SEACAR code review

Marcus W. Beck, [mbeck@tbep.org](mailto:mbeck@tbep.org)

2021-11-05 17:01:12

Table of Contents

View as [HTML document](https://FloridaSEACAR.github.io/SEACAR/review.html)

# 1 Document overview

SEACAR is a collaborative process which involves local, state and federal natural resource managers, data providers, researchers and partners to identify and assess ecological indicators and to develop a decision support tool to better understand the status of aquatic resources throughout the Office of Resilience and Coastal Protection managed areas. SEACAR staff are working with the University of South Florida (USF) Water Institute to create an online platform for data analysis and reporting to support science-based decisions of coastal resources in Florida. Analysis routines have been created to import, format, and summarize status and trends for various metrics of coastal condition reported by habitat, indicators, multiple parameters per indicator, and across 47 management areas. Primary habitats include water column, submerged aquatic vegetation, coral reefs, oyster reefs, and coastal wetlands.

To date, several scripts written using the R statistical programming language have been produced. This document is a review of these analysis scripts, as currently delivered in two separate GitHub repositories (available at <https://github.com/FloridaSEACAR>). This review includes a description of the purpose of each script to ensure the outputs are as expected based on preliminary goals established for the project. In general, status and trends of key indicators is to be quantified using conventional, non-parametric methods following the seasonal Kendall tau test that assesses direction and magnitude of a change within a specified period of time. Minimum requirements for the trend tests for each indicator include at least five years of data with at least two months in common across at least two consecutive years for each managed area. In addition, linear mixed-effects models have been used to assess trends in seagrass coverage.

This review is delivered as a bulleted outline for each script describing the objective, required inputs, processing steps, and output (e.g., graphic, summary table). The tests run or summaries created are described. The extent to which each indicator fulfills the requirements for the applicable trend test is also assessed, e.g., by describing available years of data or if data are insufficient. Each script is also reviewed for coding best practices to identify areas that can be improved or shortened for efficiency and to identify potential mistakes or errors in calculation. Specific lines are referenced as needed to identify areas in need of improvement or further review. Outputs, such as graphics or tabular summaries, are also reviewed to identify potential improvements in future versions of the code.

# 2 Comments by habitat type

## 2.1 Water column

### Scripts

|  |  |  |
| --- | --- | --- |
| Index | File | Link |
| 1 | 2021.08.21\_FINAL\_SEACAR Water\_Nutrients\_Phosphorus\_FINAL.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR%20Water_Nutrients_Phosphorus_FINAL.R) |
| 2 | 2021.08.21\_FINAL\_SEACAR\_Water Clarity\_CDOM\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR_Water%20Clarity_CDOM_Final.R) |
| 3 | 2021.08.21\_FINAL\_SEACAR\_Water Clarity\_chla\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR_Water%20Clarity_chla_Final.R) |
| 4 | 2021.08.21\_FINAL\_SEACAR\_Water Clarity\_Turbidity\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR_Water%20Clarity_Turbidity_Final.R) |
| 5 | 2021.08.21\_FINAL\_SEACAR\_Water Nutrients\_Nitrogen\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR_Water%20Nutrients_Nitrogen_Final.R) |
| 6 | 2021.08.22\_FINAL SEACAR\_WaterQuality\_DOsaturation\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.22_FINAL%20SEACAR_WaterQuality_DOsaturation_Final.R) |
| 7 | 2021.08.22\_FINAL\_SEACAR\_WaterQuality\_DOmgL\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.22_FINAL_SEACAR_WaterQuality_DOmgL_Final.R) |
| 8 | 2021.08.22\_FINAL\_SEACAR\_WaterQuality\_pH\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.22_FINAL_SEACAR_WaterQuality_pH_Final.R) |
| 9 | 2021.08.22\_FINAL\_SEACAR\_WaterQuality\_Salinity.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.22_FINAL_SEACAR_WaterQuality_Salinity.R) |
| 10 | 2021.08.22\_FINAL\_SEACAR\_WATERQUALITY\_Temperature\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.22_FINAL_SEACAR_WATERQUALITY_Temperature_Final.R) |
| 11 | FINAL 2021.09.19\_WaterNekton\_SppRichnessbyMA.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL%202021.09.19_WaterNekton_SppRichnessbyMA.R) |

### Summaries

1. 2021.08.21\_FINAL\_SEACAR Water\_Nutrients\_Phosphorus\_FINAL.R
   * Objective: Import and format phosphorus data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Nutrients - Nitrogen and Phosphorus-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 52: load libraries and import phosphorus and regions datasets
     + 53 - 187: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data). NA values are removed for water quality measurements and dates.
     + 188 - 206: boxplot summaries of surface phosphorus by year, year/month for complete dataset
     + 208 - 222: boxplot summaries of surface phosphorus by year/month, separate for managed areas
     + 223 - 239: boxplot summaries of surface phosphorus by year, separate for managed areas
     + 240 - 264: average phosphorus by year, month, managed area to prep for trend test
     + 265 - 322: identify number of unique years in each dataset, remove those with < 10
     + 323 - 349: plots of average year, month phosphorus, separate for managed areas
     + 350 - 368: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 369 - 440: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Nutrients\_P\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Nutrients_P_MA_surf_MonthlyBoxplots.pdf), [Nutrients\_P\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Nutrients_P_MA_surf_AnnualBoxplots.pdf), [2Nutrients\_P\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Nutrients_P_MA_surf_plots.pdf), [Nutrients\_P\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Nutrients_P_MA_surf_KToutput.csv), Nutrients\_P\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + Line 41: object regions called before import
     + Line 202 (and others): boxplot summaries should be on log-scale
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Values at MDL are not removed, these need to be treated separately before running trend tests, i.e., estimate averages using appropriate statistics that account for left-censored data.
2. 2021.08.21\_FINAL\_SEACAR\_Water Clarity\_CDOM\_Final.R
   * Objective: Import and format CDOM data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: CDOM and FDOM-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 49: load libraries and import CDOM and regions datasets
     + 50 - 186: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data). NA values are removed for water quality measurements and dates.
     + 187 - 205: boxplot summaries of surface CDOM by year, year/month for complete dataset
     + 206 - 216: boxplot summaries of surface CDOM by year/month, separate for managed areas
     + 217 - 238: boxplot summaries of surface CDOM by year, separate for managed areas
     + 239 - 262: average CDOM by year, month, managed area to prep for trend test
     + 263 - 320: identify number of unique years in each dataset, remove those with < 10
     + 321 - 347: plots of average year, month CDOM, separate for managed areas
     + 350 - 366: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 367 - 438: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Clarity\_CDOM\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_CDOM_MA_surf_MonthlyBoxplots.pdf), [Clarity\_CDOM\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_CDOM_MA_surf_AnnualBoxplots.pdf), [2Clarity\_CDOM\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Clarity_CDOM_MA_surf_plots.pdf), [Clarity\_CDOM\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_CDOM_MA_surf_KToutput.csv), Clarity\_CDOM\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + Line 201 (and others): boxplot summaries should be on log-scale
     + Line 269: object name has a reserved character (-) and cannot be created
     + Line 359: object name has a reserved character (-) and cannot be created
     + Line 371: object name has a reserved character (-) and cannot be created
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Values at MDL are not removed, these need to be treated separately before running trend tests, i.e., estimate averages using appropriate statistics that account for left-censored data.
3. 2021.08.21\_FINAL\_SEACAR\_Water Clarity\_chla\_Final.R
   * Objective: Import and format chlorophyll data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Chlorophyll-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 49: load libraries and import chlorophyll and regions datasets
     + 50 - 185: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data). NA values are removed for water quality measurements and dates.
     + 186 - 205: boxplot summaries of surface chlorophyll by year, year/month for complete dataset
     + 206 - 220: boxplot summaries of surface chlorophyll by year/month, separate for managed areas
     + 221 - 237: boxplot summaries of surface chlorophyll by year, separate for managed areas
     + 238 - 261: average chlorophyll by year, month, managed area to prep for trend test
     + 262 - 319: identify number of unique years in each dataset, remove those with < 10
     + 320 - 346: plots of average year, month chlorophyll, separate for managed areas
     + 347 - 365: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 366 - 437: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Clarity\_chla\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_chla_MA_surf_MonthlyBoxplots.pdf), [Clarity\_chla\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_chla_MA_surf_AnnualBoxplots.pdf), [2Clarity\_chla\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Clarity_chla_MA_surf_plots.pdf), [Clarity\_chla\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_chla_MA_surf_KToutput.csv), Clarity\_chla\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + Line 205 (and others): boxplot summaries should be on log-scale
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Values at MDL are not removed, these need to be treated separately before running trend tests, i.e., estimate averages using appropriate statistics that account for left-censored data.
4. 2021.08.21\_FINAL\_SEACAR\_Water Clarity\_Turbidity\_Final.R
   * Objective: Import and format turbidity data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Turbidity and TSS-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 50: load libraries and import turbidity and regions datasets
     + 51 - 186: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data). NA values are removed for water quality measurements and dates.
     + 187 - 205: boxplot summaries of surface turbidity by year, year/month for complete dataset
     + 206 - 221: boxplot summaries of surface turbidity by year/month, separate for managed areas
     + 222 - 238: boxplot summaries of surface turbidity by year, separate for managed areas
     + 239 - 262: average turbidity by year, month, managed area to prep for trend test
     + 263 - 320: identify number of unique years in each dataset, remove those with < 10
     + 321 - 347: plots of average year, month turbidity, separate for managed areas
     + 348 - 366: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 367 - 438: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Clarity\_Turbidity\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_Turbidity_MA_surf_MonthlyBoxplots.pdf), [Clarity\_Turbidity\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_Turbidity_MA_surf_AnnualBoxplots.pdf), [2Clarity\_Turbidity\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Clarity_Turbidity_MA_surf_plots.pdf), [Clarity\_Turbidity\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Clarity_Turbidity_MA_surf_KToutput.csv), Clarity\_Turbidity\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + Line 201 (and others): boxplot summaries should be on log-scale
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Values at MDL are not removed, these need to be treated separately before running trend tests, i.e., estimate averages using appropriate statistics that account for left-censored data.
5. 2021.08.21\_FINAL\_SEACAR\_Water Nutrients\_Nitrogen\_Final.R
   * Objective: Import and format nitrogen data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Nitrogen-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 48: load libraries and import nitrogen and regions datasets
     + 49 - 219: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data). NA values are removed for water quality measurements and dates. Creates TN as sum of TKN and NOx. Values less than zero are flagged as 999999999, which is different from the other water quality variables.
     + 220 - 239: boxplot summaries of surface nitrogen by year, year/month for complete dataset
     + 240 - 255: boxplot summaries of surface nitrogen by year/month, separate for managed areas
     + 256 - 272: boxplot summaries of surface nitrogen by year, separate for managed areas
     + 273 - 296: average nitrogen by year, month, managed area to prep for trend test
     + 297 - 354: identify number of unique years in each dataset, remove those with < 10
     + 355 - 381: plots of average year, month nitrogen, separate for managed areas
     + 382 - 400: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 401 - 472: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Nutrients\_N\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Nutrients_N_MA_surf_MonthlyBoxplots.pdf), [Nutrients\_N\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Nutrients_N_MA_surf_AnnualBoxplots.pdf), [2Nutrients\_N\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Nutrients_N_MA_surf_plots.pdf), [Nutrients\_N\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Nutrients_N_MA_surf_KToutput.csv), [Nutrients\_N\_MA\_surf\_resultssummarystats2.csv](https://FloridaSEACAR.github.io/SEACAR/output/Nutrients_N_MA_surf_resultssummarystats2.csv)
   * Potential issues:
     + Line 115: Date conversion is incorrect, creates all NA preventing downstream analysis
     + Line 235 (and others): boxplot summaries should be on log-scale
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Values at MDL are not removed, these need to be treated separately before running trend tests, i.e., estimate averages using appropriate statistics that account for left-censored data.
     + Many of the plots of the observed or averaged data are suspect, with large jumps in values between years, possibly related to changing lab methods or incorrect sum to get TN.
6. 2021.08.22\_FINAL SEACAR\_WaterQuality\_DOsaturation\_Final.R
   * Objective: Import and format dissolved oxygen saturation data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Dissolved Oxygen-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 49: load libraries and import dissolved oxygen saturation and regions datasets
     + 50 - 193: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data)
     + 194 - 212: boxplot summaries of surface dissolved oxygen saturation by year, year/month for complete dataset
     + 213 - 228: boxplot summaries of surface dissolved oxygen saturation by year/month, separate for managed areas
     + 229 - 245: boxplot summaries of surface dissolved oxygen saturation by year, separate for managed areas
     + 246 - 269: average dissolved oxygen saturation by year, month, managed area to prep for trend test
     + 270 - 327: identify number of unique years in each dataset, remove those with < 10
     + 328 - 354: plots of average year, month dissolved oxygen saturation, separate for managed areas
     + 355 - 373: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 374 - 445: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Quality\_DOsat\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_DOsat_MA_surf_MonthlyBoxplots.pdf), [Quality\_DOsat\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_DOsat_MA_surf_AnnualBoxplots.pdf), [2Quality\_DOsat\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Quality_DOsat_MA_surf_plots.pdf), [Quality\_DOsat\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Quality_DOsat_MA_surf_KToutput.csv), Quality\_DOsat\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Many of the plots of the observed or averaged data are suspect, with large jumps in values between years.
7. 2021.08.22\_FINAL\_SEACAR\_WaterQuality\_DOmgL\_Final.R
   * Objective: Import and format dissolved oxygen concentration data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Dissolved Oxygen-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 49: load libraries and import dissolved oxygen concentration and regions datasets
     + 50 - 193: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data)
     + 194 - 212: boxplot summaries of surface dissolved oxygen concentration by year, year/month for complete dataset
     + 213 - 228: boxplot summaries of surface dissolved oxygen concentration by year/month, separate for managed areas
     + 229 - 245: boxplot summaries of surface dissolved oxygen concentration by year, separate for managed areas
     + 246 - 269: average dissolved oxygen concentration by year, month, managed area to prep for trend test
     + 270 - 327: identify number of unique years in each dataset, remove those with < 10
     + 328 - 354: plots of average year, month dissolved oxygen concentration, separate for managed areas
     + 355 - 373: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 374 - 445: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Quality\_DOmgL\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_DOmgL_MA_surf_MonthlyBoxplots.pdf), [Quality\_DOmgL\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_DOmgL_MA_surf_AnnualBoxplots.pdf), [2Quality\_DOmgL\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Quality_DOmgL_MA_surf_plots.pdf), [Quality\_DOmgL\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Quality_DOmgL_MA_surf_KToutput.csv), Quality\_DOmgL\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Many of the plots of the observed or averaged data are suspect, with large jumps in values between years.
8. 2021.08.22\_FINAL\_SEACAR\_WaterQuality\_pH\_Final.R
   * Objective: Import and format pH data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Water Temperature and pH-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 49: load libraries and import pH and regions datasets
     + 50 - 193: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data)
     + 194 - 212: boxplot summaries of surface pH by year, year/month for complete dataset
     + 213 - 228: boxplot summaries of surface pH by year/month, separate for managed areas
     + 229 - 245: boxplot summaries of surface pH by year, separate for managed areas
     + 246 - 269: average pH by year, month, managed area to prep for trend test
     + 270 - 327: identify number of unique years in each dataset, remove those with < 10
     + 328 - 354: plots of average year, month pH, separate for managed areas
     + 355 - 373: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 374 - 445: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Quality\_pH\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_pH_MA_surf_MonthlyBoxplots.pdf), [Quality\_pH\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_pH_MA_surf_AnnualBoxplots.pdf), [2Quality\_pH\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Quality_pH_MA_surf_plots.pdf), [Quality\_pH\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Quality_pH_MA_surf_KToutput.csv), Quality\_pH\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + The pH file is 1Gb, not sure this is a huge issue but import time is longer than most files.
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Many of the plots of the observed or averaged data are suspect, with large jumps in values between years.
9. 2021.08.22\_FINAL\_SEACAR\_WaterQuality\_Salinity.R
   * Objective: Import and format salinity data, create summary boxplots, evaluate trends with seasonal Kendall tests.
   * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
   * File inputs: Salinity and Specific Conductivity-2021-Jul-26.csv, SEACARMARegionList.csv
   * Steps by line number:
     + 1 - 49: load libraries and import salinity and regions datasets
     + 50 - 193: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data)
     + 194 - 212: boxplot summaries of surface salinity by year, year/month for complete dataset
     + 213 - 228: boxplot summaries of surface salinity by year/month, separate for managed areas
     + 229 - 245: boxplot summaries of surface salinity by year, separate for managed areas
     + 246 - 269: average salinity by year, month, managed area to prep for trend test
     + 270 - 327: identify number of unique years in each dataset, remove those with < 10
     + 328 - 354: plots of average year, month salinity, separate for managed areas
     + 355 - 373: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
     + 374 - 445: format all trend tests in a simple table, combine with region names, save output as csv
   * File outputs: [Quality\_Salinity\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_Salinity_MA_surf_MonthlyBoxplots.pdf), [Quality\_Salinity\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_Salinity_MA_surf_AnnualBoxplots.pdf), [2Quality\_Salinity\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Quality_Salinity_MA_surf_plots.pdf), [Quality\_Salinity\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Quality_Salinity_MA_surf_KToutput.csv), Quality\_Salinity\_MA\_surf\_resultssummarystats2.csv
   * Potential issues:
     + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
     + Many of the plots of the observed or averaged data are suspect, with large jumps in values between years.
10. 2021.08.22\_FINAL\_SEACAR\_WATERQUALITY\_Temperature\_Final.R
    * Objective: Import and format temperature data, create summary boxplots, evaluate trends with seasonal Kendall tests.
    * Packages: dplyr, EnvStats, gt, tidymodels, tidyr, tidyverse
    * File inputs: Water Temperature and pH-2021-Jul-26.csv, SEACARMARegionList.csv
    * Steps by line number:
      + 1 - 49: load libraries and import temperature and regions datasets
      + 50 - 193: Format data for analysis, including removal of NA values, conversion of date column to date objects, and removing entries with flagged QA codes (from WIN and NERRS data)
      + 194 - 212: boxplot summaries of surface temperature by year, year/month for complete dataset
      + 213 - 228: boxplot summaries of surface temperature by year/month, separate for managed areas
      + 229 - 245: boxplot summaries of surface temperature by year, separate for managed areas
      + 246 - 269: average temperature by year, month, managed area to prep for trend test
      + 270 - 327: identify number of unique years in each dataset, remove those with < 10
      + 328 - 354: plots of average year, month temperature, separate for managed areas
      + 355 - 373: run trend tests, first on a “test” site then for all, results saved as raw output from the trend test
      + 374 - 445: format all trend tests in a simple table, combine with region names, save output as csv
    * File outputs: [Quality\_Temperature\_MA\_surf\_MonthlyBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_Temperature_MA_surf_MonthlyBoxplots.pdf), [Quality\_Temperature\_MA\_surf\_AnnualBoxplots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Quality_Temperature_MA_surf_AnnualBoxplots.pdf), [2Quality\_Temperature\_MA\_surf\_plots.pdf](https://FloridaSEACAR.github.io/SEACAR/output/2Quality_Temperature_MA_surf_plots.pdf), [Quality\_Temperature\_MA\_surf\_KToutput.csv](https://FloridaSEACAR.github.io/SEACAR/output/Quality_Temperature_MA_surf_KToutput.csv), Quality\_Temperature\_MA\_surf\_resultssummarystats2.csv
    * Potential issues:
      + The temperature file is 1Gb, not sure this is a huge issue but import time is longer than most files.
      + NA values were also removed for dates that were incomplete. This may possibly be corrected through additional formatting but is not done here.
      + Many of the plots of the observed or averaged data are suspect, with large jumps in values between years.
11. FINAL 2021.09.19\_WaterNekton\_SppRichnessbyMA.R
    * Objective: Import and format nekton data, calculate richness, assess trends with linear mixed effects models.
    * Packages: dplyr, gt, nlme, RColorBrewer, tidyverse
    * File inputs: Nekton\_NE-2021-Jul-26.csv
    * Steps by line number:
      + 1 - 19: load libraries and import nekton data
      + 20 - 30: combine genus, species columns as one, remove any rows with missing genus or management area names
      + 31 - 52: calculate richness metrics management area, region, year
      + 53 - 69: identify number of unique years in each dataset, remove those with < 5
      + 70 - 86: create linear mixed models evaluating richness vs year, using a random region effect
      + 87 - 113: create plot of richness vs year by management area using model output
      + 114 - 141: tabular summary of richness vs year by management area
      + 142 - 166: tabular summary of richness vs year by management area, similar info as previous table
      + 167 - 200: join results from the previous two tables, save output as csv
    * File outputs: [Nekton\_SppRichnessbyMAmodel.csv](https://FloridaSEACAR.github.io/SEACAR/output/Nekton_SppRichnessbyMAmodel.csv), [Nekton\_SppRichnessbyMAmodelPlot.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Nekton_SppRichnessbyMAmodelPlot.pdf), Nekton\_SppRichnessbyMAmodelTable.pdf.csv
    * Potential issues:
      + Line 38: NA values are incorrectly identified, should do this with a regular expression search.
      + If the goal is to evaluate diversity, this was not done. Only richness was evaluated.
      + There are redundant models created for the plots and tables. These should only be created once and recycled as needed.
      + There are some outlier values for one of the managed areas in the species richness vs year plot that should be verified.

## 2.2 Coral reef

### Scripts

|  |  |  |
| --- | --- | --- |
| Index | File | Link |
| 1 | 2021.09.19\_CoralNektonRichness\_FINAL.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.09.19_CoralNektonRichness_FINAL.R) |
| 2 | FINAL\_2021.09.18\_DryTort Region\_Any Corals.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_DryTort%20Region_Any%20Corals.R) |
| 3 | FINAL\_2021.09.18\_DryTortugas Region\_Genus.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_DryTortugas%20Region_Genus.R) |
| 4 | FINAL\_2021.09.18\_DryTortugas Region\_Group1.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_DryTortugas%20Region_Group1.R) |
| 5 | FINAL\_2021.09.18\_FLAKEYS\_AnyCorals.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_FLAKEYS_AnyCorals.R) |
| 6 | FINAL\_2021.09.18\_FLKeysRegion\_Group1.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_FLKeysRegion_Group1.R) |
| 7 | FINAL\_2021.09.18\_Keys Region\_Genus.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_Keys%20Region_Genus.R) |
| 8 | FINAL\_2021.09.18\_SE Region\_AnyCorals.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_SE%20Region_AnyCorals.R) |
| 9 | FINAL\_2021.09.18\_SE\_Genus.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_SE_Genus.R) |
| 10 | FINAL\_2021.09.18\_SERegion\_Group1.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL_2021.09.18_SERegion_Group1.R) |

### Summaries

1. 2021.09.19\_CoralNektonRichness\_FINAL.R
   * Objective: import nekton and coral data, combine, assess relationships between nekton richness and coral cover by managed area.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: CorSE.csv, CorFLKeys.csv, CorDryTortugas.csv, Nekton2.csv
   * Steps by line number:
     + 1 - 43: import coral data for each region (Dry Tortugas, Florida Keys, SE Florida), filter by managed area names, combine all into one object
     + 44 - 65: import nekton data, filter by grazers and reef dependent species
     + 66 - 90: estimate nekton richness by managed area, year, estimate coral percent cover by managed area, year, combine into one dataset
     + 91 - 103: create linear mixed effects model of nekton richness by mean percent coral cover, year as random effect, output as csv
     + 104 - 124: create linear mixed effects model of nekton richness by mean percent coral cover, year as random effect, separate models by managed area, output as csv
     + 125 - 149: create plots of nekton richness vs coral percent cover from model output, save as pdf
     + 150 - 236: format and summarize model output to create summary table of model fit, saved as csv
   * File outputs: [CoralNektonlmeSERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralNektonlmeSERegion.csv), [CoralNektonlmebyMA.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralNektonlmebyMA.csv), [NektonCoral\_SppRichnessbyMAmodelPlot.pdf](https://FloridaSEACAR.github.io/SEACAR/output/NektonCoral_SppRichnessbyMAmodelPlot.pdf), [NektonCoral\_SppRichnessbyMAmodelTable.csv](https://FloridaSEACAR.github.io/SEACAR/output/NektonCoral_SppRichnessbyMAmodelTable.csv)
   * Potential issues:
     + Requires data inputs created in files FINAL 2021.09.18\_SE Region\_AnyCorals.R, FINAL\_2021.09.18\_FLAKEYS\_AnyCorals.R, FINAL\_2021.09.18\_DryTort Region\_Any Corals.R. The Nekton2.csv file is created from Nekton\_SE-2021-Jul-26.csv.
2. FINAL\_2021.09.18\_DryTort Region\_Any Corals.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends for any corals across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - DRY TORT-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 20: load libraries and import percent cover data for Dry Tortugas
     + 21 - 79: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 80 - 101: create a summary table for percent cover statistics by year, region across all species, save output
     + 102 - 112: create linear model year vs percent cover, save model summary plots to pdf
     + 113 - 128: create boxplots of percent cover for any coral by year, save as pdf
     + 129 - 145: create linear mixed effects model of percent cover for any coral vs year, random effect for program location ID
     + 146 - 179: create linear mixed effects model of percent cover for any coral vs year, random effect for program location ID, separate for coral region, save output as csv
     + 180 - 202: plot model output as estimated linear trend across years by coral region, save as pdf
     + 203 - 301: format and summarize model output to create summary table of model fit, saved as csv
   * File outputs: [CoralSummaryStatisticsTable\_AnyCoral\_DRYTORTUGASRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_AnyCoral_DRYTORTUGASRegion.csv), [CoralSummaryStatisticsPlots\_AnyCoral\_DRYTORTUGASRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsPlots_AnyCoral_DRYTORTUGASRegion.pdf), [CoralBoxplots\_AnyCoral\_DRYTORTUGASRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_AnyCoral_DRYTORTUGASRegion.pdf), CorDryTortugas.csv, [Corallme\_AnyCoral\_DRYTORTUGASRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_AnyCoral_DRYTORTUGASRegion.csv), [CorallmePlot\_AnyCoral\_DRYTORTUGASRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_AnyCoral_DRYTORTUGASRegion.pdf), [Coral\_lmeResults\_AnyCoral\_DRYTORTUGASRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Coral_lmeResults_AnyCoral_DRYTORTUGASRegion.csv)
   * Potential issues:
     + 106, 109: Year is dependent variable in the linear model, it should be the independent variable.
     + Model results are separate for each coral region, yet the file includes only one region (Dry Tortugas). This does not create incorrect output, but the intention is unclear.
3. FINAL\_2021.09.18\_DryTortugas Region\_Genus.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends by genus across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - DRY TORT-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 18: load libraries and import percent cover data for Dry Tortugas
     + 19 - 83: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 84 - 103: create a summary table for percent cover statistics by year, genus, region, save output
     + 104 - 110: create linear model year vs percent cover, not separated by genus
     + 111 - 125: create boxplots of percent cover by genus, year, save as pdf
     + 126 - 161: remove genera with less than ten years of data, create linear mixed effects model of percent cover for all genera vs year, random effect for program location ID. No output created because no genus meets the minimum year requirement.
     + 162 - 184: create linear mixed effects model of percent cover for each genus vs year, random effect for program location ID, save output as csv. No output created because no genus meets the minimum year requirement.
     + 185 - 210: plot model output as estimated linear trend across years by genus, save as pdf. No output created because no genus meets the minimum year requirement.
     + 211 - 309: format and summarize model output to create summary table of model fit, saved as csv. No output created because no genus meets the minimum year requirement.
   * File outputs: [CoralSummaryStatisticsTable\_Genus\_DRYRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_Genus_DRYRegion.csv), [CoralBoxplots\_ByGenus\_DRYRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_ByGenus_DRYRegion.pdf), Corallme\_ByGenus\_DRYRegion.csv, [CorallmePlot\_ByGenus\_DRYRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_ByGenus_DRYRegion.pdf), Coral\_lmeResults\_ByGenus\_DRYRegion.csv
   * Potential issues:
     + Line 108: Year is dependent variable in the linear model, it should be the independent variable.
     + No output created because no genus meets the minimum year requirement.
4. FINAL\_2021.09.18\_DryTortugas Region\_Group1.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends by species group across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - DRY TORT-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 17: load libraries and import percent cover data for Dry Tortugas
     + 18 - 81: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 82 - 101: create a summary table for percent cover statistics by year, species group, region, save output
     + 102 - 107: create linear model year vs percent cover, not separated by species group
     + 108 - 122: create boxplots of percent cover by species group, year, save as pdf
     + 123 - 158: create linear mixed effects model of percent cover for each species group vs year, random effect for program location ID, save as csv
     + 159 - 181: plot model output as estimated linear trend across years by species group, save as pdf
     + 182 - 280: format and summarize model output to create summary table of model fit, saved as csv
   * File outputs: [CoralSummaryStatisticsTable\_Group1\_DryTortugasRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_Group1_DryTortugasRegion.csv), [CoralBoxplots\_ByGroup1\_DRYRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_ByGroup1_DRYRegion.pdf), [Corallme\_ByGroup1\_DRYRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_ByGroup1_DRYRegion.csv), [CorallmePlot\_ByGroup1\_DRYRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_ByGroup1_DRYRegion.pdf), [Coral\_lmeResults\_ByGroup1\_DryTortugasRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Coral_lmeResults_ByGroup1_DryTortugasRegion.csv)
   * Potential issues:
     + Line 106: Year is dependent variable in the linear model, it should be the independent variable.
5. FINAL\_2021.09.18\_FLAKEYS\_AnyCorals.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends for any corals across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - FLA KEYS-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 20: load libraries and import percent cover data for Florida Keys
     + 21 - 79: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 80 - 101: create a summary table for percent cover statistics by year, region across all species, save output
     + 102 - 112: create linear model year vs percent cover, save model summary plots to pdf
     + 113 - 127: create boxplots of percent cover for any coral by year, save as pdf
     + 128 - 144: create linear mixed effects model of percent cover for any coral vs year, random effect for program location ID
     + 145 - 178: create linear mixed effects model of percent cover for any coral vs year, random effect for program location ID, separate for coral region, save output as csv
     + 179 - 201: plot model output as estimated linear trend across years by coral region, save as pdf
     + 202 - 300: format and summarize model output to create summary table of model fit, saved as csv
   * File outputs: [CoralSummaryStatisticsTable\_AnyCoral\_FLKEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_AnyCoral_FLKEYSRegion.csv), [CoralSummaryStatisticsPlots\_AnyCoral\_FLKEYSRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsPlots_AnyCoral_FLKEYSRegion.pdf), [CorFLKeys.csv](https://FloridaSEACAR.github.io/SEACAR/output/CorFLKeys.csv), [CoralBoxplots\_AnyCoral\_FLKEYSRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_AnyCoral_FLKEYSRegion.pdf), [Corallme\_AnyCoral\_FLKEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_AnyCoral_FLKEYSRegion.csv), [CorallmePlot\_AnyCoral\_FLKEYSRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_AnyCoral_FLKEYSRegion.pdf), [Coral\_lmeResults\_AnyCoral\_FLKEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Coral_lmeResults_AnyCoral_FLKEYSRegion.csv)
   * Potential issues:
     + 106, 109: Year is dependent variable in the linear model, it should be the independent variable.
     + Model results are separate for each coral region, yet the file includes only one region (Florida Keys). This does not create incorrect output, but the intention is unclear.
6. FINAL\_2021.09.18\_FLKeysRegion\_Group1.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends by species group across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - FLA KEYS-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 17: load libraries and import percent cover data for Florida Keys
     + 18 - 82: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 83 - 102: create a summary table for percent cover statistics by year, species group, region, save output
     + 103 - 108: create linear model year vs percent cover, not separated by species group
     + 109 - 123: create boxplots of percent cover by species group, year, save as pdf
     + 124 - 159: create linear mixed effects model of percent cover for each species group vs year, random effect for program location ID, save as csv
     + 160 - 182: plot model output as estimated linear trend across years by species group, save as pdf
     + 183 - 281: format and summarize model output to create summary table of model fit, saved as csv
   * File outputs: [CoralSummaryStatisticsTable\_Group1\_KEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_Group1_KEYSRegion.csv), [CoralBoxplots\_ByGroup1\_KEYSRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_ByGroup1_KEYSRegion.pdf), [Corallme\_ByGroup1\_KEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_ByGroup1_KEYSRegion.csv), [CorallmePlot\_ByGroup1\_KEYSRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_ByGroup1_KEYSRegion.pdf), [Coral\_lmeResults\_ByGroup1\_KEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Coral_lmeResults_ByGroup1_KEYSRegion.csv)
   * Potential issues:
     + Line 107: Year is dependent variable in the linear model, it should be the independent variable.
7. FINAL\_2021.09.18\_Keys Region\_Genus.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends by genus across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - FLA KEYS-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 17: load libraries and import percent cover data for Florida Keys
     + 18 - 82: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 83 - 102: create a summary table for percent cover statistics by year, genus, region, save output
     + 103 - 109: create linear model year vs percent cover, not separated by genus
     + 110 - 124: create boxplots of percent cover by genus, year, save as pdf
     + 125 - 160: remove genera with less than ten years of data, create linear mixed effects model of percent cover for all genera vs year, random effect for program location ID
     + 161 - 183: create linear mixed effects model of percent cover for each genus vs year, random effect for program location ID, save output as csv
     + 184 - 209: plot model output as estimated linear trend across years by genus, save as pdf
     + 210 - 308: format and summarize model output to create summary table of model fit, saved as csv
   * File outputs: [CoralSummaryStatisticsTable\_Genus\_KEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_Genus_KEYSRegion.csv), [CoralBoxplots\_ByGenus\_KEYSRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_ByGenus_KEYSRegion.pdf), [Corallme\_ByGenus\_KEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_ByGenus_KEYSRegion.csv), [CorallmePlot\_ByGenus\_KEYSRegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_ByGenus_KEYSRegion.pdf), [Coral\_lmeResults\_ByGenus\_KEYSRegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Coral_lmeResults_ByGenus_KEYSRegion.csv)
   * Potential issues:
     + Line 107: Year is dependent variable in the linear model, it should be the independent variable.
8. FINAL\_2021.09.18\_SE Region\_AnyCorals.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends for any corals across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - SE FL-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 19: load libraries and import percent cover data for SE Florida
     + 20 - 78: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 79 - 100: create a summary table for percent cover statistics by year, region across all species, save output
     + 101 - 111: create linear model year vs percent cover, save model summary plots to pdf
     + 112 - 126: create boxplots of percent cover for any coral by year, save as pdf
     + 127 - 143: create linear mixed effects model of percent cover for any coral vs year, random effect for program location ID
     + 144 - 177: create linear mixed effects model of percent cover for any coral vs year, random effect for program location ID, separate for coral region, save output as csv
     + 178 - 200: plot model output as estimated linear trend across years by coral region, save as pdf
     + 201 - 299: format and summarize model output to create summary table of model fit, saved as csv
   * File outputs: [CoralSummaryStatisticsTable\_AnyCoral\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_AnyCoral_SERegion.csv), [CoralSummaryStatisticsPlots\_AnyCoral\_SERegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsPlots_AnyCoral_SERegion.pdf), CorSE.csv, [CoralBoxplots\_AnyCoral\_SERegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_AnyCoral_SERegion.pdf), [Corallme\_AnyCoral\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_AnyCoral_SERegion.csv), [CorallmePlot\_AnyCoral\_SERegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_AnyCoral_SERegion.pdf), [Coral\_lmeResults\_AnyCoral\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Coral_lmeResults_AnyCoral_SERegion.csv)
   * Potential issues:
     + 105, 108: Year is dependent variable in the linear model, it should be the independent variable.
     + Model results are separate for each coral region, yet the file includes only one region (SE Florida). This does not create incorrect output, but the intention is unclear.
9. FINAL\_2021.09.18\_SE\_Genus.R
   * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends by genus across years with linear models.
   * Packages: dplyr, gt, nlme, tidyverse
   * File inputs: Percent Cover - FLA SE-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 17: load libraries and import percent cover data for SE Florida
     + 18 - 81: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
     + 82 - 101: create a summary table for percent cover statistics by year, genus, region, save output
     + 102 - 108: create linear model year vs percent cover, not separated by genus
     + 109 - 123: create boxplots of percent cover by genus, year, save as pdf
     + 124 - 159: remove genera with less than ten years of data, create linear mixed effects model of percent cover for all genera vs year, random effect for program location ID. No output created because no genus meets the minimum year requirement.
     + 160 - 182: create linear mixed effects model of percent cover for each genus vs year, random effect for program location ID, save output as csv. No output created because no genus meets the minimum year requirement.
     + 183 - 208: plot model output as estimated linear trend across years by genus, save as pdf. No output created because no genus meets the minimum year requirement.
     + 209 - 307: format and summarize model output to create summary table of model fit, saved as csv. No output created because no genus meets the minimum year requirement.
   * File outputs: [CoralSummaryStatisticsTable\_Genus\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_Genus_SERegion.csv), [CoralBoxplots\_ByGenus\_SERegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_ByGenus_SERegion.pdf), [Corallme\_ByGenus\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_ByGenus_SERegion.csv), [CorallmePlot\_ByGenus\_SERegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_ByGenus_SERegion.pdf), Coral\_lmeResults\_ByGenus\_SERegion.csv
   * Potential issues:
     + Line 16: Input file name is incorrect, should be Percent Cover - SE FL-2021-Jul-26.csv
     + Line 106: Year is dependent variable in the linear model, it should be the independent variable.
     + No output created because no genus meets the minimum year requirement.
10. FINAL\_2021.09.18\_SERegion\_Group1.R
    * Objective: Import and format percent cover data, create summary boxplots, evaluate percent cover trends by species group across years with linear models.
    * Packages: dplyr, gt, nlme, tidyverse
    * File inputs: Percent Cover - SE FL-2021-Jul-26.csv
    * Steps by line number:
      + 1 - 17: load libraries and import percent cover data for SE Florida
      + 18 - 82: format data for analysis including removal of non-coral taxa, renaming columns, removing NA percent cover data, combining genus/species columns, and removing missing rows for genus/species
      + 83 - 102: create a summary table for percent cover statistics by year, species group, region, save output
      + 103 - 108: create linear model year vs percent cover, not separated by species group
      + 109 - 123: create boxplots of percent cover by species group, year, save as pdf
      + 124 - 159: create linear mixed effects model of percent cover for each species group vs year, random effect for program location ID, save as csv
      + 160 - 182: plot model output as estimated linear trend across years by species group, save as pdf
      + 183 - 281: format and summarize model output to create summary table of model fit, saved as csv
    * File outputs: [CoralSummaryStatisticsTable\_Group1\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/CoralSummaryStatisticsTable_Group1_SERegion.csv), [CoralBoxplots\_ByGroup1\_SERegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CoralBoxplots_ByGroup1_SERegion.pdf), [Corallme\_ByGroup1\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Corallme_ByGroup1_SERegion.csv), [CorallmePlot\_ByGroup1\_SERegion.pdf](https://FloridaSEACAR.github.io/SEACAR/output/CorallmePlot_ByGroup1_SERegion.pdf), [Coral\_lmeResults\_ByGroup1\_SERegion.csv](https://FloridaSEACAR.github.io/SEACAR/output/Coral_lmeResults_ByGroup1_SERegion.csv)
    * Potential issues:
      + Line 107: Year is dependent variable in the linear model, it should be the independent variable.

## 2.3 Coastal wetlands

### Scripts

|  |  |  |
| --- | --- | --- |
| Index | File | Link |
| 1 | Final 2021.09.19\_SEACAR\_CoastalWetlands\_GroupRichness.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/Final%202021.09.19_SEACAR_CoastalWetlands_GroupRichness.R) |
| 2 | FINAL 2021.09.19\_SEACAR\_CoastalWetlands\_SppRich\_Region.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/FINAL%202021.09.19_SEACAR_CoastalWetlands_SppRich_Region.R) |

### Summaries

1. Final 2021.09.19\_SEACAR\_CoastalWetlands\_GroupRichness.R
   * Objective: Import and format richness data, summarize group richness (Marsh, Marsh succulents, Mangroves and associate) by year, month for management areas including plots and linear models
   * Packages: dplyr, ggplot2, hrbrthemes, tidyverse
   * File inputs: All Parameters but Hecatres-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 23: load libraries and import wetland richness data for all regions
     + 24 - 119: format data for analysis including removal of NA data for genus, species, dates, removal of programs with insufficient data, removal of duplicates, formatting of genus, species, date columns, correct misspelled taxa, create unique species list by removing those doubly counted by different programs
     + 120 - 166: remove managed areas that don’t have at least five years of data, estimate group richness by management area, year, month
     + 167 - 187: create plots of group richness by management area, year, month
     + 188 - 207: create plots of group richness by management area, year
     + 208 - 224: select one managed area (GTMNERR), create and plot a linear model of group richness by month
     + 225 - 235: create plots of group richness by year, month and by year for selected managed area
     + 236 - 243: create a histogram of group richness for the selected managed area
     + 244 - 279: create a linear model of group richness by year for the selected managed area, plot the results
     + 280 - 298: select one managed area (Guana River Marsh), create and plot a linear model of group richness by month
     + 299 - 309: create plots of group richness by year, month and by year for selected managed area
     + 310 - 317: create a histogram of group richness for the selected managed area
     + 318 - 354: create a linear model of group richness by year for the selected managed area, plot the results
   * File outputs: [Coastal Wetlands Group Richness by YearMonth, Managed Area.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Coastal%20Wetlands%20Group%20Richness%20by%20YearMonth,%20Managed%20Area.pdf), [Coastal Wetlands Group Richness by Year, Managed Area.pdf.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Coastal%20Wetlands%20Group%20Richness%20by%20Year,%20Managed%20Area.pdf.pdf), [GTMNERR Coastal Wetlands Group Richness Data.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GTMNERR%20Coastal%20Wetlands%20Group%20Richness%20Data.pdf), [GTMNERR Coastal Wetlands Group Richness by Year\_Diagnostics.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GTMNERR%20Coastal%20Wetlands%20Group%20Richness%20by%20Year_Diagnostics.pdf), [GuanaRiverMarsh Coastal Wetlands Group Richness Data.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GuanaRiverMarsh%20Coastal%20Wetlands%20Group%20Richness%20Data.pdf), [GuanaRiverMarsh Coastal Wetlands Group Richness by Year\_Diagnostics.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GuanaRiverMarsh%20Coastal%20Wetlands%20Group%20Richness%20by%20Year_Diagnostics.pdf)
   * Potential issues:
     + The plots of richness by year in lines 188 - 207 use the same data to plot richness by month, year and is an incorrect measure of richness within a year. Unique groups in each year need to be separately calculated.
     + Plots in lines 225 - 235, 299 - 309 are redundant with those in 167 - 207
     + 270 - 278, 344 - 352 are commented code to evaluate linear models with different distribution families, e.g., Poisson, etc. These are worth exploring.
2. FINAL 2021.09.19\_SEACAR\_CoastalWetlands\_SppRich\_Region.R
   * Objective: Import and format richness data, summarize species richness by year, month for management areas including plots and linear models
   * Packages: dplyr, ggplot2, hrbrthemes, tidyverse
   * File inputs: All Parameters but Hecatres-2021-Jul-26.csv
   * Steps by line number:
     + 1 - 23: load libraries and import wetland richness data for all regions
     + 24 - 121: format data for analysis including removal of NA data for genus, species, dates, removal of programs with insufficient data, removal of duplicates, formatting of genus, species, date columns, correct misspelled taxa, create unique species list by removing those doubly counted by different programs
     + 122 - 168: remove managed areas that don’t have at least five years of data, estimate species richness by management area, year, month
     + 169 - 189: create plots of species richness by management area, year, month
     + 190 - 209: create plots of species richness by management area, year
     + 210 - 226: select one managed area (GTMNERR), create and plot a linear model of species richness by month
     + 227 - 237: create plots of species richness by year, month and by year for selected managed area
     + 238 - 245: create a histogram of species richness for the selected managed area
     + 246 - 281: create a linear model of species richness by year for the selected managed area, plot the results
     + 282 - 300: select one managed area (Guana River Marsh), create and plot a linear model of species richness by month
     + 301 - 311: create plots of species richness by year, month and by year for selected managed area
     + 312 - 319: create a histogram of species richness for the selected managed area
     + 320 - 356: create a linear model of species richness by year for the selected managed area, plot the results
   * File outputs: [wetlandsspprichReg.csv](https://FloridaSEACAR.github.io/SEACAR/output/wetlandsspprichReg.csv), [Coastal Wetlands Species Richness by YearMonth, Managed Area.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Coastal%20Wetlands%20Species%20Richness%20by%20YearMonth,%20Managed%20Area.pdf), [Coastal Wetlands Species Richness by Year, Managed Area.pdf.pdf](https://FloridaSEACAR.github.io/SEACAR/output/Coastal%20Wetlands%20Species%20Richness%20by%20Year,%20Managed%20Area.pdf.pdf), [GTMNERR Coastal Wetlands Species Richness Data.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GTMNERR%20Coastal%20Wetlands%20Species%20Richness%20Data.pdf), [GTMNERR Coastal Wetlands Species Richness by Year\_Diagnostics.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GTMNERR%20Coastal%20Wetlands%20Species%20Richness%20by%20Year_Diagnostics.pdf), [GuanaRiverMarsh Coastal Wetlands Species Richness Data.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GuanaRiverMarsh%20Coastal%20Wetlands%20Species%20Richness%20Data.pdf), [GuanaRiverMarsh Coastal Wetlands Species Richness by Year\_Diagnostics.pdf](https://FloridaSEACAR.github.io/SEACAR/output/GuanaRiverMarsh%20Coastal%20Wetlands%20Species%20Richness%20by%20Year_Diagnostics.pdf)
   * Potential issues:
     + The plots of richness by year in lines 190 - 209 use the same data to plot richness by month, year and is an incorrect measure of richness within a year. Unique species in each year need to be separately calculated.
     + Plots in lines 227 - 237, 301 - 311 are redundant with those in 169 - 209
     + 272 - 280, 346 -354 are commented code to evaluate linear models with different distribution families, e.g., Poisson, etc. These are worth exploring.

## 2.4 Submered aquatic vegetation

### Scripts

|  |  |  |
| --- | --- | --- |
| Index | File | Link |
| 1 | SEACAR\_SAV\_BB\_script.R | [link](https://github.com/FloridaSEACAR/SAV_CLR/blob/main/SEACAR_SAV_BB_script.R) - private repo |

### Summaries

1. SEACAR\_SAV\_BB\_script.R
   * Objective: Import and format SAV data, create summary plots and maps of changes over time, model changes over time using Bayesian and mixed-effects models. Results are separate for each managed area, species, and parameter (e.g., Braun Blanquet, percent cover, etc.).
   * Packages: bayesplot, brms, broom.mixed, data.table, grid, gridExtra, gtable, nlme, scales, sf, tictoc, tidybayes, tidyverse
   * File inputs: Combined\_SAV\_column\_All-2021-Sep-20.csv, seacar\_dbo\_SampleLocation\_Point.shp, seacar\_dbo\_SampleLocation\_Line.shp, ORCP\_Managed\_Areas.shp, Counties\_-\_Detailed\_Shoreline.shp, MApolygons\_corners.csv
   * Steps by line number:
     + 1 - 17: load libraries and import SAV file
     + 18 - 104: format SAV data including renaming columns, removing NA values, and reformatting abundance/cover values. For the latter, this included removal of NA values and those out of range, and ensuring appropriate values for Braun Blanquet, modified Braun Blaunqet, percent occurrence, and percent cover.
     + 105 - 243: commented (not used) code for creating a percent occurrence table and functions for plotting model results
     + 244 - 293: function for plotting model predictions
     + 294 - 342: function for rotating a spatial object, used for plotting to create a presentation of stacked 2-d layers.
     + 343 - 366: import spatial data objects, project all to WGS 1984 datum. Spatial data objects include sample locations and relevant boundaries (counties, management area boundaries)
     + 367 - 400: setup parameter list and objects for looping through parameters to create models and summary output, parameters include Braun Blanquet, median percent cover, visual percent cover, percent occurrence, frequency of occurrence
     + 401 - 1712: loop through parameters to create models and summary output, the following is an outline of steps in this loop.
       - 416 -1710: loop through managed areas using parameter from outside loop
       - 435 - 449: create and save plot of parameter score for managed area over time by species
       - 450 - 464: create and save plot of parameter score for managed area over time by program ID
       - 465 - 479: create and save plot of parameter score for managed area over time grouped by species, by program ID
       - 480 - 494: create and save plot of quadrat sizes for managed area over time by species
       - 495 - 509: create and save plot of quadrat sizes for managed area over time by program ID
       - 510 - 524: create and save plot of method for managed area over time by species
       - 525 - 539: create and save plot of method for managed area over time by program ID
       - 540 - 555: create and save plot of method for managed area by quadrat size and species
       - 556 - 571: create and save plot of method for managed area by quadrat size and program ID
       - 572 - 603: create and save plots of grid values over time by species and program ID if data available
       - 604 - 636: create and save plots of depth values over time by species and program ID if data available
       - 637 - 655: create and save a plot legend of species
       - 675 - 681: loop through species to create and save a plot of parameter score over time
       - 681 - 704: create and save a plot of totals for the species of parameter score over time
       - 705 - 723: create and save a plot legend of species
       - 724 - 749: loop through species to create and save a plot of parameter score as boxplots over time
       - 750 - 772: create and save a plot of boxplots for the species of parameter score over time
       - 773 - 833: create and save stacked maps of parameter values for the managed area by year
       - 834 - 886: setup empty objects for model results
       - 888 - 1603: Loop through species to fit models, with separate exception statements for different parameters. The modeling workflow is similar for each parameter, with minor exceptions. The general goal of each is to assess trends in a parameter over time for a particular species and managed area. Each workflow includes error handling if models did not converge, produces summary tables of model fit, and summary plots showing model result. The models vary in the Gaussian distribution family for the response variable depending on parameter. Random effects (e.g., for LocationID) are used for all models.
       - 1604 - 1648: create and save barplots of parameter results for managed area over time by species, only for Braun Blanquet and percent cover, save model results from prior loops
       - 1649 - 1666: save model results from prior loops for frequency occurrence and percent occurrence
       - 1667 - 1704: create and save barplots of parameter results for managed area over time by species, only for frequency occurrence, save model results from prior loops
       - 1705 - 1709: print statements for the console
     + 1713 - 1716: save list of failed models and R session info
   * File outputs: Multiple binary RDS files.
   * Potential issues:
     + Functions use objects created in the global environment, not passed as arguments to the function, e.g., line 244 uses SAV4 inside the function, but it is not used as an argument to the function. The script works, but it can cause problems for modularity (i.e., if functions were moved to another script).
     + All output, including plots, are in binary format, which I assume was to reduce file size, but it is difficult to evaluate results in this format. There are also hundreds of files, so a summary or a discussion of how these results are presented is needed.

## 2.5 Oyster reef

### Scripts

|  |  |  |
| --- | --- | --- |
| Index | File | Link |
| 1 | SEACAR\_Oyster\_Analyses\_Final.R | [link](https://github.com/FloridaSEACAR/SEACAR/blob/master/SEACAR_Oyster_Analyses_Final.R) |

### Summaries

1. SEACAR\_Oyster\_Analyses\_Final.R

* Objective: Import and format oyster data, model changes over time using Bayesian and mixed-effects models. Results are separate for each managed area, size classes (25 - 75, > 75), and parameter (e.g., shell height, density, percent live).
  + Packages: brms, data.table, doFuture, doRNG, ggplot2, here, lubridate, mapview, modelr, piecewiseSEM, rcompanion, Rmisc, sf, tictoc, tidybayes, tidyverse
  + File inputs: All\_Parameters\_but\_Hectares-2021-Sep-20.csv, Oyster\_Beds\_in\_Florida.shp, Florida\_Aquatic\_Preserves.shp, Florida\_National\_Estuarine\_Resarch\_Reserves\_\_NERR\_\_Boundaries.shp, GTM\_RB\_2016\_Merge.kml, seacar\_dbo\_SampleLocation\_Point.shp, GTMNERR\_Regions.kml
  + Steps by line number:
    - 1 - 36: load libraries and import oyster data
    - 37 - 244: create universal reef IDs
    - 245 - 279: add a second region column for areas of the Guana Tolomato NERR to the oyster data
    - 280 - 402: formatting of oyster data including correct date column, converting columns to correct format (i.e., numeric, character, etc.), manual correction of data entries that were in error, filling missing values that can be calculated from non-missing columns (i.e., density), verifying unique quadrat identifiers, and removing size outliers. An empty table for storing model results is also created.
    - 403 - 419: create a summary table of oyster shell height
    - 420 - 1660: these lines are repeated code to evaluate trends in oyster shell height over time by managed area, where the code is repeated by managed area. Separate models are created for size ranges for those from 25 - 75 mm and those greater than 75. Trends are assessed using Bayesian regression models, that also incorporate measurement error of predictors (Sample\_age\_stdev) and random effects (UniversalReefID). Summary outputs for each model include diagnostic plots and estimated trends.
    - 1661 - 1698: create a summarized oyster density table
    - 1699 - 1931: estimate densities by different size classes
    - 1932 - 2317: these lines are repeated code to evaluate trends in oyster density over time by managed area, where the code is repeated by managed area. Trends are assessed using Bayesian regression models, with random effects for relative year conditional on UniversalReefID. Summary outputs for each model include diagnostic plots and estimated trends.
    - 2318 - 2534: similar models and results as those from the previous lines are created, but for oyster densities in size classes 25 - 75mm
    - 2535 - 2686: similar models and results as those from the previous lines are created, but for oyster densities in size classes > 75mm
    - 2687 - 2732: create a data table for percent live for models in the following lines
    - 2733 - 2947: these lines are repeated code to evaluate trends in percent live oysters over time by managed area, where the code is repeated by managed area. Trends are assessed using Bayesian regression models, with random effects for relative year conditional on UniversalReefID. Summary outputs for each model include diagnostic plots and estimated trends.
  + File outputs: Multiple binary RDS files. Some plot results are created as jpeg files but I could not reproduce.
  + Potential issues:
    - I could not recreate any of the results because the formatting on lines 280 - 402 did not produce data appropriate for analysis. This is likely a simple issue with data inputs and package versioning (i.e., my local setup is not the same as the setup used to write the scripts).
    - I could not recreate the universal reef IDs on lines 37 - 244, as noted by S. Durham. These IDs should have been added to the database by Claude, so this code is not necessary but may be good to keep for a record of how the IDs were created.
    - The projection for the shapefiles used for creating the universal reef IDs is for a UTM zone that applies to North Carolina (EPSG: 32119). I’m not sure this matters, but it may be worth verifying no spatial inaccuracies result from applying the projection to Florida data.
    - The conversion of the LiveDate column to an integer does not work on line 297. I assume this is meant to extract a year value from a date column. Downstream analyses do not work without this column.
    - Oyster summary table on lines 403 - 419 returns NA values for all estimates of shell height because column conversions earlier in the script did not work.
    - The code on lines 1699 - 1931 is very difficult to follow and it’s not clear what’s produced.

# 3 General comments

* Leverage more tools from the tidyverse for more concise, streamline coding. For example, use case\_when instead of nested ifelse statements.
* Modularize the parts of copied code in functions. For example, the scripts for water quality variables are all very similar. Functions could be created to remove NAs, filter by QA codes, create summary plots, run trend tests, and format results.
* General issue with portability, environment for analysis is not isolated, i.e., scripts were created on a local computer and specific setup was unknown. Could be solved by package devevelopment, possibly combined with Docker container. Use of rm(list = ls() at the top of many scripts demonstrates the problem.
* Many scripts use install.packages() which is not necessary for running the scripts in production.
* Some libraries are loaded more than once (e.g., tidyverse, then dplyr which is included in tidyverse, as in lines 25-29 [here](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR%20Water_Nutrients_Phosphorus_FINAL.R)). This does not impact performance but creates challenges for tracking dependencies. All relevant libraries should be entered at the top of each script.
* Use of logical values entered as text strings (e.g., line 151 [here](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR%20Water_Nutrients_Phosphorus_FINAL.R)) is incorrect and should use the actual atomic values, i.e., TRUE, FALSE.
* Water quality trend tests are based on average year, month values. Is it possible to run these tests without averaging first? This reduces sample size and removes real variability that could influence the outcome.
* Month, year summary boxplots for water column variables are difficult to interpret as all values are chronological, maybe better to show just month on x-axis, then facet by year.
* Output files created with sink create a csv but this would be more appropriate as a text file since these are not tabular data (e.g., line 361 [here](https://github.com/FloridaSEACAR/SEACAR/blob/master/2021.08.21_FINAL_SEACAR_Water%20Clarity_chla_Final.R)).
* Formatted water column tables with trend statistics for water column data should have number of years for which the trend applies, in addition to min/max years. The time series for each managed area are all different and it’s useful to provide this information.
* Data I/O needs to be evaluated. All scripts use files from a local path and outputs are saved in the working directory. This is an issue to be solved in production.
* For corals, analyses were requested separately by sampling type: fixed or random. Note that this refers to sampling method, not a statistical test. However, the scripts indicate that sample information is unavailable in the input data (e.g., line 22 [here](https://github.com/FloridaSEACAR/SEACAR/blob/main/FINAL_2021.09.18_DryTort%20Region_Any%20Corals.R)). The analyses are not done at this time.
* For corals, trends in percent cover are assessed using linear mixed-effects models. It may be worth confirming trends using non-parametric analogs, particular since percent cover is generally a right-skewed variable. There also is not a minimum year requirement for tests in percent cover evaluating any coral and for species groups, unlike trends tests for other habitats (i.e., water column). However, there is a minimum year requirement for tests evaluating genera.
* Trend tests for percent cover of coral by genus require a minimum of ten years of data. For the Dry Tortugas and SE Florida region, no genus meet the criteria. The analysis script returns errors as opposed to informative output indicating insufficient data. This is an issue that could be solved through package development, e.g., using stopifnot(), or other error handling so that analyses don’t “break” in production when test criteria are unmet.
* As with other habitat types, the coral analyses could be greatly simplified by creating functions that accomplish routine tasks. Each analysis is similar - for each region, create a linear model of percent cover across years, with separate models for all taxa combined, by genus, or by species groups. Standardized outputs are also created, e.g., plots of model predictions, tables of model summaries, plots of observed data, etc. All of this can be turned into functions, including appropriate exceptions within each function to handle unique cases (e.g., filter genus data for those with at least ten years of data).
* Coastal wetlands scripts only evaluate richness. There were no analyses evaluating percent cover, which were indicated as an outcome in supplementary documents.
* Coastal wetlands analyses filter managed areas that do not have at least five years of data. Other analyses use ten years as a criteria. Is there a justification for choosing a minimum number of years?
* For seagrasses, the modelling work is by far the most comprehensive, but it will be very challenging to distill the results into meaningful summaries. I also question if the modelling framework can be simplified. Bayesian models with prior distributions are used for many of the parameters, which likely creates a more robust model, but it may be difficult to scale these models for automation. Other habitat analyses use simple non-parametric tests, which have their downfalls, but it may be worth exploring these simpler models, i.e., do the more complex models provide an advantage or do they give the same answer at the end of the day? The summary plots are also generally diagnostic and not a synthesis of results for quick interpretation. The stacked maps are cool, but the idea doesn’t translate well for more than a handful of years (i.e., each layer in the stack is a year).
* For oysters, many of my comments about model complexity for the seagrass analyses apply equally to these.