SEACAR Discrete Water Quality Analysis: Surface Dissolved Oxygen Saturation

Last compiled on 18 May, 2022

Contents

Purpose	1
Adjustable Inputs	2
Libraries	2
File Import	3
Data Filtering and Data Impacted by Specific Value Qualifiers	4
Managed Area Statistics	6
Monitoring Location Statistics	7
Seasonal Kendall Tau Analysis	8
Appendix I: Scatter Plot of Entire Dataset	11
Appendix II: Dataset Summary Box Plots	14
Appendix III: Excluded Managed Areas	20
Appendix IV: Managed Area Trendlines	26
Appendix V: Managed Area Summary Box Plots	44

Purpose

The purpose of this script is to analyze the discrete surface dissolved oxygen saturation data that is created from the SEACAR database, apply filtering criteria, create summary plots, and perform seasonal Kendall Tau analysis for each program location and summary statistics for values measured at the desired depth.

All scripts and outputs can be found on the SEACAR GitHub repository:

https://github.com/FloridaSEACAR/SEACAR Panzik

Note: The top 2% of data is excluded when computing mean and standard deviations in plotting sections solely for the purpose of getting y-axis scales. The exclusion of the top 2% is not used in any statistics that are exported.

Adjustable Inputs

This is placed early so that is is easier to edit parameters that users may want to adjust.

The first variable is whether you want to create the summary plots in the appendices. If you want to see all appendix plots, set APP_Plots to TRUE. If you would like to only perform the analysis and export the data files with minimal plots, set APP_Plots to FALSE. This option is available because generating the plots in the appendices increases the processing time significantly.

Since the file names all have similar structure with only the parameter name being varied, the code below sets variables to include standard string information that is the same across all data files.

This includes: the raw data directory (in_dir), output file directory (out_dit), file prefix (file_pref), date the files were created from the database (file_date), the name of the parameter of interest (param_name), and the relative depth of interest (depth). The complete file name is created by pasting all of the strings together with the specific parameter name without spaces (paste0 command).

```
APP_Plots <- TRUE
in_dir <- "data/"
out_dir <- "output/"
file_pref <- "Combined_WQ_WC_NUT_"
file_date <- "2022-Apr-13"
param_name <- "Dissolved_Oxygen_Saturation"
depth <- "Surface"
activity <- "Field"
```

Libraries

Loads libraries used in the script. The inclusion of scipen option limits how frequently R defaults to scientific notation.

```
library(knitr)
library(data.table)

## Warning: package 'data.table' was built under R version 4.2.0
library(dplyr)
```

Warning: package 'dplyr' was built under R version 4.2.0

```
library(lubridate)
library(ggplot2)
library(ggpubr)
library(scales)
library(EnvStats)

## Warning: package 'EnvStats' was built under R version 4.2.0

library(tidyr)
options(scipen = 999)
```

File Import

Creates file name from inputs above and read in the file from txt format with pipe delimiters.

The code creates output directories for the output files if they don't exist in the directory.

The command fread is used because of its improved speed while handling large data files. Only columns that are used by the script are imported from the file, and are designated in the select input.

The script then gets the units of the parameter, sets the SampleDate as a date object, and creates various scales of the date to be used by plotting functions.

```
if(!file.exists("output")){
   dir.create("output")}
file_in <- paste0(in_dir, file_pref, param_name, "-", file_date, ".txt")</pre>
data <- fread(file_in, sep = "|", header = TRUE, stringsAsFactors = FALSE,</pre>
             select = c("ManagedAreaName", "ProgramID", "ProgramName",
                         "ProgramLocationID", "SampleDate", "Year", "Month",
                         "RelativeDepth", "ActivityType", "ParameterName",
                         "ResultValue", "ParameterUnits", "ValueQualifier",
                         "SEACAR_QAQCFlagCode", "Include"), na.strings = "")
MA_All <- fread("../../ManagedArea.csv", sep = ",", header = TRUE, stringsAsFactors = FALSE,
             na.strings = "")
data <- merge.data.frame(MA_All[,c("AreaID", "ManagedAreaName")],</pre>
                           data, by = "ManagedAreaName")
parameter <- unique(data$ParameterName)</pre>
unit <- unique(data$ParameterUnits)</pre>
data$SampleDate <- as.Date(data$SampleDate)</pre>
data$YearMonth <- pasteO(data$Month, "-", data$Year)</pre>
data$YearMonthDec <- data$Year + ((data$Month-0.5) / 12)</pre>
data$DecDate <- decimal_date(data$SampleDate)</pre>
```

Data Filtering and Data Impacted by Specific Value Qualifiers

Most data filtering is performed on export from the database, and is indicated by the Include variable. Include values of 1 indicate the data should be used for analysis, values of 0 indicate the data should not be used for analysis. Documentation on the database filtering is provided here: SEACAR Documentation-Analysis Filters and Calculations.docx

The filtering that is performed by the script at this point removes rows that are missing values for ResultValue, and only keeps data that is measured at the relative depth (surface, bottom, etc.) of interest. This is partly handled on export with the RelativeDepth variable, but there are some measurements that are considered both surface and bottom based on measurement depth and total depth. By default, these are marked as Surface for RelativeDepth and receive a SEACAR_QAQCFlag indicator of 12Q. Data passes the filtering the process if it is from the correct depth and has an Include value of 1.

After the initial filtering, a second filter variable is created to determine whether enough time is represented in the managed area, which is that each managed area has 10 year or more of unique year entries for observation that pass the initial filter. If data passes the first set of filtering criteria and the time criteria, they are used in the analysis.

After filtering, the amount of data impacted by the H (for dissolved oxygen & pH in program 476), I, Q, and U value qualifiers. A variable is also created that determines if scatter plot points should be a different color based on value qualifiers of interest.

```
if(depth=="Bottom"){
   data$RelativeDepth[grep("12Q", data$SEACAR_QAQCFlagCode[
      data$RelativeDepth == "Surface"])] <- "Bottom"</pre>
}
data$Include <- as.logical(data$Include)</pre>
data <- data[!is.na(data$ResultValue),]</pre>
data <- data[!is.na(data$RelativeDepth),]</pre>
data <- data[data$RelativeDepth==depth,]</pre>
if(length(grep("Blank", data$ActivityType)) >0){
      data <- data[-grep("Blank", data$ActivityType),]</pre>
}
if(param_name == "Chlorophyll_a_uncorrected_for_pheophytin" |
   param_name == "Salinity" | param_name == "Turbidity"){
   data <- data[grep(activity, data$ActivityType[!is.na(data$ActivityType)]),]</pre>
}
if(param_name == "Water_Temperature"){
   data <- data[data$ResultValue>=-2,]
} else{
   data <- data[data$ResultValue>=0,]
data$Include[grep("H", data$ValueQualifier[data$ProgramID==476])] <- TRUE
MA_Years <- data[data$Include == TRUE, ] %>%
   group_by(ManagedAreaName) %>%
   summarize(N = length(unique(Year)))
MA_Years <- as.data.table(MA_Years[order(MA_Years$ManagedAreaName), ])
MA_Years$Enough_Time <- ifelse(MA_Years$N < 10, FALSE, TRUE)
data <- merge.data.frame(data, MA_Years[,c("ManagedAreaName", "Enough_Time")],
                          by = "ManagedAreaName")
```

```
data$Use_In_Analysis <- ifelse(data$Include == TRUE &</pre>
                                    data$Enough_Time == TRUE,
                                 TRUE, FALSE)
total <- length(data$Include)</pre>
pass_filter <- length(data$Include[data$Include==TRUE])</pre>
count H <- length(grep("H", data$ValueQualifier[data$ProgramID==476]))</pre>
perc H <- 100*count H/length(data$ValueQualifier)</pre>
count_I <- length(grep("I", data$ValueQualifier))</pre>
perc_I <- 100*count_I/length(data$ValueQualifier)</pre>
count_Q <- length(grep("Q", data$ValueQualifier))</pre>
perc_Q <- 100*count_Q/length(data$ValueQualifier)</pre>
count_U <- length(grep("U", data$ValueQualifier))</pre>
perc_U <- 100*count_U/length(data$ValueQualifier)</pre>
data$VQ_Plot <- data$ValueQualifier</pre>
inc_H <- ifelse(param_name=="pH" | param_name=="Dissolved_Oxygen" |</pre>
                    param_name=="Dissolved_Oxygen_Saturation", TRUE, FALSE)
if (inc H==TRUE){
   data$VQ_Plot <- gsub("[^HU]+", "", data$VQ_Plot)</pre>
   data$VQ_Plot <- gsub("UH", "HU", data$VQ_Plot)</pre>
   data$VQ_Plot[data$ProgramID!=476] <- gsub("[^U]+", "",</pre>
                                                data$VQ_Plot[data$ProgramID!=476])
   data$VQ_Plot[data$VQ_Plot==""] <- NA</pre>
   cat(paste0("Number of Measurements: ", total,
               ", Number Passed Filter: ", pass_filter, "\n",
               "Program 476 H Codes: ", count_H, " (", round(perc_H, 6), "%)\n",
               "I Codes: ", count_I, " (", round(perc_I, 6), "%)\n",
               "Q Codes: ", count_Q, " (", round(perc_Q, 6), "%)\n",
               "U Codes: ", count_U, " (", round(perc_U, 6), "%)"))
} else{
   data$VQ_Plot <- gsub("[^U]+", "", data$VQ_Plot)</pre>
   data$VQ_Plot[data$VQ_Plot==""] <- NA</pre>
   cat(paste0("Number of Measurements: ", total,
               ", Number Passed Filter: ", pass_filter, "\n",
               "I Codes: ", count_I, " (", round(perc_I, 6), "%)\n",
               "Q Codes: ", count_Q, " (", round(perc_Q, 6), "%)\n",
               "U Codes: ", count_U, " (", round(perc_U, 6), "%)"))
}
## Number of Measurements: 88515, Number Passed Filter: 87937
## Program 476 H Codes: 0 (0%)
## I Codes: 0 (0%)
## Q Codes: 0 (0%)
## U Codes: 0 (0%)
```

Managed Area Statistics

Gets summary statistics for each managed area. Excluded managed areas are not included into whether the data should be used or not. Uses piping from dplyr package to feed into subsequent steps. The following steps are performed:

- 1. Take the data variable and only include rows that have a Use_In_Analysis value of TRUE
- 2. Group data that have the same ManagedAreaName, Year, and Month.
 - Second summary statistics do not use the Month grouping and are only for ManagedAreaName and Year.
 - Third summary statistics do not use Year grouping and are only for ManagedAreaName and Month
- 3. For each group, provide the following information: Number of Entries (N), Lowest Value (Min), Largest Value (Max), Median, Mean, Standard Deviation, and a list of all Program IDs included in these measurements.
- 4. Sort the data in ascending (A to Z and 0 to 9) order based on ManagedAreaName then Year then Month
- 5. Write summary stats to a pipe-delimited .txt file in the output directory
 - Click this text to open Git directory with output files

```
MA_names <- unique(data$ManagedAreaName[data$Use_In_Analysis == TRUE])
MA_names <- MA_names[order(MA_names)]</pre>
n <- length(MA_names)</pre>
MA_YM_Stats <- data[data$Use_In_Analysis == TRUE, ] %>%
   group_by(ManagedAreaName, Year, Month) %>%
   summarize(N = length(ResultValue),
             Min = min(ResultValue),
             Max = max(ResultValue),
             Median = median(ResultValue),
             Mean = mean(ResultValue),
             StandardDeviation = sd(ResultValue),
             ProgramIDs = paste(sort(unique(ProgramID), decreasing = FALSE),
                                collapse = ', '))
MA_YM_Stats <- as.data.table(MA_YM_Stats[order(MA_YM_Stats$ManagedAreaName,
                                                MA YM Stats$Year,
                                                MA YM Stats$Month), ])
fwrite(MA_YM_Stats, paste0(out_dir,"/", param_name, "_", file_date, "_",
                           activity, "_", depth,
                           "_ManagedArea_YearMonth_Stats.txt"), sep = "|")
MA_Y_Stats <- data[data$Use_In_Analysis == TRUE, ] %>%
   group_by(ManagedAreaName, Year) %>%
   summarize(N = length(ResultValue),
             Min = min(ResultValue),
             Max = max(ResultValue),
             Median = median(ResultValue),
             Mean = mean(ResultValue),
             StandardDeviation = sd(ResultValue),
             ProgramIDs = paste(sort(unique(ProgramID), decreasing = FALSE),
                                collapse = ', '))
MA_Y_Stats <- as.data.table(MA_Y_Stats[order(MA_Y_Stats$ManagedAreaName,
                                              MA_Y_Stats$Year), ])
fwrite(MA_Y_Stats, paste0(out_dir,"/", param_name, "_", file_date, "_",
```

```
activity, "_", depth,
                          "_ManagedArea_Year_Stats.txt"), sep = "|")
MA M Stats <- data[data$Use In Analysis == TRUE, ] %>%
   group by (ManagedAreaName, Month) %>%
   summarize(N = length(ResultValue),
             Min = min(ResultValue),
             Max = max(ResultValue),
             Median = median(ResultValue),
             Mean = mean(ResultValue),
             StandardDeviation = sd(ResultValue),
             ProgramIDs = paste(sort(unique(ProgramID), decreasing = FALSE),
                                collapse = ', '))
MA_M_Stats <- as.data.table(MA_M_Stats[order(MA_M_Stats$ManagedAreaName,
                                             MA_M_Stats$Month), ])
fwrite(MA_M_Stats, paste0(out_dir,"/", param_name, "_", file_date, "_",
                          activity, "_", depth,
                          " ManagedArea Month Stats.txt"), sep = "|")
```

Monitoring Location Statistics

Gets monitoring location statistics, which is defined as a unique combination of ManagedAreaName, ProgramID, ProgramAreaName, and ProgramLocationID, using piping from dplyr package. The following steps are performed:

- 1. Take the data variable and only include rows that have a Use_In_Analysis value of TRUE
- 2. Group data that have the same ManagedAreaName, ProgramID, ProgramName, and ProgramLocationID.
- 3. For each group, provide the following information: Earliest Sample Date (EarliestSampleDate), Latest Sample Date (LastSampleDate), Number of Entries (N), Lowest Value (Min), Largest Value (Max), Median, Mean, and Standard Deviation.
- 4. Sort the data in ascending (A to Z and 0 to 9) order based on ManagedAreaName then ProgramName then ProgramLocationID
- 5. Write summary stats to a pipe-delimited .txt file in the output directory
 - Click this text to open Git directory with output files

Seasonal Kendall Tau Analysis

Gets seasonal Kendall Tau statistics using the kendallSeasonalTrendTest from the EnvStats package. The Trend parameter is determined from a user-defined function based on the median, Senn slope, and p values from the data. Analysis modified from code created by Jason Scolaro that performed at The Water Atlas: https://sarasota.wateratlas.usf.edu/water-quality-trends/#analysis-overview

The following steps are performed:

- 1. Define the functions used in the analysis
- 2. Check to see if there are any groups to run analysis on.
- 3. Take the data variable and only include rows that have a Use_In_Analysis value of TRUE
- 4. Group data that have the same ManagedAreaName.
- 5. For each group, provides the following information: Earliest Sample Date (EarliestSampleDate), Latest Sample Date (LastSampleDate), Number of Entries (N), Lowest Value (Min), Largest Value (Max), Median, Mean, Standard Deviation, tau, Senn Slope (SennSlope), Senn Intercept (SennIntercept), and p.
 - The analysis is run with the kendallSeasonalTrendTest function using the Year values for year, and Month as the seasonal qualifier, and Trend.
 - An independent obs value of TRUE indicates that the data should be treated as not being serially auto-correlated. An independent obs value of FALSE indicates that it is treated as being serially auto-correlated, but also requires one observation per season per year for the full time of observation.
- 6. Reformat columns in the data frame from export.
- 7. Write summary stats to a pipe-delimited .txt file in the output directory
 - Click this text to open Git directory with output files

```
trend <- trend_calculator(slope, stats.median, p)</pre>
   }, warning = function(w) {
      print(w)
   }, error = function(e) {
      print(e)
   }, finally = {
      if (!exists("tau")) {
         tau <- NULL
      if (!exists("p")) {
         p <- NULL
      if (!exists("slope")) {
         slope <- NULL</pre>
      if (!exists("intercept")) {
         intercept <- NULL</pre>
      if (!exists("trend")) {
         trend <- NULL
   })
   KT <-c(unique(data$AreaID),</pre>
          unique(data$ManagedAreaName),
          parameter,
          depth,
          activity,
          independent,
          stats.median,
          nrow(data),
          stats.minYear,
          stats.maxYear,
          tau,
          p,
          slope,
          intercept,
          trend)
   return(KT)
}
runStats <- function(data) {</pre>
   data$Index <- as.Date(data$SampleDate) # , "%Y-%m-%d")</pre>
   data$ResultValue <- as.numeric(data$ResultValue)</pre>
   # Calculate basic stats
   stats.median <- median(data$ResultValue, na.rm = TRUE)</pre>
   stats.minYear <- min(data$Year, na.rm = TRUE)</pre>
   stats.maxYear <- max(data$Year, na.rm = TRUE)</pre>
   # Calculate Kendall Tau and Slope stats, then update appropriate columns and table
   KT <- tauSeasonal(data, TRUE, stats.median,</pre>
                       stats.minYear, stats.maxYear)
   if (is.null(KT[11])) {
      KT <- tauSeasonal(data, FALSE, stats.median,</pre>
                          stats.minYear, stats.maxYear)
   }
```

```
if (is.null(KT.Stats) == TRUE) {
      KT.Stats <- KT
   } else{
      KT.Stats <- rbind(KT.Stats, KT)</pre>
   return(KT.Stats)
trend calculator <- function(slope, median value, p) {</pre>
   trend <-
      if (p < .05 \& abs(slope) > abs(median_value) / 10.) {
         if (slope > 0) {
             2
         }
         else {
             -2
   else if (p < .05 & abs(slope) < abs(median_value) / 10.) {</pre>
      if (slope > 0) {
         1
      }
      else {
         -1
   }
   else
      0
   return(trend)
KT.Stats <- NULL
# Loop that goes through each managed area. List of managed areas stored in MA_Years$ManagedAreaName
c_names <- c("AreaID", "ManagedAreaName", "ParameterName", "RelativeDepth",</pre>
              "ActivityType", "Independent", "Median", "N", "EarliestYear",
              "LatestYear", "tau", "p", "SennSlope", "SennIntercept", "Trend")
if(n==0){
   KT.Stats <- data.frame(matrix(ncol=15, nrow=0))</pre>
   colnames(KT.Stats) <- c_names</pre>
   fwrite(KT.Stats, paste0(out_dir,"/", param_name, "_", file_date, "_",
                            activity, "-", depth,
                            "_KendallTau_Stats.txt"), sep = "|")
} else{
   for (i in 1:n) {
      values <- data[data$Use_In_Analysis == TRUE &</pre>
                         data$ManagedAreaName == MA_names[i], ]
      if (nrow(values) > 0) {
         KT.Stats <- runStats(values)</pre>
      }
   }
   KT.Stats <- as.data.frame(KT.Stats)</pre>
   c_names <- c("AreaID", "ManagedAreaName", "ParameterName", "RelativeDepth",</pre>
                 "ActivityType", "Independent", "Median", "N", "EarliestYear",
                 "LatestYear", "tau", "p", "SennSlope", "SennIntercept", "Trend")
   if(dim(KT.Stats)[2] == 1){
```

```
KT.Stats <- as.data.frame(t(KT.Stats))</pre>
}
colnames(KT.Stats) <- c_names</pre>
rownames(KT.Stats) <- seq(1:nrow(KT.Stats))</pre>
KT.Stats$Median <- as.numeric(KT.Stats$Median)</pre>
KT.Stats$N <- as.integer(KT.Stats$N)</pre>
KT.Stats$EarliestYear <- as.integer(KT.Stats$EarliestYear)</pre>
KT.Stats$LatestYear <- as.integer(KT.Stats$LatestYear)</pre>
KT.Stats$tau <- round(as.numeric(KT.Stats$tau), digits=4)</pre>
KT.Stats$p <- round(as.numeric(KT.Stats$p), digits=4)</pre>
KT.Stats$SennSlope <- as.numeric(KT.Stats$SennSlope)</pre>
KT.Stats$SennIntercept <- as.numeric(KT.Stats$SennIntercept)</pre>
KT.Stats$Trend <- as.integer(KT.Stats$Trend)</pre>
fwrite(KT.Stats, paste0(out_dir,"/", param_name, "_", file_date, "_",
                          activity, "_", depth,
                           "_KendallTau_Stats.txt"), sep = "|")
```

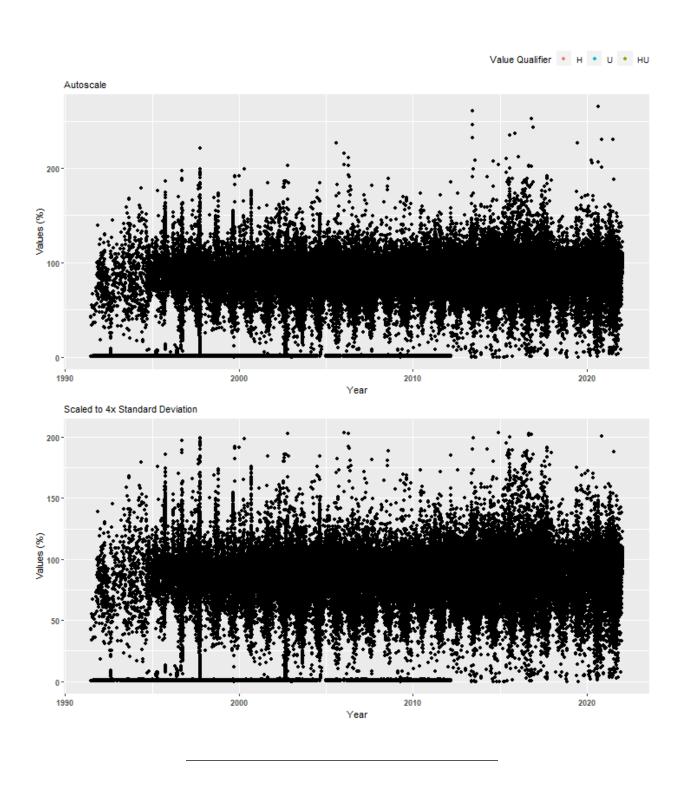
Appendix I: Scatter Plot of Entire Dataset

This part will create a scatter plot of the all data that passed initial filtering criteria with points colored based on specific value qualifiers. The values determined at the beginning (year_lower, year_upper, min_RV, mn_RV, x_scale, and y_scale) are solely for use by the plotting functions and are not output as part of the computed statistics.

```
year_lower <- min(data$Year)</pre>
year_upper <- max(data$Year)</pre>
min_RV <- min(data$ResultValue)</pre>
mn_RV <- mean(data$ResultValue[data$ResultValue <</pre>
                                    quantile(data$ResultValue, 0.98)])
sd_RV <- sd(data$ResultValue[data$ResultValue <</pre>
                                  quantile(data$ResultValue, 0.98)])
x_scale <- ifelse(year_upper - year_lower > 30, 10, 5)
y_scale \leftarrow mn_RV + 4 * sd_RV
p1 <- ggplot(data = data[data$Include==TRUE,],</pre>
             aes(x = SampleDate, y = ResultValue,
                  color=VQ_Plot)) +
   geom_point(size = 1.5) +
   labs(subtitle = "Autoscale",
        x = "Year", y = paste0("Values (", unit, ")"),
        color="Value Qualifier") +
   theme(legend.position = "top", legend.box = "horizontal",
         legend.justification = "right",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element text(face="bold")) +
   scale_x_date(labels = date_format("%Y")) +
   {if(inc H==TRUE){
```

```
scale_color_manual(values = c("H"= "#F8766D", "U"= "#00BFC4",
                                     "HU" = "#7CAE00"), na.value="black")
   } else {
      scale_color_manual(values = c("U"= "#00BFC4"), na.value="black")
   }}
p2 <- ggplot(data = data[data$Include==TRUE,],</pre>
             aes(x = SampleDate, y = ResultValue,
                 color=VQ_Plot)) +
   geom_point(size = 1.5) +
   ylim(min_RV, y_scale) +
   labs(subtitle = "Scaled to 4x Standard Deviation",
        x = "Year", y = paste0("Values (", unit, ")")) +
   theme(legend.position = "none",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold")) +
   scale_x_date(labels = date_format("%Y")) +
   {if(inc_H==TRUE){
      scale_color_manual(values = c("H"= "#F8766D", "U"= "#00BFC4",
                                     "HU" = "#7CAE00"), na.value="black")
   } else {
      scale_color_manual(values = c("U"= "#00BFC4"), na.value="black")
   }}
leg <- get_legend(p1)</pre>
pset <- ggarrange(leg, p1 + theme(legend.position = "none"), p2,</pre>
                  ncol = 1, heights = c(0.1, 1, 1))
p0 <- ggplot() + labs(title = "Scatter Plot for Entire Dataset") +</pre>
   theme_bw() + theme(plot.title = element_text(face="bold"),
                      panel.border = element_blank(),
                      panel.grid.major = element_blank(),
                      panel.grid.minor = element_blank(),
                      axis.line = element_blank())
ggarrange(p0, pset, ncol = 1, heights = c(0.1, 1))
```

Scatter Plot for Entire Dataset



Appendix II: Dataset Summary Box Plots

Box plots are created by using the entire data set and excludes any data that has been previously filtered out. The scripts that create plots follow this format

- 1. Use the data set that only has Use_In_Analysis of TRUE
- 2. Set what values are to be used for the x-axis, y-axis, and the variable that should determine groups for the box plots
- 3. Set the plot type as a box plot with the size of the outlier points
- 4. Create the title, x-axis, y-axis, and color fill labels
- 5. Set the v and x limits
- 6. Make the axis labels bold
- 7. Plot the arrangement as a set of panels

This set of box plots are grouped by year.

```
min RV <- min(data$ResultValue[data$Include == TRUE])</pre>
mn RV <- mean(data$ResultValue[data$Include == TRUE &</pre>
                                   data$ResultValue <
                                   quantile(data$ResultValue, 0.98)])
sd RV <- sd(data$ResultValue[data$Include == TRUE &</pre>
                                 data$ResultValue <</pre>
                                 quantile(data$ResultValue, 0.98)])
y_scale \leftarrow mn_RV + 4 * sd_RV
p1 <- ggplot(data = data[data$Include == TRUE, ],</pre>
             aes(x = Year, y = ResultValue, group = Year)) +
   geom_boxplot(outlier.size = 0.5) +
  labs(subtitle = "Autoscale", x = "Year",
        y = paste0("Values (", unit, ")")) +
   theme(axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p2 <- ggplot(data = data[data$Include == TRUE, ],</pre>
             aes(x = Year, y = ResultValue, group = Year)) +
   geom boxplot(outlier.size = 0.5) +
  labs(subtitle = "Scaled to 4x Standard Deviation", x = "Year",
        y = paste0("Values (", unit, ")")) +
   ylim(min_RV, y_scale) +
   theme(axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p3 <- ggplot(data = data[data$Include == TRUE, ],
             aes(x = as.integer(Year), y = ResultValue, group = Year)) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation, Last 10 Years",
        x = "Year", y = paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(max(data$Year) - 10.5, max(data$Year)+1),
                      breaks = seq(max(data$Year) - 10, max(data$Year), 2)) +
   theme(axis.text.x = element text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
```

This set of box plots are grouped by year and month with the color being related to the month.

```
p1 <- ggplot(data = data[data$Include == TRUE, ],</pre>
             aes(x = YearMonthDec, y = ResultValue,
                 group = YearMonth, color = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
  labs(subtitle = "Autoscale", x = "Year",
        y = paste0("Values (", unit, ")"), color="Month") +
   theme(legend.position = "top", legend.box = "horizontal",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold")) +
   guides(color = guide_legend(nrow = 1))
p2 <- ggplot(data = data[data$Include == TRUE, ],</pre>
             aes(x = YearMonthDec, y = ResultValue,
                 group = YearMonth, color = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation",
        x = "Year", y = paste0("Values (", unit, ")")) +
   ylim(min_RV, y_scale) +
   theme(legend.position = "none", axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p3 <- ggplot(data = data[data$Include == TRUE, ],
             aes(x = YearMonthDec, y = ResultValue,
                 group = YearMonth, color = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
  labs(subtitle = "Scaled to 4x Standard Deviation, Last 10 Years",
        x = "Year", y = paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(max(data$Year) - 10.5, max(data$Year)+1),
                      breaks = seq(max(data$Year) - 10, max(data$Year), 2)) +
   theme(legend.position = "none", axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
leg <- get_legend(p1)</pre>
set <- ggarrange(leg, p1 + theme(legend.position = "none"), p2, p3, ncol = 1,</pre>
                 heights = c(0.1, 1, 1, 1)
p0 <- ggplot() + labs(title = "Summary Box Plots for Entire Data",
                      subtitle = "By Year & Month") + theme_bw() +
   theme(plot.title = element_text(face="bold"),
         panel.border = element blank(), panel.grid.major = element blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank())
```

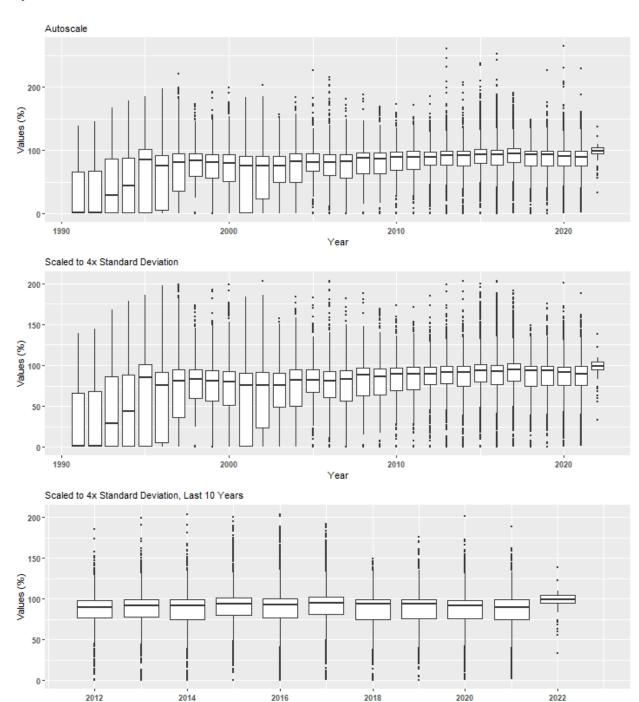
```
YMset <- ggarrange(p0, set, ncol=1, heights = c(0.07, 1))
```

The following box plots are grouped by month with fill color being related to the month. This is designed to view potential seasonal trends.

```
p1 <- ggplot(data = data[data$Include == TRUE, ],
             aes(x = Month, y = ResultValue,
                 group = Month, fill = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Autoscale", x = "Month",
        y = paste0("Values (", unit, ")"), fill="Month") +
   scale_x_continuous(limits = c(0, 13), breaks = seq(3, 12, 3)) +
   theme(legend.position = "top", legend.box = "horizontal",
         axis.text.x = element text(face = "bold"),
         axis.text.y = element_text(face = "bold")) +
   guides(fill = guide_legend(nrow = 1))
p2 <- ggplot(data = data[data$Include == TRUE, ],</pre>
             aes(x = Month, y = ResultValue,
                 group = Month, fill = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation",
        x = "Month", y = paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
   scale_x continuous(limits = c(0, 13), breaks = seq(3, 12, 3)) +
   theme(legend.position = "none", axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p3 <- ggplot(data = data[data$Include == TRUE &
                            data$Year >= max(data$Year) - 10, ],
             aes(x = Month, y = ResultValue,
                 group = Month, fill = as.factor(Month))) +
   geom boxplot(outlier.size = 0.5) +
  labs(subtitle = "Scaled to 4x Standard Deviation, Last 10 Years",
        x = "Month", y = paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(0, 13), breaks = seq(3, 12, 3)) +
   theme(legend.position = "none", axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
leg <- get_legend(p1)</pre>
set <- ggarrange(leg, p1 + theme(legend.position = "none"), p2, p3, ncol = 1,</pre>
                 heights = c(0.1, 1, 1, 1)
p0 <- ggplot() + labs(title = "Summary Box Plots for Entire Data",
                      subtitle = "By Month") + theme_bw() +
   theme(plot.title = element_text(face="bold"),
         panel.border = element_blank(), panel.grid.major = element_blank(),
         panel.grid.minor = element blank(), axis.line = element blank())
Mset <- ggarrange(p0, set, ncol=1, heights = c(0.07, 1))</pre>
```

Summary Box Plots for Entire Data

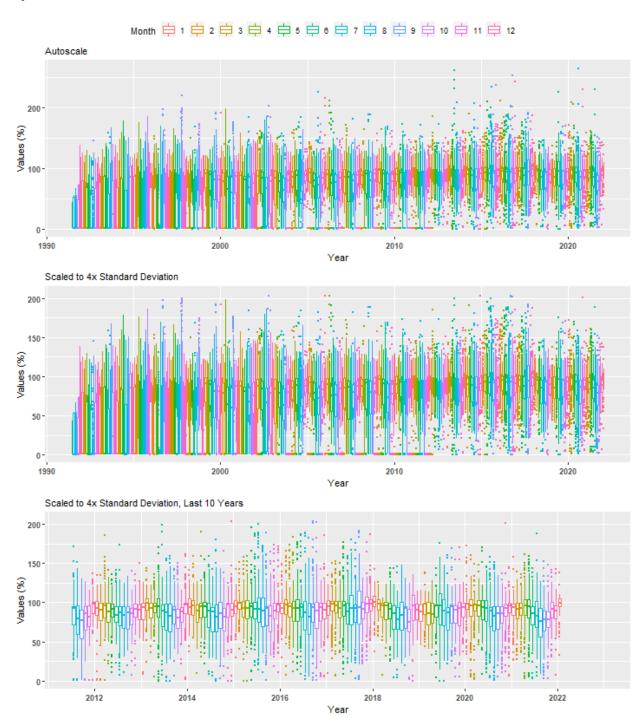
By Year



Year

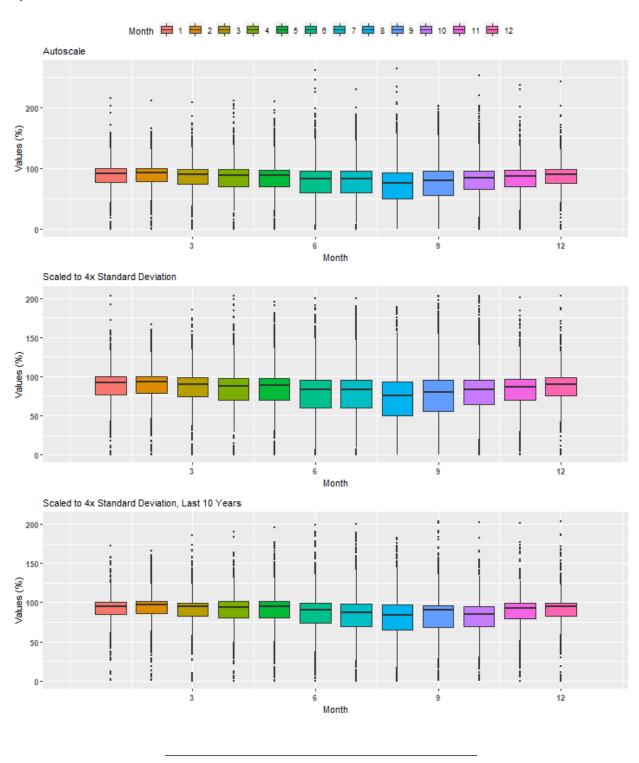
Summary Box Plots for Entire Data

By Year & Month



Summary Box Plots for Entire Data

By Month

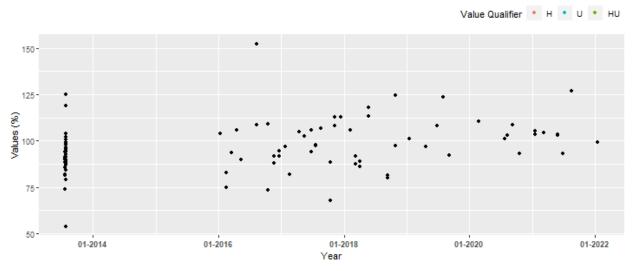


Appendix III: Excluded Managed Areas

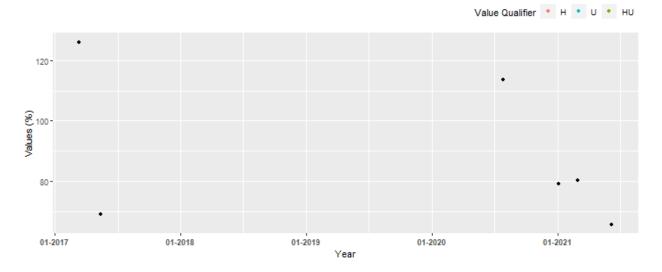
Scatter plots of data values are created for managed areas that have fewer than 10 separate years of data entries. Data points are colored based on specific value qualifiers of interest.

```
MA_Exclude <- MA_Years[MA_Years$Enough_Time==FALSE,]</pre>
MA_Exclude <- MA_Exclude[order(MA_Exclude$ManagedAreaName),]</pre>
z=length(MA Exclude$ManagedAreaName)
if(z==0){
  print("There are no managed areas that qualify.")
} else {
  for(i in 1:z){
      p1<-ggplot(data=data[data$ManagedAreaName==MA_Exclude$ManagedAreaName[i]&
                              data$Include == TRUE, ],
                 aes(x = SampleDate, y = ResultValue, color=VQ_Plot)) +
         geom_point() +
         labs(title = paste0("Scatter Plot of Excluded Managed Area\n",
                             MA_Exclude$ManagedAreaName[i], " (",
                             MA_Exclude$N[i], " Unique Years)"),
              subtitle="Autoscale", x = "Year",
              y = paste0("Values (", unit, ")"), color="Value Qualifier") +
         theme(legend.position = "top", legend.box = "horizontal",
               legend.justification = "right",
               axis.text.x = element_text(face = "bold")) +
         scale_x_date(labels = date_format("%m-%Y")) +
         {if(inc_H==TRUE){
            scale_color_manual(values = c("H"= "#F8766D", "U"= "#00BFC4",
                                           "HU" = "#7CAE00"), na.value="black")
            scale_color_manual(values = c("U"= "#00BFC4"), na.value="black")
         }}
      print(p1)
  }
```

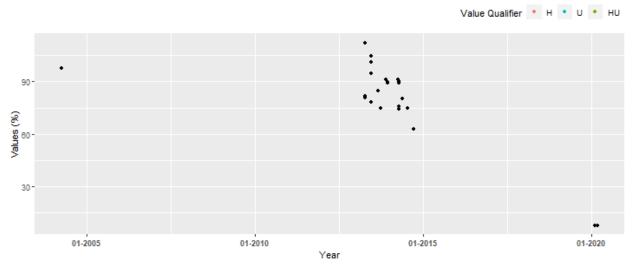
Scatter Plot of Excluded Managed Area Alligator Harbor Aquatic Preserve (8 Unique Years) Autoscale



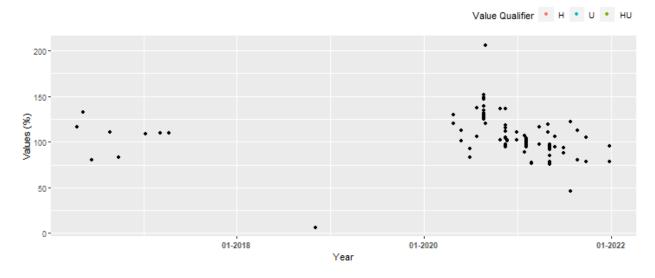
Scatter Plot of Excluded Managed Area Cape Haze Aquatic Preserve (3 Unique Years) Autoscale



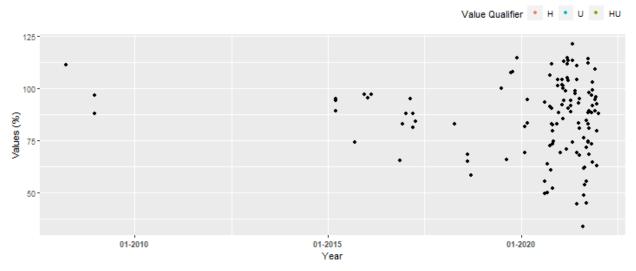
Scatter Plot of Excluded Managed Area Fort Clinch State Park Aquatic Preserve (4 Unique Years)



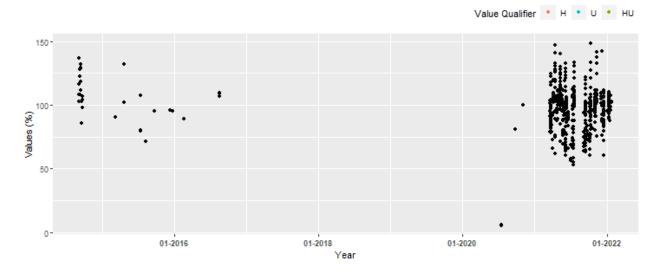
Scatter Plot of Excluded Managed Area Lignumvitae Key Aquatic Preserve (5 Unique Years) Autoscale



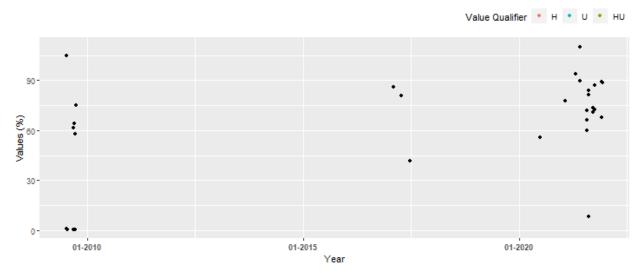
Scatter Plot of Excluded Managed Area Matlacha Pass Aquatic Preserve (9 Unique Years) Autoscale



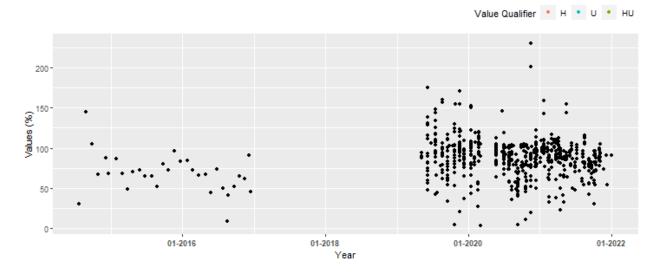
Scatter Plot of Excluded Managed Area Nature Coast Aquatic Preserve (6 Unique Years) Autoscale



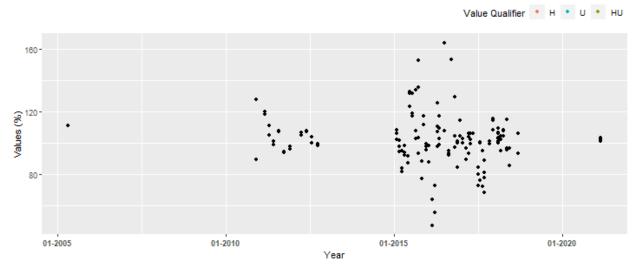
Scatter Plot of Excluded Managed Area North Fork St. Lucie Aquatic Preserve (4 Unique Years) Autoscale



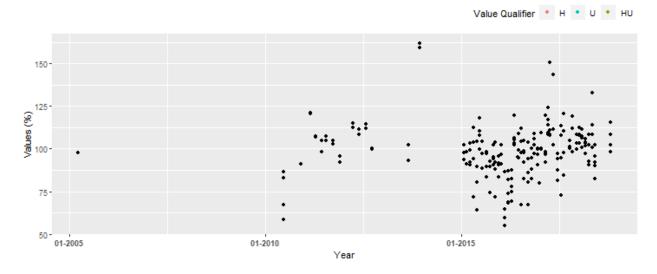
Scatter Plot of Excluded Managed Area Pine Island Sound Aquatic Preserve (7 Unique Years) Autoscale



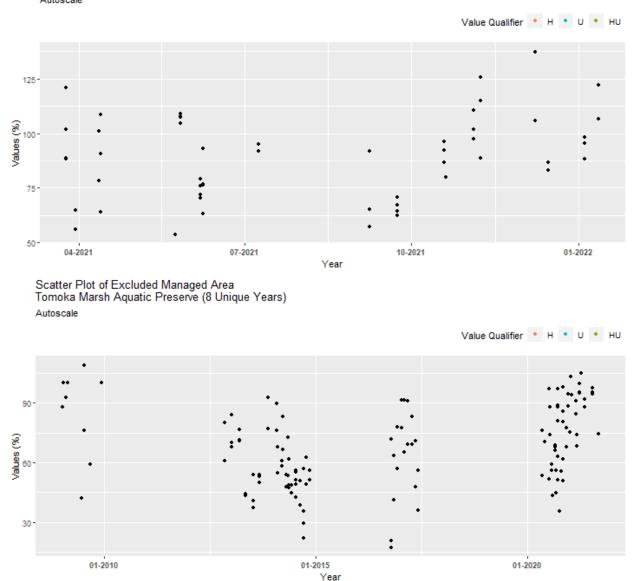
Scatter Plot of Excluded Managed Area St. Andrews State Park Aquatic Preserve (9 Unique Years) Autoscale



Scatter Plot of Excluded Managed Area St. Joseph Bay Aquatic Preserve (9 Unique Years) Autoscale



Scatter Plot of Excluded Managed Area St. Martins Marsh Aquatic Preserve (2 Unique Years)



Appendix IV: Managed Area Trendlines

The plots created in this section are designed to show the general trend of the data. Data is taken and grouped by ManagedAreaName. The trendlines on the plots are created using the Senn slope and intercept from the seasonal Kendall Tau analysis. The scripts that create plots follow this format

- 1. Use the data set that only has Use_In_Analysis of TRUE for the desired managed area
- 2. Determine the earliest and latest year of the data to create x-axis scale and intervals
- 3. Determine the minimum, mean, and standard deviation for the data to be used for y-axis scales

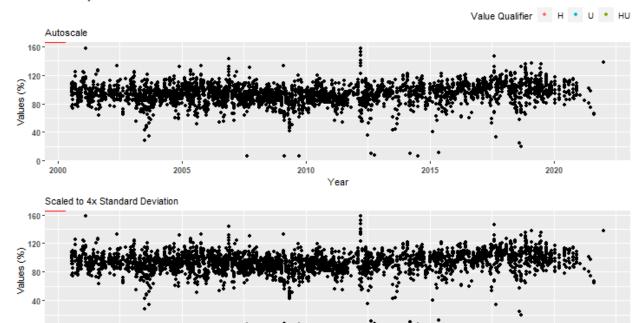
- Excludes the top 2% of values to reduce the impact of extreme outliers on the y-axis scale
- 4. Set what values are to be used for the x-axis, y-axis, and the variable that should determine groups for the plots
- 5. Set the plot type as a point plot with the size of the points
- 6. Add the linear trend
- 7. Create the title, x-axis, y-axis, and color fill labels
- 8. Set the v and x limits
- 9. Make the axis labels bold
- 10. Plot the arrangement as a set of panels

```
if(n==0){
   print("There are no managed areas that qualify.")
} else {
   for (i in 1:n) {
      plot data <- data[data$Use In Analysis == TRUE &</pre>
                                       data$ManagedAreaName == MA names[i],]
      year_lower <- min(plot_data$Year)</pre>
      year upper <- max(plot data$Year)</pre>
      min_RV <- min(plot_data$ResultValue)</pre>
      mn_RV <- mean(plot_data$ResultValue[data$ResultValue <</pre>
                                          quantile(data$ResultValue, 0.98)])
      sd_RV <- sd(plot_data$ResultValue[data$ResultValue <</pre>
                                        quantile(data$ResultValue, 0.98)])
      x_scale <- ifelse(year_upper - year_lower > 30, 10, 5)
      y_scale <- mn_RV + 4 * sd_RV</pre>
      tau <- KT.Stats$tau[KT.Stats$ManagedAreaName==MA_names[i]]</pre>
      s slope <- KT.Stats$SennSlope[KT.Stats$ManagedAreaName==MA names[i]]
      s_int <- KT.Stats$SennIntercept[KT.Stats$ManagedAreaName==MA_names[i]]</pre>
      trend <- KT.Stats$Trend[KT.Stats$ManagedAreaName==MA_names[i]]</pre>
      p <- KT.Stats$p[KT.Stats$ManagedAreaName==MA_names[i]]</pre>
      model <- lm(ResultValue ~ DecDate,
                   data = plot_data)
      m int <- coef(model)[[1]]</pre>
      m_slope <- coef(model)[[2]]</pre>
      p1 <- ggplot(data = plot_data,</pre>
                    aes(x = DecDate, y = ResultValue,
                        color=VQ Plot)) +
         geom_point(size = 1.5) +
         geom_abline(aes(slope=s_slope, intercept=s_int),
                      color="red", size=1.5) +
         labs(subtitle = "Autoscale",
               x = "Year", y = paste0("Values (", unit, ")"),
               color="Value Qualifier") +
         theme(legend.position = "top", legend.box = "horizontal",
               legend.justification = "right",
               axis.text.x = element_text(face = "bold"),
                axis.text.y = element_text(face="bold")) +
         {if(inc H==TRUE){
            scale_color_manual(values = c("H"= "#F8766D", "U"= "#00BFC4",
                                             "HU" = "#7CAE00"), na.value="black")
         } else {
```

```
scale_color_manual(values = c("U"= "#00BFC4"), na.value="black")
         }}
      p2 <- ggplot(data = plot_data,</pre>
                   aes(x = DecDate, y = ResultValue,
                       color=VQ_Plot)) +
         geom_point(size = 1.5) +
         geom_abline(aes(slope=s_slope, intercept=s_int),
                     color="red", size=1.5) +
         ylim(min_RV, y_scale) +
         labs(subtitle = "Scaled to 4x Standard Deviation",
              x = "Year", y = paste0("Values (", unit, ")")) +
         theme(legend.position = "none",
               axis.text.x = element_text(face = "bold"),
               axis.text.y = element_text(face="bold")) +
         {if(inc_H==TRUE){
            scale_color_manual(values = c("H"= "#F8766D", "U"= "#00BFC4",
                                           "HU" = "#7CAE00"), na.value="black")
         } else {
            scale_color_manual(values = c("U"= "#00BFC4"), na.value="black")
      leg <- get_legend(p1)</pre>
      KTset <- ggarrange(leg, p1 + theme(legend.position = "none"), p2,</pre>
                         ncol = 1, heights = c(0.1, 1, 1))
      p0 <- ggplot() + labs(title = paste0("Data Points with Trendlines for ",</pre>
                                            MA_names[i]),
                             subtitle =paste0("Senn Slope = ", s_slope,
                                                     Senn Intercept = ", s_int,
                                              "\nTrend = ", trend,
                                              ", tau = ", tau,
                                                    p = ", p,
                                              "\nLinear Trendline: ",
                                              "y = ", m_slope, "x + ", m_int)) +
         theme_bw() + theme(plot.title = element_text(face="bold"),
                             panel.border = element_blank(),
                             panel.grid.major = element_blank(),
                             panel.grid.minor = element_blank(),
                             axis.line = element_blank())
      print(ggarrange(p0, KTset, ncol = 1, heights = c(0.15, 1)))
  }
}
```

Data Points with Trendlines for Apalachicola Bay Aquatic Preserve

Senn Slope = 0.50833333333333334, Senn Intercept = -848.780844155844 Trend = 1, tau = 0.1434, p = 0.1444, p = 0.1444



2010

Year

2015

2020

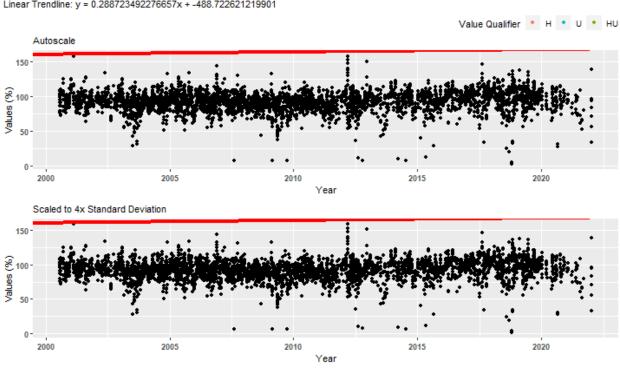
Data Points with Trendlines for Apalachicola National Estuarine Research Reserve

Senn Slope = 0.350000000000001, Senn Intercept = -539.158333333334

2005

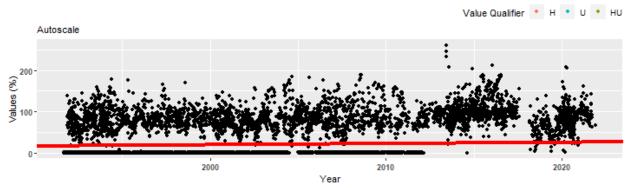
2000

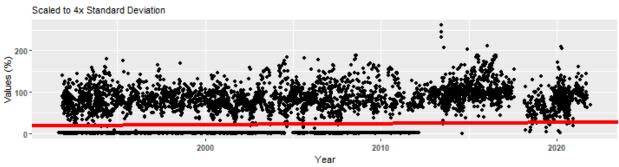
Trend = 1, tau = 0.0966, p = 0 Linear Trendline: y = 0.288723492276657x + -488.722621219901



Data Points with Trendlines for Banana River Aquatic Preserve

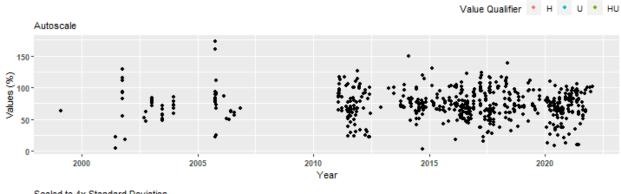
Senn Slope = 0.275, Senn Intercept = -528.937435185185Trend = 1, tau = 0.2121, p = 0 Linear Trendline: y = 2.1803374096097x + -4323.37044577855

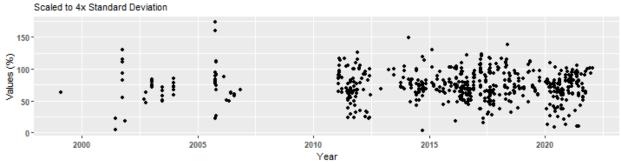




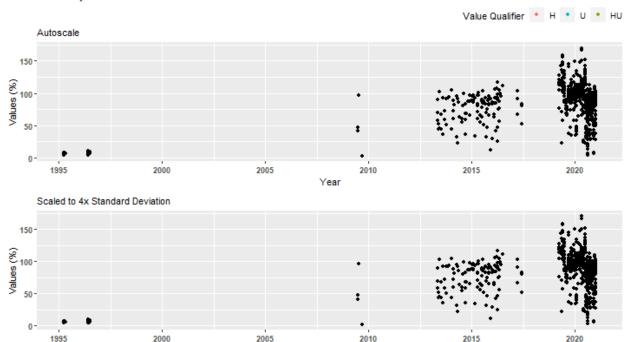
Data Points with Trendlines for Big Bend Seagrasses Aquatic Preserve

Senn Slope = 0.05, Senn Intercept = -569.980128205128 Trend = 0, tau = 0.0221, p = 0.7697 Linear Trendline: y = -0.118210038464142x + 312.308691233918



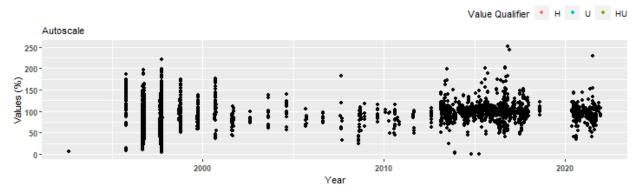


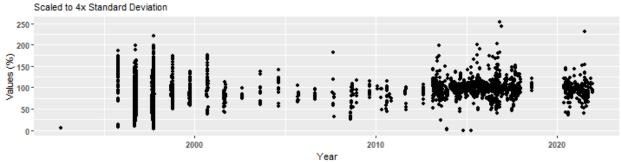
Data Points with Trendlines for Biscayne Bay Aquatic Preserve



Data Points with Trendlines for Boca Ciega Bay Aquatic Preserve

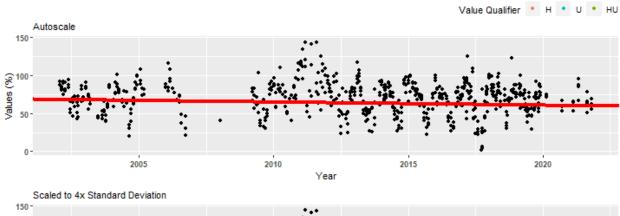
Senn Slope = 0.647058823529412, Senn Intercept = 405.92499999998 Trend = 1, tau = 0.0407, p = 0 Linear Trendline: y = 0.586899341573524x + -1086.42498934385

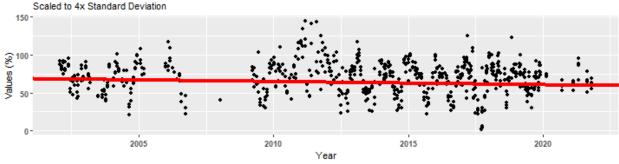




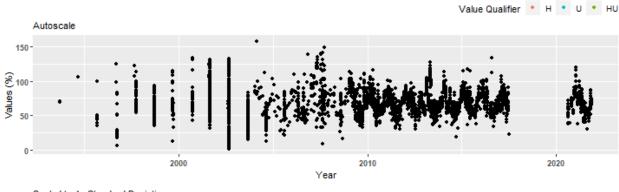
Data Points with Trendlines for Cape Romano-Ten Thousand Islands Aquatic Preserve

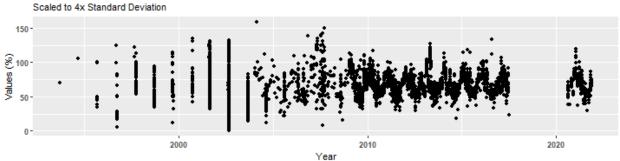
Senn Slope = -0.41, Senn Intercept = 889.43333333335 Trend = -1, tau = -0.0814, p = 0.0001 Linear Trendline: y = -0.455957402265215x + 988.582203319446





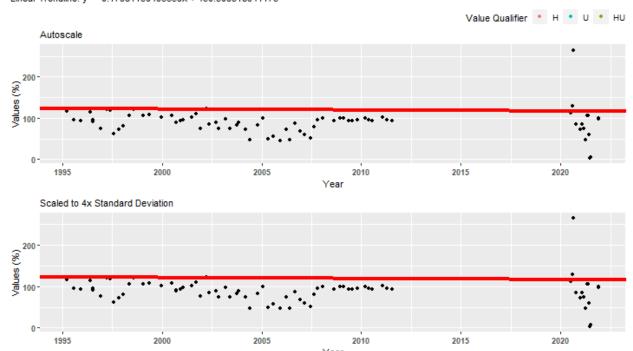
Data Points with Trendlines for Cockroach Bay Aquatic Preserve





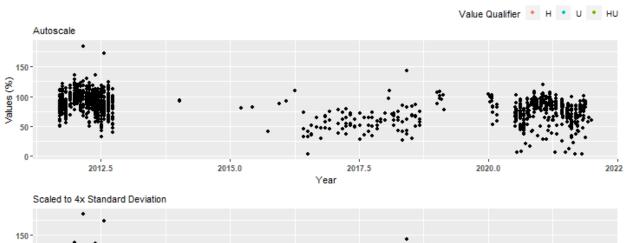
Data Points with Trendlines for Coupon Bight Aquatic Preserve

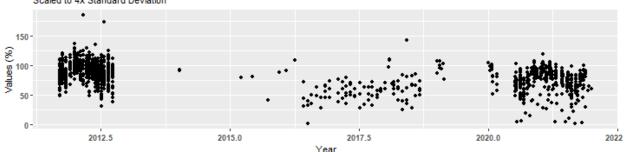
Senn Slope = -0.235259265652174, Senn Intercept = 593.851891364668 Trend = 0, tau = -0.0768, p = 0.3152 Linear Trendline: y = -0.17981130496659x + 450.509318941173



Data Points with Trendlines for Estero Bay Aquatic Preserve

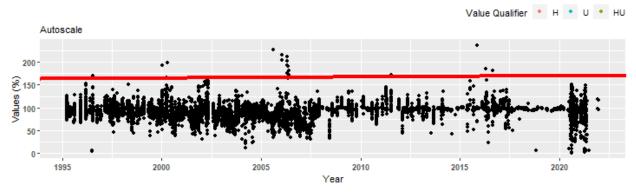
Senn Slope = -1.7, Senn Intercept = 3048.23573333333 Trend = -1, tau = -0.2393, p = 0 Linear Trendline: y = -1.94230477210825x + 3997.91994162807

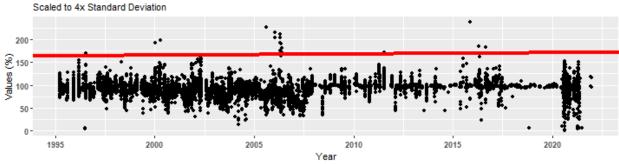




Data Points with Trendlines for Florida Keys National Marine Sanctuary

Senn Slope = 0.270295072000002, Senn Intercept = -375.078394267913 Trend = 1, tau = 0.1734, p = 0.335546742978506x + <math>-582.340812668057



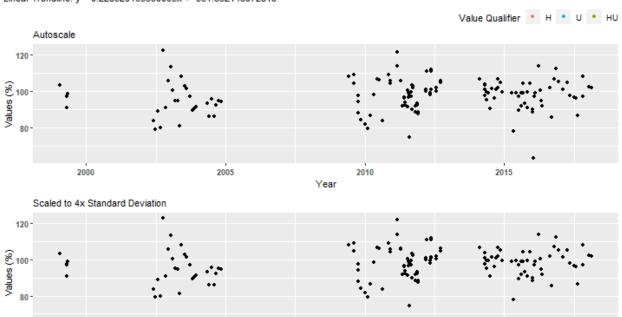


Data Points with Trendlines for Fort Pickens State Park Aquatic Preserve

2005

Senn Slope = 0.202941176470588, Senn Intercept = -216.188833333334 Trend = 0, tau = 0.0839, p = 0.2165 Linear Trendline: y = 0.228529166330095x + -361.832148972819

2000



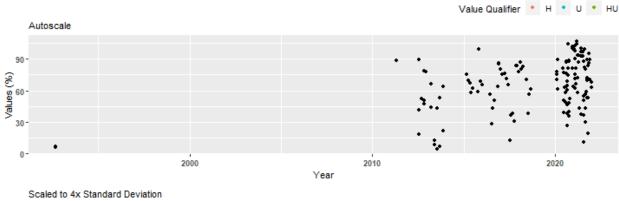
Year

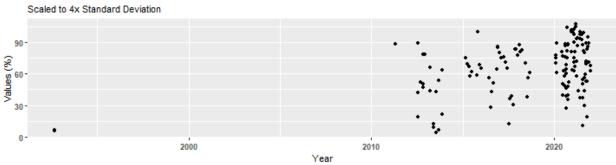
2010

2015

Data Points with Trendlines for Gasparilla Sound-Charlotte Harbor Aquatic Preserve

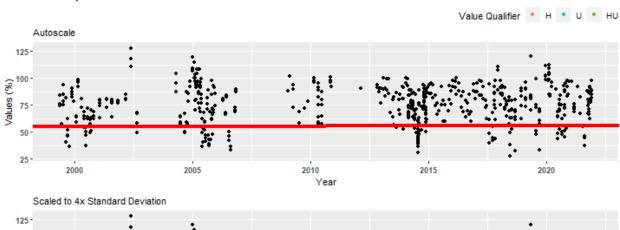
Senn Slope = 2.5, Senn Intercept = -4590.3715979064Trend = 1, tau = 0.2321, p = 0.0002 Linear Trendline: y = 2.3847682216714x + -4748.72085113039

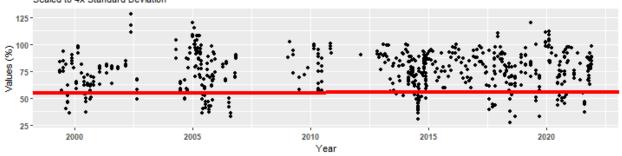




Data Points with Trendlines for Guana River Marsh Aquatic Preserve

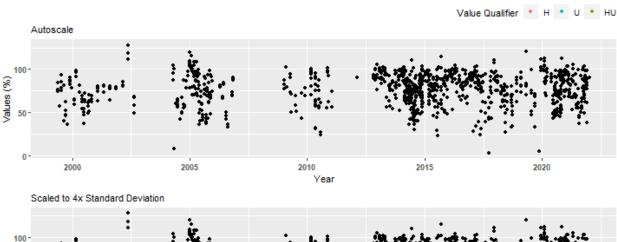
Senn Slope = 0.02500000000000004, Senn Intercept = 5.238461538462 Trend = 0, tau = 0.0182, p = 0.7075 Linear Trendline: y = 0.150219995290366x + -227.221841073999

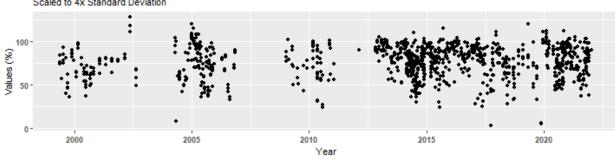




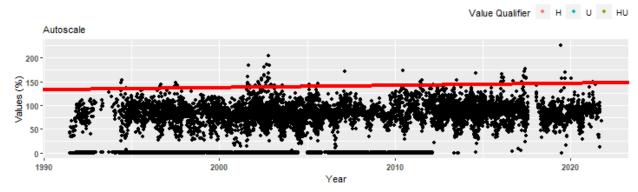
Data Points with Trendlines for Guana Tolomato Matanzas National Estuarine Research Reserve

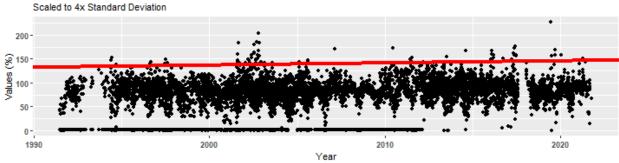
Senn Slope = 0.004545454545454545461, Senn Intercept = -156.669322344322 Trend = 0.0036, p = 0.8743 Linear Trendline: y = 0.182203665930078x + -290.309117675694





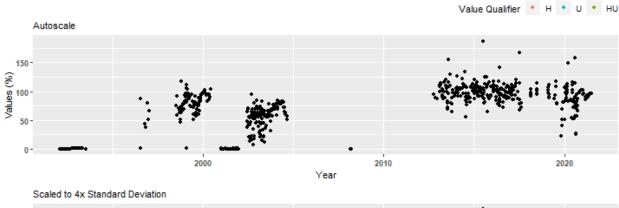
Data Points with Trendlines for Indian River-Malabar to Vero Beach Aquatic Preserve

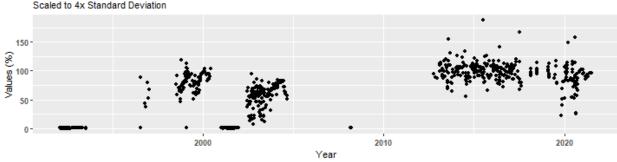




Data Points with Trendlines for Indian River-Vero Beach to Ft. Pierce Aquatic Preserve

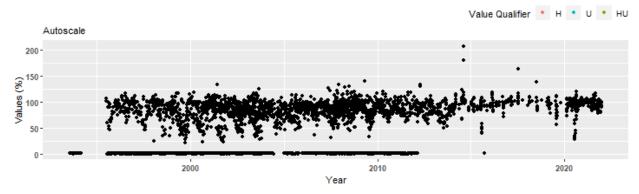
Senn Slope = 2.44058823529412, Senn Intercept = -4618.6966666667 Trend = 1, tau = 0.4343, p = 0 Linear Trendline: y = 2.91450638779029x + -5782.34277963125

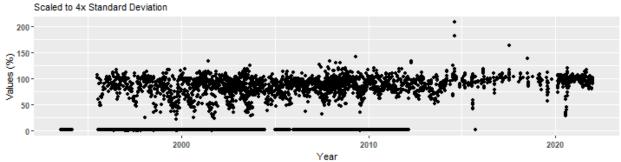




Data Points with Trendlines for Jensen Beach to Jupiter Inlet Aquatic Preserve

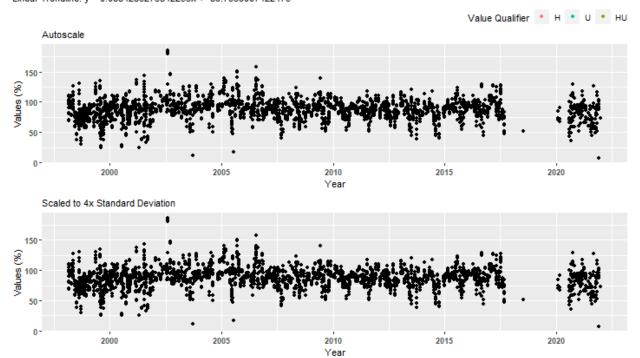
Senn Slope = 0.84375, Senn Intercept = -1279.85892941176Trend = 1, tau = 0.2324, p = 0Linear Trendline: y = 2.5388644510096x + -5038.21781560281





Data Points with Trendlines for Lemon Bay Aquatic Preserve

Senn Slope = 0.171428571428572, Senn Intercept = -414.5 Trend = 1, tau = 0.059, p = 0 Linear Trendline: y = 0.0884286276842253x + -89.7369907422475

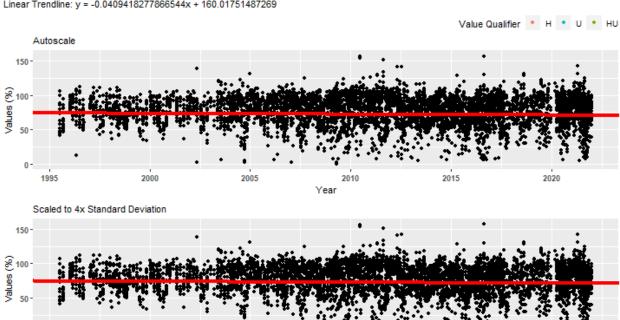


Data Points with Trendlines for Loxahatchee River-Lake Worth Creek Aquatic Preserve

Senn Slope = -0.122222222222222222, Senn Intercept = 318.863727272728 Trend = -1, tau = -0.0385, p = 0.0003 Linear Trendline: y = -0.0409418277866544x + 160.01751487269

2000

0 -1995



2010

Year

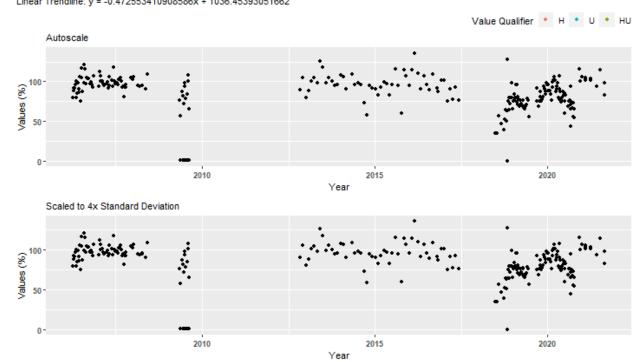
2015

2020

2005

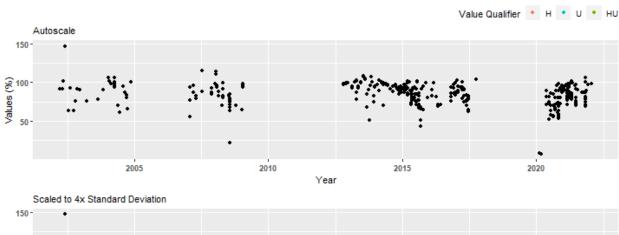
Data Points with Trendlines for Mosquito Lagoon Aquatic Preserve

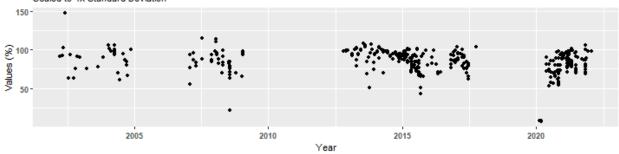
Senn Slope = -0.909401709401709, Senn Intercept = 1683.46482948718 Trend = -1, tau = -0.1853, p = 0.0001 Linear Trendline: y = -0.472553410908586x + 1036.45393051662



Data Points with Trendlines for Nassau River-St. Johns River Marshes Aquatic Preserve

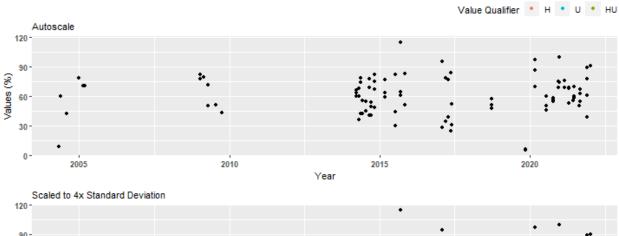
Senn Slope = -0.523529411764706, Senn Intercept = 928.321428571428 Trend = -1, tau = -0.161, p = 0 Linear Trendline: y = -0.645841248051005x + 1386.48877038216

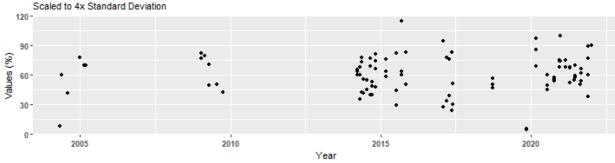




Data Points with Trendlines for Pellicer Creek Aquatic Preserve

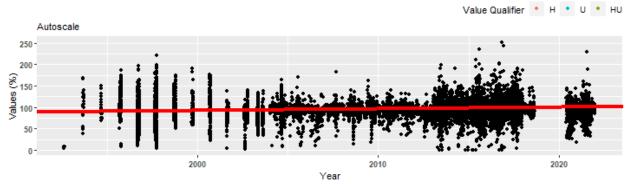
Senn Slope = 0.84659090909090909, Senn Intercept = -1271.55101809955 Trend = 0, tau = 0.1038, p = 0.1515 Linear Trendline: y = 0.316354834786362x + -577.421974453245

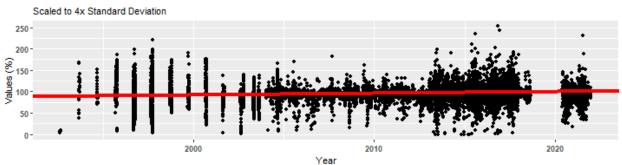




Data Points with Trendlines for Pinellas County Aquatic Preserve

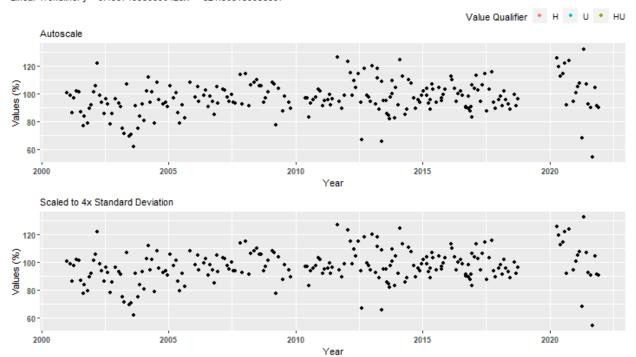
Senn Slope = 0.392307692307692, Senn Intercept = -691.04090909091 Trend = 1, tau = 0.0837, p = 0 Linear Trendline: y = 0.41344549943314x + -739.162870124131





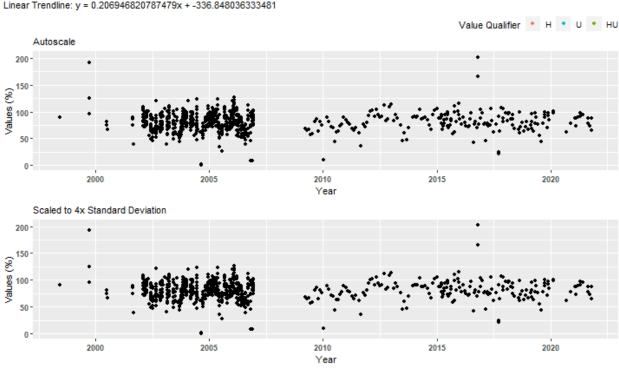
Data Points with Trendlines for Rocky Bayou State Park Aquatic Preserve

Senn Slope = 0.400000000000000006, Senn Intercept = -920.317222222222 Trend = 1, tau = 0.1356, p = 0.0006 Linear Trendline: y = 0.456746350690426x + -821.303189958597



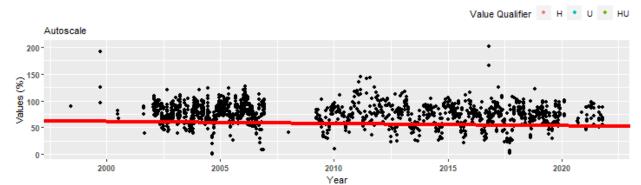
Data Points with Trendlines for Rookery Bay Aquatic Preserve

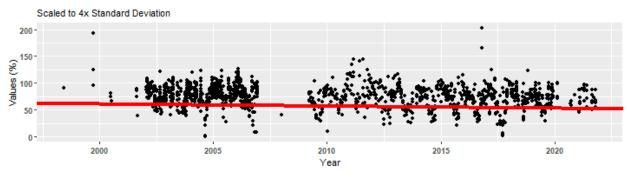
Senn Slope = 0.049999999999995, Senn Intercept = -127.759285714286 Trend = 0, tau = 0.0146, p = 0.5961 Linear Trendline: y = 0.206946820787479x + -336.848036333481



Data Points with Trendlines for Rookery Bay National Estuarine Research Reserve

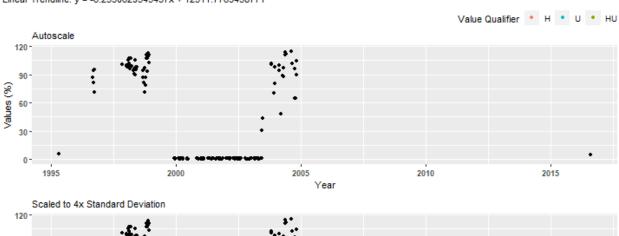
Senn Slope = -0.401, Senn Intercept = 863.532053571427Trend = -1, tau = -0.0956, p = 0 Linear Trendline: y = -0.410805448989987x + 900.339058087688

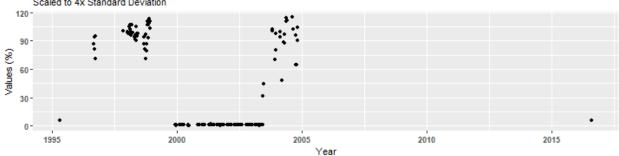




Data Points with Trendlines for Southeast Florida Coral Reef Ecosystem Conservation Area

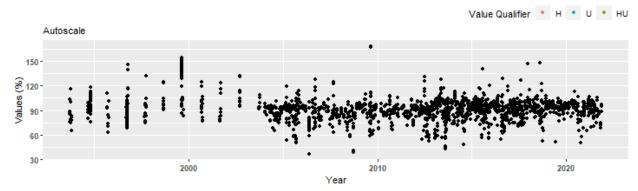
Senn Slope = -0.1110666666666667, Senn Intercept = 365.011225 Trend = 0, tau = -0.0936, p = 0.0634 Linear Trendline: y = -6.2330629545437x + 12511.7763458771

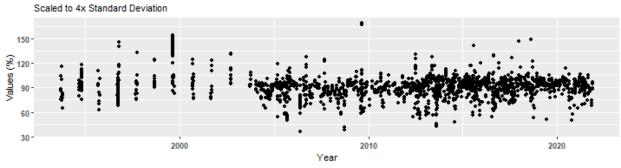




Data Points with Trendlines for Terra Ceia Aquatic Preserve

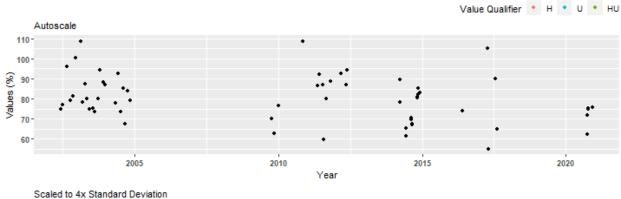
Senn Slope = 0, Senn Intercept = -421.323928571428 Trend = 0, tau = 0.0463, p = 0.9935 Linear Trendline: y = -0.13880492271777x + 370.342199944609

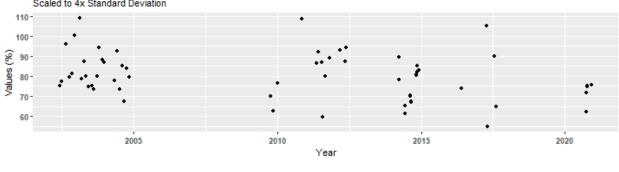




Data Points with Trendlines for Yellow River Marsh Aquatic Preserve

Senn Slope = -0.8220583333333333, Senn Intercept = 1513.7740288961Trend = -1, tau = -0.2586, p = 0.0078Linear Trendline: y = -0.549980048023466x + 1185.81187897948





Appendix V: Managed Area Summary Box Plots

Data is taken and grouped by ManagedAreaName. The scripts that create plots follow this format

- 1. Use the data set that only has Use_In_Analysis of TRUE for the desired managed area
- 2. Determine the earliest and latest year of the data to create x-axis scale and intervals
- 3. Determine the minimum, mean, and standard deviation for the data to be used for y-axis scales
 - Excludes the top 2% of values to reduce the impact of extreme outliers on the y-axis scale
- 4. Set what values are to be used for the x-axis, y-axis, and the variable that should determine groups for the box plots
- 5. Set the plot type as a box plot with the size of the outlier points
- 6. Create the title, x-axis, y-axis, and color fill labels
- 7. Set the y and x limits
- 8. Make the axis labels bold
- 9. Plot the arrangement as a set of panels

The following plots are arranged by ManagedAreaName with data grouped by Year, then Year and Month, then finally Month only. Each managed area will have 3 sets of plots, each with 3 panels in them. Each panel goes as follows:

- 1. Y-axis autoscaled
- 2. Y-axis set to be mean + 5 time the standard deviation
- 3. Y-axis set to be mean + 5 time the standard deviation for most recent 10 years of data

```
if(n==0){
   print("There are no managed areas that qualify.")
} else {
   for (i in 1:n) {
      year_lower <- min(data$Year[data$Use_In_Analysis == TRUE &</pre>
                                       data$ManagedAreaName == MA_names[i]])
      year_upper <- max(data$Year[data$Use_In_Analysis == TRUE &</pre>
                                       data$ManagedAreaName == MA_names[i]])
      min_RV \leftarrow min(data\$ResultValue[data\$Use_In_Analysis == TRUE &
                                          data$ManagedAreaName == MA names[i]])
      mn_RV <- mean(data$ResultValue[data$Use_In_Analysis == TRUE &</pre>
                                          data$ManagedAreaName == MA names[i] &
                                          data$ResultValue <</pre>
                                          quantile(data$ResultValue, 0.98)])
      sd_RV <- sd(data$ResultValue[data$Use_In_Analysis == TRUE &</pre>
                                        data$ManagedAreaName == MA_names[i] &
                                        data$ResultValue <</pre>
                                        quantile(data$ResultValue, 0.98)])
      x_scale <- ifelse(year_upper - year_lower > 30, 10, 5)
      y_scale <- mn_RV + 4 * sd_RV</pre>
      ##Year plots
      p1 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                                   data$ManagedAreaName == MA_names[i], ],
                    aes(x = Year, y = ResultValue, group = Year)) +
         geom_boxplot(outlier.size = 0.5) +
         labs(subtitle = "Autoscale",
              x = "Year", y = paste0("Values (", unit, ")")) +
```

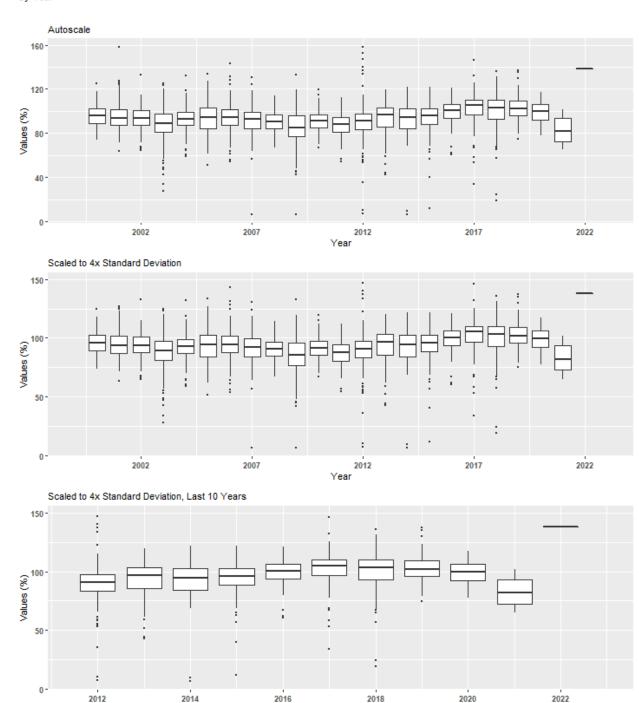
```
scale_x_continuous(limits = c(year_lower - 1, year_upper + 1),
                      breaks = rev(seq(year_upper,
                                       year_lower, -x_scale))) +
   theme(axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p2 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                            data$ManagedAreaName == MA names[i], ],
             aes(x = Year, y = ResultValue, group = Year)) +
   geom boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation",
        x = "Year", y = paste0("Values (", unit, ")")) +
   ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(year_lower - 1, year_upper + 1),
                      breaks = rev(seq(year_upper,
                                       year_lower, -x_scale))) +
   theme(axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p3 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                            data$ManagedAreaName == MA_names[i] &
                            data$Year>=year_upper-10, ],
             aes(x = Year, y = ResultValue, group = Year)) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation, Last 10 Years",
        x = "Year", y = paste0("Values (", unit, ")")) +
   ylim(min RV, y scale) +
   scale_x_continuous(limits = c(year_upper - 10.5, year_upper + 1),
                      breaks = rev(seq(year_upper, year_upper - 10,-2))) +
   theme(axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
Yset <- ggarrange(p1, p2, p3, ncol = 1)</pre>
p0 <- ggplot() + labs(title = paste0("Summary Box Plots for ",
                                     MA_names[i]), subtitle = "By Year") +
   theme_bw() + theme(plot.title = element_text(face="bold"),
                      panel.border = element_blank(),
                      panel.grid.major = element_blank(),
                      panel.grid.minor = element_blank(), axis.line = element_blank())
## Year & Month Plots
p4 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                            data$ManagedAreaName == MA_names[i], ],
             aes(x = YearMonthDec, y = ResultValue,
                 group = YearMonth, color = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Autoscale",
        x = "Year", y = pasteO("Values (", unit, ")"), color = "Month") +
   scale_x_continuous(limits = c(year_lower - 1, year_upper + 1),
                      breaks = rev(seq(year_upper,
                                       year_lower, -x_scale))) +
```

```
theme(legend.position = "none",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p5 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                            data$ManagedAreaName == MA names[i], ],
             aes(x = YearMonthDec, y = ResultValue,
                 group = YearMonth, color = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation",
        x = "Year", y = paste0("Values (", unit, ")"), color = "Month") +
   ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(year_lower - 1, year_upper + 1),
                      breaks = rev(seq(year_upper,
                                       year_lower, -x_scale))) +
   theme(legend.position = "top", legend.box = "horizontal",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold")) +
   guides(color = guide_legend(nrow = 1))
p6 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                            data$ManagedAreaName == MA_names[i], ],
             aes(x = YearMonthDec, y = ResultValue,
                 group = YearMonth, color = as.factor(Month)
             )) +
   geom boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation, Last 10 Years",
        x = "Year", y = paste0("Values (", unit, ")"), color = "Month") +
   ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(year_upper - 10.5, year_upper + 1),
                      breaks = rev(seq(year_upper, year_upper - 10,-2))) +
   theme(legend.position = "none",
        axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
leg1 <- get_legend(p5)</pre>
YMset <- ggarrange(leg1, p4, p5 + theme(legend.position = "none"), p6,
                   ncol = 1, heights = c(0.1, 1, 1, 1)
p00 <- ggplot() + labs(title = paste0("Summary Box Plots for ",
                                      MA_names[i]),
                       subtitle = "By Year & Month") + theme_bw() +
   theme(plot.title = element_text(face="bold"),
         panel.border = element_blank(),
         panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), axis.line = element_blank())
## Month Plots
p7 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                            data$ManagedAreaName == MA_names[i], ],
             aes(x = Month, y = ResultValue,
                 group = Month, fill = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
```

```
labs(subtitle = "Autoscale",
        x = "Month", y = paste0("Values (", unit, ")"), fill = "Month") +
   scale_x_continuous(limits = c(0, 13), breaks = seq(3, 12, 3)) +
   theme(legend.position = "none",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
p8 <- ggplot(data = data[data$Use In Analysis == TRUE &
                            data$ManagedAreaName == MA_names[i], ],
             aes(x = Month, y = ResultValue,
                 group = Month, fill = as.factor(Month))) +
   geom_boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation",
        x = "Month", y = paste0("Values (", unit, ")"), fill = "Month") +
   ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(0, 13), breaks = seq(3, 12, 3)) +
   theme(legend.position = "top", legend.box = "horizontal",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold")) +
   guides(fill = guide_legend(nrow = 1))
p9 <- ggplot(data = data[data$Use_In_Analysis == TRUE &
                            data$ManagedAreaName == MA names[i] &
                            data$Year >= year_upper - 10, ],
             aes(x = Month, y = ResultValue,
                 group = Month, fill = as.factor(Month))) +
   geom boxplot(outlier.size = 0.5) +
   labs(subtitle = "Scaled to 4x Standard Deviation, Last 10 Years",
        x = "Month", y = paste0("Values (", unit, ")"), fill = "Month") +
   ylim(min_RV, y_scale) +
   scale_x_continuous(limits = c(0, 13), breaks = seq(3, 12, 3)) +
   theme(legend.position = "none",
         axis.text.x = element_text(face = "bold"),
         axis.text.y = element_text(face = "bold"))
leg2 <- get_legend(p8)</pre>
Mset <- ggarrange(leg2, p7, p8 + theme(legend.position = "none"), p9,</pre>
                  ncol = 1, heights = c(0.1, 1, 1, 1)
p000 <- ggplot() + labs(title = paste0("Summary Box Plots for ",</pre>
                                       MA_names[i]),
                        subtitle = "By Month") + theme bw() +
   theme(plot.title = element text(face="bold"),
         panel.border = element_blank(),
         panel.grid.major = element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank())
print(ggarrange(p0, Yset, ncol = 1, heights = c(0.07, 1)))
print(ggarrange(p00, YMset, ncol = 1, heights = c(0.07, 1)))
print(ggarrange(p000, Mset, ncol = 1, heights = c(0.07, 1, 0.7)))
```

}

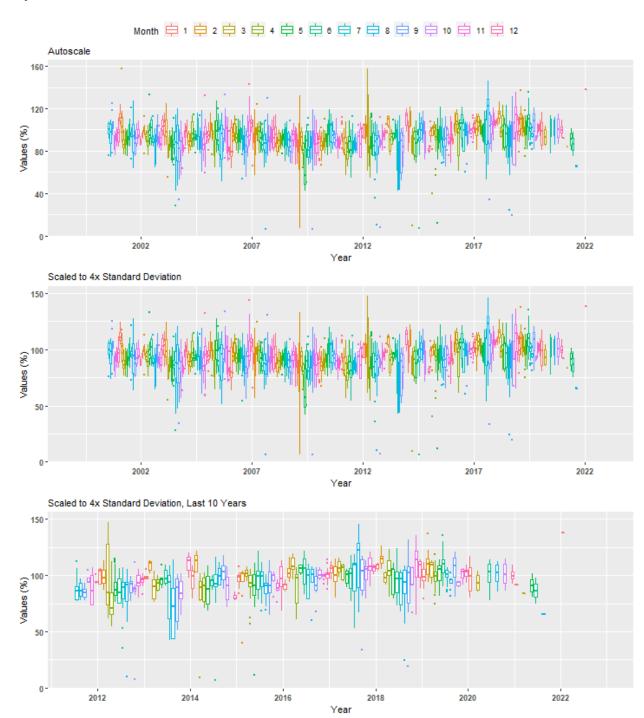
Summary Box Plots for Apalachicola Bay Aquatic Preserve $\ensuremath{\mathsf{By\,Year}}$



Year

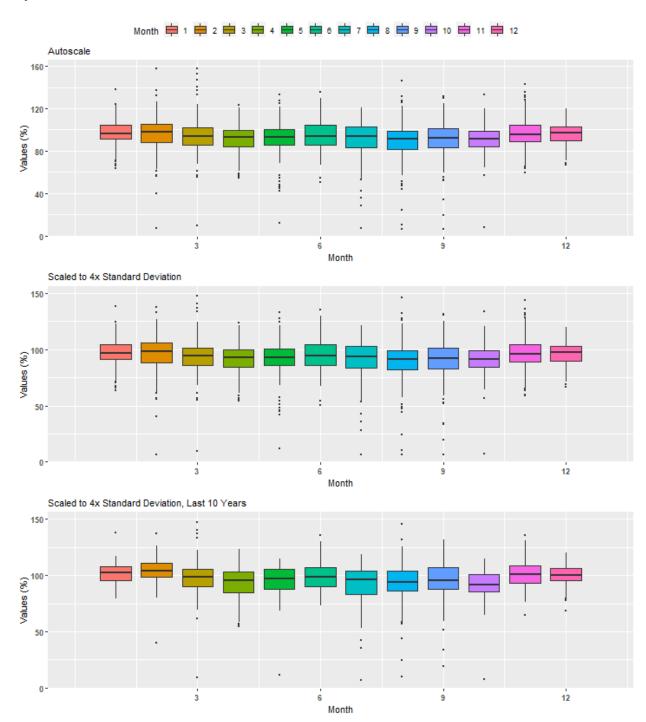
Summary Box Plots for Apalachicola Bay Aquatic Preserve

By Year & Month

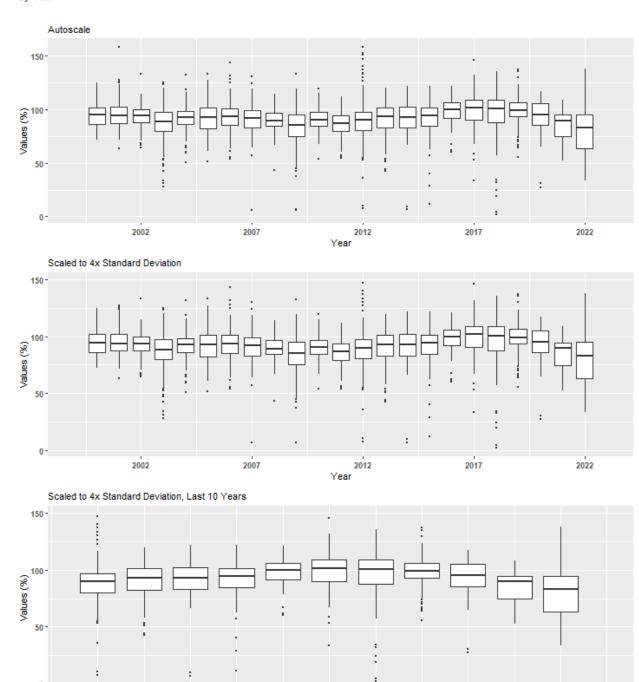


Summary Box Plots for Apalachicola Bay Aquatic Preserve

By Month



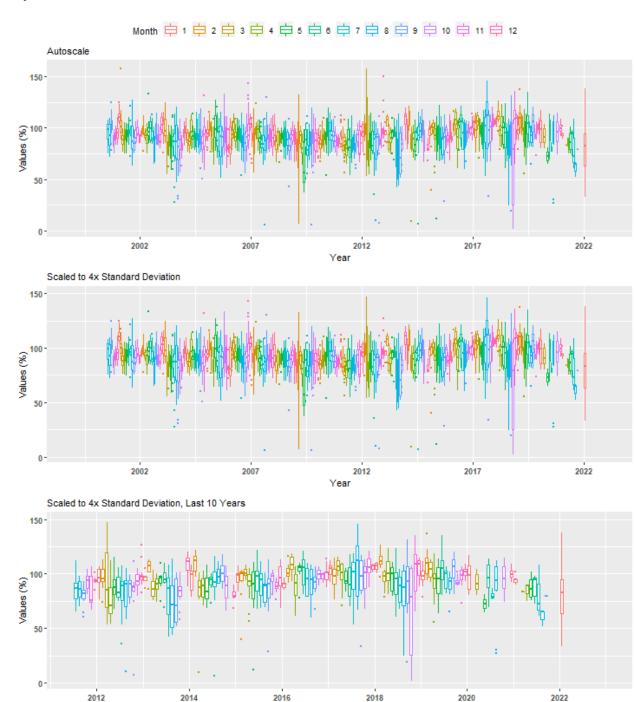
Summary Box Plots for Apalachicola National Estuarine Research Reserve $\ensuremath{\mathsf{By\,Year}}$



Year

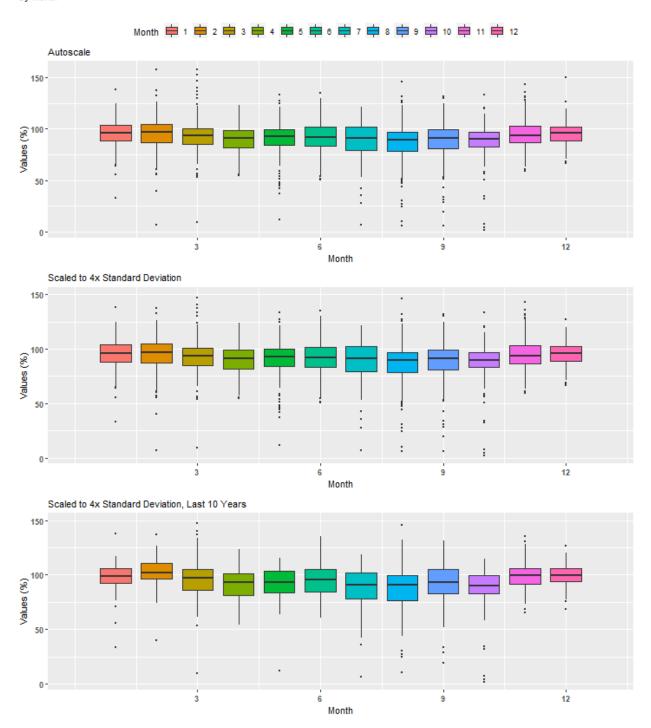
Summary Box Plots for Apalachicola National Estuarine Research Reserve

By Year & Month



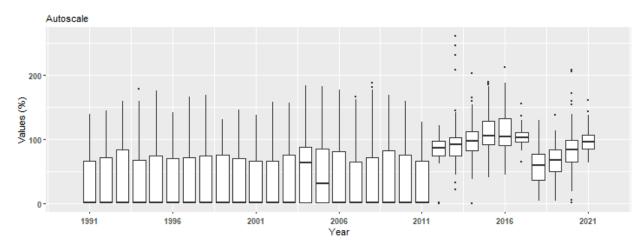
Year

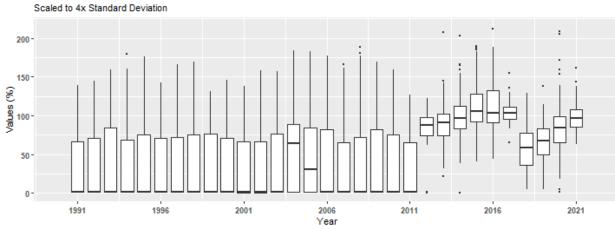
Summary Box Plots for Apalachicola National Estuarine Research ReserveBy Month

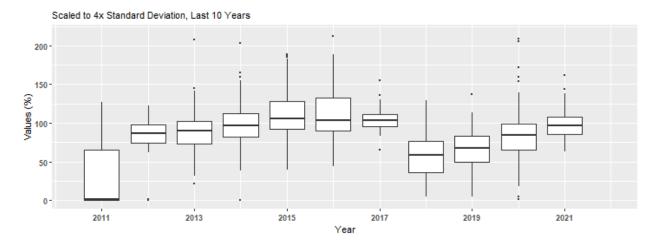


Summary Box Plots for Banana River Aquatic Preserve

By Year

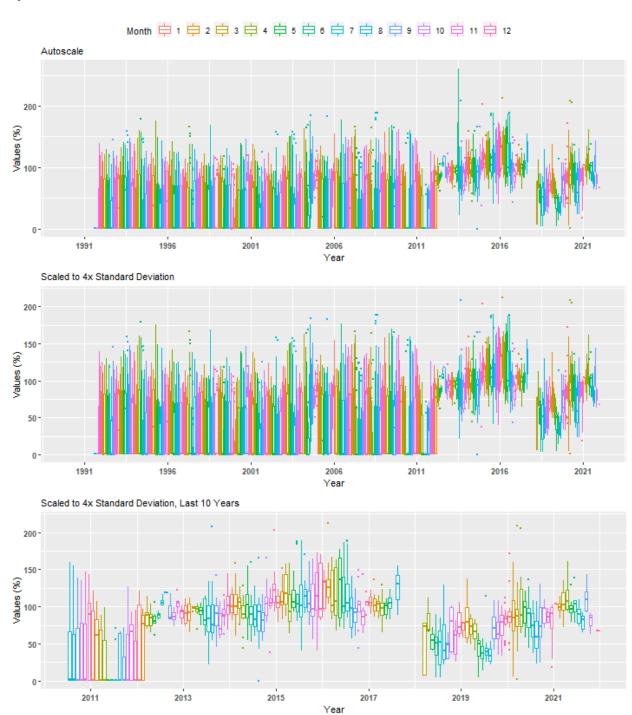






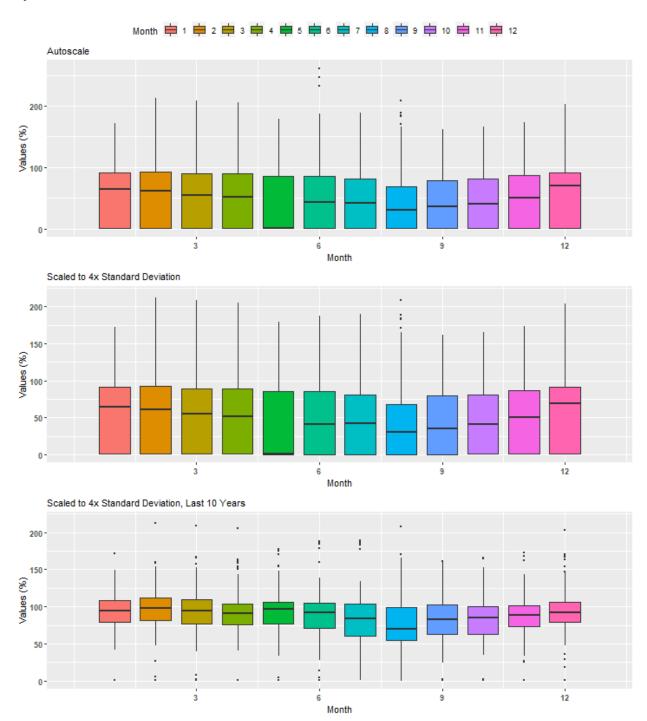
Summary Box Plots for Banana River Aquatic Preserve

By Year & Month



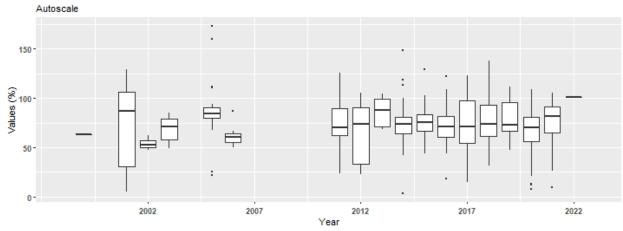
Summary Box Plots for Banana River Aquatic Preserve

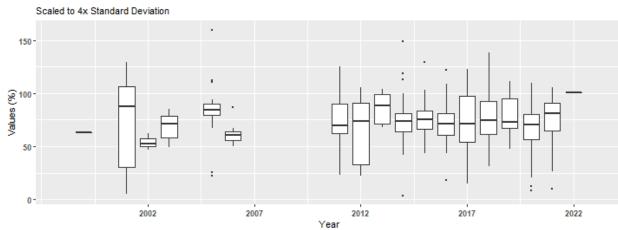
By Month

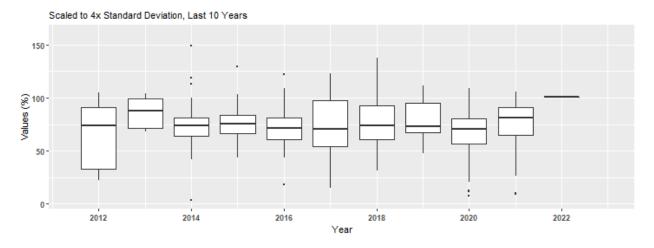


Summary Box Plots for Big Bend Seagrasses Aquatic Preserve

By Year

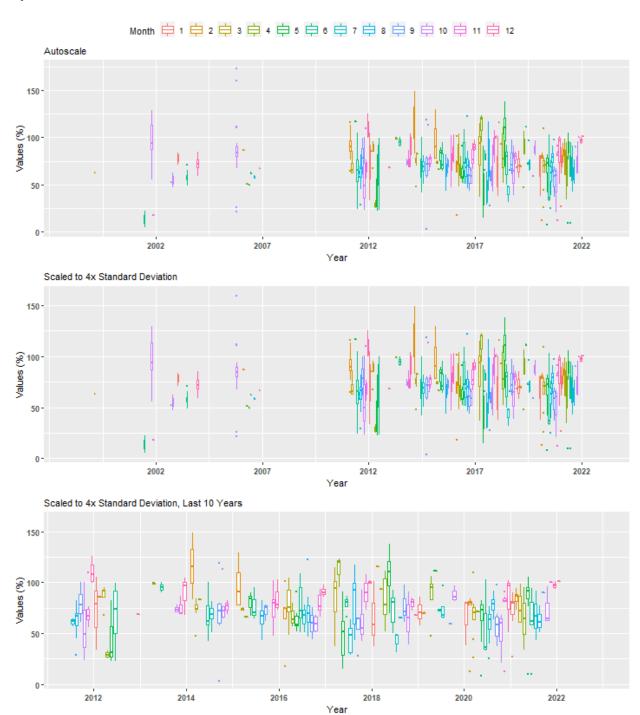






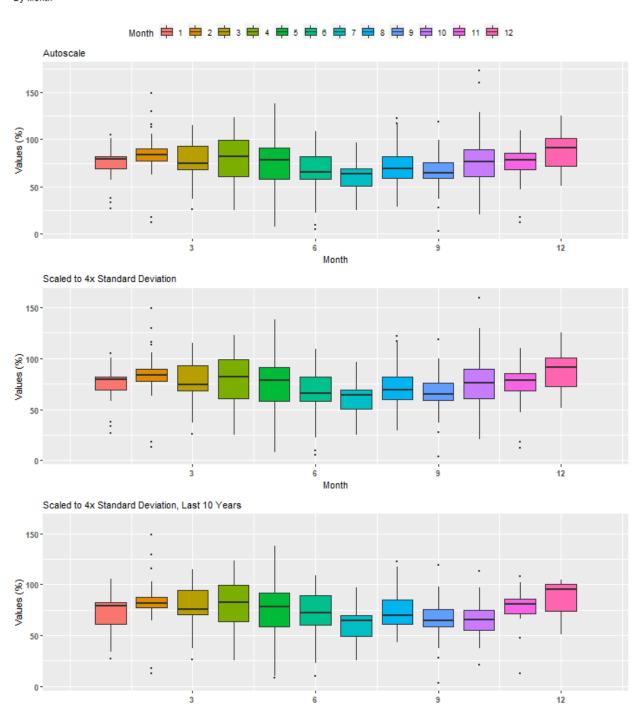
Summary Box Plots for Big Bend Seagrasses Aquatic Preserve

By Year & Month



Summary Box Plots for Big Bend Seagrasses Aquatic Preserve

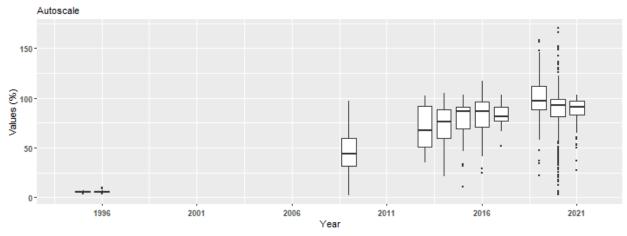
By Month

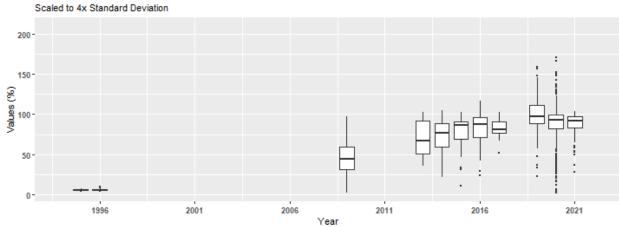


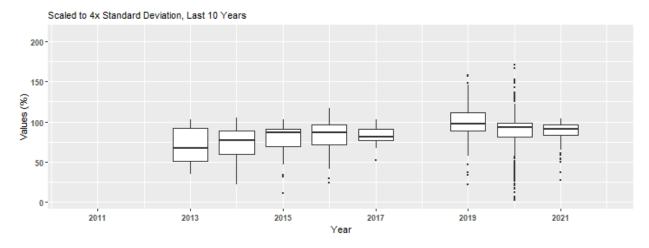
Month

Summary Box Plots for Biscayne Bay Aquatic Preserve

By Year







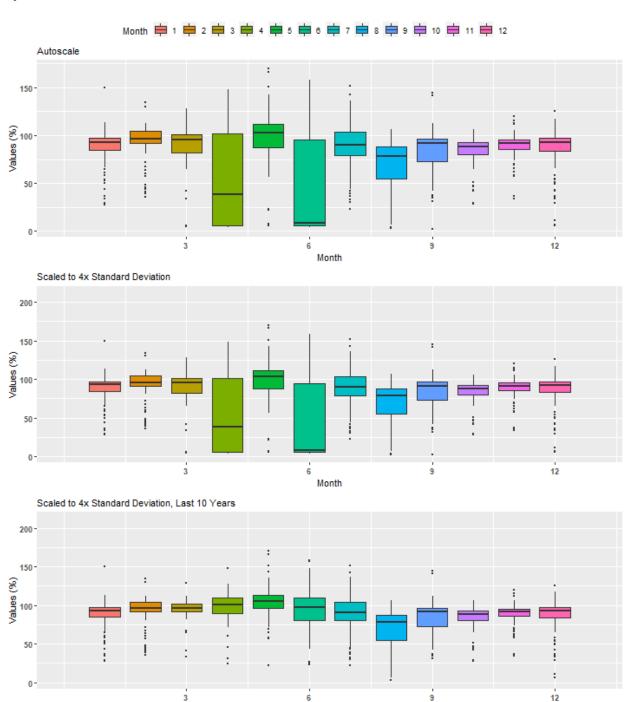
Summary Box Plots for Biscayne Bay Aquatic Preserve

By Year & Month



Summary Box Plots for Biscayne Bay Aquatic Preserve

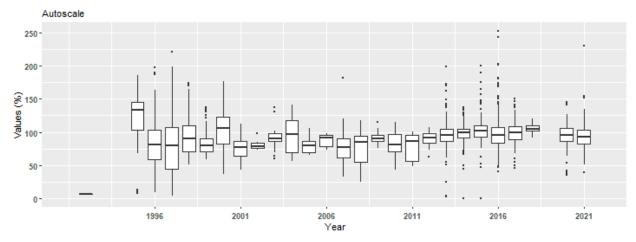
By Month

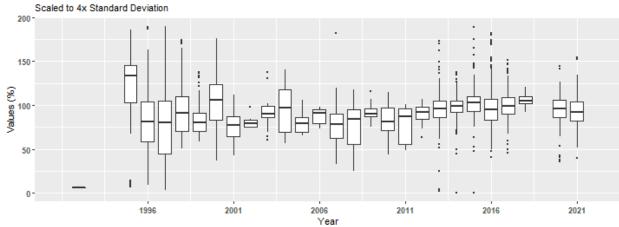


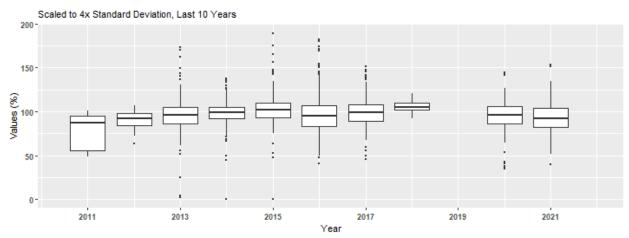
Month

Summary Box Plots for Boca Ciega Bay Aquatic Preserve

By Year

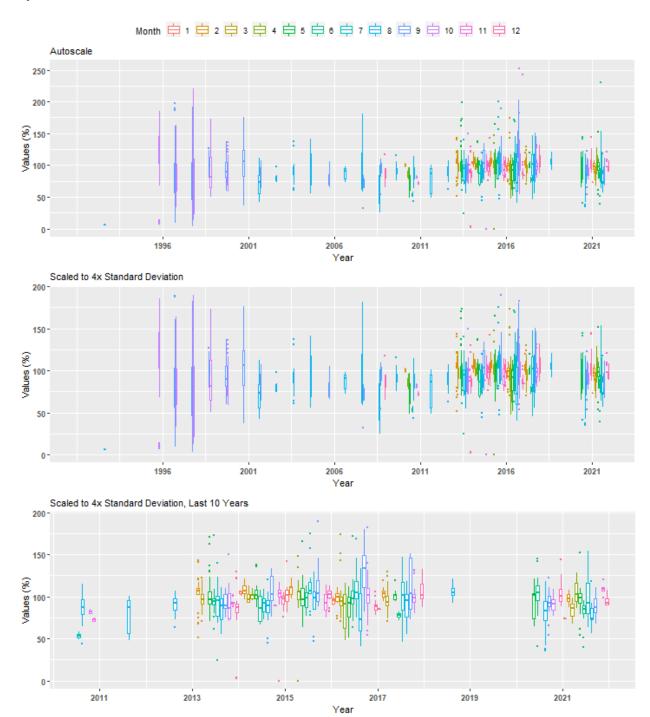






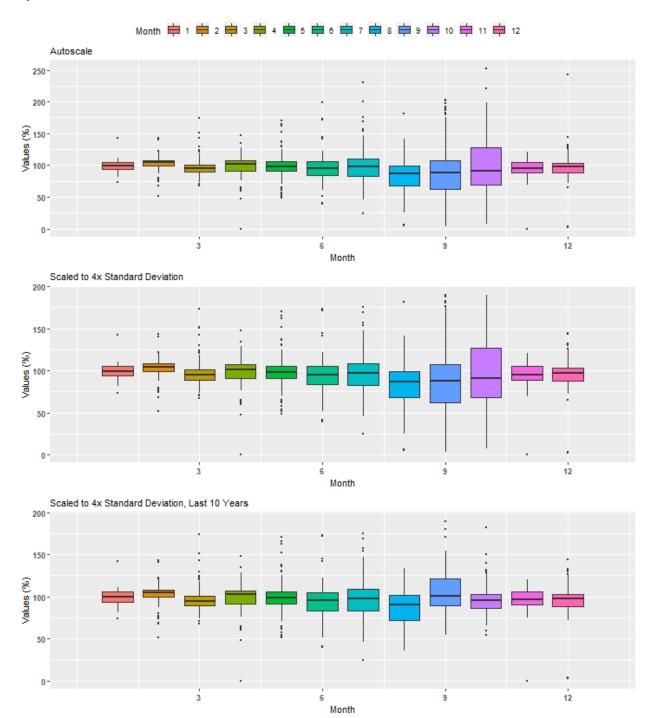
Summary Box Plots for Boca Ciega Bay Aquatic Preserve

By Year & Month

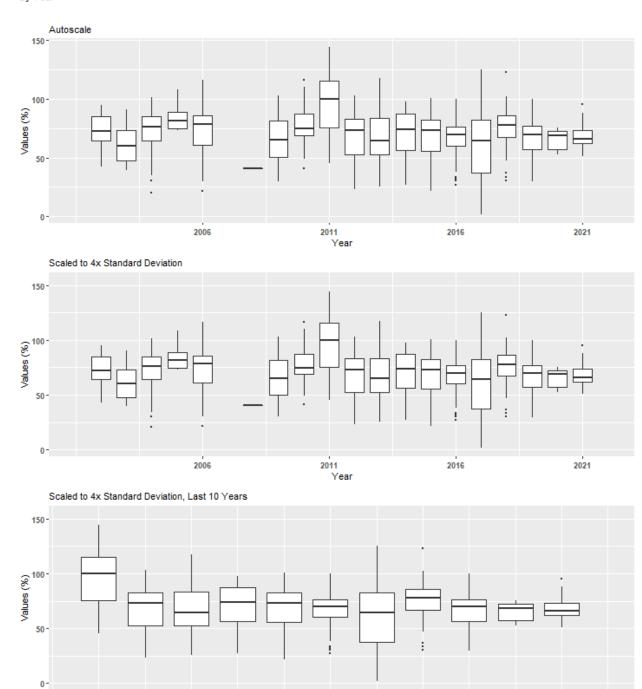


Summary Box Plots for Boca Ciega Bay Aquatic Preserve

By Month

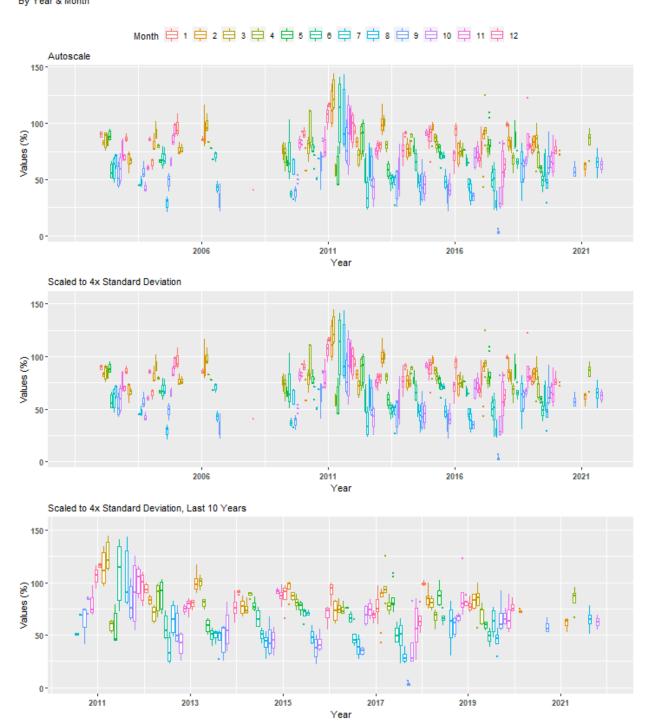


Summary Box Plots for Cape Romano-Ten Thousand Islands Aquatic Preserve $\ensuremath{\mathsf{By\,Year}}$

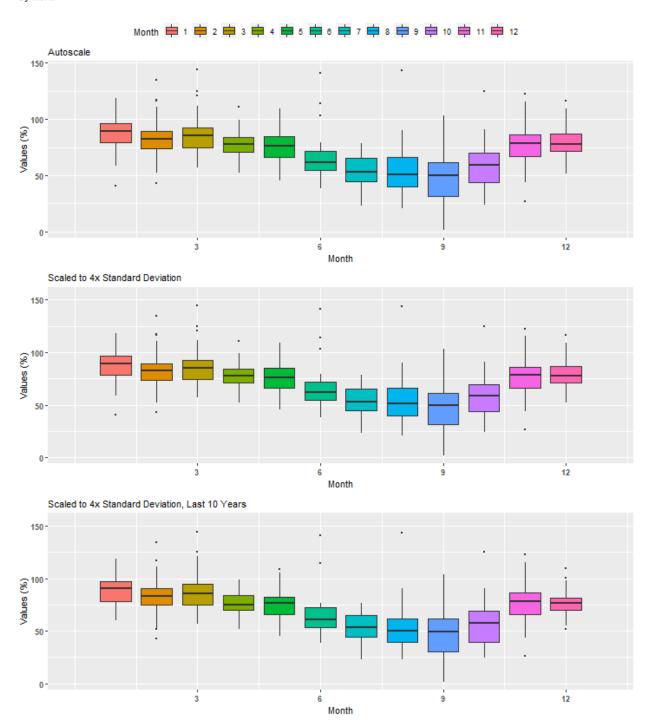


Year

Summary Box Plots for Cape Romano-Ten Thousand Islands Aquatic Preserve By Year & Month

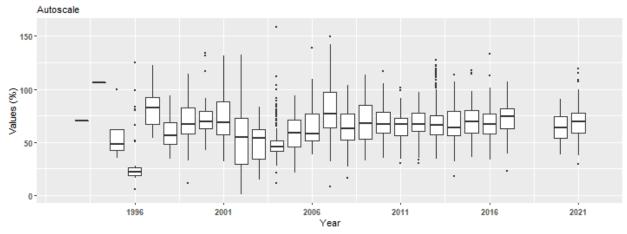


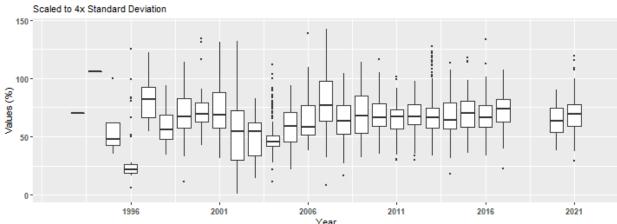
Summary Box Plots for Cape Romano-Ten Thousand Islands Aquatic Preserve By Month

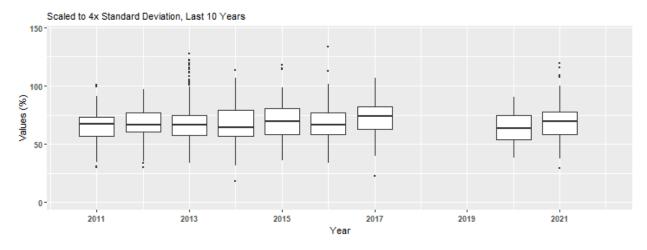


Summary Box Plots for Cockroach Bay Aquatic Preserve

By Year

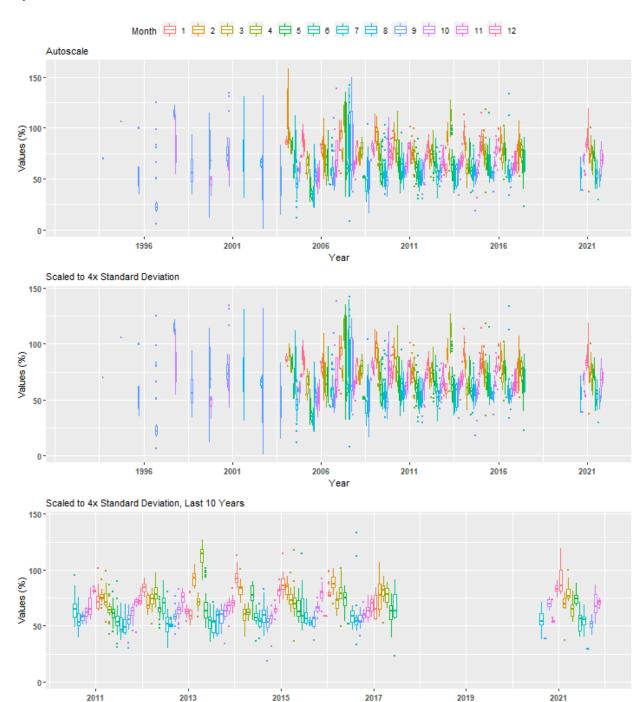






Summary Box Plots for Cockroach Bay Aquatic Preserve

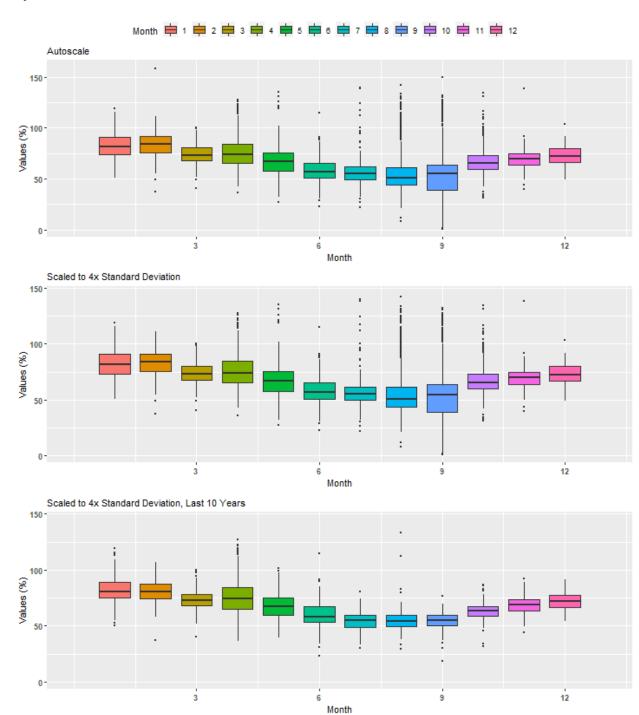
By Year & Month



Year

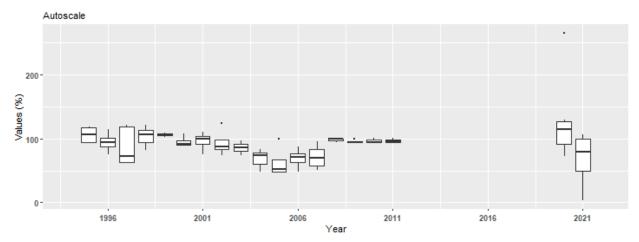
Summary Box Plots for Cockroach Bay Aquatic Preserve

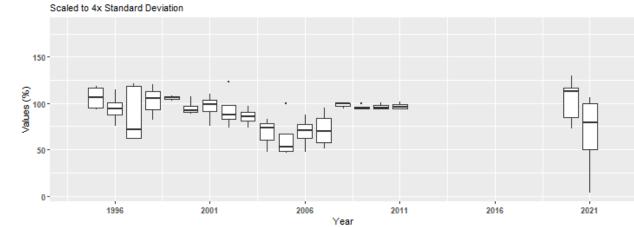
By Month

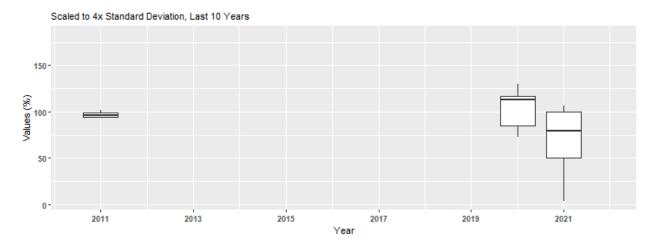


Summary Box Plots for Coupon Bight Aquatic Preserve

By Year

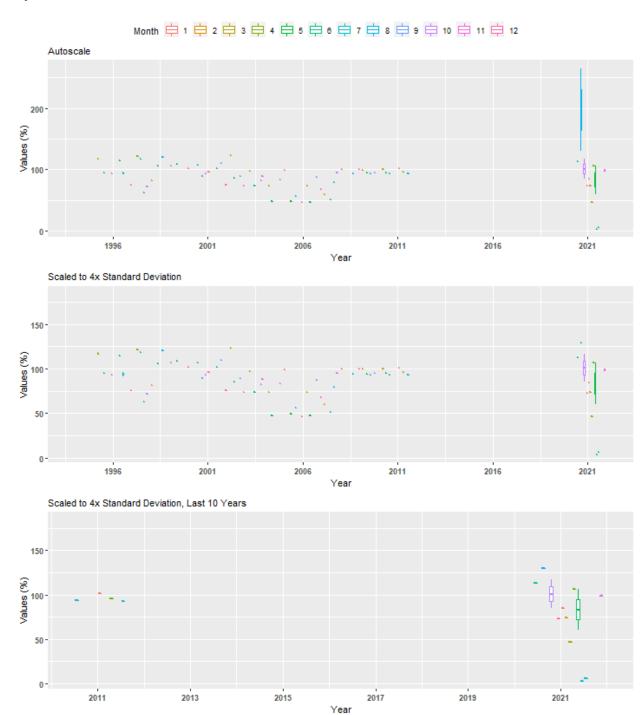






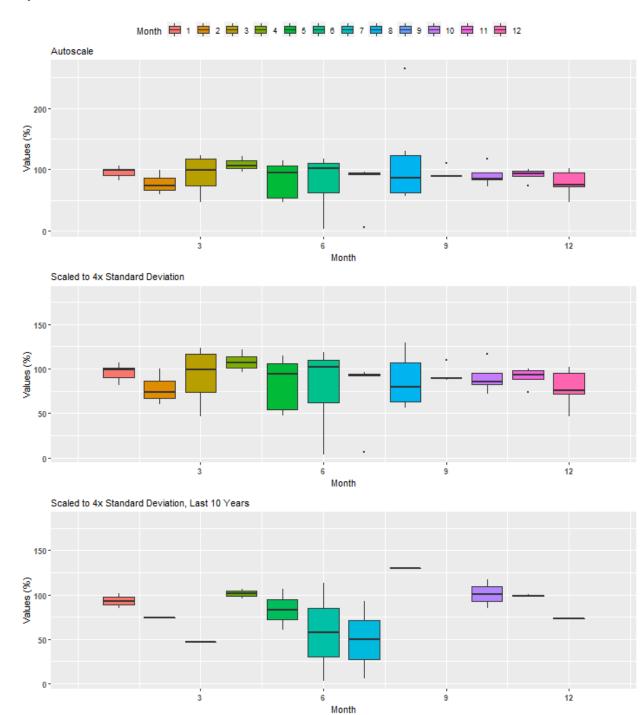
Summary Box Plots for Coupon Bight Aquatic Preserve

By Year & Month



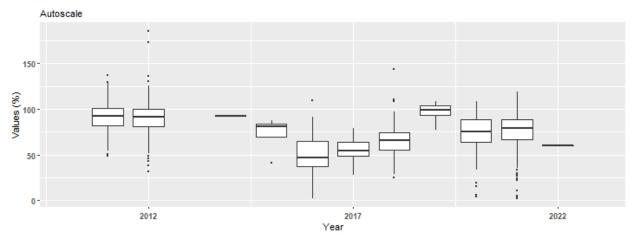
Summary Box Plots for Coupon Bight Aquatic Preserve

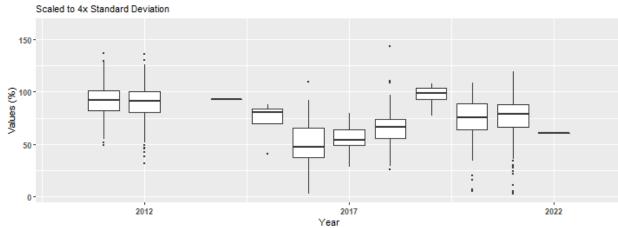
By Month

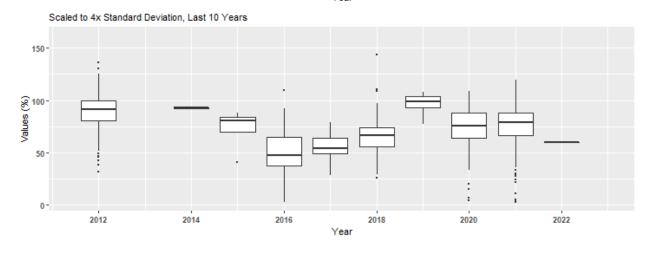


Summary Box Plots for Estero Bay Aquatic Preserve

By Year

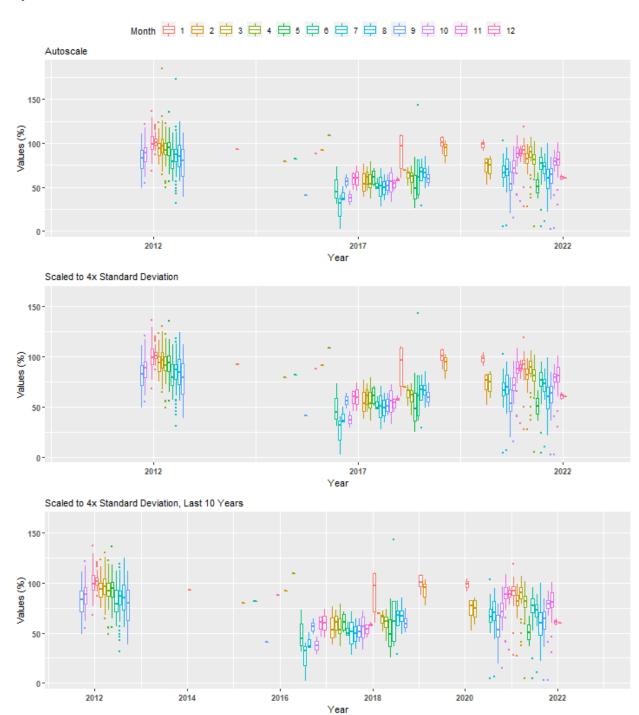






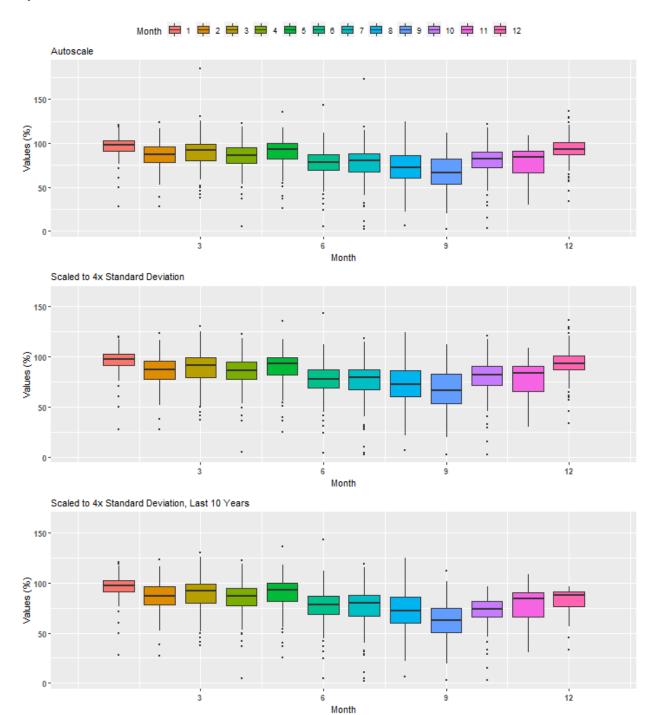
Summary Box Plots for Estero Bay Aquatic Preserve

By Year & Month



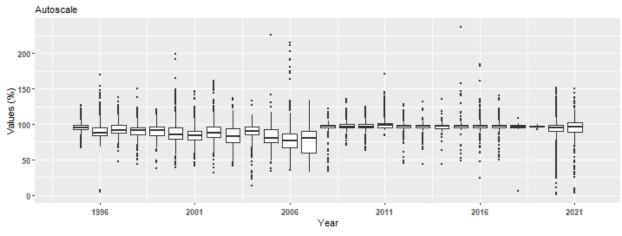
Summary Box Plots for Estero Bay Aquatic Preserve

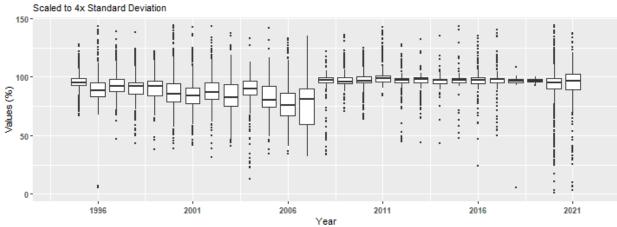
By Month

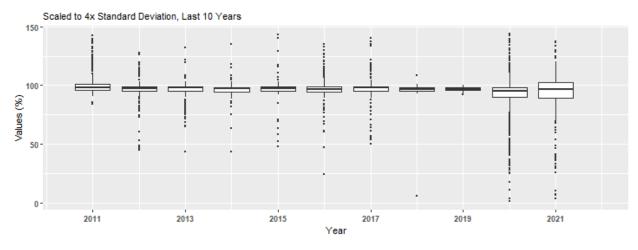


Summary Box Plots for Florida Keys National Marine Sanctuary

By Year

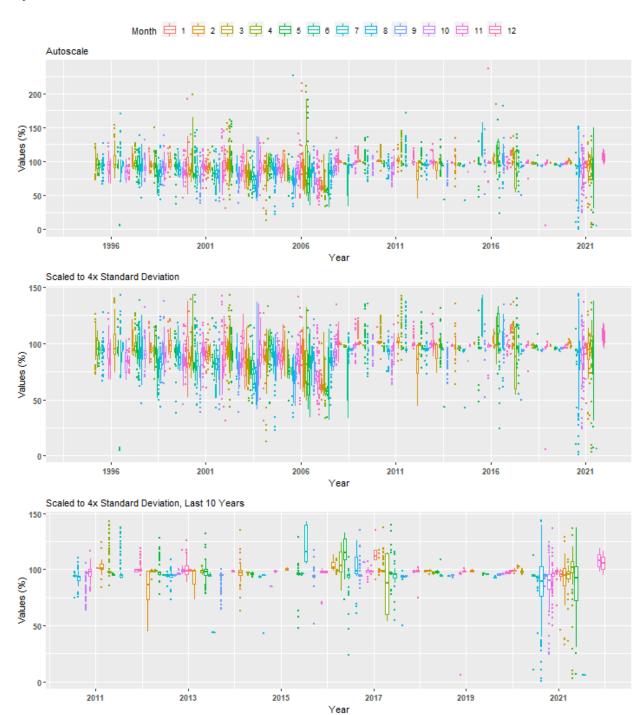






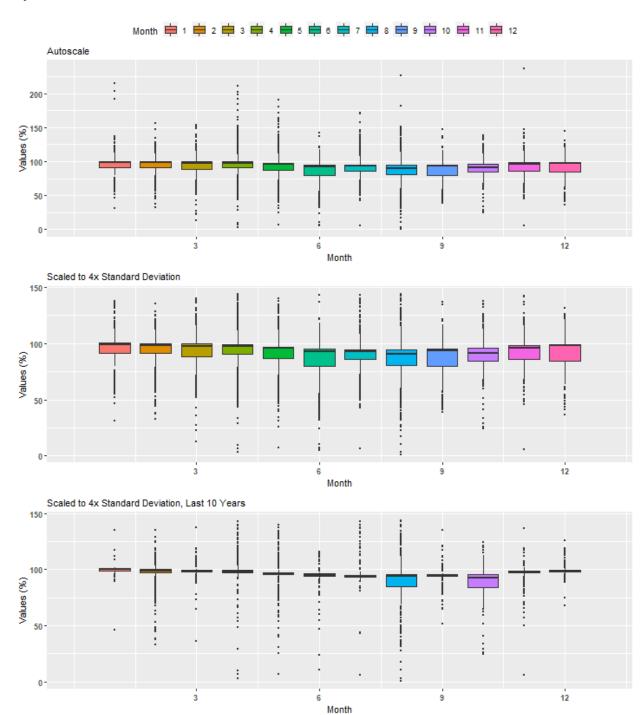
Summary Box Plots for Florida Keys National Marine Sanctuary

By Year & Month



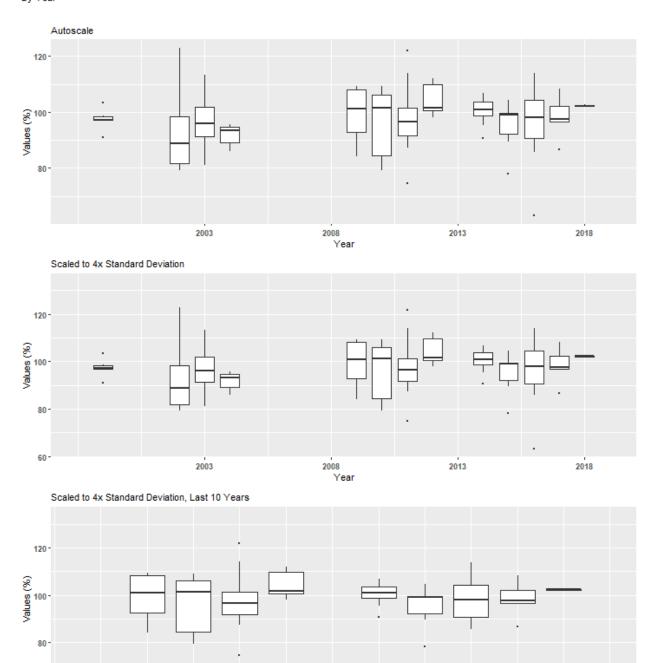
Summary Box Plots for Florida Keys National Marine Sanctuary

By Month



Summary Box Plots for Fort Pickens State Park Aquatic Preserve $\ensuremath{\mathsf{By}}\ \ensuremath{\mathsf{Year}}$

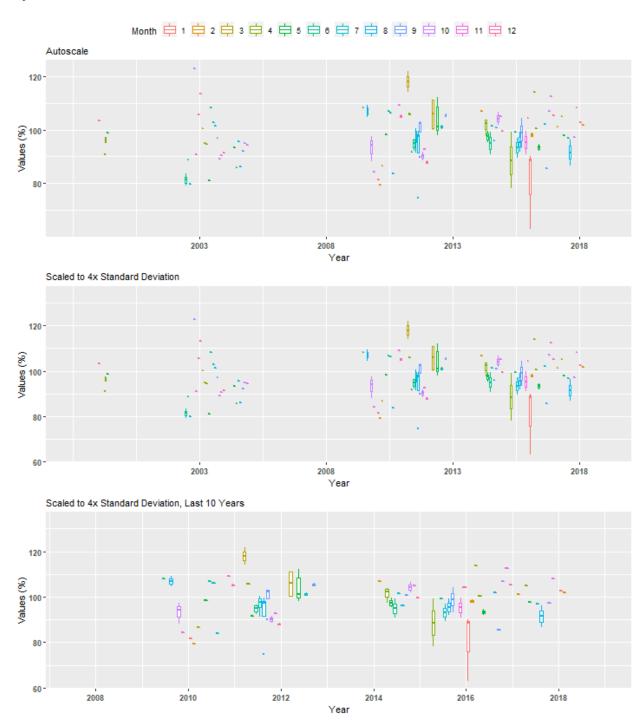
60 -



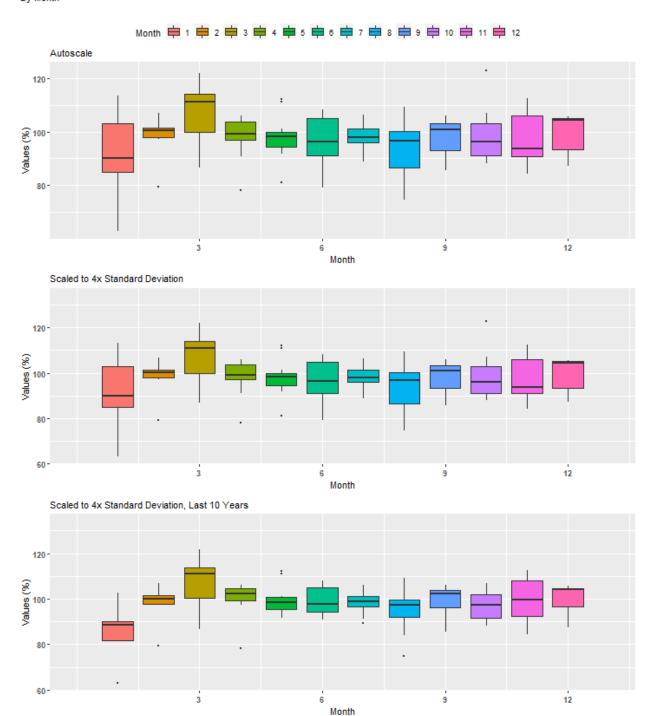
Year

Summary Box Plots for Fort Pickens State Park Aquatic Preserve

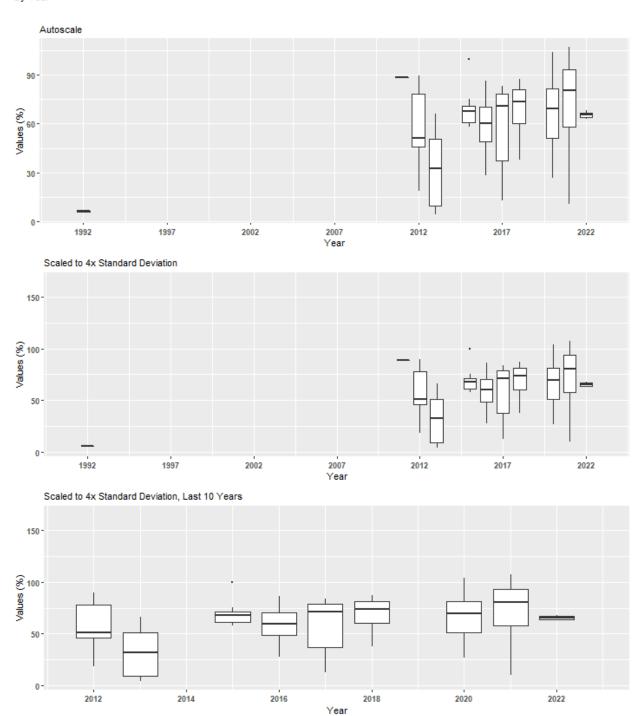
By Year & Month



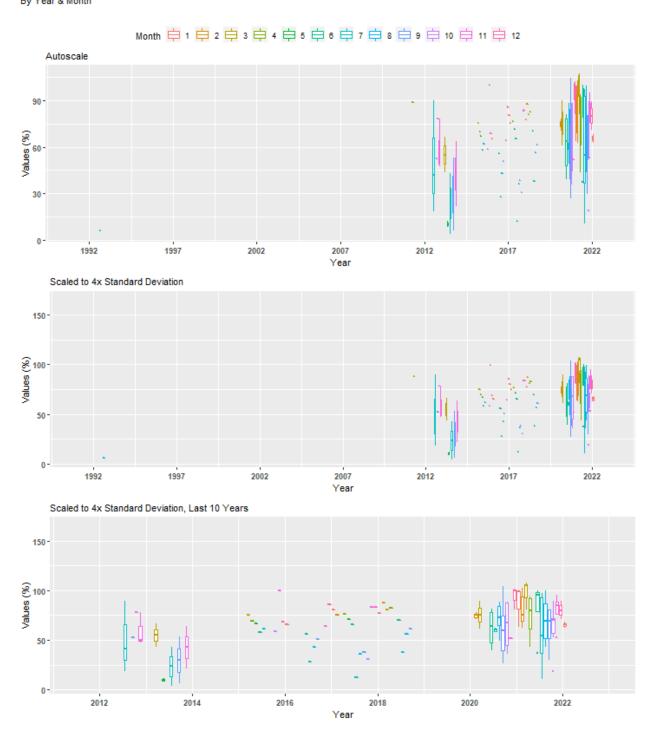
Summary Box Plots for Fort Pickens State Park Aquatic Preserve By Month



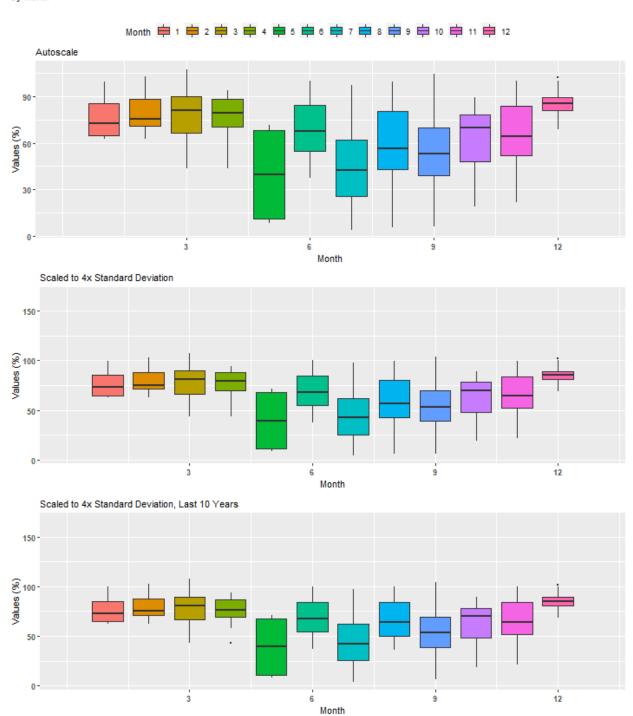
Summary Box Plots for Gasparilla Sound-Charlotte Harbor Aquatic Preserve $\ensuremath{\mathsf{By\,Year}}$



Summary Box Plots for Gasparilla Sound-Charlotte Harbor Aquatic Preserve By Year & Month

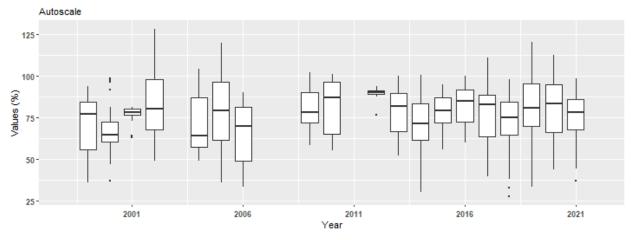


Summary Box Plots for Gasparilla Sound-Charlotte Harbor Aquatic PreserveBy Month

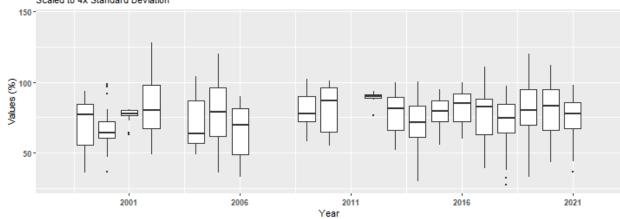


Summary Box Plots for Guana River Marsh Aquatic Preserve

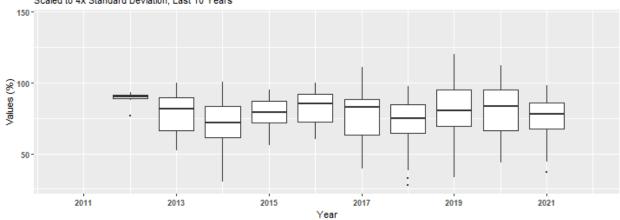
By Year





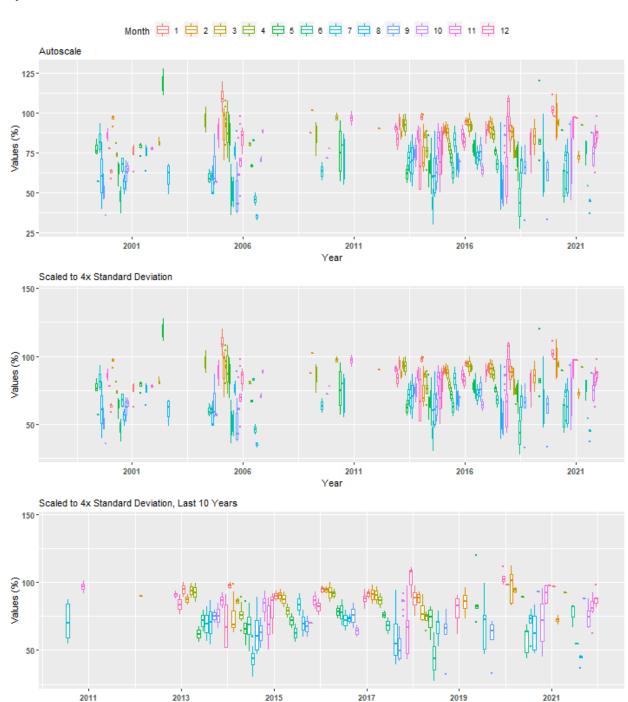






Summary Box Plots for Guana River Marsh Aquatic Preserve

By Year & Month

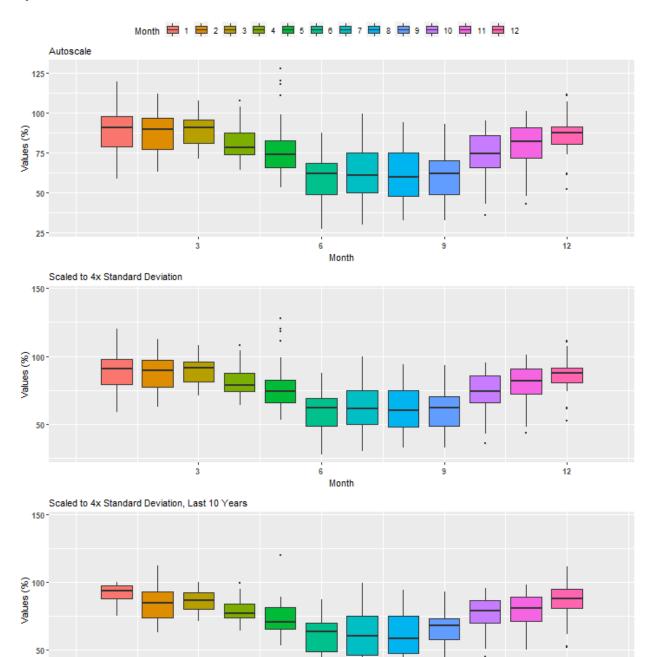


Year

Summary Box Plots for Guana River Marsh Aquatic Preserve

3

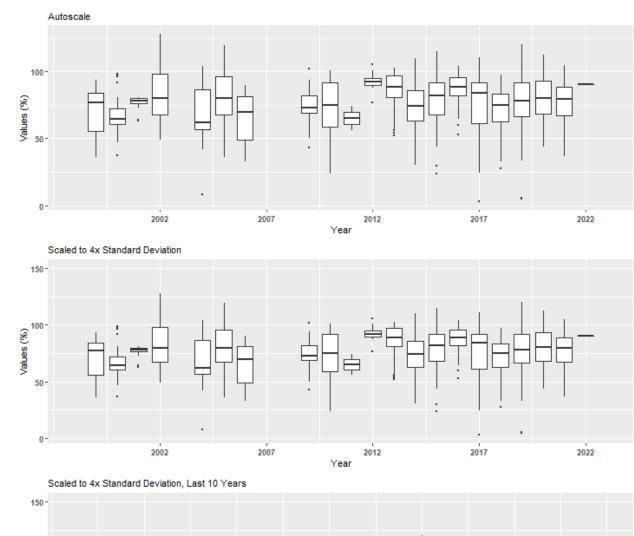
By Month

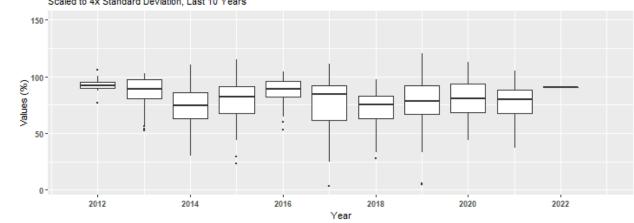


Month

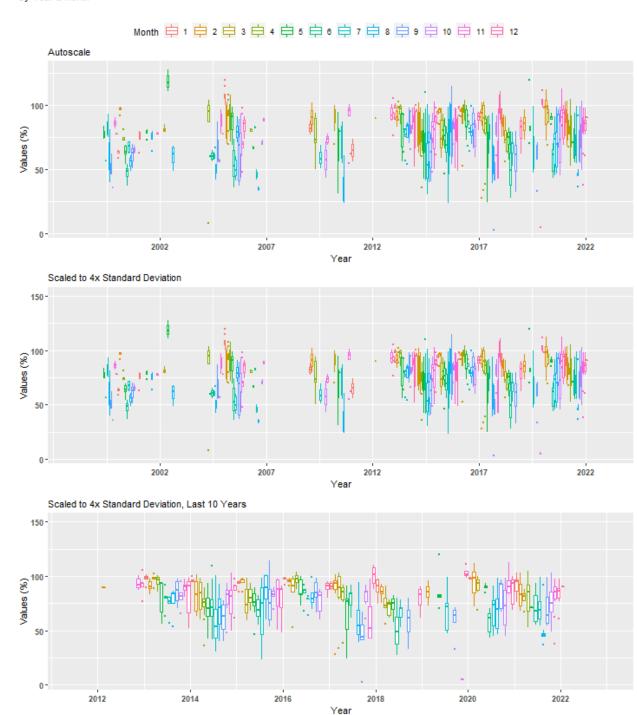
12

Summary Box Plots for Guana Tolomato Matanzas National Estuarine Research Reserve By Year

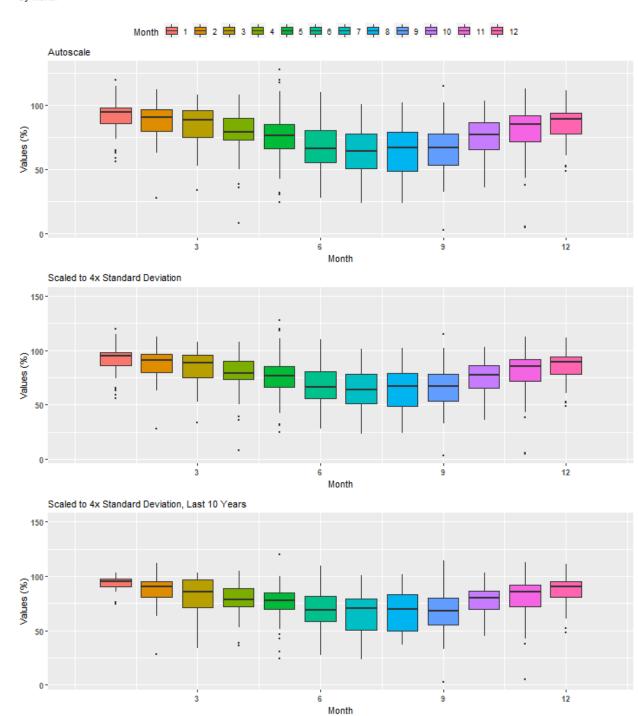




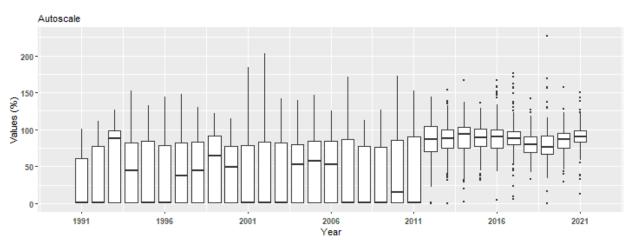
Summary Box Plots for Guana Tolomato Matanzas National Estuarine Research Reserve By Year & Month

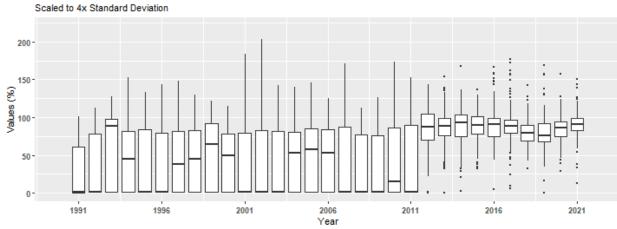


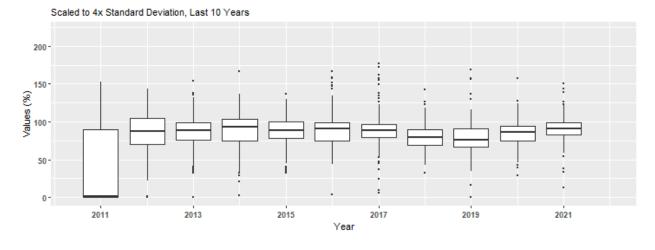
Summary Box Plots for Guana Tolomato Matanzas National Estuarine Research Reserve By Month



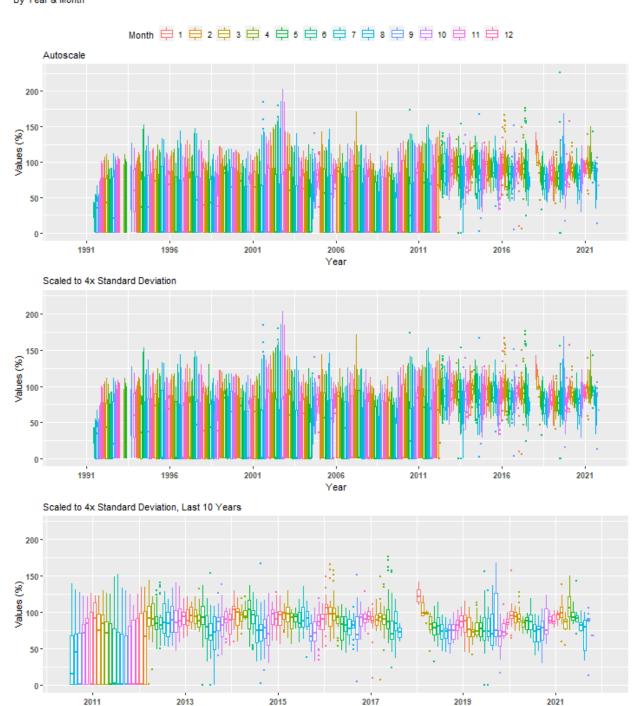
Summary Box Plots for Indian River-Malabar to Vero Beach Aquatic Preserve $\ensuremath{\mathsf{By}}\ \ensuremath{\mathsf{Year}}$





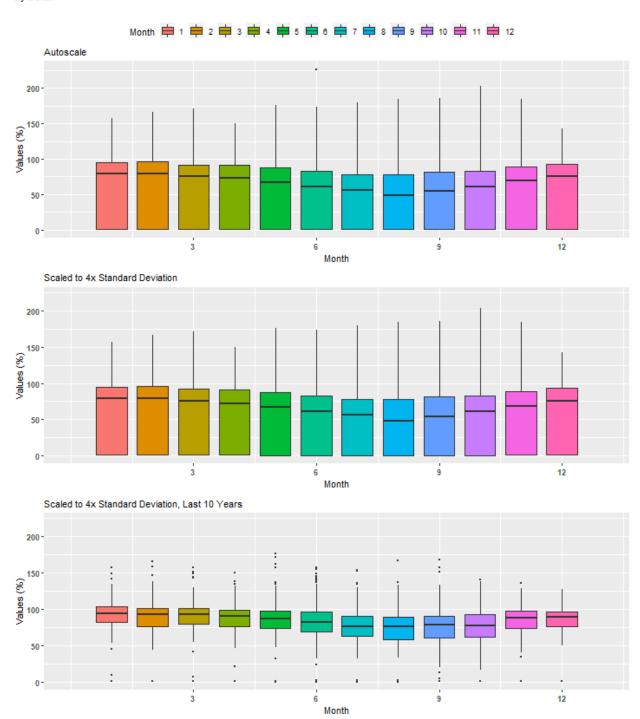


Summary Box Plots for Indian River-Malabar to Vero Beach Aquatic Preserve By Year & Month

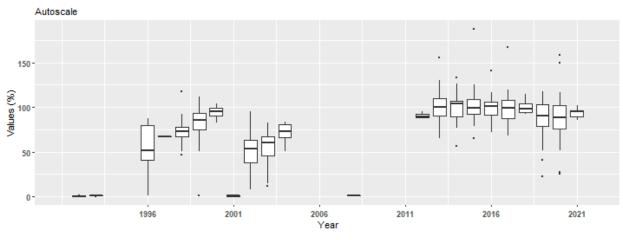


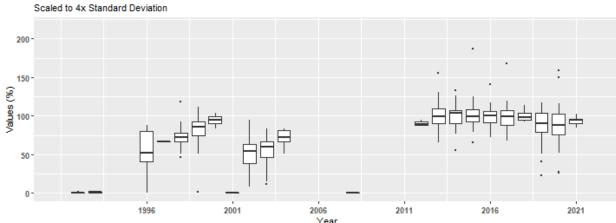
Year

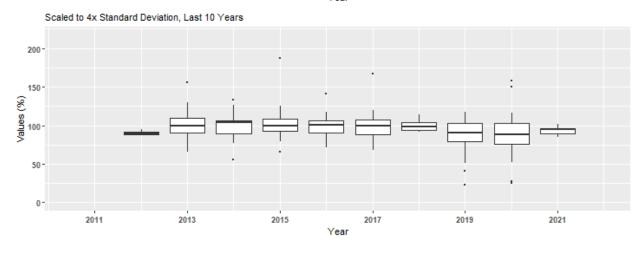
Summary Box Plots for Indian River-Malabar to Vero Beach Aquatic Preserve By Month



Summary Box Plots for Indian River-Vero Beach to Ft. Pierce Aquatic Preserve $\ensuremath{\mathsf{By\,Year}}$



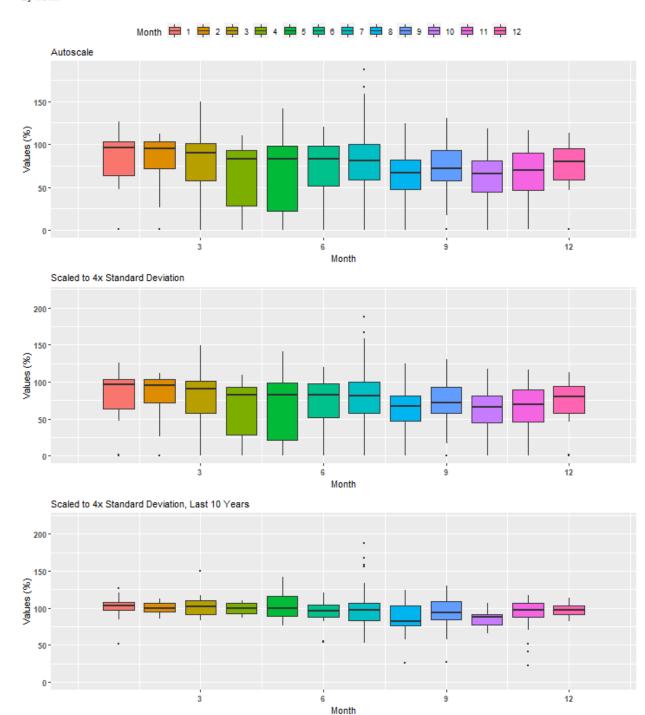




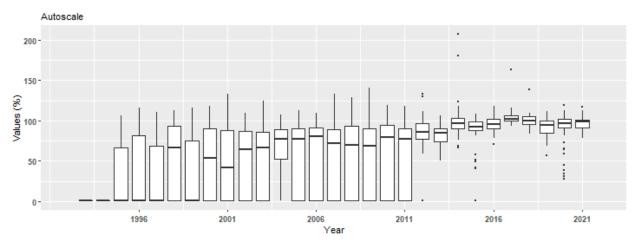
Summary Box Plots for Indian River-Vero Beach to Ft. Pierce Aquatic Preserve By Year & Month

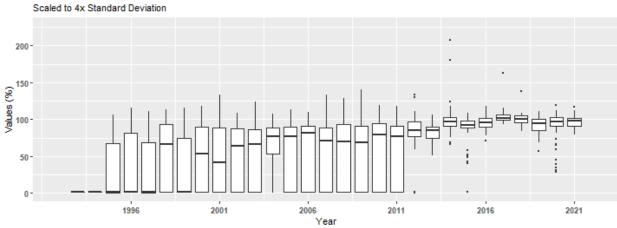


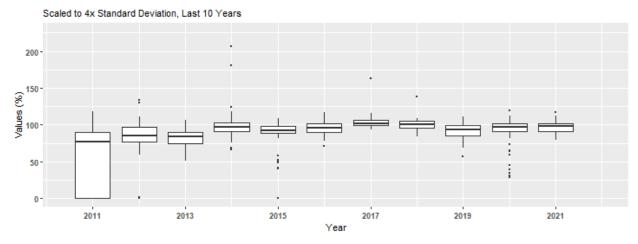
Summary Box Plots for Indian River-Vero Beach to Ft. Pierce Aquatic Preserve By Month



Summary Box Plots for Jensen Beach to Jupiter Inlet Aquatic Preserve $\ensuremath{\mathsf{By}}\, \ensuremath{\mathsf{Year}}$

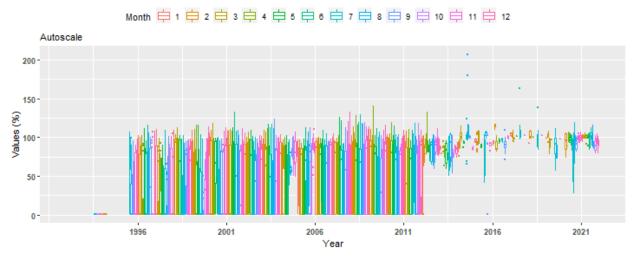


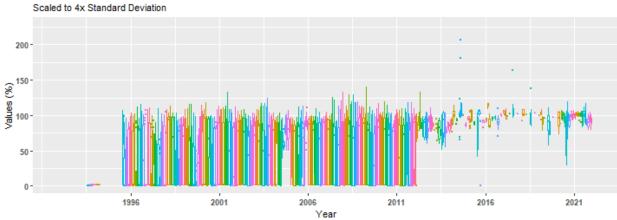


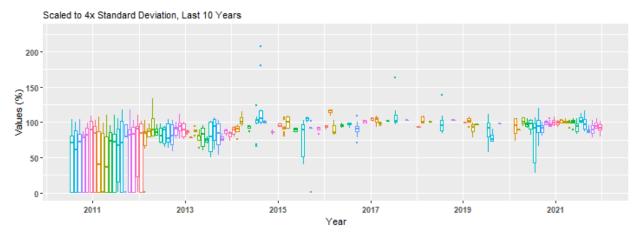


Summary Box Plots for Jensen Beach to Jupiter Inlet Aquatic Preserve

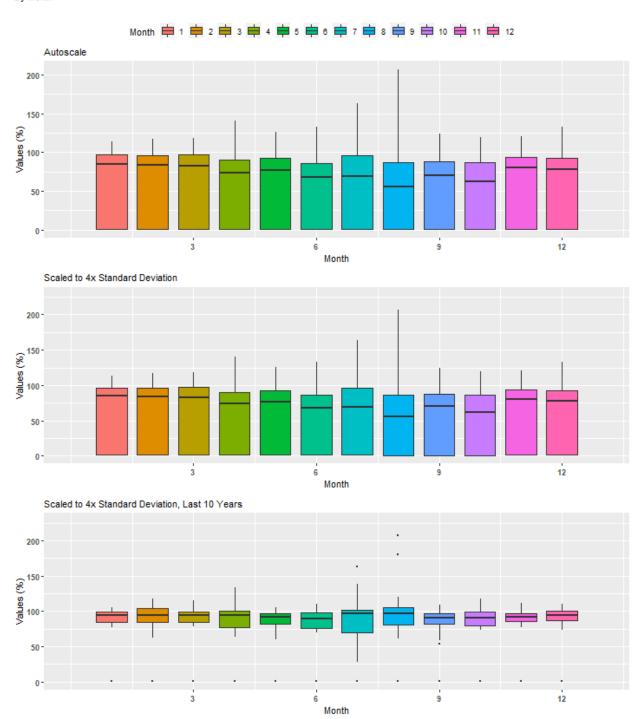
By Year & Month





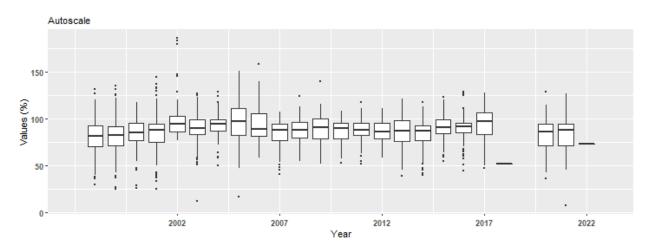


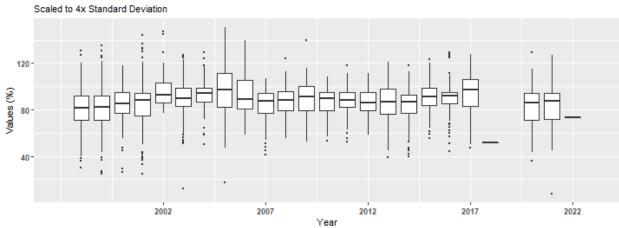
Summary Box Plots for Jensen Beach to Jupiter Inlet Aquatic Preserve By Month

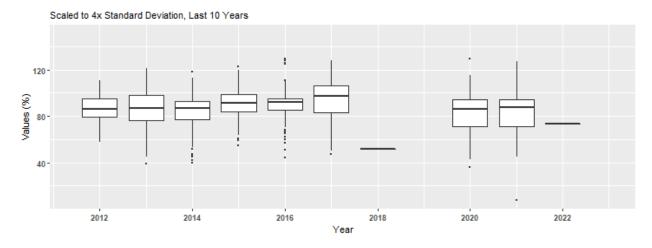


Summary Box Plots for Lemon Bay Aquatic Preserve

By Year

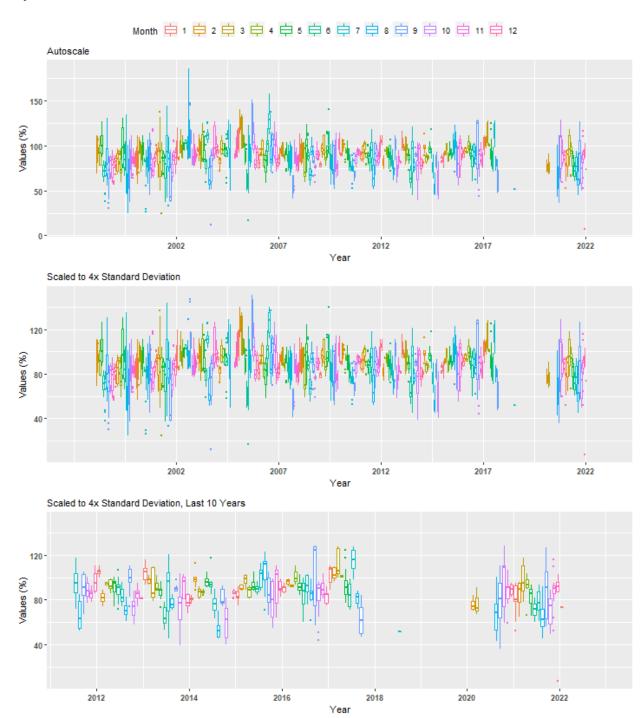






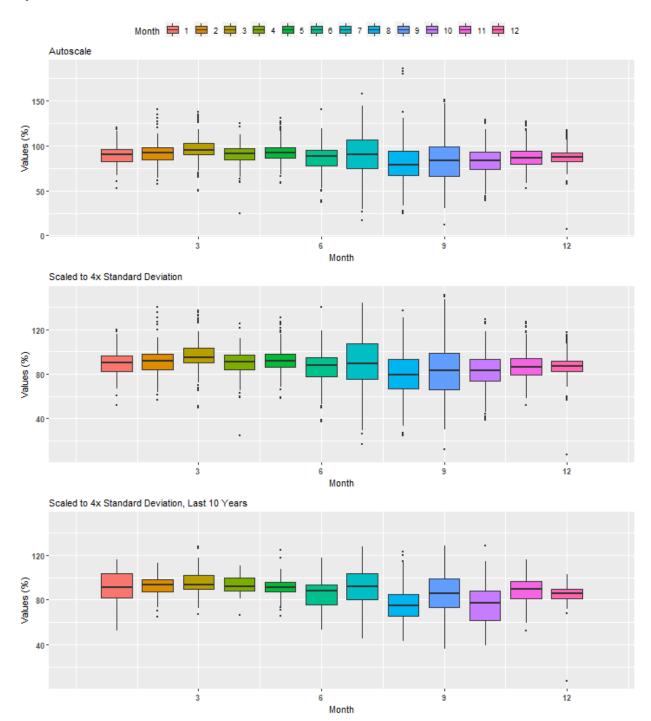
Summary Box Plots for Lemon Bay Aquatic Preserve

By Year & Month

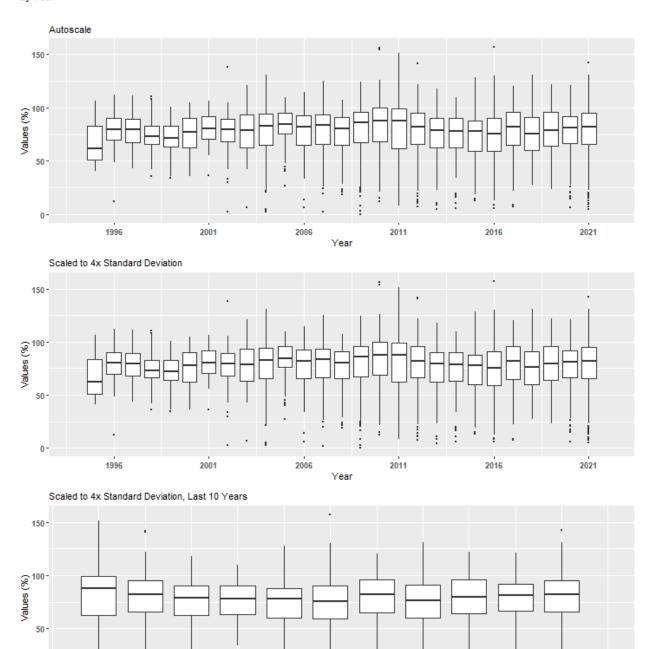


Summary Box Plots for Lemon Bay Aquatic Preserve

By Month



Summary Box Plots for Loxahatchee River-Lake Worth Creek Aquatic Preserve By Year

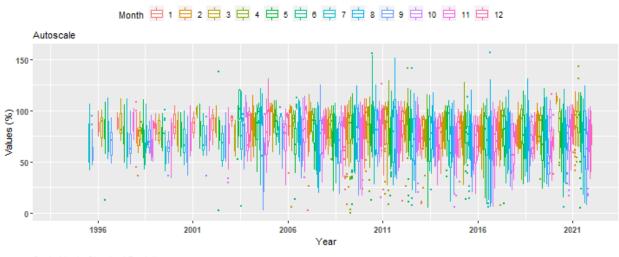


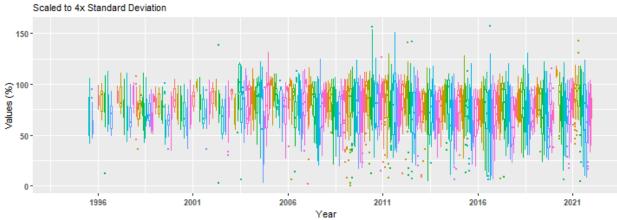
Year

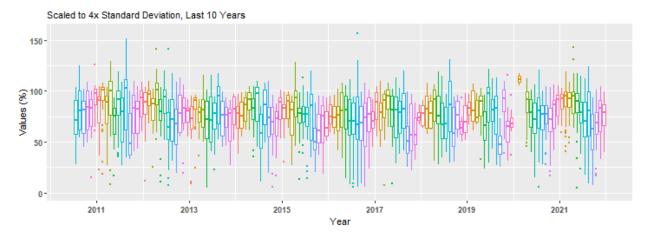
0-

Summary Box Plots for Loxahatchee River-Lake Worth Creek Aquatic Preserve

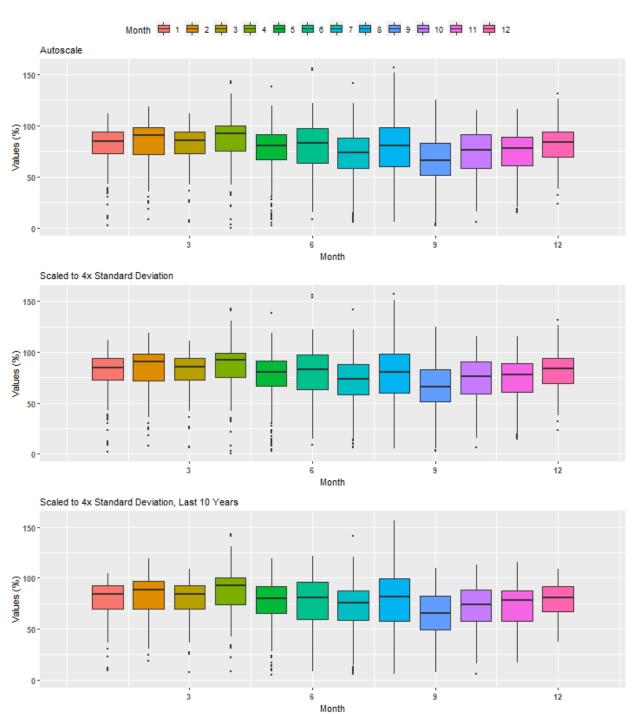
By Year & Month





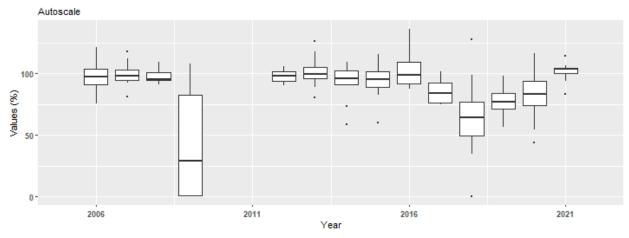


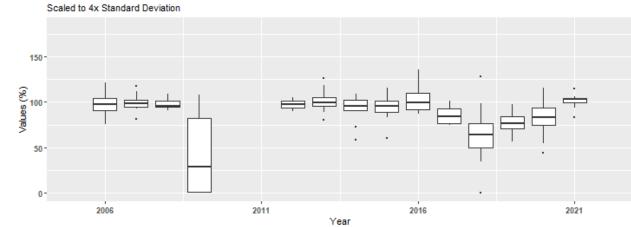
Summary Box Plots for Loxahatchee River-Lake Worth Creek Aquatic Preserve By Month

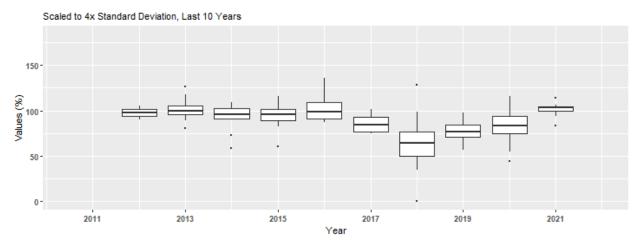


Summary Box Plots for Mosquito Lagoon Aquatic Preserve

By Year





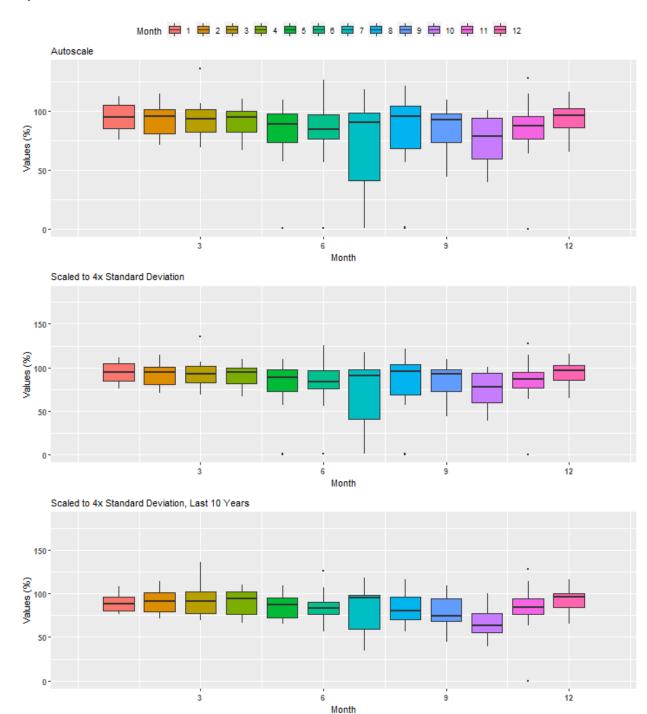


Summary Box Plots for Mosquito Lagoon Aquatic Preserve

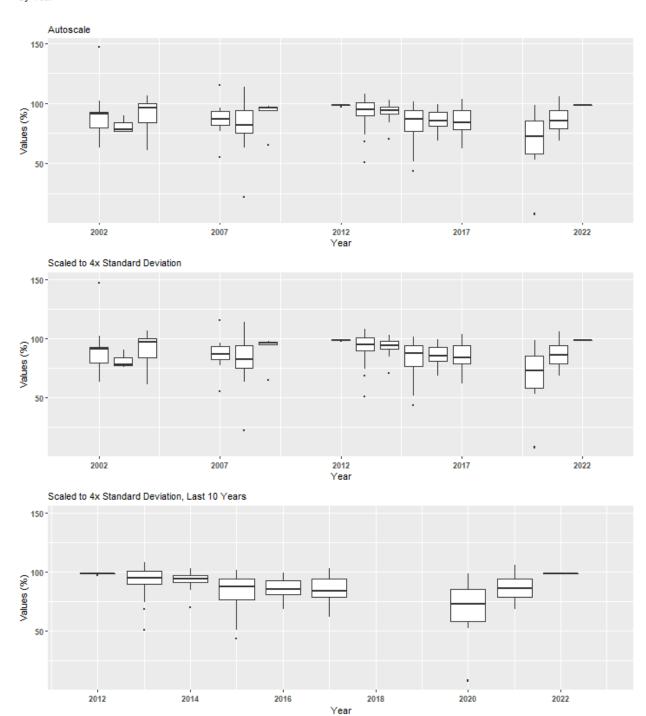


Summary Box Plots for Mosquito Lagoon Aquatic Preserve

By Month



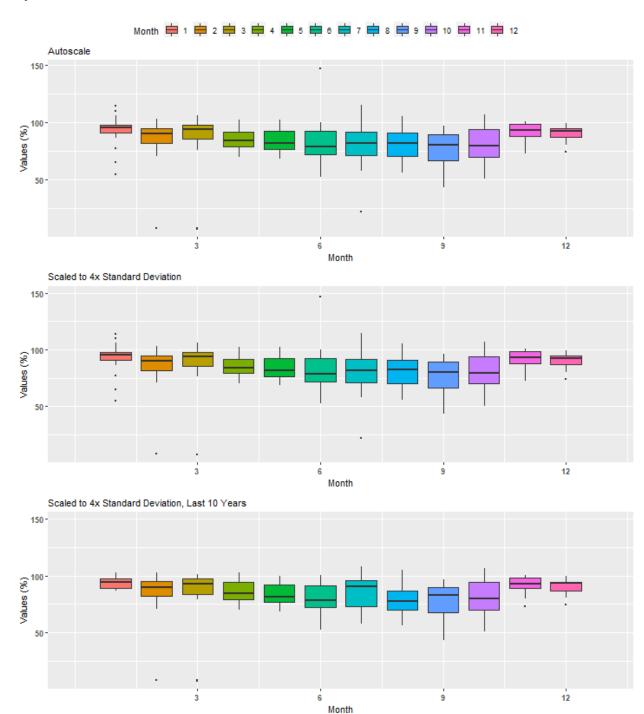
Summary Box Plots for Nassau River-St. Johns River Marshes Aquatic Preserve $\ensuremath{\mathsf{By\,Year}}$



Summary Box Plots for Nassau River-St. Johns River Marshes Aquatic Preserve By Year & Month

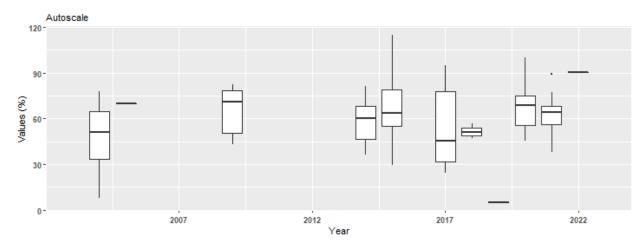


Summary Box Plots for Nassau River-St. Johns River Marshes Aquatic PreserveBy Month

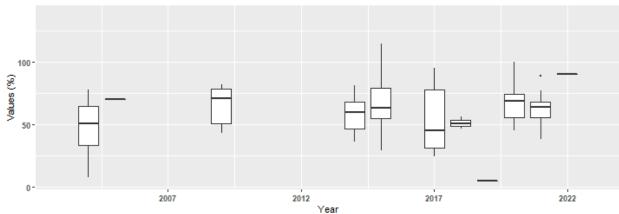


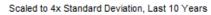
Summary Box Plots for Pellicer Creek Aquatic Preserve

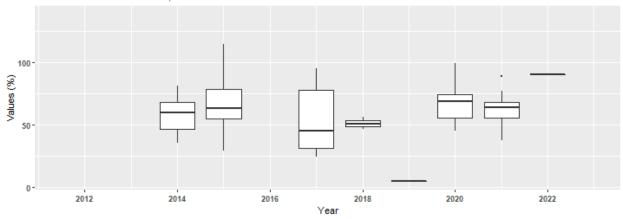
By Year



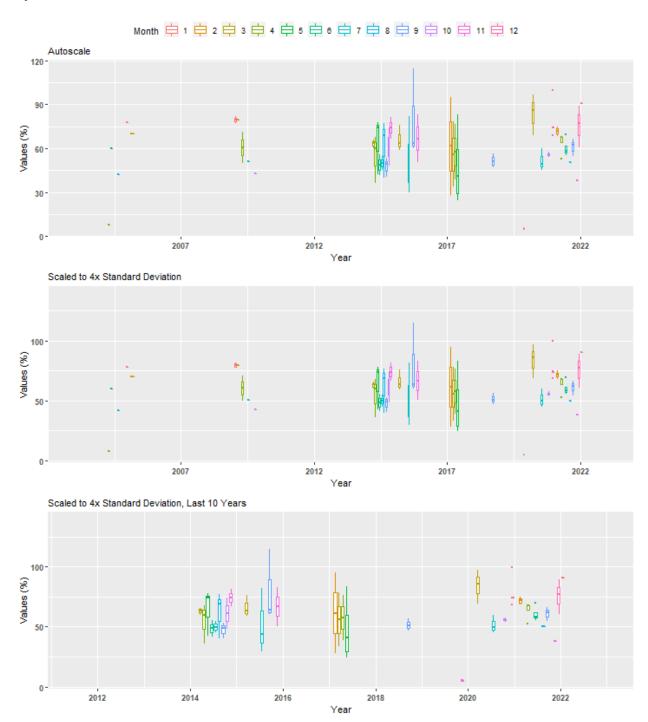






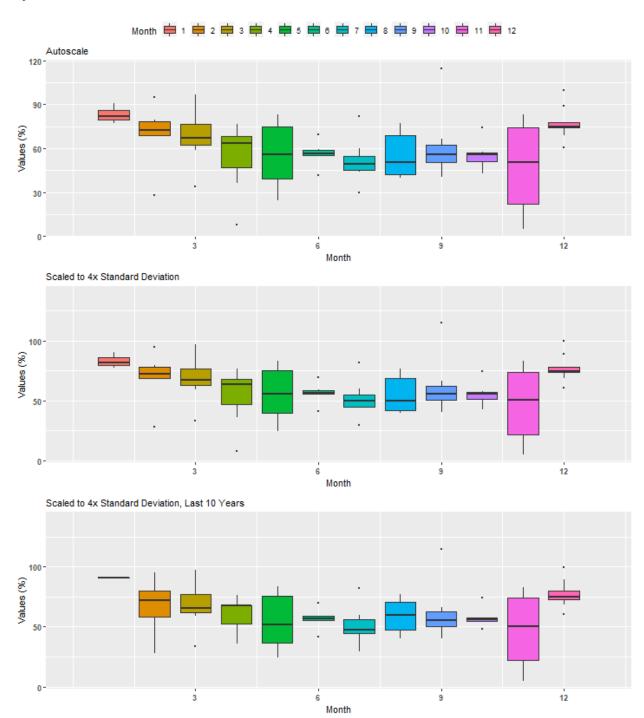


Summary Box Plots for Pellicer Creek Aquatic Preserve



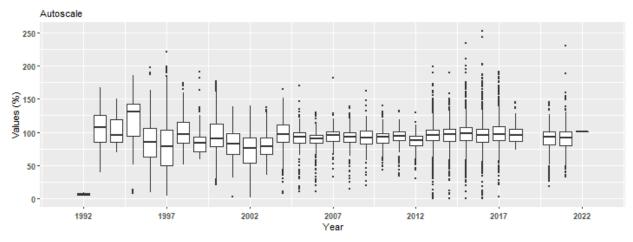
Summary Box Plots for Pellicer Creek Aquatic Preserve

By Month

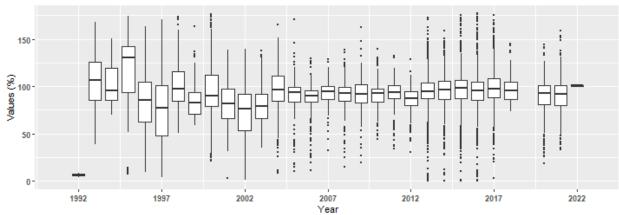


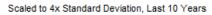
Summary Box Plots for Pinellas County Aquatic Preserve

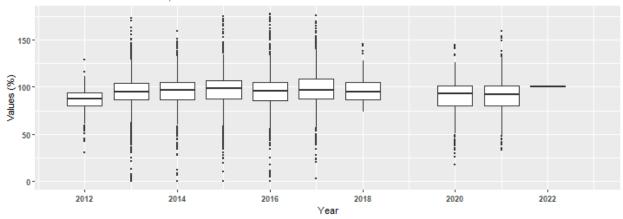
By Year



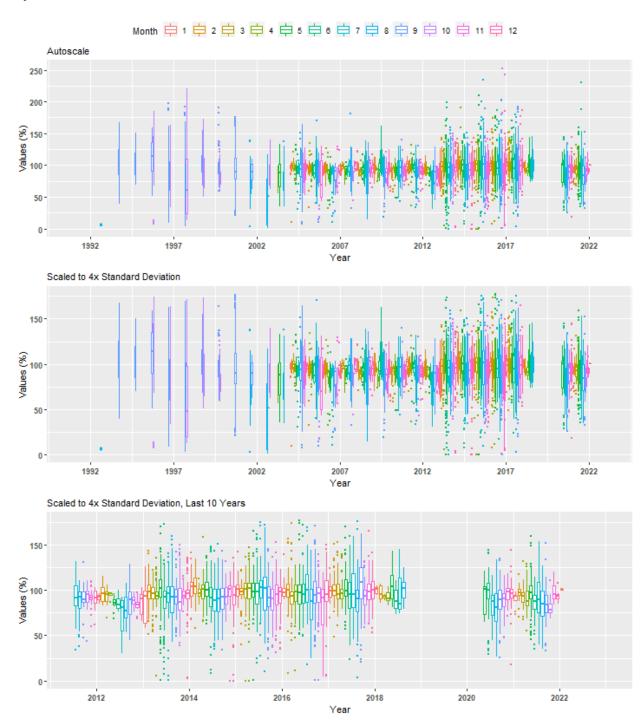






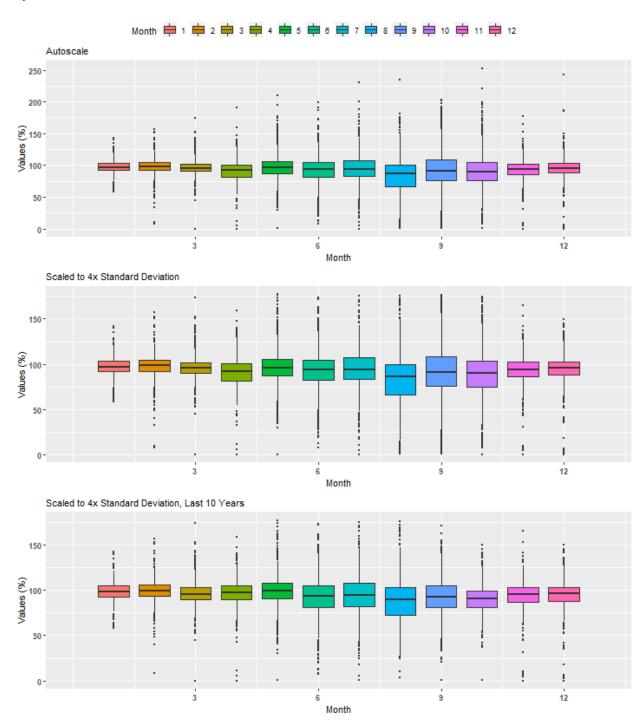


Summary Box Plots for Pinellas County Aquatic Preserve

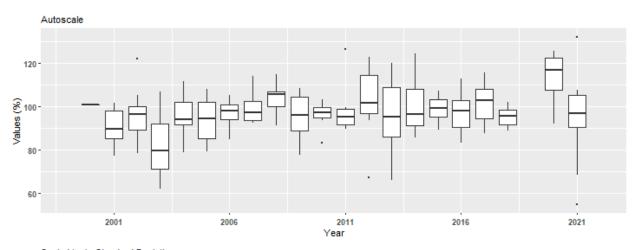


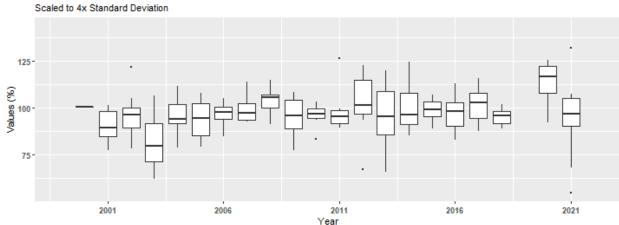
Summary Box Plots for Pinellas County Aquatic Preserve

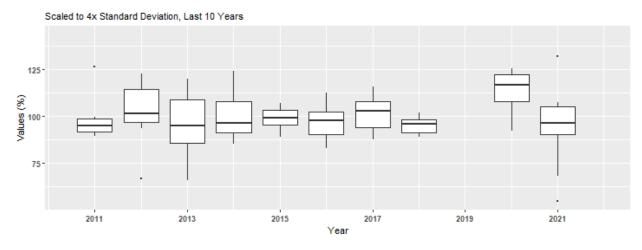
By Month



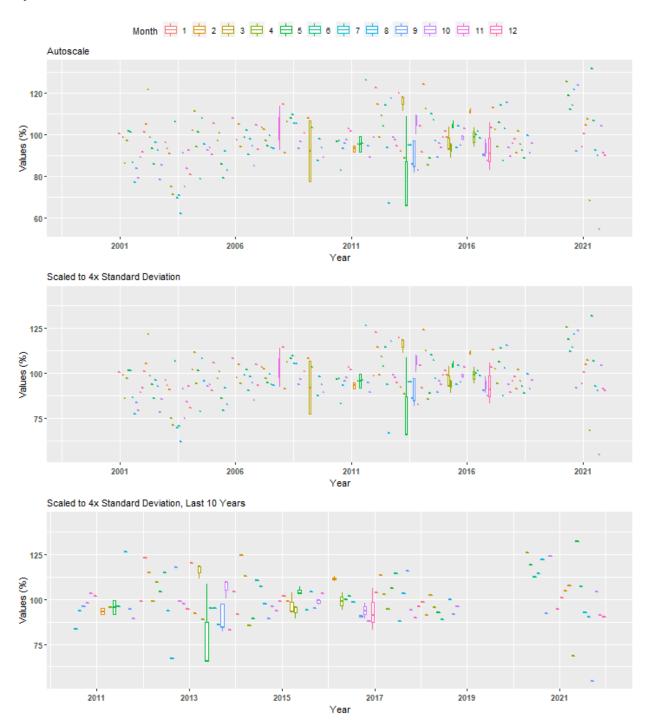
Summary Box Plots for Rocky Bayou State Park Aquatic Preserve By Year



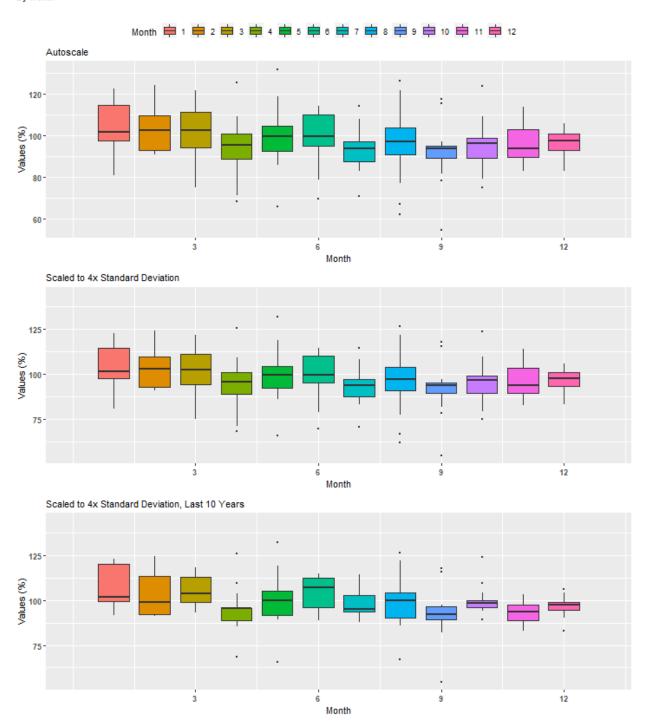




Summary Box Plots for Rocky Bayou State Park Aquatic Preserve

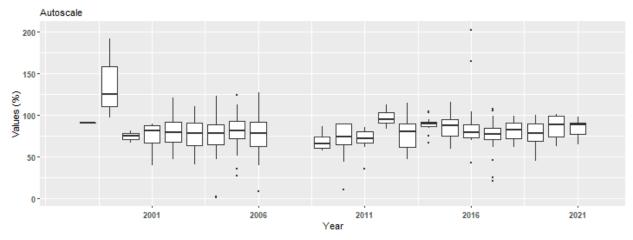


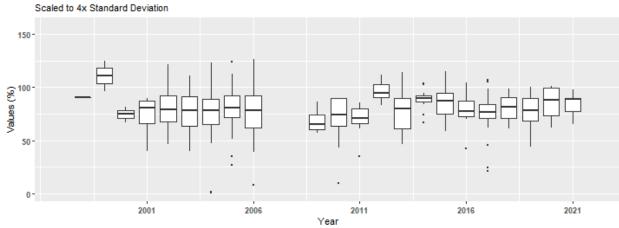
Summary Box Plots for Rocky Bayou State Park Aquatic Preserve By Month

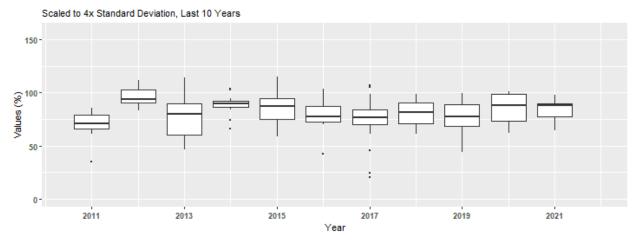


Summary Box Plots for Rookery Bay Aquatic Preserve

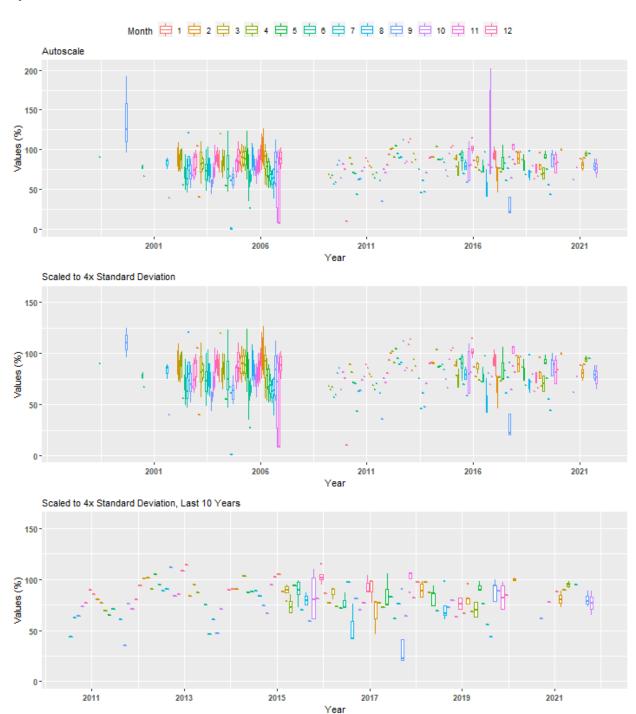
By Year





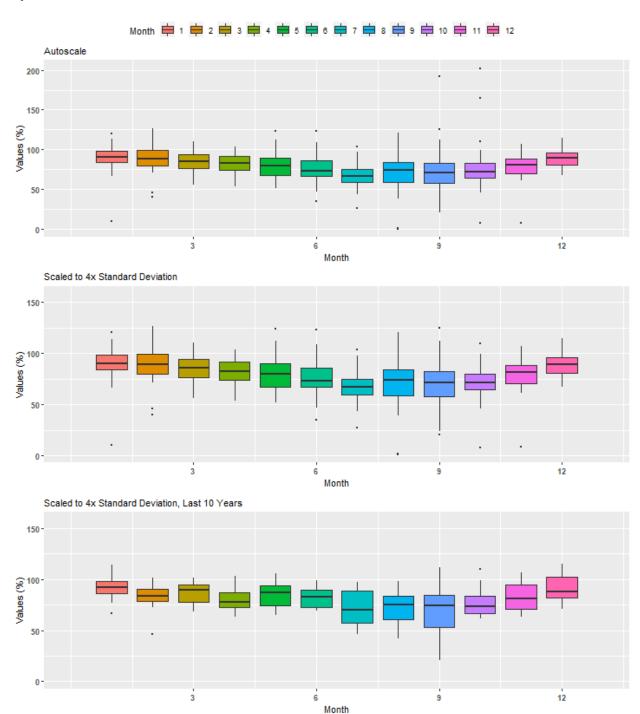


Summary Box Plots for Rookery Bay Aquatic Preserve

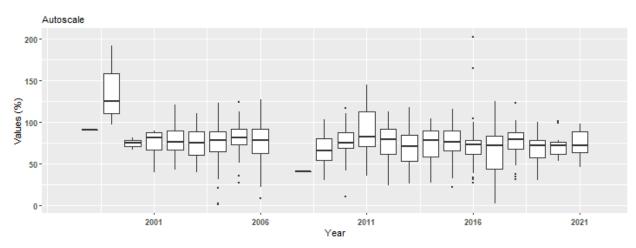


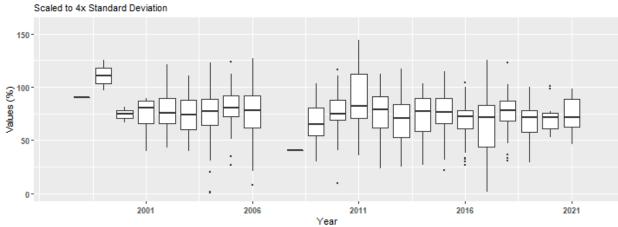
Summary Box Plots for Rookery Bay Aquatic Preserve

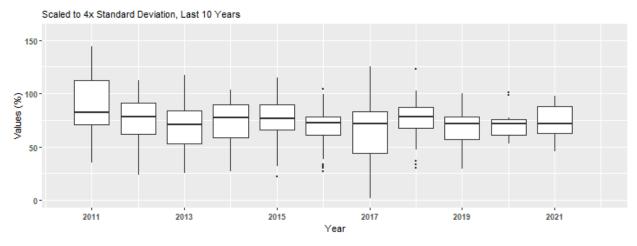
By Month



Summary Box Plots for Rookery Bay National Estuarine Research Reserve $\ensuremath{\mathsf{By\,Year}}$



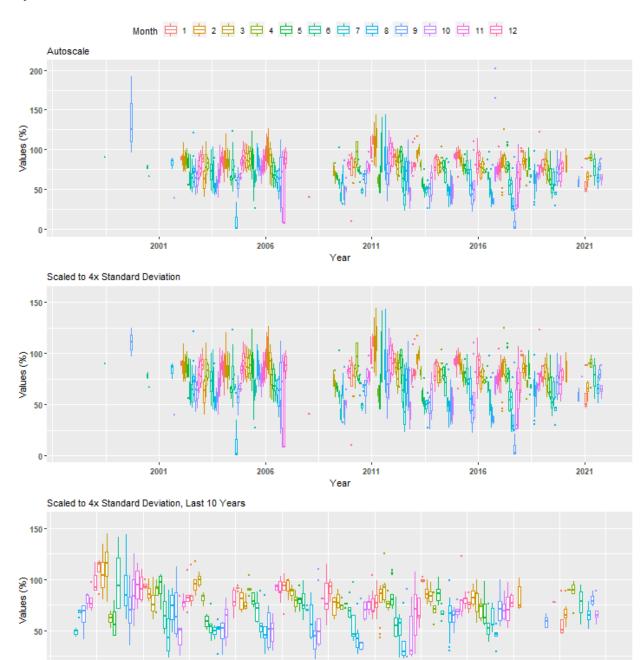




Summary Box Plots for Rookery Bay National Estuarine Research Reserve

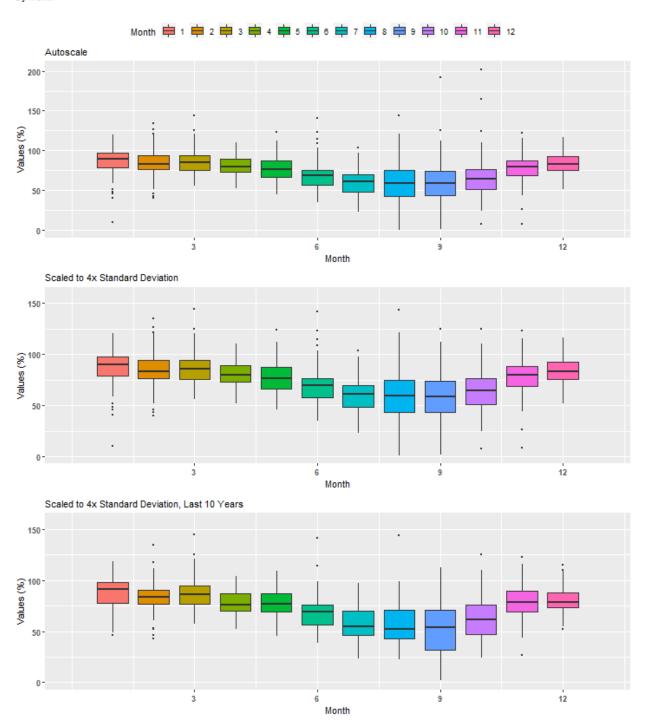
By Year & Month

0 -

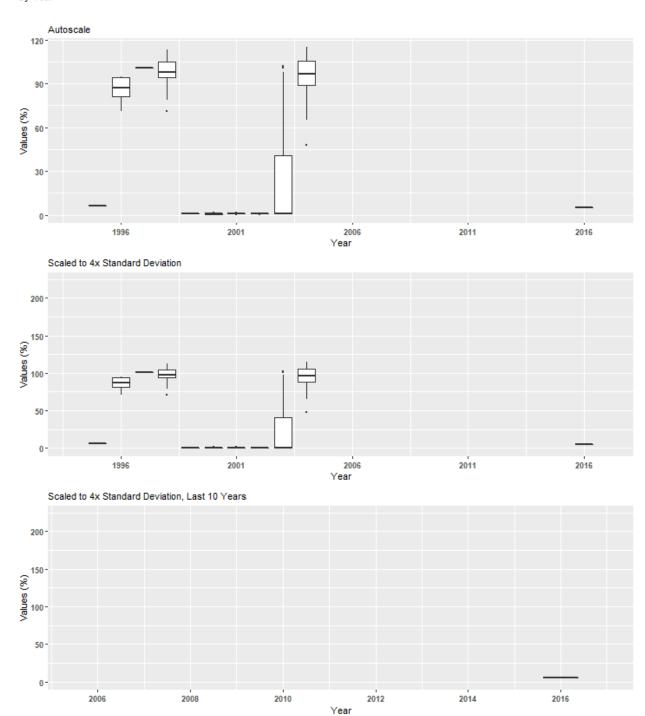


Year

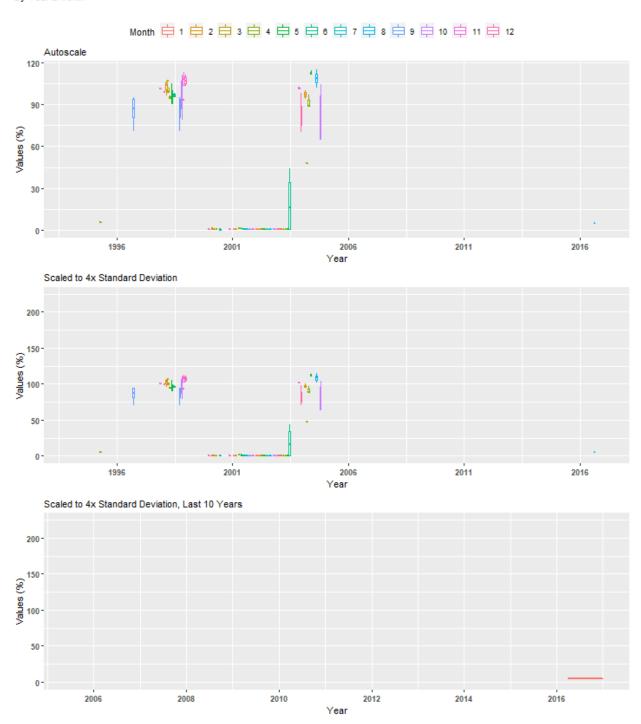
Summary Box Plots for Rookery Bay National Estuarine Research Reserve By Month



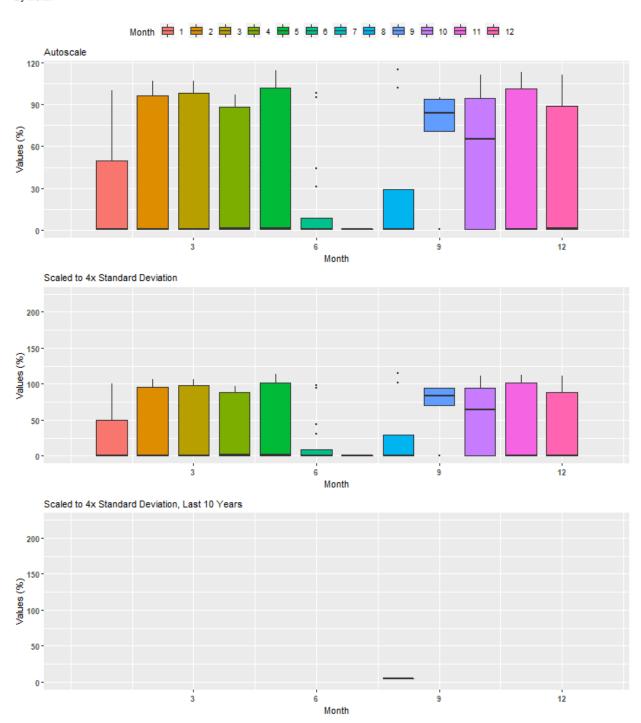
Summary Box Plots for Southeast Florida Coral Reef Ecosystem Conservation Area $\ensuremath{\mathsf{By}}\ \ensuremath{\mathsf{Year}}$



Summary Box Plots for Southeast Florida Coral Reef Ecosystem Conservation Area By Year & Month

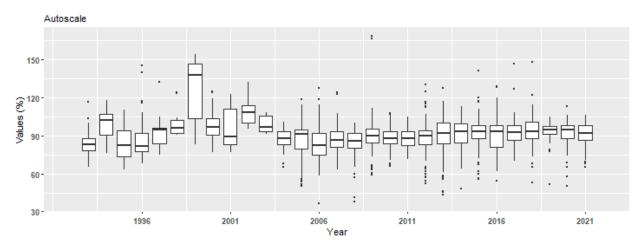


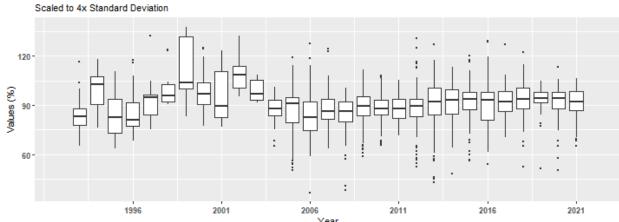
Summary Box Plots for Southeast Florida Coral Reef Ecosystem Conservation Area By Month

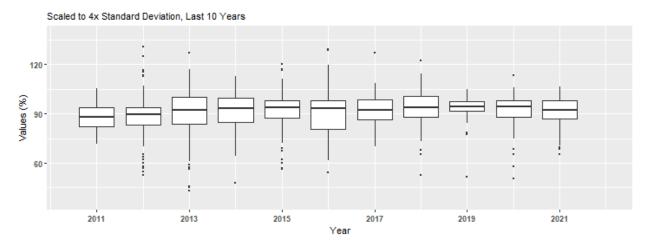


Summary Box Plots for Terra Ceia Aquatic Preserve

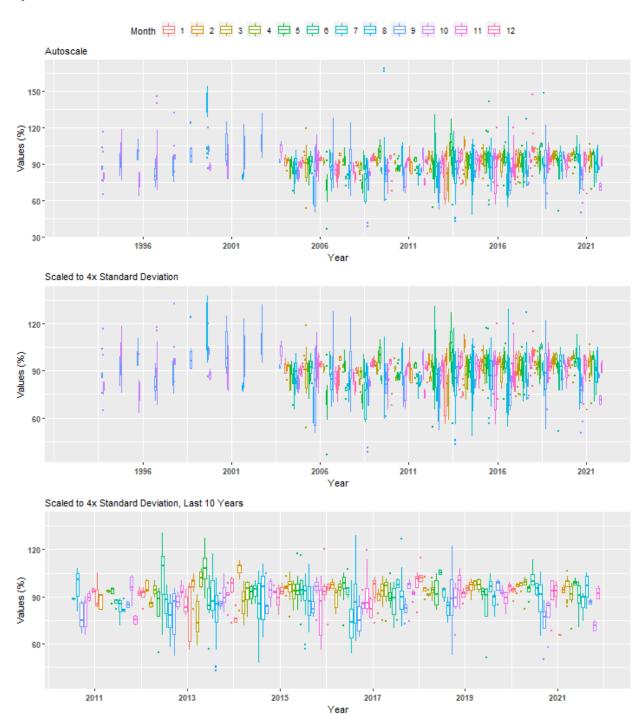
By Year





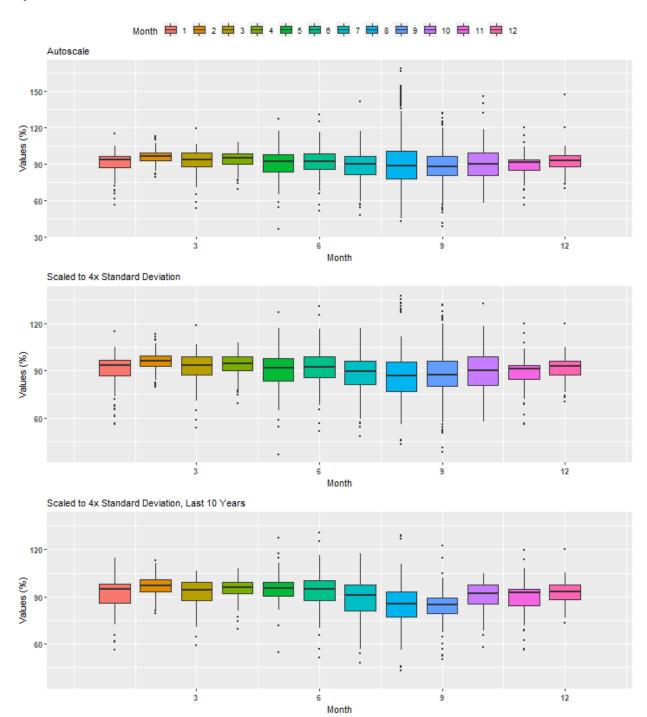


Summary Box Plots for Terra Ceia Aquatic Preserve



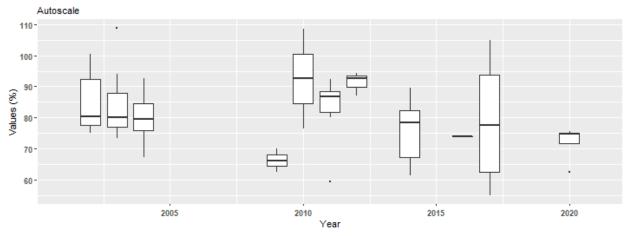
Summary Box Plots for Terra Ceia Aquatic Preserve

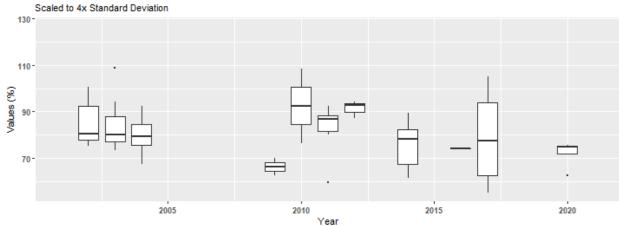
By Month

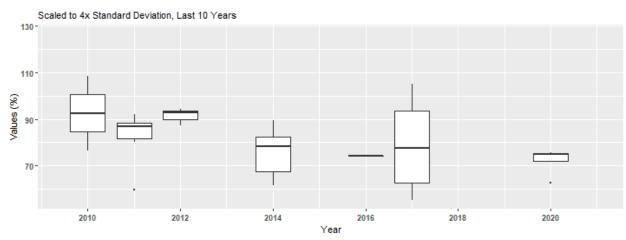


Summary Box Plots for Yellow River Marsh Aquatic Preserve

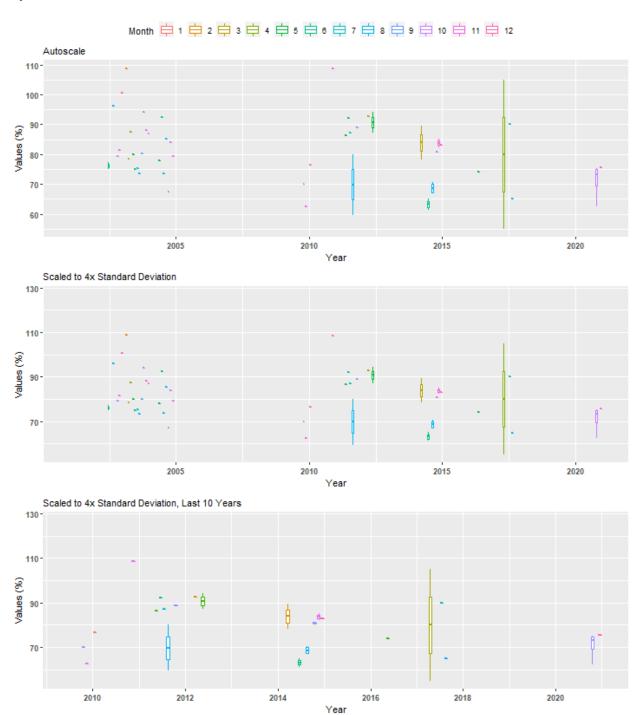
By Year







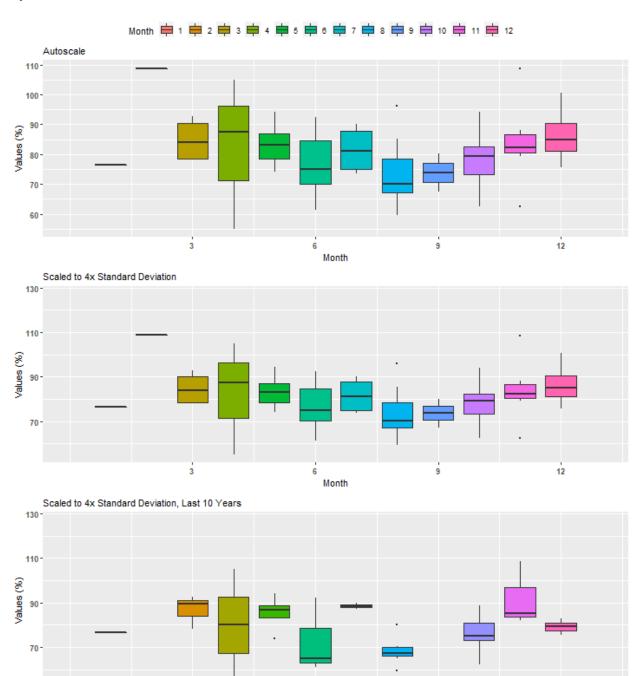
Summary Box Plots for Yellow River Marsh Aquatic Preserve



Summary Box Plots for Yellow River Marsh Aquatic Preserve

3

By Month



Month

12