

SEACAR Discrete Water Quality Analysis: Sample Bottom Total Phosphorus

Last compiled on 30 May, 2022

Contents

Libraries	1
File Import	2
Data Filtering and Data Impacted by Specific Value Qualifiers	2
Managed Area Statistics	5
Monitoring Location Statistics	7
Seasonal Kendall Tau Analysis	8
Appendix I: Scatter Plot of Entire Dataset	11
Appendix II: Dataset Summary Box Plots	13
Appendix III: Excluded Managed Areas	19
Appendix IV: Managed Area Trendlines	34
Appendix V: Managed Area Summary Box Plots	41

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

```
library(EnvStats)
library(tidyr)
options(scipen=999)
opts_chunk$set(warning=FALSE, message=FALSE)
```

File Import

Imports file that is determined in the WC_Discrete_parameter_ReportCompile.R script.

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 name of the parameter as it appears in the data file, units of the parameter, sets the `SampleDate` as a date object, and creates various scales of the date to be used by plotting functions.

```
data <- fread(file_in, sep="|", header=TRUE, stringsAsFactors=FALSE,
              select=c("ManagedAreaName", "ProgramID", "ProgramName",
                        "ProgramLocationID", "SampleDate", "Year", "Month",
                        "RelativeDepth", "ActivityType", "ParameterName",
                        "ResultValue", "ParameterUnits", "ValueQualifier",
                        "SEACAR_QAQCFlagCode", "Include"), na.strings="")

parameter <- unique(data$ParameterName)
unit <- unique(data$ParameterUnits)
data$SampleDate <- as.Date(data$SampleDate)
data$YearMonth <- paste0(data$Month, "-", data$Year)
data$YearMonthDec <- data$Year + ((data$Month-0.5) / 12)
data$DecDate <- decimal_date(data$SampleDate)
```

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.) and activity type (field or sample) 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. The script also only looks at data of the desired `ActivityType` which indicates whether it was measured in the field (`Field`) or in the lab (`Sample`).

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, S (for Secchi depth), 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$RelativeDepth[grepl("12Q", data$SEACAR_QAQCFlagCode[
    data$RelativeDepth=="Surface"])] <- "Bottom"
}

data$Include <- as.logical(data$Include)
data$Include[grepl("H", data$ValueQualifier[data$ProgramID==476])] <- TRUE
data <- data[!is.na(data$ResultValue),]

if(param_name!="Secchi_Depth"){
  data <- data[!is.na(data$RelativeDepth),]
  data <- data[data$RelativeDepth==depth,]
}

if(length(grep("Blank", data$ActivityType))>0){
  data <- data[-grep("Blank", data$ActivityType),]
}

if(param_name=="Chlorophyll_a_uncorrected_for_pheophytin" |
  param_name=="Salinity" | param_name=="Turbidity"){
  data <- data[grepl(activity, data$ActivityType[!is.na(data$ActivityType)]),]
}

if(param_name=="Water_Temperature"){
  data <- data[data$ResultValue>=-2,]
} else{
  data <- data[data$ResultValue>=0,]
}

data <- merge.data.frame(MA_All[,c("AreaID", "ManagedAreaName")],
  data, by="ManagedAreaName", all=TRUE)

MA_Summ <- data %>%
  group_by(AreaID, ManagedAreaName) %>%
  summarize(ParameterName=parameter,
    RelativeDepth=depth,
    ActivityType=activity,
    N_Data=length(ResultValue[Include==TRUE & !is.na(ResultValue)]),
    N_Years=length(unique(Year[Include==TRUE & !is.na(Year)])),
    EarliestYear=min(Year[Include==TRUE]),
    LatestYear=max(Year[Include==TRUE]),
    SufficientData=ifelse(N_Data>0 & N_Years>=10, TRUE, FALSE))

data <- merge.data.frame(data, MA_Summ[,c("ManagedAreaName", "SufficientData")],
  by="ManagedAreaName")

data$Use_In_Analysis <- ifelse(data$Include==TRUE & data$SufficientData==TRUE,
  TRUE, FALSE)

MA_Summ <- MA_Summ %>%
  select(AreaID, ManagedAreaName, ParameterName, RelativeDepth, ActivityType,
    SufficientData, everything())

```

```

MA_Summ <- as.data.frame(MA_Summ[order(MA_Summ$ManagedAreaName), ])

total <- length(data$Include)
pass_filter <- length(data$Include[data$Include==TRUE])

count_H <- length(grep("H", data$ValueQualifier[data$ProgramID==476]))
perc_H <- 100*count_H/length(data$ValueQualifier)

count_I <- length(grep("I", data$ValueQualifier))
perc_I <- 100*count_I/length(data$ValueQualifier)

count_Q <- length(grep("Q", data$ValueQualifier))
perc_Q <- 100*count_Q/length(data$ValueQualifier)

count_S <- length(grep("S", data$ValueQualifier))
perc_S <- 100*count_S/length(data$ValueQualifier)

count_U <- length(grep("U", data$ValueQualifier))
perc_U <- 100*count_U/length(data$ValueQualifier)

data$VQ_Plot <- data$ValueQualifier

inc_H <- ifelse(param_name=="pH" | param_name=="Dissolved_Oxygen" |
               param_name=="Dissolved_Oxygen_Saturation", TRUE, FALSE)

if (inc_H==TRUE){
  data$VQ_Plot <- gsub("[^HU]+", "", data$VQ_Plot)
  data$VQ_Plot <- gsub("UH", "HU", data$VQ_Plot)
  data$VQ_Plot[na.omit(data$ProgramID!=476)] <- gsub("[^U]+", "",
                                                    data$VQ_Plot[na.omit(data$ProgramID!=476)])
  data$VQ_Plot[data$VQ_Plot==""] <- NA

  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 if (param_name=="Secchi_Depth") {
  count_S <- length(grep("S", data$ValueQualifier))
  perc_S <- 100*count_S/length(data$ValueQualifier)
  data$VQ_Plot <- gsub("[^SU]+", "", data$VQ_Plot)
  data$VQ_Plot <- gsub("US", "SU", data$VQ_Plot)
  data$VQ_Plot[data$VQ_Plot==""] <- NA
  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",
             "S Codes: ", count_S, " (", round(perc_S, 6), "%)\n",
             "U Codes: ", count_U, " (", round(perc_U, 6), "%)"))
}

```

```

} else{
  data$VQ_Plot <- gsub("[^U]+", "", data$VQ_Plot)
  data$VQ_Plot[data$VQ_Plot==""] <- NA
  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: 15030, Number Passed Filter: 15012
## I Codes: 498 (3.313373%)
## Q Codes: 431 (2.867598%)
## U Codes: 237 (1.576846%)

```

```

data_summ <- data %>%
  group_by(AreaID, ManagedAreaName) %>%
  summarize(ParameterName=parameter,
            RelativeDepth=depth,
            ActivityType=activity,
            N_Total=length(ResultValue),
            N_AnalysisUse=length(ResultValue[SufficientData==TRUE]),
            N_H=length(grep("H", data$ValueQualifier[data$ProgramID==476])),
            perc_H=100*N_H/length(data$ValueQualifier),
            N_I=length(grep("I", data$ValueQualifier)),
            perc_I=100*N_I/length(data$ValueQualifier),
            N_Q=length(grep("Q", data$ValueQualifier)),
            perc_Q=100*N_Q/length(data$ValueQualifier),
            N_S=length(grep("S", data$ValueQualifier)),
            perc_S=100*N_S/length(data$ValueQualifier),
            N_U=length(grep("U", data$ValueQualifier)),
            perc_U=100*N_U/length(data$ValueQualifier))

data_summ <- as.data.table(data_summ[order(data_summ$ManagedAreaName), ])
fwrite(data_summ, paste0(out_dir, "/", param_name, "_", activity, "_", depth,
                        "_DataSummary.csv"), sep=",")

rm(data_summ)
MA_Include <- MA_Summ$ManagedAreaName[MA_Summ$SufficientData==TRUE &
                                       MA_Summ$N_Data<2000000]

n <- length(MA_Include)
MA_Exclude <- MA_Summ[MA_Summ$N_Years<10 & MA_Summ$N_Years>0,]
MA_Exclude <- MA_Exclude[,c("ManagedAreaName", "N_Years")]
z <- nrow(MA_Exclude)

```

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 `SufficientData` 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_YM_Stats <- data[data$Use_In_Analysis==TRUE, ] %>%
  group_by(AreaID, ManagedAreaName, Year, Month) %>%
  summarize(ParameterName=parameter,
             RelativeDepth=depth,
             ActivityType=activity,
             N_Data=length(ResultValue),
             Min=min(ResultValue),
             Max=max(ResultValue),
             Median=median(ResultValue),
             Mean=mean(ResultValue),
             StandardDeviation=sd(ResultValue),
             ProgramIDs=paste(sort(unique(ProgramID), decreasing=FALSE),
                              collapse=', '))
MA_YM_Stats <- as.data.table(MA_YM_Stats[order(MA_YM_Stats$ManagedAreaName,
                                              MA_YM_Stats$Year,
                                              MA_YM_Stats$Month), ])
fwrite(MA_YM_Stats, paste0(out_dir, "/", param_name, "_", activity, "_", depth,
                           "_ManagedArea_YearMonth_Stats.txt"), sep="|")
rm(MA_YM_Stats)

MA_Y_Stats <- data[data$Use_In_Analysis==TRUE, ] %>%
  group_by(AreaID, ManagedAreaName, Year) %>%
  summarize(ParameterName=parameter,
             RelativeDepth=depth,
             ActivityType=activity,
             N=length(ResultValue),
             Min=min(ResultValue),
             Max=max(ResultValue),
             Median=median(ResultValue),
             Mean=mean(ResultValue),
             StandardDeviation=sd(ResultValue),
             ProgramIDs=paste(sort(unique(ProgramID), decreasing=FALSE),
                              collapse=', '))
MA_Y_Stats <- as.data.table(MA_Y_Stats[order(MA_Y_Stats$ManagedAreaName,
                                              MA_Y_Stats$Year), ])
fwrite(MA_Y_Stats, paste0(out_dir, "/", param_name, "_", activity, "_", depth,
                           "_ManagedArea_Year_Stats.txt"), sep="|")
rm(MA_Y_Stats)

MA_M_Stats <- data[data$Use_In_Analysis==TRUE, ] %>%
  group_by(AreaID, ManagedAreaName, Month) %>%
```

```

summarize(ParameterName=parameter,
           RelativeDepth=depth,
           ActivityType=activity,
           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, "_", activity, "_", depth,
                           "_ManagedArea_Month_Stats.txt"), sep="|")
rm(MA_M_Stats)

```

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 `SufficientData` 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 `ProgramID` 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](#)

```

Mon_Stats <- data[data$Use_In_Analysis==TRUE, ] %>%
  group_by(AreaID, ManagedAreaName, ProgramID, ProgramName, ProgramLocationID) %>%
  summarize(ParameterName=parameter,
           RelativeDepth=depth,
           ActivityType=activity,
           EarliestSampleDate=min(SampleDate),
           LastSampleDate=max(SampleDate),
           N=length(ResultValue),
           Min=min(ResultValue),
           Max=max(ResultValue),
           Median=median(ResultValue),
           Mean=mean(ResultValue),
           StandardDeviation=sd(ResultValue))

Mon_Stats <- as.data.table(Mon_Stats[order(Mon_Stats$ManagedAreaName,
                                           Mon_Stats$ProgramName,
                                           Mon_Stats$ProgramID,
                                           Mon_Stats$ProgramLocationID), ])

```

```
fwrite(Mon_Stats, paste0(out_dir,"/", param_name, "_", activity, "_", depth,
                          "_MonitoringLoc_Stats.txt"), sep="|")
rm(Mon_Stats)
```

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 `SufficientData` 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

```
tauSeasonal <- function(data, independent, stats.median, stats.minYear,
                        stats.maxYear) {
  tau <- NULL
  tryCatch({ken <- kendallSeasonalTrendTest(
    y = data$ResultValue,
    season = data$Month,
    year = data$Year,
    independent.obs = independent)

  tau <- ken$estimate[1]
  p <- ken$p.value[2]
  slope <- ken$estimate[2]
  intercept <- ken$estimate[3]
  trend <- trend_calculator(slope, stats.median, p)
  rm(ken)
}, warning = function(w) {
  print(w)
}, error = function(e) {
  print(e)
```



```

    }, finally = {
      if (!exists("tau")) {
        tau <- NA
      }
      if (!exists("p")) {
        p <- NA
      }
      if (!exists("slope")) {
        slope <- NA
      }
      if (!exists("intercept")) {
        intercept <- NA
      }
      if (!exists("trend")) {
        trend <- NA
      }
    })
    KT <- c(unique(data$AreaID),
            unique(data$ManagedAreaName),
            stats.median,
            independent,
            tau,
            p,
            slope,
            intercept,
            trend)
    return(KT)
  }
runStats <- function(data) {
  data$Index <- as.Date(data$SampleDate) # , "%Y-%m-%d")
  data$ResultValue <- as.numeric(data$ResultValue)
  # Calculate basic stats
  stats.median <- median(data$ResultValue, na.rm = TRUE)
  stats.minYear <- min(data$Year, na.rm = TRUE)
  stats.maxYear <- max(data$Year, na.rm = TRUE)
  # Calculate Kendall Tau and Slope stats, then update appropriate columns and table
  KT <- tauSeasonal(data, TRUE, stats.median,
                    stats.minYear, stats.maxYear)
  if (is.null(KT[9])) {
    KT <- tauSeasonal(data, FALSE, stats.median,
                      stats.minYear, stats.maxYear)
  }
  if (is.null(KT.Stats) == TRUE) {
    KT.Stats <- KT
  } else{
    KT.Stats <- rbind(KT.Stats, KT)
  }
  return(KT.Stats)
}
trend_calculator <- function(slope, median_value, p) {
  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.) {
    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$ManagedAreaName
c_names <- c("AreaID", "ManagedAreaName", "Median", "Independent",
             "tau", "p", "SennSlope", "SennIntercept", "Trend")
if(n==0){
    KT.Stats <- data.frame(matrix(ncol=length(c_names),
                                nrow=length(MA_Summ$ManagedAreaName)))
    colnames(KT.Stats) <- c_names
    KT.Stats[, c("AreaID", "ManagedAreaName")] <-
        MA_Summ[, c("AreaID", "ManagedAreaName")]
} else{
    for (i in 1:n) {
        x <- nrow(data[data$Use_In_Analysis == TRUE &
                       data$ManagedAreaName == MA_Include[i], ])
        if (x>0) {
            KT.Stats <- runStats(data[data$Use_In_Analysis == TRUE &
                                     data$ManagedAreaName ==
                                     MA_Include[i], ])
        }
    }
    KT.Stats <- as.data.frame(KT.Stats)
    c_names <- c("AreaID", "ManagedAreaName", "Median", "Independent",
                "tau", "p", "SennSlope", "SennIntercept", "Trend")
    if(dim(KT.Stats)[2]==1){
        KT.Stats <- as.data.frame(t(KT.Stats))
    }
    colnames(KT.Stats) <- c_names
    rownames(KT.Stats) <- seq(1:nrow(KT.Stats))
    KT.Stats$tau <- round(as.numeric(KT.Stats$tau), digits=4)
    KT.Stats$p <- round(as.numeric(KT.Stats$p), digits=4)
    KT.Stats$SennSlope <- as.numeric(KT.Stats$SennSlope)
    KT.Stats$SennIntercept <- as.numeric(KT.Stats$SennIntercept)
    KT.Stats$Trend <- as.integer(KT.Stats$Trend)
}

```

```

KT.Stats <- merge.data.frame(MA_Summ, KT.Stats,
                             by=c("AreaID", "ManagedAreaName"), all=TRUE)

KT.Stats <- as.data.table(KT.Stats[order(KT.Stats$ManagedAreaName), ])

fwrite(KT.Stats, paste0(out_dir, "/", param_name, "_", activity, "_", depth,
                          "_KendallTauStats.txt"), sep="|")

data <- data[!is.na(data$ResultValue),]

```

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.

```

plot_theme <- theme_bw() + theme(text=element_text(family="Segoe UI"),
                                   title=element_text(face="bold"),
                                   plot.title=element_text(hjust=0.5, size=14, color="#314963"),
                                   plot.subtitle=element_text(hjust=0.5, size=10, color="#314963"),
                                   axis.text.x=element_text(face="bold"),
                                   axis.text.y=element_text(face="bold"))

year_lower <- min(data$Year)
year_upper <- max(data$Year)
min_RV <- min(data$ResultValue)
mn_RV <- mean(data$ResultValue[data$ResultValue <
                                quantile(data$ResultValue, 0.98)])
sd_RV <- sd(data$ResultValue[data$ResultValue <
                              quantile(data$ResultValue, 0.98)])
x_scale <- ifelse(year_upper - year_lower > 30, 10, 5)
y_scale <- mn_RV + 4 * sd_RV

p1 <- ggplot(data=data[data$Include==TRUE,],
             aes(x=SampleDate, y=ResultValue, fill=VQ_Plot)) +
  geom_point(shape=21, size=3, color="#333333", alpha=0.75) +
  labs(subtitle="Autoscale",
       x="Year", y=paste0("Values (", unit, ")"),
       fill="Value Qualifier") +
  plot_theme +
  theme(legend.position="top", legend.box="horizontal",
        legend.justification="right") +
  scale_x_date(labels=date_format("%Y")) +
  {if(inc_H==TRUE){
    scale_fill_manual(values=c("H"= "#F8766D", "U"= "#00BFC4",
                               "HU"= "#7CAE00"), na.value="#cccccc")
  } else if(param_name=="Secchi_Depth"){
    scale_fill_manual(values=c("S"= "#F8766D", "U"= "#00BFC4",
                               "SU"= "#7CAE00"), na.value="#cccccc")
  } else {

```

```

    scale_fill_manual(values=c("U"= "#00BFC4"), na.value="#cccccc")
  }}

p2 <- ggplot(data=data[data$Include==TRUE,],
             aes(x=SampleDate, y=ResultValue, fill=VQ_Plot)) +
  geom_point(shape=21, size=3, color="#333333", alpha=0.75) +
  ylim(min_RV, y_scale) +
  labs(subtitle="Scaled to 4x Standard Deviation",
       x="Year", y=paste0("Values (", unit, ")")) +
  plot_theme +
  theme(legend.position="none") +
  scale_x_date(labels=date_format("%Y")) +
  {if(inc_H==TRUE){
    scale_fill_manual(values=c("H"= "#F8766D", "U"= "#00BFC4",
                              "HU"="#7CAE00"), na.value="#cccccc")
  } else if(param_name=="Secchi_Depth"){
    scale_fill_manual(values=c("S"= "#F8766D", "U"= "#00BFC4",
                              "SU"="#7CAE00"), na.value="#cccccc")
  } else {
    scale_fill_manual(values=c("U"= "#00BFC4"), na.value="#cccccc")
  }}

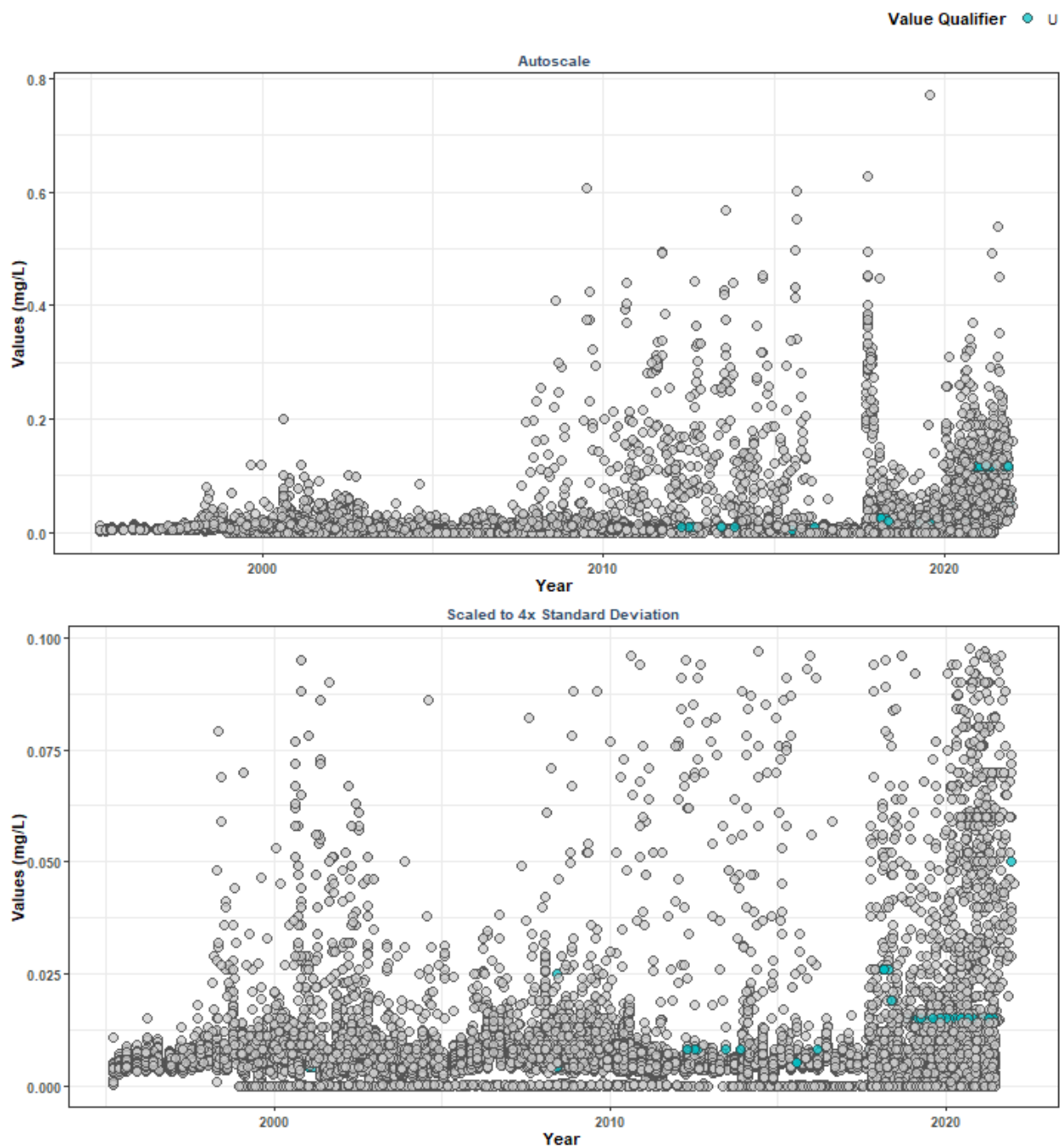
leg <- get_legend(p1)
pset <- ggarrange(leg, p1 + theme(legend.position="none"), p2,
                 ncol=1, heights=c(0.1, 1, 1))

p0 <- ggplot() + labs(title="Scatter Plot for Entire Dataset") +
  plot_theme + theme(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 `SufficientData` 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 y 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])
mn_RV <- mean(data$ResultValue[data$Include==TRUE &
                                data$ResultValue <
                                quantile(data$ResultValue, 0.98)])
sd_RV <- sd(data$ResultValue[data$Include==TRUE &
                              data$ResultValue <
                              quantile(data$ResultValue, 0.98)])
y_scale <- mn_RV + 4 * sd_RV

p1 <- ggplot(data=data[data$Include==TRUE, ],
             aes(x=Year, y=ResultValue, group=Year)) +
  geom_boxplot(color="#333333", fill="#cccccc", outlier.shape=21,
               outlier.size=3, outlier.color="#333333",
               outlier.fill="#cccccc", outlier.alpha=0.75) +
  labs(subtitle="Autoscale", x="Year",
       y=paste0("Values (", unit, ")")) +
  plot_theme

p2 <- ggplot(data=data[data$Include==TRUE, ],
             aes(x=Year, y=ResultValue, group=Year)) +
  geom_boxplot(color="#333333", fill="#cccccc", outlier.shape=21,
               outlier.size=3, outlier.color="#333333",
               outlier.fill="#cccccc", outlier.alpha=0.75) +
  labs(subtitle="Scaled to 4x Standard Deviation", x="Year",
       y=paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
  plot_theme

p3 <- ggplot(data=data[data$Include==TRUE, ],
             aes(x=as.integer(Year), y=ResultValue, group=Year)) +
  geom_boxplot(color="#333333", fill="#cccccc", outlier.shape=21,
               outlier.size=3, outlier.color="#333333",
               outlier.fill="#cccccc", outlier.alpha=0.75) +
  labs(subtitle="Scaled to 4x Standard Deviation, Last 10 Years",
       x="Year", y=paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
  scale_x_continuous(limits=c(max(data$Year) - 10.5, max(data$Year)+1),
                    breaks=seq(max(data$Year) - 10, max(data$Year), 2)) +
  plot_theme

set <- ggarrange(p1, p2, p3, ncol=1)

p0 <- ggplot() + labs(title="Summary Box Plots for Entire Data",
```

```

        subtitle="By Year") + plot_theme +
    theme(panel.border=element_blank(), panel.grid.major=element_blank(),
          panel.grid.minor=element_blank(), axis.line=element_blank())

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

```

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, ],
             aes(x=YearMonthDec, y=ResultValue,
                 group=YearMonth, color=as.factor(Month))) +
  geom_boxplot(fill="#cccccc", outlier.size=1.5, outlier.alpha=0.75) +
  labs(subtitle="Autoscale", x="Year",
       y=paste0("Values (", unit, ")"), color="Month") +
  plot_theme +
  theme(legend.position="top", legend.box="horizontal") +
  guides(color=guide_legend(nrow=1))

p2 <- ggplot(data=data[data$Include==TRUE, ],
             aes(x=YearMonthDec, y=ResultValue,
                 group=YearMonth, color=as.factor(Month))) +
  geom_boxplot(fill="#cccccc", outlier.size=1.5, outlier.alpha=0.75) +
  labs(subtitle="Scaled to 4x Standard Deviation",
       x="Year", y=paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
  plot_theme +
  theme(legend.position="none", axis.text.x=element_text(face="bold"),
        axis.text.y=element_text(face="bold"))

p3 <- ggplot(data=data[data$Include==TRUE, ],
             aes(x=YearMonthDec, y=ResultValue,
                 group=YearMonth, color=as.factor(Month))) +
  geom_boxplot(fill="#cccccc", outlier.size=1.5, outlier.alpha=0.75) +
  labs(subtitle="Scaled to 4x Standard Deviation, Last 10 Years",
       x="Year", y=paste0("Values (", unit, ")")) +
  ylim(min_RV, y_scale) +
  scale_x_continuous(limits=c(max(data$Year) - 10.5, max(data$Year)+1),
                    breaks=seq(max(data$Year) - 10, max(data$Year), 2)) +
  plot_theme +
  theme(legend.position="none")
leg <- get_legend(p1)
set <- ggarrange(leg, p1 + theme(legend.position="none"), p2, p3, ncol=1,
                 heights=c(0.1, 1, 1, 1))

p0 <- ggplot() + labs(title="Summary Box Plots for Entire Data",
                    subtitle="By Year & Month") + plot_theme +
  theme(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(color="#333333", outlier.shape=21, outlier.size=3,
              outlier.color="#333333", outlier.alpha=0.75) +
  labs(subtitle="Autoscale", x="Month",
       y=paste0("Values (", unit, ")"), fill="Month") +
  scale_x_continuous(limits=c(0, 13), breaks=seq(3, 12, 3)) +
  plot_theme +
  theme(legend.position="top", legend.box="horizontal") +
  guides(fill=guide_legend(nrow=1))

p2 <- ggplot(data=data[data$Include==TRUE, ],
             aes(x=Month, y=ResultValue,
                 group=Month, fill=as.factor(Month))) +
  geom_boxplot(color="#333333", outlier.shape=21, outlier.size=3,
              outlier.color="#333333", outlier.alpha=0.75) +
  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)) +
  plot_theme +
  theme(legend.position="none")

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(color="#333333", outlier.shape=21, outlier.size=3,
              outlier.color="#333333", outlier.alpha=0.75) +
  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)) +
  plot_theme +
  theme(legend.position="none")

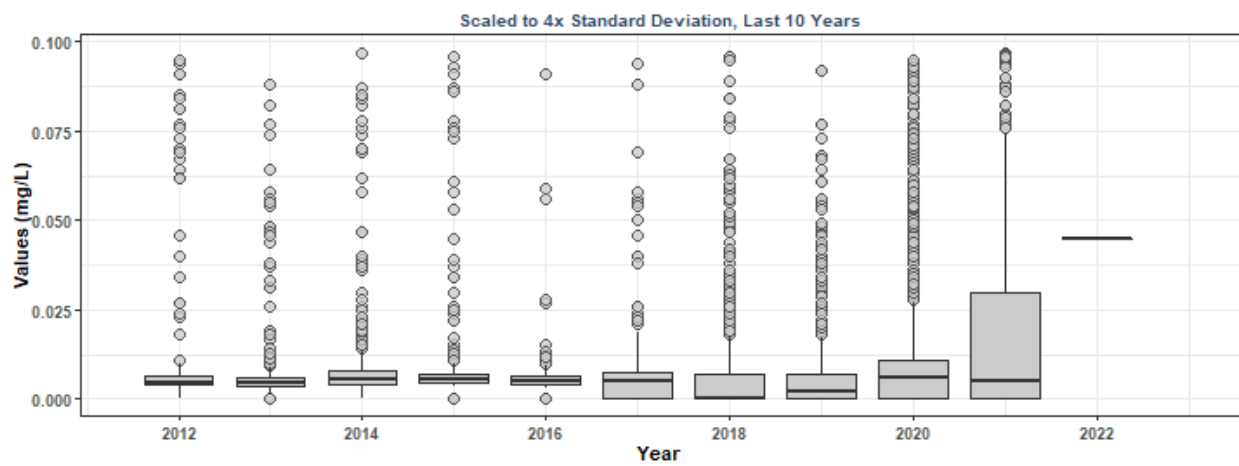
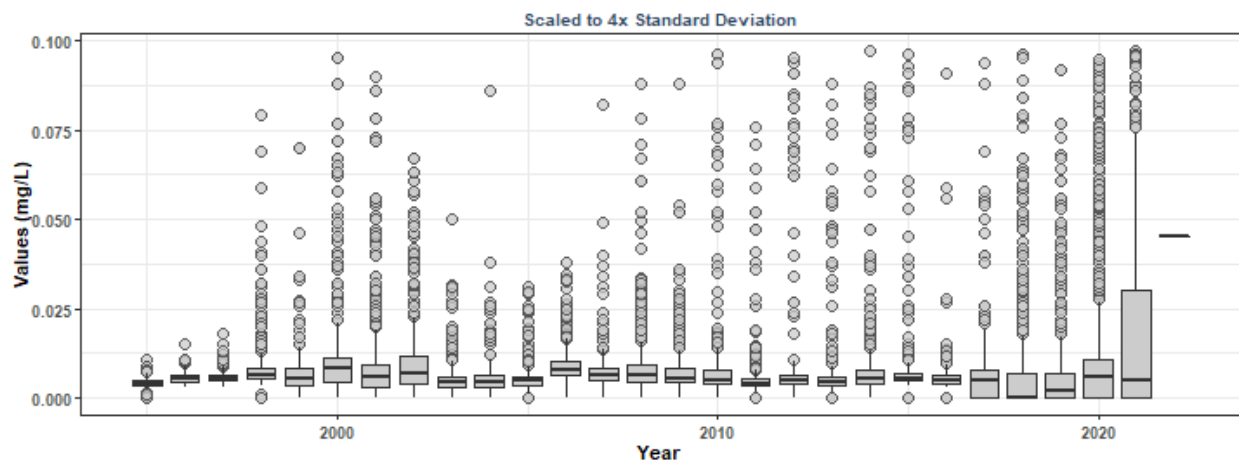
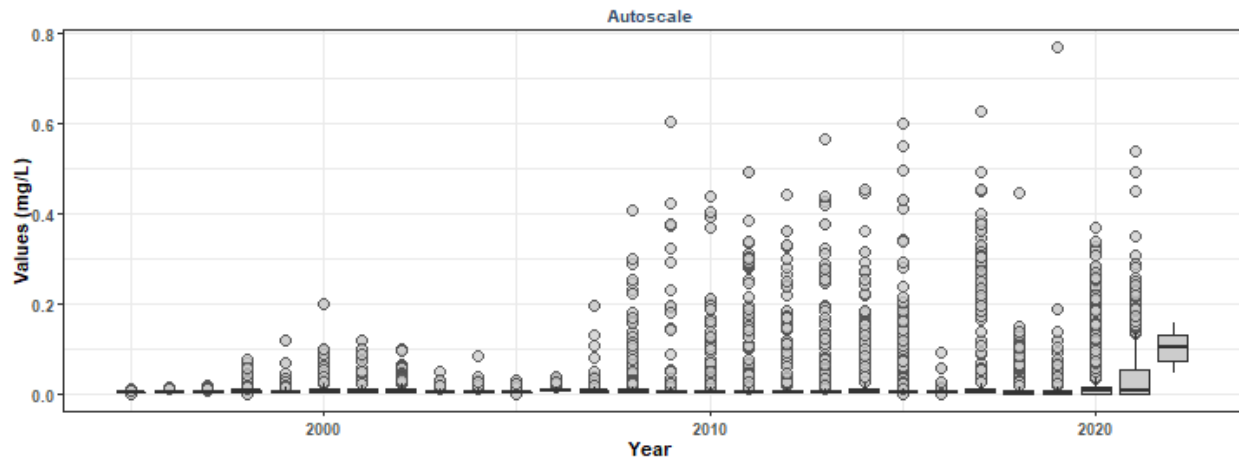
leg <- get_legend(p1)
set <- ggarrange(leg, p1 + theme(legend.position="none"), p2, p3, ncol=1,
                heights=c(0.1, 1, 1, 1))

p0 <- ggplot() + labs(title="Summary Box Plots for Entire Data",
                    subtitle="By Month") + plot_theme +
  theme(panel.border=element_blank(), panel.grid.major=element_blank(),
        panel.grid.minor=element_blank(), axis.line=element_blank())

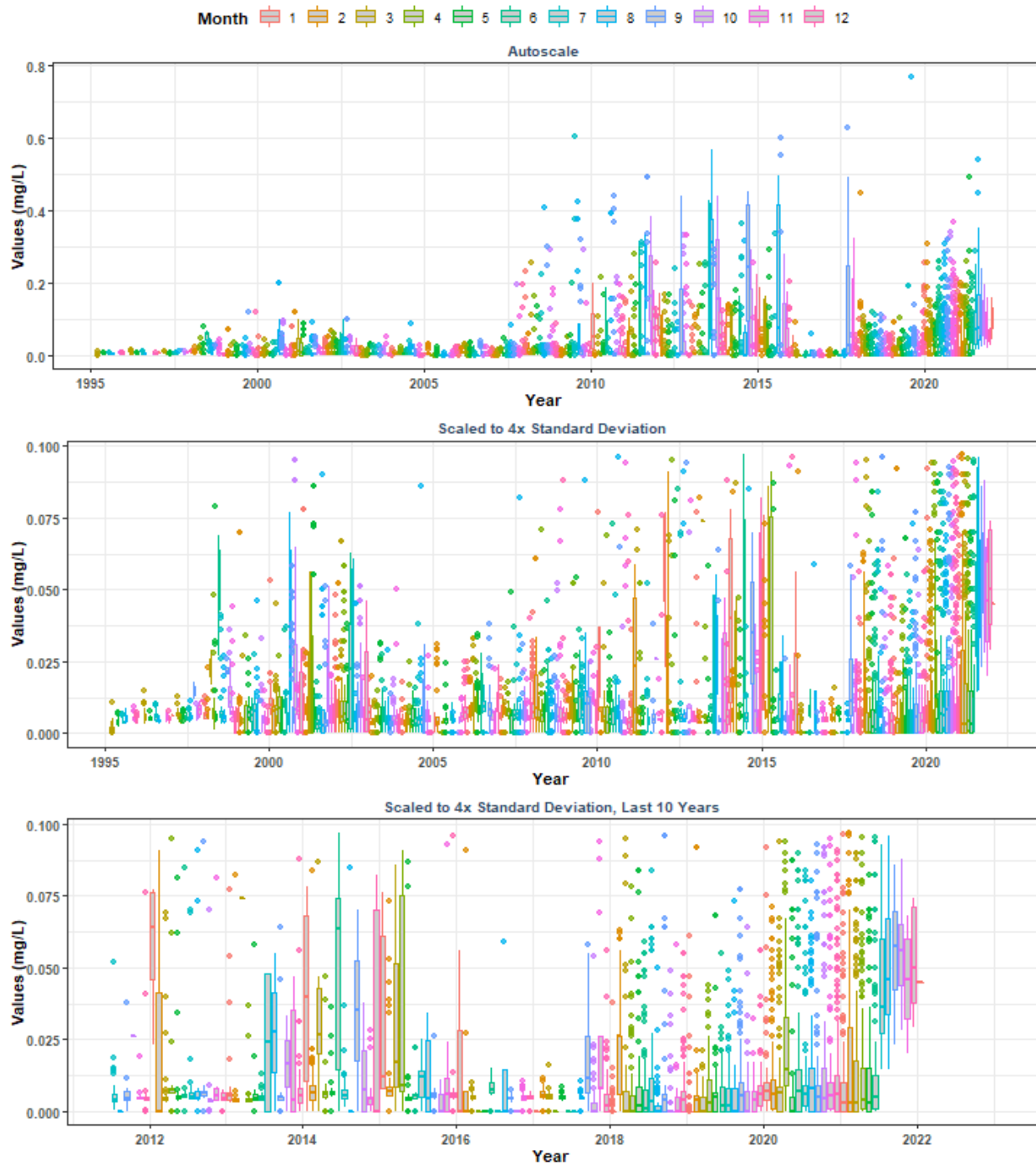
Mset <- ggarrange(p0, set, ncol=1, heights=c(0.07, 1))

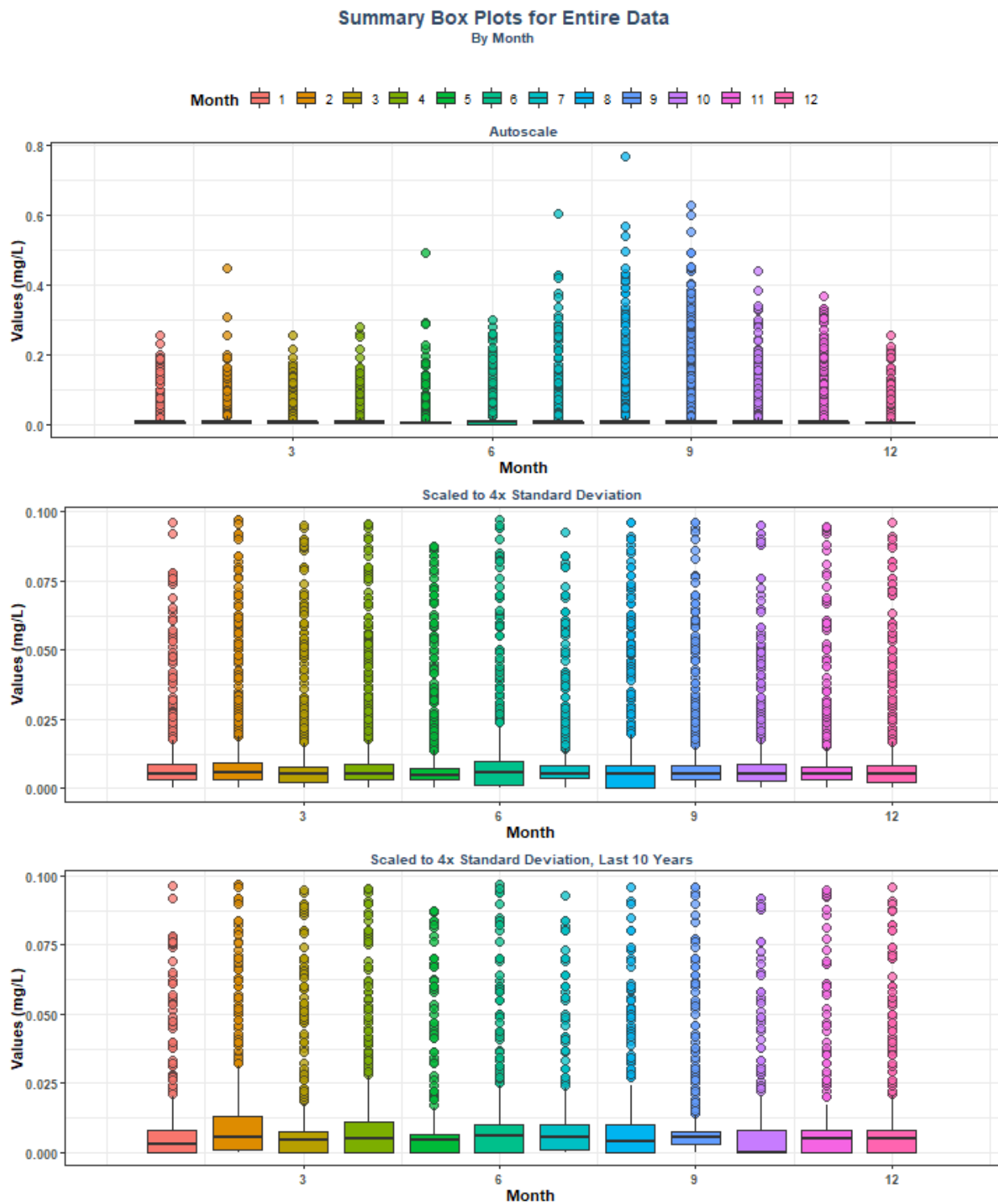
```


Summary Box Plots for Entire Data By Year



Summary Box Plots for Entire Data By Year & Month





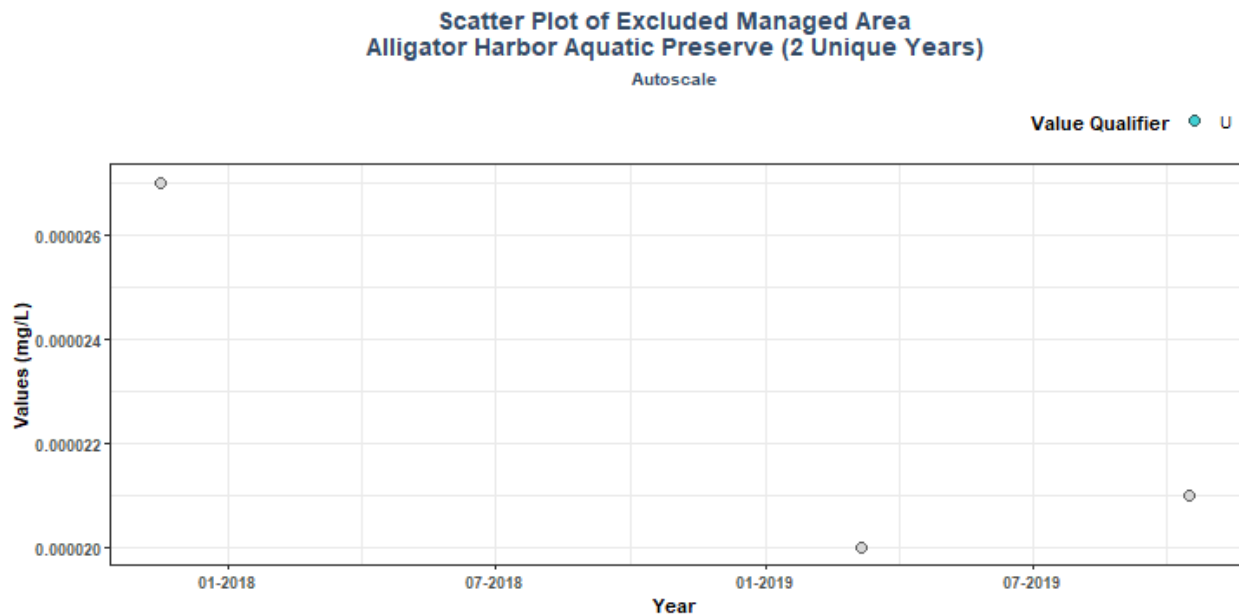
Appendix III: Excluded Managed Areas

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.

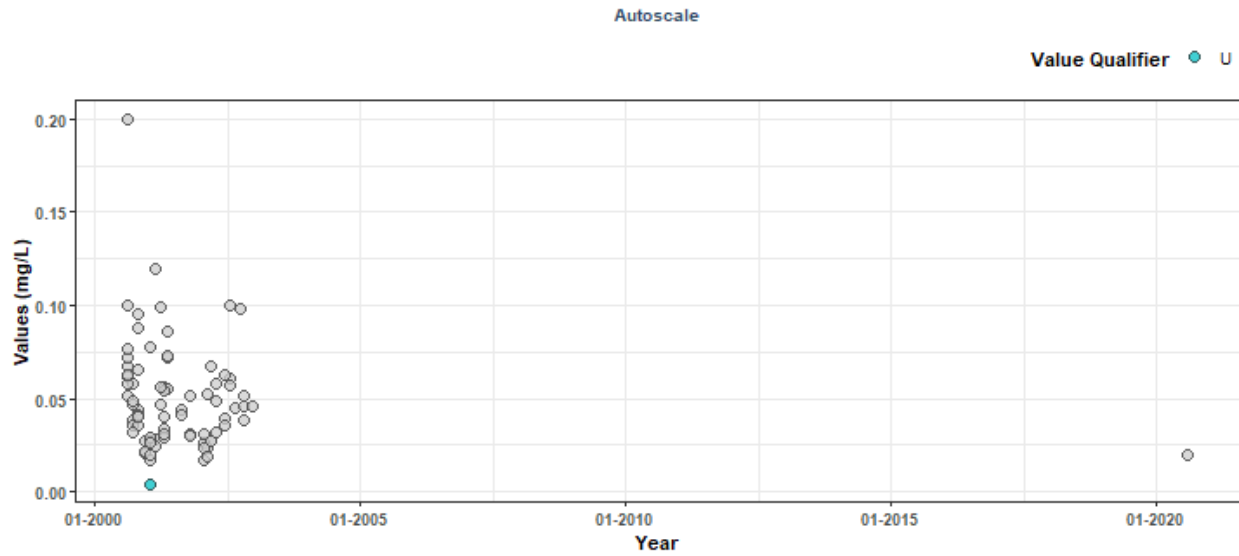
```

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$Include==TRUE, ],
      aes(x=SampleDate, y=ResultValue, fill=VQ_Plot)) +
    geom_point(shape=21, size=3, color="#333333", alpha=0.75) +
    labs(title=paste0("Scatter Plot of Excluded Managed Area\n",
      MA_Exclude$ManagedAreaName[i], " (",
      MA_Exclude$N_Years[i], " Unique Years)"),
      subtitle="Autoscale", x="Year",
      y=paste0("Values (", unit, ")"), fill="Value Qualifier") +
    plot_theme +
    theme(legend.position="top", legend.box="horizontal",
      legend.justification="right") +
    scale_x_date(labels=date_format("%m-%Y")) +
    {if(inc_H==TRUE){
      scale_fill_manual(values=c("H"= "#F8766D", "U"= "#00BFC4",
        "HU"="#7CAE00"), na.value="#cccccc")
    } else if(param_name=="Secchi_Depth"){
      scale_fill_manual(values=c("S"= "#F8766D", "U"= "#00BFC4",
        "SU"="#7CAE00"), na.value="#cccccc")
    } else {
      scale_fill_manual(values=c("U"= "#00BFC4"), na.value="#cccccc")
    }}
    print(p1)
  }
}

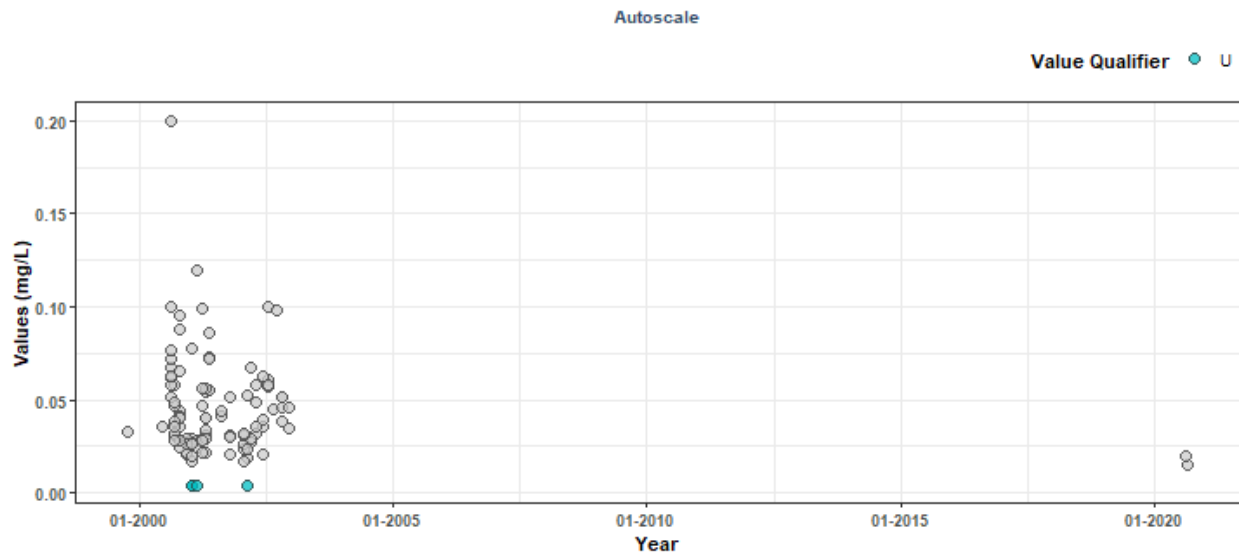
```



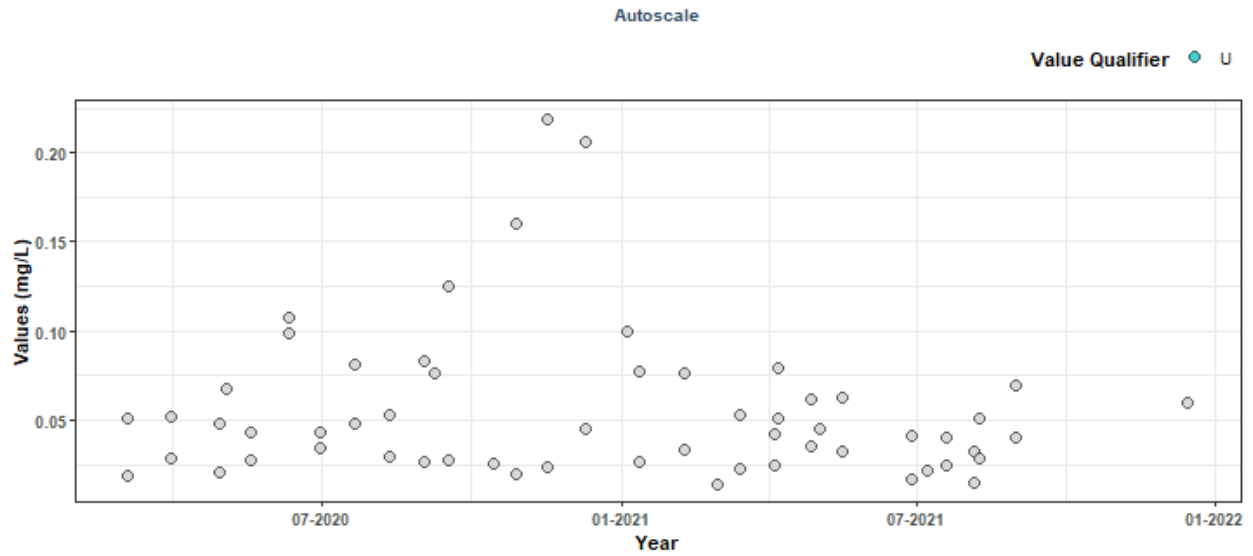
Scatter Plot of Excluded Managed Area
Apalachicola Bay Aquatic Preserve (4 Unique Years)



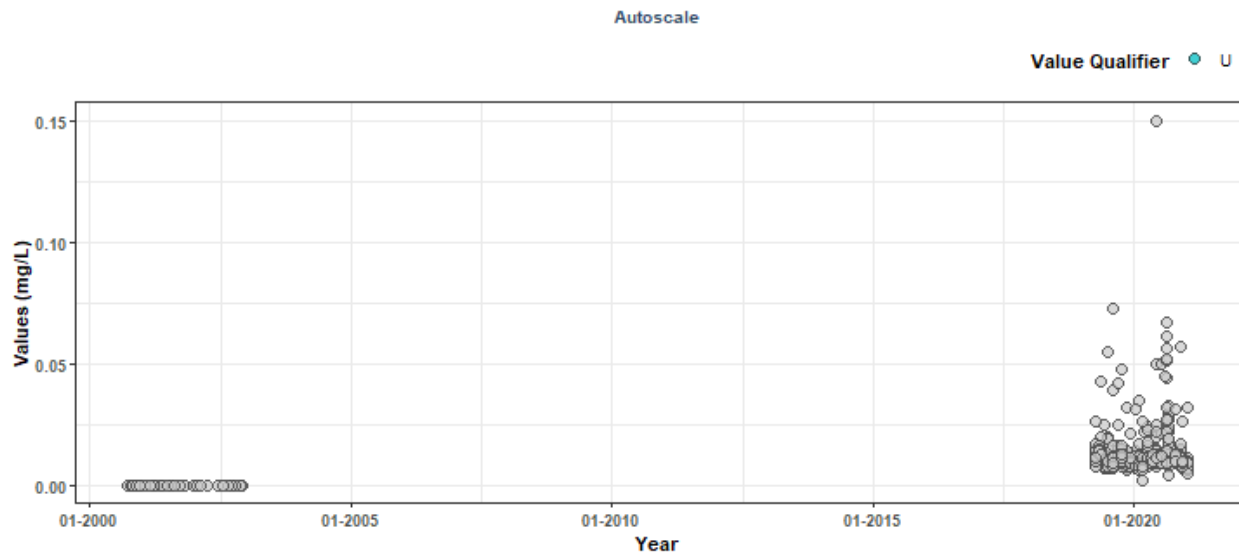
Scatter Plot of Excluded Managed Area
Apalachicola National Estuarine Research Reserve (5 Unique Years)



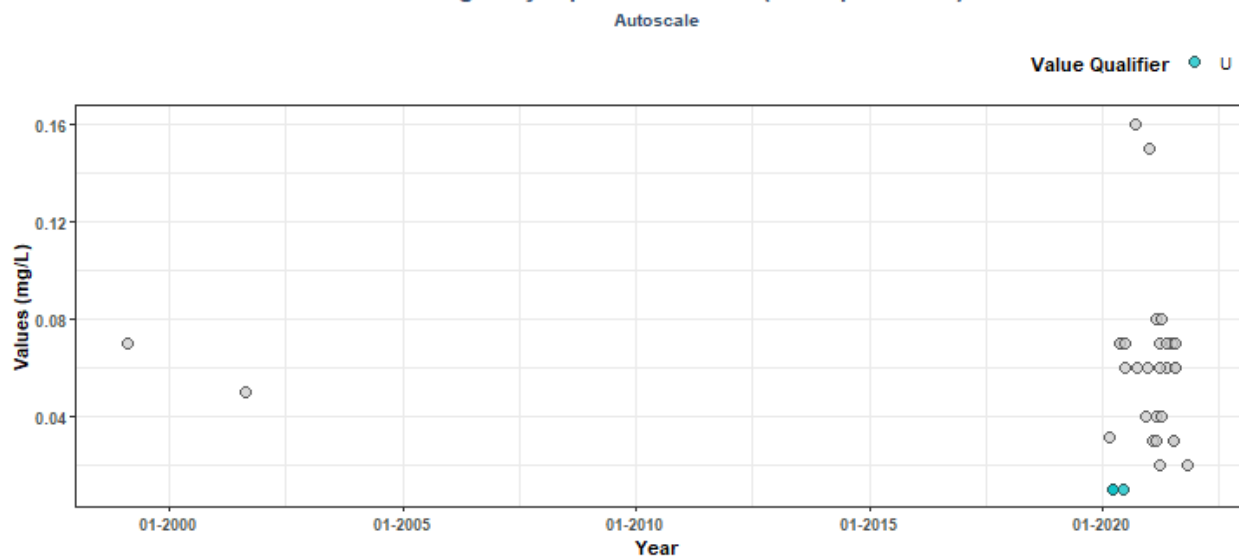
Scatter Plot of Excluded Managed Area
Banana River Aquatic Preserve (2 Unique Years)



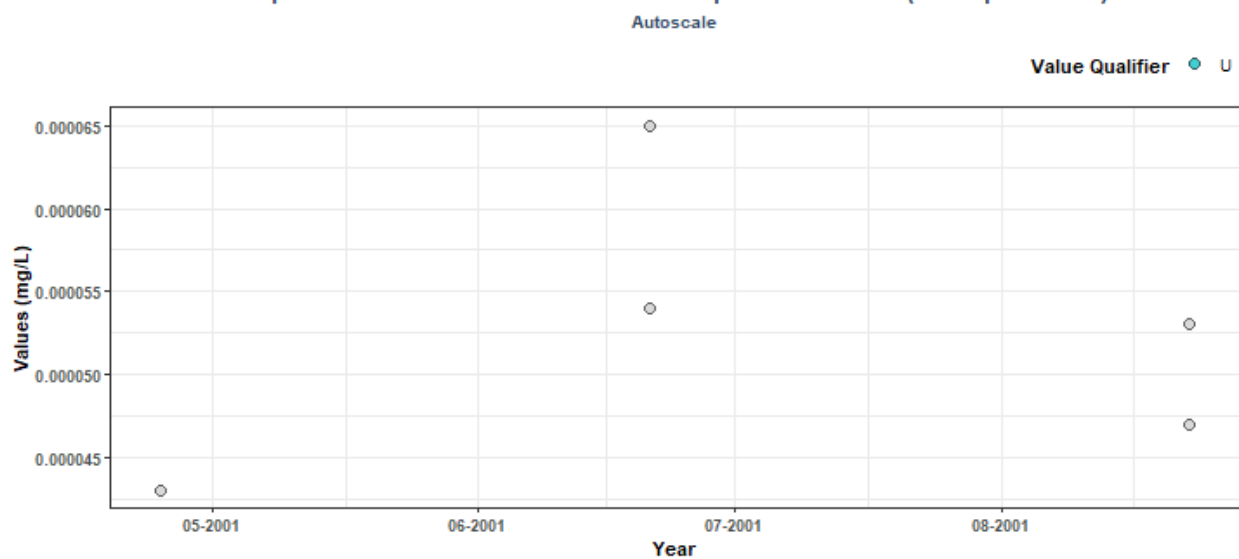
Scatter Plot of Excluded Managed Area
Biscayne Bay Aquatic Preserve (6 Unique Years)



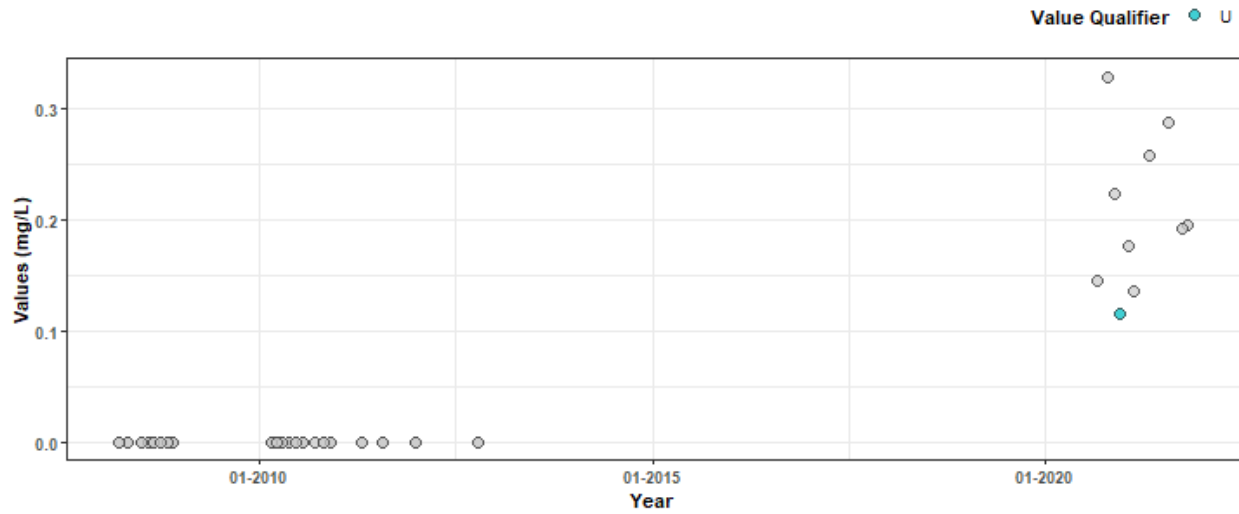
Scatter Plot of Excluded Managed Area
Boca Ciega Bay Aquatic Preserve (4 Unique Years)



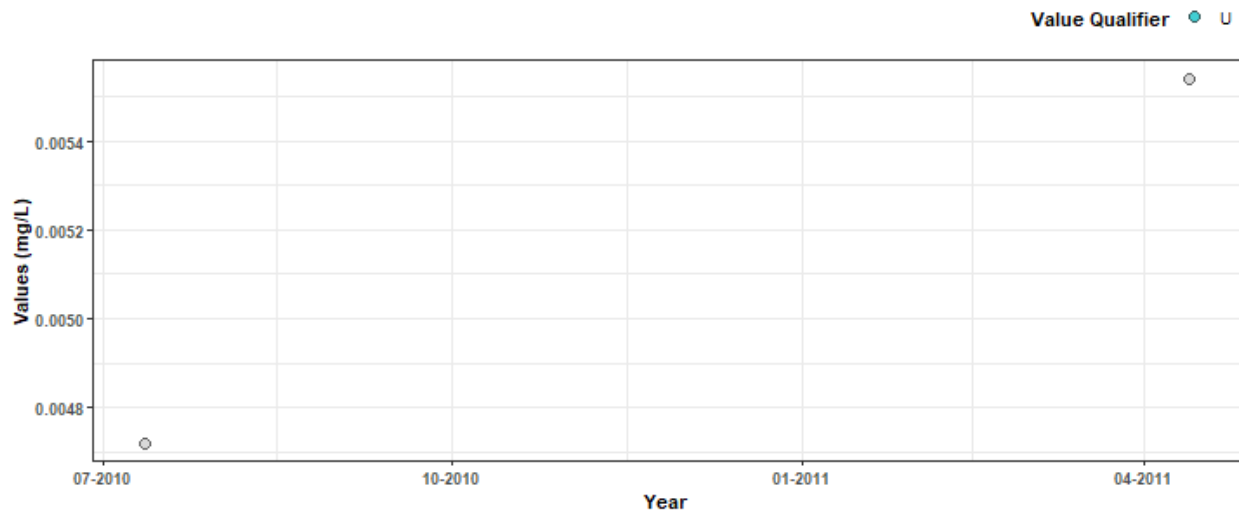
Scatter Plot of Excluded Managed Area
Cape Romano-Ten Thousand Islands Aquatic Preserve (1 Unique Years)



Autoscale



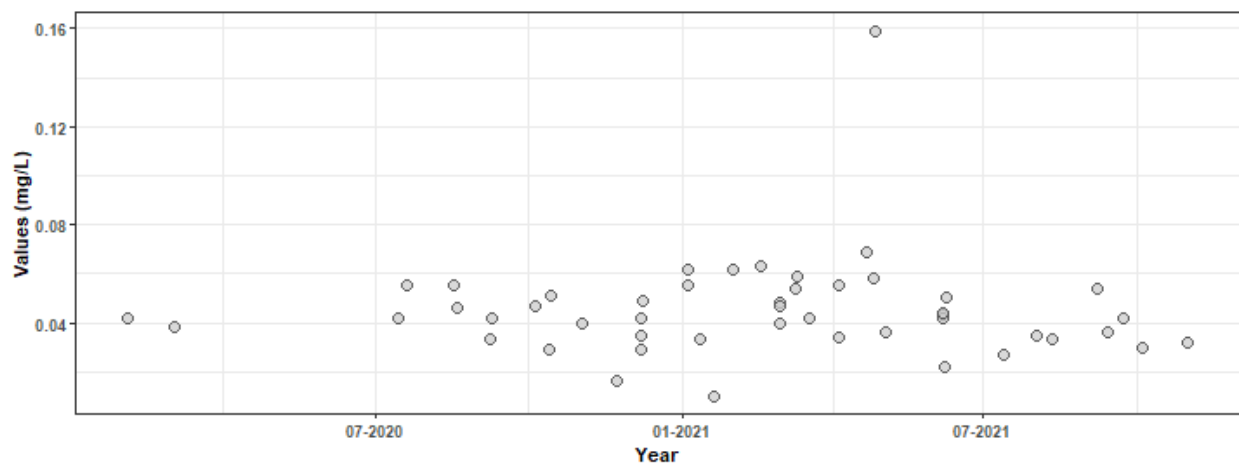
Autoscale



Scatter Plot of Excluded Managed Area
 Estero Bay Aquatic Preserve (2 Unique Years)

Autoscale

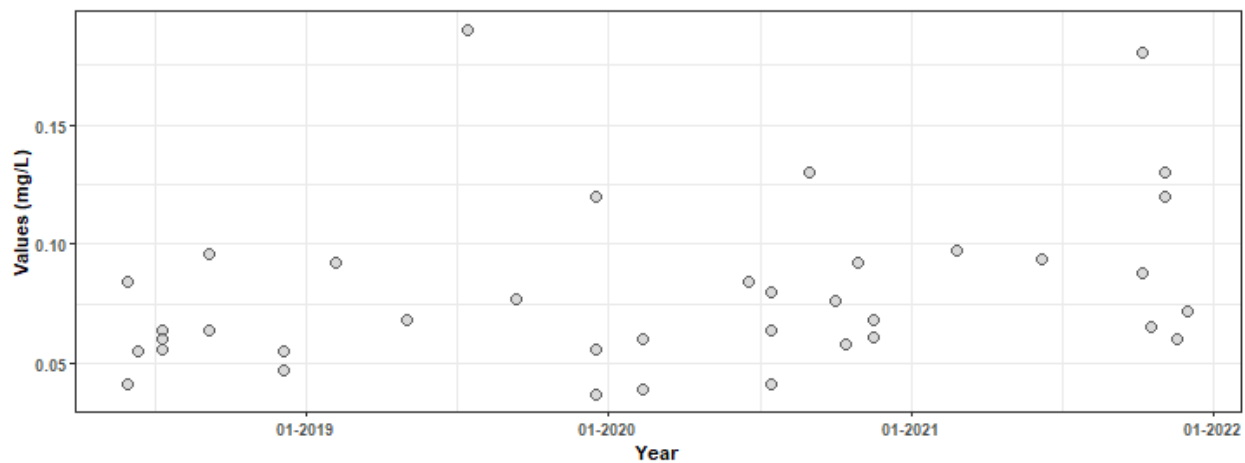
Value Qualifier ● U



Scatter Plot of Excluded Managed Area
 Guana River Marsh Aquatic Preserve (4 Unique Years)

Autoscale

Value Qualifier ● U



Autoscale

The scatter plot displays the concentration of nitrate (mg/L) in the water column over time from 2018 to 2022. The y-axis ranges from 0.05 to 0.15 mg/L. The x-axis shows years from 2018 to 2022. Data points are scattered, with notable peaks in late 2018 and late 2021.

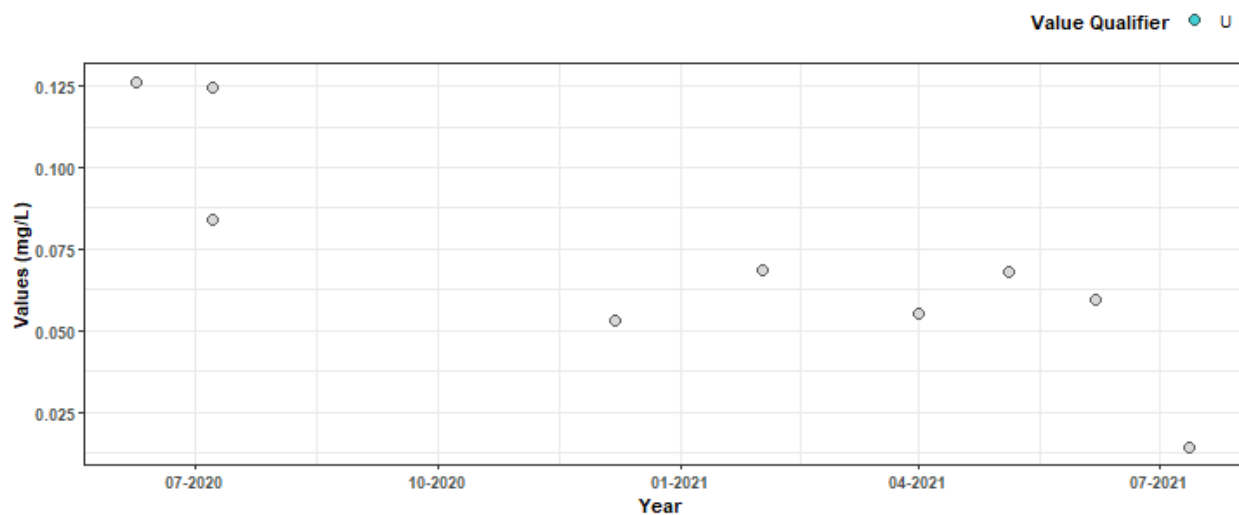
Autoscale

A scatter plot showing the relationship between 'Year' (X-axis) and 'Values (mg/L)' (Y-axis). The X-axis ranges from 2020 to 2021, with major ticks at 07-2020, 10-2020, 01-2021, 04-2021, and 07-2021. The Y-axis ranges from 0.0 to 0.4, with major ticks at 0.1, 0.2, 0.3, and 0.4. The plot displays several data points, most of which are clustered between 0.05 and 0.1 mg/L. There is a significant outlier at the end of the timeline (around 08-2021) with a value of approximately 0.45 mg/L.

Year	Values (mg/L)
~05-2020	0.09
~08-2020	0.06
~09-2020	0.08
~09-2020	0.07
~09-2020	0.10
~11-2020	0.05
~01-2021	0.06
~01-2021	0.04
~02-2021	0.08
~02-2021	0.05
~02-2021	0.07
~03-2021	0.09
~04-2021	0.10
~07-2021	0.11
~07-2021	0.10
~07-2021	0.11
~08-2021	0.45

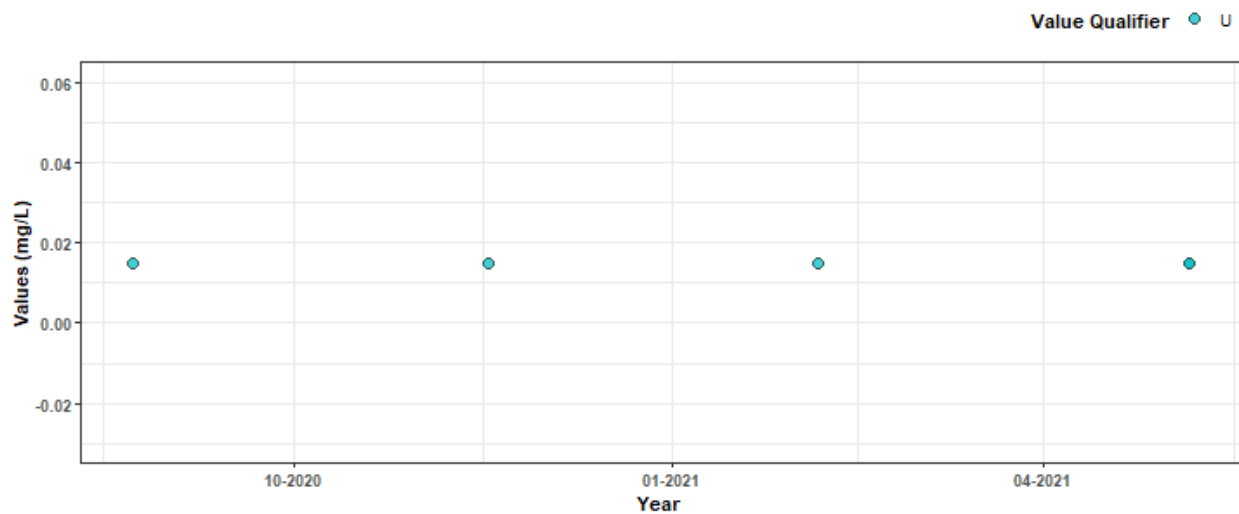
Scatter Plot of Excluded Managed Area
Indian River-Vero Beach to Ft. Pierce Aquatic Preserve (2 Unique Years)

Autoscale

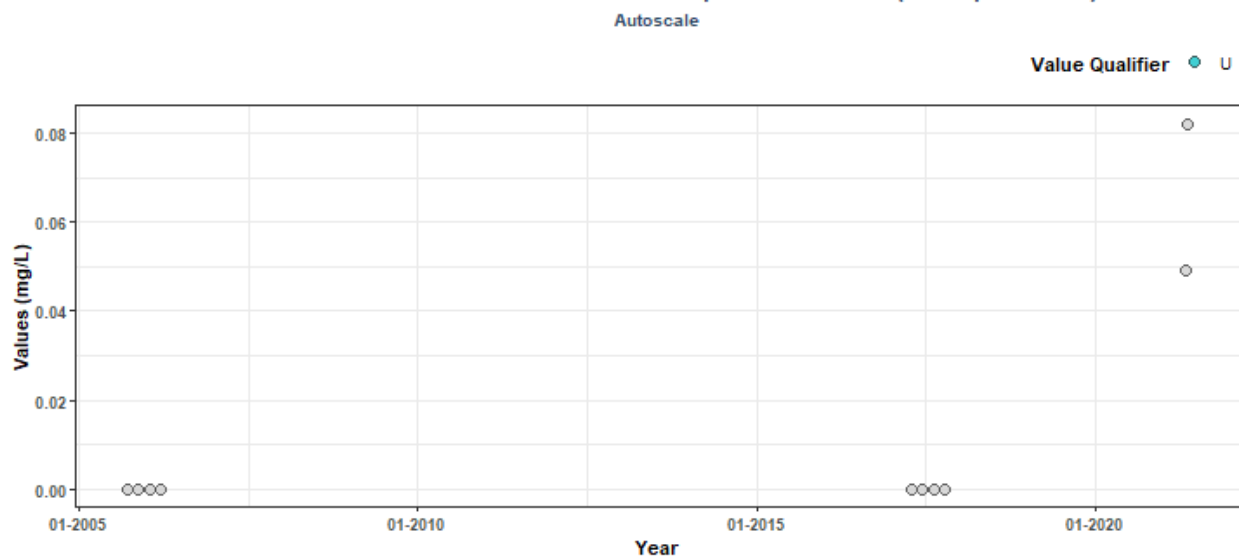


Scatter Plot of Excluded Managed Area
Lignumvitae Key Aquatic Preserve (2 Unique Years)

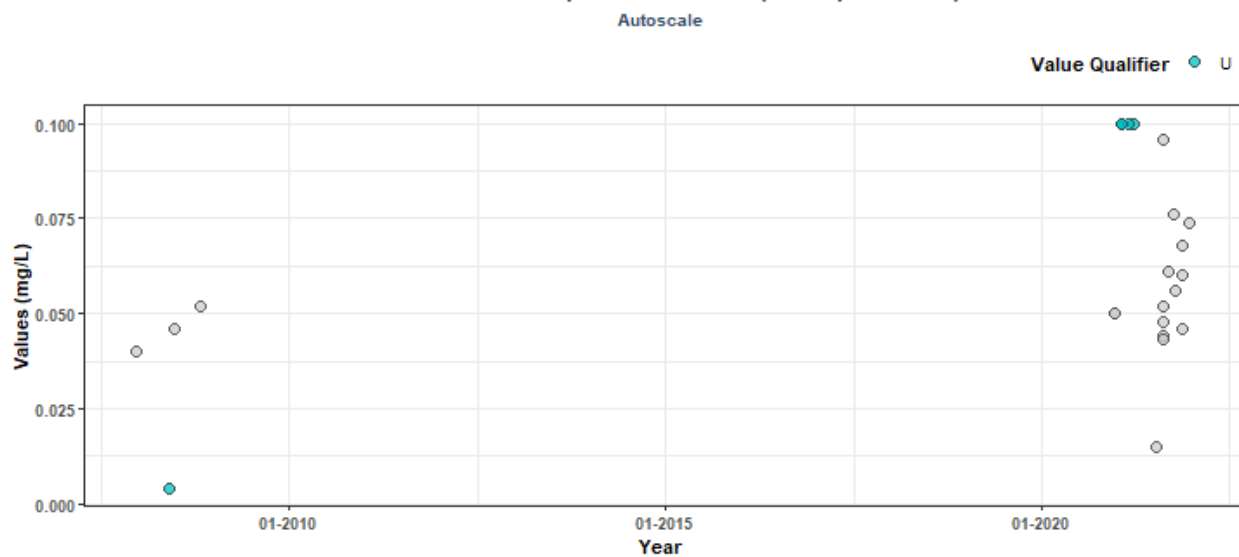
Autoscale



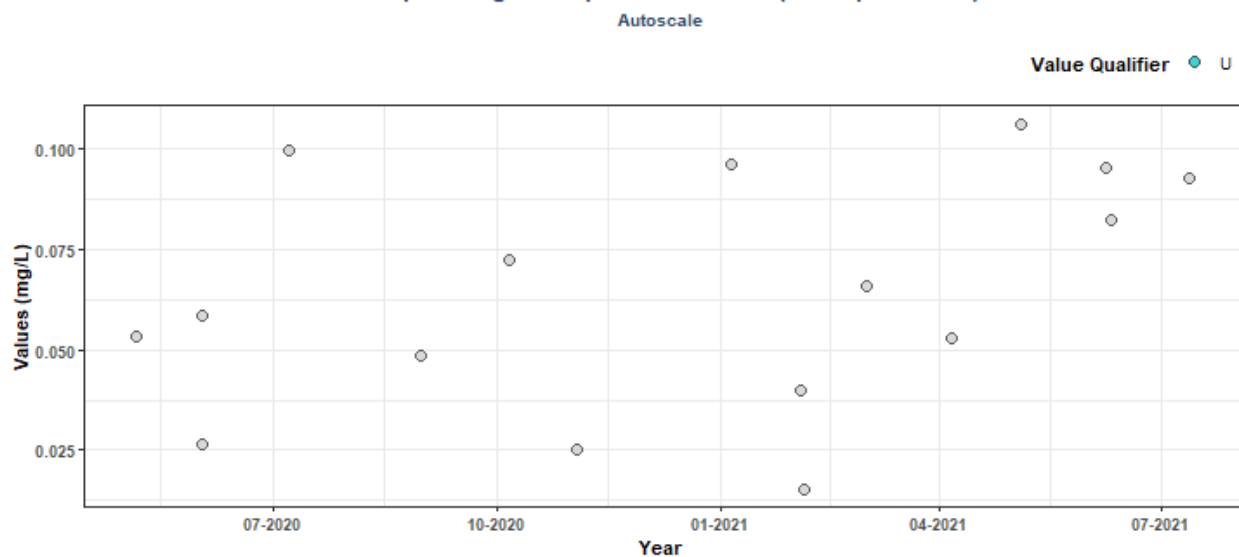
Scatter Plot of Excluded Managed Area
Loxahatchee River-Lake Worth Creek Aquatic Preserve (4 Unique Years)



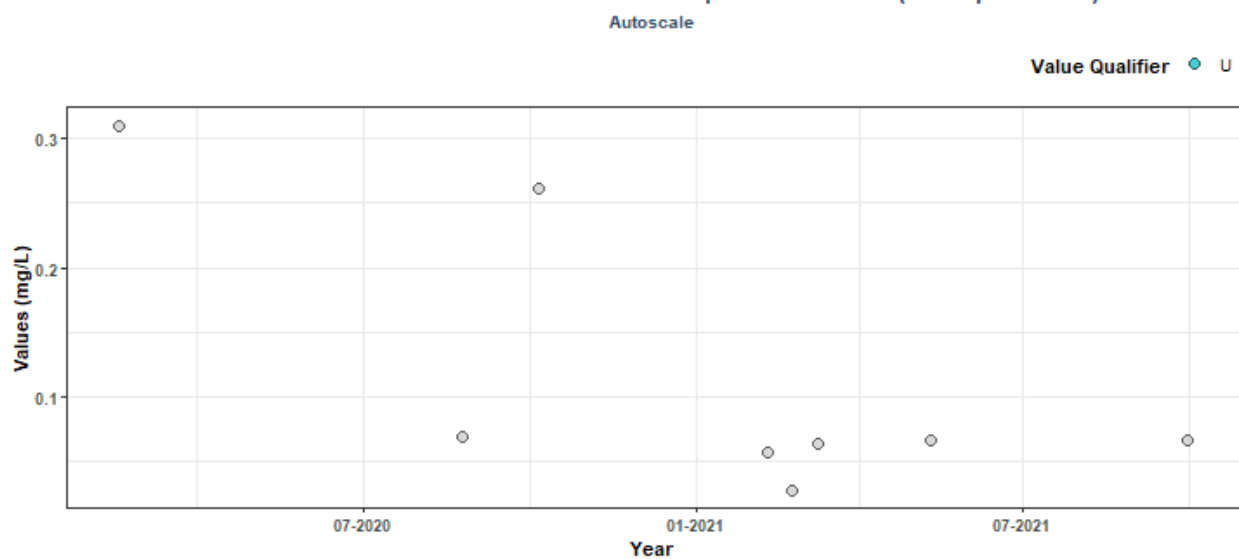
Scatter Plot of Excluded Managed Area
Matlacha Pass Aquatic Preserve (4 Unique Years)



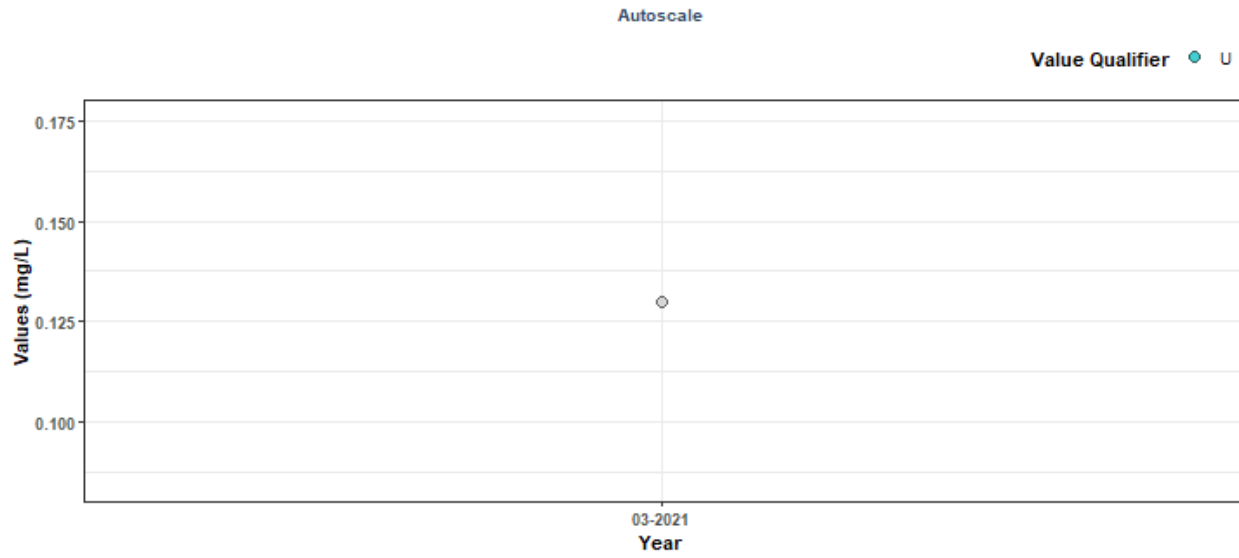
Scatter Plot of Excluded Managed Area
Mosquito Lagoon Aquatic Preserve (2 Unique Years)



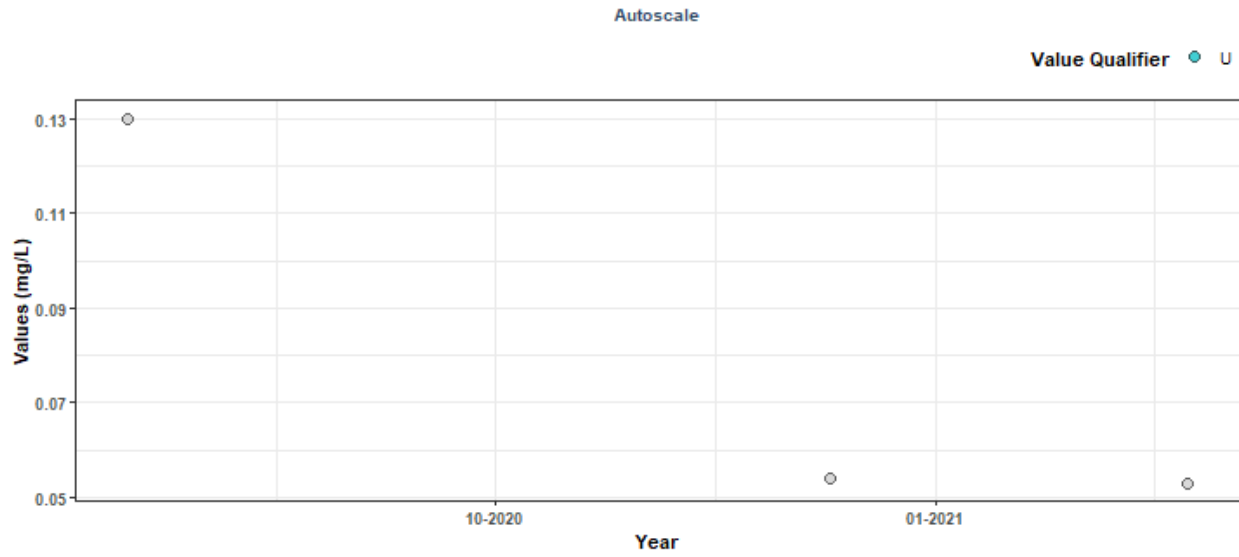
Scatter Plot of Excluded Managed Area
Nassau River-St. Johns River Marshes Aquatic Preserve (2 Unique Years)



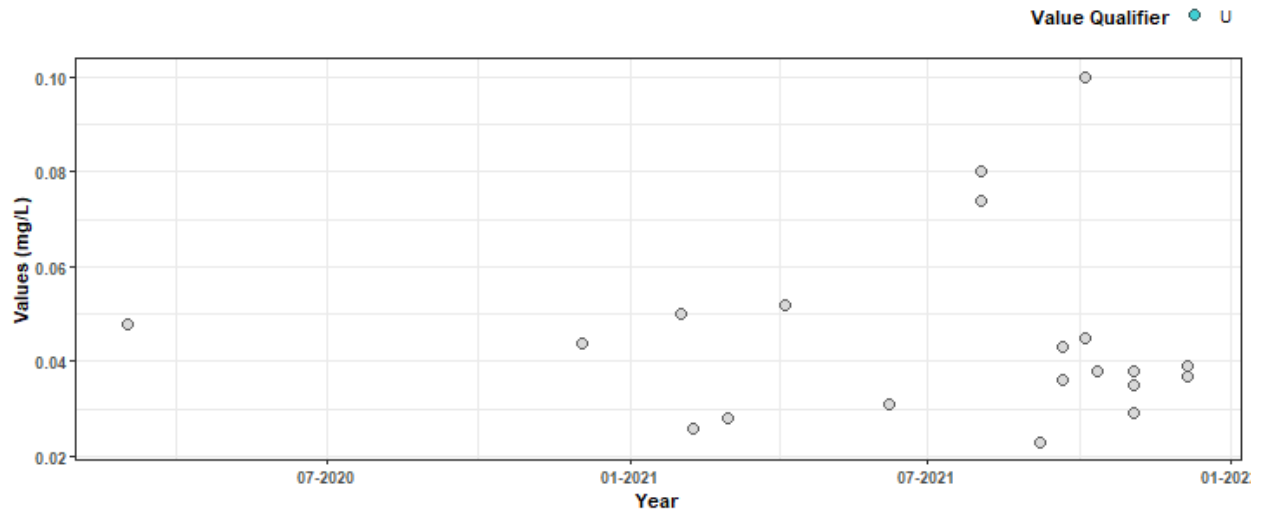
Scatter Plot of Excluded Managed Area
North Fork St. Lucie Aquatic Preserve (1 Unique Years)



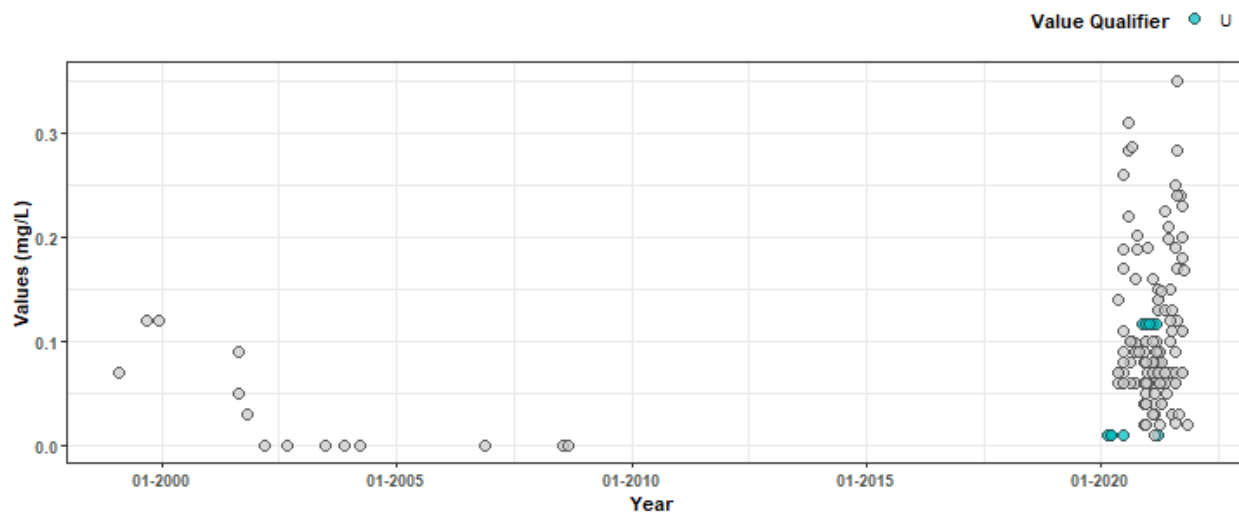
Scatter Plot of Excluded Managed Area
Pellicer Creek Aquatic Preserve (2 Unique Years)



Autoscale



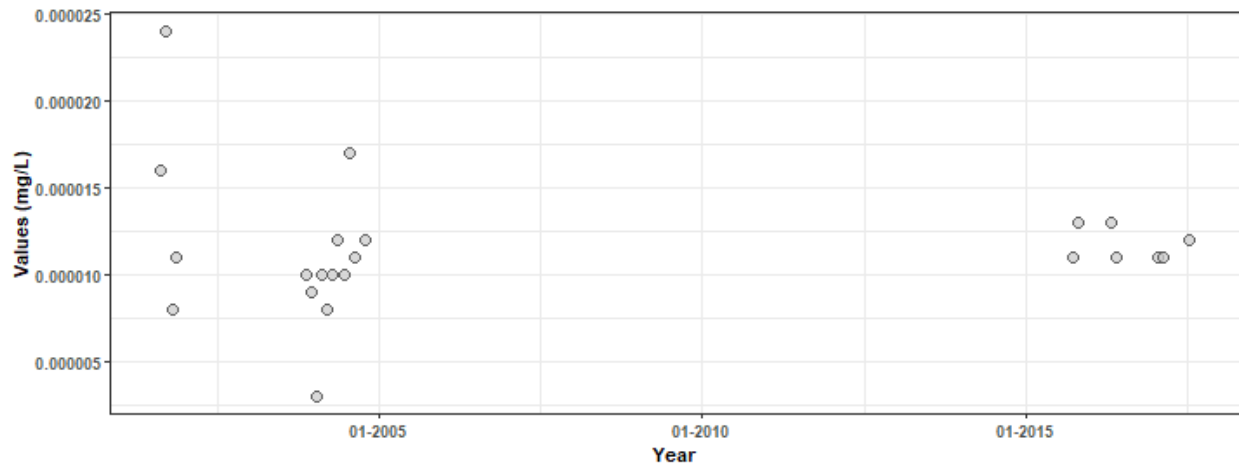
Autoscale



Scatter Plot of Excluded Managed Area Rocky Bayou State Park Aquatic Preserve (6 Unique Years)

Autoscale

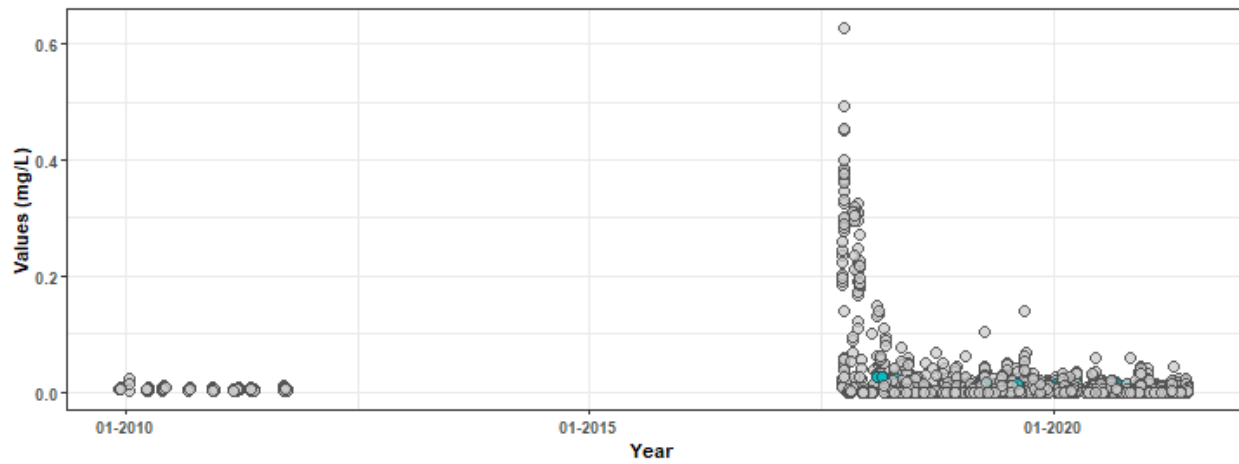
Value Qualifier ● U



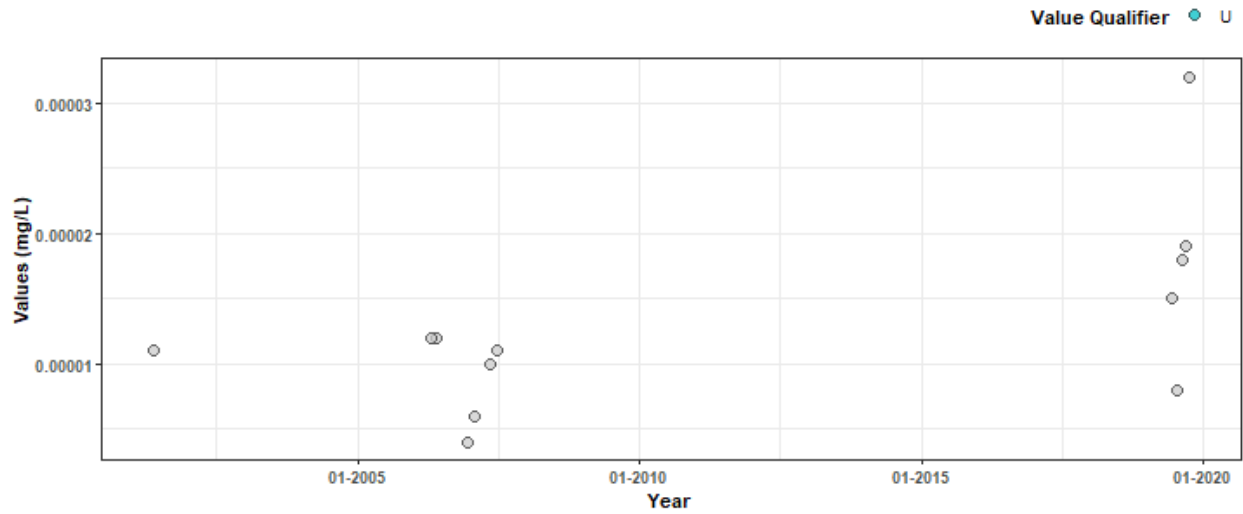
Scatter Plot of Excluded Managed Area Southeast Florida Coral Reef Ecosystem Conservation Area (8 Unique Years)

Autoscale

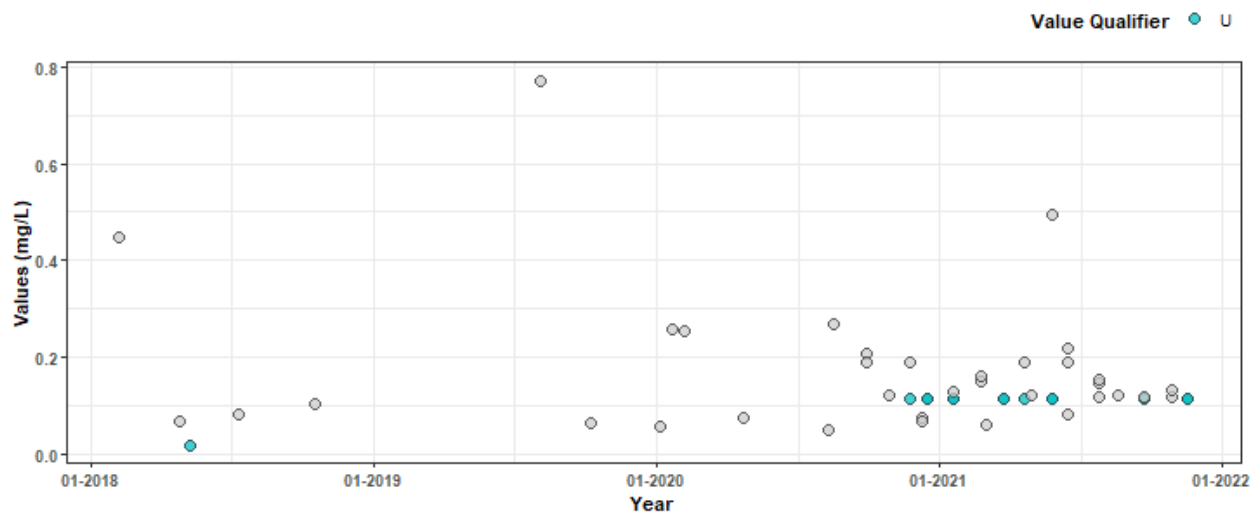
Value Qualifier ● U

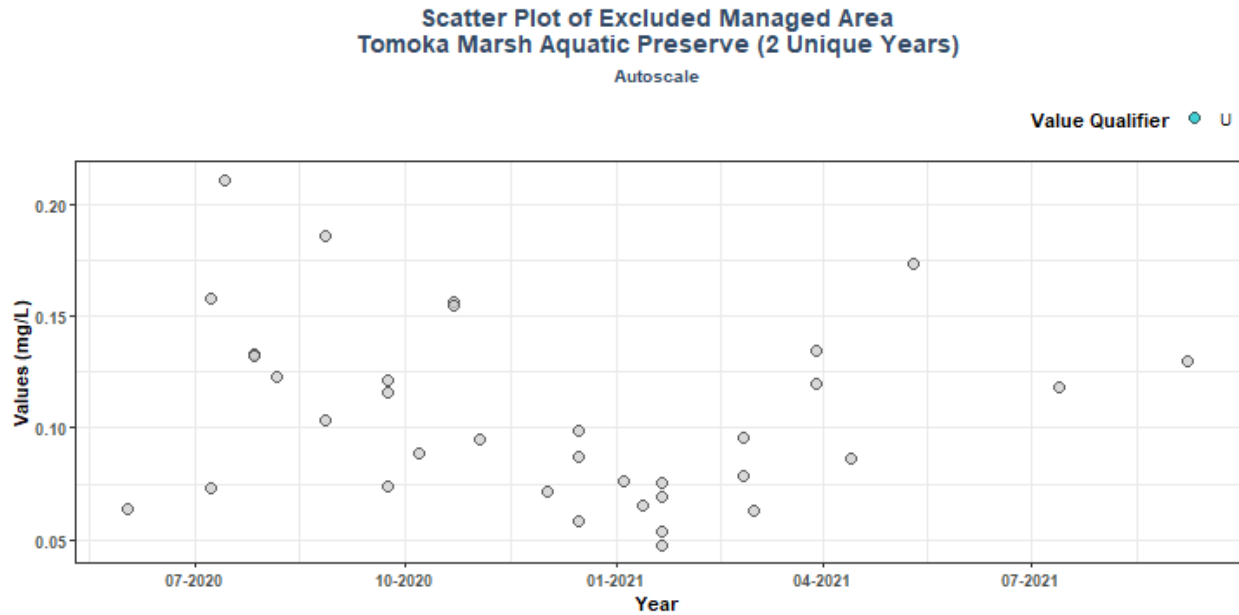


Autoscale



Autoscale





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 **SufficientData** 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 y 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$SufficientData==TRUE &
                      data$ManagedAreaName==MA_Include[i],]
    year_lower <- min(plot_data$Year)
    year_upper <- max(plot_data$Year)
    min_RV <- min(plot_data$ResultValue)
    mn_RV <- mean(plot_data$ResultValue[plot_data$ResultValue <
                                         quantile(data$ResultValue, 0.98)])
```

```

sd_RV <- sd(plot_data$ResultValue[plot_data$ResultValue <
                                quantile(data$ResultValue, 0.98)])
x_scale <- ifelse(year_upper - year_lower > 30, 10, 5)
y_scale <- mn_RV + 4 * sd_RV

tau <- KT.Stats$tau[KT.Stats$ManagedAreaName==MA_Include[i]]
s_slope <- KT.Stats$SennSlope[KT.Stats$ManagedAreaName==MA_Include[i]]
s_int <- KT.Stats$SennIntercept[KT.Stats$ManagedAreaName==MA_Include[i]]
trend <- KT.Stats$Trend[KT.Stats$ManagedAreaName==MA_Include[i]]
p <- KT.Stats$p[KT.Stats$ManagedAreaName==MA_Include[i]]

model <- lm(ResultValue ~ DecDate,
             data=plot_data)
m_int <- coef(model)[[1]]
m_slope <- coef(model)[[2]]
rm(model)
p1 <- ggplot(data=plot_data,
             aes(x=DecDate, y=ResultValue, fill=VQ_Plot)) +
  geom_point(shape=21, size=3, color="#333333", alpha=0.75) +
  geom_abline(aes(slope=s_slope, intercept=s_int),
             color="blue", size=1.2) +
  labs(subtitle="Autoscale",
       x="Year", y=paste0("Values (", unit, ")"),
       fill="Value Qualifier") +
  plot_theme +
  theme(legend.position="top", legend.box="horizontal",
       legend.justification="right") +
  {if(inc_H==TRUE){
    scale_fill_manual(values=c("H"= "#F8766D", "U"= "#00BFC4",
                              "HU"="#7CAE00"), na.value="#cccccc")
  } else if(param_name=="Secchi_Depth"){
    scale_fill_manual(values=c("S"= "#F8766D", "U"= "#00BFC4",
                              "SU"="#7CAE00"), na.value="#cccccc")
  } else {
    scale_fill_manual(values=c("U"= "#00BFC4"), na.value="#cccccc")
  }}

p2 <- ggplot(data=plot_data,
             aes(x=DecDate, y=ResultValue, fill=VQ_Plot)) +
  geom_point(shape=21, size=3, color="#333333", alpha=0.75) +
  geom_abline(aes(slope=s_slope, intercept=s_int),
             color="blue", size=1.2) +
  ylim(min_RV, y_scale) +
  labs(subtitle="Scaled to 4x Standard Deviation",
       x="Year", y=paste0("Values (", unit, ")")) +
  plot_theme +
  theme(legend.position="none") +
  {if(inc_H==TRUE){
    scale_fill_manual(values=c("H"= "#F8766D", "U"= "#00BFC4",
                              "HU"="#7CAE00"), na.value="#cccccc")
  } else if(param_name=="Secchi_Depth"){
    scale_fill_manual(values=c("S"= "#F8766D", "U"= "#00BFC4",
                              "SU"="#7CAE00"), na.value="#cccccc")
  }}

```

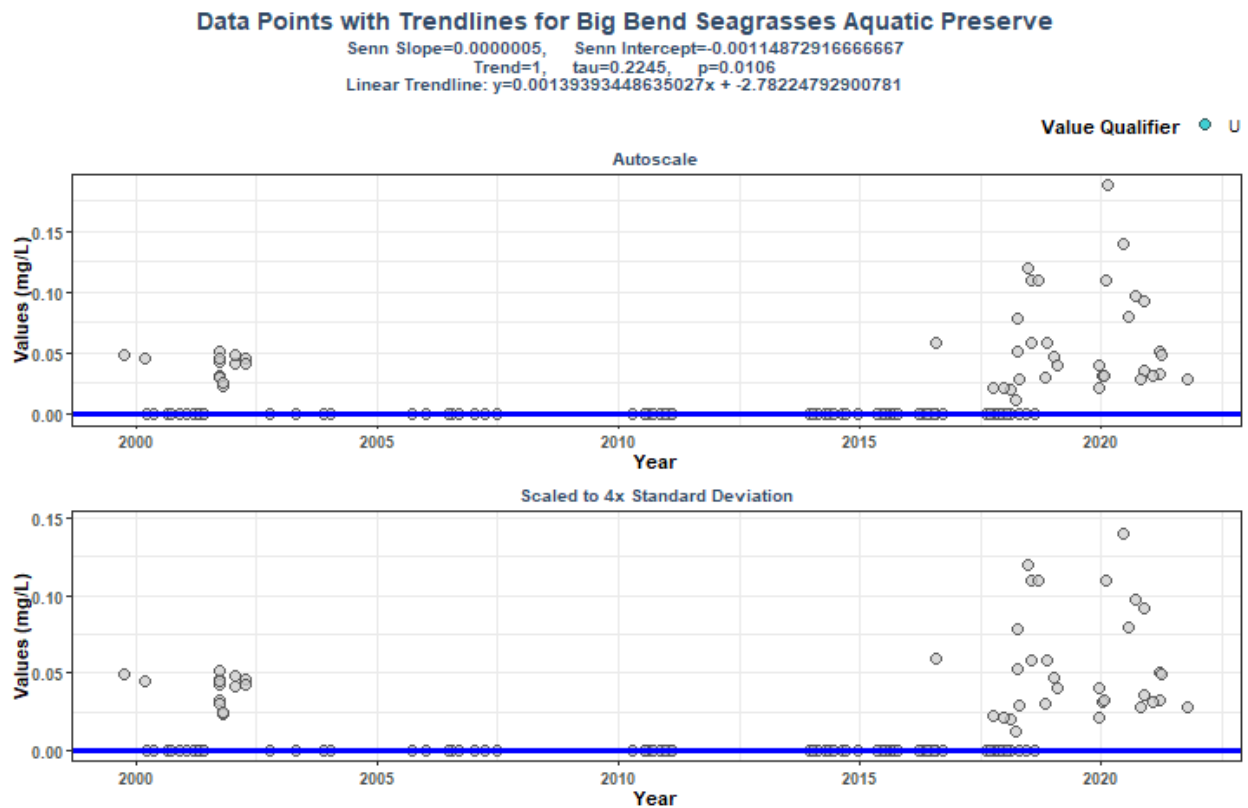
```

    } else {
      scale_fill_manual(values=c("U"= "#00BFC4"), na.value="#cccccc")
    }
  }
  leg <- get_legend(p1)
  KTset <- ggarrange(leg, p1 + theme(legend.position="none"), p2,
                    ncol=1, heights=c(0.1, 1, 1))

  p0 <- ggplot() + labs(title=paste0("Data Points with Trendlines for ",
                                     MA_Include[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)) +
    plot_theme + theme(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)))
  rm(plot_data)
  rm(KTset, leg)
}
}

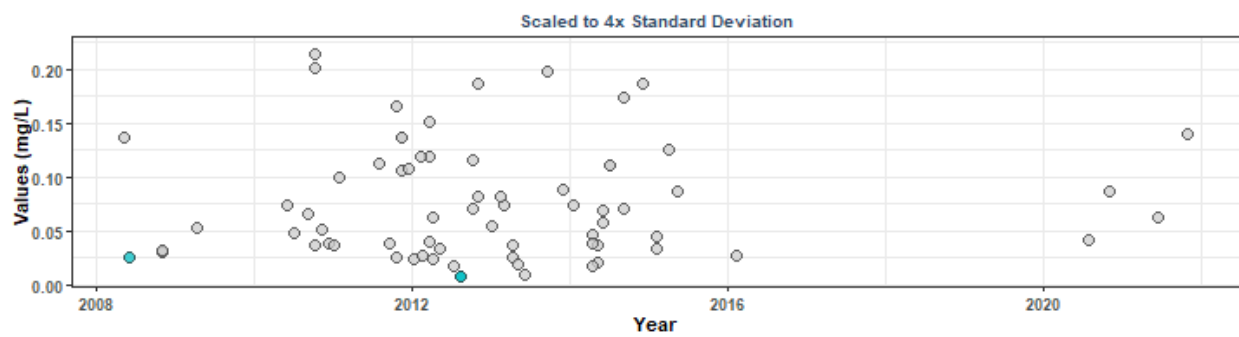
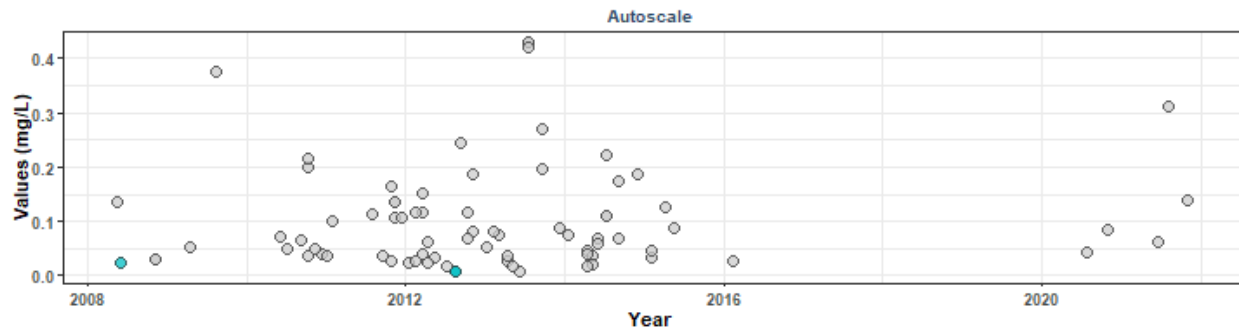
```



Data Points with Trendlines for Cape Haze Aquatic Preserve

Senn Slope=0.002, Senn Intercept=1.34742445054945
 Trend=0, tau=0.0605, p=0.5032
 Linear Trendline: $y=0.00323645831208939x - 6.41303450354353$

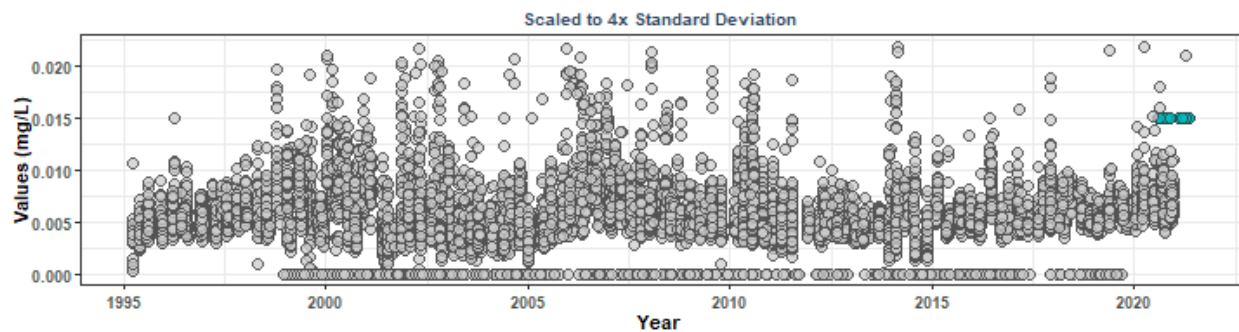
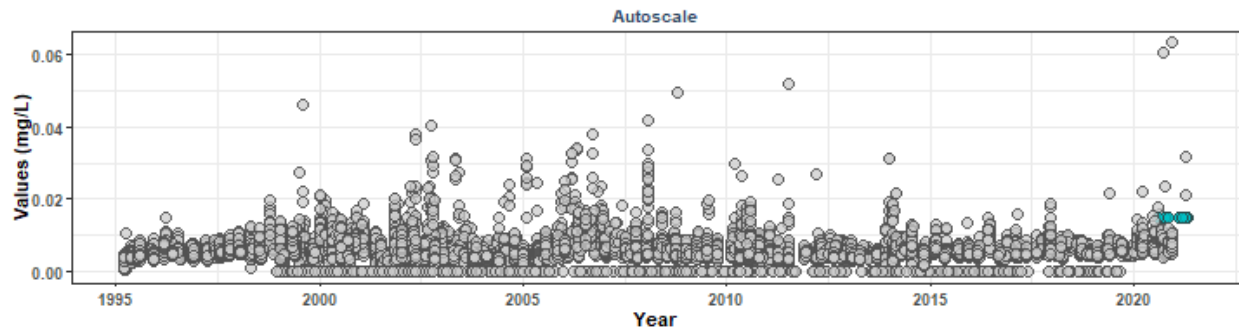
Value Qualifier ● U



Data Points with Trendlines for Florida Keys National Marine Sanctuary

Senn Slope=0.000015485, Senn Intercept=-0.0669200684375
 Trend=1, tau=0.0371, p=0
 Linear Trendline: $y=0.0000277551822720919x - 0.04965640558818$

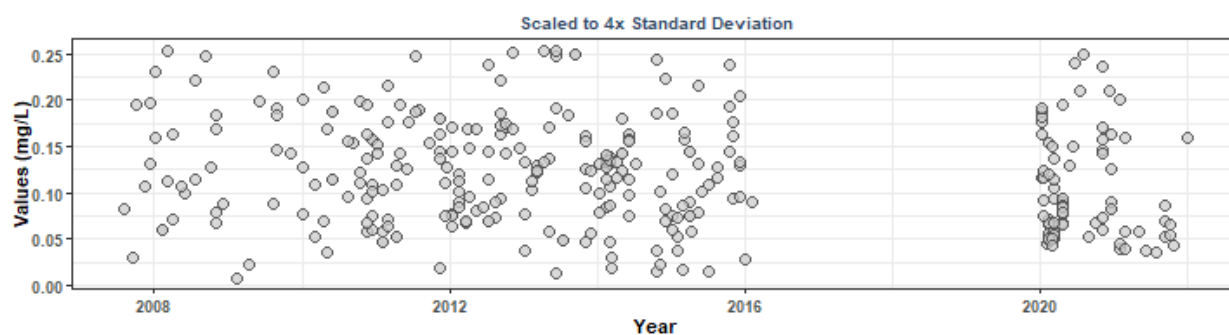
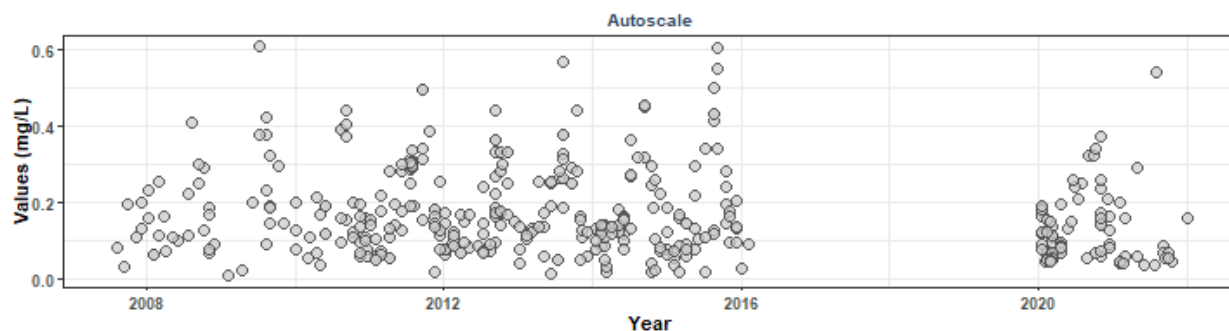
Value Qualifier ● U



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

Senn Slope=-0.00099999999999995, Senn Intercept=1.3104
Trend=0, tau=0.0376, p=0.339
Linear Trendline: $y = -0.00474041316623702x + 9.71431239619615$

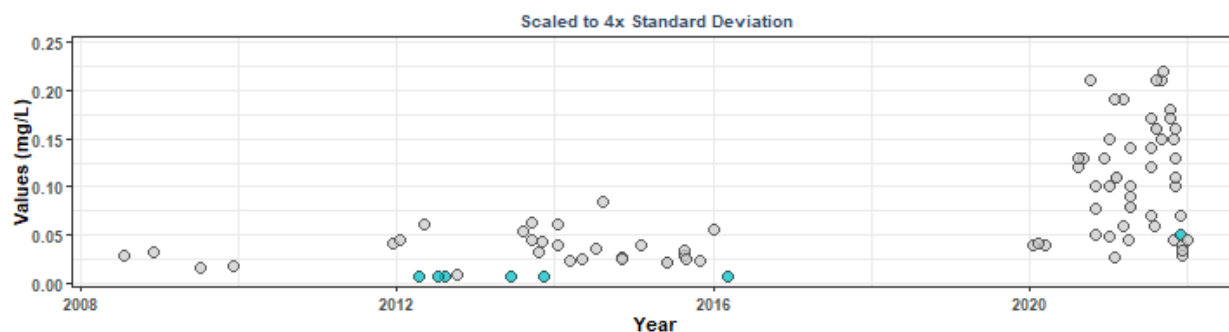
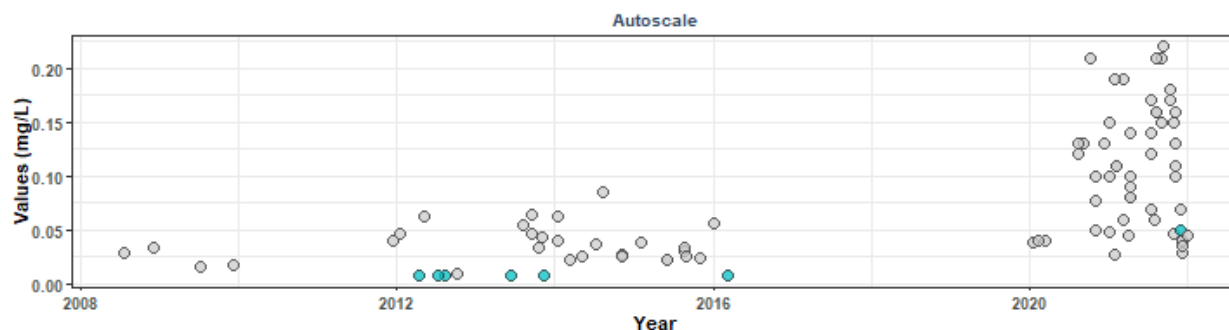
Value Qualifier ● U



Data Points with Trendlines for Lemon Bay Aquatic Preserve

Senn Slope=0.0101818181818182, Senn Intercept=-19.1350865384615
Trend=2, tau=0.4222, p=0
Linear Trendline: $y = 0.00918164665575325x - 18.4523437956218$

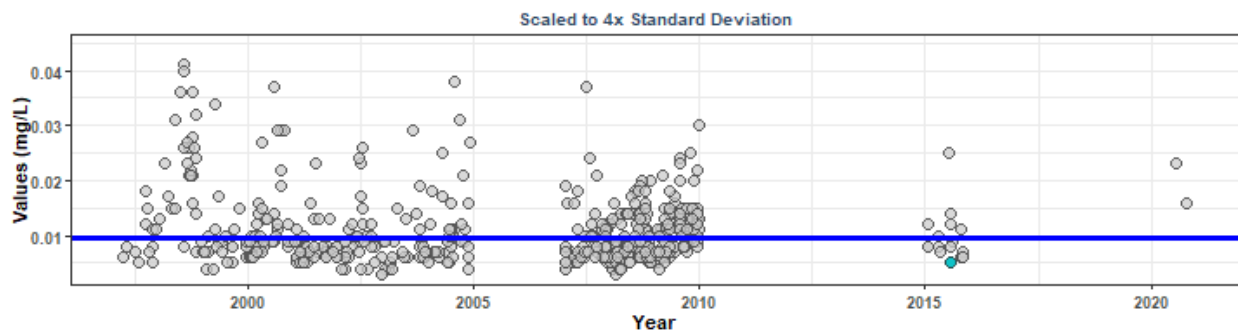
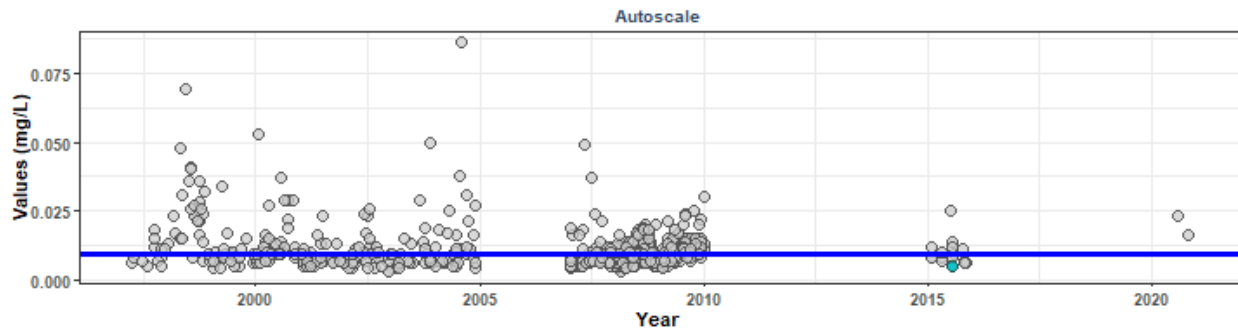
Value Qualifier ● U



Data Points with Trendlines for Nature Coast Aquatic Preserve

Senn Slope=0, Senn Intercept=0.0095
Trend=0, tau=-0.0363, p=0.2004
Linear Trendline: $y = -0.000284661597099223x + 0.582194473651978$

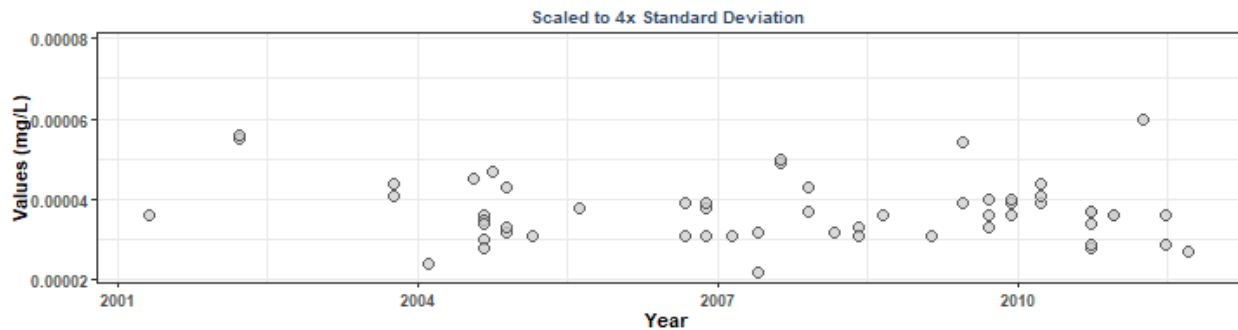
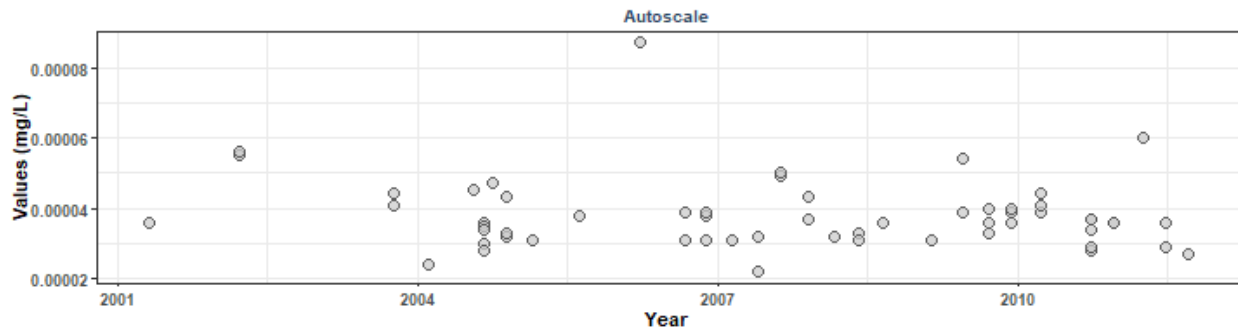
Value Qualifier ● U



Data Points with Trendlines for Rookery Bay Aquatic Preserve

Senn Slope=-0.00000005, Senn Intercept=0.000881500000000002
Trend=0, tau=-0.0237, p=0.6511
Linear Trendline: $y = -0.000000509231024177618x + 0.00106044082397307$

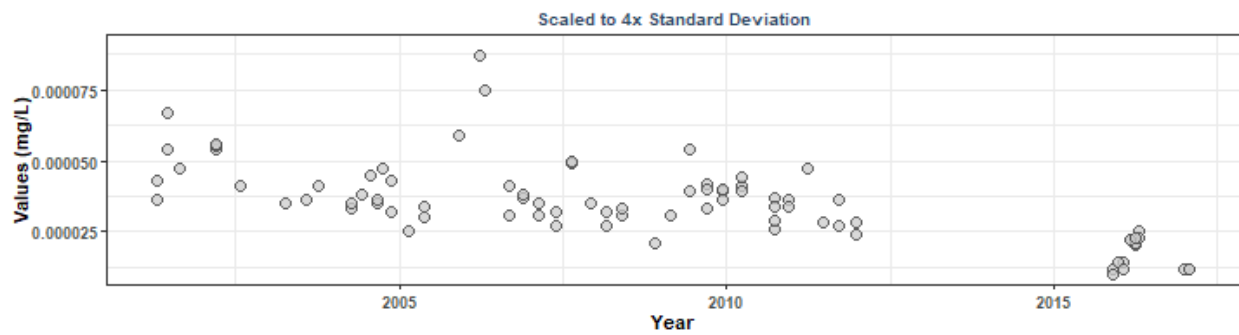
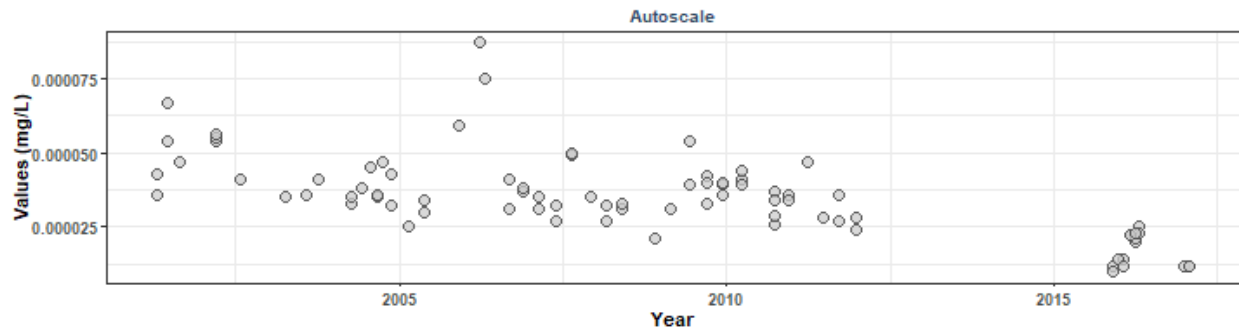
Value Qualifier ● U



Data Points with Trendlines for Rookery Bay National Estuarine Research Reserve

Senn Slope=-0.000002, Senn Intercept=0.00203699999999999
Trend=1, tau=0.4117, p=0
Linear Trendline: $y = -0.00000202143387844598x + 0.00409585638757994$

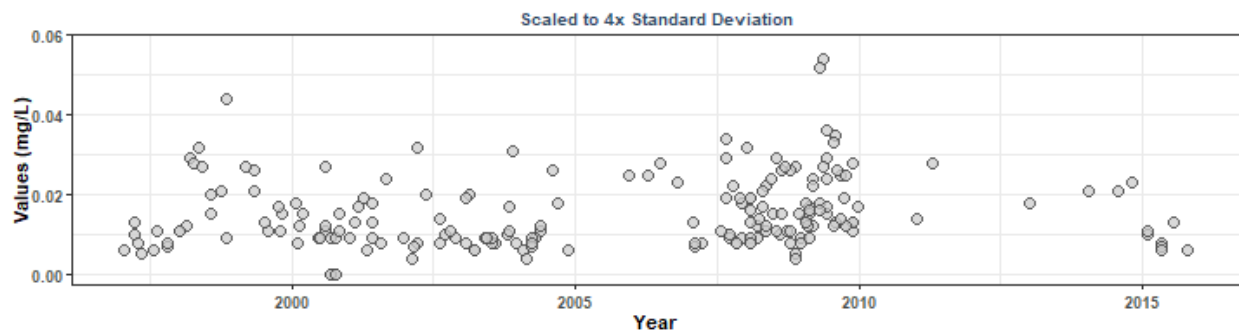
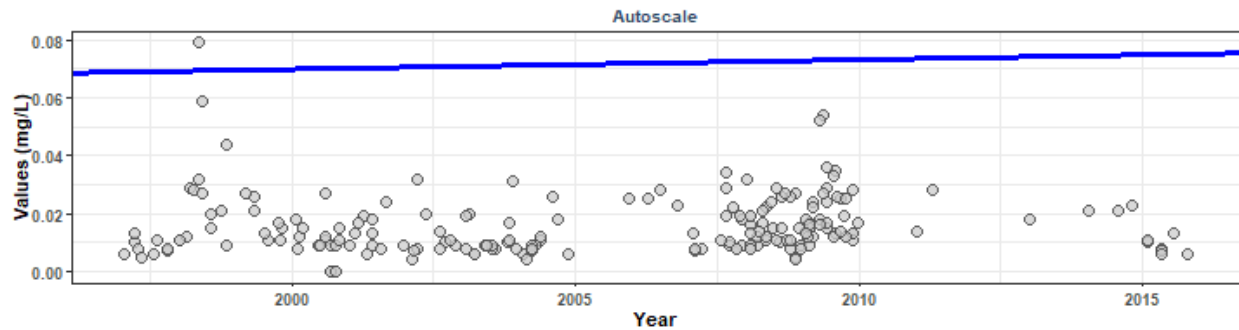
Value Qualifier ● U



Data Points with Trendlines for St. Martins Marsh Aquatic Preserve

Senn Slope=0.000333333333333333, Senn Intercept=-0.596583333333333
Trend=1, tau=0.1561, p=0.0028
Linear Trendline: $y = 0.000127601849164824x - 0.240062175932052$

Value Qualifier ● U



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 `SufficientData` 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 + 4 times the standard deviation
3. Y-axis set to be mean + 4 times 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) {
    plot_data <- data[data$SufficientData==TRUE &
                      data$ManagedAreaName==MA_Include[i],]
    year_lower <- min(plot_data$Year)
    year_upper <- max(plot_data$Year)
    min_RV <- min(plot_data$ResultValue)
    mn_RV <- mean(plot_data$ResultValue[plot_data$ResultValue <
                                         quantile(data$ResultValue, 0.98)])
    sd_RV <- sd(plot_data$ResultValue[plot_data$ResultValue <
                                       quantile(data$ResultValue, 0.98)])
    x_scale <- ifelse(year_upper - year_lower > 30, 10, 5)
    y_scale <- mn_RV + 4 * sd_RV

    ##Year plots
    p1 <- ggplot(data=plot_data,
                 aes(x=Year, y=ResultValue, group=Year)) +
      geom_boxplot(color="#333333", fill="#cccccc", outlier.shape=21,
                  outlier.size=3, outlier.color="#333333",
                  outlier.fill="#cccccc", outlier.alpha=0.75) +
      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))) +
      plot_theme
```

```

p2 <- ggplot(data=plot_data,
             aes(x=Year, y=ResultValue, group=Year)) +
  geom_boxplot(color="#333333", fill="#cccccc", outlier.shape=21,
              outlier.size=3, outlier.color="#333333",
              outlier.fill="#cccccc", outlier.alpha=0.75) +
  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))) +
  plot_theme

p3 <- ggplot(data=plot_data[plot_data$Year >= year_upper - 10, ],
             aes(x=Year, y=ResultValue, group=Year)) +
  geom_boxplot(color="#333333", fill="#cccccc", outlier.shape=21,
              outlier.size=3, outlier.color="#333333",
              outlier.fill="#cccccc", outlier.alpha=0.75) +
  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))) +
  plot_theme

Yset <- ggarrange(p1, p2, p3, ncol=1)

p0 <- ggplot() + labs(title=paste0("Summary Box Plots for ",
                                   MA_Include[i]),
                    subtitle="By Year") +
  plot_theme + theme(panel.border=element_blank(),
                    panel.grid.major=element_blank(),
                    panel.grid.minor=element_blank(),
                    axis.line=element_blank())

## Year & Month Plots
p4 <- ggplot(data=plot_data,
             aes(x=YearMonthDec, y=ResultValue,
                 group=YearMonth, color=as.factor(Month))) +
  geom_boxplot(fill="#cccccc", outlier.size=1.5, outlier.alpha=0.75) +
  labs(subtitle="Autoscale",
       x="Year", y=paste0("Values (", unit, ")"), color="Month") +
  scale_x_continuous(limits=c(year_lower - 1, year_upper + 1),
                    breaks=rev(seq(year_upper,
                                   year_lower, -x_scale))) +
  plot_theme +
  theme(legend.position="none")

p5 <- ggplot(data=plot_data,
             aes(x=YearMonthDec, y=ResultValue,
                 group=YearMonth, color=as.factor(Month))) +
  geom_boxplot(fill="#cccccc", outlier.size=1.5, outlier.alpha=0.75) +

```

```

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))) +

plot_theme +
theme(legend.position="top", legend.box="horizontal") +
guides(color=guide_legend(nrow=1))

p6 <- ggplot(data=plot_data[plot_data$Year >= year_upper - 10, ],
             aes(x=YearMonthDec, y=ResultValue,
                 group=YearMonth, color=as.factor(Month))) +
geom_boxplot(fill="#cccccc", outlier.size=1.5, outlier.alpha=0.75) +
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))) +

plot_theme +
theme(legend.position="none")

leg1 <- get_legend(p5)
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_Include[i]),
                      subtitle="By Year & Month") + plot_theme +
theme(panel.border=element_blank(),
      panel.grid.major=element_blank(),
      panel.grid.minor=element_blank(), axis.line=element_blank())

## Month Plots
p7 <- ggplot(data=plot_data,
             aes(x=Month, y=ResultValue,
                 group=Month, fill=as.factor(Month))) +
geom_boxplot(color="#333333", outlier.shape=21, outlier.size=3,
             outlier.color="#333333", outlier.alpha=0.75) +
labs(subtitle="Autoscale",
      x="Month", y=paste0("Values (", unit, ")"), fill="Month") +
scale_x_continuous(limits=c(0, 13), breaks=seq(3, 12, 3)) +
plot_theme +
theme(legend.position="none")

p8 <- ggplot(data=plot_data,
             aes(x=Month, y=ResultValue,
                 group=Month, fill=as.factor(Month))) +
geom_boxplot(color="#333333", outlier.shape=21, outlier.size=3,
             outlier.color="#333333", outlier.alpha=0.75) +
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)) +
    plot_theme +
    theme(legend.position="top", legend.box="horizontal") +
    guides(fill=guide_legend(nrow=1))

p9 <- ggplot(data=plot_data[plot_data$Year >= year_upper - 10, ],
            aes(x=Month, y=ResultValue,
                group=Month, fill=as.factor(Month))) +
    geom_boxplot(color="#333333", outlier.shape=21, outlier.size=3,
                outlier.color="#333333", outlier.alpha=0.75) +
    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)) +
    plot_theme +
    theme(legend.position="none")

leg2 <- get_legend(p8)
Mset <- ggarrange(leg2, p7, p8 + theme(legend.position="none"), p9,
                 ncol=1, heights=c(0.1, 1, 1, 1))

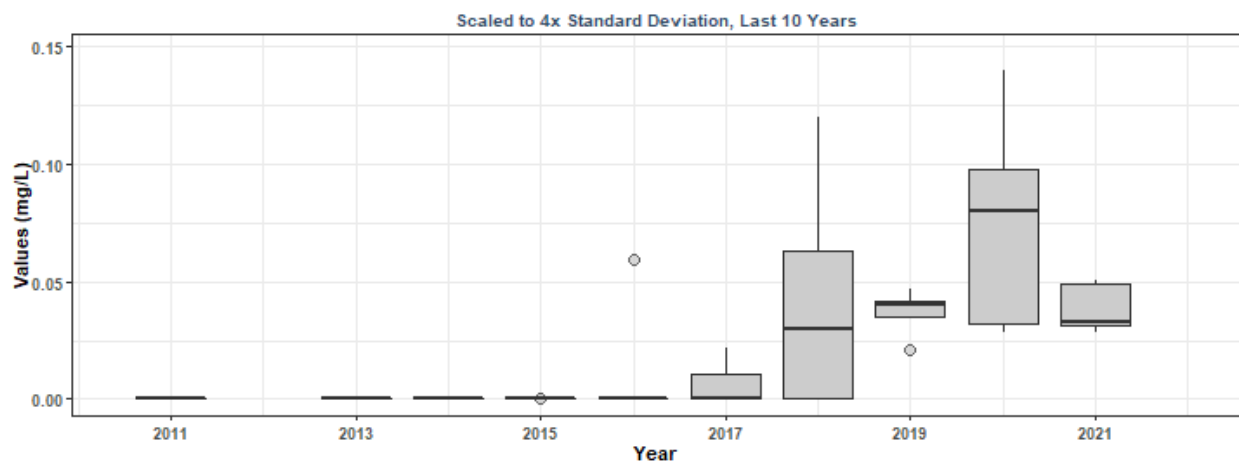
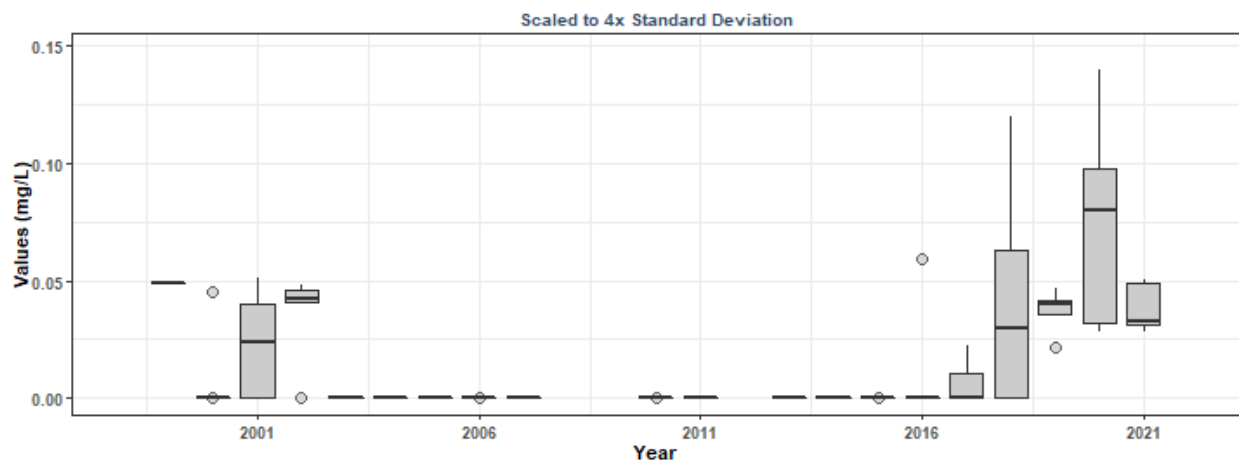
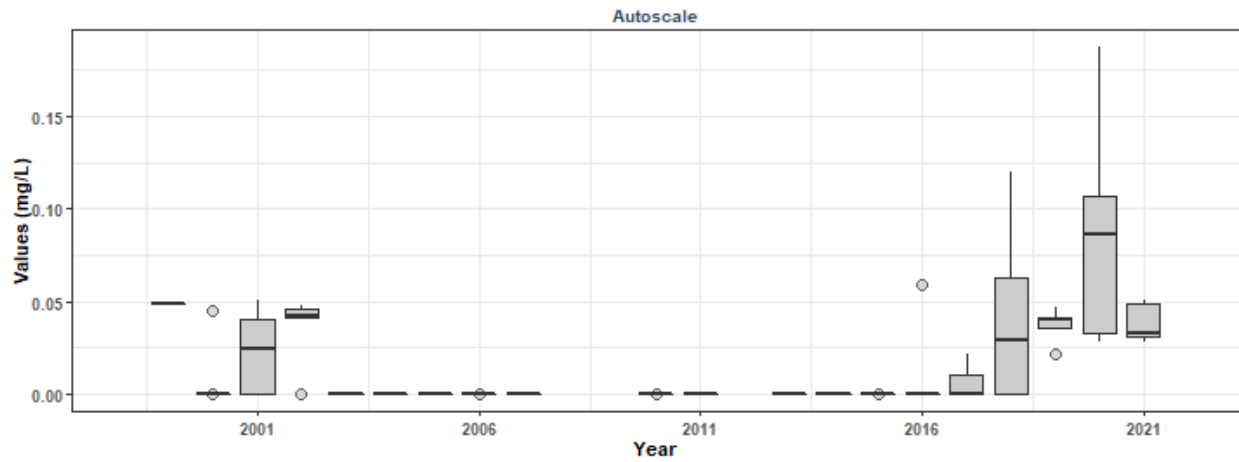
p000 <- ggplot() + labs(title=paste0("Summary Box Plots for ",
                                     MA_Include[i]),
                       subtitle="By Month") + plot_theme +
    theme(panel.border=element_blank(),
          panel.grid.major=element_blank(),
          panel.grid.minor=element_blank(), 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)))

rm(plot_data)
rm(p1, p2, p3, p4, p5, p6, p7, p8, p9, p0, p00, p000, leg1, leg2,
    Yset, YMset, Mset)
}
}

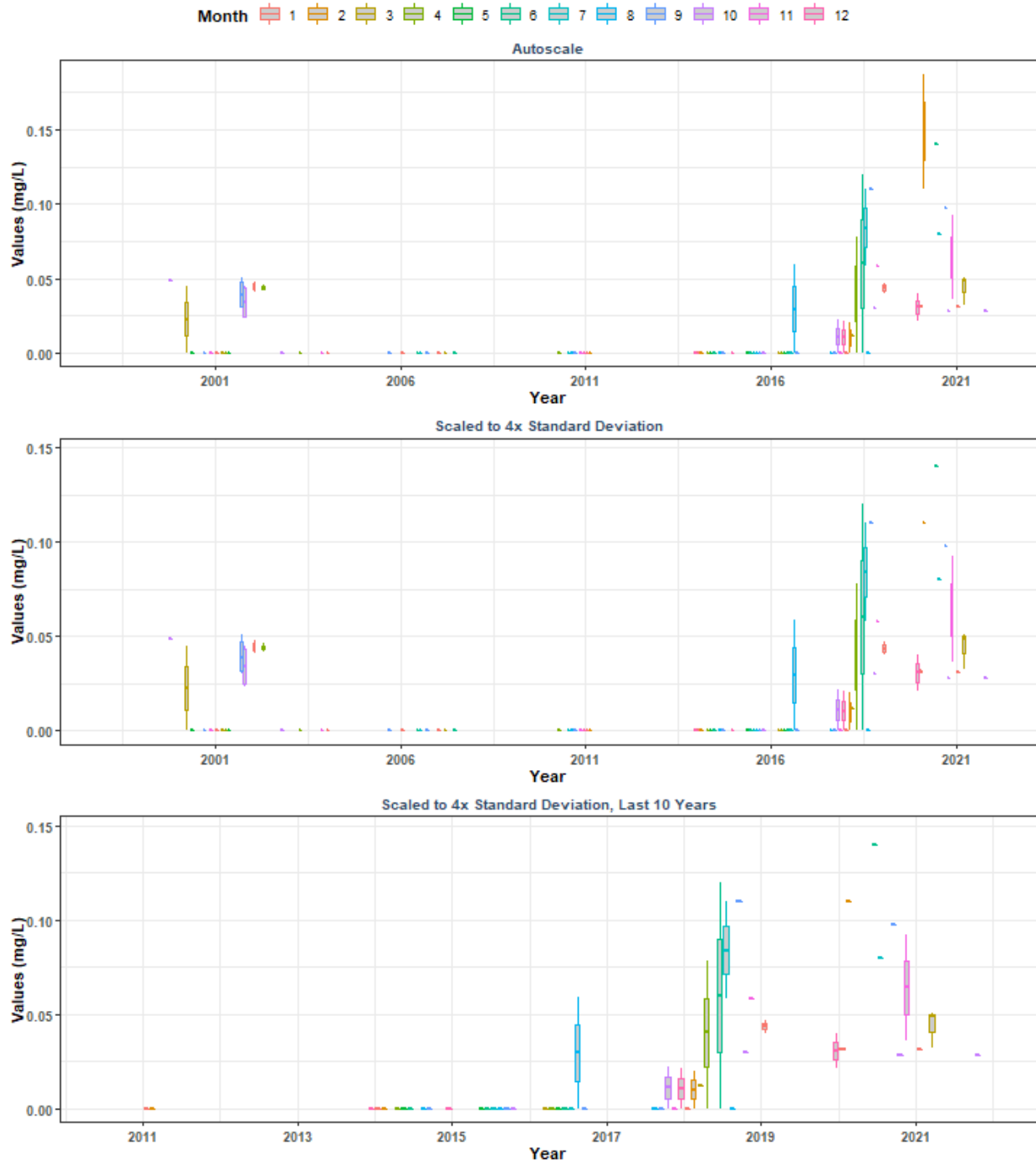
```

Summary Box Plots for Big Bend Seagrasses Aquatic Preserve By Year



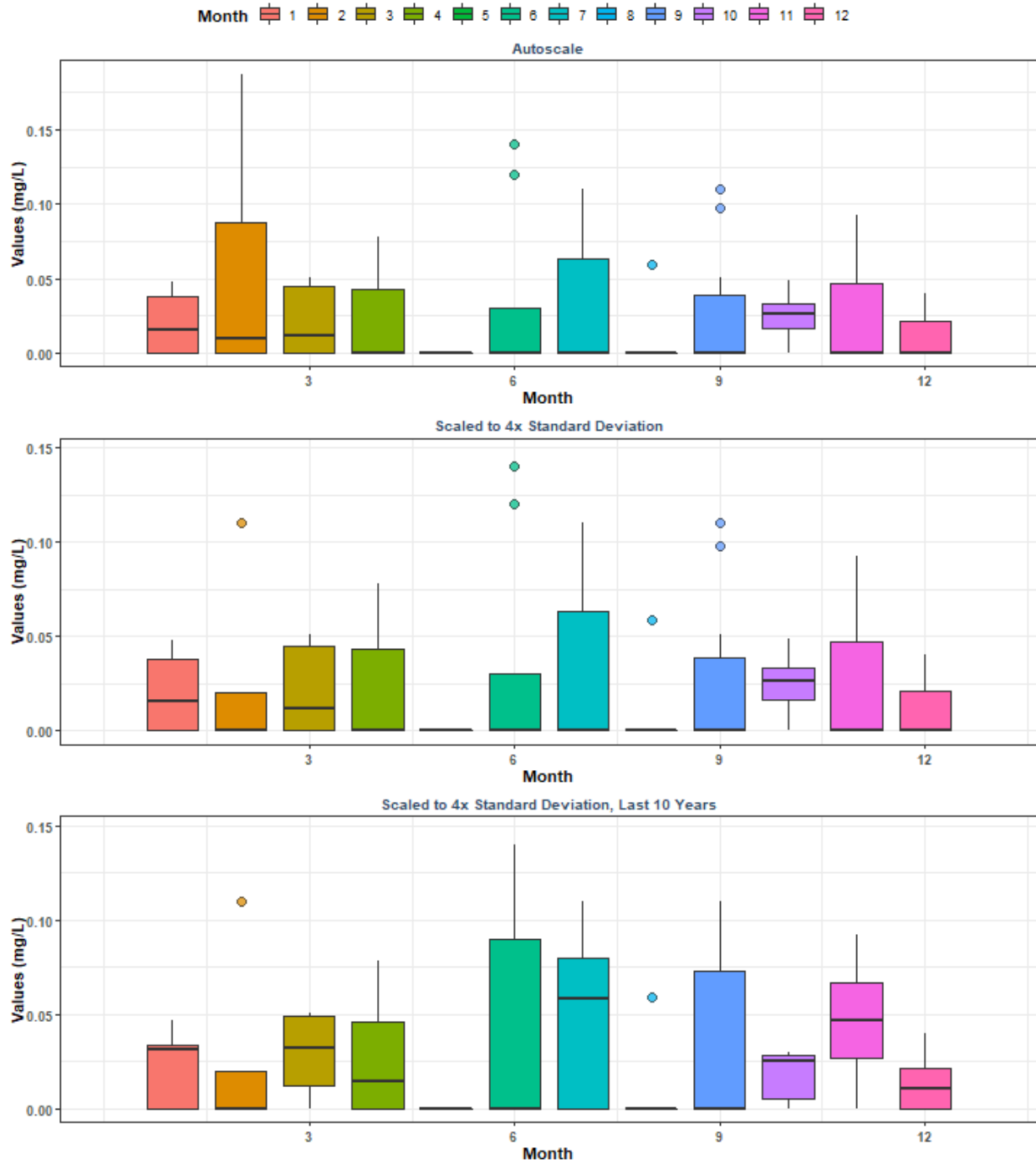
Summary Box Plots for Big Bend Seagrasses Aquatic Preserve

By Year & Month



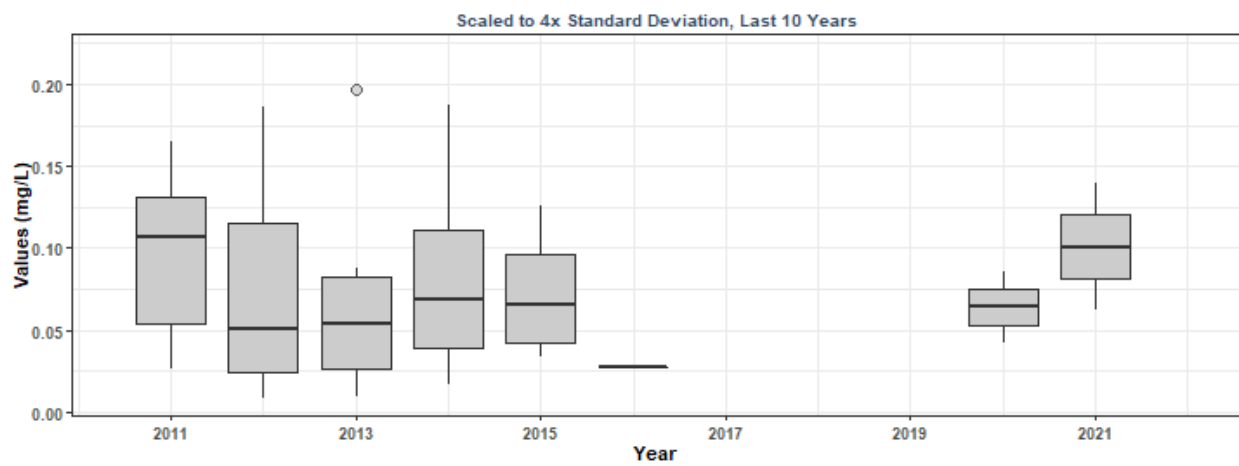
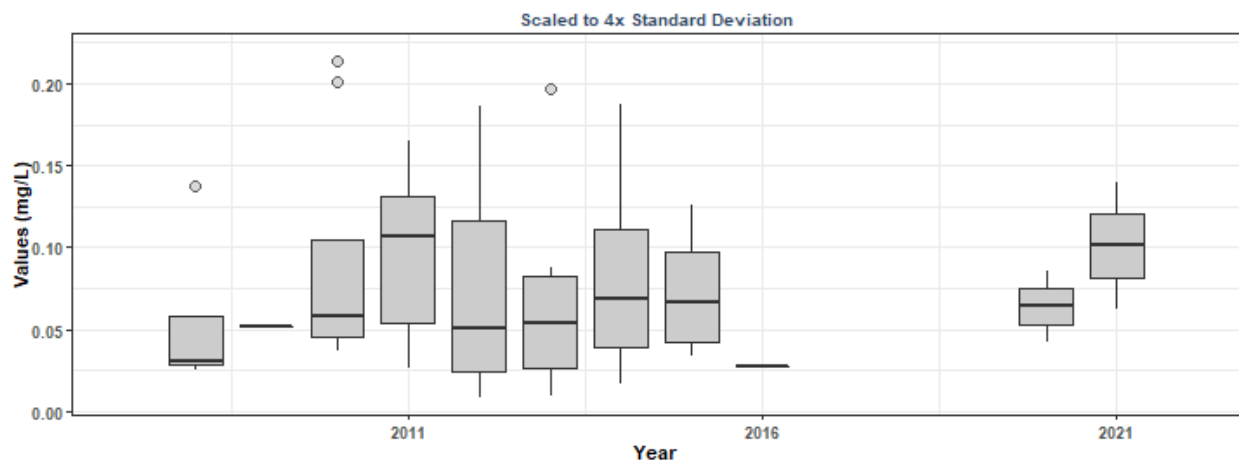
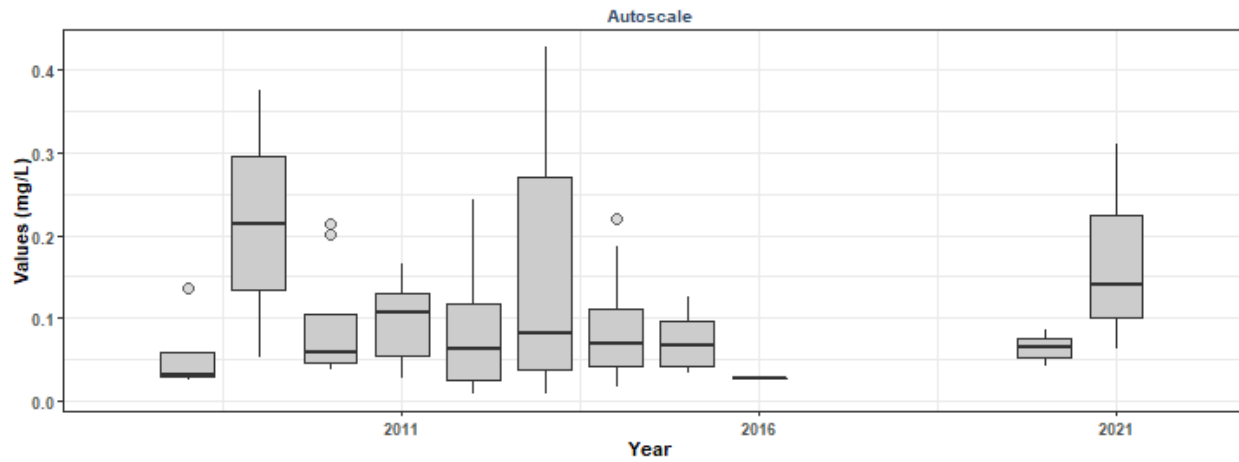
Summary Box Plots for Big Bend Seagrasses Aquatic Preserve

By Month



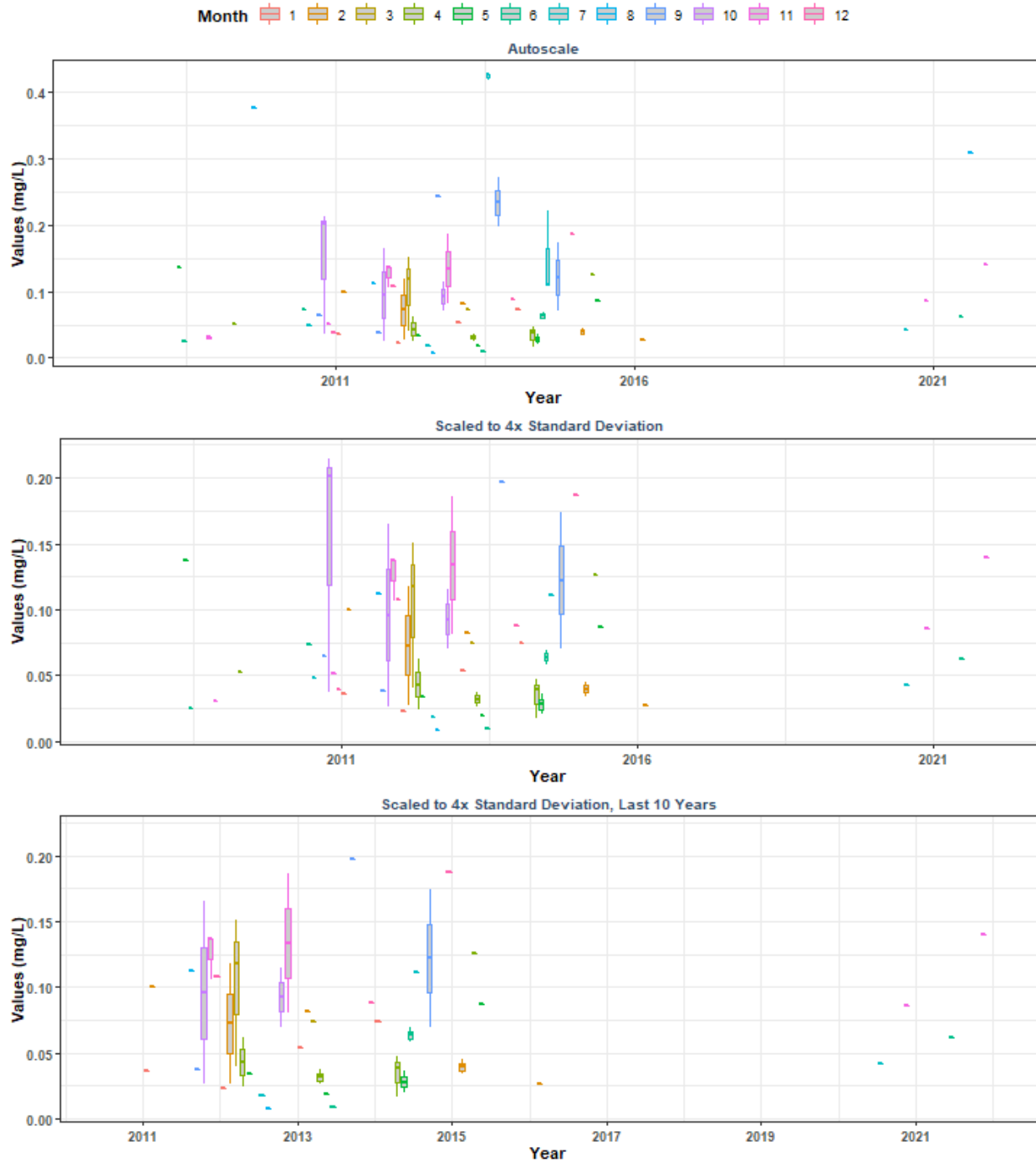
Summary Box Plots for Cape Haze Aquatic Preserve

By Year



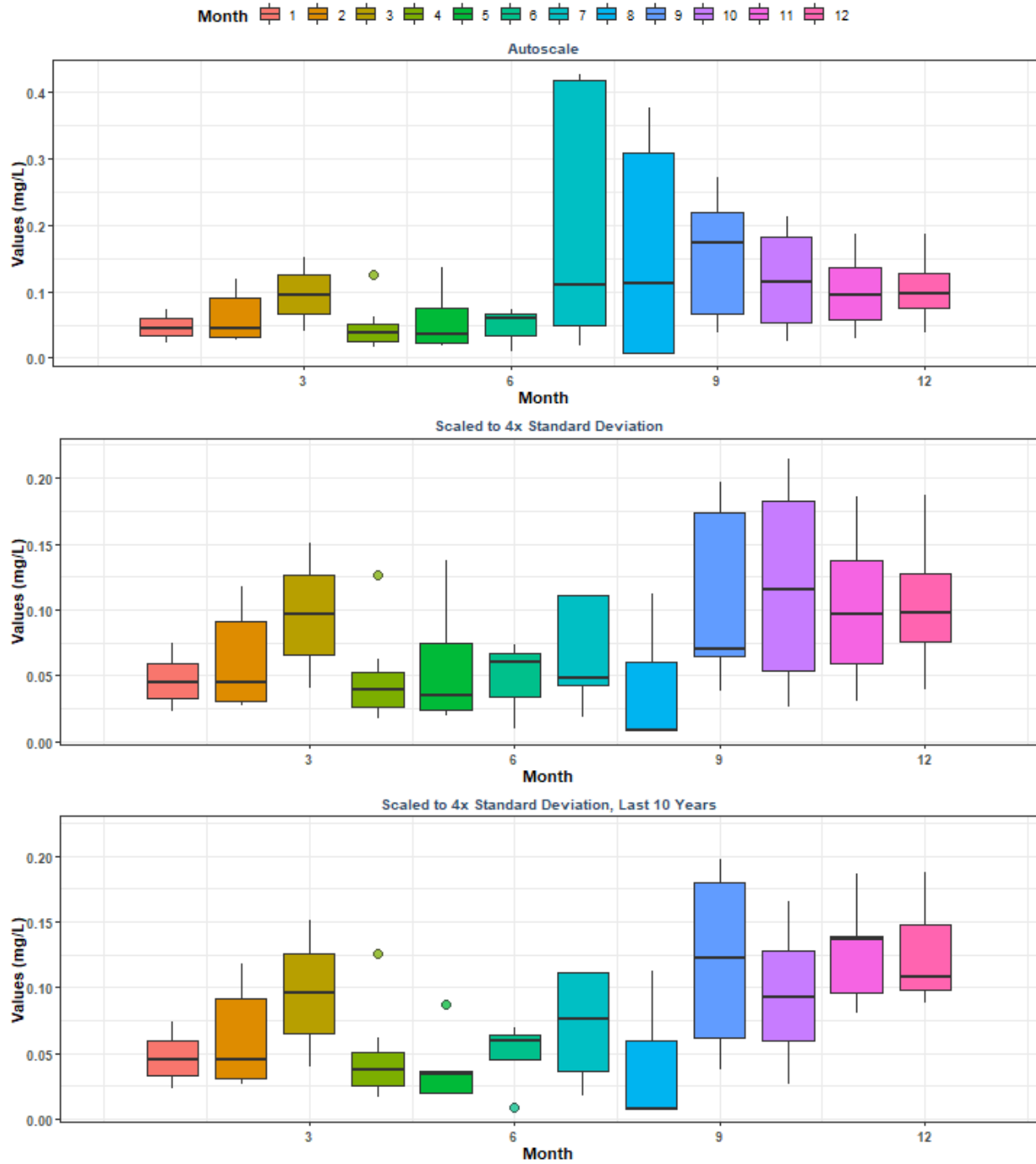
Summary Box Plots for Cape Haze Aquatic Preserve

By Year & Month

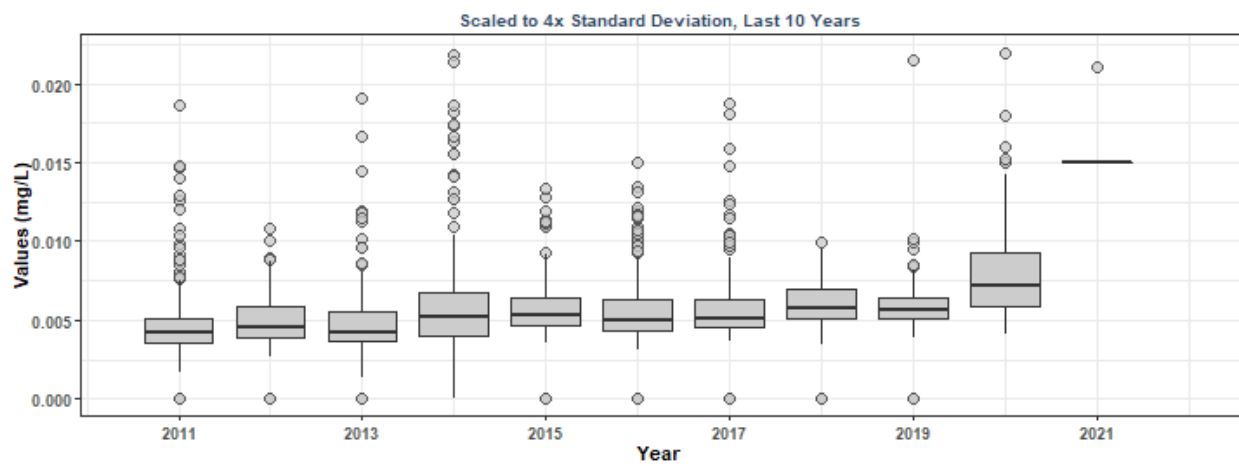
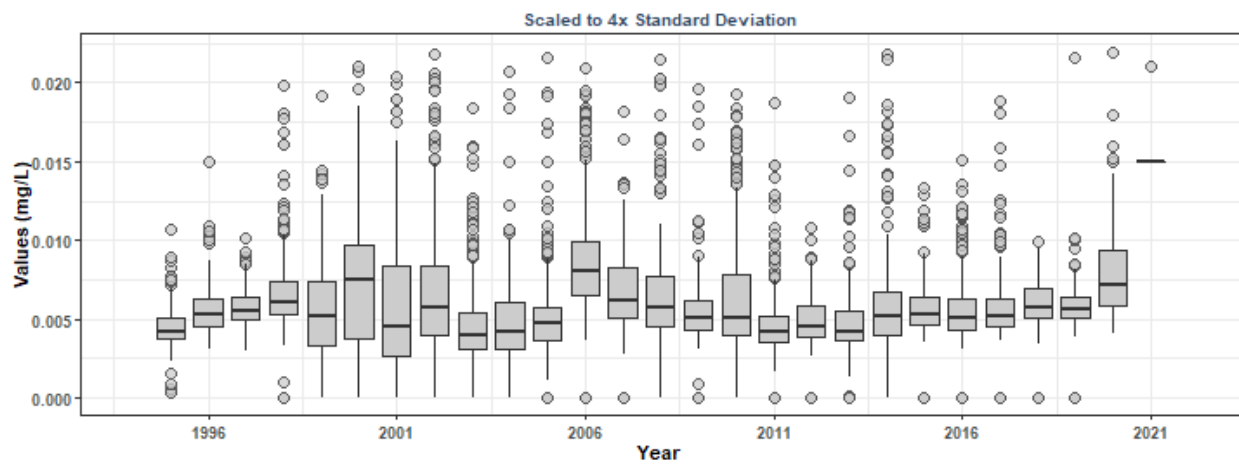
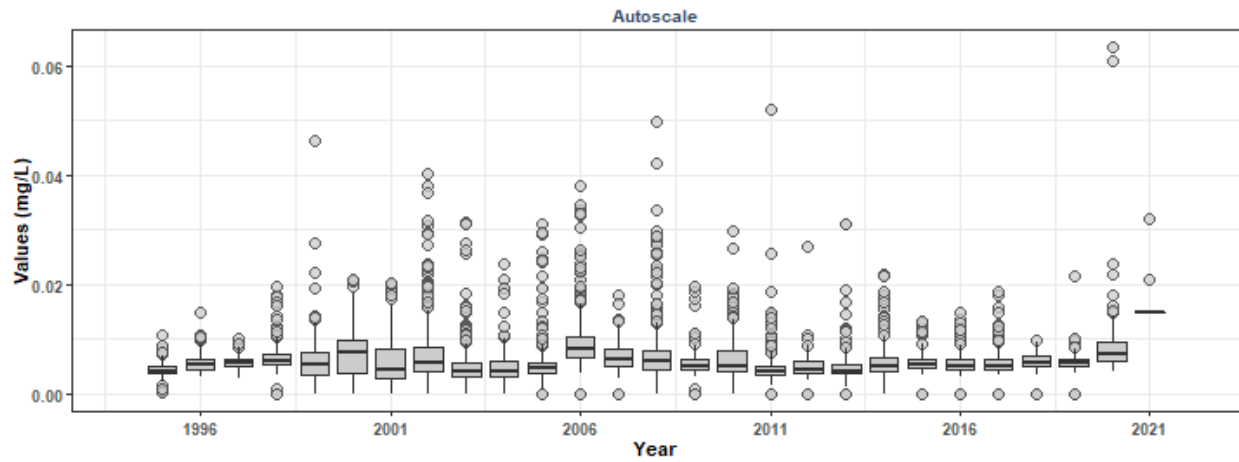


Summary Box Plots for Cape Haze Aquatic Preserve

By Month

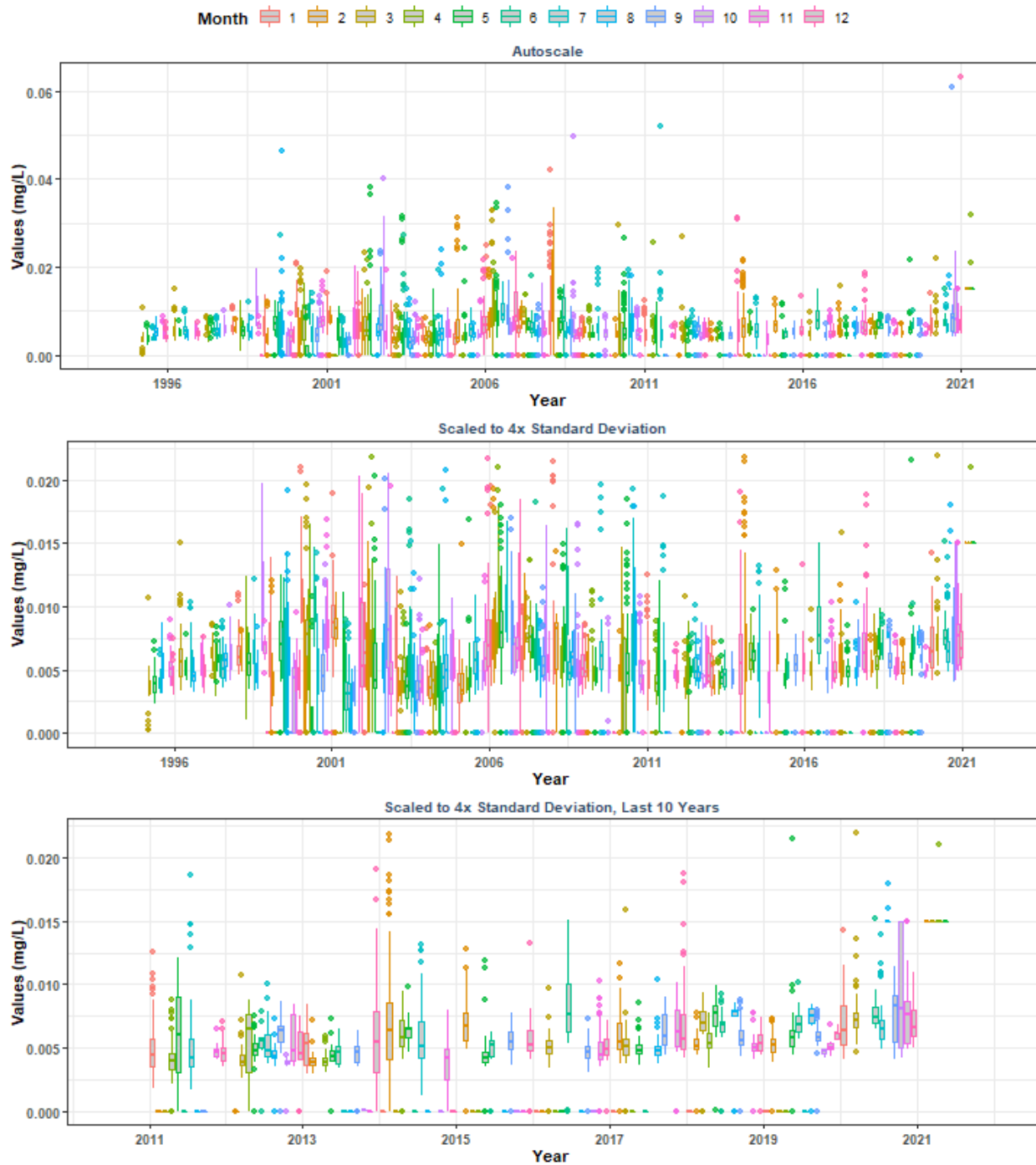


Summary Box Plots for Florida Keys National Marine Sanctuary By Year



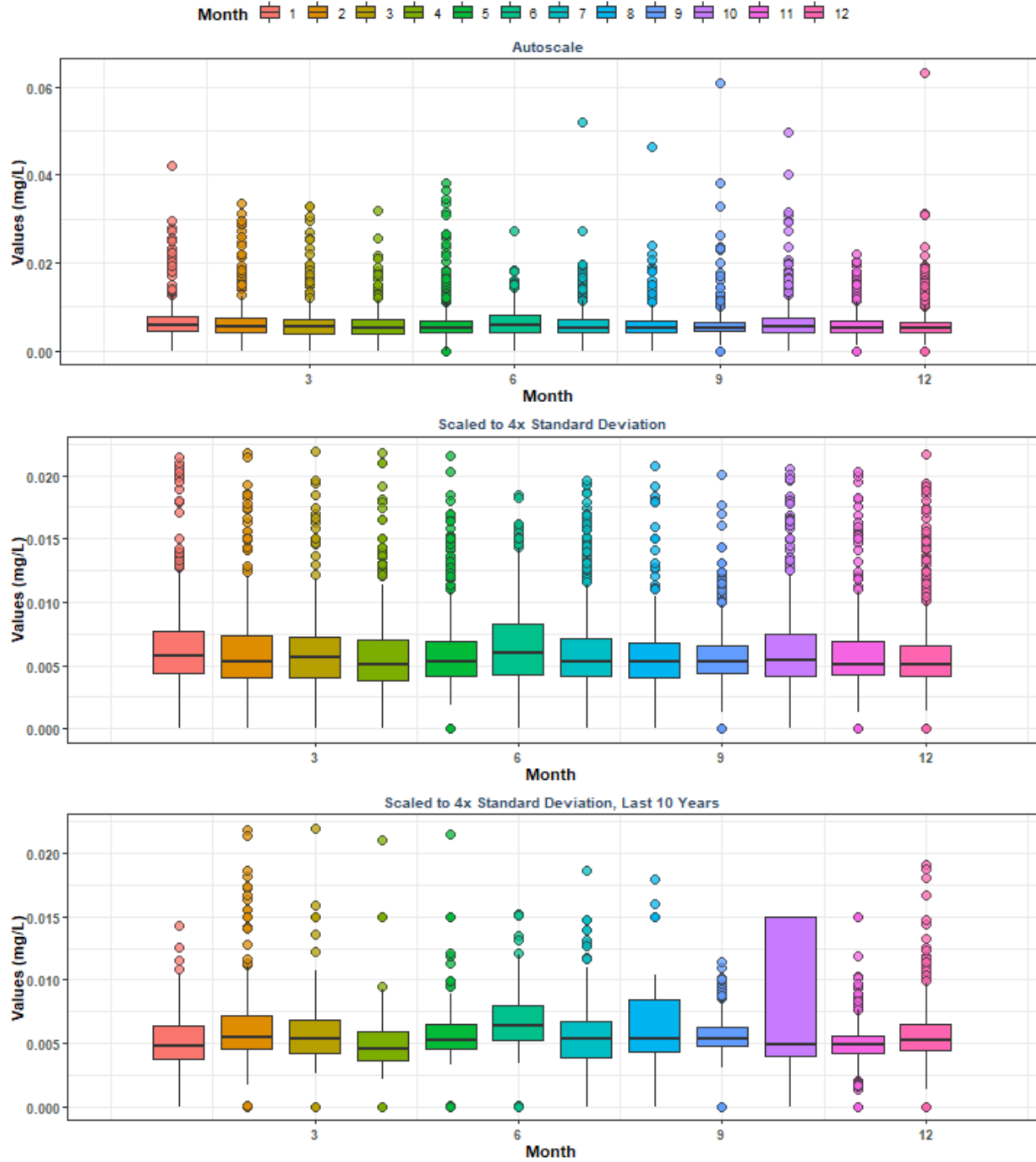
Summary Box Plots for Florida Keys National Marine Sanctuary

By Year & Month

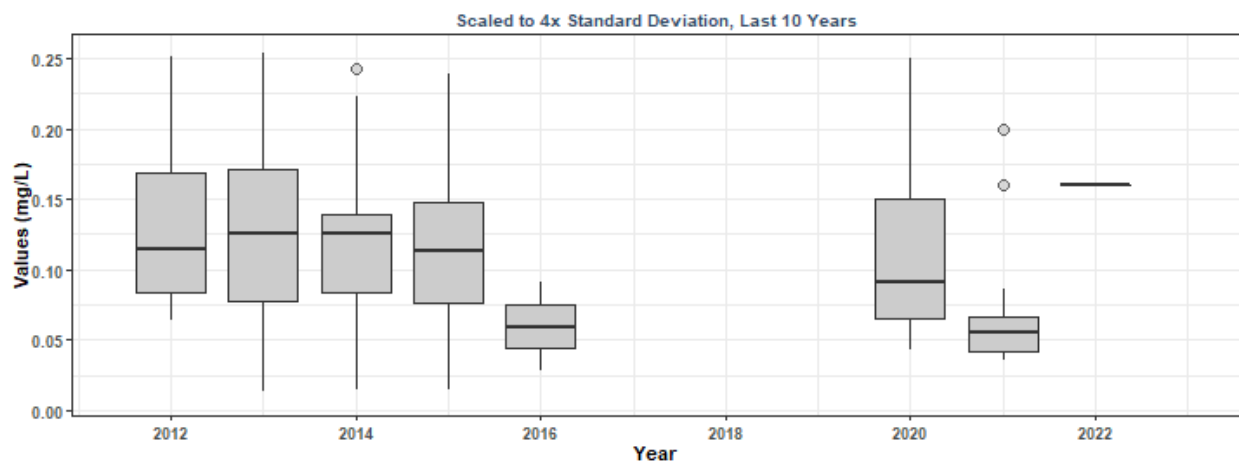
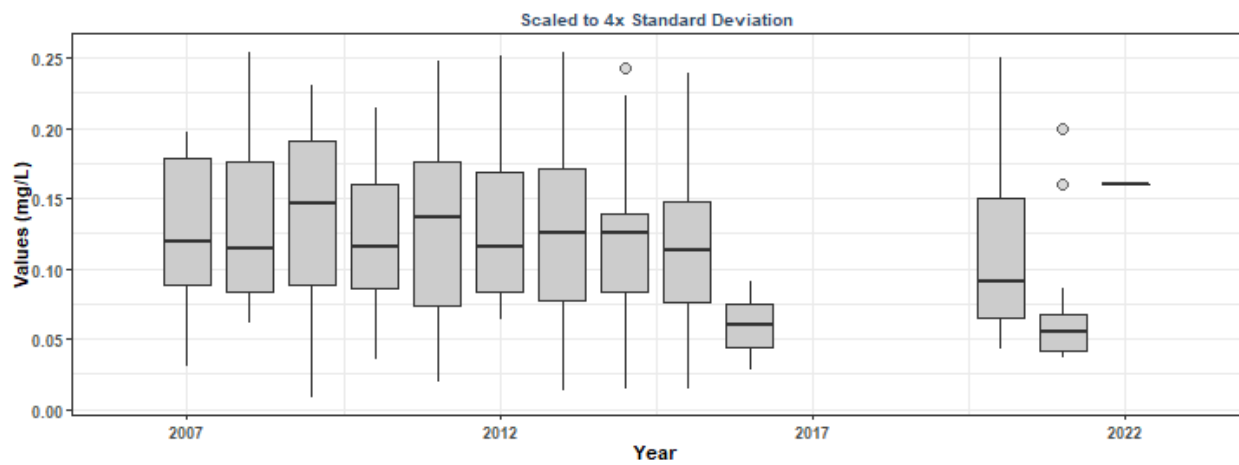
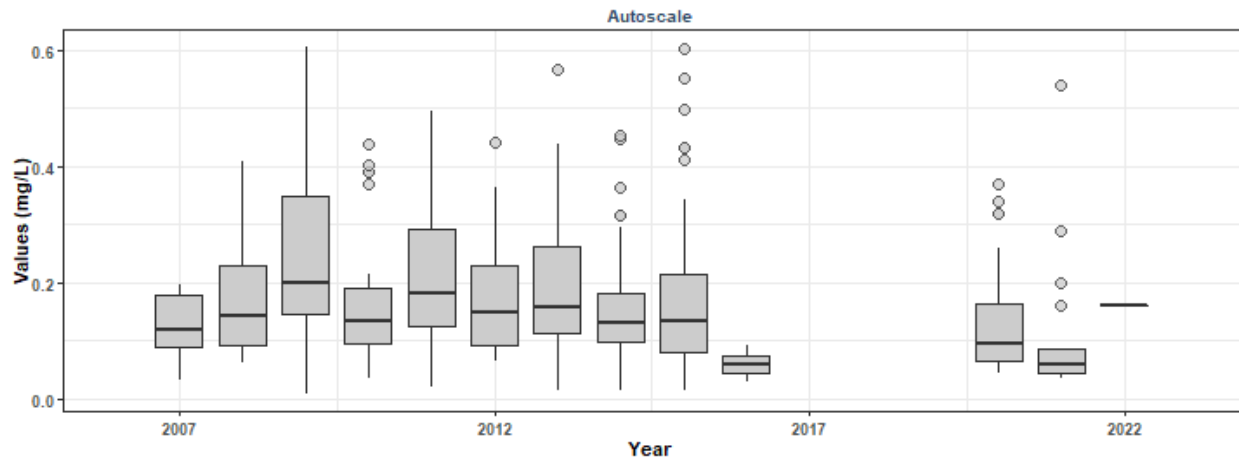


Summary Box Plots for Florida Keys National Marine Sanctuary

By Month

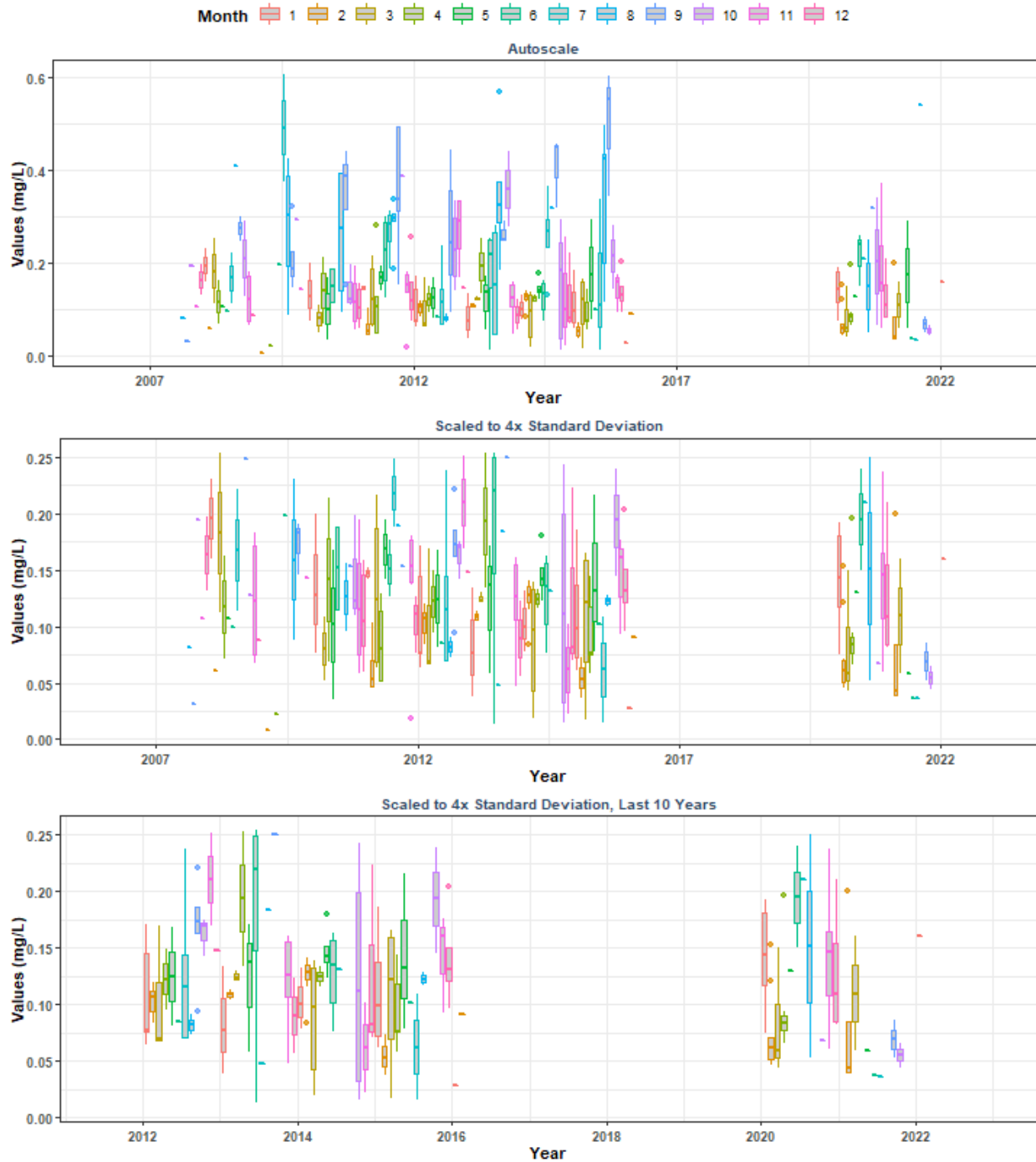


Summary Box Plots for Gasparilla Sound-Charlotte Harbor Aquatic Preserve By Year



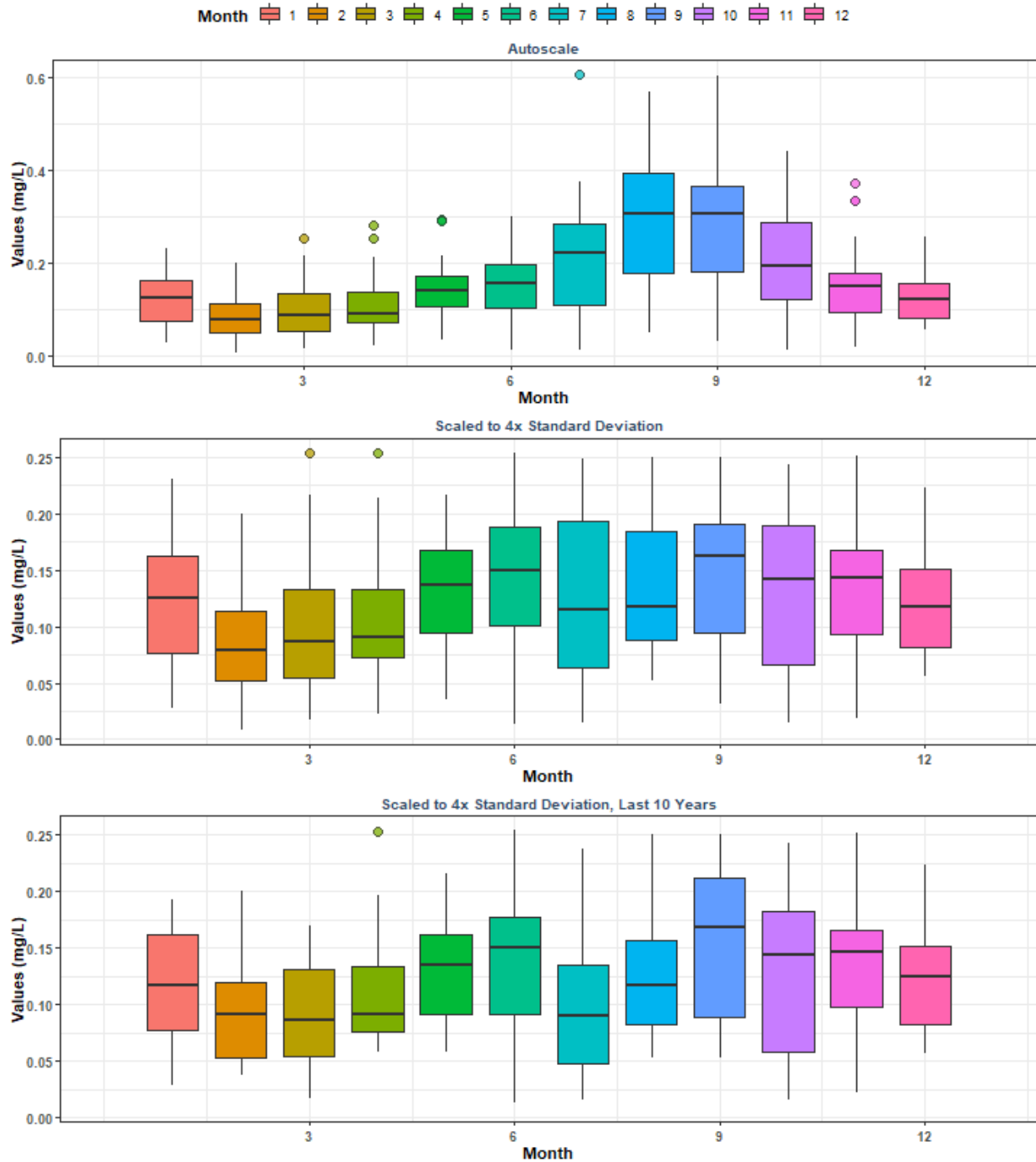
Summary Box Plots for Gasparilla Sound-Charlotte Harbor Aquatic Preserve

By Year & Month



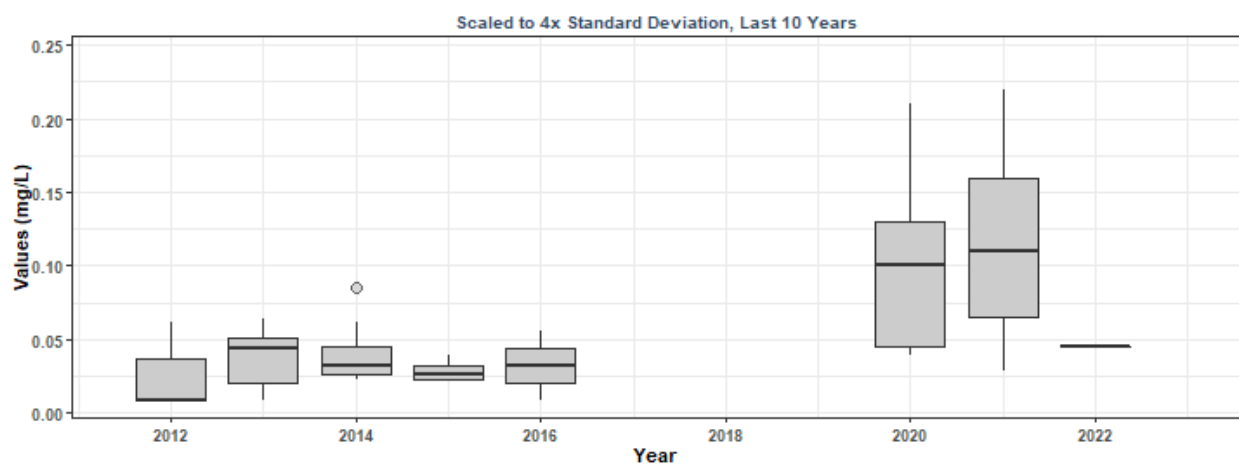
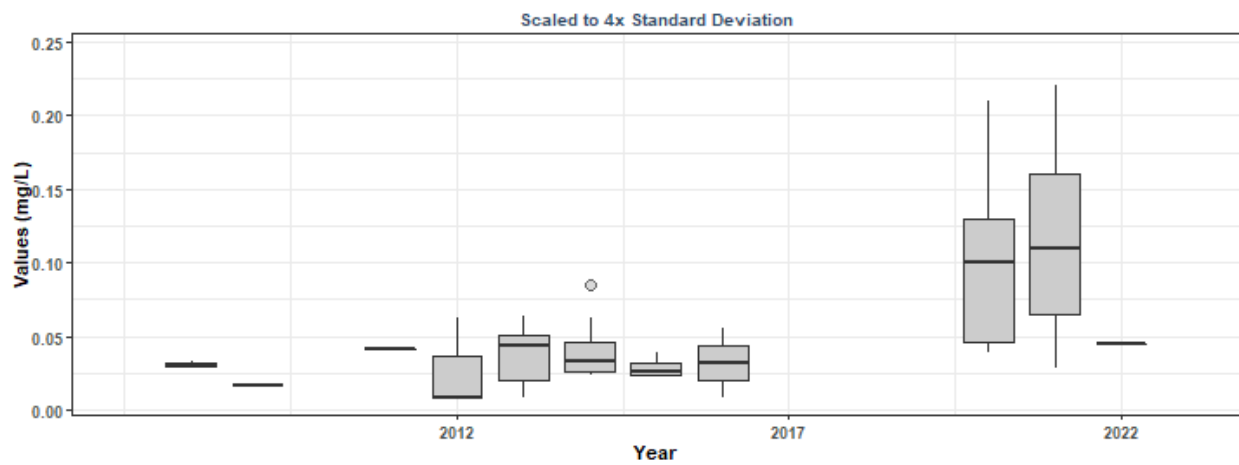
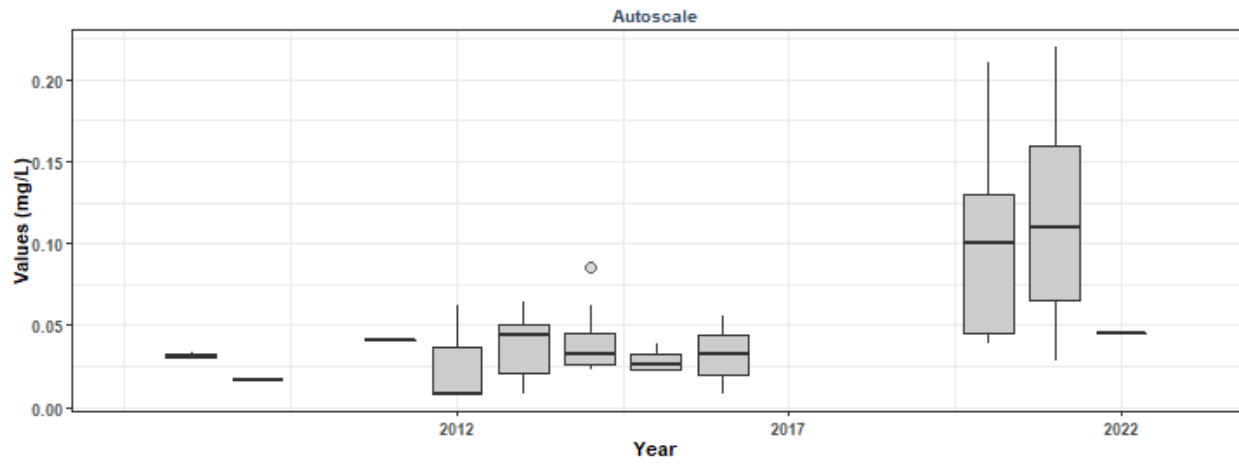
Summary Box Plots for Gasparilla Sound-Charlotte Harbor Aquatic Preserve

By Month



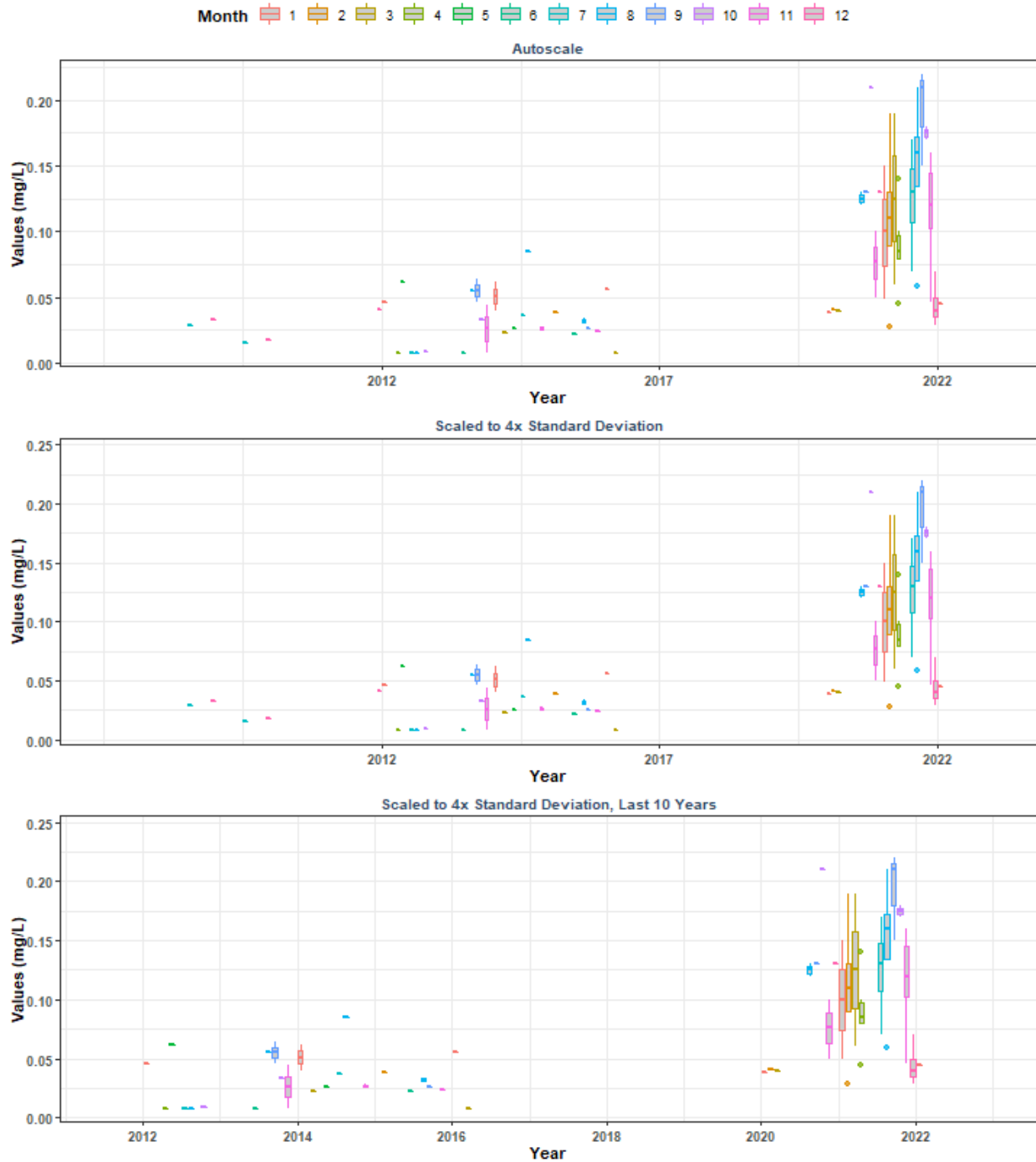
Summary Box Plots for Lemon Bay Aquatic Preserve

By Year



Summary Box Plots for Lemon Bay Aquatic Preserve

By Year & Month

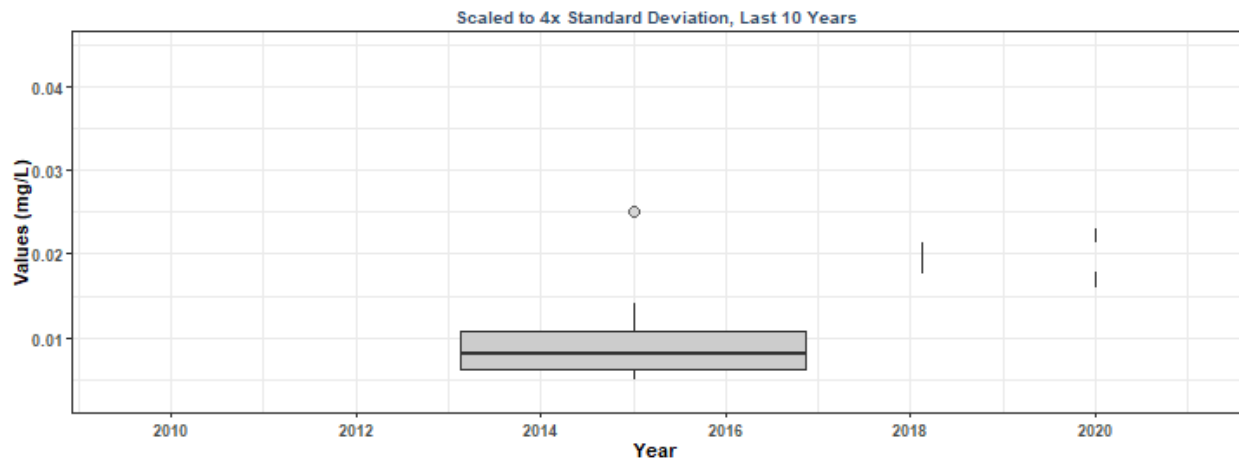
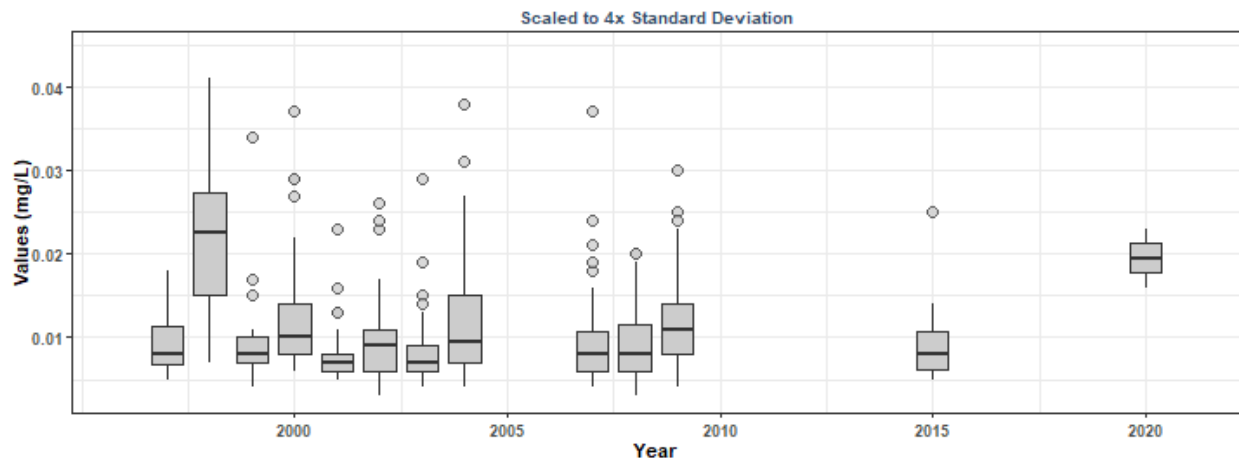
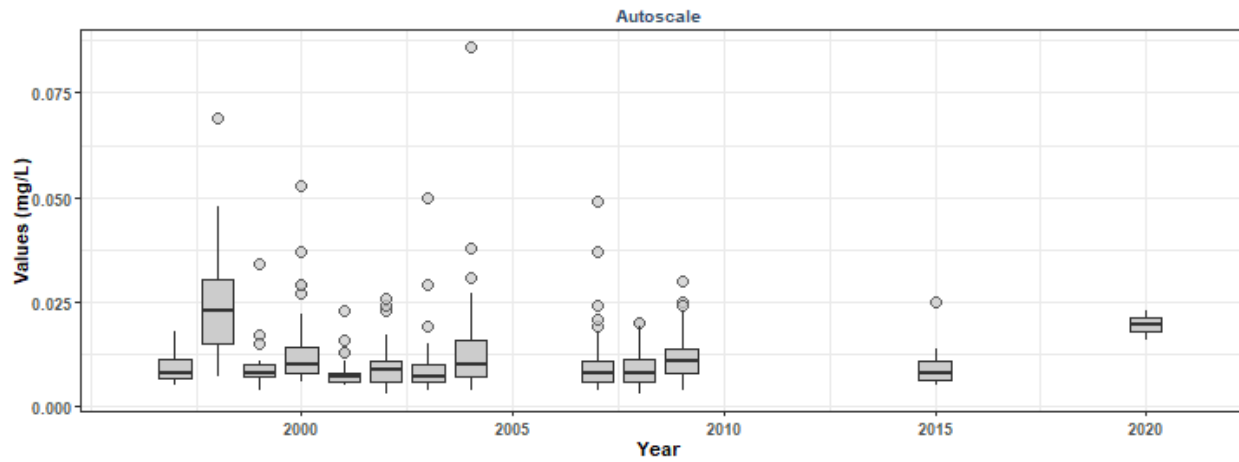


By Month



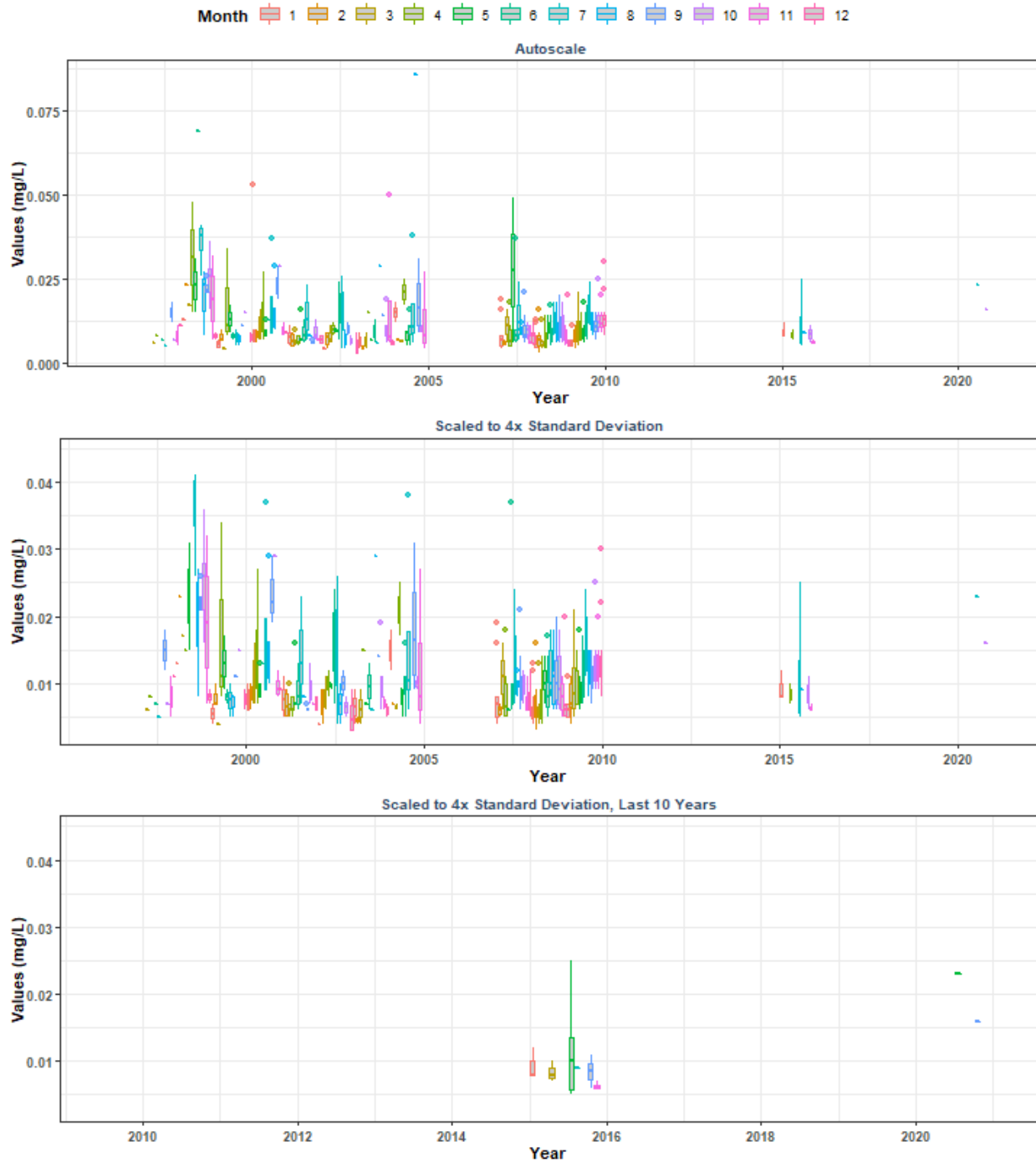
Summary Box Plots for Nature Coast Aquatic Preserve

By Year



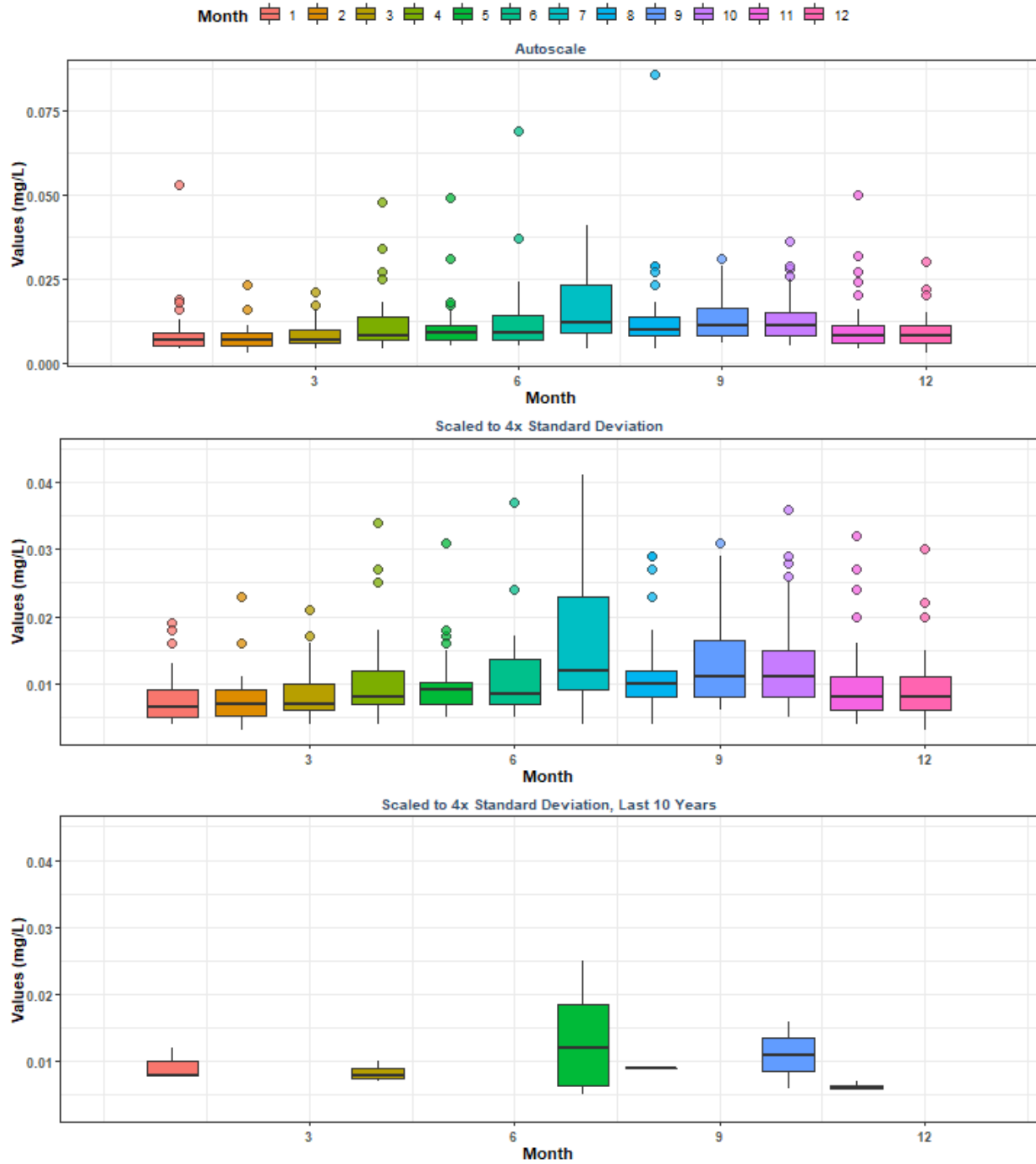
Summary Box Plots for Nature Coast Aquatic Preserve

By Year & Month

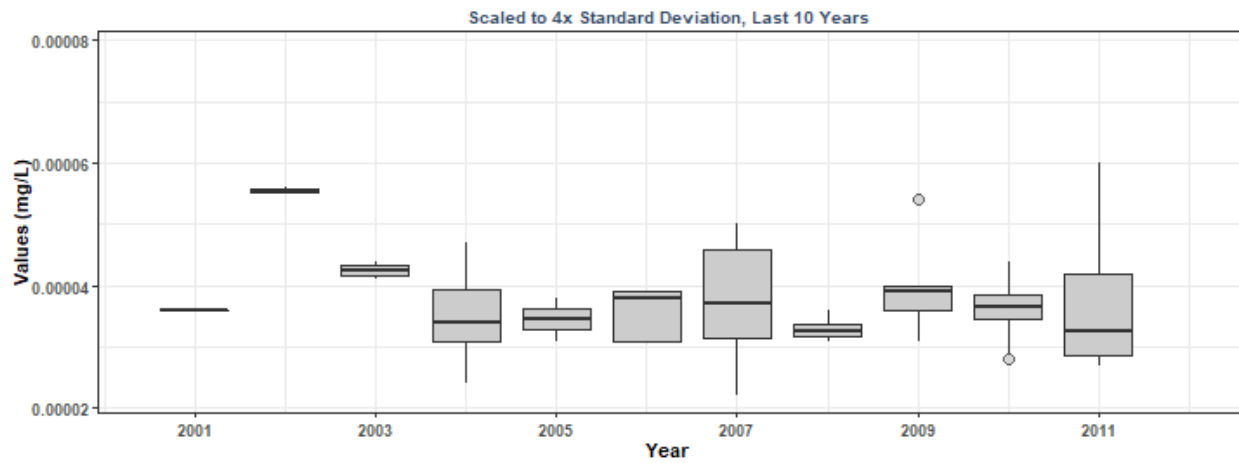
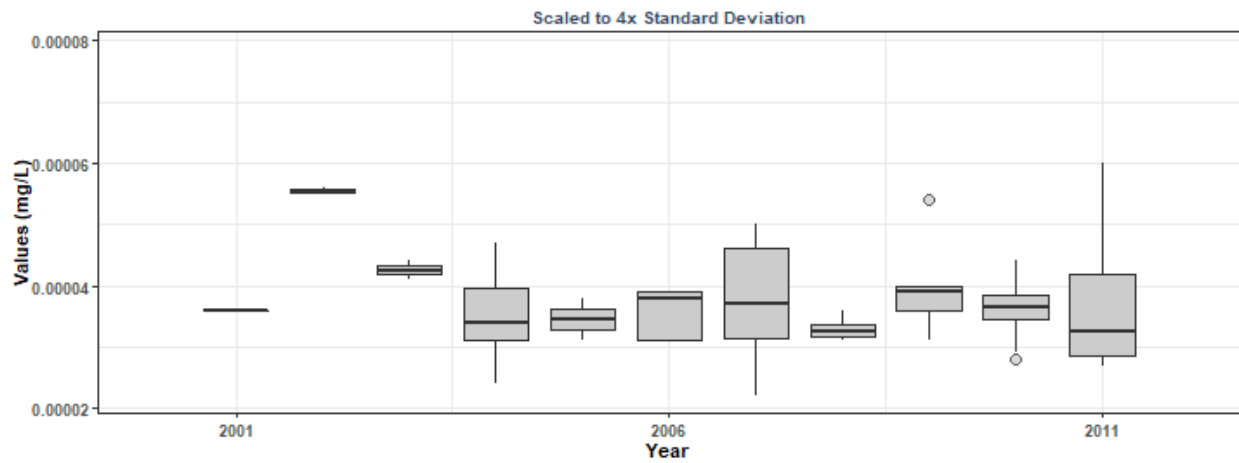
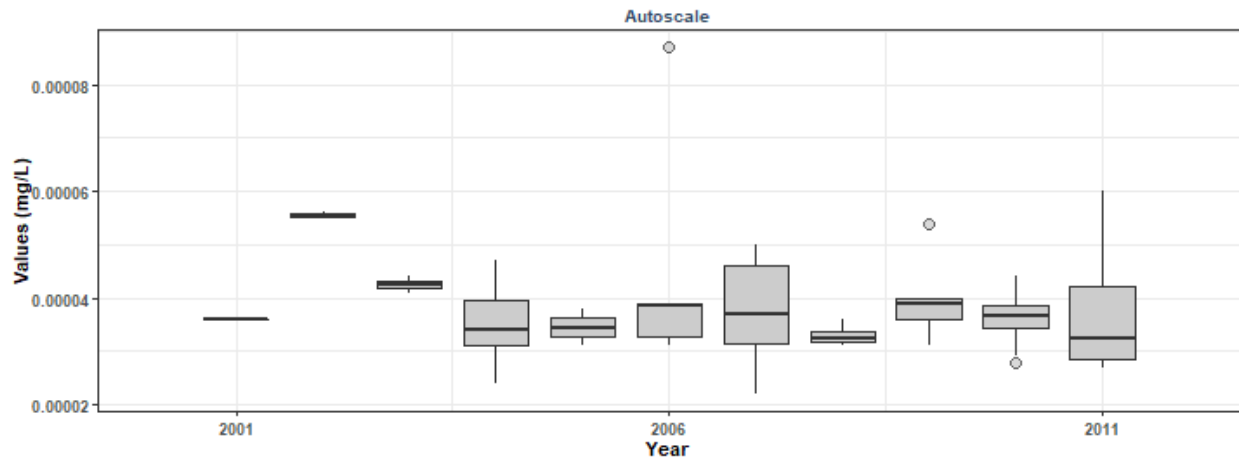


Summary Box Plots for Nature Coast Aquatic Preserve

By Month

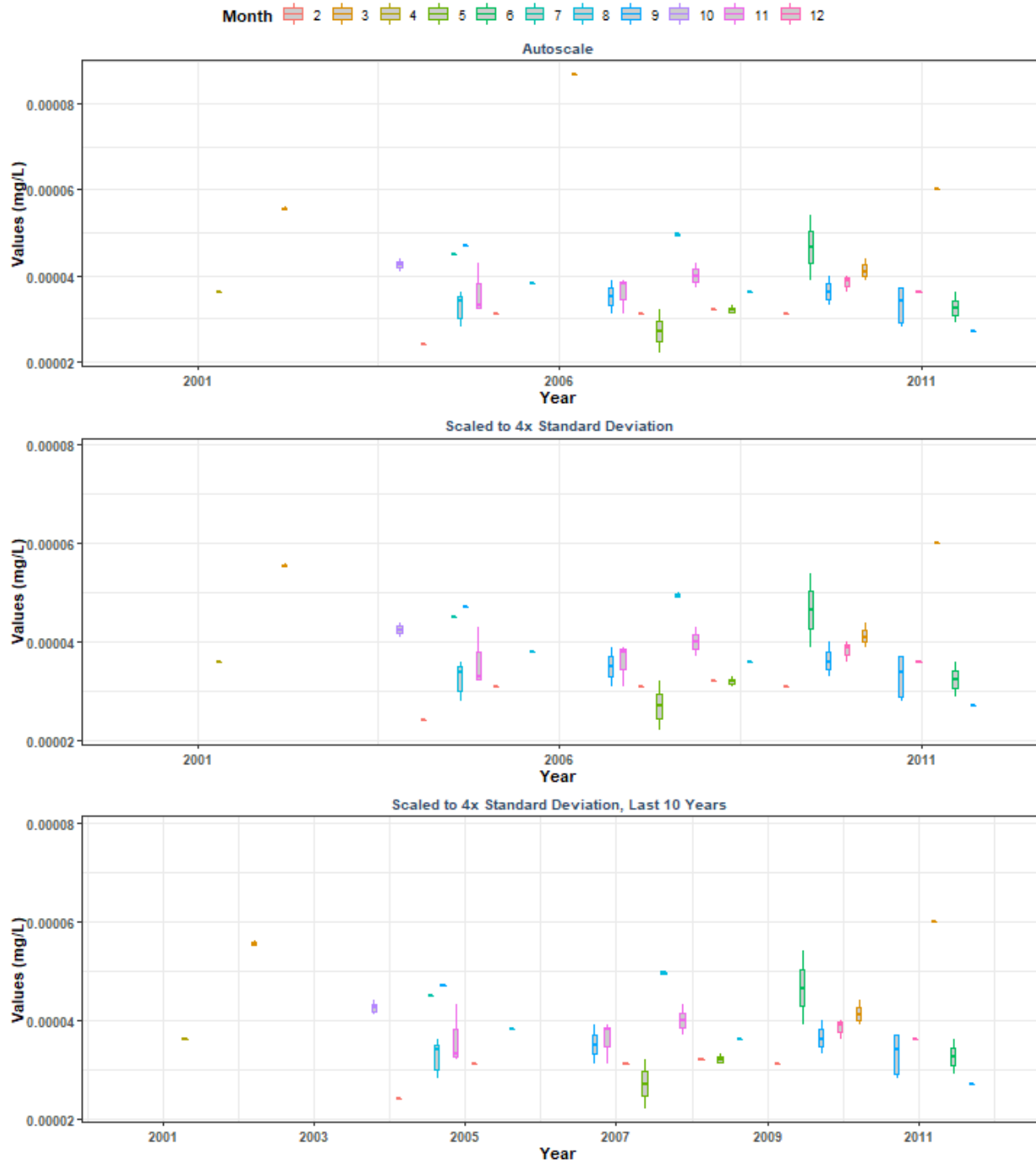


Summary Box Plots for Rookery Bay Aquatic Preserve By Year



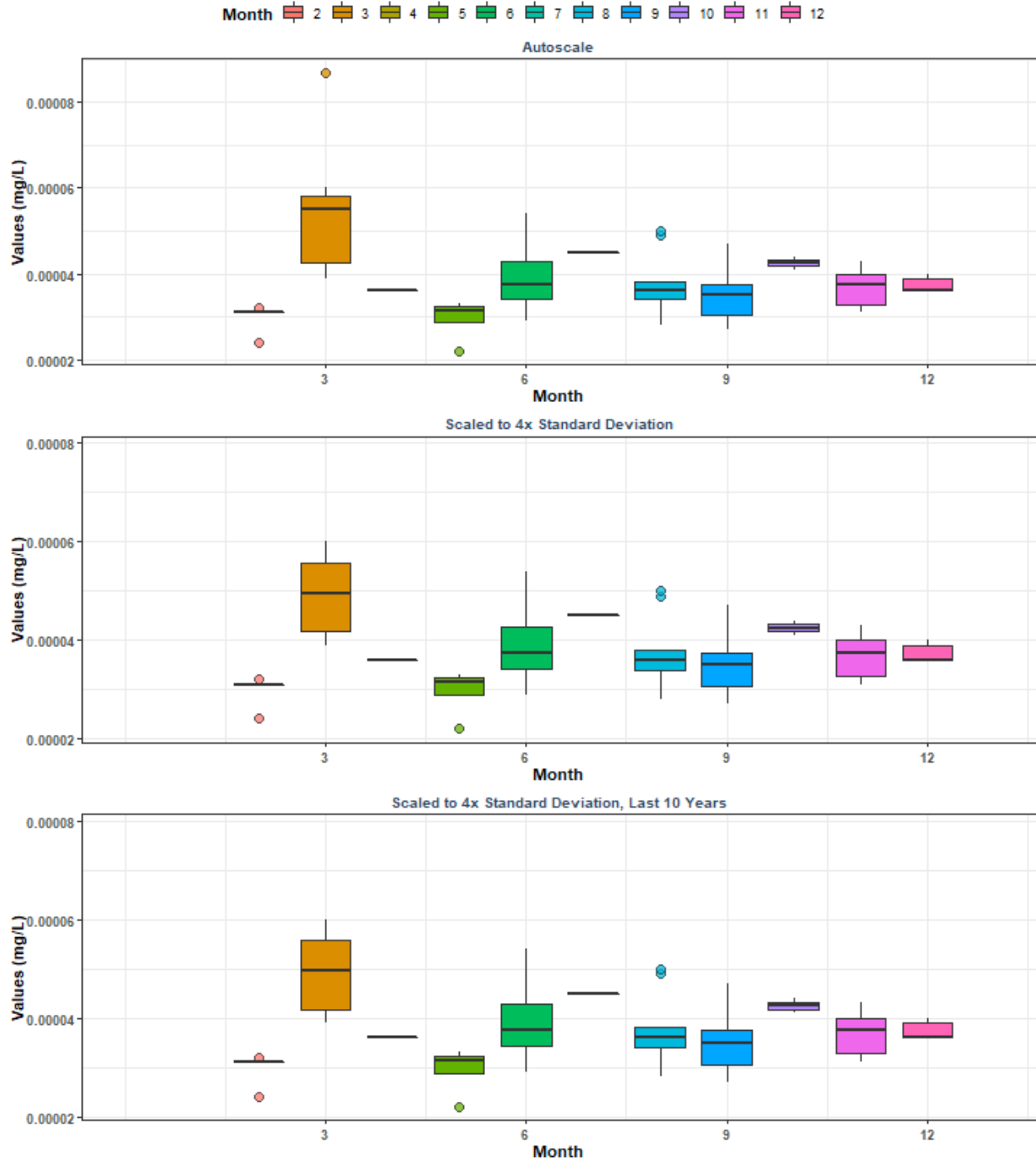
Summary Box Plots for Rookery Bay Aquatic Preserve

By Year & Month

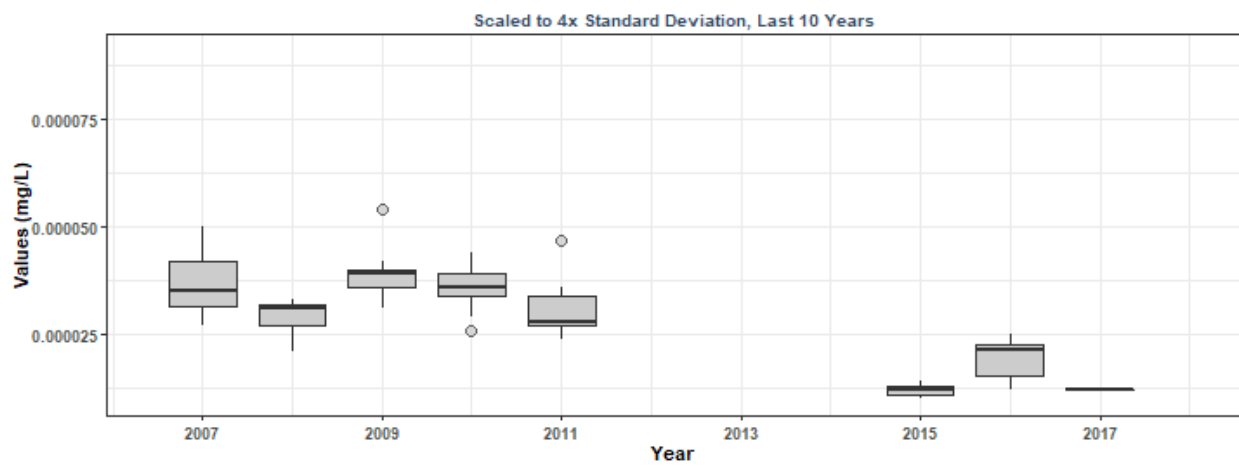
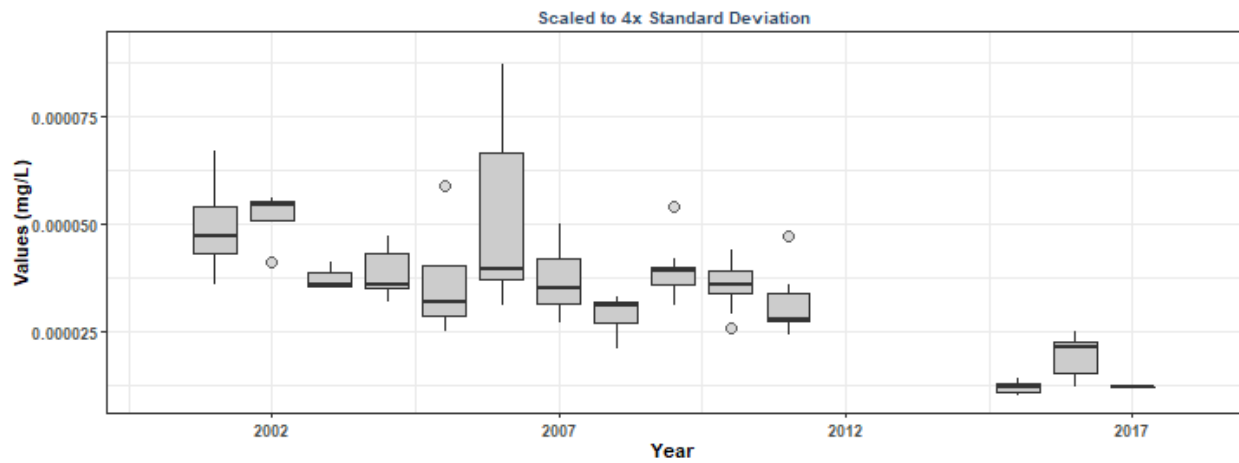
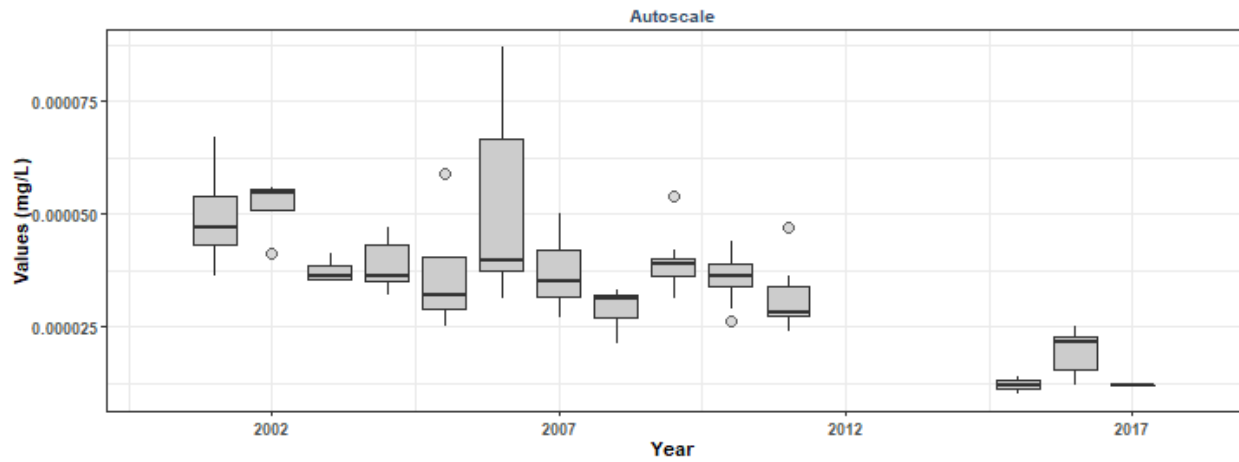


Summary Box Plots for Rookery Bay Aquatic Preserve

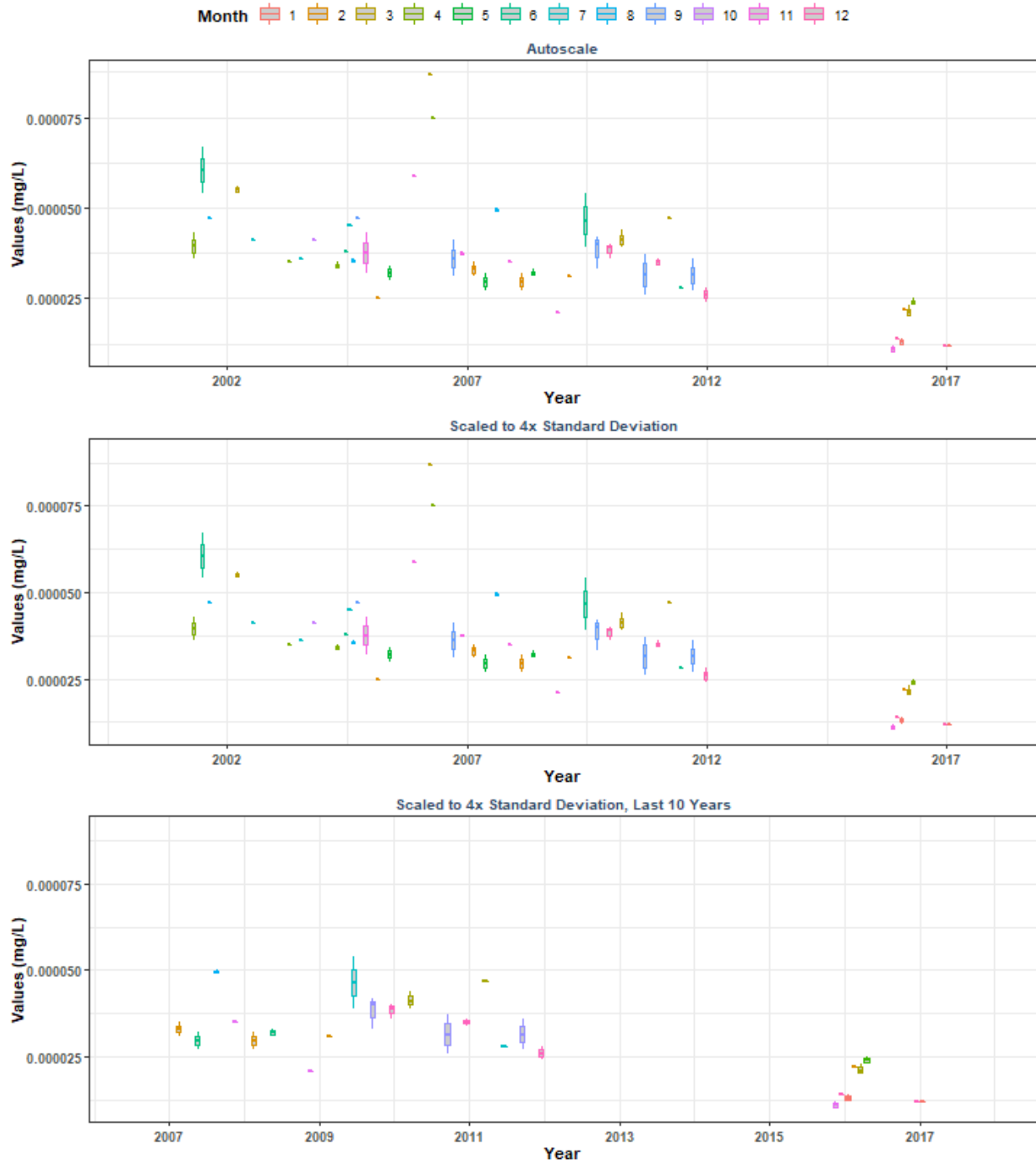
By Month



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

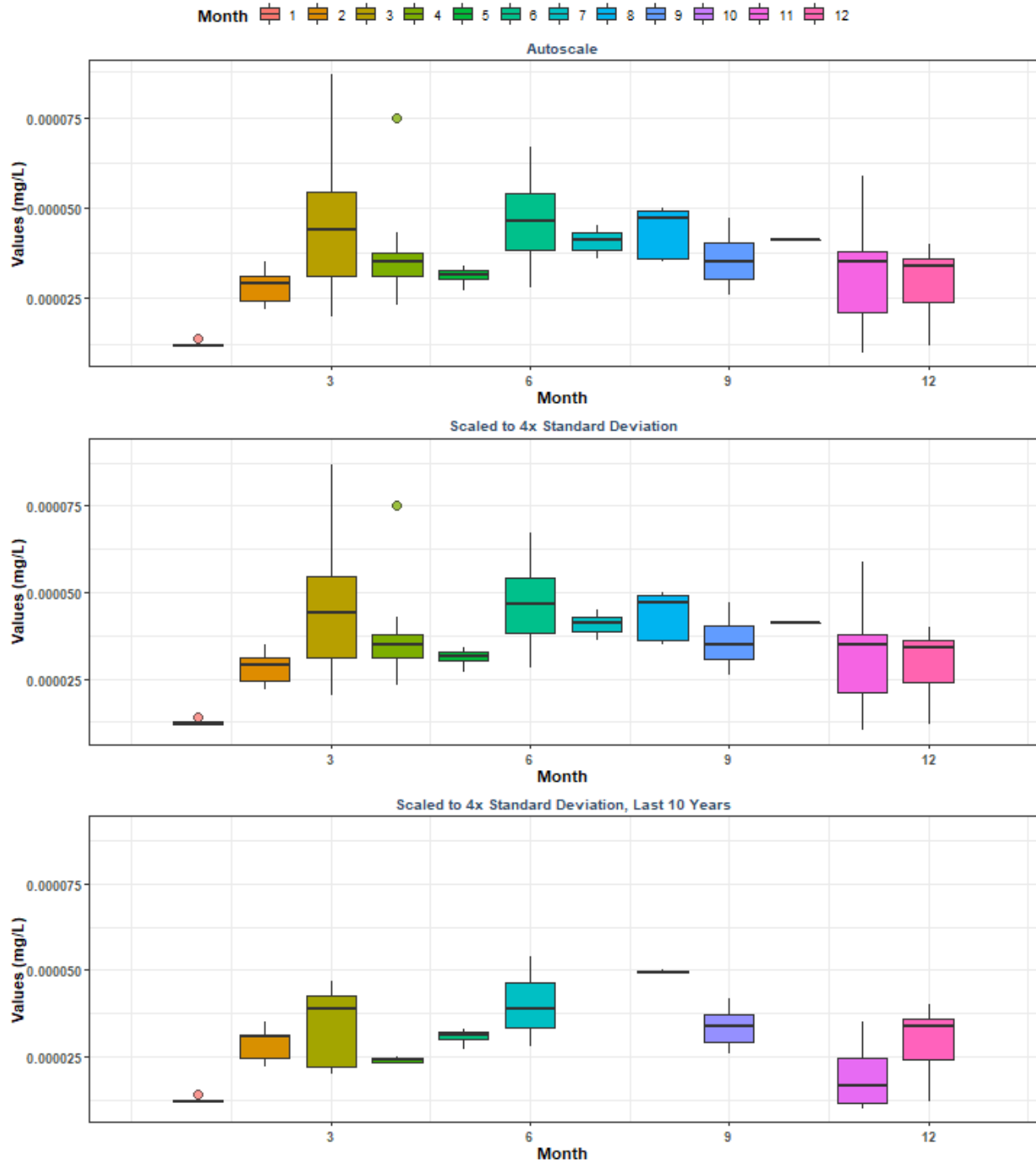


Summary Box Plots for Rookery Bay National Estuarine Research Reserve By Year & Month

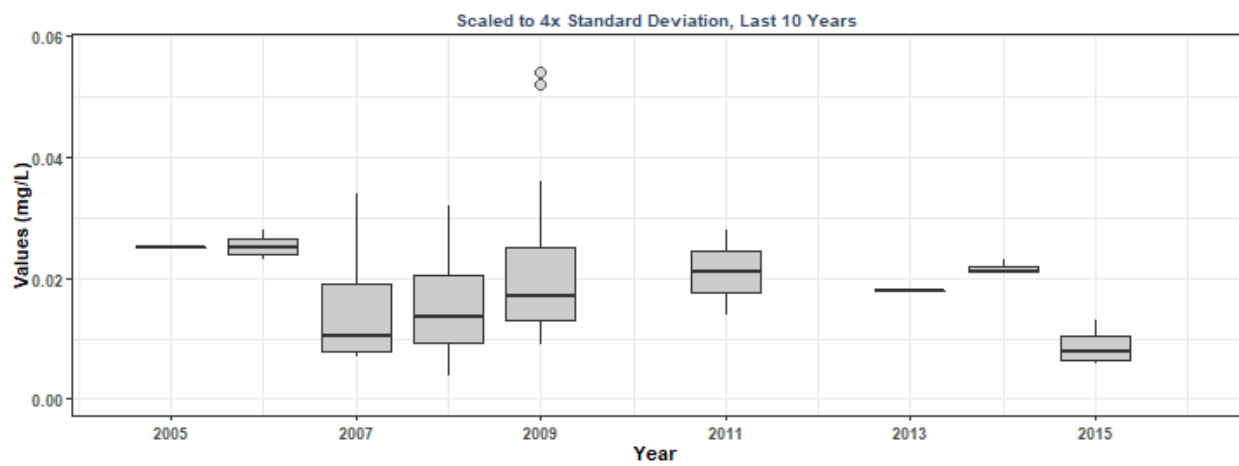
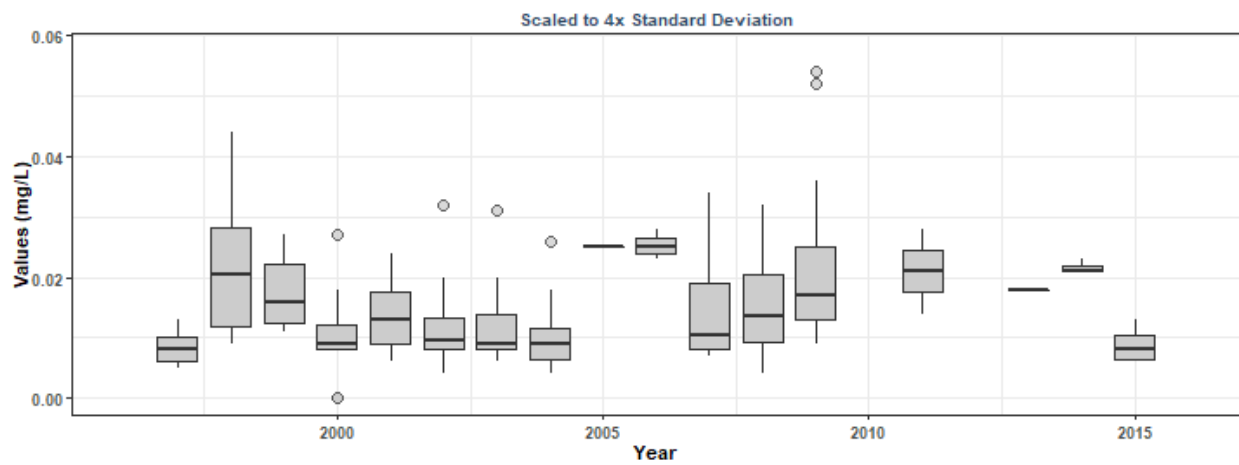
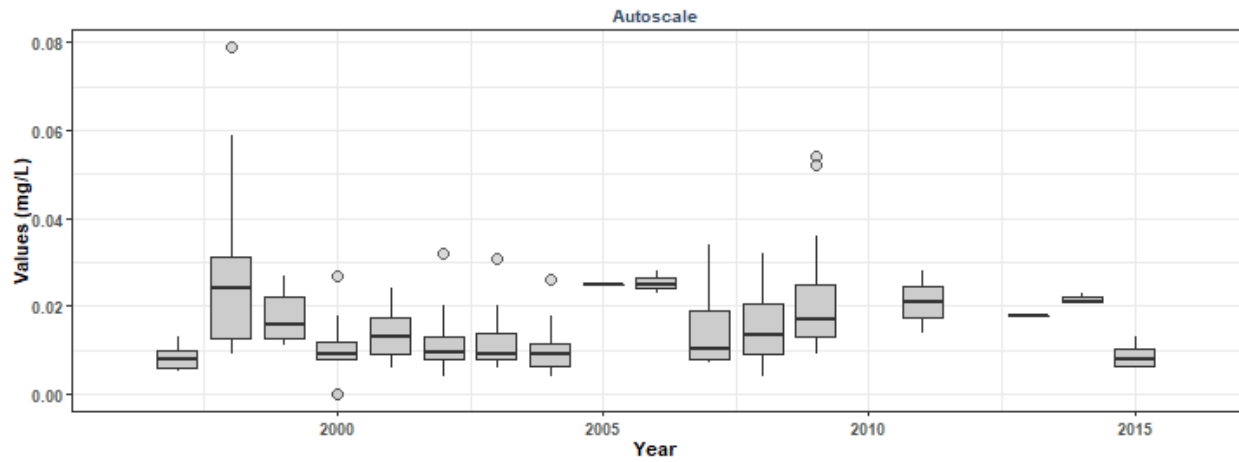


Summary Box Plots for Rookery Bay National Estuarine Research Reserve

By Month

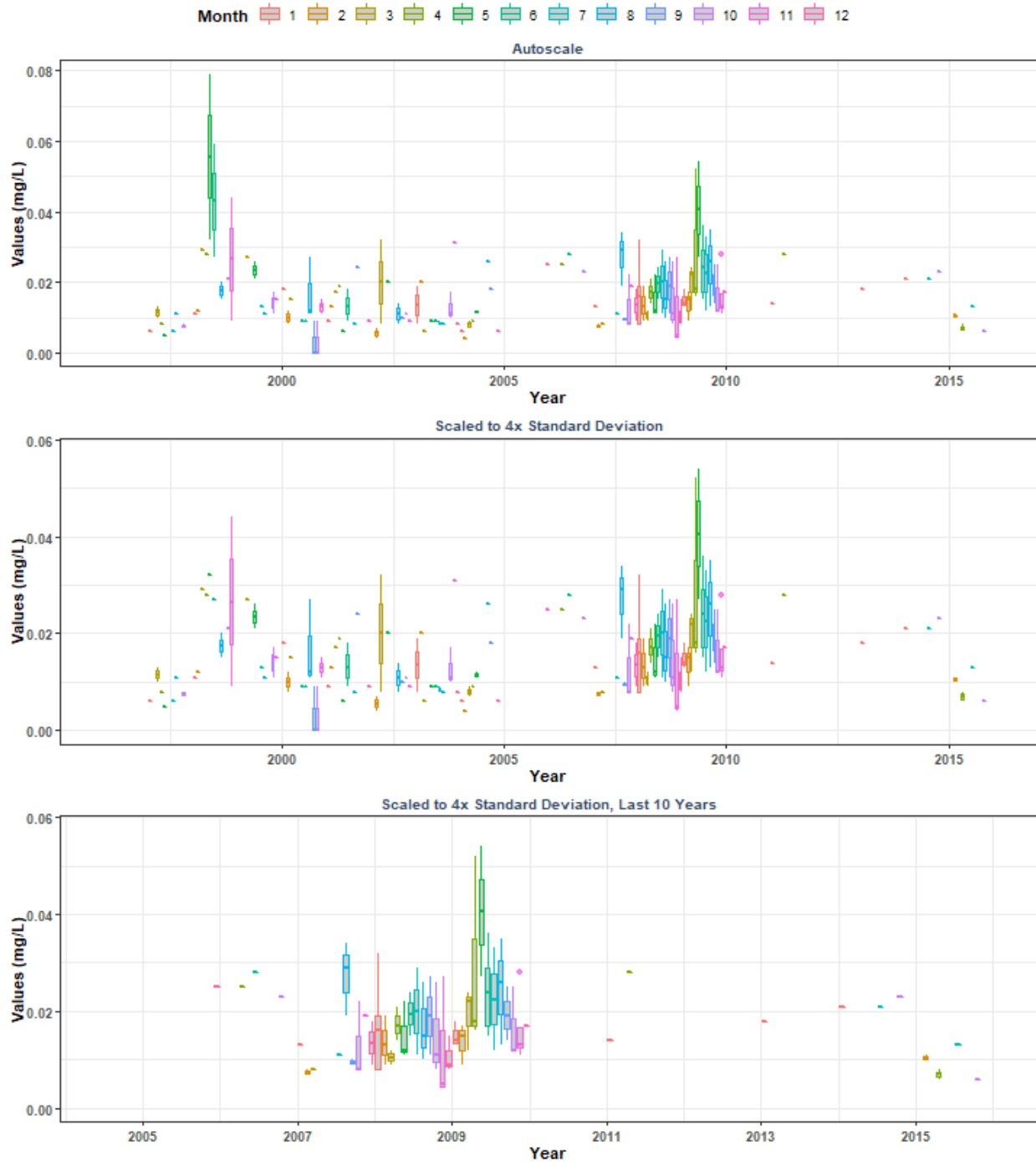


Summary Box Plots for St. Martins Marsh Aquatic Preserve By Year



Summary Box Plots for St. Martins Marsh Aquatic Preserve

By Year & Month



Summary Box Plots for St. Martins Marsh Aquatic Preserve

By Month

