

Draft ER390 Project Proposal

Visualizing HAT's GOE Monitoring Data

Wendy Anthony

2025/09/04

Contents

The Idea behind submitting this project proposal	2
Background	2
Project Partner	2
Project Goals and Overall Purpose	2
Project Objectives (SMART actionable targets)	2
Project Methods for HAT Data visualization	3
Proposed Deliverables	3
Workflow Graph	4
Proposed Timeline (Look at SMART Objectives)	5
Example Data Visualization Results	6
Data download and formatting	6
Load Site Data	6
Data Cleaning and Transformation	6
Data Visualization Example Results	7
Data Table	7
Map of Site Locations	8
Subregion Plots	9
Site Plots	10

The Idea behind submitting this project proposal

Nancy Shackelford (RNS Director) suggested that creating a data visualization tool for HAT GOE Monitoring for a suitable ER390 Final Project. I propose starting Fall 2025 term. While waiting for HAT to prepare to share data in October, I could continue the coding work I started in August (36+ hours), to create reproducible code snippets for use when HAT data is ready to be shared. I could also create R code to help with October data cleaning task. See *Proposed Gantt Timeline Chart*

Background

Data used for this proposal is from two tables in the report that HAT used for modelling their GOE Monitoring Project (Malloff & Shackelford, 2024). Comparing monitoring results from multiple restoration projects requires consistent measurements of vegetation composition (e.g. native and exotic species cover), which may help identify responses to stressors affecting already compromised and rare Garry Oak Ecosystems. Representing data in a visual manner can help make complex data more understandable than a tables of numbers, and may reveal patterns or trends resulting from different restoration management actions (Malloff & Shackelford, 2024). Data visualization can represent abstract information visually, by allowing for data exploration and analysis, modelling of restoration management alternatives, data-based story-telling, and communicating potential insights, to help understand and make restoration management decisions. Using RMarkdown to generate reports helps create reproducible research methods (McKay, 2020).

° Malloff, J., & Shackelford, N. (2024). Feeling the Pulse: Monitoring methods and initial outcomes in oak meadow ecosystems. Restoration Futures Lab at the University of Victoria.

° McKay, S. K. (2020). Data visualization for ecological analysis and restoration. US Army Corp of Engineers:Ecosystem Management and Restoration Research Program (EMRRP) Webinar Series. <https://cw-environment.erdc.dren.mil/webinars/20Feb5-EcoMod-DataViz.pdf>

Project Partner

I spoke to Vanessa Brownlee, Habitat Restoration Coordinator at Habitat Acquisition Trust (HAT) Social on 2025-08-27, who was very interested in my proposed project idea, and seemed willing to consider letting me start in Fall 2025, with data available later in October. I also later spoke briefly to Board Chair Andy MacKinnon, who seemed keen to know of my interest in doing my RNS Final Project with HAT and looks forward to learning more. I'm prepared to be guided by HAT's preferences, timing, and data protocols.

Project Goals and Overall Purpose

I propose to create a customized, automatic coding and reporting tool, using open-source R package libraries, for use by non-coders for data formatting and interactively visualizing data collected at multiple sites for HAT Garry Oak Ecosystem (GOE) Monitoring Program, including tutorial instructions, to help determine any patterns in collected data, and provide an interactive way to communication information about the data. This project proposal was created using code in RMarkdown, to showcase types of visualizations tools and reports that I propose to offer HAT, after adapting code to reflect HAT's data.

Project Objectives (SMART actionable targets)

- See *WorkFlowchart, Figure 1*, and *Gantt Chart Proposed Timeline, Figure 2*
- **Analyze collected data** to understand underlying structure, ensure use of consistent naming protocols for variables, and format data to be usable for use with data visualization code.
- **Write data visualization code**, using RStudio, RMarkdown, and RShiny, to create appropriate visualizations to enable a better understanding of the collected data.
- **Customize data tools** to reflect consistent styles using HAT's colours, logos, and plotting themes.

- **Create RMarkdown documents**, with reproducible code, for non-coders, to automatically download and format data, create data visualizations, and produce customized reports.
- **Create scaleable data visualization tools** that can have more sites added, with different measuring and sensitivity criteria (e.g. First Nations GOE sites), and multi-year data.
- **Track individual site differences** using variables and other metrics or indices that HAT measures.
- **Create custom R functions** to read, analyze, and plot data, as basis for HAT R package.
- **Create tools for multi-season use** to ensure consistent use of data variable names, and to incorporate any lessons learned in the process of working with it data in year one.
- **Create an interactive, browser-based Shiny App** User-chosen, custom-make data visualization plots, maps, and statistical analysis; select location(s) and variables; download results.
- **Recognize sensitive and confidential nature of any First Nations data** by creating separate data visualization files, code, and reports appropriate to individual land project goals.

Project Methods for HAT Data visualization

- See *WorkFlowchart, Figure 1*, and *Gantt Chart Proposed Timeline, Figure 2*
- Continue research for Shiny reactive code to make interactive data choices (e.g. drop-down lists), respond to choices, filter results. Interactive choices will eventually react to variable selections to make custom-made visualizations and analysis with only a few clicks
- Sign any HAT data confidentiality agreements as may be required
- Consult with HAT restoration monitor team to explore and understand the nature of the collected data, data format (forms, spreadsheets), the variables that are being measured, any indices to be calculated using the collected data, what questions they would like to explore with the data
- Create a Project folder, with subfolders for R scripts, raw data, processed data, images, reports, meta data (units, descriptions, etc); create file and variable naming protocols
- Analyze and clean data, check for errors, inconsistencies; calculate indices values from raw data
- Create data visualization plot charts that appropriately reflect nature of variables being measured
- Documents my methods, successes, and failures in the process of the project
- Continued research to find appropriate R coding methods, and to contribute to writing my final paper

Proposed Deliverables

See *WorkFlowchart, Figure 1*, and *Proposed Timeline, Figure 2*; All deliverables created using RStudio

Data Visualizations: Including data tables, plots, interactive maps; for individual sites, comparing sites, or confidentially for First Nations; See *Example Visualization* in next section

RMarkdown Data Processing Document: Data downloading, data cleaning and preparation; create data visualizations; Create reports in HTML, pdf, or Word with plot output images, or code snippets

Customizations: Create customized colour palette for Plot themes, App, Reports; add HAT logo

Shiny App Interactive Data Viz Tool: Data processed with RMarkdown document; choose data variables to visualize in tables, plot, maps; <https://goe-interact.shinyapps.io/Test-Shiny-GOE-Monitor/>

Tutorials for using Data visualization Tools: customized, easy to understand

Create Report Templates: Customized fpr individual site, total project reports, or as requested by HAT

Best practises for data files: Ensure consistency in naming files, variables, file folder structure, etc.

Create Customized Functions for HAT: Functions could form core of a HAT-specific R Package library

Presentation: Completed project presented to community partner HAT; including slide show, live demonstrations of Data Visualization Tools, and final copy of my written report when approved to be published.

Written Report: Due 2026-08-15, document project; publish in Ecorestoration: RNS Technical Series

Workflow Graph

- See *Objectives, Methods, and Deliverables*
- The workflow shows how the raw data, from HAT monitoring measurements, would be transformed and included in code to produce a variety of deliverable products.
- This graph is made with code from DiagrammeR, using a colour-blind friendly colour palette.

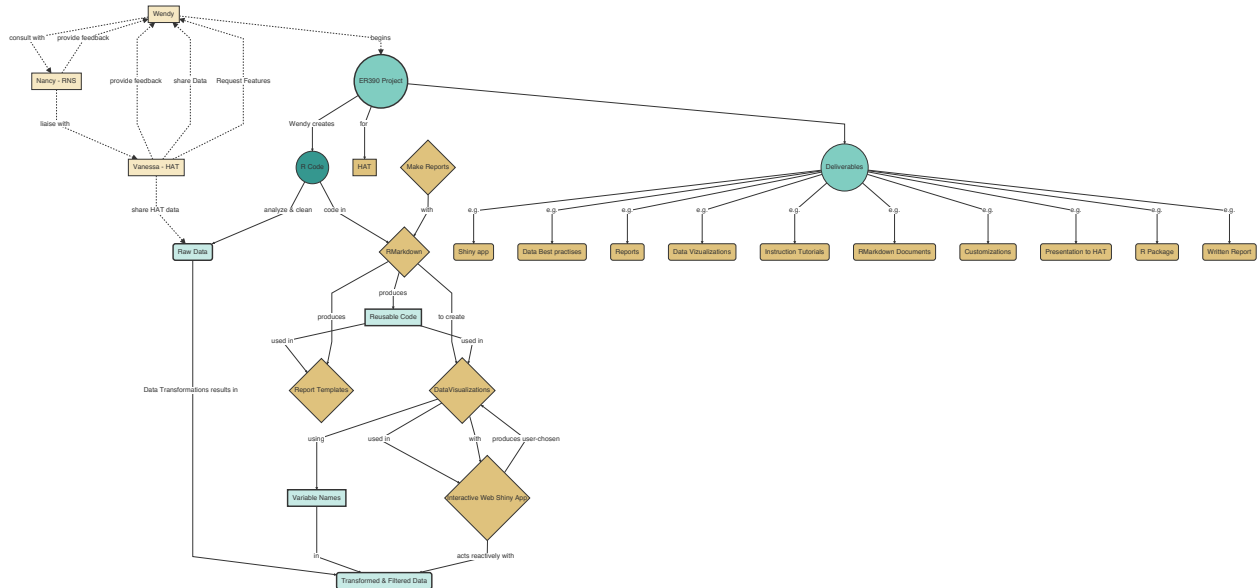


Figure 1: Project method workflow diagram for ER390 data visualization for HAT

Proposed Timeline (Look at SMART Objectives)

- red coloured items are considered critical
- grey items are completed
- Note: This chart was coded with DiagrammeR using an imported csv file of data
- See Objectives, Methods, and Deliverables

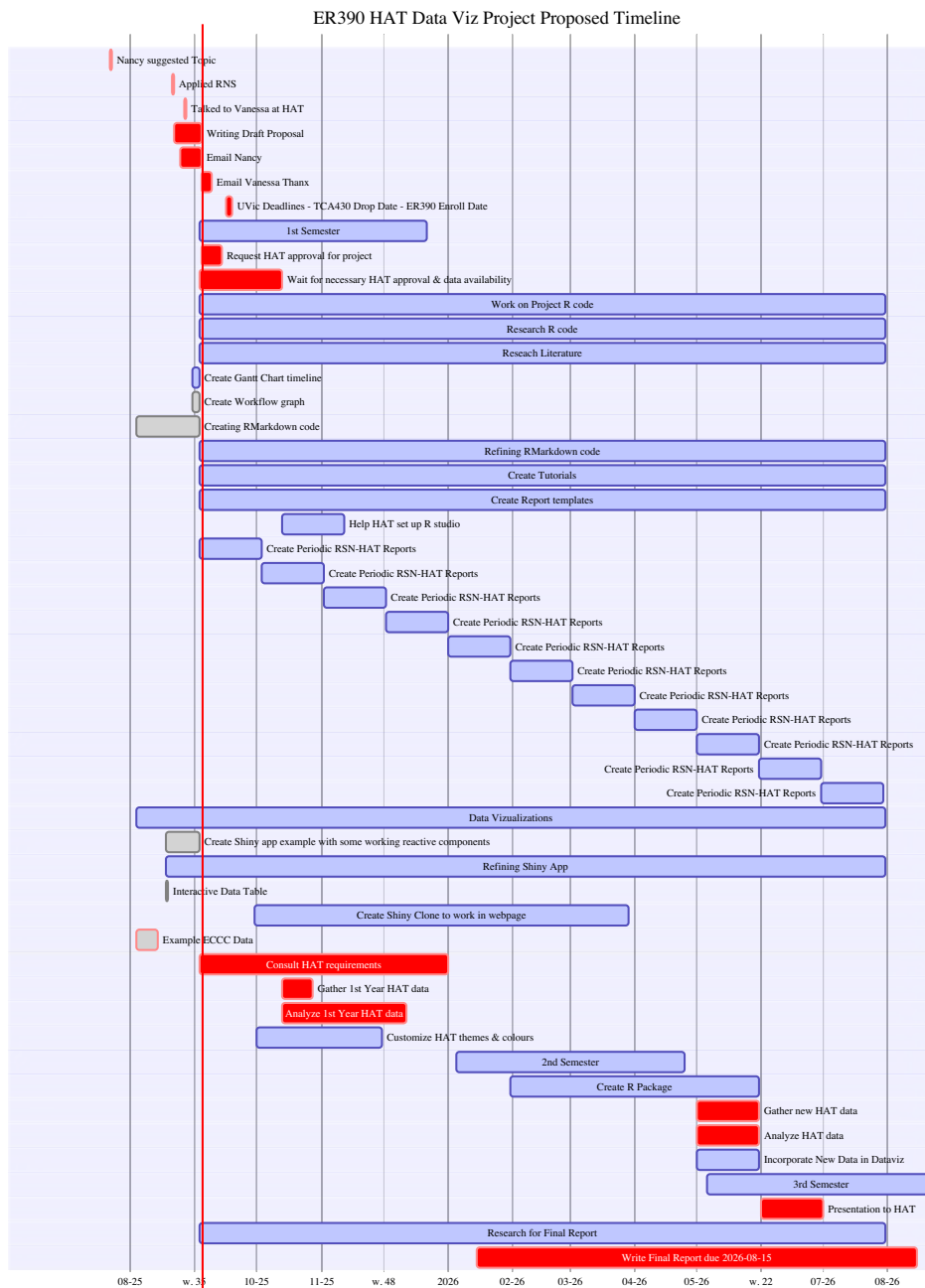


Figure 2: Gantt Proposed ER390 Timeline

Example Data Visualization Results

Data download and formatting

- These example visualizations use GOE monitoring data from report by Malloff & Shackelford (2024)
- Data (in excel or csv format) is uploaded
- Data is formatted, saved, and managed to be useful for creating visualizations
- The code has been hidden for the HTML report, highlighting only the resulting plots, though is visible in PDF document. The R code has been documented using hidden comments.

Load Site Data

Click open to see R code for Loading Data

```
# Load data and create 2 data frames
# Load Site Data
geo_eccc <- read.csv("data/SiteData-TEST-GOE-Monitor.csv", header = TRUE, sep = ",")
# Load Site Details
Sites_eccc <- read.csv("data/SiteDetails-TEST-GOE-Monitor.csv", header = TRUE, sep = ",")
```

Data Cleaning and Transformation

Click open to see R code for Data Cleaning and Transformation

```
# create new column with proportions changed to percentage as a whole number
Percentage_ns <- geo_eccc$Proportion_of_native_species * 100

# add-new column for Percentages
geo_eccc <- cbind(geo_eccc, Percentage_ns)

# arrange order of site names
Sites_eccc_1 <- Sites_eccc %>% arrange(Sites_eccc$Site)
geo_eccc_1 <- geo_eccc %>% arrange(geo_eccc$Site)

# bind 2 dataframes together
# this requires Sites rows from both data frames to be in alphabetical order
## to assign the data rows correctly
geo_eccc_sites <- cbind(Sites_eccc_1, geo_eccc_1)

# Remove duplicate columns resulting from binding 2 data frames
geo_eccc_sites_1 <- geo_eccc_sites[c(-9,-10)]

# Rename columns to remove dot-format
names(geo_eccc_sites_1)[names(geo_eccc_sites_1) == "Land.Manager"] <- "Land_Manager"
names(geo_eccc_sites_1)[names(geo_eccc_sites_1) == "Main.Restoration.Type"] <- "Main_Restoration_Type"
names(geo_eccc_sites_1)[names(geo_eccc_sites_1) == "Restoration.Intensity"] <- "Restoration_Intensity"
names(geo_eccc_sites_1)[names(geo_eccc_sites_1) == "Area.ha"] <- "Area_ha"

# write csv file with combined data
## to create new data file which is used for creating the visualizations
write.csv(geo_eccc_sites_1, "data/geo_eccc_all_site_data.csv", row.names = FALSE)
```

Data Visualization Example Results

Data Table

- using data from Malloff & Shackelford (2024)

Full Data Table

- The full table is too wide for use in pdf because current variable names are too long.

Subregion Site Land_Manager Main_Restoration_Type Restoration_Intensity Area_ha Lat Lng Proportion_of_native_species Cultural_species_richness Exotic_species Trampling Herbivory Composite_Index Year Percentage_ns

Table Head of Site Information

Subregion	Site	Land_Manager	Main_Restoration_Type	Restoration_Intensity	Area_ha	Lat	Lng
Gulf Islands	Anniversary Island	GINPR	herbivore reduction	high	4.39	48.82292	-123.1823
Gulf Islands	AVNR	Saltspring conservancy	herbivore reduction	minimal	20.54	48.80351	-123.4425
Saanich Peninsula	Bear Hill Park	CRD	invasive removal	low	3.80	48.54639	-123.4078
Gulf Islands	Brackman Island	GINPR	invasive removal	high	4.41	48.71897	-123.3864
Saanich Peninsula	Camas Hill	HAT	invasive removal	high	10.10	48.40173	-123.5975
Gulf Islands	Crows Nest	Trinity Western University	herbivore reduction	high	15.34	48.78237	-123.4612

Table Head of Site Data

Subregion	Site	Proportion_of_native_species	Cultural_species_richness	Exotic_species	Trampling	Herbivory	Composite
Gulf Islands	Anniversary Island	0.94	2.71	0.17	5.00	1.14	
Gulf Islands	AVNR	0.55	1.30	57.19	4.70	36.50	
Saanich Peninsula	Bear Hill Park	0.62	1.86	24.46	7.57	9.50	
Gulf Islands	Brackman Island	0.77	2.80	11.17	5.40	0.00	
Saanich Peninsula	Camas Hill	0.61	2.43	27.19	7.71	4.71	
Gulf Islands	Crows Nest	0.66	1.07	31.51	1.64	26.71	

Figures 3,4,5: Example Data Tables

Map of Site Locations

Leaflet Map with Circle Markers

- The Leaflet Map is created using R code in RMarkdown.
- Data is summarized in map marker popups with summary of site data

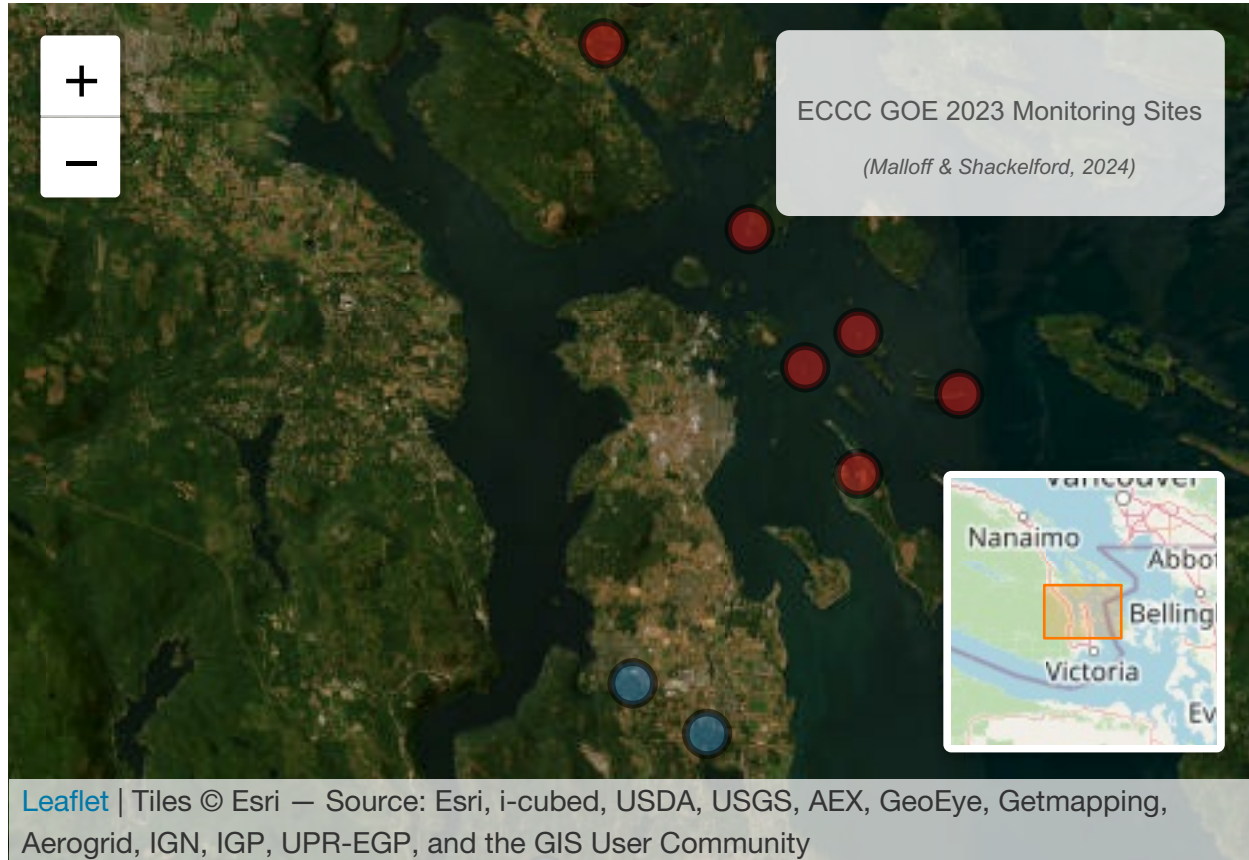


Figure 6: Interactive Leaflet Map with Example Sites Data

Subregion Plots

- These plot compare monitoring variable measurement results between two Subregions.

Violin/Bar Plot: Compare Subregions Exotic and Native Species

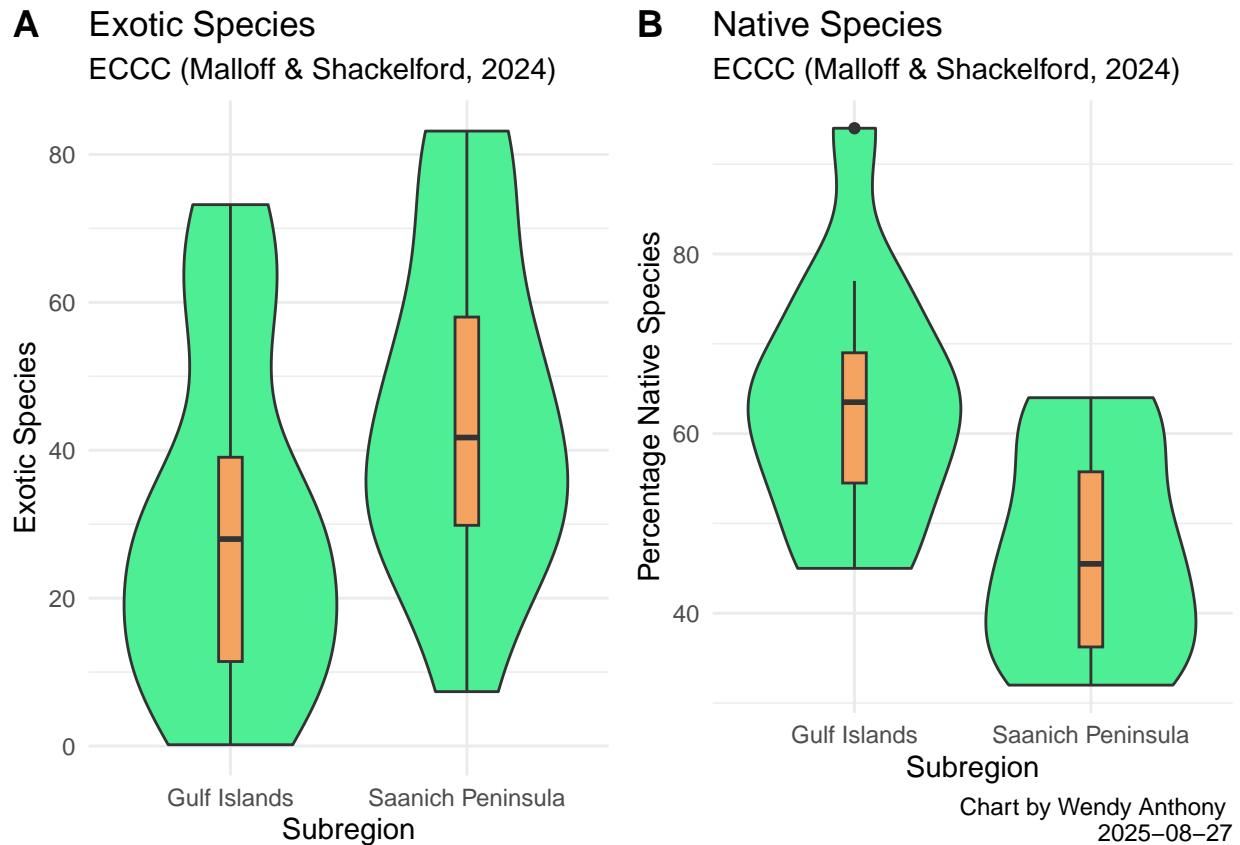


Figure 7: Violin / bar plot comparing variables of 2 Subregions

Site Plots

Point Plot: Compare Individual Sites by Restoration Intensity

- Sites are ordered by the level of restoration intensity, ranging from high to minimal
- This plotly-style chart becomes interactive by clicking data points for view a tool-tip of data results.

Plot Site Restoration Intensity

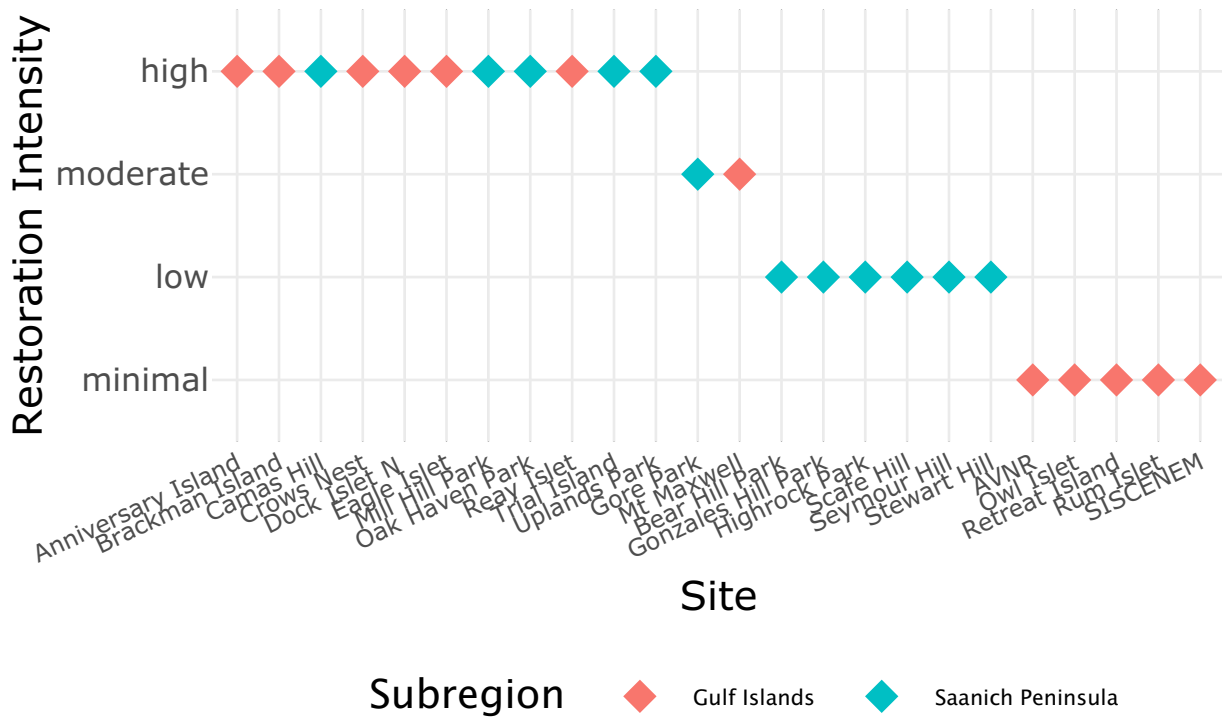


Figure 8: Point Plot of Restoration Intensity of Example Sites