sievart, C., 2020 Dunnington, D. 2023 Iannone et al., 2023	Visualization
ENMeval (2.0)	Kass, J.M. et al., 2021.	Testing
shiny (1.7.4) **shinyWidgets** (0.8.1) **leaflet** (2.2.0) **shinydashboard** (0.7.2) **shinycssloaders** (1.0.0) **DT** (0.27) **markdown** (1.5) **fontawesome** (0.5.2) **sass** (0.4.7) **htmltools** (0.5.6.1) **wallace** (2.1.1) **scales** (1.2.1) **rsconnect** (1.2.2)	Chang et al., 2022 Perrier et al., 2024 Cheng et al., 2023 Chang et al., 2021 Sali et al., 2020 Xie et al., 2023 Allaire et al., 2023 Tannone et al., 2023 Cheng et al., 2023 Cheng et al., 2023 Kass et al., 2023 Wickham & Seidel, 2022 Atkins et al., 2024	Dashboard development & testing

Table 2. Summary of Software & Tools. Does not include all package dependencies. For full computing environment specifications, please refer to the `session_info.txt` files in each repository.

Appendix II: Shared Data File Structure

```
├── clean_data
│   ├── bioclim
│   │   └── wallace_bioclim.tif
│   ├── canopy_cover
│   │   └── canopy_cover_cleaned.tif
│   ├── dem
│   │   ├── eastness.tif
│   │   ├── lpnf_aspect.tif
│   │   ├── lpnf_dem.tif
│   │   ├── lpnf_slope.tif
│   │   └── northness.tif
│   ├── lpnf_boundary
│   │   ├── lpnf_boundary
│   │   │   ├── lpnf_boundary.shp
│   │   │   └── ...
│   │   ├── lpnf_boundary_buffered
│   │   │   ├── lpnf_boundary_buffered.shp
│   │   │   └── ...
│   │   ├── lpnf_boundary_north
│   │   │   ├── lpnf_boundary_north.shp
│   │   │   └── ...
│   │   ├── lpnf_boundary_north_buffered
│   │   │   ├── lpnf_boundary_north_buffered.shp
│   │   │   └── ...
│   │   ├── lpnf_boundary_south
│   │   │   ├── lpnf_boundary_south.shp
│   │   │   └── ...
│   │   └── lpnf_boundary_south_buffered
│   │       ├── lpnf_boundary_south_buffered.shp
│   │       └── ...
│   ├── lpnf_land_ownership
│   │   ├── lpnf_land_ownership.shp
│   │   └── ...
│   ├── milkweed_data
│   │   ├── sdm_milkweed_points
│   │   │   ├── californica_points.csv
│   │   │   ├── eriocarpa_points.csv
```

```
│   │   │   ├── erosa_points.csv
│   │   │   └── vestita_points.csv
│   │   └── survey_location_centroids
│   │       ├── all_species_points.shp
│   │       └── ...
│   ├── sdm_env_stack
│   │   ├── env_stack2.tif
│   │   ├── env_stack_north.tif
│   │   └── env_stack.tif
│   ├── site_accessibility
│   └── trails_and_roads
│       ├── forest_watch
│       │   ├── forest_roads_south.shp
│       │   ├── ...
│       │   ├── forest_trails_south.shp
│       │   └── ...
│       └── usgs
│           ├── usgs_roads_north.dbf
│           └── ...
├── outputs
│   ├── priority_sites_outputs
│   ├── sdm_outputs
│   │   ├── allpoints_bioclim_canopy_dem.rda
│   │   ├── allpoints_bioclim_canopy_dem.tif
│   │   └── californica_bioclim_canopy_dem.tif
│   └── site_accessibility_outputs
└── raw_data
    ├── California_Land_Ownership
    │   ├── ownership23_1.shp
    │   └── ...
    ├── canopy_cover
    │   ├── KernCounty-Vegetation-CanopyCover-2020-Summer-00010m.tif
    │   └── ...
    ├── dem
    │   ├── USGS_1_n35w119_20190919.tif
    │   └── ...
    ├── milkweed_polygon_data
    │   ├── MilkweedPolygon_Data.shp
    │   └── ...
    ├── S_USA_AdministrativeForest.gdb
    │   ├── gdb
    │   └── ...
    └── trails_and_roads
```

```
        ├── 2023_Regional_Trails_and_Roads_lines
        │   ├── 2023_Regional_Trails_and_Roads_lines.shp
        │   └── ...
        └── CA_transportation
            ├── Meta_DatasetDetail.dbf
            ├── Trans_RoadSegment_0.shp
            ├── Trans_TrailSegment.shp
            └── ...
```

Appendix III: milkweed-mod GitHub Repository File Structure

```
├── data_cleaning                # data cleaning and preparation
│   ├── accessibility_template
│   │   └── template_raster.qmd
│   ├── bioclim
│   │   └── bioclim.R
│   ├── boundary
│   │   └── lpnf_boundary.qmd
│   ├── canopy_cover
│   │   └── canopy_cover.qmd
│   ├── combine_layers
│   │   └── crop_stack.R
│   ├── dem
│   │   └── dem_cleaning.qmd
│   ├── land_ownership
│   │   └── land_ownership.qmd
│   ├── milkweed_polygon            # milkweed data cleaning & processing
│   │   ├── milkweed_subsets_points.qmd
│   │   ├── points_clean_export.R
│   │   └── survey_points_map.R
│   ├── solar_rad
│   │   └── solar_radiation.qmd
│   └── trails_and_roads
│       └── trails_and_roads.qmd
├── legends                      # legend development for maps
│   └── legend_infographic.qmd
├── LICENSE
├── maxent                       # species distribution modeling/habitat
suitability map development
│   ├── A_californica_sdm.qmd
│   ├── A_eriocarpa_sdm.qmd
│   ├── A_erosa_sdm.qmd
```

```
│   ├── A_vestita_sdm.qmd
│   └── max_suitability_sdm.qmd
├── milkweed-mod.Rproj
├── outputs
│   ├── dashboard                      # outputs used in
milkweed-site-finder dashboard
│   │   ├── accessibility_legend.png
│   │   ├── all_species_points.rda
│   │   ├── californica_sdm.tif
│   │   ├── eriocarpa_sdm.tif
│   │   ├── erosa_sdm.tif
│   │   ├── max_suitable_sdm.tif
│   │   ├── priority_legend.png
│   │   ├── suitability_legend.png
│   │   └── vestita_sdm.tif
│   └── figs
│       └── MilkweedMod-transparent.png   # hex sticker
├── priority_sites                        # survey site priority index
development
│   ├── priority_sites.qmd
│   └── priority_sites_table.qmd
├── R                               # scripts
│   ├── accessibility_setup.R
│   ├── addLegend_decreasing.R
│   ├── milkweed_maxent.R
│   ├── rescale_raster.R
│   └── setup.R
├── README.md
├── session_info.txt            # computing environment specs
├── site_accessibility          # survey site accessibility index
development
│   ├── create_accessibility_index.qmd
│   ├── distance_calculations.qmd
│   └── rescale_all_layers.qmd
└── walkthroughs                 # walkthroughs for updating this repo
    └── updating_milkweed-mod.qmd
```

Appendix IV: milkweed-site-finder GitHub Repository File Structure.

```
├── .gitignore
├── figs                        # folder containing figures used in README.md
│   ├── data.png
```

```
│   ├── habitat_suitability_models.png
│   ├── high_priority.png
│   ├── home.png
│   ├── milkweed_locations.png
│   ├── MilkweedMod.jpeg
│   └── survey_site_accessibility.png
├── global.R
├── LICENSE
├── milkweed-site-finder.Rproj
├── R
│   └── addLegend_decreasing.R        # function to create legend in leaflet
├── README.md
├── rsconnect                         # folder for app deployment
│   └── shinyapps.io
│       └── mwidas                    # this will change to username of account
│           └── milkweed-site-finder.dcf
├── server.R
├── session_info.txt      # session_info and operating system information
├── text    # folder containing all markdown files used and where text is stored
│   ├── authors.md
│   ├── background_context.md
│   ├── background_info.md
│   ├── disclaimer.md
│   ├── habitat_suitability_all.md
│   ├── high_priority_table.md
│   ├── overview_data.md
│   ├── overview_habitat_suitability.md
│   ├── overview_milkweed_locations.md
│   ├── overview_site_accessibility.md
│   ├── overview_site_finder.md
├── ui.R
└── www                                # graphics folder
    ├── fonts
    │   ├── 5cfb33b712cbdaf9310b.woff2
    │   ├── ab9aea6faeaea5115410.woff2
    │   ├── bace964bccc02d90d74e.woff2
    │   ├── da1fdeda6d5756ee6227.ttf
    │   ├── ebec98131cf900cb698e.woff2
    │   └── edfac2e6370304cd74ae.woff2
    ├── monarch_milkweed.jpeg
    ├── sass-style.css
    ├── sass-style.scss               # sass file for styling
    └── SBBG_logo.png
```

Appendix V: Data Archive File Structure

```
├── clean_data
│   ├── bioclim
│   │   └── wallace_bioclim.tif
│   ├── canopy_cover
│   │   └── canopy_cover_cleaned.tif
│   ├── dem
│   │   ├── eastness.tif
│   │   ├── lpnf_aspect.tif
│   │   ├── lpnf_dem.tif
│   │   ├── lpnf_slope.tif
│   │   └── northness.tif
│   ├── lpnf_land_ownership
│   │   ├── lpnf_land_ownership.shp
│   │   └── ...
│   ├── milkweed_data
│   │   ├── sdm_milkweed_points
│   │   │   ├── californica_points.csv
│   │   │   ├── eriocarpa_points.csv
│   │   │   ├── erosa_points.csv
│   │   │   └── vestita_points.csv
│   │   └── survey_location_centroids
│   │       ├── all_species_points.shp
│   │       └── ...
│   ├── sdm_env_stack
│   │   ├── env_stack2.tif
│   │   ├── env_stack_north2.tif
│   │   ├── env_stack_north.tif
│   │   └── env_stack.tif
│   ├── site_accessibility
│   │   ├── roads_distance_raster.tif
│   │   ├── template_raster.tif
│   │   └── trails_distance_raster.tif
│   └── trails_and_roads
│       ├── forest_open_roads_south.shp
│       ├── ...
│       ├── forest_open_trails_south.shp
│       ├── ...
│       ├── forest_roads_south.shp
│       ├── ...
│       ├── forest_trails_south.shp
│       └── ...
├── outputs
│   ├── priority_sites_outputs
```

```
│   │   ├── californica_priority.tif
│   │   ├── eriocarpa_priority.tif
│   │   ├── erosa_priority.tif
│   │   └── vestita_priority.tif
│   ├── sdm_outputs
│   │   ├── californica_sdm.rda
│   │   ├── californica_sdm.tif
│   │   ├── eriocarpa_sdm.rda
│   │   ├── eriocarpa_sdm.tif
│   │   ├── erosa_sdm.rda
│   │   ├── erosa_sdm.tif
│   │   ├── max_suitable_sdm.tif
│   │   ├── vestita_sdm.rda
│   │   └── vestita_sdm.tif
│   └── site_accessibility_outputs
│       ├── access_index_final.tif
│       ├── canopy_rescaled.tif
│       ├── ownership_rescaled.tif
│       ├── roads_rescaled.tif
│       ├── slope_rescaled.tif
│       └── trails_rescaled.tif
└── raw_data
    └── trails_and_roads
        └── 2023_Regional_Trails_and_Roads_lines
            ├── 2023_Regional_Trails_and_Roads_lines.shp
            └── ...
```