User Guide for Reserve-Level Tools

Developed as part of

‘A National Synthesis of Tidal Marshes to Detect Impacts of Climate Change across Multiple Scales’

Last update: 2025-02-04

Contents

1. Project Background

a. Abstract

Tidal marshes are facing tremendous pressures from climate change, including accelerated rates of sea-level-rise, increased storm activity and precipitation extremes, lengthened growing seasons and shifts in salinity regimes. Recent research within National Estuarine Research Reserve (NERR) sites has revealed striking changes in plant communities that seem to be caused by rising seas, including shifts toward more flood tolerant species, lower overall plant diversity, and in some cases growing swaths of bare ground.

This three-year collaborative research project builds on the results and tools generated by two catalyst projects that looked at changes in vegetation and marsh surface elevation. The project team will use a variety of NERR datasets to quantify shifts in plant species ranges and patterns of diversity and productivity across latitudes and regions, and then evaluate local and regional drivers of vegetation change. The team will engage staff from 21 reserves as well as external advisors to help integrate existing datasets from across the country and ensure that tools and products are useful for a variety of audiences. In addition to providing key insights into how climate change is affecting marshes nationwide, the project will generate reserve-specific templates and automated tools for data analysis and visualization that can be easily updated, and with modification could be transferable to other organizations with marsh monitoring or other coastal vegetation datasets (e.g., seagrasses, mangroves). Ultimately, the project will elevate the visibility and advance the application of the Reserve System's long-term monitoring program and provide a framework for facilitating other national-level synthesis efforts.

This work was sponsored by the National Estuarine Research Reserve System Science Collaborative, which supports collaborative research that addresses coastal management problems important to the reserves. The Science Collaborative is funded by the National Oceanic and Atmospheric Administration and managed by the University of Michigan Water Center (NA19NOS4190058).

b. Find us online

Project Website: https://nerrssciencecollaborative.org/project/Peter20

Dashboard: https://www.arcgis.com/apps/dashboards/2981ed793f6d44e593c3c3a050678521

c. Purpose of Reserve-level code and this guide

The project team developed a data entry template for NERRS monitoring data, and transformed existing reserve data into this new format. Reserve-level code was developed not only to perform analyses during Namaste, but also so that reserve staff can entry data into the created format and easily re-run the analyses in the future. Additionally, an interactive app was developed that allows users to re-shape their data from the Namaste format into the format required for CDMO submission. This guide is meant to walk reserve technical staff through the developed tools and enable this future data use.

2. Directory Structure

The folder ***Namaste-Reserve-Final*** will be referred to as the "project directory". It contains a few useful files and folders.

a. Folders

data – This is where your reserve's data (ReserveCode_veg.xlsx) and analysis spec (ReserveCode_veg-specs.xlsx) files are stored. You can add to or edit the data file, and make changes in the analysis spec file. If your reserve has multiple components, you may have multiples of each file type. Make sure the prefix for each component's veg file matches the prefix for that component's spec file.

output – This folder is where the html files produced by the '01_run_analyses' R script will be written. Feel free to change the names of the output files, particularly if you want to run the analysis multiple times in a single day.

R – This folder contains all the R code. You generally should not need to open this folder from file explorer, but only within the .Rproj file discussed below.

renv – This folder will not be in your initial download of the project directory, but if you take the 'renv' option of installing packages, it will be created. You should not manually modify anything in this folder.

b. Files

Namaste User Guide.pdf – The guide you are reading now!

Namaste-Reserve-Final.Rproj – This is what you should open in order to interact with any of the R code and produce this project's outputs.

renv.lock – This is a special file that works with the 'renv' package to ensure all of the R packages you install are the same versions with which this R code was developed, so in theory everything will work smoothly.

3. Data file

The data file for your reserve starts with your 3-letter SWMP Reserve code and ends with "_veg.xlsx". If you have multiple reserve components/files, a component code will be attached to the Reserve code with a dash – e.g. "APA-LSM_veg.xlsx". Should you wish to divide your own data file into multiple files, make sure to identify them before the "_veg" in the name, as in the APA example.

a. 'Metadata' Sheet

The Metadata sheet has detailed information on the other sheets' contents and how to update them. Please refer to the metadata sheet if you want more information than what is in this guide.

b. 'Cover' Sheet

The Cover sheet is the sheet from which the R code pulls data for analyses and re-shaping to CDMO format. There is one row per vegetation plot per sample date, with species as columns.

Columns A through N contain identifying information for the sample, as well as orthometric height and/or distance measurements if they were taken in a given year. There are orthometric height and distance columns in the Station Table sheet but they are meant to represent those measurements at plot establishment, and should not be updated in the Station Table. If you take updated readings, those can be recorded in the Cover sheet.

Column O is 'Total', which is the sum of all plant and abiotic cover, excluding water and overstory. This should be 100 if using ocular cover.

Species columns contain measurements or estimates of that species' cover and should only contain numeric entries. All species and abiotic columns should be listed in the Species Names sheet. Abiotic groups (e.g. bare, unvegetated, wrack) are the left-most of the 'species' columns, and then species are in alphabetical order. **Do not** enter text into these cells – if a reading was missing, mark it in the F_ column for the species with <-2>. If a reading was less than a required threshold (e.g. < 1%), choose a number to represent the situation (e.g. 0.5) and use that number in the cells. Note in the Metadata sheet, Methods Overview, Cover section that you did so.

To the right of the species columns are columns that contain species names, preceded by "Density_" and "Height_", the latter of which may be "Average Canopy Height", "Maximum Height", or another variation. There should only be one underscore in the column names, to separate the measurement type from the species. These columns are meant to contain the **average of any replicate measurements** of density and Height. Raw measurements of density and height should go into the Density and Height sheets of the data file.

To the right of these columns are additional columns, with species names preceded by "F_". These are QA/QC Flag columns and can contain text. As this guide is being written, detailed QA/QC flags and codes do not exist for vegetation data, but you should still use <-3> if a data point should be rejected for some reason.

c. 'Station_Table' Sheet

The Station Table contains one row per vegetation plot, with static information related to the plots. Latitude, longitude, and vegetation zone values are pulled from this table and tied to each unique vegetation plot for the Namaste analyses.

The combination of SiteID-TransectID-PlotID should be unique for each. For ease of future CDMO submissions, make sure you use Site, Transect, and Plot IDs that match previous CDMO submissions your reserve has made. (If you have to make changes in the station table, make sure to update existing entries in the Cover, Density, and Height sheets.)

Latitude and Longitude should be in decimal degrees. These coordinates are used for all spatial graphics in the Namaste output.

Type – does this plot represent Emergent vegetation monitoring, Submerged, or Mangrove? Towards the end of the project the point was raised that Type should possibly be associated with each vegetation species, not a monitoring plot – keep an eye on any guidance from the vegetation working group and/or CDMO. In this table and file, however, Type represents each plot.

Vegetation Zone represents the general zone the plot belonged in when it was established. Column names in the 'Ecotone Migrators' sheet of the analysis spec file should include any vegetation zones represented in your Station Table. Vegetation Zones represented in the Namaste project include:

- M-Mudflat
- S-Seaward Edge
- L-Low Marsh
- T-Transition: transition of low to high marsh
- P-Pools/Pannes
- H-High Marsh
- UE-Upland Edge
- DB-Dunes and Berms

Orthometric Height, Height at MLLW, Distance to Water, and Distance along Transect again represent these values at plot establishment. Updated values can be entered into the Cover, Density, and Height sheets when necessary, but the values in the Station Table should remain constant (unless you are fixing an error).

SSAM-1 is a yes/no column for whether each vegetation plot is part of that reserve's sentinel site monitoring.

d. 'Species_Names' Sheet

The Species Names sheet contains one row per vegetation species or abiotic grouping. This includes any variations on "unknown" and "unidentified".

The left-most column starting with "Species" should be the one that contains the most current names. This should be column E. See cell B34 of the Metadata sheet for instructions on updating species names. When you do update a species, make sure to also update it in the Cover, Density, and Height sheets. Remember that the species name may be in multiple places in each – its own column *and* the F_species column in Cover, plus potentially Density_species and Height_species columns in the Cover sheet.

Columns to the left of species, used to describe and/or aggregate by various grouping levels, are:

- Plant_Categories – Used only for live vegetation species, to group into the following categories: A-Algae, B-Brackish, F-Freshwater, H-Halophyte, U-Upland.
- Native_Classification – Used only for live vegetation species. Plants are assigned a category of: Native, Non-native, Native invasive, or Non-native invasive.
- Cover_Categories – Used for ALL rows. Specifies broad groupings of all cover types into Live vegetation (all living plants and algae), Unvegetated category (dead vegetation, other abiotic categories such as bare and live non-vegetation such as invertebrates), or Other layer (cover types that should be analyzed separately, e.g., water, overstory).

- NMST_Groupings – Used to group columns of “things other than live vegetation” into consistent categories of Bare, Rock, Dead (non-living plant material originating from the plot from prior years), Wrack (non-living plant material originating from outside the plot typically brought in by the tides), Wood (large woody debris), Water (standing water present at low tide), Overstory (canopy cover of larger vegetation, typically trees, shading the plot), ‘Other Unvegetated’ (catch all for any other abiotic/faunal cover).

Groupings from Plant_Categories, Cover_Categories, NMST_Groupings, or names the left-most Species column can be specified as choices for any analysis in the Analysis_Specs sheet of the analysis specs file. The Native_Classification category can only be used in custom metrics at this time. Capitalization, punctuation, and spelling for choices in the analysis specs file must match ***exactly*** what is in this Species Names table so that everything can be properly linked.

e. ‘Density’ and ‘Height’ Sheets

The Density and Height sheets are set up like the Cover sheet discussed above: one row per vegetation plot per sample date. Columns exist for replicates that we had access to while formatting data into the Namaste format. These values should be averaged and transferred into the “Density_species” and “Height_species” columns of the Cover sheet. There are no QA/QC F_ columns in these sheets. Any QA/QC notes should be made in the appropriate F_species column of the Cover sheet.

4. Analysis Specs file

a. ‘Ecotone Migrators’ Sheet

EMI, Ecotone Migration Index, is the proportional cover of species/covers or species/cover groupings that are expected to increase within a vegetation zone as sea level rises. The Ecotone Migrators sheet of the analysis specs file is where you can define which species/cover groupings you expect to increase within each zone. You can modify this sheet.

The Ecotone Migrators listed in the sheets developed during Namaste originated from conceptual models that each Reserve made in google slides (Fig. 1). To define Ecotone Migrators, we took the categories that were listed with an expectation of increased cover along with increased sea level rise, and listed them in the columns associated with each vegetation zone.

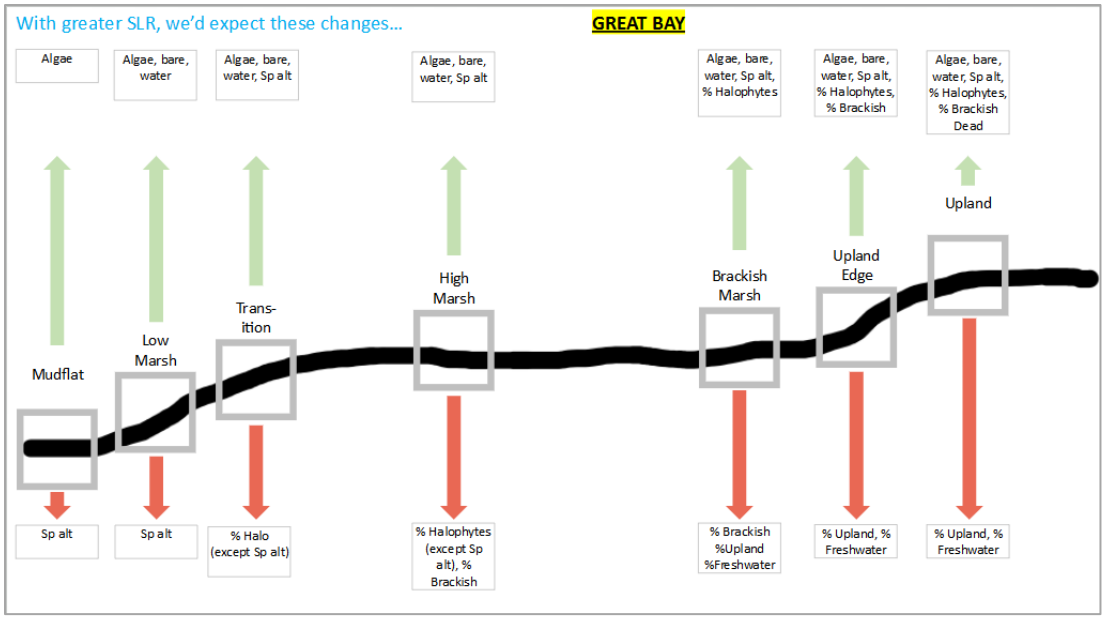

Figure 1: Conceptual Model developed by Great Bay

The columns in the Ecotone Migrators sheet represent vegetation zones from the conceptual diagrams, with the naming convention used in the Station Table of the data file. They should be ordered from low elevation to high elevation from left to right; ordering of vegetation zones in the Namaste outputs are based on this column ordering.

It is best to list single species before any groupings they may belong to, because the code is set up to avoid double-counting if the species comes before the grouping.

Species, Plant_Categories, Cover_Categories, and NMST_Groupings from the Species Names table are all valid entries in the Ecotone Migrators sheet. Please ensure their spelling, punctuation, and capitalization match the Species Names sheet ***exactly***.

Only zones that are also present in the Station Table will actually be used in the Namaste calculations and outputs, so do not worry if you see extra zones (of course, feel free to remove them). Only species or groups present in the Species Names table will be used in EMI calculation. For example, if 'Water' is entered as an ecotone migrator, but is not measured and there is no column for it in the data file, it will be disregarded. 'Water' will also be disregarded if there is a column in the data sheet but it has different capitalization – 'water' in the data sheet will not be seen as a match to 'Water'.

b. 'Analysis_Specs' Sheet

The Analysis Specs sheet is the primary sheet where you can modify Namaste outputs. The column where you should make modifications is entitled 'Choice' and is highlighted yellow. You can make choices for:

- Univariate analyses, rows 2-5. Choose up to 4 species or groups, each of which will make up a tab in the Univariate Analyses section of the resulting output file. See the 'Output File' section of this document for more detail.

- Custom metrics, rows 6-7. Optionally, develop up to 2 mathematical combinations of species/categories. If specified, will be used in univariate analyses only. Tabs in the Univariate Analyses section of the output file will be created for each custom metric that is specified.
    - Species, group, or column names must be inside backticks: ``` `, e.g. `Spartina alterniflora` + `Juncus roemerianus`. Backticks are different from apostrophes. The backtick can be found in the upper left corner of a standard keyboard, above the tab key.
    - Use parentheses as needed, if you are adding species together and dividing by something else. E.g. (`Spartina alterniflora` + `Juncus roemerianus`) / (`Unvegetated` + 1)
    - Remember not to specify a denominator that could be 0; this will cause errors in the scripts. Add 1 or some other constant in a denominator to avoid this.
- Spatial bar charts, rows 8-10. Choose up to 3. "Other" category will be calculated from the rest unless specified otherwise in the More_Options sheet (see below).
- Summary bar charts AND NMDS loading factors (arrows), rows 11-18. Choose up to 8. In averaged bar charts, "other" category will be calculated from the rest, unless specified otherwise in the More_Options sheet (see below).

Species and categories you put into the 'Choice' column must match ***exactly*** the spelling, punctuation, and capitalization of those species/categories in the Species_Names sheet of the data file.

c. 'More_Options' Sheet

This sheet provides analysis options beyond species/category choices. The column you should fill in is again entitled 'Choice' and again highlighted in yellow. Options include:

- Generating spatial bar plots by transect rather than by site. Row 4 in the sheet; this defaults to 'Site' but there is a dropdown menu and you can select 'Transect' instead. Remember this will multiply the number of spatial graphs accordingly!
- *Not* lumping species into an 'Other' category if they are not among the few choices in the Analysis_Specs sheet. This can be helpful if your key species generally have very low cover and the 'Other' category dominates the bar charts. The default is to lump species into the 'Other' category, and Row 5 of the sheet is 'Yes' by default. To "turn off 'Other'", change this to 'No' using the dropdown menu.
- Create names in rows 9 and 10 for the custom metrics you specified in the Analysis_Specs sheet. Names can have spaces, underscores, and periods, but should not use other special characters.
- Limit the starting and/or ending years used in analyses in rows 14 and 15. The year you specify *will* be included in analyses, so if you want to exclude (e.g.) 2014 and all that came before, enter 2015 as the starting year.
- Either analyze a single site or exclude a single site using rows 19 and 20. Make sure the spelling, capitalization, and punctuation of the site name ***exactly*** match the site name in the Station_Table sheet of the data file.
- Do more with NMDS:
    - Add other explanatory factors (they must be present in the Cover sheet) as loading arrows. You can specify multiple columns in row 24, and separate them by a comma. Make sure the column names ***exactly*** match those in the data file.

- Run an NMDS that includes all years, not only the first and last. The default choice in row 25 is 'No' because this can take a lot of computing resources, but change it to 'Yes' to give the bigger NMDS a try.

5. R scripts and their use

a. Setting up on a new computer

1. **Install R & RStudio.** R before RStudio. Instructions for both are here: https://posit.co/download/rstudio-desktop/
2. **Operating System additional needs.** Your computer needs to be able to compile packages from various configurations in order for these tools to all work correctly together. Please make sure you have the following software, which is involved in said package compilation.
    a. **Windows:** install **RTools.** That is *not* a typical R package; it's another download. You can get it here: https://cran.r-project.org/bin/windows/Rtools/
    b. **Mac:** make sure you have **Xcode** and **gfortran**. See instructions here; a link for gfortran is in the 2nd sentence: https://cran.r-project.org/bin/macosx/tools/
3. **Download the Namaste directory**. Make sure you download this *to your C drive.* It may work on a networked drive, but *will not work if your files are in OneDrive.* If you do not already have a copy you are transferring to your computer, download it from github: https://github.com/swmpkim/Namaste-Reserve-Final
    a. Look for the green button on the right that says 'Code'. Click it, and a dropdown menu comes up.
    b. Choose the bottom option, 'Download ZIP'.

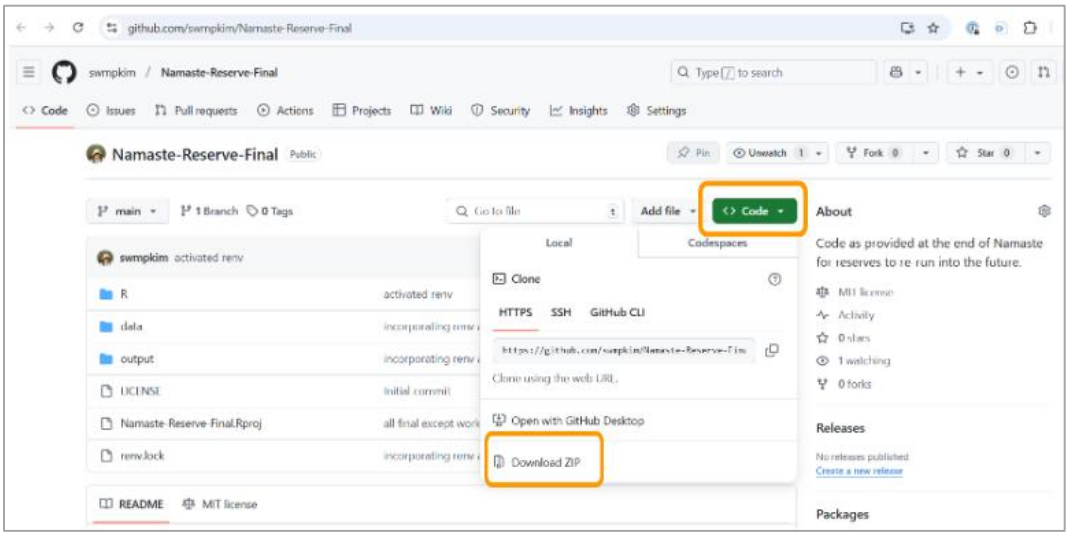

Figure 2. Code and Download ZIP options to download directory from GitHub.

    c. Unzip that somewhere on your computer - this is the main directory to work out of for all reserve-level Namaste tools.
4. **Put your reserve's files into the 'data' folder.** If you need help obtaining your data and analysis spec files, please contact Kim Cressman or Chris Peter.

5. **Install required R packages** by running **00_install_packages.R** within the .Rproj. Detailed instructions are in the list below this paragraph. This setup uses a package called 'renv', which will install the versions of the packages that were used during code development. These versions of the packages will only be installed in the project directory, and will not affect (or install) the packages for other R instances on your computer.
    a. Open the 'Namaste-Reserve-Final.RProj' file (see Fig. 3) by double-clicking from Windows Explorer.
    b. This will open up an instance of RStudio. You can verify that the Project is open by looking in the upper right corner, where you will see the same icon and name as you did in Windows Explorer (circled in Fig. 4).
    c. In the bottom right pane, the Files pane, click on the 'R' folder (circled in Fig. 4). This will work like Windows Explorer and show you files within that folder.
    d. From the Files pane, click '00_install_packages.R'.
    e. Run the line of code by placing your cursor in the line and hitting the 'Run' button in the upper right corner of the pane containing the script or by using the keyboard shortcut Cmd/Ctrl-Enter.
    f. When you run this line, renv::restore(), you will be asked in the Console pane (lower left) if you want to proceed. Type the letter Y (or y – type it as capitalized in the question; they switch it up to make sure you're paying attention to the choices you make) and hit enter to run. You should see the packages getting installed.
    g. If you see any errors in the console (common packages that will cause errors are MASS, survival, and nlme), see step 5a.2. above and make sure your system-specific software is installed. Restart R and try running this script again. If you still see errors, contact Kim Cressman.
    h. Good to go!
6. **Set up your file codes** in [***R/01_run-Analyses.R***] On line 18 of the script, make sure the file(s) you want to use are specified inside the parentheses. If you have done this before and saved the file, you won't need to do anything. If you do need to change it, put the prefix of your files inside quotation marks. If you have multiple files to use, separate them with commas, e.g.

```
files <- c("APA-LSM", "APA-PC")
```

b. Preparing to work with either Data Analysis Script or Shiny App

i. Setup

1. Verify that the 'data' folder of the project directory contains both your data file and your analysis specs file.
2. Make sure that the "_veg" and "_veg-specs" files have matching prefixes.
3. If you have multiple files, make sure all prefixes match.

ii. Open the RStudio Project

1. Open the 'Namaste-Reserve-Final.RProj' file (see Fig. 3) by double-clicking from Windows Explorer.

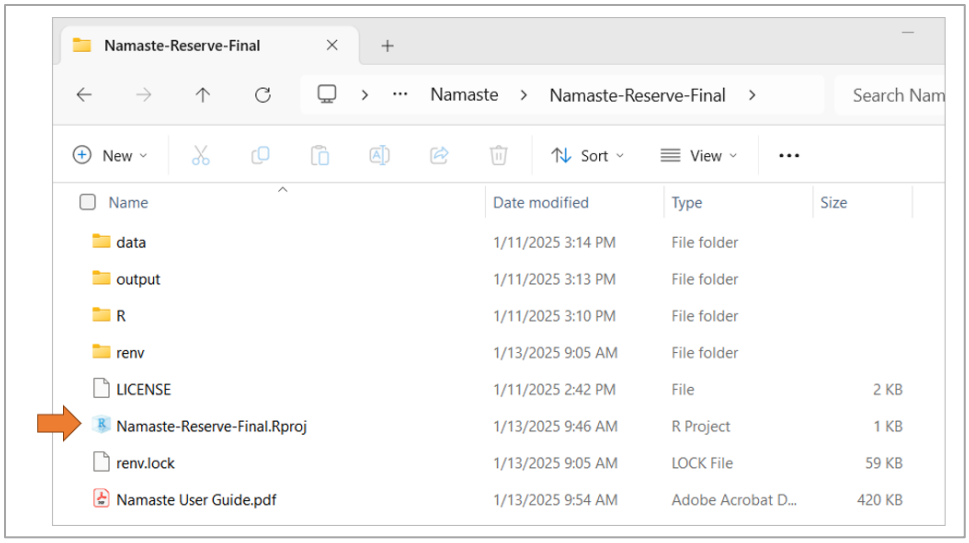

Figure 3: .Rproj file to open when working with R scripts

2. This will open up an instance of RStudio. You can verify that the Project is open by looking in the upper right corner, where you will see the same icon and name as you did in Windows Explorer (circled in Fig. 4).
3. In the bottom right pane, the Files pane, click on the 'R' folder (circled in Fig. 4). This will work like Windows Explorer and show you files within that folder. From here you can select either of the primary R files to work with.

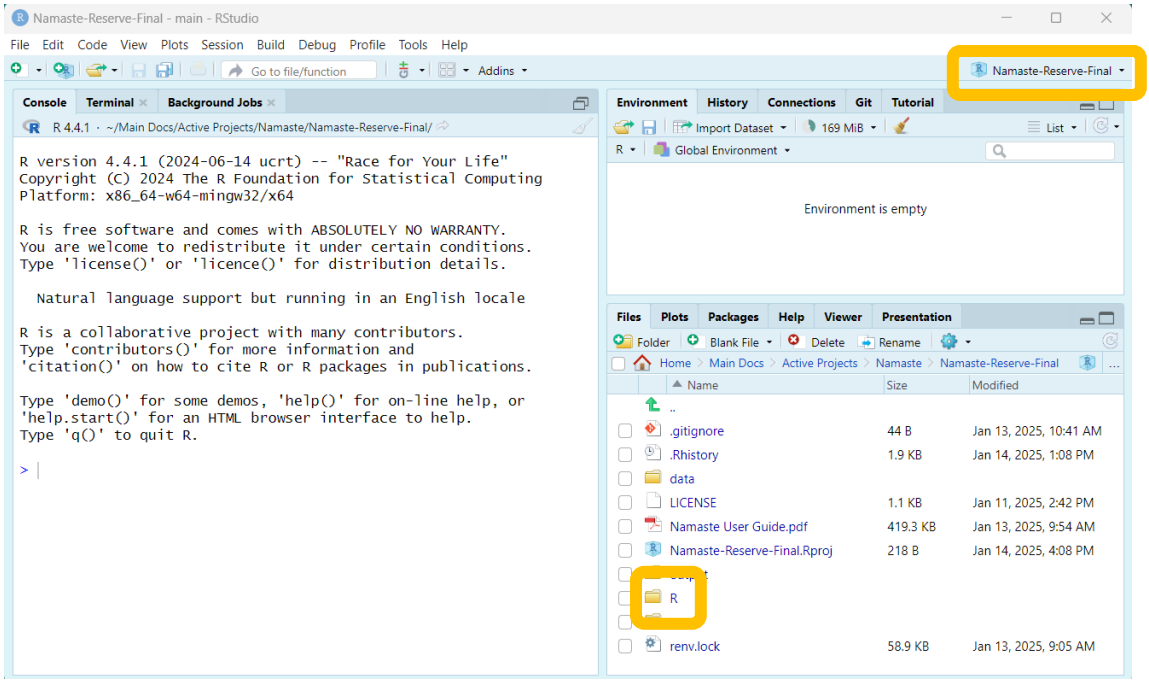

Figure 4. RStudio as opened from the .Rproj file.

c. Run the Data Analysis Script

1. Open the .RProj file and the 'R' folder as in the section above.
2. From the Files pane (bottom right), open the file '***01_run-Analyses.R***' by clicking on it once. This will open the script in a pane in the upper left.
3. If you have been using R, restart it from the 'Session' menu.
4. On line 18 of the script, make sure the file(s) you want to use are specified inside the parentheses. If you have done this before and saved the file, you won't need to do anything. If you do need to change it, put the prefix of your files inside quotation marks. If you have multiple files to use, separate them with commas, e.g. `files <- c("APA-LSM", "APA-PC")`
5. Run the script. If there are multiple files, it will loop through them.
    a. Select the entire script. The keyboard shortcut is Ctrl-a on Windows and Cmd-a on Mac.
    b. Run all, by either selecting the 'Run' button in the upper right corner or using the keyboard shortcut Ctrl/Cmd-Enter.
6. Let it run. You will see some things going by in the lower left pane ('Console', which may switch focus to 'Background Jobs').
7. When the file(s) is finished, you should see a record of the time it took in the console, along with a message saying "All files processed successfully!" If a file was problematic, the console should have a message telling you "The following files failed to process:", with file codes. Make sure those files exist, match, and meet all parameters otherwise specified in this document.
8. Go into the 'output' folder of the project directory and look at the output file!

d. Run the Shiny App to generate CDMO-formatted data

1. Open the .RProj file and the 'R' folder as in the section above.
2. From the Files pane (bottom right), open the file '***02_dataFormattingApp.R***' by clicking on it once. This will open the script in a pane in the upper left.
3. If you have been using R, restart it from the 'Session' menu.
4. There is only one line of code in this script. Put your cursor in it (anywhere) and run it, by either selecting the 'Run' button in the upper right corner or using the keyboard shortcut Ctrl/Cmd-Enter.
5. The app will pop up in a new window. In the upper left corner is a 'Browse' button so you can select the file you want to use (this should be your _veg.xlsx file), using your system's interactive explorer. When it uploads, some preview information will appear in the first tab.
6. Work your way through the tabs:

    - 1. Review data – gut check that the file looks right, has the correct sites and sampling frequency, etc.
    - 2. Shape to CDMO layout. In the left sidebar, make some choices, then hit the 'Reshape!' button at the bottom of the sidebar. This will show a preview of re-shaped data in the right pane. The choices you have are:
        - What year(s) to work with? Years present in the data file are sorted with the most recent at the top. You can select more than one option.

        - What species (if any) to exclude from CDMO reporting? These columns will be removed from the data before any re-shaping occurs. This defaults to species marked 'Other layer' in the Cover_Categories column of the Species_Names sheet in the data file. When this app was developed, categories like Overstory and Water were not reported to the CDMO. As this guide is being written, that is changing, so please check on current CDMO recommendations. You can remove species that are already selected, using the backspace key; and add any others that you don't want to report (e.g. if you measure shellfish cover) from a drop-down list.
        - What species should be included in 'Unvegetated'? This defaults to species labeled as 'Unvegetated Category' in the Cover_Categories column of the Species_Names sheet in the data file. Some reserves measure many specific categories that the CDMO does not track beyond 'Unvegetated'; this is where you can let the app add them together for you. As with the box above, you can select from a drop-down list, and remove options if needed.
    - 3. Verify plot totals – This section adds all of the cover totals for all of the species/categories remaining in the file. The table is sortable by clicking on the arrows next to the column names. We recommend sorting total, which will first sort from low to high; then click the arrows again to sort from high to low. These values should be near 100, and if any are not, you should investigate your data file, determine why, and make any necessary updates. Then come back to the app.
    - 4. Add descriptive columns – there is again a left sidebar with some choices, based on the required "Distance", "Orthometric Height", and "Height Relative to MLLW" columns. Make your selections and then hit the "Use these choices" button at the bottom of the sidebar to see a preview of the entire CDMO-shaped data file on the right.
        - What does distance represent? It is left up to Reserves to determine what the "Distance" column will represent, and specify it in the CDMO metadata. Make that choice here.
        - Then choose which worksheet to pull information from for each of Distance, Orthometric Height, and Height Relative to MLLW. For parameters that do have updated measurements taken through time, we recommend using the Cover sheet of the data file. For parameters that do not change, we recommend the Station Table.
    - 5. Download CDMO layout – Click the button to download your reshaped data as an excel file! A window will pop up that lets you name it (we recommend following CDMO naming guidelines) and choose where to save it.

6. Output File

Output files start with the file prefix specified in the 'run analyses' script and include the date generated, in yyyy-mm-dd format. Files will be overwritten if you run the script multiple times in a day, so make sure to re-name them if you want to run more that day but keep the earlier ones.

The output files are in html format because there are many interactive elements. They will open in your internet browser, but they are stored on your computer; they are not online.

Major categories are in a floating table of contents in the upper left corner. These are detailed in sections below. Subcategories within each major category are in tabs that are spread horizontally across the screen. Each major category has an 'About' tab with details of the section.

a. 'About this Document' section

This section has instructions on how to click around the file. It also has some notes on data inclusion that have been discussed elsewhere in this document, but are meant as reminders in the output.

b. 'Exploration and QA/QC' section

This section contains information on sampling effort, vegetation zone assignment, some overview time series, and overviews of Ecotone Migration Index (EMI).

i. Missing/Removed Data tab

This tab contains tables of:

- Plots/dates without samples – rows where no numeric data was entered
- Plots without enough samples – we required at least 3 years of data for a plot to be included in statistical analyses. If any plots did not meet this threshold, they are listed here. These plots are still included in exploratory graphics and summary bar charts.
- Data flagged suspect or reject – information about any data points that were removed due to information in the F_ column for the species.

ii. Sampling Info tab

This tab contains tables of:

- Number of samples at each site by year
- Number of samples in each vegetation zone by year
- Species in the file, and their overall mean and max cover values for each site. This species table is sortable and searchable.

iii. Plots and Zones tab

This tab contains a simple graphic where x represents longitude and y represents latitude. Point color and shape both represent vegetation zone as assigned in the Station_Table tab of the data file. Each site is in a separate graph panel. This is not as nice as a map (or as the Namaste dashboard) but is a good way to make sure your plots seem to be in the correct location and the vegetation zones make sense.

iv. Time-series- Species tab

This tab contains two types of time series graph, each on two sets of vegetation species.

The sets of vegetation species are:

- Up to 4 species or groupings, as identified in the 'Univariate Linear Models' section (rows 2-5) of the Analysis_Specs sheet. With or without the remaining species combined as 'Other'.
- The three species that, across the entire data file, had the highest mean cover. With remaining species combined as 'Other'.

The first graphic contains one graph panel for each vegetation zone. Time is on the x-axis, and percent cover is on the y-axis. There is one point for each vegetation plot on each date. Points are colored by species or vegetation group, and a loess smooth for each species or vegetation group is overlaid.

The second graphic also has time on the x-axis, cover on the y-axis, a point for every plot, and a line for each species or group. Instead of separate panels for each vegetation zone, there is a single panel for the entire dataset, combined.

v. EMI tab

This tab starts with a table showing which species were counted as "migrators" within each vegetation zone. If a group (e.g. 'H-Halophyte') was specified in the Ecotone_Migrators sheet, the species comprising that group are shown in this table.

The table is followed by two time series of EMI, each with time on the x-axis and EMI on the y-axis, and a panel for each vegetation zone. The first time series has a single loess smooth per zone. In the second time series, points are colored by Site and there is a loess smooth for each site. This allows you to investigate whether the same zone may be experiencing different changes at different sites.

The final graph in this tab is a histogram of EMI, with one panel per vegetation zone.

c. 'Summary Bar Charts' section

There are two main types of summary bar chart in this section, and four tabs.

- Averaged stacked bar charts - show the relative distribution of species and cover classes and how these relationships change over time. Charts have been created at the Site, Zone, and Site x Zone levels (these are 3 of the 4 tabs).
- Spatial stacked bar charts - show the relative distribution of species and cover classes as above, but for each plot. These charts are laid out spatially by site or transect, as specified in the More_Options sheet of the analysis specs file. For ease of interpretation, only 4 evenly-spaced time points are used along the x-axis in each chart, no matter how long the time series.

Each stacked bar chart was generated with species chosen in the Analysis_Specs sheet (up to 8 for averaged bar charts; up to 3 for spatial bar charts) and a second time with the three species with highest mean cover across the dataset.

d. 'Univariate Analyses' section

In the Univariate Analyses section, change through time is assessed on one response variable at a time. Each response makes up one of the tabs. Every reserve has EMI, Richness, and Diversity tabs; up to 4 tabs for species or groups identified in the Analysis_Specs sheet, Univariate Linear Models row; and up to 2 custom metrics specified in the Analysis_Specs sheet.

Key questions addressed are:

- Are there shifts in key vegetation species/groups over time?
- Do these shifts vary by vegetation zone?

The main statistical model used is a linear mixed model, via lme4::lmer(), with the form:

y ~ Vegetation Zone + Time + Zone*Time + 1|Plot

where y is the response variable given in the tab, and each individual plot has a random intercept.

If only one vegetation zone is present in the data file, Vegetation Zone is removed and the statistical model is simplified to:

y ~ Time + 1|Plot.

Each tab contains several pieces of output:

- Type III ANOVA table from the model
- Table of estimated marginal slopes (change per year for each zone) – generated whether or not the Zone*Time interaction was statistically significant
- Table of overall slope (change per year), only generated if the Zone*Time interaction was not statistically significant ($p > 0.05$)
- R^2 values, marginal and conditional
- Graphs of model-predicted values:
    - Single panel, with regression line for each vegetation zone and no points
    - Multi-panel: one panel per vegetation zone, points represent data (one point per vegetation plot), overlay is regression line for that zone
- Contrasts plot, presented two ways. If the Zone*Time interaction was significant ($p < 0.05$), pairwise comparisons of slope by zone were conducted and letters represent their groupings. If the interaction was not significant, letters are not present but the graphs are still generated. Note, these graphs are only generated if multiple vegetation zones are present in the file. On this graph, zone is along the x-axis, and change is on the y-axis. The point is the estimate of change and the whiskers represent the 95% confidence interval for that estimate.
    - The first plot has zones arranged from lowest to highest change.
    - The second plot has zones arranged in the same order in which they appear in the Ecotone_Migrators sheet of the analysis specs file, presumably low to high elevation (or moving from the water toward the upland).

e. 'Multivariate Analyses' section

i. Questions and statistical methods:

- Are there shifts in the vegetation community (as defined by the entire percent-cover matrix) over time? Do these shifts vary by vegetation zone?
    - addressed with PERMANOVA, comparing first year of monitoring to the 'last' (most recent) year of monitoring. PERMDISP results are also included for help in interpreting the PERMANOVA.

    - the first and last year of monitoring at a specific vegetation plot should be considered together. To do this, permutations were restricted to allow swapping of time point within a vegetation plot, but keep both time points of a plot together as plots are permuted across vegetation zones. Unfortunately these restricted permutations do not account for the repeated measures within a plot. Ideally we could use a random effect, as in the univariate models. At this point such a model for PERMANOVA is not possible in R.
    - if the above test indicated that differences in time between vegetation zones were significant or nearly so ($p <= 0.10$), the species matrix was split by Vegetation Zone. A PERMANOVA was run for each vegetation zone, again with restricted permutations to allow only permutation of time point within vegetation plot.
- Which species/groups contribute most to these shifts?
    - addressed using SIMPER, comparing first and last year of monitoring data within each zone. SIMPER is used to follow up on vegetation zones where the p-value for the PERMANOVA was <= 0.2.
- Where is the plant community changing, and what characteristics do those areas have in common (e.g. site, zone, distance from water, elevation)?
    - visualized via NMDS.

ii. About tab

In addition to the information given above on questions and statistical techniques, the About tab includes information on the data frame itself. Only plots that were present both in the first and last year of monitoring for a vegetation zone were retained, and counts of these plots (and what the first and last year are for each zone) are provided in a table. Additionally, a list of species included in the response matrix is included. We did not group species in the analyses; the raw species matrix was used (with any "Other layer" groupings and plots without at least 3 years of data removed). Species groupings *were* used as NMDS loading factors if specified, by finding the centroid of the species comprising the groupings.

iii. PERMANOVA tab

Overall section

The PERMANOVA was set up to test the null hypothesis of no difference in community change (or lack thereof) across vegetation zones. Categories are 'Time_group', representing the first year and last year of monitoring ('start' and 'end') and 'Vegetation_Zone'. The ANOVA table shows results for each term, sequentially (first to last). If the interaction row, 'Time_group:Vegetation_Zone' is significant ($p < 0.05$), we have evidence that community change (or lack thereof) is different in different zones. If this term is not significant, the main effect of 'Time_group' can be evaluated to determine whether there is evidence of a difference between 'start' and 'end' overall.

If only one vegetation zone was present in the file, this overall PERMANOVA was not run, and results for the single vegetation zone can be found in the Zone-wise section.

Zone-wise section

A separate PERMANOVA for 'start' vs. 'end' was run for each zone if the interaction term in the overall PERMANOVA was <= 0.1. The null hypothesis being tested is, for each zone, that there is no difference in the vegetation community between 'start' and 'end'. The summary table presented shows a p-value for each zone, followed by a Bonferroni-adjusted p-value for each zone.

Check for homogeneity of dispersion section

A PERMDISP is run in this section to test the assumption that dispersions do not differ between groups. Dispersion is the multivariate equivalent of variance. If this assumption is violated, caution should be used in interpreting PERMANOVA results, as significant results could be the result of differing dispersions rather than differences between group means. Groups in this instance are the combination of vegetation zone and time group; e.g. “Low Marsh, Start (first year of monitoring)”; “Low Marsh, End (last year of monitoring)”; “High Marsh, Start”; “High Marsh, End”.

The table displays information about the test, along with the F statistic and p-value. If this p-value is significant (< 0.05), at least one group has different dispersion from the others, but we do not know which group or groups are different, and should examine the information graphically. Below the table is a boxplot of the distances from each sample to the group centroid within each zone, with the individual sample distances overlaid as jittered points. Below the boxplot is a table with the mean distance to centroid for each group.

Finally, if zone-wise PERMANOVAs were generated, PERMDISP was also run for each zone. The mean dispersion for each Zone x Time point combination should be the same (within rounding error) as in the table above, but the test is only for start and end within each zone. PCoA plots of dispersions are provided in a folded section; click the arrow next to ‘Click to view plots of dispersion for each zone’ to see them.

iv. SIMPER tab

SIMPER was run if:

- Across all zones: the interaction term in the overall PERMANOVA was not significant (p > 0.05) and the main effect for time was significant or close to it (p <= 0.2).
- Zone-wise: the interaction term in the overall PERMANOVA was significant or close to it (p <= 0.1) and the within-zone effect for time in the zone-wise PERMANOVA was significant or close to it (p <= 0.2).
- Generally only one version of the SIMPER will be run (across all zones vs. zone-wise), but when the interaction term was near significance (0.05 < p < 0.1), SIMPER was run both ways.

The p-values determining the above logic are unadjusted. Due to the exploratory nature of these analyses, we did not adjust p-values for multiple comparisons.

Tables are presented for each SIMPER that was run. Each table contains the top 6 species contributing to differences. Columns are:

- average - the average contribution of that species to the Bray-Curtis distance between the two groups (note, this is *not* expressed in % and the column does not total to 1).
- sd - the standard deviation of the species’ contribution.
- cumulative - the cumulative % contribution for this species and all those above it in the table. Typically people only report species up to the one that brings “cumulative” over 0.7.
- p - a p-value for that species based on permutation tests.
- mean_start - the mean cover of that species in that vegetation zone in the first year of monitoring

- mean_end - the mean cover of that species in that vegetation zone in the last year of monitoring

v. NMDS – start/end tab

The start/end NMDS can be considered a visual representation of the PERMANOVA results. This NMDS used Bray-Curtis dissimilarity on the full species matrix of the reserve, and 3 dimensions. The NMDS start/end tab contains several pieces of output and information.

First is a table of NMDS stress and how to interpret stress values. The stress value for the reserve's NMDS is printed in bold above this table. Additional sections have headers:

2-dimensional NMDS plot

A biplot of the first 2 axes of the ordination. Each small point represents the community of a vegetation plot at one of the 'start' or 'end' time points. Point color represents vegetation zone. Open circles with crosshatches in the middle represent the 'start' time point (first year of monitoring). Filled circles represent the 'end' (last year of monitoring). Large points represent the centroid for each Vegetation Zone x Time Point combination. Hovering over a centroid will show the full vegetation zone name and the year represented. Black arrows and labels represent prioritized loadings, as identified by the reserve in the Analysis_Specs sheet of the analysis specs file. Hovering over an arrow will show the species or group it represents. If a group was chosen to be a loading factor, its location is the centroid of the locations of the comprising species.

Contour plots

These plots are below a fold; click on the arrow next to 'Click to expand contour plots' to see them. These plots are non-interactive versions of the 2-dimensional plot above, but instead of arrows to represent all loadings at once, there are contours to represent values of one loading at a time. This may be more informative than arrows alone. Currently this only works on individual species from the response matrix, and will not work on species groups.

3-d NMDS plot

This is an interactive plot so you can explore the entire cloud of points. As with the 2-d plot, point color represents vegetation zone. Open circles represent the 'start' time point (first year of monitoring). Filled circles represent the 'end' (last year of monitoring). There are no centroids in this plot; only individual points for each plot. Red lines and labels represent prioritized loadings, as identified by the reserve in the Analysis_Specs sheet of the analysis specs file.

You can zoom in and out, rotate the graphic, and generally explore the data. Hovering over the graphic will make a menu bar pop up in the upper right corner of the plot, and you can find more options there.

Optional additional loadings

If additional loadings were specified in the More_Options sheet of the analysis specs file, they will be graphed here. The base plot is the same as the 2-dimensional NMDS plot above, but non-interactive. Additional loadings will be displayed as both arrows and contours (separately).

vi. NMDS – all years tab

By default, this analysis is not run because it may take more computing power than is available to perform ordination on many years' worth of data points. This can be changed in the More_Options sheet of the analysis specs file.

Text in the output will tell you if the NMDS was attempted or not. If it was attempted, and worked, there will be 3 plots. The first two are interactive: you can zoom in on any part(s) of the graph to take a closer look.

- Fixed axes: one panel for each vegetation zone. Each point represents a plot's measurement for a year. The centroid for each year within each vegetation zone is a text label of that year. In this graph, the axes are fixed so emphasize the relative locations of the zones' centroids relative to each other.
- Free axes: All is as above, but the x- and y-axis limits vary based on the points in that vegetation zone. This means the starting point of the graph is zoomed in more for each vegetation zone.
- Loadings: As with the start/end NMDS, these arrows represent the loadings of up to 8 species or groups identified by the reserve in the Analysis_Specs file.

f. 'Documentation' section

This section logs the R and R package versions used to generate the output document.

7. Troubleshooting

a. Packages don't install.

If you have package installation troubles, check the following:

1. Make sure the Namaste directory is either on your computer or a networked drive. A network drive might work; it if doesn't, move the directory to your C drive. ***Nothing will work properly if you try to run this via OneDrive***.
2. Make sure you're in the RStudio Project. Look in the top right corner of the RStudio window for a blue box with the letter R inside. Next to that, you should see 'Namaste-Reserve-Final'. If you are not in the project, see Section 5b of this document.
3. See Section 5a.2 of this document and make sure you have the proper compilation software (RTools for Windows; Xcode and a Fortran compiler for Mac). Figure 5 shows examples of outputs when this software is not installed.
4. Re-run the line renv::restore(), either through the package installation script or by typing it in the console and pressing enter to run it.
5. If you continue to have problems, you can install via a different script – this other script will not control package versions. The script is in R/sourced and is named '99_original-install_packages.R'. Run it one line as a time to install packages individually.

```
Error: Error installing package 'MASS':
================================

* installing *source* package 'MASS' ...
** package 'MASS' successfully unpacked and MD5 sums checked
** using staged installation
** libs
Error in system(paste(MAKE, p1(paste("-f", shQuote(makefiles))), "compilers"),  :
  'make' not found
* removing 'C:/Users/[redacted]/Namaste-Reserve-Final-main/renv/staging/1/MASS'
install of package 'MASS' failed [error code 1]
```

```
- Downloading writexl from CRAN ...          OK [file is up to date]
Successfully downloaded 116 packages in 45 seconds.

# Installing packages --------------------------------------------------------
- Installing MASS ...                       OK [linked from cache]
- Installing Matrix ...                     OK [linked from cache]
- Installing survival ...                   OK [built from source and cached in 18s]
- Installing nlme ...                       FAILED
Error: Error installing package 'nlme':
================================

* installing *source* package 'nlme' ...
** package 'nlme' successfully unpacked and MD5 sums checked
** using staged installation
** libs
using C compiler: 'Apple clang version 16.0.0 (clang-1600.0.26.6)'
sh: /opt/gfortran/bin/gfortran: No such file or directory
using SDK: 'MacOSX15.2.sdk'
/opt/gfortran/bin/gfortran -arch arm64  -fPIC  -Wall -g -O2  -c chol.f -o chol.o
make: /opt/gfortran/bin/gfortran: No such file or directory
make: *** [chol.o] Error 1
ERROR: compilation failed for package 'nlme'
* removing '/Users/ladmin/Downloads/Namaste-Reserve-Final-main/renv/staging/1/nlme'
--------------------------------------------------------------------------------
R was unable to find the gfortran binary.
gfortran is required for the compilation of FORTRAN source files.
Please check that gfortran is installed and available on the PATH.
Please see https://stackoverflow.com/q/35999874 for more information.

Reason(s):
- 'sh: /opt/gfortran/bin/gfortran: No such file or directory'
- 'make: /opt/gfortran/bin/gfortran: No such file or directory'
install of package 'nlme' failed [error code 1]
```

Figure 5. Errors in package compilation. Highlighted text indicates the portions of the errors that indicate said problems. Top, red text: Windows. Solution: install RTools. Bottom, black text: Mac. Solution: install gfortran. See Section 5a.2 for details.

b. The output file doesn't generate.

These errors are frustrating because the messages you get don't seem to* tell you much about the problem. *you can learn more if you dig through all of it, so please copy it into an email if you need to reach out to Kim for help.

So far they have followed one of two patterns.

1. You see a message in the console that says only "*Quitting from line 2*" or similar. If this is the case, you may actually have package installation problems.
    a. Type renv::status() into the console and hit enter. If it says "No issues found -- the project is in a consistent state", move on to the next list.
    b. If it returns any other message, move up in this list to the package installation troubleshooting section.
    c. Reach out to Kim Cressman if you need to (don't waste hours trying to troubleshoot this on your own).
2. Another message may be along the lines of *some file "does not exist"*. This means you should verify the following things:
    a. Make sure both "_veg.xlsx" and "_veg-specs.xlsx" files are in the "data" folder, and that they have matching prefixes.
    b. Make sure the Namaste directory is either on your computer or a networked drive. A network drive might work; it if doesn't, move the directory to your C drive. Nothing will work properly if you try to run this via OneDrive.
    c. Make sure you're in the RStudio Project. Look in the top right corner of the RStudio window for a blue box with the letter R inside. Next to that, you should see 'Namaste-Reserve-Final'. If you are not in the project, see Section 5b of this document.
    d. Make sure that prefix is used in line 18 of the 01_run-Analyses.R script, inside quotation marks and parentheses. E.g. `files <- c("GRB")` or `files <- c("GRB", "GND")`

c. The output file generates, but I don't see output I expect.

Check the text in the area you don't see output; effort was made to document why something might not appear. General things to check are:

1. Spelling, capitalization, and punctuation in your Analysis_Specs sheet of the analysis specs file. Make sure the options you included ***exactly*** match species or group names in the Species_Names sheet of the data file.
2. If you have updated a species in the Species_Names file, make sure that update propagated to the column names in the Cover, Density, and Height sheets. Errors in the script would be caused by the Cover sheet in particular – check the species column as well as any columns starting with Density_, Height_, and F_ to make sure any suffixes of the species name are updated. Again, check your spelling! Typos are easy to make.
3. Is a species you chose in the Analysis_Specs sheet marked as 'Other layer' in the Cover_Categories column of the Species_Names sheet in the data file? If so, it was removed before any exploration and analysis. Choose another category here (Unvegetated category or Live vegetation) to keep it in the analyses.
4. Do you have enough data? Only plots with at least 3 separate years of data were included in statistical analyses.