Appendix A - ADS506-01-FA22 - Final Project

Team 1

12/05/2022

RMarkdown global setup

```
knitr::opts_chunk$set(fig.align = 'center')
```

```
library(AppliedPredictiveModeling)
library(BioStatR)
library(car)
library(caret)
library(class)
library(corrplot)
library(datasets)
library(e1071)
library(Hmisc)
library(lubridate)
library(mlbench)
library(gridExtra)
library(psych)
library(randomForest)
library(RANN)
library(reshape2)
library(rpart)
library(rpart.plot)
library(scales)
library(tidyverse)
library(timetk)
library(tsibble)
library(tseries)
library(zoo)

set.seed(1699)
```

Create custom function to generate boxplots and descriptive statistics for continuous variables

```
box_comp <- function(xcol = c(),
                     df = NA,
                     rtn_met = TRUE,
                     grph_title = "All",
                     box = TRUE) {
  # Define function to produce formatted boxplots
  sig <- 3
```

```r
metrics_df01 <- data.frame(metric = c("",
                                      "Total N:",
                                      "Count",
                                      "NA Count",
                                      "Mean",
                                      "Median",
                                      "Standard Deviation",
                                      "Variance",
                                      "Range",
                                      "Min",
                                      "Max",
                                      "25th Percentile",
                                      "75th Percentile",
                                      "Subset w/o Outliers:",
                                      "Count",
                                      "%",
                                      "Outlier %",
                                      "NA Count",
                                      "Mean",
                                      "Median",
                                      "Standard Deviation",
                                      "Variance",
                                      "Range",
                                      "Min",
                                      "Max"
                                      ))
for(var in xcol) {
  df_s1 <- df[, var]
  df_s1s1 <-data.frame(df_s1)
  df_s1_fit <- preProcess(df_s1s1,
                          method = c("center", "scale"))
  df_s1_trans <- predict(df_s1_fit, df_s1s1)

  # Calculate quartiles
  var_iqr_lim <- IQR(df_s1) * 1.5
  var_q1 <- quantile(df_s1, probs = c(.25))
  var_otlow <- var_q1 - var_iqr_lim
  var_q3 <- quantile(df_s1, probs = c(.75))
  var_othigh <- var_q3 + var_iqr_lim

  # Subset non-outlier data
  var_non_otlr_df01 <- subset(df, (abs(df_s1_trans) <= 3))
  #var_non_otlr_df01 <- subset(df, (df_s1 > var_otlow & df_s1 < var_othigh))
  df_s2 <- var_non_otlr_df01[, var]

  # Begin calculating measures of centrality & dispersion
  var_mean <- mean(df_s1, na.remove = TRUE)
  var_non_otlr_df01_trunc_mean <- mean(df_s2, na.remove = TRUE)
  var_med <- median(df_s1)
  var_non_otlr_df01_trunc_med <- median(df_s2)
  var_mode <- mode(df_s1)
  var_non_otlr_df01_trunc_mode <- mode(df_s2)
  var_stde <- sd(df_s1)
```

```
    var_non_otlr_df01_trunc_stde <- sd(df_s2)
    var_vari <- var(df_s1)
    var_non_otlr_df01_trunc_vari <- var(df_s2)
    var01_min <- min(df[, var])
    var01_max <- max(df[, var])
    var01_range <- var01_max - var01_min
    var02_min <- min(var_non_otlr_df01[, var])
    var02_max <- max(var_non_otlr_df01[, var])
    var02_range <- var02_max - var02_min

    # Configure y-axis min & max to sync graphs
    plot_min <- min(var01_min, var02_min)
    plot_max <- max(var01_max, var02_max)
    nonoutlier_perc <- round((as.numeric(dim(var_non_otlr_df01)[1] /
↪   as.numeric(dim(df)[1]))) * 100, 1)

    # Fill in metrics table
    measure_val01 <- c(paste0("Variable: ", var),
                       "",
                       as.character(dim(df)[1]),
                       sum(is.na(df_s1)),
                       round(var_mean, sig),
                       round(var_med, sig),
                       round(var_stde, sig),
                       round(var_vari, sig),
                       round(var01_range, sig),
                       round(var01_min, sig),
                       round(var01_max, sig),
                       round(var_q1, sig),
                       round(var_q3, sig),
                       "",
                       as.character(dim(var_non_otlr_df01)[1]),
                       paste0(nonoutlier_perc, "%"),
                       paste0(round(100 - nonoutlier_perc, 1), "%"),
                       sum(is.na(df_s2)),
                       round(var_non_otlr_df01_trunc_mean, sig),
                       round(var_non_otlr_df01_trunc_med, sig),
                       round(var_non_otlr_df01_trunc_stde, sig),
                       round(var_non_otlr_df01_trunc_vari, sig),
                       round(var02_range, sig),
                       round(var02_min, sig),
                       round(var02_max, sig)
                       )

    var_name <- paste0("Variable: ", var)
    metrics_df01[, ncol(metrics_df01) + 1] <- measure_val01
}
  # Format boxplot titles based on number of plots
  if(box == TRUE) {
    if(length(xcol == 1)) {
      boxplot(df,
              ylab = "Parameter Values",
              main = paste0("Boxplot for ", xcol, " (", grph_title, ")"))
```

```
    }
    else {
      boxplot(df,
              ylab = "Parameter Values",
              main = paste0("Boxplot for Multiple Parameters", " (", grph_title, ")"))

    }
  }
  # Return and print metrics table(s)
  if(rtn_met == TRUE) {
    print(metrics_df01)
    return(metrics_df01)
  }
}
```

Import and merge Datasets

```
# Import 4 separate CSV files
owt_df01a <- read.csv("../data/Ocean Water/water_quality_1990_1999_datasd.csv", header =
↪  TRUE, sep = ",")
owt_df01b <- read.csv("../data/Ocean Water/water_quality_2000_2010_datasd.csv", header =
↪  TRUE, sep = ",")
owt_df01c <- read.csv("../data/Ocean Water/water_quality_2011_2019_datasd.csv", header =
↪  TRUE, sep = ",")
owt_df01d <- read.csv("../data/Ocean Water/water_quality_2020_2021_datasd.csv", header =
↪  TRUE, sep = ",")

# Merge 4 seperate dataframes into 1
owt_df01 <- rbind(owt_df01a, owt_df01b, owt_df01c, owt_df01d)

print(head(owt_df01))
```

```
##       sample station depth_m date_sample time project   parameter qualifier
## 1 9011158743      C5       9  1990-11-15         PLOO CHLOROPHYLL
## 2 9011158743      C5       9  1990-11-15         PLOO     DENSITY
## 3 9011158743      C5       9  1990-11-15         PLOO          DO
## 4 9011158743      C5       9  1990-11-15         PLOO          PH
## 5 9011158743      C5       9  1990-11-15         PLOO    SALINITY
## 6 9011158743      C5       9  1990-11-15         PLOO        TEMP
##    value   units
## 1  0.870    ug/L
## 2 23.855 sigma-t
## 3  6.550    mg/L
## 4  8.080      pH
## 5 33.617     ppt
## 6 19.430       C
```

```
describe(owt_df01)
```

```
##             vars       n   mean      sd min     max   range   se
## sample         1 1236769    NaN      NA Inf    -Inf    -Inf   NA
## station        2 1236769    NaN      NA Inf    -Inf    -Inf   NA
## depth_m        3 1152608  19.38   25.07   1     116     115 0.02
```

```
## date_sample    4 1236769    NaN      NA Inf    -Inf    -Inf   NA
## time           5 1236769    NaN      NA Inf    -Inf    -Inf   NA
## project        6 1236769    NaN      NA Inf    -Inf    -Inf   NA
## parameter      7 1236769    NaN      NA Inf    -Inf    -Inf   NA
## qualifier      8 1236769    NaN      NA Inf    -Inf    -Inf   NA
## value          9 1231466 124.24 1785.21 -37 1100000 1100037 1.61
## units         10 1236769    NaN      NA Inf    -Inf    -Inf   NA
```

```
#write.csv(owt_df01, "C:/Users/acarr/Sync/Programming/Python and R/ADS Markdown
↪  Files/ADS-506 Markdown/Final Project/data/Ocean Water/ocean_df01.csv")
```

Factorize and format column types; print NA counts

```
# List of parameter values
param_lst01 <- c("CHLOROPHYLL",
                 "DENSITY",
                 "DO",
                 "ENTERO",
                 "FECAL",
                 "OG",
                 "PH",
                 "SALINITY",
                 "SUSO",
                 "TEMP",
                 "TOTAL",
                 "XMS")

# List of col names
col_lst01 <- c("sample",
               "station",
               "date_sample",
               "time",
               "project",
               "parameter",
               "qualifier",
               "units")

# Citation: https://www.geeksforgeeks.org/find-columns-and-rows-with-na-in-r-dataframe/
owt_df02 <- owt_df01
for(c in col_lst01) {
  owt_df02[, c] <- as.factor(owt_df02[, c] )
}

# Generate NA summary tables
# Citation: https://www.geeksforgeeks.org/replace-character-value-with-na-in-r/
owt_df02[owt_df02 == ""] <- NA
print(head(owt_df02))
```

```
##       sample station depth_m date_sample time project   parameter qualifier
## 1 9011158743      C5       9  1990-11-15 <NA>    PLOO CHLOROPHYLL      <NA>
## 2 9011158743      C5       9  1990-11-15 <NA>    PLOO     DENSITY      <NA>
## 3 9011158743      C5       9  1990-11-15 <NA>    PLOO          DO      <NA>
## 4 9011158743      C5       9  1990-11-15 <NA>    PLOO          PH      <NA>
## 5 9011158743      C5       9  1990-11-15 <NA>    PLOO    SALINITY      <NA>
```

```
## 6 9011158743      C5       9  1990-11-15 <NA>    PLOO        TEMP      <NA>
##    value   units
## 1  0.870    ug/L
## 2 23.855 sigma-t
## 3  6.550    mg/L
## 4  8.080      pH
## 5 33.617     ppt
## 6 19.430       C
```

```
owt_df02_na <- sapply(owt_df02, function(x) sum(is.na(x)))
owt_df02_notna <- sapply(owt_df02, function(x) sum(!is.na(x)))
owt_df02_tbl01 <- rbind(owt_df02_notna, owt_df02_na)
owt_df02_tbl02 <- rbind(owt_df02_tbl01, round(prop.table(owt_df02_tbl01, margin = 2), 4))
print("All parameters")
```

```
## [1] "All parameters"
```

```
print(owt_df02_tbl02)
```

```
##                 sample station     depth_m date_sample       time project
## owt_df02_notna 1236769 1236769 1152608.000     1236769 1075929.00 1236769
## owt_df02_na          0       0   84161.000           0  160840.00       0
## owt_df02_notna       1       1       0.932           1       0.87       1
## owt_df02_na          0       0       0.068           0       0.13       0
##                parameter   qualifier        value   units
## owt_df02_notna   1236769 394867.0000 1231466.0000 1236769
## owt_df02_na            0 841902.0000    5303.0000       0
## owt_df02_notna         1      0.3193       0.9957       1
## owt_df02_na            0      0.6807       0.0043       0
```

```
owt_df02a <- owt_df02[which(is.na(owt_df02), arr.ind=TRUE), ]
#print(head(owt_df02a))

for(p in param_lst01) {
  df = owt_df02[owt_df02$parameter == p, ]
  print(head(df[which(is.na(df$value), arr.ind=TRUE), ]))
  df_na <- sapply(df, function(x) sum(is.na(x)))
  df_notna <- sapply(df, function(x) sum(!is.na(x)))
  df_tbl01 <- rbind(df_notna, df_na)
  df_tbl02 <- rbind(df_tbl01, round(prop.table(df_tbl01, margin = 2), 4))
  rownames(df_tbl02) <- c("Not NA n", "NA n", "Not NA %", "NA %")
  print(p)
  print(df_tbl02) # This table w/ null counts and proportions for each feature not
  ↪  displayed for space purposes
}
```

```
##            sample station depth_m date_sample time project   parameter
## 272260 1229972635     B10     1.5  1997-12-29 <NA>    PLOO CHLOROPHYLL
## 272267 1229972638     B10    42.7  1997-12-29 <NA>    PLOO CHLOROPHYLL
## 272274 1229972633     B10    61.0  1997-12-29 <NA>    PLOO CHLOROPHYLL
## 272281 1229972634     B10    79.2  1997-12-29 <NA>    PLOO CHLOROPHYLL
## 272288 1229972637     B10    97.5  1997-12-29 <NA>    PLOO CHLOROPHYLL
## 272295 1229972636     B10   115.8  1997-12-29 <NA>    PLOO CHLOROPHYLL
##        qualifier value units
## 272260      <NA>    NA  ug/L
```

```
## 272267      <NA>    NA  ug/L
## 272274      <NA>    NA  ug/L
## 272281      <NA>    NA  ug/L
## 272288      <NA>    NA  ug/L
## 272295      <NA>    NA  ug/L
## [1] "CHLOROPHYLL"
##          sample station depth_m date_sample       time project parameter
## Not NA n  88471   88471   88471       88471 74093.0000   88471     88471
## NA n          0       0       0           0 14378.0000       0         0
## Not NA %      1       1       1           1     0.8375       1         1
## NA %          0       0       0           0     0.1625       0         0
##          qualifier      value units
## Not NA n         0 87911.0000 88471
## NA n         88471   560.0000     0
## Not NA %         0     0.9937     1
## NA %             1     0.0063     0
##            sample station depth_m date_sample         time project parameter
## 429235 1018012132     E16     1.5  2001-10-18  8:32:00 PST    PLOO   DENSITY
## 503878  513032821      C8     1.0  2003-05-13 10:30:00 PST    PLOO   DENSITY
## 547108  601042705      C4     1.0  2004-06-01 11:37:00 PST    PLOO   DENSITY
## 720423  702082492      A6     1.0  2008-07-02  9:42:00 PST    PLOO   DENSITY
## 720433  702082490      A6    12.0  2008-07-02  9:42:00 PST    PLOO   DENSITY
## 720443  702082491      A6    18.0  2008-07-02  9:42:00 PST    PLOO   DENSITY
##        qualifier value   units
## 429235      <NA>    NA sigma-t
## 503878      <NA>    NA sigma-t
## 547108      <NA>    NA sigma-t
## 720423      <NA>    NA sigma-t
## 720433      <NA>    NA sigma-t
## 720443      <NA>    NA sigma-t
## [1] "DENSITY"
##          sample station depth_m date_sample       time project parameter
## Not NA n  88317   88317   88317       88317 73999.0000   88317     88317
## NA n          0       0       0           0 14318.0000       0         0
## Not NA %      1       1       1           1     0.8379       1         1
## NA %          0       0       0           0     0.1621       0         0
##          qualifier      value units
## Not NA n         0 88256.0000 88317
## NA n         88317    61.0000     0
## Not NA %         0     0.9993     1
## NA %             1     0.0007     0
##            sample station depth_m date_sample         time project parameter
## 82185   CTD014940     A11    42.7  1993-05-06 10:45:00 PST    PLOO        DO
## 168703 1024942994      A1     1.5  1994-10-24  9:10:00 PST    PLOO        DO
## 168713 1024942995      A1     3.0  1994-10-24  9:10:00 PST    PLOO        DO
## 168723 1024942996      A1     6.1  1994-10-24  9:10:00 PST    PLOO        DO
## 168733 1024942997      A1    12.2  1994-10-24  9:10:00 PST    PLOO        DO
## 168734 1024942998      A1    12.2  1994-10-24  9:10:00 PST    PLOO        DO
##        qualifier value units
## 82185       <NA>    NA  mg/L
## 168703      <NA>    NA  mg/L
## 168713      <NA>    NA  mg/L
## 168723      <NA>    NA  mg/L
## 168733      <NA>    NA  mg/L
```

```
## 168734      <NA>    NA  mg/L
## [1] "DO"
##          sample station depth_m date_sample       time project parameter
## Not NA n 109542  109542  109542      109542 83392.0000  109542    109542
## NA n          0       0       0           0 26150.0000       0         0
## Not NA %      1       1       1           1     0.7613       1         1
## NA %          0       0       0           0     0.2387       0         0
##          qualifier      value  units
## Not NA n         0 109437.000 109542
## NA n        109542    105.000      0
## Not NA %         0      0.999      1
## NA %             1      0.001      0
##            sample station depth_m date_sample        time project parameter
## 25057    MO001828      A2     1.0  1991-10-09 8:27:00 PST    PLOO    ENTERO
## 81650    SH000903      D5      NA  1993-05-04        <NA>    PLOO    ENTERO
## 103629 1208931380      A2     3.0  1993-12-08 8:20:00 PST    PLOO    ENTERO
## 103642 1208931381      A2     6.1  1993-12-08 8:20:00 PST    PLOO    ENTERO
## 103655 1208931382      A2    12.2  1993-12-08 8:20:00 PST    PLOO    ENTERO
## 103668 1208931383      A2    18.3  1993-12-08 8:20:00 PST    PLOO    ENTERO
##        qualifier value      units
## 25057          e    NA CFU/100 mL
## 81650       <NA>    NA CFU/100 mL
## 103629      <NA>    NA CFU/100 mL
## 103642      <NA>    NA CFU/100 mL
## 103655      <NA>    NA CFU/100 mL
## 103668      <NA>    NA CFU/100 mL
## [1] "ENTERO"
##          sample station     depth_m date_sample        time project parameter
## Not NA n 144341  144341 116186.0000      144341 144012.0000  144341    144341
## NA n          0       0  28155.0000           0    329.0000       0         0
## Not NA %      1       1      0.8049           1      0.9977       1         1
## NA %          0       0      0.1951           0      0.0023       0         0
##           qualifier       value  units
## Not NA n 136251.000 143075.0000 144341
## NA n       8090.000   1266.0000      0
## Not NA %      0.944      0.9912      1
## NA %          0.056      0.0088      0
##            sample station depth_m date_sample        time project parameter
## 39760    MO003378      B3    61.0  1992-03-04 9:23:00 PST    PLOO     FECAL
## 73538    MO002093      A4    79.2  1993-02-03 9:15:00 PST    PLOO     FECAL
## 81651    SH000903      D5      NA  1993-05-04        <NA>    PLOO     FECAL
## 82671    SH001257      D7      NA  1993-05-11 9:43:00 PST    PLOO     FECAL
## 103630 1208931380      A2     3.0  1993-12-08 8:20:00 PST    PLOO     FECAL
## 103643 1208931381      A2     6.1  1993-12-08 8:20:00 PST    PLOO     FECAL
##        qualifier value      units
## 39760          <    NA CFU/100 mL
## 73538          <    NA CFU/100 mL
## 81651       <NA>    NA CFU/100 mL
## 82671       <NA>    NA CFU/100 mL
## 103630      <NA>    NA CFU/100 mL
## 103643      <NA>    NA CFU/100 mL
## [1] "FECAL"
##          sample station     depth_m date_sample        time project parameter
## Not NA n 137649  137649 109633.0000      137649 137322.0000  137649    137649
```

```
## NA n          0       0  28016.0000           0    327.0000       0         0
## Not NA %      1       1      0.7965           1      0.9976       1         1
## NA %          0       0      0.2035           0      0.0024       0         0
##            qualifier      value  units
## Not NA n 127994.0000 136405.000 137649
## NA n       9655.0000   1244.000      0
## Not NA %      0.9299      0.991      1
## NA %          0.0701      0.009      0
##           sample station depth_m date_sample         time project parameter
## 112401 112941961      A1    12.2  1994-01-12  8:36:00 PST    PLOO        OG
## 125289 309941665      A4    79.2  1994-03-09 10:06:00 PST    PLOO        OG
## 136304 503941704      B5     1.5  1994-05-03 10:24:00 PST    PLOO        OG
## 136346 503941705      B5    24.4  1994-05-03 10:24:00 PST    PLOO        OG
## 136396 503941706      B5    61.0  1994-05-03 10:24:00 PST    PLOO        OG
## 136772 504941833      A2    24.4  1994-05-04  8:02:00 PST    PLOO        OG
##        qualifier value units
## 112401      <NA>    NA  mg/L
## 125289      <NA>    NA  mg/L
## 136304      <NA>    NA  mg/L
## 136346      <NA>    NA  mg/L
## 136396      <NA>    NA  mg/L
## 136772      <NA>    NA  mg/L
## [1] "OG"
##          sample station depth_m date_sample      time project parameter
## Not NA n   7944    7944    7944        7944 7940.0000    7944      7944
## NA n          0       0       0           0    4.0000       0         0
## Not NA %      1       1       1           1    0.9995       1         1
## NA %          0       0       0           0    0.0005       0         0
##          qualifier     value units
## Not NA n 7673.0000 7922.0000  7944
## NA n      271.0000   22.0000     0
## Not NA %    0.9659    0.9972     1
## NA %        0.0341    0.0028     0
##           sample station depth_m date_sample        time project parameter
## 168706 1024942994      A1     1.5  1994-10-24 9:10:00 PST    PLOO        PH
## 168716 1024942995      A1     3.0  1994-10-24 9:10:00 PST    PLOO        PH
## 168726 1024942996      A1     6.1  1994-10-24 9:10:00 PST    PLOO        PH
## 168737 1024942997      A1    12.2  1994-10-24 9:10:00 PST    PLOO        PH
## 168738 1024942998      A1    12.2  1994-10-24 9:10:00 PST    PLOO        PH
## 168751 1024942999      A1    18.3  1994-10-24 9:10:00 PST    PLOO        PH
##        qualifier value units
## 168706      <NA>    NA    pH
## 168716      <NA>    NA    pH
## 168726      <NA>    NA    pH
## 168737      <NA>    NA    pH
## 168738      <NA>    NA    pH
## 168751      <NA>    NA    pH
## [1] "PH"
##          sample station depth_m date_sample       time project parameter
## Not NA n 107818  107818  107818      107818 81668.0000  107818    107818
## NA n          0       0       0           0 26150.0000       0         0
## Not NA %      1       1       1           1     0.7575       1         1
## NA %          0       0       0           0     0.2425       0         0
##          qualifier       value  units
```

```
## Not NA n         0 107564.0000 107818
## NA n        107818    254.0000      0
## Not NA %         0      0.9976      1
## NA %             1      0.0024      0
##           sample station depth_m date_sample         time project parameter
## 82188   CTD014940     A11    42.7  1993-05-06 10:45:00 PST    PLOO  SALINITY
## 168707 1024942994      A1     1.5  1994-10-24  9:10:00 PST    PLOO  SALINITY
## 168717 1024942995      A1     3.0  1994-10-24  9:10:00 PST    PLOO  SALINITY
## 168727 1024942996      A1     6.1  1994-10-24  9:10:00 PST    PLOO  SALINITY
## 168739 1024942997      A1    12.2  1994-10-24  9:10:00 PST    PLOO  SALINITY
## 168740 1024942998      A1    12.2  1994-10-24  9:10:00 PST    PLOO  SALINITY
##        qualifier value units
## 82188       <NA>    NA   ppt
## 168707      <NA>    NA   ppt
## 168717      <NA>    NA   ppt
## 168727      <NA>    NA   ppt
## 168739      <NA>    NA   ppt
## 168740      <NA>    NA   ppt
## [1] "SALINITY"
##          sample station depth_m date_sample       time project parameter
## Not NA n 109492  109492  109492      109492 83351.0000  109492    109492
## NA n          0       0       0           0 26141.0000       0         0
## Not NA %      1       1       1           1     0.7613       1         1
## NA %          0       0       0           0     0.2387       0         0
##          qualifier       value  units
## Not NA n         0 109318.0000 109492
## NA n        109492    174.0000      0
## Not NA %         0      0.9984      1
## NA %             1      0.0016      0
##           sample station depth_m date_sample         time project parameter
## 182188 119951890      B1    61.0  1995-01-19  7:25:00 PST    PLOO      SUSO
## 197840 413951394      B9    97.5  1995-04-13 11:32:00 PST    PLOO      SUSO
## 236137 821961617      B9     1.5  1996-08-21  8:23:00 PST    PLOO      SUSO
## 236146 821961618      B9    42.7  1996-08-21  8:23:00 PST    PLOO      SUSO
## 236171 821961619      B9    97.5  1996-08-21  8:23:00 PST    PLOO      SUSO
## 236208 821961620     E10     1.5  1996-08-21 10:58:00 PST    PLOO      SUSO
##        qualifier value units
## 182188      <NA>    NA  mg/L
## 197840      <NA>    NA  mg/L
## 236137      <NA>    NA  mg/L
## 236146      <NA>    NA  mg/L
## 236171      <NA>    NA  mg/L
## 236208      <NA>    NA  mg/L
## [1] "SUSO"
##          sample station depth_m date_sample       time project parameter
## Not NA n  27543   27543   27543       27543 27538.0000   27543     27543
## NA n          0       0       0           0     5.0000       0         0
## Not NA %      1       1       1           1     0.9998       1         1
## NA %          0       0       0           0     0.0002       0         0
##           qualifier     value units
## Not NA n  1290.0000 27515.000 27543
## NA n     26253.0000    28.000     0
## Not NA %     0.0468     0.999     1
## NA %         0.9532     0.001     0
```

```
##          sample station depth_m date_sample         time project parameter
## 81483 CTD013369      K4    18.3  1993-05-03  7:07:00 PST    PLOO      TEMP
## 82189 CTD014940     A11    42.7  1993-05-06 10:45:00 PST    PLOO      TEMP
## 82454 CTD013374      K4    18.3  1993-05-06  8:52:00 PST    PLOO      TEMP
## 82871 CTD013384      K4    18.3  1993-05-12  7:36:00 PST    PLOO      TEMP
## 83293 CTD008245      C6     1.5  1993-05-19 12:10:00 PST    PLOO      TEMP
## 83297 CTD008246      C6     3.0  1993-05-19 12:10:00 PST    PLOO      TEMP
##       qualifier value units
## 81483      <NA>    NA     C
## 82189      <NA>    NA     C
## 82454      <NA>    NA     C
## 82871      <NA>    NA     C
## 83293      <NA>    NA     C
## 83297      <NA>    NA     C
## [1] "TEMP"
##          sample station depth_m date_sample        time project parameter
## Not NA n 139066  139066  139066      139066 112709.0000  139066    139066
## NA n          0       0       0           0  26357.0000       0         0
## Not NA %      1       1       1           1      0.8105       1         1
## NA %          0       0       0           0      0.1895       0         0
##          qualifier       value  units
## Not NA n         0 139046.0000 139066
## NA n        139066     20.0000      0
## Not NA %         0      0.9999      1
## NA %             1      0.0001      0
##            sample station depth_m date_sample        time project parameter
## 81652    SH000903      D5      NA  1993-05-04        <NA>    PLOO     TOTAL
## 82672    SH001257      D7      NA  1993-05-11 9:43:00 PST    PLOO     TOTAL
## 101335 1124931032      C6     1.5  1993-11-24 8:23:00 PST    PLOO     TOTAL
## 103637 1208931380      A2     3.0  1993-12-08 8:20:00 PST    PLOO     TOTAL
## 103650 1208931381      A2     6.1  1993-12-08 8:20:00 PST    PLOO     TOTAL
## 103663 1208931382      A2    12.2  1993-12-08 8:20:00 PST    PLOO     TOTAL
##        qualifier value      units
## 81652       <NA>    NA CFU/100 mL
## 82672       <NA>    NA CFU/100 mL
## 101335      <NA>    NA CFU/100 mL
## 103637      <NA>    NA CFU/100 mL
## 103650      <NA>    NA CFU/100 mL
## 103663      <NA>    NA CFU/100 mL
## [1] "TOTAL"
##          sample station     depth_m date_sample        time project parameter
## Not NA n 137584  137584 109594.0000      137584 137257.0000  137584    137584
## NA n          0       0  27990.0000           0    327.0000       0         0
## Not NA %      1       1      0.7966           1      0.9976       1         1
## NA %          0       0      0.2034           0      0.0024       0         0
##            qualifier       value  units
## Not NA n 121659.0000 136141.0000 137584
## NA n      15925.0000   1443.0000      0
## Not NA %      0.8843      0.9895      1
## NA %          0.1157      0.0105      0
##          sample station depth_m date_sample         time project parameter
## 81485 CTD013369      K4    18.3  1993-05-03  7:07:00 PST    PLOO       XMS
## 82191 CTD014940     A11    42.7  1993-05-06 10:45:00 PST    PLOO       XMS
## 82456 CTD013374      K4    18.3  1993-05-06  8:52:00 PST    PLOO       XMS
```

```
## 82873 CTD013384      K4    18.3  1993-05-12  7:36:00 PST    PLOO       XMS
## 83497 CTD013399      K4    18.3  1993-05-20  7:20:00 PST    PLOO       XMS
## 83564 CTD004913      A7     1.5  1993-05-24  8:31:00 PST    PLOO       XMS
##       qualifier value units
## 81485      <NA>    NA     %
## 82191      <NA>    NA     %
## 82456      <NA>    NA     %
## 82873      <NA>    NA     %
## 83497      <NA>    NA     %
## 83564      <NA>    NA     %
## [1] "XMS"
##          sample station depth_m date_sample        time project parameter
## Not NA n 139002  139002  139002      139002 112648.0000  139002    139002
## NA n          0       0       0           0  26354.0000       0         0
## Not NA %      1       1       1           1      0.8104       1         1
## NA %          0       0       0           0      0.1896       0         0
##          qualifier       value  units
## Not NA n         0 138876.0000 139002
## NA n        139002    126.0000      0
## Not NA %         0      0.9991      1
## NA %             1      0.0009      0
```

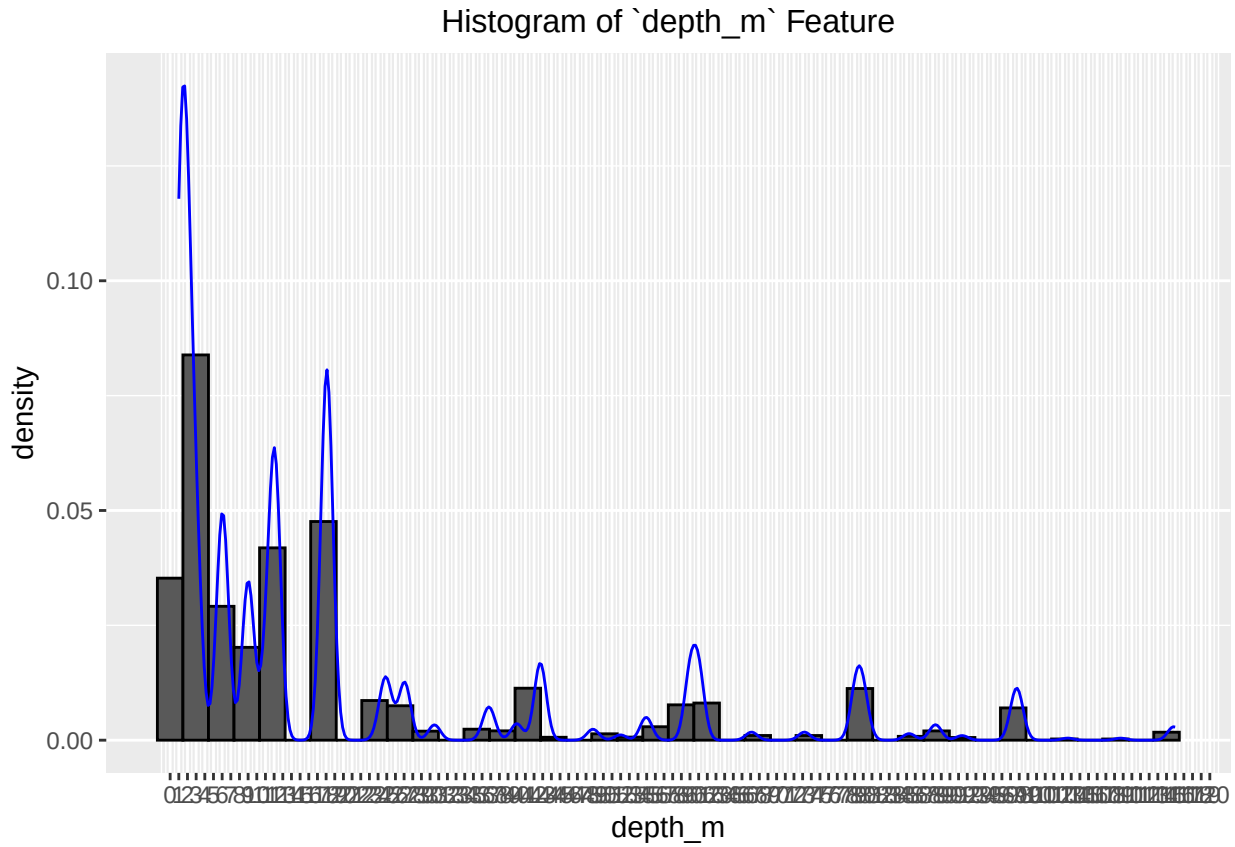
```
# Display value summaries
# Citation: Dave Hurst
params <- owt_df02 %>%
  mutate(unit_def = paste(parameter, qualifier, units)) %>%
  count(unit_def)
print(params)
```

```
##                unit_def      n
## 1   CHLOROPHYLL NA ug/L  88471
## 2    DENSITY NA sigma-t  88317
## 3            DO NA mg/L 109542
## 4   ENTERO < CFU/100 mL  95848
## 5   ENTERO > CFU/100 mL    452
## 6   ENTERO e CFU/100 mL  39766
## 7  ENTERO LA CFU/100 mL     23
## 8  ENTERO NA CFU/100 mL   8090
## 9  ENTERO ND CFU/100 mL      1
## 10 ENTERO NR CFU/100 mL      1
## 11 ENTERO NS CFU/100 mL    160
## 12   FECAL < CFU/100 mL  85819
## 13   FECAL > CFU/100 mL    746
## 14   FECAL e CFU/100 mL  41255
## 15  FECAL LA CFU/100 mL      8
## 16  FECAL NA CFU/100 mL   9655
## 17  FECAL ND CFU/100 mL      1
## 18  FECAL NR CFU/100 mL      4
## 19  FECAL NS CFU/100 mL    161
## 20            OG < mg/L   7673
## 21           OG NA mg/L    271
## 22             PH NA pH 107818
## 23      SALINITY NA ppt 109492
## 24          SUSO < mg/L   1290
```

```
## 25         SUSO NA mg/L  26253
## 26            TEMP NA C 139066
## 27   TOTAL < CFU/100 mL  66655
## 28   TOTAL > CFU/100 mL   2844
## 29  TOTAL >= CFU/100 mL      3
## 30   TOTAL e CFU/100 mL  51924
## 31  TOTAL LA CFU/100 mL     14
## 32  TOTAL NA CFU/100 mL  15925
## 33  TOTAL ND CFU/100 mL     54
## 34  TOTAL NR CFU/100 mL      4
## 35  TOTAL NS CFU/100 mL    161
## 36             XMS NA % 139002
```

Bin `depth_m` variable

```
owt_df02$date_sample <- as.Date(owt_df02$date_sample, "%Y-%m-%d")

# Create bins for depth_m values
depth_lvls01 <- c("[0,8)",
                  "[8,33)",
                  "[33,47)",
                  "[47,70)",
                  "[70,90)",
                  "[90,112)",
                  "[112,120]",
                  "Unknown")

# Plot distribution of depth_m values
#Citation:
↪  https://community.rstudio.com/t/ggplot-x-axis-y-axis-ticks-labels-breaks-and-limits/119123/2
ggplot(owt_df02, aes(x = depth_m)) +
  geom_histogram(color = "black", bins = 40, aes(y = stat(density))) +
  geom_density(col = "blue") +
  labs(title = "Histogram of `depth_m` Feature") +
  scale_x_continuous(breaks=seq(0,120,1)) +
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```

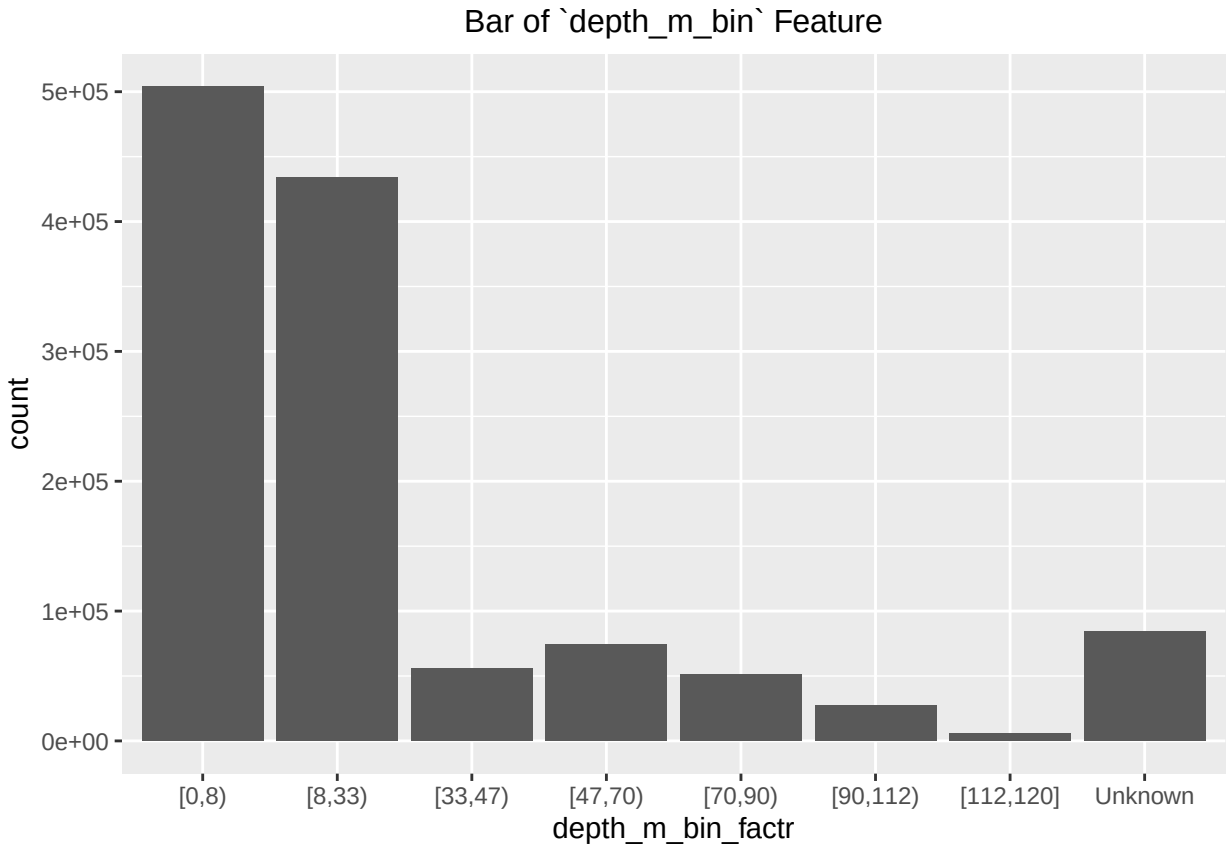



```
# Create new column with bins
# Citation:
↪  https://www.marsja.se/r-add-column-to-dataframe-based-on-other-columns-conditions-dplyr/
owt_df02 <- mutate(owt_df02, depth_m_bin = case_when(depth_m < 8 ~ "[0,8)",
                                                     depth_m < 33 ~ "[8,33)",
                                                     depth_m < 47 ~ "[33,47)",
                                                     depth_m < 70 ~ "[47,70)",
                                                     depth_m < 90 ~ "[70,90)",
                                                     depth_m < 112 ~ "[90,112)",
                                                     depth_m >= 112 ~ "[112,120]"))

# Replace NAs with "Unknown"
# Citation: https://statisticsglobe.com/r-replace-na-with-0/
owt_df02$depth_m_bin <- replace_na(owt_df02$depth_m_bin, "Unknown")

# Citation: https://www.statology.org/order-bars-ggplot2-bar-chart/
owt_df02$depth_m_bin_factr = factor(owt_df02$depth_m_bin, levels = depth_lvls01)
print(head(owt_df02))
```

```
##       sample station depth_m date_sample time project   parameter qualifier
## 1 9011158743      C5       9  1990-11-15 <NA>    PLOO CHLOROPHYLL      <NA>
## 2 9011158743      C5       9  1990-11-15 <NA>    PLOO     DENSITY      <NA>
## 3 9011158743      C5       9  1990-11-15 <NA>    PLOO          DO      <NA>
## 4 9011158743      C5       9  1990-11-15 <NA>    PLOO          PH      <NA>
## 5 9011158743      C5       9  1990-11-15 <NA>    PLOO    SALINITY      <NA>
## 6 9011158743      C5       9  1990-11-15 <NA>    PLOO        TEMP      <NA>
```

```
##    value   units depth_m_bin depth_m_bin_factr
## 1  0.870    ug/L      [8,33)            [8,33)
## 2 23.855 sigma-t      [8,33)            [8,33)
## 3  6.550    mg/L      [8,33)            [8,33)
## 4  8.080      pH      [8,33)            [8,33)
## 5 33.617     ppt      [8,33)            [8,33)
## 6 19.430       C      [8,33)            [8,33)
```

```
# Display transformed bar chart
ggplot(owt_df02, aes(x = depth_m_bin_factr)) +
  geom_bar() +
  labs(title = "Bar of `depth_m_bin` Feature") +
  theme(plot.title = element_text(hjust = 0.5, size = 12))
```



Perform several aggregations on the data for performing EDA at multiple levels

```
# Display aggregations by different features
owt_df02_gb01 <- owt_df02 %>%
  group_by(station) %>%
  summarise(Count = n())

print(owt_df02_gb01[owt_df02_gb01$Count == min(owt_df02_gb01$Count), ])
```

```
## # A tibble: 2 x 2
##   station Count
##   <fct>   <int>
```

```
## 1 A15        30
## 2 A16        30
```

```
print(owt_df02_gb01[owt_df02_gb01$Count == max(owt_df02_gb01$Count), ])
```

```
## # A tibble: 1 x 2
##   station Count
##   <fct>   <int>
## 1 A1      55217
```

```
owt_df02_gb02 <- owt_df02 %>%
  group_by(project) %>%
  summarise(Count = n())

owt_df02_gb03 <- owt_df02 %>%
  group_by(date_sample) %>%
  summarise(Count = n())

# Main DF1
owt_df02_gb04 <- owt_df02 %>%
  group_by(date_sample, parameter) %>%
  summarise(Avg = mean(value, na.remove = TRUE))
```

```
## `summarise()` has grouped output by 'date_sample'. You can override using the
## `.groups` argument.
```

```
owt_df02_gb05 <- owt_df02 %>%
  group_by(parameter) %>%
  summarise(Count = n())

owt_df02_gb06 <- owt_df02 %>%
  group_by(depth_m) %>%
  summarise(Count = n())

# Main DF3
owt_df02_gb07 <- owt_df02 %>%
  group_by(date_sample, project, depth_m_bin, parameter) %>%
  summarise(Avg = mean(value, na.remove = TRUE))
```

```
## `summarise()` has grouped output by 'date_sample', 'project', 'depth_m_bin'.
## You can override using the `.groups` argument.
```

```
owt_df02_gb08 <- owt_df02 %>%
  group_by(depth_m_bin) %>%
  summarise(Count = n())

# Main DF2
owt_df02_gb09 <- owt_df02 %>%
  group_by(date_sample, project, parameter) %>%
  summarise(Avg = mean(value, na.remove = TRUE))
```

```
## `summarise()` has grouped output by 'date_sample', 'project'. You can override
## using the `.groups` argument.
```

```
print(owt_df02_gb01)
```

```
## # A tibble: 157 x 2
##    station Count
##    <fct>   <int>
##  1 A1      55217
##  2 A10      5852
##  3 A11      6602
##  4 A12      5812
##  5 A13      6599
##  6 A14      5886
##  7 A15        30
##  8 A16        30
##  9 A17      2961
## 10 A2       7572
## # ... with 147 more rows
```

```
print(dim(owt_df02_gb01))
```

```
## [1] 157   2
```

```
print(owt_df02_gb02)
```

```
## # A tibble: 2 x 2
##   project  Count
##   <fct>    <int>
## 1 PLOO    805915
## 2 SBOO    430854
```

```
print(dim(owt_df02_gb02))
```

```
## [1] 2 2
```

```
print(head(owt_df02_gb03))
```

```
## # A tibble: 6 x 2
##   date_sample Count
##   <date>      <int>
## 1 1990-11-15     14
## 2 1991-01-02    195
## 3 1991-01-03    195
## 4 1991-01-07    190
## 5 1991-01-08    181
## 6 1991-01-09    577
```

```
print(dim(owt_df02_gb03))
```

```
## [1] 5580    2
```

```
print(head(owt_df02_gb04))
```

```
## # A tibble: 6 x 3
## # Groups:   date_sample [1]
##   date_sample parameter     Avg
##   <date>      <fct>       <dbl>
## 1 1990-11-15  CHLOROPHYLL  1.07
## 2 1990-11-15  DENSITY     23.9
## 3 1990-11-15  DO           6.98
## 4 1990-11-15  PH           8.13
```

```
## 5 1990-11-15  SALINITY    33.6
## 6 1990-11-15  TEMP        19.4
```

```
print(dim(owt_df02_gb04))
```

```
## [1] 36811     3
```

```
print(owt_df02_gb05)
```

```
## # A tibble: 12 x 2
##    parameter    Count
##    <fct>        <int>
##  1 CHLOROPHYLL  88471
##  2 DENSITY      88317
##  3 DO          109542
##  4 ENTERO      144341
##  5 FECAL       137649
##  6 OG            7944
##  7 PH          107818
##  8 SALINITY    109492
##  9 SUSO         27543
## 10 TEMP        139066
## 11 TOTAL       137584
## 12 XMS         139002
```

```
print(dim(owt_df02_gb05))
```

```
## [1] 12  2
```

```
print(owt_df02_gb06)
```

```
## # A tibble: 51 x 2
##    depth_m  Count
##      <dbl>  <int>
##  1     1   119858
##  2     1.5  93937
##  3     2   120007
##  4     3    71058
##  5     6    55762
##  6     6.1  43265
##  7     9    65616
##  8     9.1   3052
##  9    11    23069
## 10    12    80308
## # ... with 41 more rows
```

```
print(dim(owt_df02_gb06))
```

```
## [1] 51  2
```

```
print(head(owt_df02_gb07))
```

```
## # A tibble: 6 x 5
## # Groups:   date_sample, project, depth_m_bin [1]
##   date_sample project depth_m_bin parameter     Avg
##   <date>      <fct>   <chr>       <fct>       <dbl>
## 1 1990-11-15  PLOO    [8,33)      CHLOROPHYLL  0.87
```

```
## 2 1990-11-15  PLOO    [8,33)      DENSITY     23.9
## 3 1990-11-15  PLOO    [8,33)      DO           6.55
## 4 1990-11-15  PLOO    [8,33)      PH           8.08
## 5 1990-11-15  PLOO    [8,33)      SALINITY    33.6
## 6 1990-11-15  PLOO    [8,33)      TEMP        19.4
```

```
print(dim(owt_df02_gb07))
```

```
## [1] 121286      5
```

```
print(owt_df02_gb08)
```

```
## # A tibble: 8 x 2
##   depth_m_bin  Count
##   <chr>        <int>
## 1 [0,8)       503887
## 2 [112,120]     5908
## 3 [33,47)      55629
## 4 [47,70)      74036
## 5 [70,90)      51358
## 6 [8,33)      434169
## 7 [90,112)     27621
## 8 Unknown      84161
```

```
print(dim(owt_df02_gb08))
```

```
## [1] 8 2
```

```
print(head(owt_df02_gb09))
```

```
## # A tibble: 6 x 4
## # Groups:   date_sample, project [1]
##   date_sample project parameter     Avg
##   <date>      <fct>   <fct>       <dbl>
## 1 1990-11-15  PLOO    CHLOROPHYLL  0.87
## 2 1990-11-15  PLOO    DENSITY     23.9
## 3 1990-11-15  PLOO    DO           6.55
## 4 1990-11-15  PLOO    PH           8.08
## 5 1990-11-15  PLOO    SALINITY    33.6
## 6 1990-11-15  PLOO    TEMP        19.4
```

```
print(dim(owt_df02_gb09))
```

```
## [1] 48395     4
```

Create custom function to df mutation/casting and boxplot display

```
ts_eda <- function(ts_df = NA,
                   form_lead = c(),
                   form_cast = c(),
                   param_lst = c(),
                   l1_col = c(),
                   l1_param = c(),
                   l2_col = c(),
                   l2_param = c(),
                   rtn_met = TRUE,
```

```
                grph_title = "All",
                box = TRUE) {
# Define function to dcast, print boxplot, and print desc stats for each parameter
dcast_form <- as.formula(paste(paste(form_lead, collapse = "+"), form_cast, sep = "~"))
#print(dcast_form)
#print(head(ts_df))

counter <- 1
# Loop to dcast parameters to cols
for(p in param_lst) {
  sym_param <- sym(p)
  print(sym_param)
  param_df <- ts_df[ts_df[, form_cast] == p, ]
  param_df <- param_df %>%
    drop_na()
  #print(param_df)

  # Citation: https://www.datasciencemadesimple.com/melting-casting-r/
  param_df_cast <- dcast(param_df, dcast_form, mean)
  #print(param_df_cast)

  if(counter  == 1) {
    ts_df_mrgd <- param_df_cast
  }
  else {
    # Merge individual casted parameter df's into 1 df
    # Citation: https://www.geeksforgeeks.org/joining-of-dataframes-in-r-programming/
    ts_df_mrgd <- merge(x = ts_df_mrgd,
                        y = param_df_cast,
                        by = form_lead,
                        all = TRUE)
  }

  # Run custom function to ID outliers and generate boxplot
  #ts_df_mrgd01a <- subset(x = ts_df_mrgd, select = num_var_lst01)
  ts_df_mrgd01a <- na.omit(param_df_cast)
  #print(head(ts_df_mrgd01a))
  dfs1 <- subset(x = ts_df_mrgd01a, select = p)
  dfs1a <- dfs1
  dfs1a$Groups <- grph_title
  box_comp(xcol = p,
           df = dfs1,
           rtn_met = rtn_met,
           box = box)
  #dfs_lst <- list(dfs1)
  #print(dfs_lst)

  #_____________________________________________________________________
  # Create subplots based on additional agg layers
  print(paste0("L2 length = ", length(l2_param)))
  if(length(l2_param) == 0) {
    for(l1 in l1_param) {
      ts_df_mrgd_chl_dmb1 <- ts_df_mrgd01a[(ts_df_mrgd01a[, l1_col] == l1), ]
```

```
        # Run custom function to ID outliers and generate boxplot
        #print(head(ts_df_mrgd_chl_dmb1))
        # Citation: https://www.geeksforgeeks.org/handling-errors-in-r-programming/
        dfs2 <- subset(x = ts_df_mrgd_chl_dmb1, select = p)
        dfs2a <- dfs2
        try(dfs2a$Groups <- l1)
        #append(dfs_lst, dfs2)
        try(box_comp(xcol = p,
                     df = dfs2,
                     rtn_met = rtn_met,
                     grph_title = l1,
                     box = box),
            silent = TRUE)
        dfs1a <- rbind(dfs1a, dfs2a)
      }
    }
    else {
      for(l1 in l1_param) {
        for(l2 in l2_param) {
          ts_df_mrgd_chl_dmb1 <- ts_df_mrgd01a[(ts_df_mrgd01a[, l1_col] == l1) &
↪   (ts_df_mrgd01a[, l2_col] == l2), ]

          # Run custom function to ID outliers and generate boxplot
          #print(head(ts_df_mrgd_chl_dmb1))
          # Citation: https://www.geeksforgeeks.org/handling-errors-in-r-programming/
          dfs3 <- subset(x = ts_df_mrgd_chl_dmb1, select = p)
          dfs3a <- dfs3
          try(dfs3a$Groups <- paste0(l1, ": ", l2))
          #append(dfs_lst, dfs3)
          try(box_comp(xcol = p,
                       df = dfs3,
                       rtn_met = rtn_met,
                       grph_title = paste0(l1, ": ", l2),
                       box = box),
              silent = TRUE)
          dfs1a <- rbind(dfs1a, dfs3a)
        }
      }
    }
    #-------------------------------------------------------------------------
    #print(dim(dfs1a))
    #print(head(dfs1a))
    bx_form <- as.formula(paste(p, "Groups", sep = "~"))
    #print(bx_form)
    boxplot(bx_form,
            dfs1a,
            main = "Boxplot(s)")

    counter <- counter + 1
  }
  print(head(ts_df_mrgd, 10))
  print(describe(ts_df_mrgd))
```

```
  # Display summary of NA count
  ts_df_na <- sapply(ts_df, function(x) sum(is.na(x)))
  ts_df_notna <- sapply(ts_df, function(x) sum(!is.na(x)))
  ts_df_tbl01 <- rbind(ts_df_notna, ts_df_na)
  ts_df_tbl02 <- rbind(ts_df_tbl01, round(prop.table(ts_df_tbl01, margin = 2), 4))
  rownames(ts_df_tbl02) <- c("Not NA n", "NA n", "Not NA %", "NA %")
  print(ts_df_tbl02) # This table w/ null counts and proportions for each feature not
  ↪  displayed for space purposes
  return(ts_df_mrgd)
}
```

Run custom function on data aggregated by `date_sample` and `parameter`

```
param_lst02 <- c("CHLOROPHYLL",
                 "DENSITY",
                 "DO",
                 "ENTERO",
                 "FECAL",
                 "OG",
                 "PH",
                 "SALINITY",
                 "SUSO",
                 "TEMP",
                 "XMS")

owt_df02_gb04_mrgd = ts_eda(ts_df = owt_df02_gb04,
                            form_lead = "date_sample",
                            form_cast = "parameter",
                            param_lst = param_lst02,
                            rtn_met = TRUE,
                            box = FALSE
                            )
```

```
## CHLOROPHYLL
```

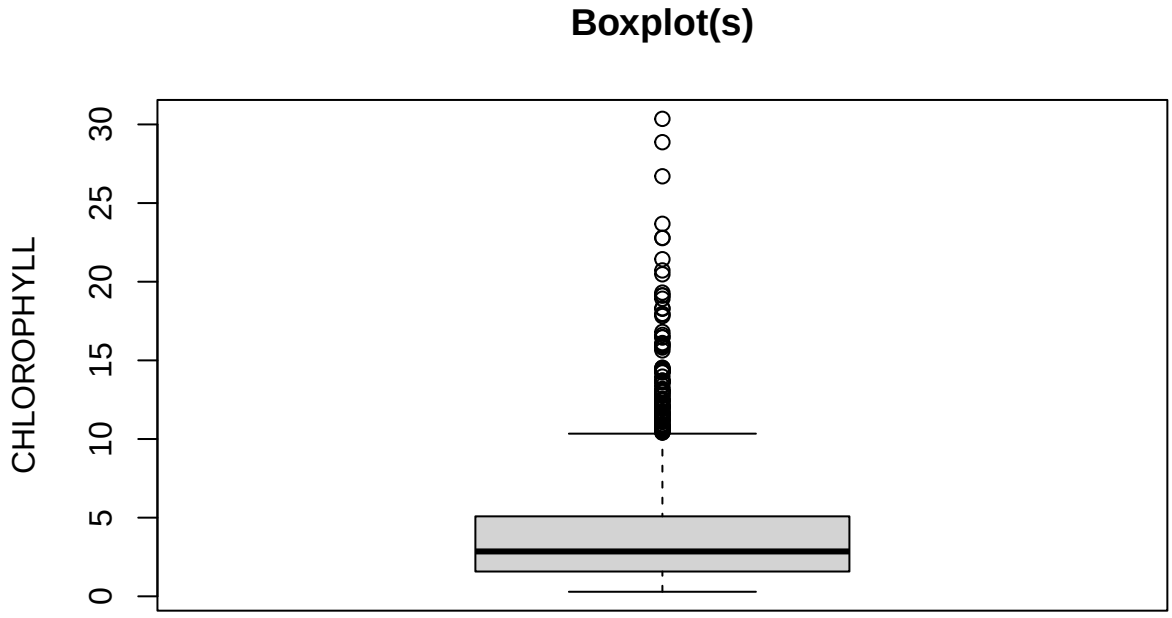
```
## Using Avg as value column: use value.var to override.
```

```
##                  metric                    V2
## 1                       Variable: CHLOROPHYLL
## 2              Total N:
## 3                 Count                  2092
## 4              NA Count                     0
## 5                  Mean                  3.91
## 6                Median                 2.853
## 7    Standard Deviation                 3.411
## 8              Variance                11.633
## 9                 Range                30.064
## 10                  Min                  0.29
## 11                  Max                30.353
## 12      25th Percentile                 1.576
## 13      75th Percentile                 5.083
## 14 Subset w/o Outliers:
## 15                Count                  2050
```

```
## 16                    %                  98%
## 17            Outlier %                   2%
## 18             NA Count                    0
## 19                 Mean                3.626
## 20               Median                2.791
## 21   Standard Deviation                 2.74
## 22             Variance                7.507
## 23                Range               13.676
## 24                  Min                0.29
## 25                  Max               13.965
## [1] "L2 length = 0"
```

```
## DENSITY
```

```
## Using Avg as value column: use value.var to override.
```

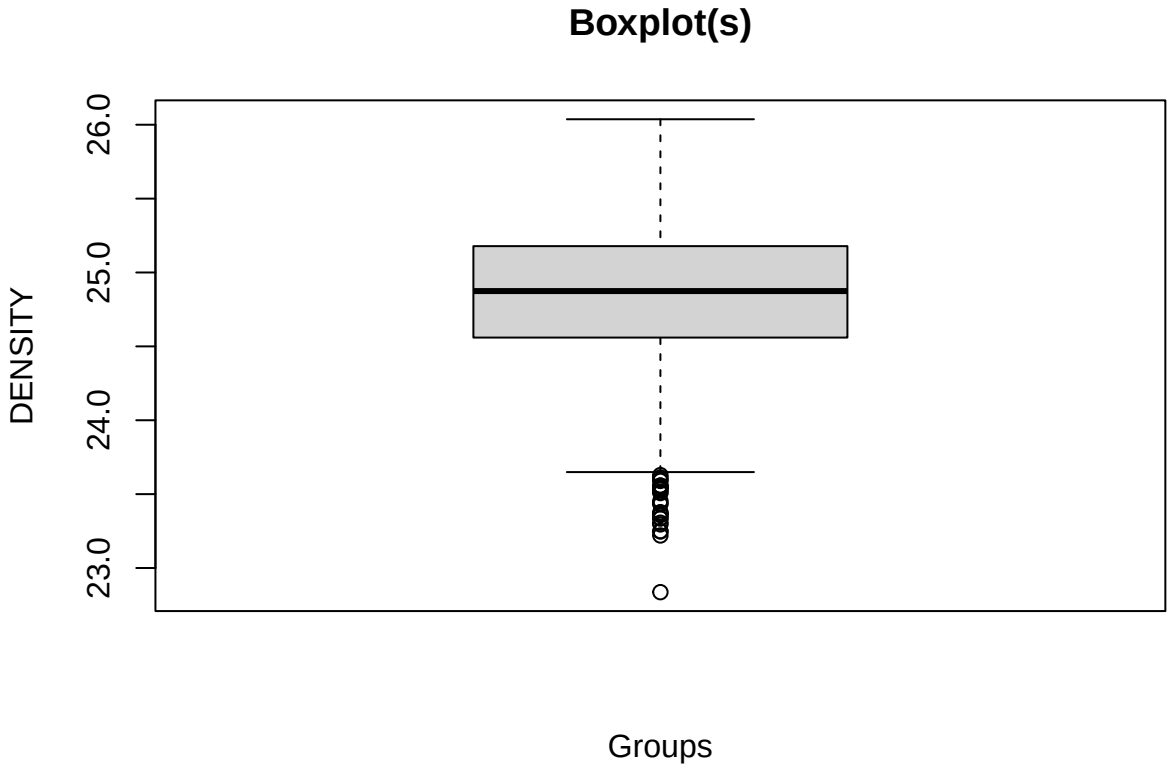



```
##                 metric                V2
## 1                      Variable: DENSITY
## 2             Total N:
## 3                Count              2246
## 4             NA Count                 0
## 5                 Mean             24.85
## 6               Median            24.874
## 7   Standard Deviation             0.466
## 8             Variance             0.217
## 9                Range               3.2
## 10                 Min            22.837
```

```
## 11                  Max           26.037
## 12      25th Percentile            24.56
## 13      75th Percentile           25.178
## 14 Subset w/o Outliers:
## 15                Count            2232
## 16                    %           99.4%
## 17            Outlier %            0.6%
## 18             NA Count               0
## 19                 Mean           24.86
## 20               Median           24.877
## 21   Standard Deviation            0.451
## 22             Variance            0.203
## 23                Range            2.583
## 24                  Min           23.454
## 25                  Max           26.037
## [1] "L2 length = 0"
```

```
## DO
```

```
## Using Avg as value column: use value.var to override.
```

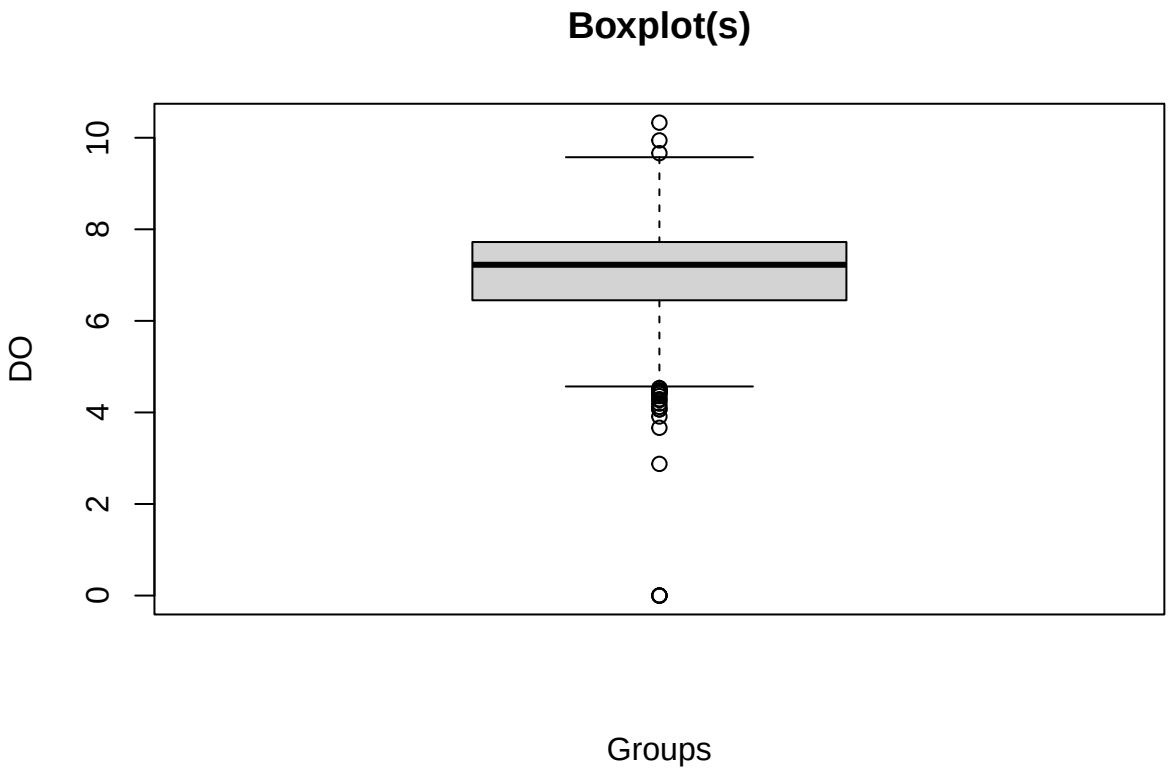



```
##                 metric          V2
## 1                      Variable: DO
## 2            Total N:
## 3               Count        2504
## 4            NA Count           0
## 5                Mean       7.041
```

```
## 6                Median        7.226
## 7    Standard Deviation         0.99
## 8              Variance         0.98
## 9                 Range       10.329
## 10                  Min            0
## 11                  Max       10.329
## 12      25th Percentile        6.451
## 13      75th Percentile        7.722
## 14 Subset w/o Outliers:
## 15                Count         2494
## 16                    %        99.6%
## 17            Outlier %         0.4%
## 18             NA Count            0
## 19                 Mean        7.057
## 20               Median        7.232
## 21   Standard Deviation        0.937
## 22             Variance        0.877
## 23                Range        5.822
## 24                  Min         4.12
## 25                  Max        9.942
## [1] "L2 length = 0"
```

```
## ENTERO
```

```
## Using Avg as value column: use value.var to override.
```

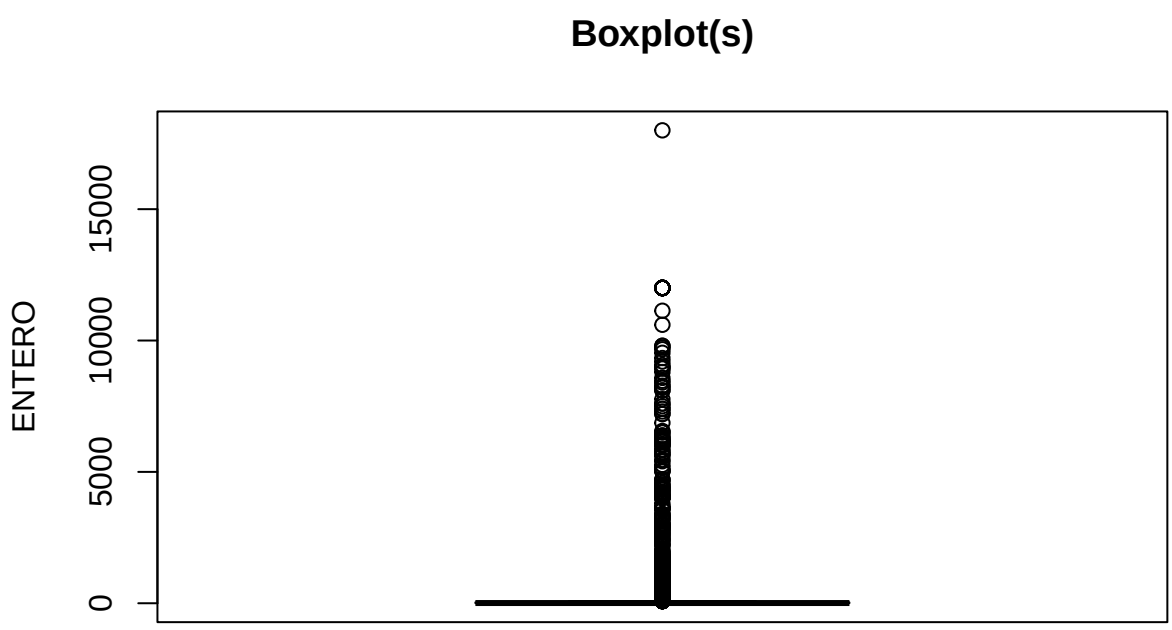



```
##                 metric             V2
```

```
## 1                       Variable: ENTERO
## 2              Total N:
## 3                 Count             5197
## 4              NA Count                0
## 5                  Mean          284.213
## 6                Median            7.333
## 7    Standard Deviation         1325.881
## 8              Variance      1757960.571
## 9                 Range            18000
## 10                  Min                0
## 11                  Max            18000
## 12      25th Percentile             2.75
## 13      75th Percentile           35.455
## 14 Subset w/o Outliers:
## 15                Count             5090
## 16                    %            97.9%
## 17            Outlier %             2.1%
## 18             NA Count                0
## 19                 Mean          112.006
## 20               Median                7
## 21   Standard Deviation          417.536
## 22             Variance       174336.299
## 23                Range         4233.333
## 24                  Min                0
## 25                  Max         4233.333
## [1] "L2 length = 0"
```

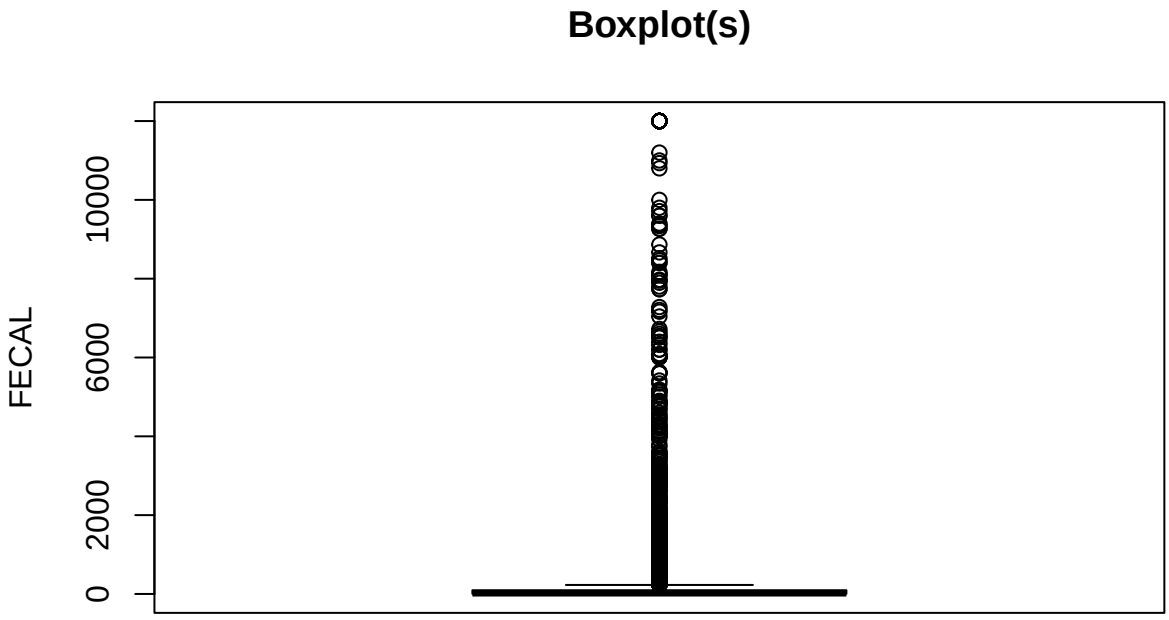
```
## FECAL
```

```
## Using Avg as value column: use value.var to override.
```



	metric	V2
1		Variable: FECAL
2	Total N:	
3	Count	5078
4	NA Count	0
5	Mean	409.934
6	Median	10.564
7	Standard Deviation	1513.552
8	Variance	2290838.653
9	Range	12000
10	Min	0
11	Max	12000
12	25th Percentile	3.404
13	75th Percentile	94.095
14	Subset w/o Outliers:	
15	Count	4965
16	%	97.8%
17	Outlier %	2.2%
18	NA Count	0
19	Mean	208.248
20	Median	10
21	Standard Deviation	600.094
22	Variance	360113.198
23	Range	4900
24	Min	0
25	Max	4900

```
## [1] "L2 length = 0"
## OG
## Using Avg as value column: use value.var to override.
```

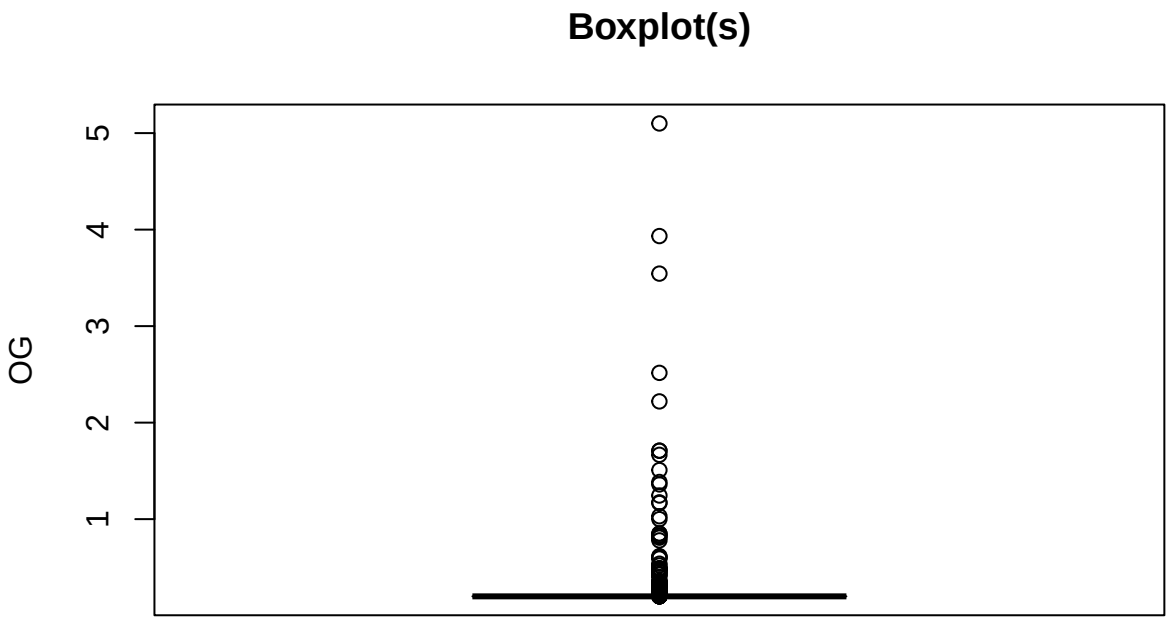


	metric	V2
1		Variable: OG
2	Total N:	
3	Count	937
4	NA Count	0
5	Mean	0.244
6	Median	0.2
7	Standard Deviation	0.286
8	Variance	0.082
9	Range	4.9
10	Min	0.2
11	Max	5.1
12	25th Percentile	0.2
13	75th Percentile	0.2
14	Subset w/o Outliers:	
15	Count	923
16	%	98.5%
17	Outlier %	1.5%
18	NA Count	0
19	Mean	0.215
20	Median	0.2

```
## 21   Standard Deviation        0.078
## 22              Variance        0.006
## 23                 Range         0.83
## 24                   Min          0.2
## 25                   Max         1.03
## [1] "L2 length = 0"
```

```
## PH
```

```
## Using Avg as value column: use value.var to override.
```

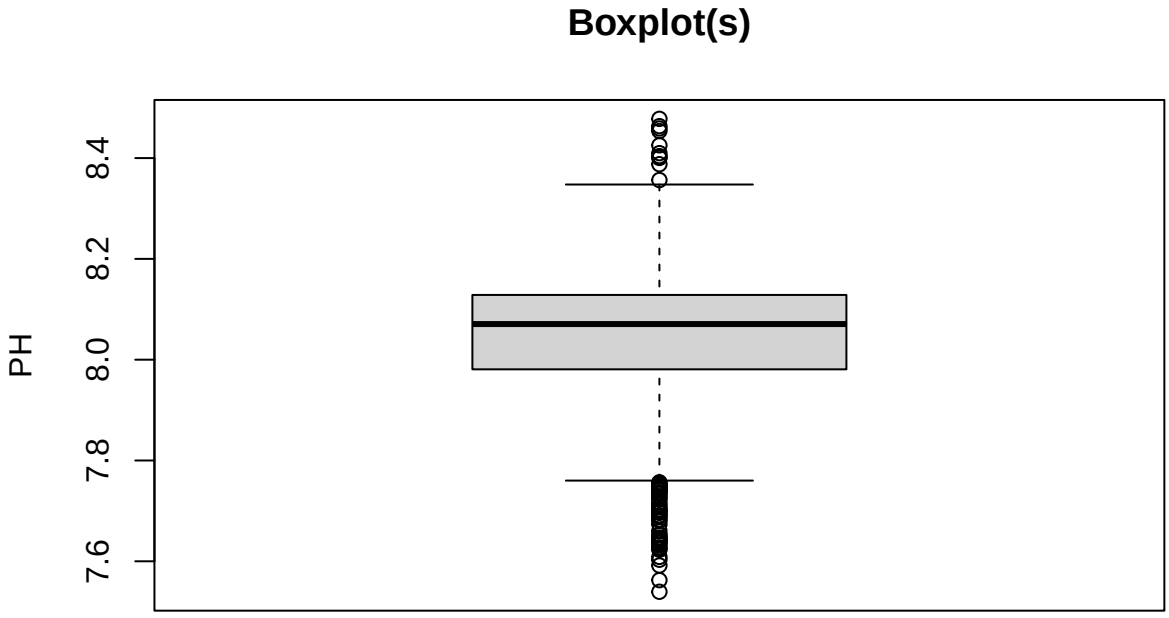



```
##                  metric           V2
## 1                       Variable: PH
## 2             Total N:
## 3                Count         2469
## 4             NA Count            0
## 5                 Mean         8.05
## 6               Median        8.071
## 7    Standard Deviation        0.117
## 8             Variance        0.014
## 9                Range        0.938
## 10                 Min         7.54
## 11                 Max        8.478
## 12      25th Percentile        7.981
## 13      75th Percentile        8.128
## 14 Subset w/o Outliers:
## 15               Count         2435
```

```
## 16                    %        98.6%
## 17           Outlier %         1.4%
## 18            NA Count            0
## 19                Mean        8.053
## 20              Median        8.071
## 21  Standard Deviation        0.108
## 22            Variance        0.012
## 23               Range        0.702
## 24                 Min          7.7
## 25                 Max        8.402
## [1] "L2 length = 0"
```

```
## SALINITY
```

```
## Using Avg as value column: use value.var to override.
```



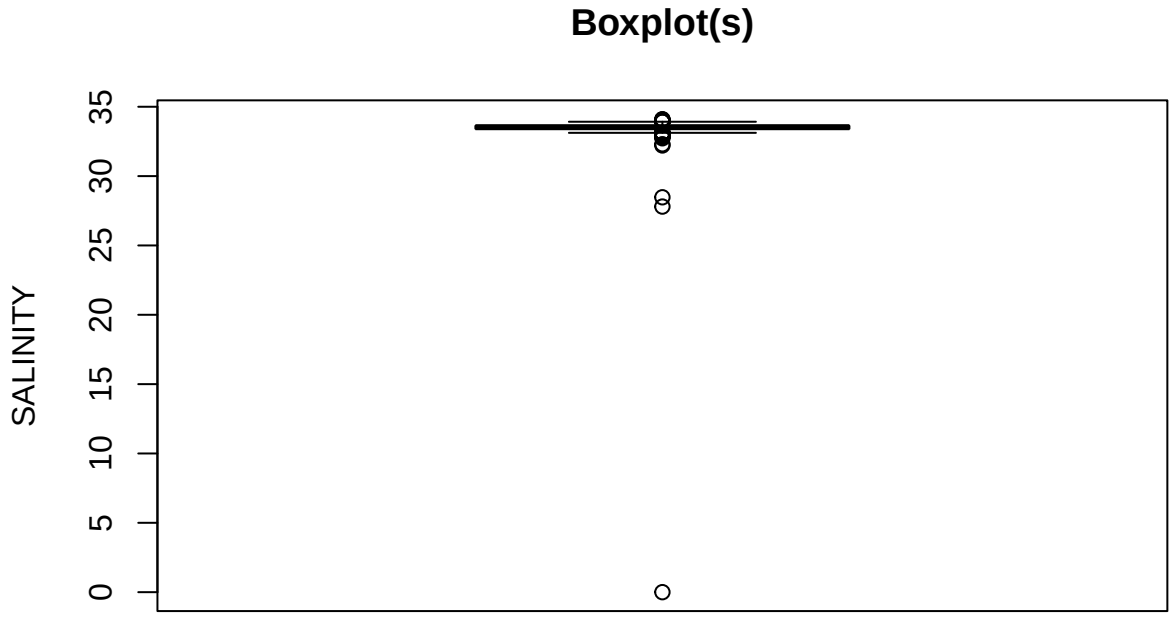
Groups

```
##                 metric                V2
## 1                      Variable: SALINITY
## 2            Total N:
## 3               Count              2505
## 4            NA Count                 0
## 5                Mean            33.502
## 6              Median            33.527
## 7  Standard Deviation             0.706
## 8            Variance             0.499
## 9               Range            34.098
## 10                Min                 0
```

```
## 11                  Max            34.098
## 12      25th Percentile            33.418
## 13      75th Percentile             33.62
## 14 Subset w/o Outliers:
## 15                Count             2502
## 16                    %             99.9%
## 17            Outlier %              0.1%
## 18             NA Count                0
## 19                 Mean            33.519
## 20               Median            33.527
## 21   Standard Deviation             0.165
## 22             Variance             0.027
## 23                Range             1.888
## 24                  Min             32.21
## 25                  Max            34.098
## [1] "L2 length = 0"
```

```
## SUSO
```

```
## Using Avg as value column: use value.var to override.
```

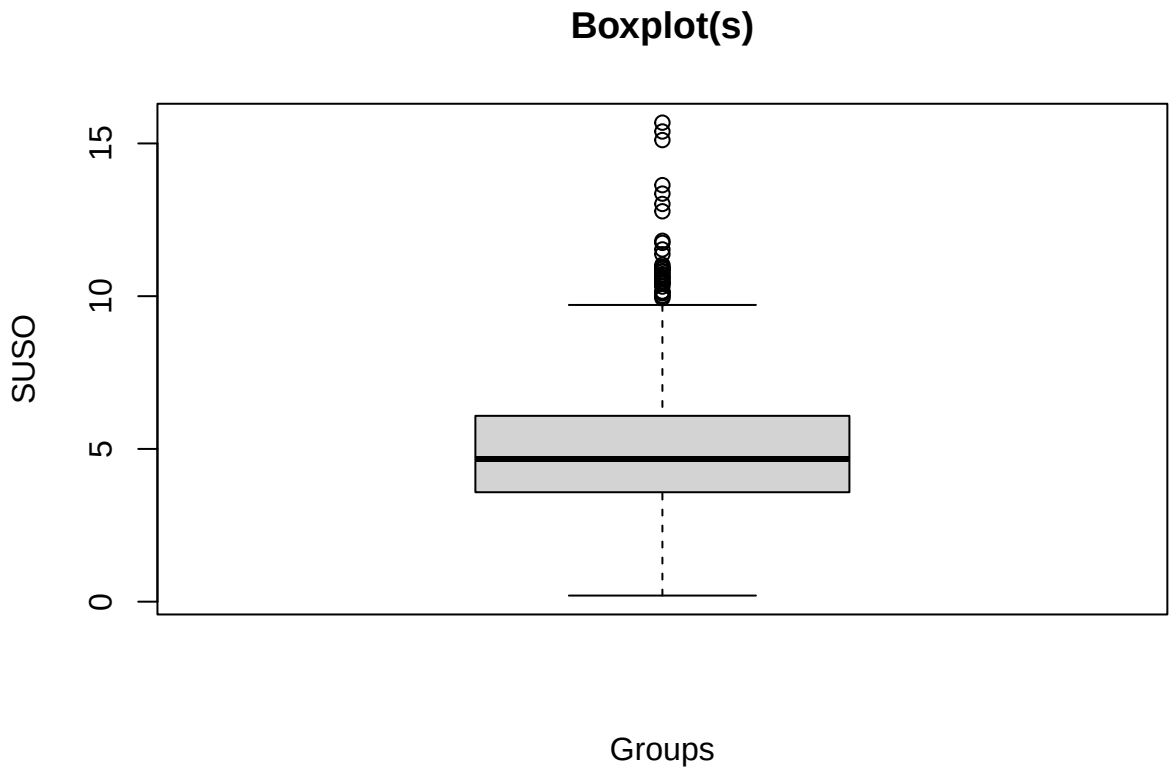



```
##                 metric             V2
## 1                      Variable: SUSO
## 2            Total N:
## 3               Count            976
## 4            NA Count              0
## 5                Mean          4.993
```

```
## 6                Median           4.67
## 7    Standard Deviation          2.164
## 8              Variance          4.681
## 9                 Range          15.48
## 10                  Min            0.2
## 11                  Max          15.68
## 12      25th Percentile          3.583
## 13      75th Percentile          6.083
## 14 Subset w/o Outliers:
## 15                Count            966
## 16                    %            99%
## 17            Outlier %             1%
## 18             NA Count              0
## 19                 Mean          4.906
## 20               Median          4.639
## 21   Standard Deviation          1.991
## 22             Variance          3.966
## 23                Range         11.172
## 24                  Min            0.2
## 25                  Max         11.372
## [1] "L2 length = 0"
```

```
## TEMP
```

```
## Using Avg as value column: use value.var to override.
```

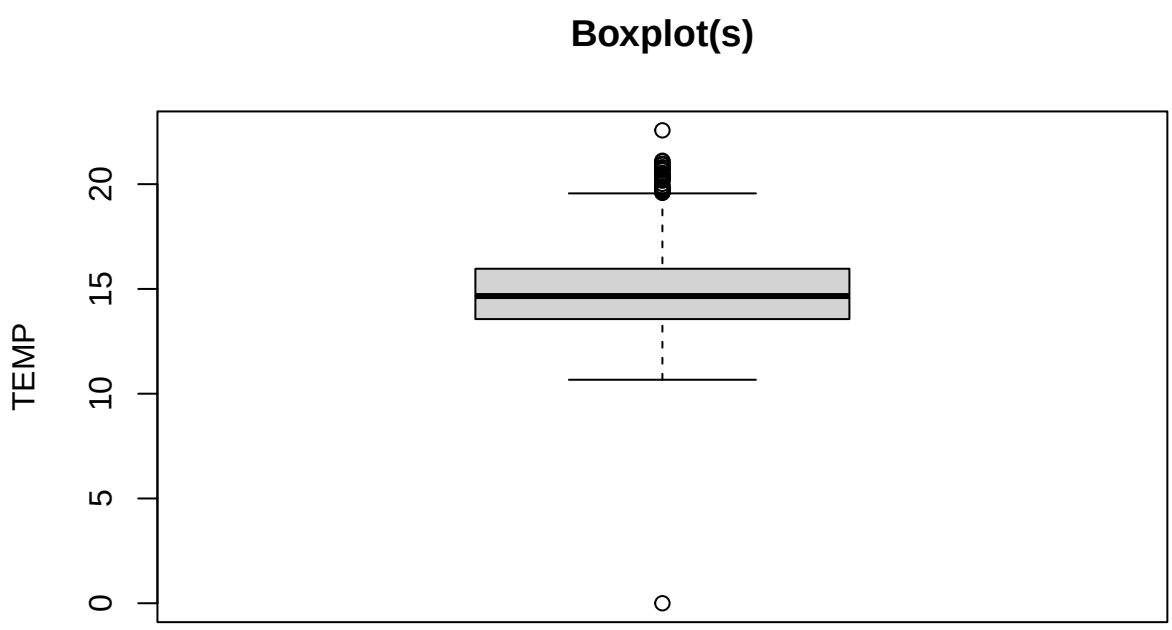



```
##                 metric            V2
```

```
## 1                       Variable: TEMP
## 2               Total N:
## 3                  Count           3373
## 4               NA Count              0
## 5                   Mean         14.843
## 6                 Median         14.661
## 7    Standard Deviation          1.879
## 8               Variance           3.53
## 9                  Range         22.568
## 10                   Min              0
## 11                   Max         22.568
## 12      25th Percentile         13.561
## 13      75th Percentile         15.964
## 14 Subset w/o Outliers:
## 15                 Count           3357
## 16                     %          99.5%
## 17            Outlier %           0.5%
## 18              NA Count              0
## 19                  Mean          14.82
## 20                Median          14.65
## 21   Standard Deviation          1.821
## 22              Variance          3.315
## 23                 Range          9.794
## 24                   Min         10.666
## 25                   Max          20.46
## [1] "L2 length = 0"
```

```
## XMS
```

```
## Using Avg as value column: use value.var to override.
```



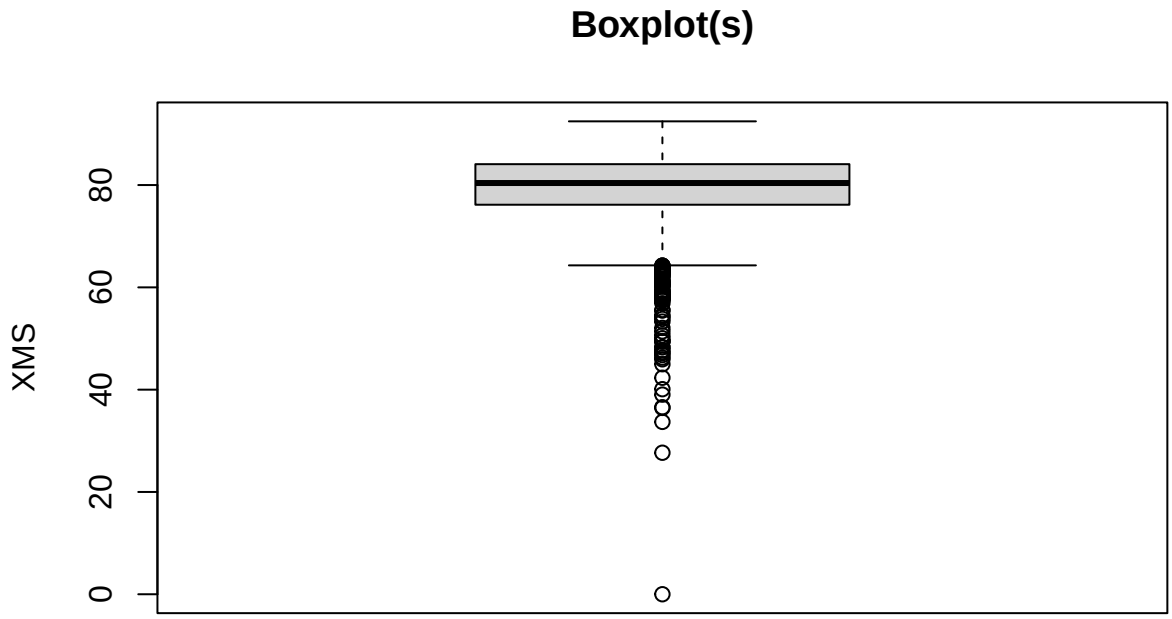
	metric	V2
1		Variable: XMS
2	Total N:	
3	Count	3344
4	NA Count	0
5	Mean	79.402
6	Median	80.4
7	Standard Deviation	6.979
8	Variance	48.713
9	Range	92.456
10	Min	0
11	Max	92.456
12	25th Percentile	76.156
13	75th Percentile	84.084
14	Subset w/o Outliers:	
15	Count	3302
16	%	98.7%
17	Outlier %	1.3%
18	NA Count	0
19	Mean	79.791
20	Median	80.486
21	Standard Deviation	5.988
22	Variance	35.857
23	Range	33.88
24	Min	58.577
25	Max	92.456

```
## [1] "L2 length = 0"
```

```
##    date_sample CHLOROPHYLL DENSITY      DO    ENTERO     FECAL OG    PH
## 1   1990-11-15        1.07  23.854 6.98000        NA        NA NA 8.130
## 2   1991-01-02          NA      NA      NA  32.05128 125.64103 NA    NA
## 3   1991-01-03          NA      NA      NA  18.71795  51.28205 NA    NA
## 4   1991-01-07          NA      NA      NA 106.15385  80.76923 NA    NA
## 5   1991-01-08          NA      NA      NA  40.00000  40.55556 NA    NA
## 6   1991-01-09          NA      NA 5.47625  20.33898  40.50847 NA 8.195
## 7   1991-01-10          NA      NA 5.50000  40.22727  78.18182 NA    NA
## 8   1991-01-11          NA      NA 5.36000  83.92857 446.07143 NA    NA
## 9   1991-01-14          NA      NA      NA  33.84615  95.64103 NA    NA
## 10  1991-01-16          NA      NA      NA  12.56410  23.58974 NA    NA
##    SALINITY SUSO     TEMP      XMS
## 1  33.59550   NA 19.37000 60.22500
## 2        NA   NA 14.50769 80.02564
## 3        NA   NA 14.39231 83.89744
## 4        NA   NA 14.54103 78.97436
## 5        NA   NA 13.43000 80.20000
## 6  33.51938   NA 14.26889 86.21111
## 7  33.49300   NA 13.75000 80.45833
## 8  33.49100   NA 14.38400 84.40000
## 9        NA   NA 14.51282 72.02564
## 10       NA   NA 14.57692 68.35897
```

```
## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning Inf
```

```
## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning -Inf
```




```
##              vars    n   mean      sd median trimmed   mad   min      max
## date_sample     1 5450    NaN      NA     NA     NaN    NA   Inf     -Inf
## CHLOROPHYLL     2 2092   3.91    3.41   2.85    3.33  2.28  0.29    30.35
## DENSITY         3 2246  24.85    0.47  24.87   24.87  0.46 22.84    26.04
## DO              4 2504   7.04    0.99   7.23    7.11  0.86  0.00    10.33
## ENTERO          5 5197 284.21 1325.88   7.33   25.74  7.91  0.00 18000.00
## FECAL           6 5078 409.93 1513.55  10.56   69.45 12.57  0.00 12000.00
## OG              7  937   0.24    0.29   0.20    0.20  0.00  0.20     5.10
## PH              8 2469   8.05    0.12   8.07    8.06  0.10  7.54     8.48
## SALINITY        9 2505  33.50    0.71  33.53   33.52  0.15  0.00    34.10
## SUSO           10  976   4.99    2.16   4.67    4.80  1.80  0.20    15.68
## TEMP           11 3373  14.84    1.88  14.66   14.75  1.78  0.00    22.57
## XMS            12 3344  79.40    6.98  80.40   80.08  5.76  0.00    92.46
##                range   skew kurtosis    se
## date_sample     -Inf     NA       NA    NA
## CHLOROPHYLL    30.06   2.25     7.84  0.07
## DENSITY         3.20  -0.42     0.25  0.01
## DO             10.33  -1.05     3.72  0.02
## ENTERO      18000.00   6.95    53.47 18.39
## FECAL       12000.00   5.81    36.94 21.24
## OG              4.90  10.89   144.21  0.01
## PH              0.94  -0.66     1.32  0.00
## SALINITY       34.10 -42.97  2021.03  0.01
## SUSO           15.48   1.09     2.23  0.07
## TEMP           22.57   0.35     1.15  0.03
## XMS            92.46  -1.86     9.43  0.12
```

```
##          date_sample parameter       Avg
## Not NA n       36811     36811 35669.000
## NA n               0         0  1142.000
## Not NA %           1         1     0.969
## NA %               0         0     0.031
```

Run custom function on data aggregated by `date_sample`, `project` and `parameter`

```
owt_df02_gb09_mrgd = ts_eda(ts_df = owt_df02_gb09,
                            form_lead = c("date_sample", "project"),
                            form_cast = "parameter",
                            param_lst = param_lst02,
                            l1_col = c("project"),
                            l1_param = c("PLOO", "SBOO"),
                            rtn_met = TRUE,
                            box = FALSE
                            )
```

```
## CHLOROPHYLL
```

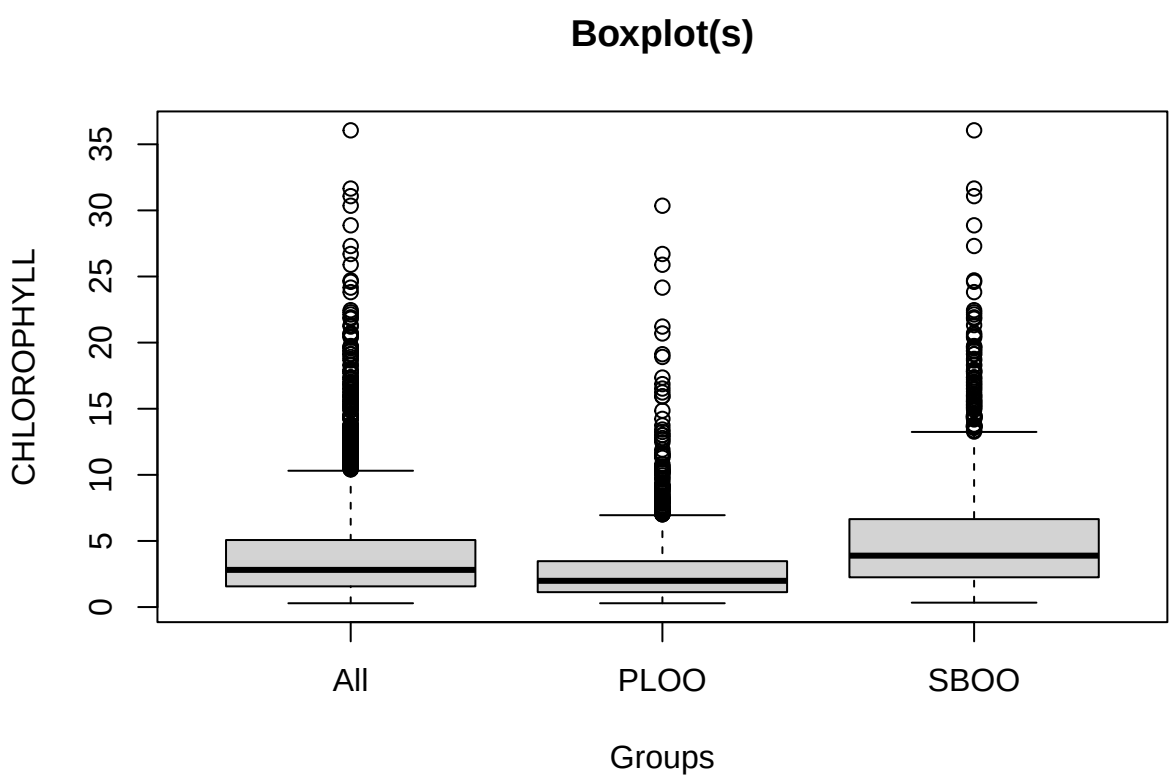
```
## Using Avg as value column: use value.var to override.
```

```
##                   metric                    V2
## 1                        Variable: CHLOROPHYLL
## 2              Total N:
## 3                 Count                  3023
## 4              NA Count                     0
## 5                  Mean                 4.013
## 6                Median                 2.812
## 7    Standard Deviation                 3.812
## 8              Variance                14.529
## 9                 Range                35.764
## 10                  Min                  0.29
## 11                  Max                36.053
## 12      25th Percentile                 1.566
## 13      75th Percentile                  5.07
## 14 Subset w/o Outliers:
## 15                Count                  2956
## 16                    %                 97.8%
## 17            Outlier %                  2.2%
## 18             NA Count                     0
## 19                 Mean                  3.65
## 20               Median                 2.757
## 21   Standard Deviation                 2.911
## 22             Variance                 8.475
## 23                Range                15.089
## 24                  Min                  0.29
## 25                  Max                15.379
## [1] "L2 length = 0"
##                   metric                    V2
## 1                        Variable: CHLOROPHYLL
## 2              Total N:
## 3                 Count                  1518
## 4              NA Count                     0
## 5                  Mean                 2.847
```

```
## 6                Median                 1.985
## 7    Standard Deviation                 2.839
## 8              Variance                 8.059
## 9                 Range                30.064
## 10                  Min                 0.29
## 11                  Max                30.353
## 12      25th Percentile                 1.127
## 13      75th Percentile                 3.472
## 14 Subset w/o Outliers:
## 15                Count                  1488
## 16                    %                   98%
## 17            Outlier %                    2%
## 18             NA Count                     0
## 19                 Mean                 2.576
## 20               Median                 1.956
## 21   Standard Deviation                 1.999
## 22             Variance                 3.998
## 23                Range                    11
## 24                  Min                 0.29
## 25                  Max                11.289
##                  metric                    V2
## 1                       Variable: CHLOROPHYLL
## 2              Total N:
## 3                 Count                  1505
## 4              NA Count                     0
## 5                  Mean                 5.188
## 6                Median                 3.888
## 7    Standard Deviation                 4.279
## 8              Variance                18.311
## 9                 Range                 35.72
## 10                  Min                 0.333
## 11                  Max                36.053
## 12      25th Percentile                 2.252
## 13      75th Percentile                 6.652
## 14 Subset w/o Outliers:
## 15                Count                  1475
## 16                    %                   98%
## 17            Outlier %                    2%
## 18             NA Count                     0
## 19                 Mean                 4.838
## 20               Median                 3.777
## 21   Standard Deviation                 3.485
## 22             Variance                12.148
## 23                Range                17.628
## 24                  Min                 0.333
## 25                  Max                17.961
```

```
## DENSITY
```

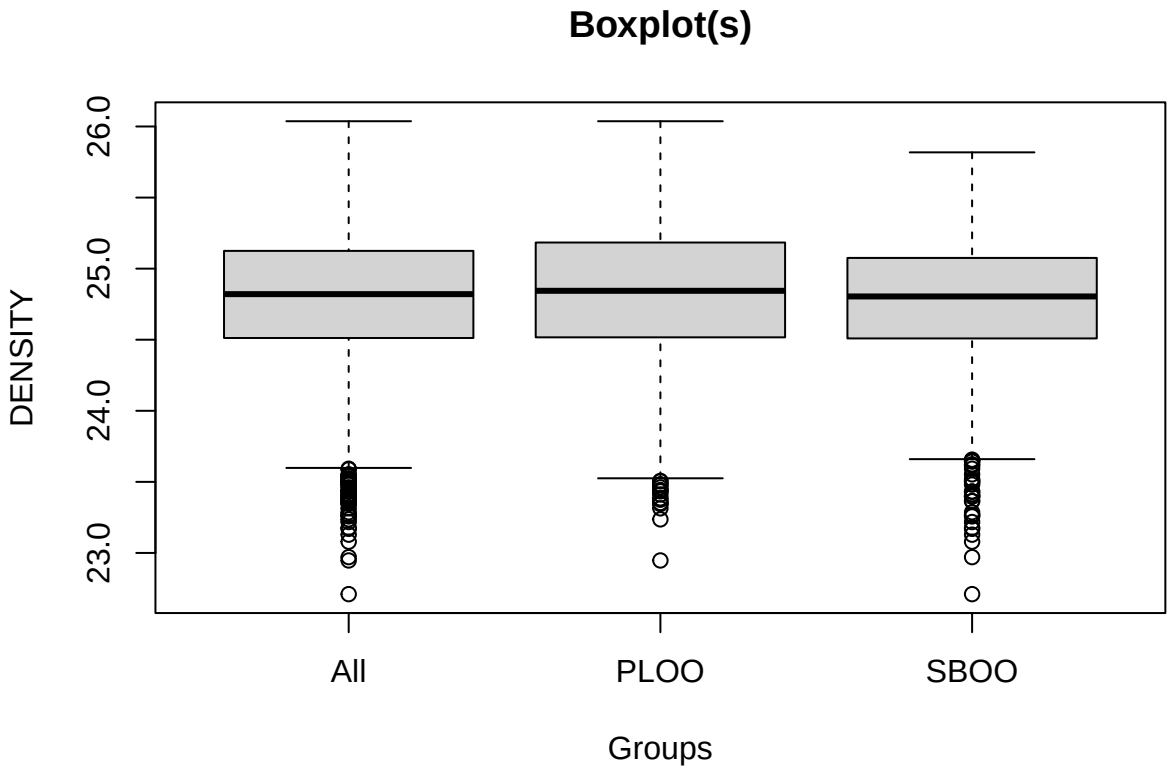
```
## Using Avg as value column: use value.var to override.
```



	metric	V2
1		Variable: DENSITY
2	Total N:	
3	Count	3177
4	NA Count	0
5	Mean	24.794
6	Median	24.821
7	Standard Deviation	0.475
8	Variance	0.225
9	Range	3.327
10	Min	22.71
11	Max	26.037
12	25th Percentile	24.512
13	75th Percentile	25.125
14	Subset w/o Outliers:	
15	Count	3158
16	%	99.4%
17	Outlier %	0.6%
18	NA Count	0
19	Mean	24.804
20	Median	24.825
21	Standard Deviation	0.459
22	Variance	0.211
23	Range	2.663
24	Min	23.374
25	Max	26.037

```
## [1] "L2 length = 0"
##                   metric                V2
## 1                        Variable: DENSITY
## 2              Total N:
## 3                 Count              1613
## 4              NA Count                 0
## 5                  Mean            24.824
## 6                Median            24.844
## 7    Standard Deviation             0.497
## 8              Variance             0.247
## 9                 Range              3.09
## 10                  Min            22.947
## 11                  Max            26.037
## 12      25th Percentile            24.517
## 13      75th Percentile            25.184
## 14 Subset w/o Outliers:
## 15                Count              1610
## 16                    %             99.8%
## 17            Outlier %              0.2%
## 18             NA Count                 0
## 19                 Mean            24.827
## 20               Median            24.845
## 21   Standard Deviation             0.492
## 22             Variance             0.242
## 23                Range             2.693
## 24                  Min            23.344
## 25                  Max            26.037
##                   metric                V2
## 1                        Variable: DENSITY
## 2              Total N:
## 3                 Count              1564
## 4              NA Count                 0
## 5                  Mean            24.764
## 6                Median            24.804
## 7    Standard Deviation             0.448
## 8              Variance             0.201
## 9                 Range             3.108
## 10                  Min             22.71
## 11                  Max            25.818
## 12      25th Percentile            24.509
## 13      75th Percentile            25.075
## 14 Subset w/o Outliers:
## 15                Count              1547
## 16                    %             98.9%
## 17            Outlier %              1.1%
## 18             NA Count                 0
## 19                 Mean            24.781
## 20               Median            24.807
## 21   Standard Deviation              0.42
## 22             Variance             0.177
## 23                Range             2.388
## 24                  Min             23.43
## 25                  Max            25.818
```

```
## DO
## Using Avg as value column: use value.var to override.
```



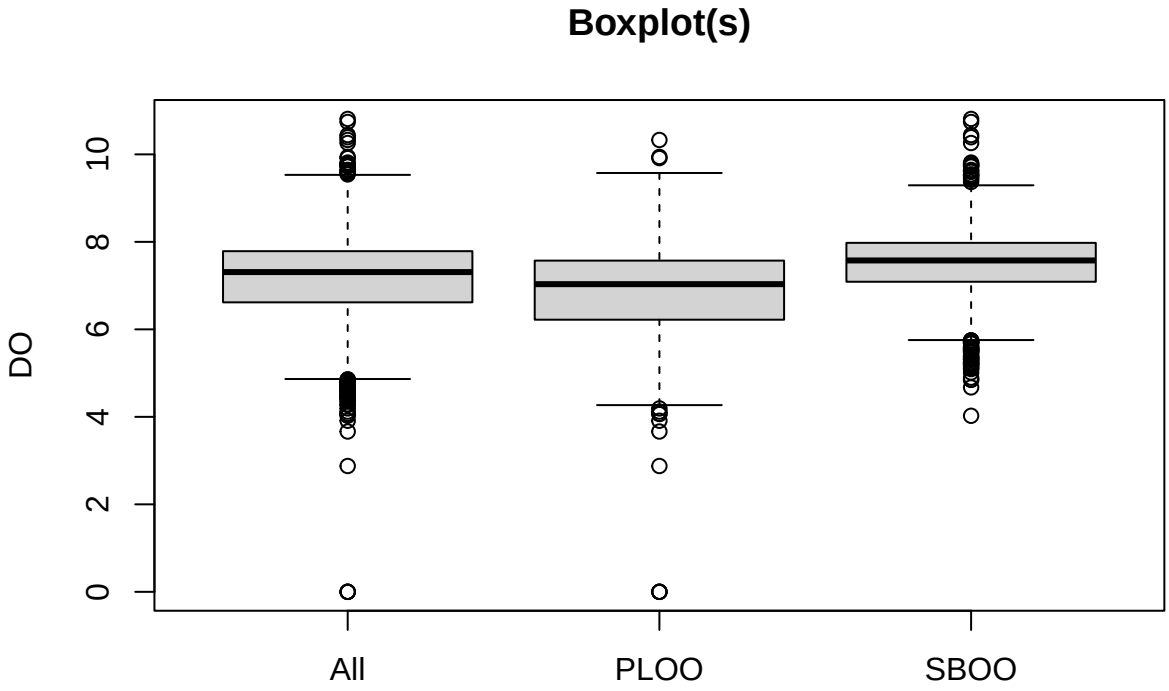
	metric	V2
1		Variable: DO
2	Total N:	
3	Count	3434
4	NA Count	0
5	Mean	7.155
6	Median	7.31
7	Standard Deviation	0.972
8	Variance	0.946
9	Range	10.81
10	Min	0
11	Max	10.81
12	25th Percentile	6.618
13	75th Percentile	7.787
14	Subset w/o Outliers:	
15	Count	3416
16	%	99.5%
17	Outlier %	0.5%
18	NA Count	0
19	Mean	7.165
20	Median	7.311
21	Standard Deviation	0.919
22	Variance	0.845

```
## 23                Range        5.674
## 24                  Min        4.268
## 25                  Max        9.942
## [1] "L2 length = 0"
##                  metric           V2
## 1                       Variable: DO
## 2             Total N:
## 3                 Count         1879
## 4             NA Count            0
## 5                  Mean        6.869
## 6                Median        7.033
## 7    Standard Deviation        1.015
## 8             Variance         1.03
## 9                 Range       10.329
## 10                  Min            0
## 11                  Max       10.329
## 12     25th Percentile         6.22
## 13     75th Percentile        7.571
## 14 Subset w/o Outliers:
## 15                Count         1870
## 16                    %        99.5%
## 17           Outlier %         0.5%
## 18            NA Count            0
## 19                Mean        6.882
## 20              Median        7.034
## 21   Standard Deviation         0.95
## 22             Variance        0.903
## 23                Range        5.664
## 24                  Min        3.911
## 25                  Max        9.575
##                  metric           V2
## 1                       Variable: DO
## 2             Total N:
## 3                 Count         1555
## 4             NA Count            0
## 5                  Mean          7.5
## 6                Median        7.575
## 7    Standard Deviation        0.792
## 8             Variance        0.627
## 9                 Range        6.786
## 10                  Min        4.024
## 11                  Max        10.81
## 12     25th Percentile        7.088
## 13     75th Percentile        7.977
## 14 Subset w/o Outliers:
## 15                Count         1541
## 16                    %        99.1%
## 17           Outlier %         0.9%
## 18            NA Count            0
## 19                Mean        7.505
## 20              Median        7.576
## 21   Standard Deviation        0.749
## 22             Variance        0.561
## 23                Range        4.651
```

```
## 24                  Min        5.162
## 25                  Max        9.813
```

```
## ENTERO
```

```
## Using Avg as value column: use value.var to override.
```



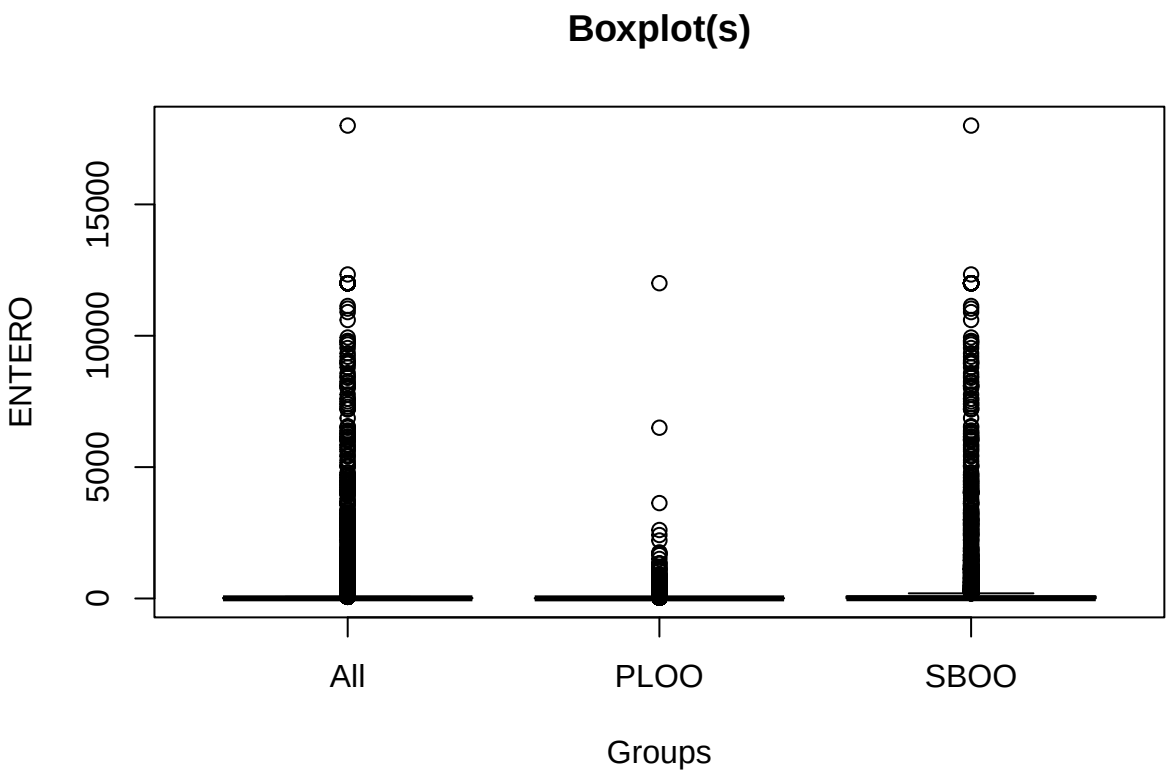

```
##                  metric              V2
## 1                       Variable: ENTERO
## 2              Total N:
## 3                 Count            6776
## 4              NA Count               0
## 5                  Mean         253.636
## 6                Median           5.266
## 7    Standard Deviation        1249.911
## 8              Variance     1562277.114
## 9                 Range           18000
## 10                  Min               0
## 11                  Max           18000
## 12      25th Percentile           2.222
## 13      75th Percentile          25.838
## 14 Subset w/o Outliers:
## 15                Count            6640
## 16                    %              98%
## 17            Outlier %              2%
## 18             NA Count               0
## 19                 Mean          94.427
```

```
## 20               Median               5
## 21   Standard Deviation          359.113
## 22             Variance       128962.481
## 23                Range         4001.333
## 24                  Min                0
## 25                  Max         4001.333
## [1] "L2 length = 0"
##                   metric               V2
## 1                       Variable: ENTERO
## 2              Total N:
## 3                 Count             3784
## 4              NA Count                0
## 5                  Mean           37.018
## 6                Median            4.206
## 7    Standard Deviation          264.127
## 8              Variance        69763.238
## 9                 Range            12000
## 10                  Min                0
## 11                  Max            12000
## 12      25th Percentile            2.222
## 13      75th Percentile           13.236
## 14 Subset w/o Outliers:
## 15                Count             3757
## 16                    %            99.3%
## 17            Outlier %             0.7%
## 18             NA Count                0
## 19                 Mean           22.753
## 20               Median            4.148
## 21   Standard Deviation           69.201
## 22             Variance          4788.75
## 23                Range            824.5
## 24                  Min                0
## 25                  Max            824.5
##                   metric               V2
## 1                       Variable: ENTERO
## 2              Total N:
## 3                 Count             2992
## 4              NA Count                0
## 5                  Mean          527.593
## 6                Median            9.095
## 7    Standard Deviation         1821.011
## 8              Variance      3316080.888
## 9                 Range            17998
## 10                  Min                2
## 11                  Max            18000
## 12      25th Percentile            2.222
## 13      75th Percentile           79.625
## 14 Subset w/o Outliers:
## 15                Count             2903
## 16                    %              97%
## 17            Outlier %               3%
## 18             NA Count                0
## 19                 Mean          245.432
## 20               Median                8
```

```
## 21   Standard Deviation          748.339
## 22              Variance       560011.269
## 23                 Range           5856.4
## 24                   Min                2
## 25                   Max           5858.4
```

```
## FECAL
```

```
## Using Avg as value column: use value.var to override.
```



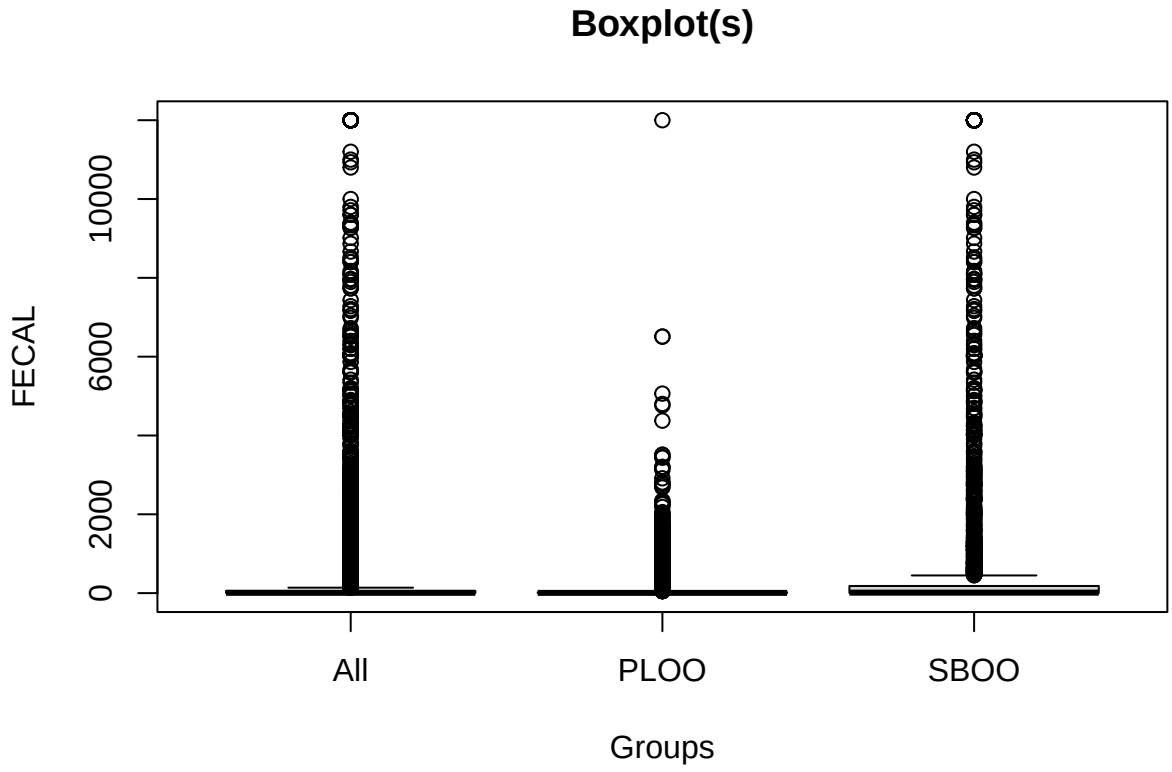

```
##                  metric             V2
## 1                       Variable: FECAL
## 2              Total N:
## 3                 Count           6594
## 4              NA Count              0
## 5                  Mean        355.647
## 6                Median          7.068
## 7    Standard Deviation       1401.818
## 8              Variance    1965093.115
## 9                 Range          12000
## 10                  Min              0
## 11                  Max          12000
## 12      25th Percentile            2.5
## 13      75th Percentile         56.886
## 14 Subset w/o Outliers:
## 15                Count           6455
## 16                    %          97.9%
```

```
## 17            Outlier %            2.1%
## 18             NA Count               0
## 19                 Mean         174.944
## 20               Median            6.75
## 21   Standard Deviation         535.812
## 22             Variance      287094.659
## 23                Range            4550
## 24                  Min               0
## 25                  Max            4550
## [1] "L2 length = 0"
##                   metric              V2
## 1                        Variable: FECAL
## 2             Total N:
## 3                 Count            3621
## 4              NA Count               0
## 5                  Mean         104.597
## 6                Median               6
## 7    Standard Deviation         429.193
## 8              Variance      184207.032
## 9                 Range           12000
## 10                  Min               0
## 11                  Max           12000
## 12      25th Percentile           2.583
## 13      75th Percentile          23.083
## 14 Subset w/o Outliers:
## 15                Count            3557
## 16                    %           98.2%
## 17            Outlier %            1.8%
## 18             NA Count               0
## 19                 Mean          60.953
## 20               Median            5.75
## 21   Standard Deviation         174.513
## 22             Variance       30454.732
## 23                Range        1390.192
## 24                  Min               0
## 25                  Max        1390.192
##                   metric              V2
## 1                        Variable: FECAL
## 2             Total N:
## 3                 Count            2973
## 4              NA Count               0
## 5                  Mean         661.416
## 6                Median          11.455
## 7    Standard Deviation        1991.136
## 8              Variance     3964620.809
## 9                 Range           11998
## 10                  Min               2
## 11                  Max           12000
## 12      25th Percentile           2.444
## 13      75th Percentile         180.944
## 14 Subset w/o Outliers:
## 15                Count            2884
## 16                    %             97%
## 17            Outlier %              3%
```

```
## 18             NA Count              0
## 19                 Mean         358.792
## 20               Median          10.364
## 21   Standard Deviation         958.274
## 22             Variance      918289.112
## 23                Range            6598
## 24                  Min               2
## 25                  Max            6600
```

```
## OG
```

```
## Using Avg as value column: use value.var to override.
```



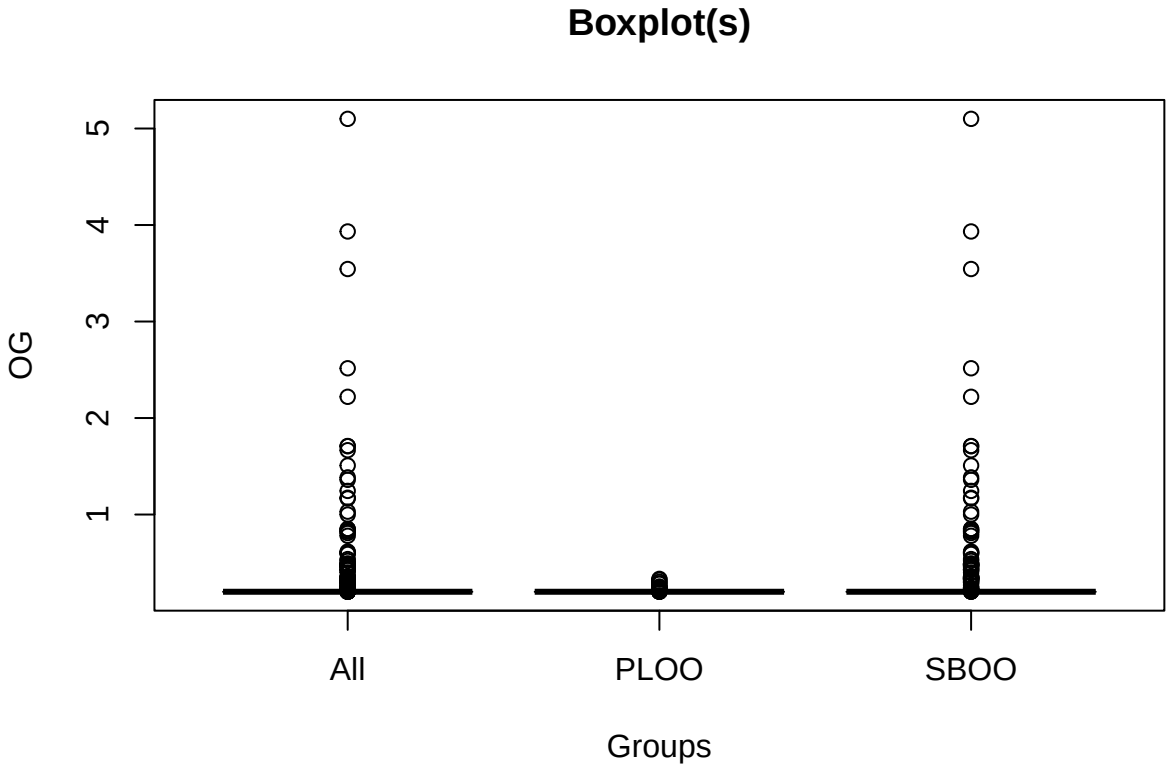

```
##                 metric           V2
## 1                      Variable: OG
## 2            Total N:
## 3               Count          937
## 4            NA Count            0
## 5                Mean        0.244
## 6              Median          0.2
## 7  Standard Deviation        0.286
## 8            Variance        0.082
## 9               Range          4.9
## 10                Min          0.2
## 11                Max          5.1
## 12     25th Percentile          0.2
## 13     75th Percentile          0.2
```

```
## 14 Subset w/o Outliers:
## 15                 Count          923
## 16                     %        98.5%
## 17             Outlier %         1.5%
## 18              NA Count            0
## 19                  Mean        0.215
## 20                Median          0.2
## 21   Standard Deviation        0.078
## 22              Variance        0.006
## 23                 Range         0.83
## 24                   Min          0.2
## 25                   Max         1.03
## [1] "L2 length = 0"
##                    metric           V2
## 1                         Variable: OG
## 2               Total N:
## 3                  Count          345
## 4               NA Count            0
## 5                   Mean        0.203
## 6                 Median          0.2
## 7    Standard Deviation        0.014
## 8               Variance            0
## 9                  Range        0.133
## 10                   Min          0.2
## 11                   Max        0.333
## 12      25th Percentile          0.2
## 13      75th Percentile          0.2
## 14 Subset w/o Outliers:
## 15                 Count          338
## 16                     %          98%
## 17             Outlier %           2%
## 18              NA Count            0
## 19                  Mean        0.201
## 20                Median          0.2
## 21   Standard Deviation        0.005
## 22              Variance            0
## 23                 Range        0.044
## 24                   Min          0.2
## 25                   Max        0.244
##                    metric           V2
## 1                         Variable: OG
## 2               Total N:
## 3                  Count          592
## 4               NA Count            0
## 5                   Mean        0.269
## 6                 Median          0.2
## 7    Standard Deviation        0.358
## 8               Variance        0.128
## 9                  Range          4.9
## 10                   Min          0.2
## 11                   Max          5.1
## 12      25th Percentile          0.2
## 13      75th Percentile          0.2
## 14 Subset w/o Outliers:
```

```
## 15                Count          581
## 16                    %        98.1%
## 17            Outlier %         1.9%
## 18             NA Count            0
## 19                 Mean        0.228
## 20               Median          0.2
## 21   Standard Deviation        0.119
## 22             Variance        0.014
## 23                Range        1.045
## 24                  Min          0.2
## 25                  Max        1.245
```

```
## PH
```

```
## Using Avg as value column: use value.var to override.
```



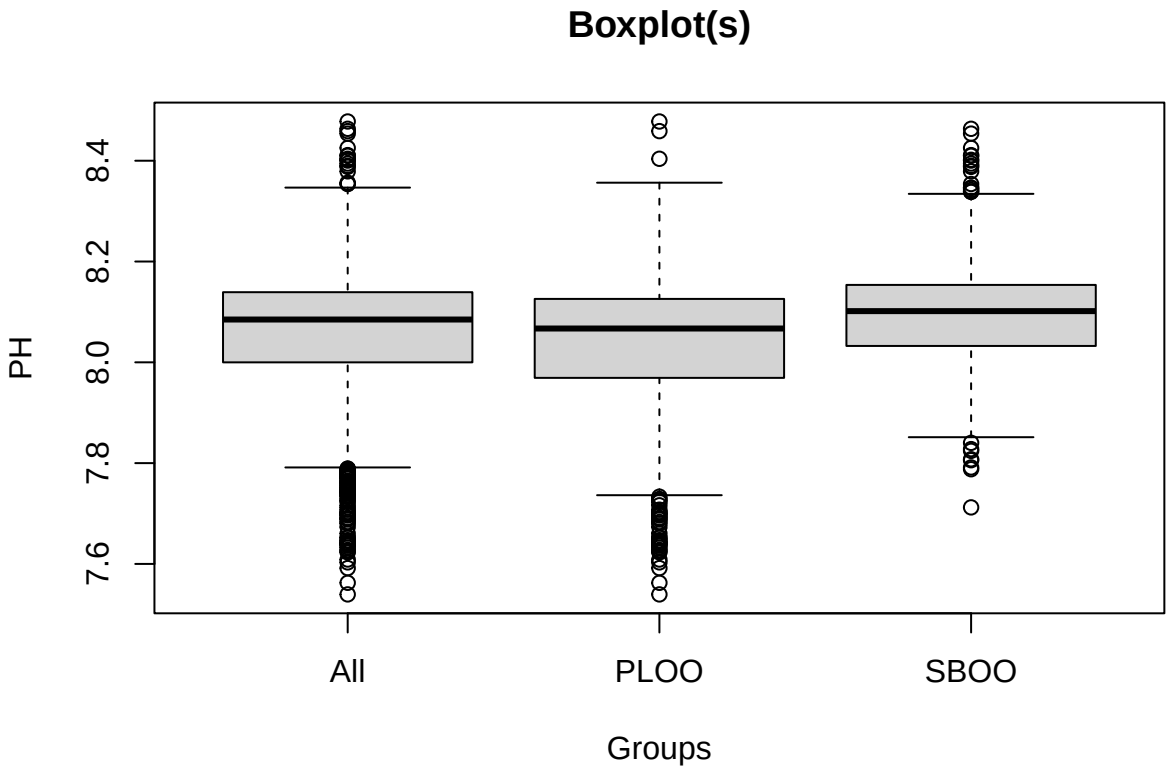

```
##                 metric           V2
## 1                      Variable: PH
## 2             Total N:
## 3                Count         3396
## 4             NA Count            0
## 5                 Mean        8.066
## 6               Median        8.085
## 7    Standard Deviation        0.114
## 8             Variance        0.013
## 9                Range        0.938
## 10                 Min         7.54
```

```
## 11                  Max        8.478
## 12      25th Percentile           8
## 13      75th Percentile        8.139
## 14 Subset w/o Outliers:
## 15                Count         3355
## 16                    %        98.8%
## 17            Outlier %         1.2%
## 18             NA Count           0
## 19                 Mean        8.069
## 20               Median        8.086
## 21   Standard Deviation        0.105
## 22             Variance        0.011
## 23                Range        0.678
## 24                  Min        7.726
## 25                  Max        8.404
## [1] "L2 length = 0"
##                   metric           V2
## 1                        Variable: PH
## 2              Total N:
## 3                 Count         1837
## 4              NA Count           0
## 5                  Mean        8.043
## 6                Median        8.067
## 7    Standard Deviation        0.122
## 8              Variance        0.015
## 9                 Range        0.938
## 10                  Min         7.54
## 11                  Max        8.478
## 12      25th Percentile        7.969
## 13      75th Percentile        8.126
## 14 Subset w/o Outliers:
## 15                Count         1815
## 16                    %        98.8%
## 17            Outlier %         1.2%
## 18             NA Count           0
## 19                 Mean        8.047
## 20               Median        8.069
## 21   Standard Deviation        0.114
## 22             Variance        0.013
## 23                Range        0.724
## 24                  Min         7.68
## 25                  Max        8.404
##                   metric           V2
## 1                        Variable: PH
## 2              Total N:
## 3                 Count         1559
## 4              NA Count           0
## 5                  Mean        8.093
## 6                Median        8.102
## 7    Standard Deviation        0.095
## 8              Variance        0.009
## 9                 Range        0.751
## 10                  Min        7.712
## 11                  Max        8.463
```

```
## 12      25th Percentile        8.032
## 13      75th Percentile        8.154
## 14 Subset w/o Outliers:
## 15                 Count         1545
## 16                     %        99.1%
## 17             Outlier %         0.9%
## 18              NA Count            0
## 19                  Mean        8.092
## 20                Median        8.101
## 21   Standard Deviation        0.091
## 22              Variance        0.008
## 23                 Range        0.571
## 24                   Min        7.808
## 25                   Max        8.379
```

```
## SALINITY
```

```
## Using Avg as value column: use value.var to override.
```



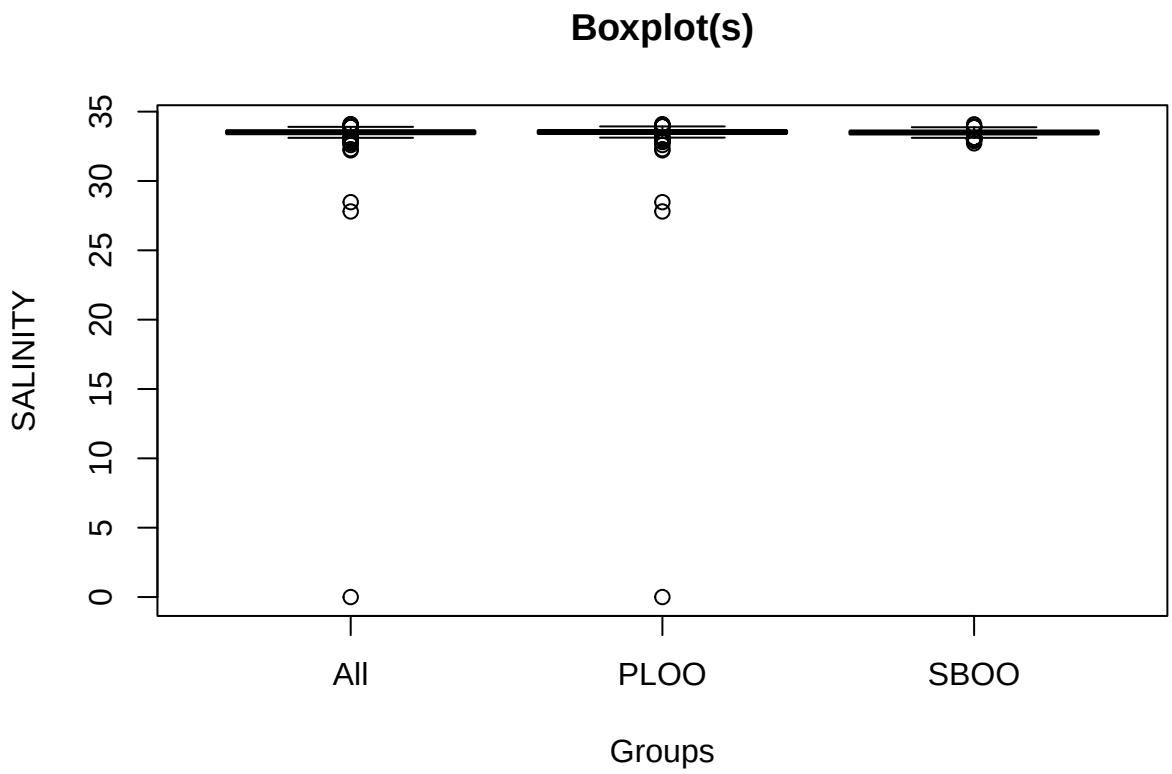

```
##                 metric                V2
## 1                      Variable: SALINITY
## 2             Total N:
## 3                Count              3434
## 4             NA Count                 0
## 5                 Mean            33.497
## 6               Median            33.517
## 7    Standard Deviation             0.609
```

```
## 8              Variance              0.37
## 9                 Range            34.098
## 10                  Min                0
## 11                  Max            34.098
## 12      25th Percentile            33.412
## 13      75th Percentile            33.611
## 14 Subset w/o Outliers:
## 15                Count              3431
## 16                    %             99.9%
## 17            Outlier %              0.1%
## 18             NA Count                 0
## 19                 Mean             33.51
## 20               Median            33.517
## 21   Standard Deviation             0.163
## 22             Variance             0.027
## 23                Range             1.888
## 24                  Min             32.21
## 25                  Max            34.098
## [1] "L2 length = 0"
##                   metric                V2
## 1                        Variable: SALINITY
## 2              Total N:
## 3                 Count              1874
## 4              NA Count                 0
## 5                  Mean            33.503
## 6                Median            33.532
## 7    Standard Deviation             0.812
## 8              Variance              0.66
## 9                 Range            34.098
## 10                  Min                0
## 11                  Max            34.098
## 12      25th Percentile            33.427
## 13      75th Percentile            33.634
## 14 Subset w/o Outliers:
## 15                Count              1871
## 16                    %             99.8%
## 17            Outlier %              0.2%
## 18             NA Count                 0
## 19                 Mean            33.527
## 20               Median            33.532
## 21   Standard Deviation             0.171
## 22             Variance             0.029
## 23                Range             1.888
## 24                  Min             32.21
## 25                  Max            34.098
##                   metric                V2
## 1                        Variable: SALINITY
## 2              Total N:
## 3                 Count              1560
## 4              NA Count                 0
## 5                  Mean            33.489
## 6                Median            33.496
## 7    Standard Deviation              0.15
## 8              Variance             0.022
```

```
## 9                  Range             1.369
## 10                   Min            32.715
## 11                   Max            34.084
## 12       25th Percentile            33.395
## 13       75th Percentile            33.587
## 14 Subset w/o Outliers:
## 15                 Count              1553
## 16                     %             99.6%
## 17             Outlier %              0.4%
## 18              NA Count                 0
## 19                  Mean            33.491
## 20                Median            33.497
## 21    Standard Deviation             0.145
## 22              Variance             0.021
## 23                 Range             0.891
## 24                   Min            33.045
## 25                   Max            33.936
```

```
## SUSO
```

```
## Using Avg as value column: use value.var to override.
```



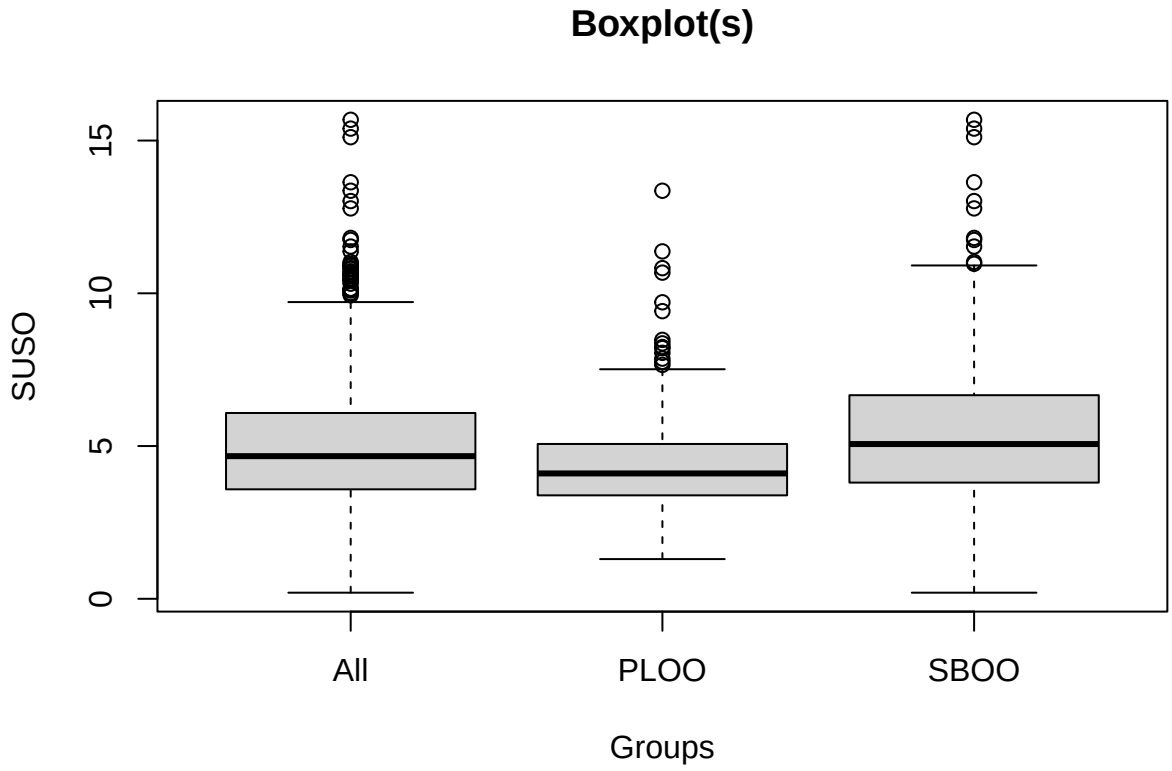

```
##                  metric             V2
## 1                       Variable: SUSO
## 2              Total N:
## 3                 Count            977
## 4              NA Count              0
```

```
## 5                  Mean          4.993
## 6                Median          4.669
## 7    Standard Deviation          2.163
## 8              Variance          4.677
## 9                 Range          15.48
## 10                  Min            0.2
## 11                  Max          15.68
## 12      25th Percentile          3.583
## 13      75th Percentile          6.082
## 14 Subset w/o Outliers:
## 15                Count            967
## 16                    %            99%
## 17            Outlier %             1%
## 18             NA Count              0
## 19                 Mean          4.906
## 20               Median          4.638
## 21   Standard Deviation          1.991
## 22             Variance          3.963
## 23                Range         11.172
## 24                  Min            0.2
## 25                  Max         11.372
## [1] "L2 length = 0"
##                   metric             V2
## 1                       Variable: SUSO
## 2              Total N:
## 3                 Count            386
## 4              NA Count              0
## 5                  Mean          4.383
## 6                Median          4.103
## 7    Standard Deviation          1.571
## 8              Variance          2.469
## 9                 Range         12.058
## 10                  Min            1.3
## 11                  Max         13.358
## 12      25th Percentile           3.39
## 13      75th Percentile          5.057
## 14 Subset w/o Outliers:
## 15                Count            380
## 16                    %          98.4%
## 17            Outlier %           1.6%
## 18             NA Count              0
## 19                 Mean           4.28
## 20               Median          4.073
## 21   Standard Deviation          1.342
## 22             Variance          1.801
## 23                Range          7.176
## 24                  Min            1.3
## 25                  Max          8.476
##                   metric             V2
## 1                       Variable: SUSO
## 2              Total N:
## 3                 Count            591
## 4              NA Count              0
## 5                  Mean          5.392
```

```
## 6                Median          5.065
## 7    Standard Deviation          2.392
## 8              Variance          5.722
## 9                 Range          15.48
## 10                  Min            0.2
## 11                  Max          15.68
## 12      25th Percentile          3.804
## 13      75th Percentile          6.665
## 14 Subset w/o Outliers:
## 15                Count            585
## 16                    %            99%
## 17            Outlier %             1%
## 18             NA Count              0
## 19                 Mean          5.301
## 20               Median          5.044
## 21   Standard Deviation          2.225
## 22             Variance          4.949
## 23                Range         11.613
## 24                  Min            0.2
## 25                  Max         11.813
```

```
## TEMP
```

```
## Using Avg as value column: use value.var to override.
```



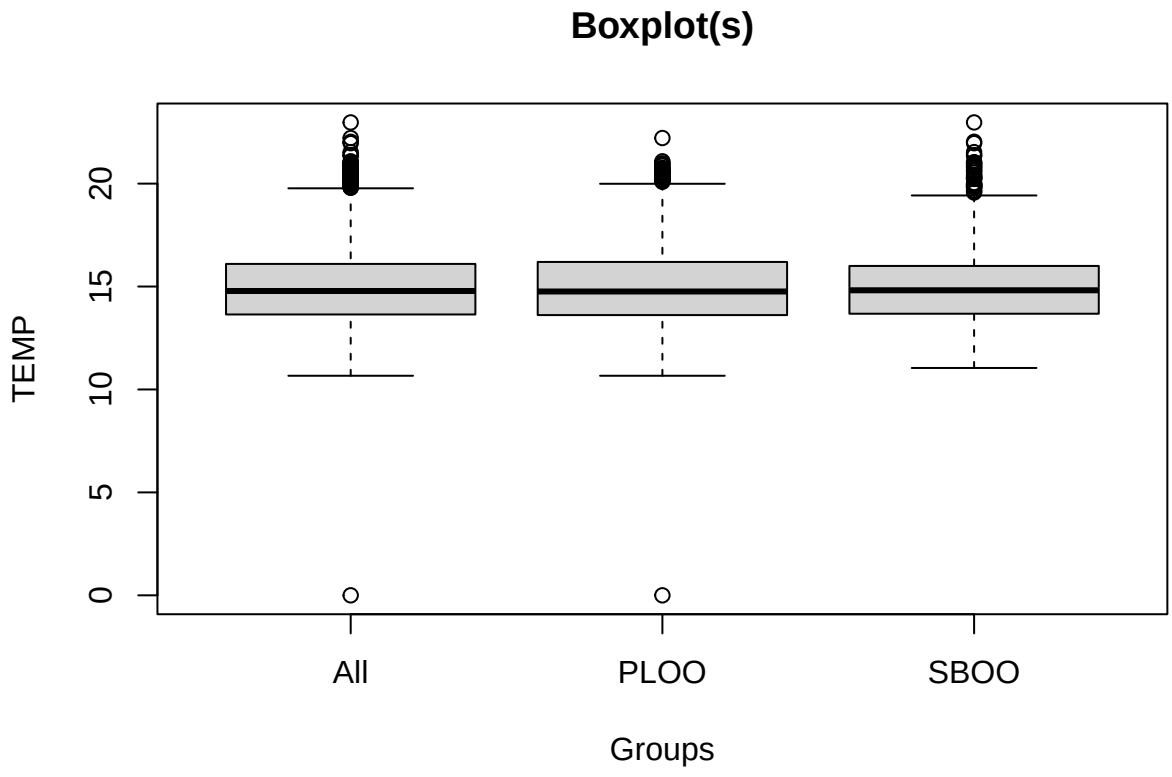

```
##                 metric            V2
## 1                      Variable: TEMP
```

```
## 2              Total N:
## 3                 Count           4524
## 4              NA Count              0
## 5                  Mean         14.969
## 6                Median         14.783
## 7    Standard Deviation          1.906
## 8              Variance          3.633
## 9                 Range         22.972
## 10                  Min              0
## 11                  Max         22.972
## 12      25th Percentile         13.643
## 13      75th Percentile         16.098
## 14 Subset w/o Outliers:
## 15                Count           4503
## 16                    %          99.5%
## 17            Outlier %           0.5%
## 18             NA Count              0
## 19                 Mean         14.945
## 20               Median         14.773
## 21   Standard Deviation           1.85
## 22             Variance          3.423
## 23                Range         10.014
## 24                  Min         10.666
## 25                  Max          20.68
## [1] "L2 length = 0"
##                   metric             V2
## 1                      Variable: TEMP
## 2              Total N:
## 3                 Count           2737
## 4              NA Count              0
## 5                  Mean         14.958
## 6                Median          14.76
## 7    Standard Deviation          1.951
## 8              Variance          3.808
## 9                 Range         22.215
## 10                  Min              0
## 11                  Max         22.215
## 12      25th Percentile         13.613
## 13      75th Percentile         16.195
## 14 Subset w/o Outliers:
## 15                Count           2731
## 16                    %          99.8%
## 17            Outlier %           0.2%
## 18             NA Count              0
## 19                 Mean         14.952
## 20               Median         14.758
## 21   Standard Deviation          1.914
## 22             Variance          3.662
## 23                Range         10.104
## 24                  Min         10.666
## 25                  Max          20.77
##                   metric             V2
## 1                      Variable: TEMP
## 2              Total N:
```

```
## 3                 Count           1787
## 4              NA Count              0
## 5                  Mean         14.987
## 6                Median         14.818
## 7    Standard Deviation          1.835
## 8              Variance          3.367
## 9                 Range         11.927
## 10                  Min         11.044
## 11                  Max         22.972
## 12      25th Percentile          13.68
## 13      75th Percentile             16
## 14 Subset w/o Outliers:
## 15                Count           1771
## 16                    %          99.1%
## 17            Outlier %           0.9%
## 18             NA Count              0
## 19                 Mean         14.931
## 20               Median         14.795
## 21   Standard Deviation          1.744
## 22             Variance          3.041
## 23                Range          9.386
## 24                  Min         11.044
## 25                  Max          20.43
```

```
## XMS
```

```
## Using Avg as value column: use value.var to override.
```



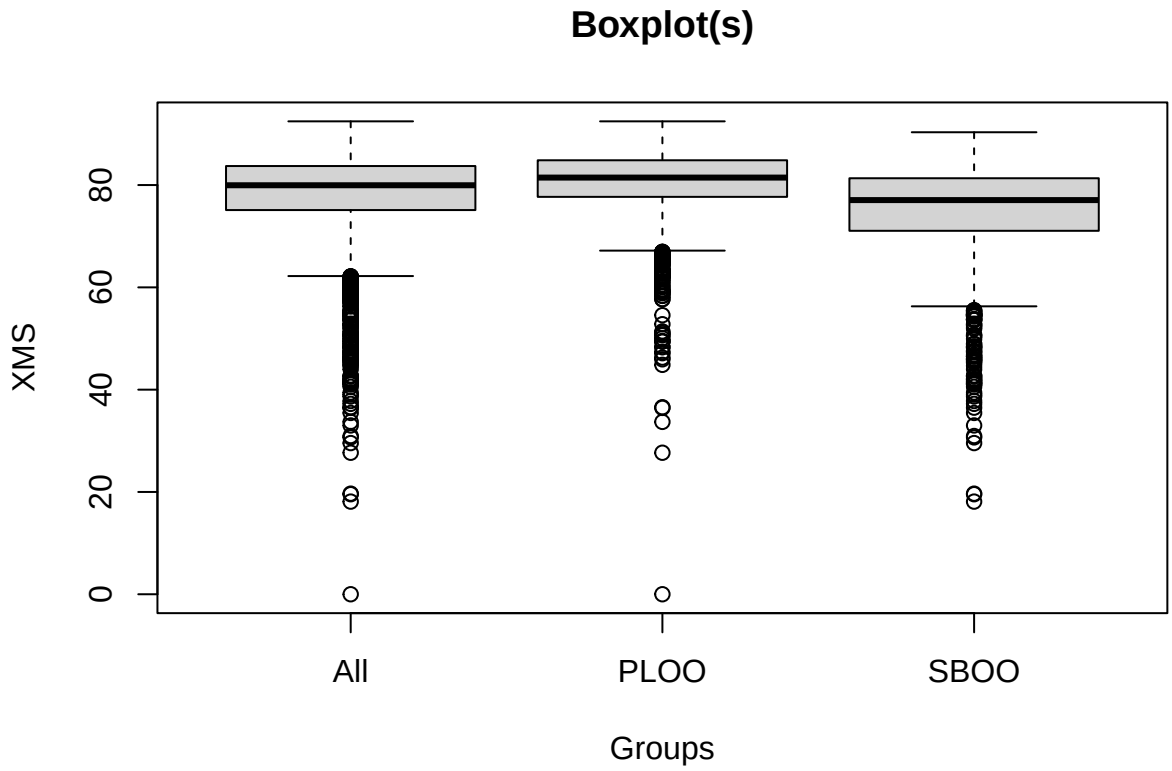
```
##                   metric            V2
## 1                        Variable: XMS
## 2               Total N:
## 3                  Count          4494
## 4               NA Count             0
## 5                   Mean        78.307
## 6                 Median        79.958
## 7    Standard Deviation         8.292
## 8               Variance        68.754
## 9                  Range        92.456
## 10                   Min             0
## 11                   Max        92.456
## 12      25th Percentile        75.112
## 13      75th Percentile        83.716
## 14 Subset w/o Outliers:
## 15                 Count          4417
## 16                     %         98.3%
## 17             Outlier %          1.7%
## 18              NA Count             0
## 19                  Mean        78.923
## 20                Median        80.085
## 21   Standard Deviation          6.81
## 22              Variance        46.377
## 23                 Range        38.826
## 24                   Min         53.63
## 25                   Max        92.456
## [1] "L2 length = 0"
##                   metric            V2
## 1                        Variable: XMS
## 2               Total N:
## 3                  Count          2714
## 4               NA Count             0
## 5                   Mean        80.521
## 6                 Median        81.462
## 7    Standard Deviation         6.662
## 8               Variance        44.384
## 9                  Range        92.456
## 10                   Min             0
## 11                   Max        92.456
## 12      25th Percentile        77.697
## 13      75th Percentile        84.845
## 14 Subset w/o Outliers:
## 15                 Count          2679
## 16                     %         98.7%
## 17             Outlier %          1.3%
## 18              NA Count             0
## 19                  Mean        80.924
## 20                Median        81.612
## 21   Standard Deviation          5.53
## 22              Variance         30.58
## 23                 Range        31.867
## 24                   Min         60.59
## 25                   Max        92.456
##                   metric            V2
```

```
## 1                       Variable: XMS
## 2              Total N:
## 3                 Count          1780
## 4              NA Count             0
## 5                  Mean        74.931
## 6                Median        77.063
## 7    Standard Deviation         9.331
## 8              Variance        87.069
## 9                 Range        72.217
## 10                  Min        18.111
## 11                  Max        90.328
## 12      25th Percentile        71.059
## 13      75th Percentile        81.333
## 14 Subset w/o Outliers:
## 15                Count          1742
## 16                    %         97.9%
## 17            Outlier %          2.1%
## 18             NA Count             0
## 19                 Mean        75.707
## 20               Median        77.226
## 21   Standard Deviation         7.716
## 22             Variance        59.531
## 23                Range        43.044
## 24                  Min        47.283
## 25                  Max        90.328
```

```
##    date_sample project CHLOROPHYLL DENSITY      DO    ENTERO     FECAL OG    PH
## 1   1990-11-15    PLOO        0.87  23.855 6.55000        NA        NA NA 8.080
## 2   1990-11-15    SBOO        1.27  23.853 7.41000        NA        NA NA 8.180
## 3   1991-01-02    PLOO          NA      NA      NA  32.05128 125.64103 NA    NA
## 4   1991-01-03    PLOO          NA      NA      NA  18.71795  51.28205 NA    NA
## 5   1991-01-07    PLOO          NA      NA      NA 106.15385  80.76923 NA    NA
## 6   1991-01-08    PLOO          NA      NA      NA  40.00000  40.55556 NA    NA
## 7   1991-01-09    PLOO          NA      NA 5.47625  20.33898  40.50847 NA 8.195
## 8   1991-01-10    PLOO          NA      NA 5.50000  40.22727  78.18182 NA    NA
## 9   1991-01-11    PLOO          NA      NA 5.36000  83.92857 446.07143 NA    NA
## 10  1991-01-14    PLOO          NA      NA      NA  33.84615  95.64103 NA    NA
##    SALINITY SUSO     TEMP      XMS
## 1  33.61700   NA 19.43000 44.85000
## 2  33.57400   NA 19.31000 75.60000
## 3        NA   NA 14.50769 80.02564
## 4        NA   NA 14.39231 83.89744
## 5        NA   NA 14.54103 78.97436
## 6        NA   NA 13.43000 80.20000
## 7  33.51938   NA 14.26889 86.21111
## 8  33.49300   NA 13.75000 80.45833
## 9  33.49100   NA 14.38400 84.40000
## 10       NA   NA 14.51282 72.02564
```

```
## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning Inf
```

```
## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning -Inf
```



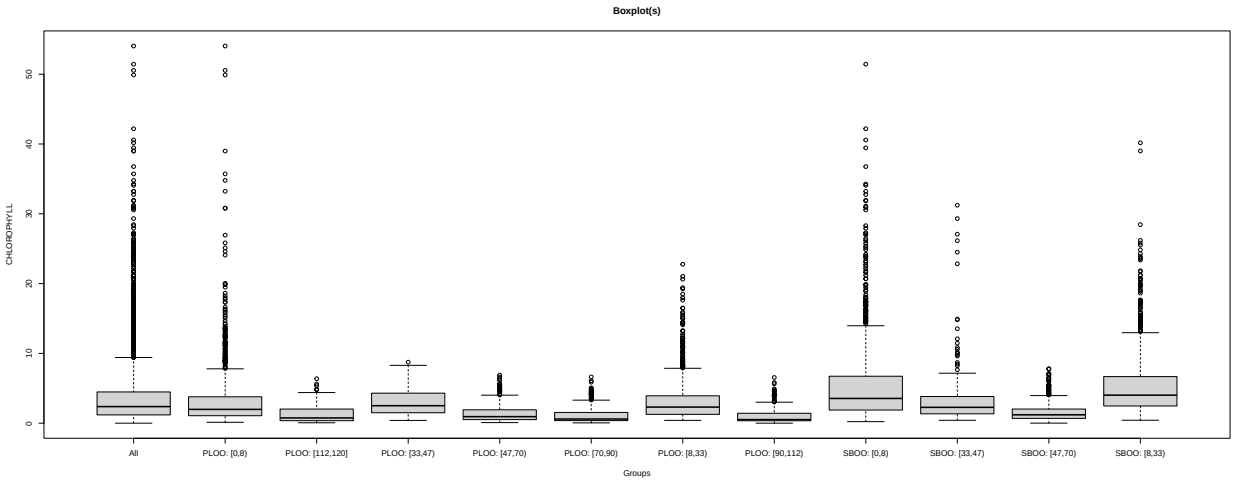

```
##              vars    n   mean      sd median trimmed  mad   min      max
## date_sample     1 7037    NaN      NA     NA     NaN   NA   Inf     -Inf
## project*        2 7037   1.43    0.50   1.00    1.42 0.00  1.00     2.00
## CHLOROPHYLL     3 3023   4.01    3.81   2.81    3.31 2.27  0.29    36.05
## DENSITY         4 3177  24.79    0.47  24.82   24.81 0.45 22.71    26.04
## DO              5 3434   7.15    0.97   7.31    7.22 0.82  0.00    10.81
## ENTERO          6 6776 253.64 1249.91   5.27   18.53 4.84  0.00 18000.00
## FECAL           7 6594 355.65 1401.82   7.07   48.93 7.51  0.00 12000.00
## OG              8  937   0.24    0.29   0.20    0.20 0.00  0.20     5.10
## PH              9 3396   8.07    0.11   8.08    8.07 0.10  7.54     8.48
## SALINITY       10 3434  33.50    0.61  33.52   33.51 0.15  0.00    34.10
## SUSO           11  977   4.99    2.16   4.67    4.80 1.80  0.20    15.68
## TEMP           12 4524  14.97    1.91  14.78   14.87 1.80  0.00    22.97
## XMS            13 4494  78.31    8.29  79.96   79.37 6.18  0.00    92.46
##                range   skew kurtosis    se
## date_sample     -Inf     NA       NA    NA
## project*        1.00   0.26    -1.93  0.01
## CHLOROPHYLL    35.76   2.56     9.61  0.07
## DENSITY         3.33  -0.46     0.38  0.01
## DO             10.81  -0.89     3.49  0.02
## ENTERO      18000.00   7.41    61.01 15.18
## FECAL       12000.00   6.25    43.27 17.26
## OG              4.90  10.89   144.21  0.01
## PH              0.94  -0.67     1.45  0.00
## SALINITY       34.10 -48.98  2674.00  0.01
## SUSO           15.48   1.09     2.23  0.07
```

```
## TEMP           22.97   0.43     0.95  0.03
## XMS            92.46  -2.00     7.54  0.12
##          date_sample project parameter        Avg
## Not NA n       48395   48395     48395 47222.0000
## NA n               0       0         0  1173.0000
## Not NA %           1       1         1     0.9758
## NA %               0       0         0     0.0242
```

Run custom function on data aggregated by `date_sample`, `project`, `depth_m_bin` and `parameter`

```
owt_df02_gb07_mrgd = ts_eda(ts_df = owt_df02_gb07,
                            form_lead = c("date_sample", "project", "depth_m_bin"),
                            form_cast = "parameter",
                            param_lst = param_lst02,
                            l1_col = c("project"),
                            l1_param = c("PLOO", "SBOO"),
                            l2_col = c("depth_m_bin"),
                            l2_param = depth_lvls01,
                            rtn_met = FALSE,
                            box = FALSE
                            )
```

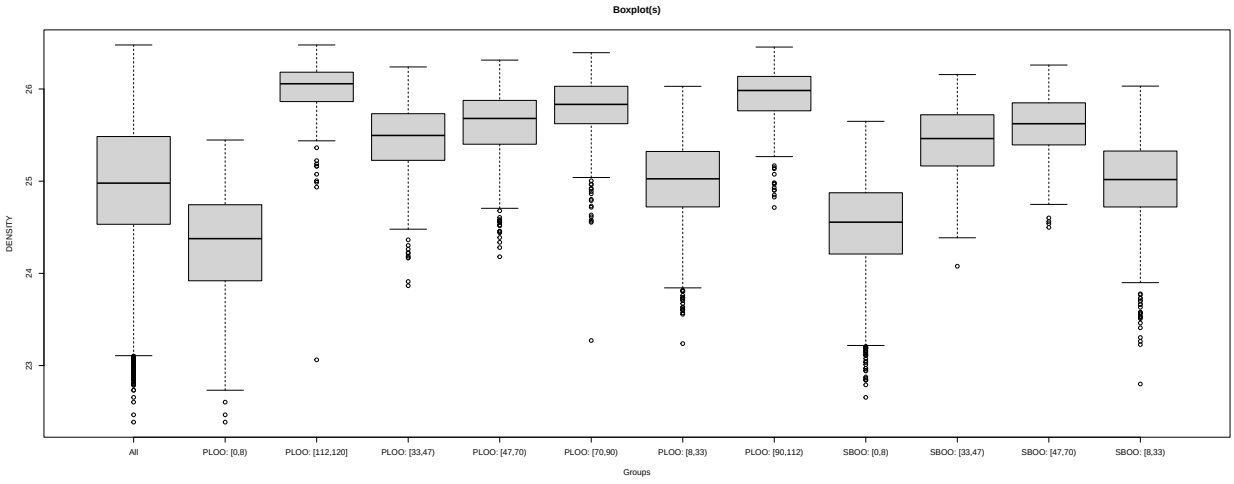
```
## CHLOROPHYLL

## Using Avg as value column: use value.var to override.

## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0

## DENSITY

## Using Avg as value column: use value.var to override.
```




```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```

```
## DO
```

```
## Using Avg as value column: use value.var to override.
```

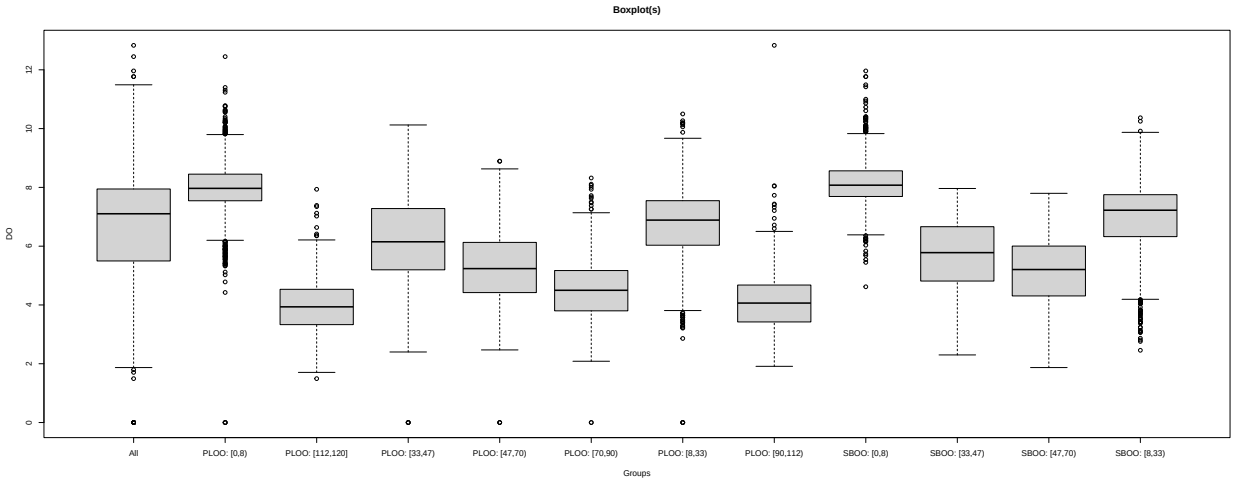



```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
```

```
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```

```
## ENTERO
```

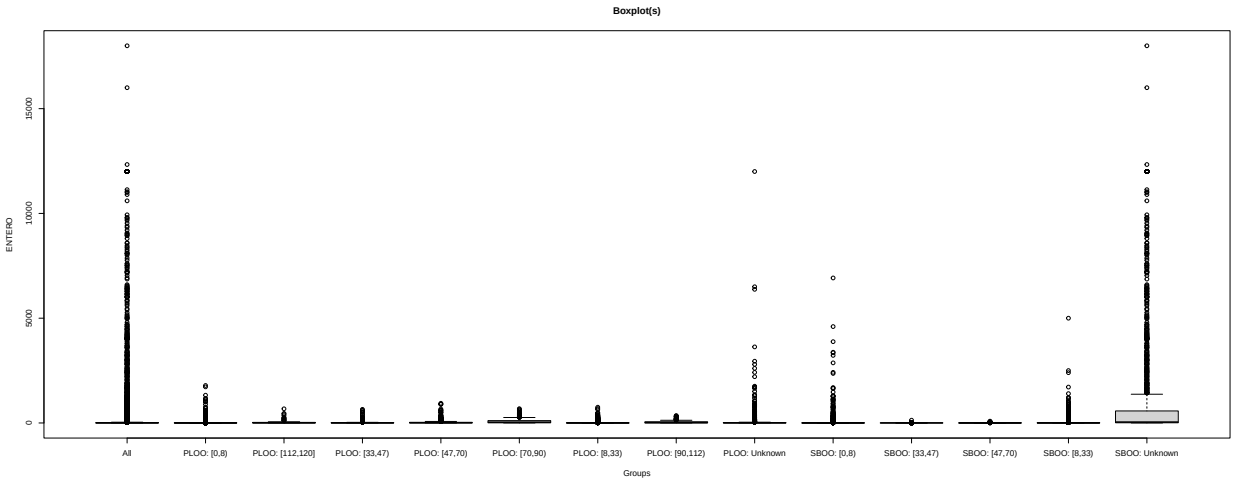
```
## Using Avg as value column: use value.var to override.
```




```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
```

```
## FECAL
```

```
## Using Avg as value column: use value.var to override.
```

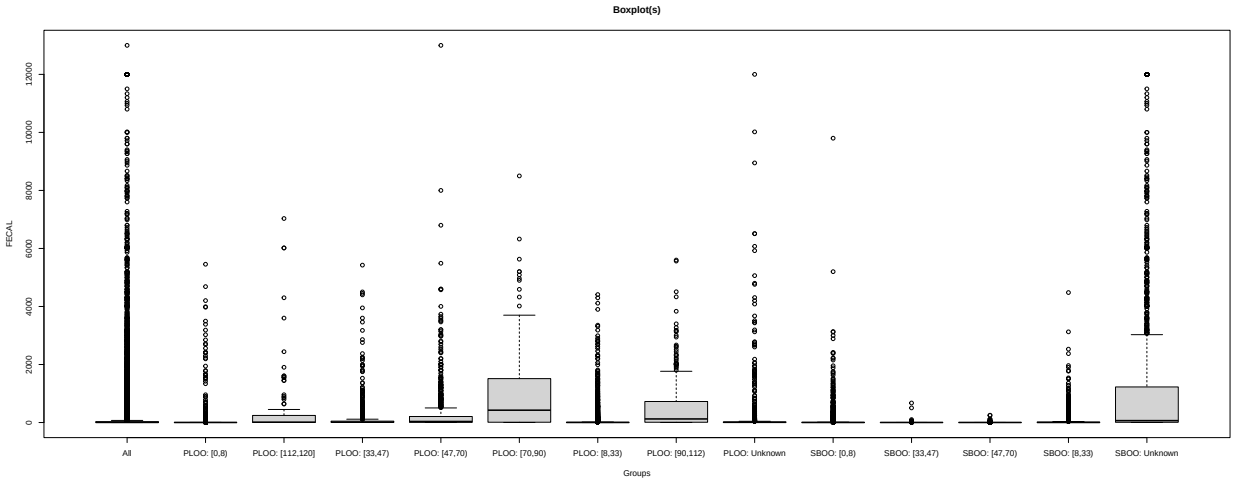



```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
```

```
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
```

```
## OG
```

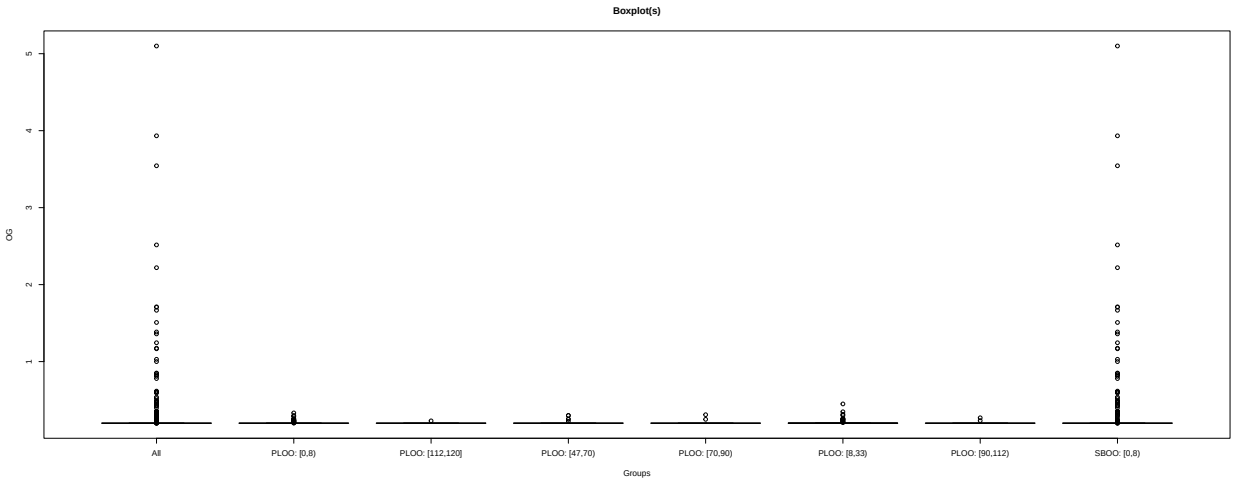
```
## Using Avg as value column: use value.var to override.
```




```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: [33,47)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [8,33)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [33,47)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [47,70)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```

```
## PH
```

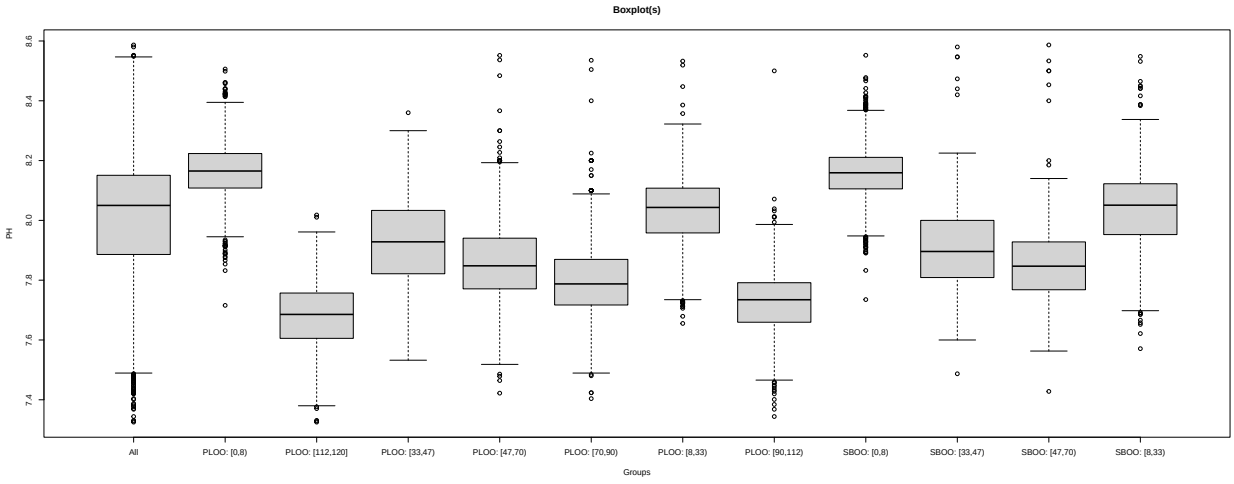
```
## Using Avg as value column: use value.var to override.
```




```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```

```
## SALINITY
```

```
## Using Avg as value column: use value.var to override.
```

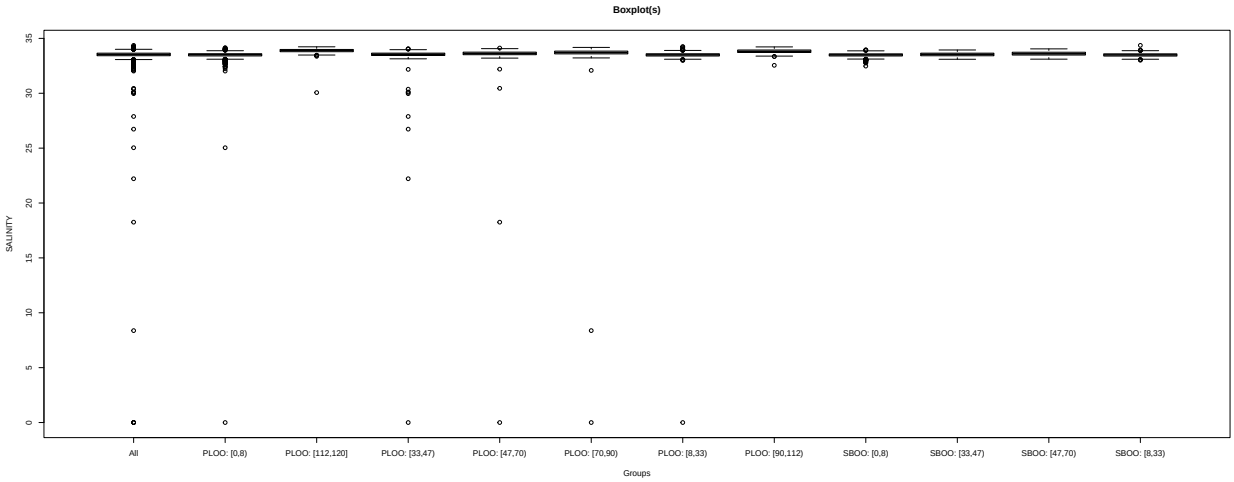



```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
```

```
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```

```
## SUSO
```

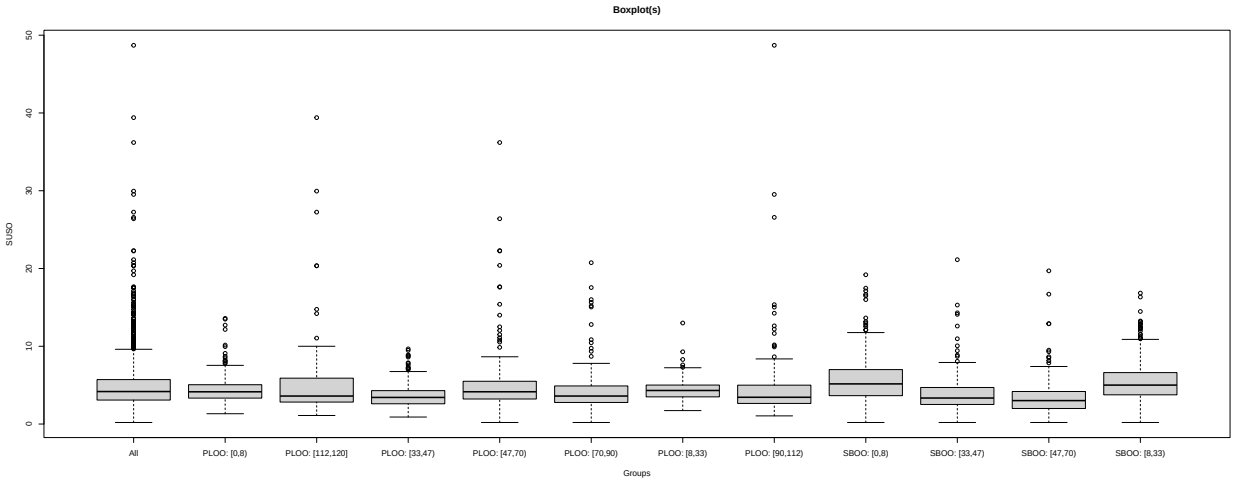
```
## Using Avg as value column: use value.var to override.
```




```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```

```
## TEMP
```

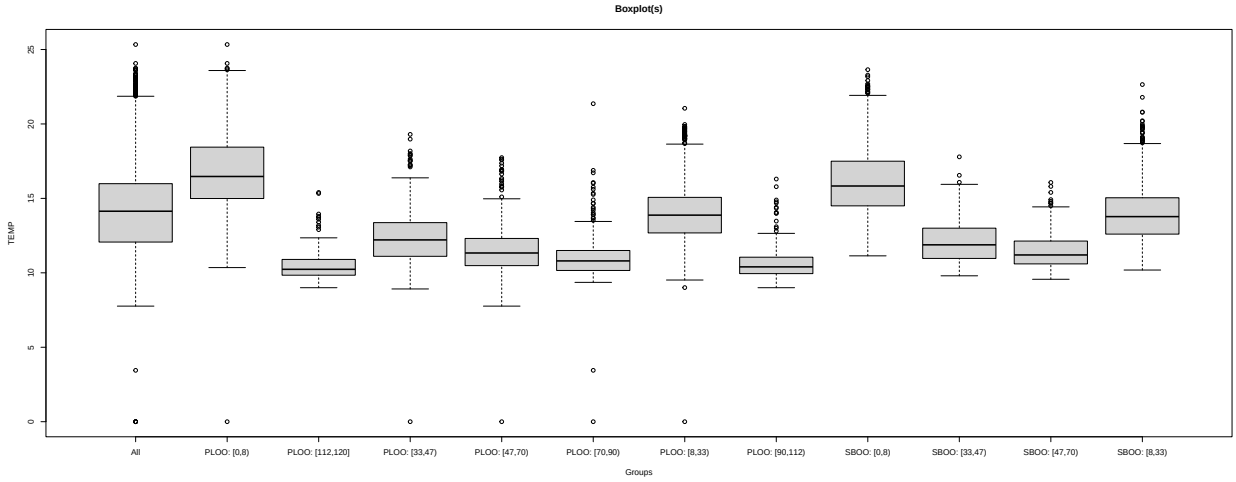
```
## Using Avg as value column: use value.var to override.
```



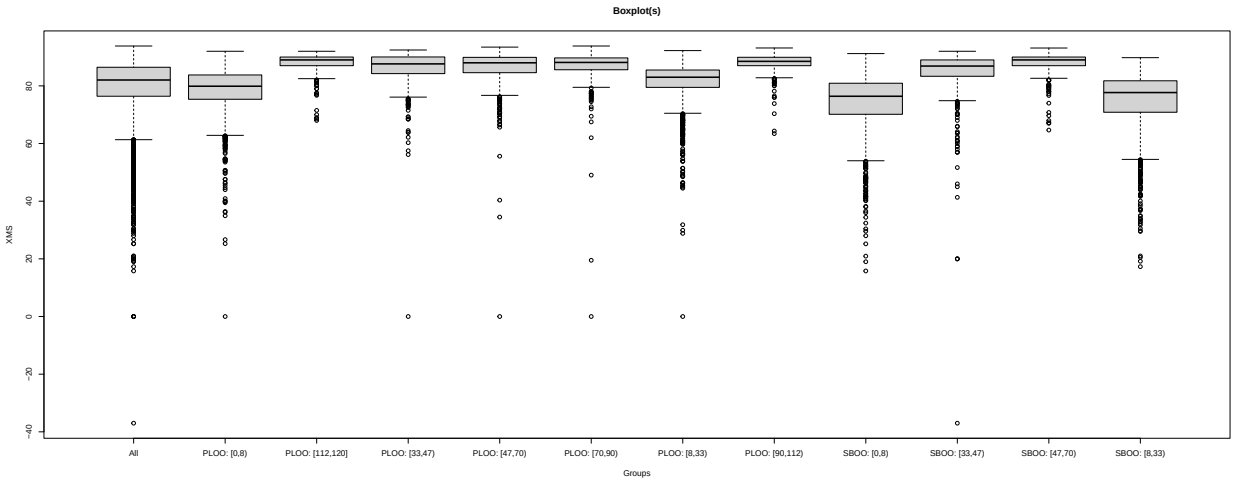
```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```

```
## XMS
```

```
## Using Avg as value column: use value.var to override.
```




```
## [1] "L2 length = 8"
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "PLOO: Unknown") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [70,90)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [90,112)") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: [112,120]") :
##   replacement has 1 row, data has 0
## Error in `$<-.data.frame`(`*tmp*`, "Groups", value = "SBOO: Unknown") :
##   replacement has 1 row, data has 0
```




```
##    date_sample project depth_m_bin CHLOROPHYLL DENSITY   DO    ENTERO     FECAL
## 1   1990-11-15    PLOO       [8,33)        0.87  23.855 6.55        NA        NA
## 2   1990-11-15    SBOO        [0,8)        1.27  23.853 7.41        NA        NA
## 3   1991-01-02    PLOO        [0,8)          NA      NA   NA  34.07407 134.81481
## 4   1991-01-02    PLOO       [8,33)          NA      NA   NA  27.50000 105.00000
## 5   1991-01-03    PLOO        [0,8)          NA      NA   NA  17.40741  45.55556
## 6   1991-01-03    PLOO       [8,33)          NA      NA   NA  21.66667  64.16667
## 7   1991-01-07    PLOO        [0,8)          NA      NA   NA  77.40741  44.07407
## 8   1991-01-07    PLOO       [8,33)          NA      NA   NA 170.83333 163.33333
## 9   1991-01-08    PLOO        [0,8)          NA      NA   NA  36.25000  31.66667
## 10  1991-01-08    PLOO      [33,47)          NA      NA   NA  10.00000  12.50000
##    OG   PH SALINITY SUSO     TEMP      XMS
## 1  NA 8.08   33.617   NA 19.43000 44.85000
## 2  NA 8.18   33.574   NA 19.31000 75.60000
## 3  NA   NA       NA   NA 14.56667 77.03704
## 4  NA   NA       NA   NA 14.37500 86.75000
## 5  NA   NA       NA   NA 14.48519 83.55556
## 6  NA   NA       NA   NA 14.18333 84.66667
## 7  NA   NA       NA   NA 14.61852 77.62963
## 8  NA   NA       NA   NA 14.36667 82.00000
## 9  NA   NA       NA   NA 13.36667 81.00000
## 10 NA   NA       NA   NA       NA       NA
##              vars     n   mean      sd median trimmed  mad    min      max
## date_sample     1 17003    NaN      NA     NA     NaN   NA    Inf     -Inf
## project*        2 17003   1.37    0.48   1.00    1.34 0.00   1.00     2.00
## depth_m_bin*    3 17003   4.49    2.61   5.00    4.48 2.97   1.00     8.00
## CHLOROPHYLL     4  8386   3.60    4.09   2.38    2.82 2.13   0.00    54.04
## DENSITY         5  8627  24.96    0.69  24.98   24.99 0.70  22.39    26.48
## DO              6  9842   6.70    1.69   7.10    6.80 1.58   0.00    12.83
## ENTERO          7 15533 150.11  925.42   2.80   10.66 1.19   0.00 18000.00
## FECAL           8 14874 269.82 1129.98   3.85   28.41 2.74   0.00 13000.00
## OG              9  1115   0.24    0.26   0.20    0.20 0.00   0.20     5.10
## PH             10  9695   8.01    0.18   8.05    8.02 0.18   7.33     8.59
## SALINITY       11  9844  33.52    0.86  33.54   33.54 0.17   0.00    34.36
## SUSO           12  3023   4.77    2.99   4.18    4.40 1.89   0.20    48.70
## TEMP           13 12010  14.27    2.80  14.14   14.11 2.90   0.00    25.34
## XMS            14 11969  80.22    9.07  82.04   81.41 7.25 -37.00    93.83
##                range   skew kurtosis   se
```

```
## date_sample      -Inf     NA       NA   NA
## project*         1.00   0.54    -1.71 0.00
## depth_m_bin*     7.00  -0.11    -1.42 0.02
## CHLOROPHYLL     54.04   3.70    22.87 0.04
## DENSITY          4.09  -0.36    -0.15 0.01
## DO              12.83  -0.55    -0.16 0.02
## ENTERO       18000.00   9.88   111.40 7.43
## FECAL        13000.00   7.12    59.92 9.27
## OG               4.90  11.87   171.74 0.01
## PH               1.26  -0.50    -0.10 0.00
## SALINITY        34.36 -34.13  1283.78 0.01
## SUSO            48.50   3.91    33.72 0.05
## TEMP            25.34   0.42     0.06 0.03
## XMS            130.83  -2.14     9.79 0.08
##          date_sample project depth_m_bin parameter         Avg
## Not NA n      121286  121286      121286    121286 119635.0000
## NA n               0       0           0         0   1651.0000
## Not NA %           1       1           1         1      0.9864
## NA %               0       0           0         0      0.0136
```

Display new dataframes

```
#print(head(owt_df02_gb04_mrgd, 10))
#print(head(owt_df02_gb09_mrgd, 10))
#print(head(owt_df02_gb07_mrgd, 10))

owt_df02_gb04_tbbl <- as_tibble(owt_df02_gb04_mrgd, index = date_sample)
owt_df02_gb09_tbbl <- as_tibble(owt_df02_gb09_mrgd, key = project, index = date_sample)
owt_df02_gb07_tbbl <- as_tibble(owt_df02_gb07_mrgd, key = c(project, depth_m_bin), index
↪  = date_sample)

print(head(owt_df02_gb04_tbbl, 10))
```

```
## # A tibble: 10 x 12
##    date_sam~1 CHLOR~2 DENSITY    DO ENTERO FECAL    OG    PH SALIN~3  SUSO  TEMP
##    <date>       <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl>
##  1 1990-11-15    1.07    23.9  6.98   NA    NA      NA  8.13    33.6    NA  19.4
##  2 1991-01-02   NA       NA    NA      32.1 126.     NA NA       NA      NA  14.5
##  3 1991-01-03   NA       NA    NA      18.7  51.3    NA NA       NA      NA  14.4
##  4 1991-01-07   NA       NA    NA     106.   80.8    NA NA       NA      NA  14.5
##  5 1991-01-08   NA       NA    NA      40    40.6    NA NA       NA      NA  13.4
##  6 1991-01-09   NA       NA     5.48   20.3  40.5    NA  8.19    33.5    NA  14.3
##  7 1991-01-10   NA       NA     5.5    40.2  78.2    NA NA       33.5    NA  13.8
##  8 1991-01-11   NA       NA     5.36   83.9 446.     NA NA       33.5    NA  14.4
##  9 1991-01-14   NA       NA    NA      33.8  95.6    NA NA       NA      NA  14.5
## 10 1991-01-16   NA       NA    NA      12.6  23.6    NA NA       NA      NA  14.6
## # ... with 1 more variable: XMS <dbl>, and abbreviated variable names
## #   1: date_sample, 2: CHLOROPHYLL, 3: SALINITY
```

```
print(head(owt_df02_gb09_tbbl, 10))
```

```
## # A tibble: 10 x 13
##    date_sample project CHLOROPH~1 DENSITY    DO ENTERO FECAL    OG    PH SALIN~2
##    <date>      <fct>        <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>   <dbl>
```

```
##  1 1990-11-15  PLOO          0.87    23.9  6.55   NA    NA      NA  8.08    33.6
##  2 1990-11-15  SBOO          1.27    23.9  7.41   NA    NA      NA  8.18    33.6
##  3 1991-01-02  PLOO         NA       NA   NA      32.1 126.     NA NA       NA
##  4 1991-01-03  PLOO         NA       NA   NA      18.7  51.3    NA NA       NA
##  5 1991-01-07  PLOO         NA       NA   NA     106.   80.8    NA NA       NA
##  6 1991-01-08  PLOO         NA       NA   NA      40    40.6    NA NA       NA
##  7 1991-01-09  PLOO         NA       NA    5.48   20.3  40.5    NA  8.19    33.5
##  8 1991-01-10  PLOO         NA       NA    5.5    40.2  78.2    NA NA       33.5
##  9 1991-01-11  PLOO         NA       NA    5.36   83.9 446.     NA NA       33.5
## 10 1991-01-14  PLOO         NA       NA   NA      33.8  95.6    NA NA       NA
## # ... with 3 more variables: SUSO <dbl>, TEMP <dbl>, XMS <dbl>, and abbreviated
## #   variable names 1: CHLOROPHYLL, 2: SALINITY
```

```
print(head(owt_df02_gb07_tbbl, 10))
```

```
## # A tibble: 10 x 14
##    date_sample project depth_m_~1 CHLOR~2 DENSITY    DO ENTERO FECAL    OG    PH
##    <date>      <fct>   <chr>        <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>
##  1 1990-11-15  PLOO    [8,33)        0.87    23.9  6.55   NA    NA      NA  8.08
##  2 1990-11-15  SBOO    [0,8)         1.27    23.9  7.41   NA    NA      NA  8.18
##  3 1991-01-02  PLOO    [0,8)        NA       NA   NA      34.1 135.     NA NA
##  4 1991-01-02  PLOO    [8,33)       NA       NA   NA      27.5 105      NA NA
##  5 1991-01-03  PLOO    [0,8)        NA       NA   NA      17.4  45.6    NA NA
##  6 1991-01-03  PLOO    [8,33)       NA       NA   NA      21.7  64.2    NA NA
##  7 1991-01-07  PLOO    [0,8)        NA       NA   NA      77.4  44.1    NA NA
##  8 1991-01-07  PLOO    [8,33)       NA       NA   NA     171.  163.     NA NA
##  9 1991-01-08  PLOO    [0,8)        NA       NA   NA      36.2  31.7    NA NA
## 10 1991-01-08  PLOO    [33,47)      NA       NA   NA      10    12.5    NA NA
## # ... with 4 more variables: SALINITY <dbl>, SUSO <dbl>, TEMP <dbl>, XMS <dbl>,
## #   and abbreviated variable names 1: depth_m_bin, 2: CHLOROPHYLL
```

Convert dataframes to tsibbles

```
#_______________________________________________________________________________
owt_df02_gb04_tbbl01 <- owt_df02_gb04_mrgd[c("date_sample", "CHLOROPHYLL")] %>%
  drop_na() %>%
  group_by(date_sample) %>%
  summarise_by_time(.date_var = date_sample,
                    .by = "week",
                    .week_start = 1,
                    CHLOROPHYLL = mean(CHLOROPHYLL, na.remove = TRUE)
                    )

print(head(owt_df02_gb04_tbbl01, 10))
```

```
## # A tibble: 10 x 2
##    date_sample CHLOROPHYLL
##    <date>            <dbl>
##  1 1990-11-12        1.07
##  2 1991-11-11        0.93
##  3 1992-11-09        0.91
##  4 1993-11-15        0.88
##  5 1994-11-14        0.86
##  6 1995-11-13        0.89
```

```
##  7 1996-01-22        0.401
##  8 1996-11-11        1.05
##  9 1997-01-06        0.416
## 10 1997-01-20        0.416
```

```
print(duplicates(owt_df02_gb04_tbbl01, index = date_sample))
```

```
## # A tibble: 0 x 2
## # ... with 2 variables: date_sample <date>, CHLOROPHYLL <dbl>
```

```
#_______________________________________________________________________________
owt_df02_gb04_tbbl02 <- owt_df02_gb04_mrgd[c("date_sample", "ENTERO")] %>%
  drop_na() %>%
  group_by(date_sample) %>%
  summarise_by_time(.date_var = date_sample,
                    .by = "week",
                    .week_start = 1,
                    ENTERO = mean(ENTERO, na.remove = TRUE)
                    )

print(head(owt_df02_gb04_tbbl02, 10))
```

```
## # A tibble: 10 x 2
##    date_sample ENTERO
##    <date>       <dbl>
##  1 1990-12-31   25.4
##  2 1991-01-07   58.1
##  3 1991-01-14   23.2
##  4 1991-01-21   31.8
##  5 1991-01-28  154.
##  6 1991-02-04   16.8
##  7 1991-02-11   45.3
##  8 1991-02-18    9.08
##  9 1991-02-25   16.5
## 10 1991-03-04   43.5
```

```
print(duplicates(owt_df02_gb04_tbbl02, index = date_sample))
```

```
## # A tibble: 0 x 2
## # ... with 2 variables: date_sample <date>, ENTERO <dbl>
```

```
owt_df02_gb04_tbb_mrgd <- merge(x = owt_df02_gb04_tbbl01,
                    y = owt_df02_gb04_tbbl02,
                    by = "date_sample",
                    all = TRUE)
print(head(owt_df02_gb04_tbb_mrgd, 10))
```

```
##    date_sample CHLOROPHYLL     ENTERO
## 1   1990-11-12        1.07         NA
## 2   1990-12-31          NA  25.384615
## 3   1991-01-07          NA  58.129735
## 4   1991-01-14          NA  23.205128
## 5   1991-01-21          NA  31.818910
## 6   1991-01-28          NA 154.230769
## 7   1991-02-04          NA  16.838141
## 8   1991-02-11          NA  45.317879
```

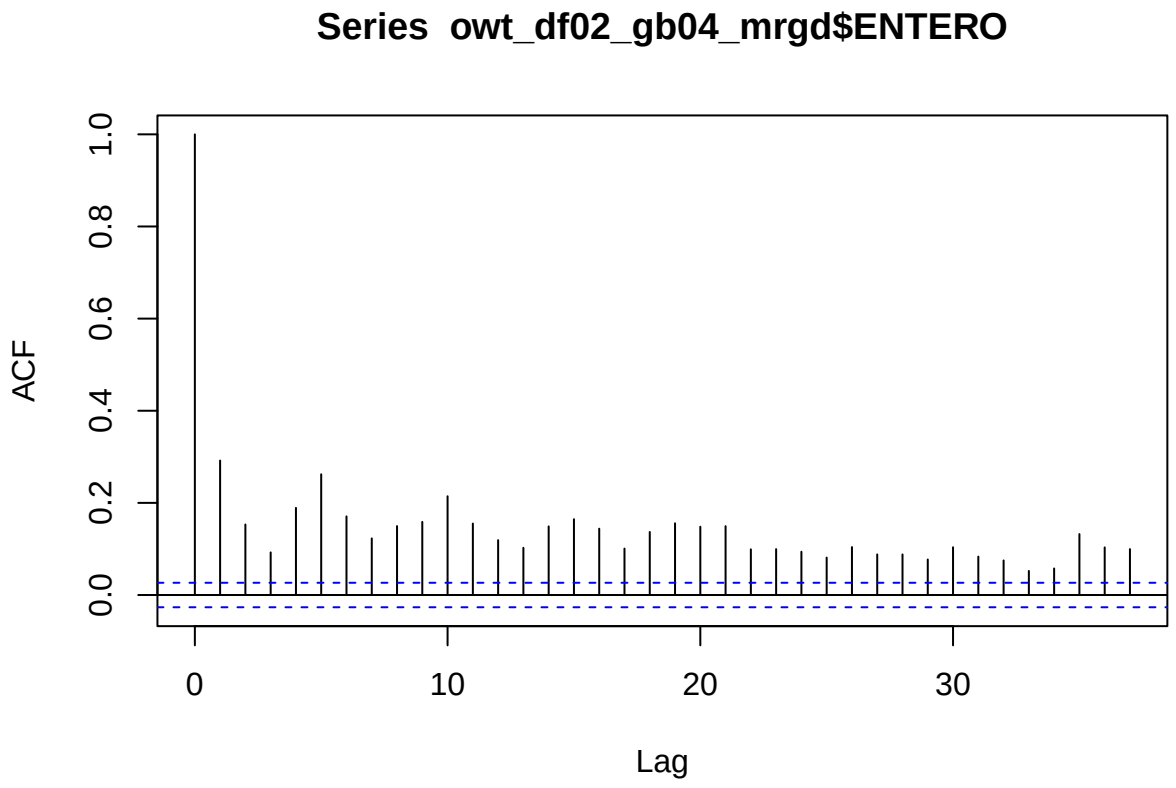
```
## 9   1991-02-18          NA   9.083333
## 10  1991-02-25          NA  16.512821
```

{r} owt_df02_gb09_tbbl01a <- owt_df02_gb09_tbbl01[1:3] print(head(owt_df02_gb09_tbbl01a)) owt_df02_gb09_tbbl01a %>% group_by_key() %>% index_by(year_week = ~ yearweek(date_sample)) %>% # monthly aggregates summarise(CHLOROPHYLL_avg = mean(CHLOROPHYLL, na.rm = TRUE))

#print(head(owt_df02_gb09_tbbl02, 100))

Plot time series for selected parameters and tsibbles

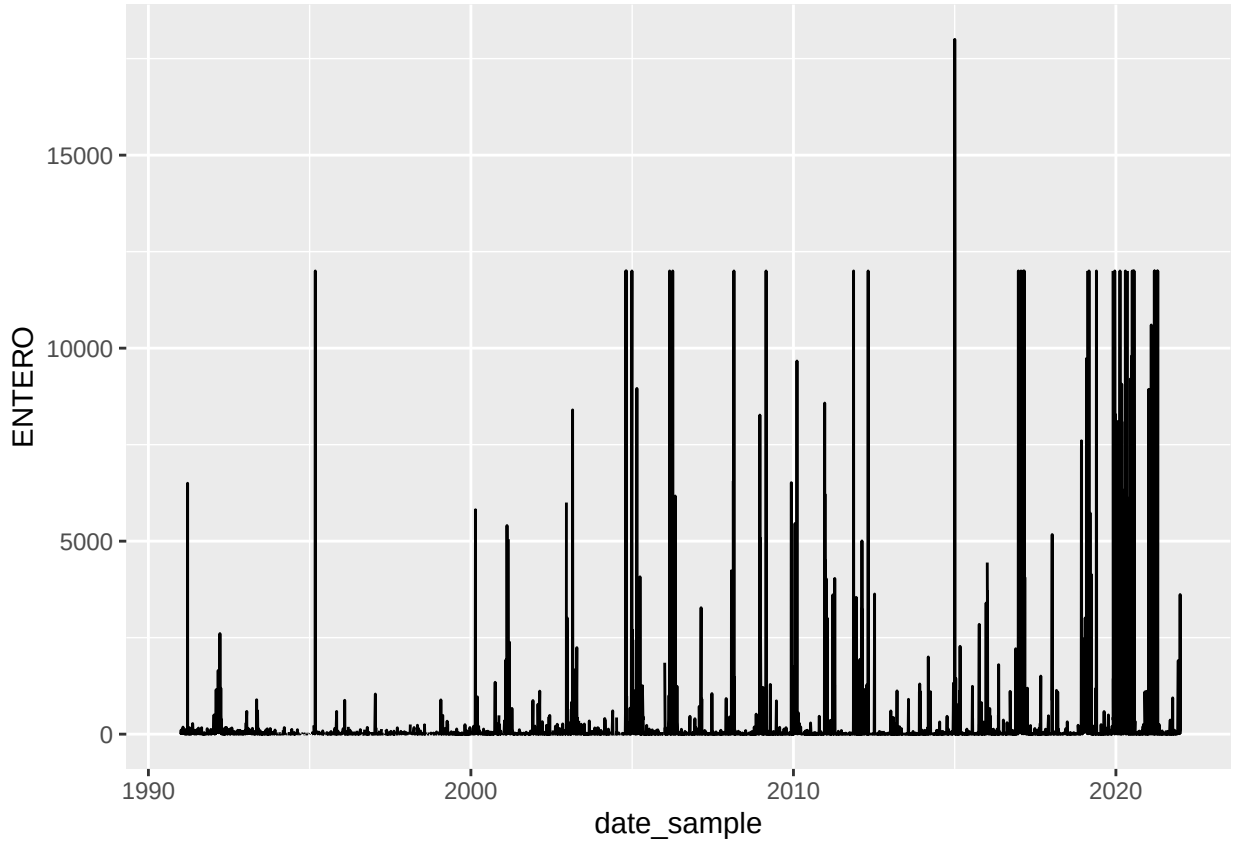
```
#, fig.height=10, fig.width=15
#print(head(owt_df02_gb04_tbbl01$CHLOROPHYLL))
#print(tail(owt_df02_gb04_tbbl01$CHLOROPHYLL))

# Viz time series data for ENTERO parameter
#owt_df02_gb04a <- owt_df02_gb04[owt_df02_gb04$parameter == "ENTERO", ]
# & owt_df02_gb04$station == "A1"
#aps_df01_ts01 <- ts(owt_df02_gb04a$Avg, start = c(1990, 1), freq = 184)
#, start = c(2020, 1), freq = 52
#print(aps_df01_ts01)

#ship_fore_avg <- tslm(aps_df01_ts01 ~ trend)
#ship_fore_trnd <- tslm(aps_df01_ts01 ~ trend + I(trend^2))

#plot(owt_df02_gb04_mrgd$ENTERO,
#     xlab = "Time",
#     ylab = "ENTERO Levels",
#     type = "o",
#     main = "ENTERO Levels Over Time")
#grid()
print(acf(owt_df02_gb04_mrgd$ENTERO, pl=TRUE, na.action = na.pass))
```



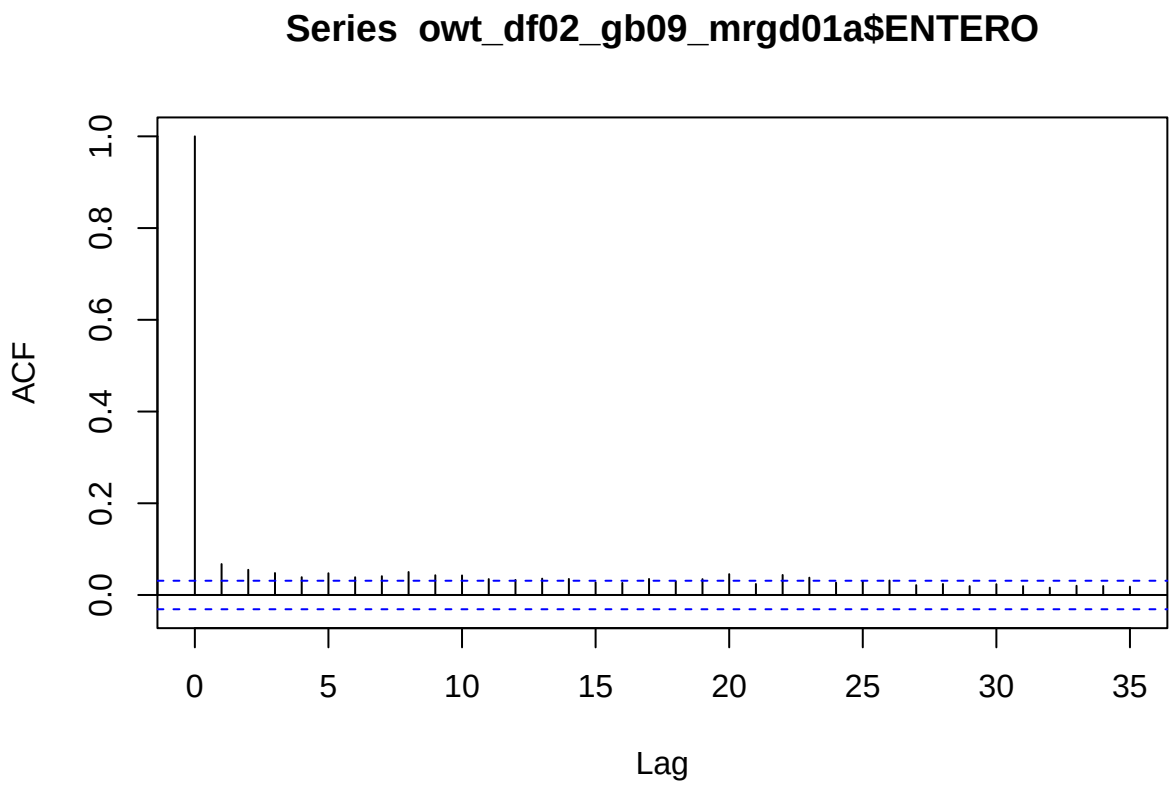

```
##
## Autocorrelations of series 'owt_df02_gb04_mrgd$ENTERO', by lag
##
##     0     1     2     3     4     5     6     7     8     9    10    11    12
## 1.000 0.292 0.153 0.093 0.189 0.262 0.171 0.123 0.150 0.159 0.214 0.155 0.119
##    13    14    15    16    17    18    19    20    21    22    23    24    25
## 0.103 0.149 0.165 0.144 0.101 0.137 0.156 0.148 0.149 0.099 0.100 0.094 0.081
##    26    27    28    29    30    31    32    33    34    35    36    37
## 0.104 0.088 0.088 0.077 0.104 0.083 0.075 0.052 0.058 0.132 0.103 0.100
```

```
# Citation: https://www.geeksforgeeks.org/time-series-visualization-with-ggplot2-in-r/
ggplot(owt_df02_gb04_mrgd, aes(x=date_sample, y=ENTERO, group = 1)) +
  geom_line()
```

```
## Warning: Removed 1 row containing missing values (`geom_line()`).
```



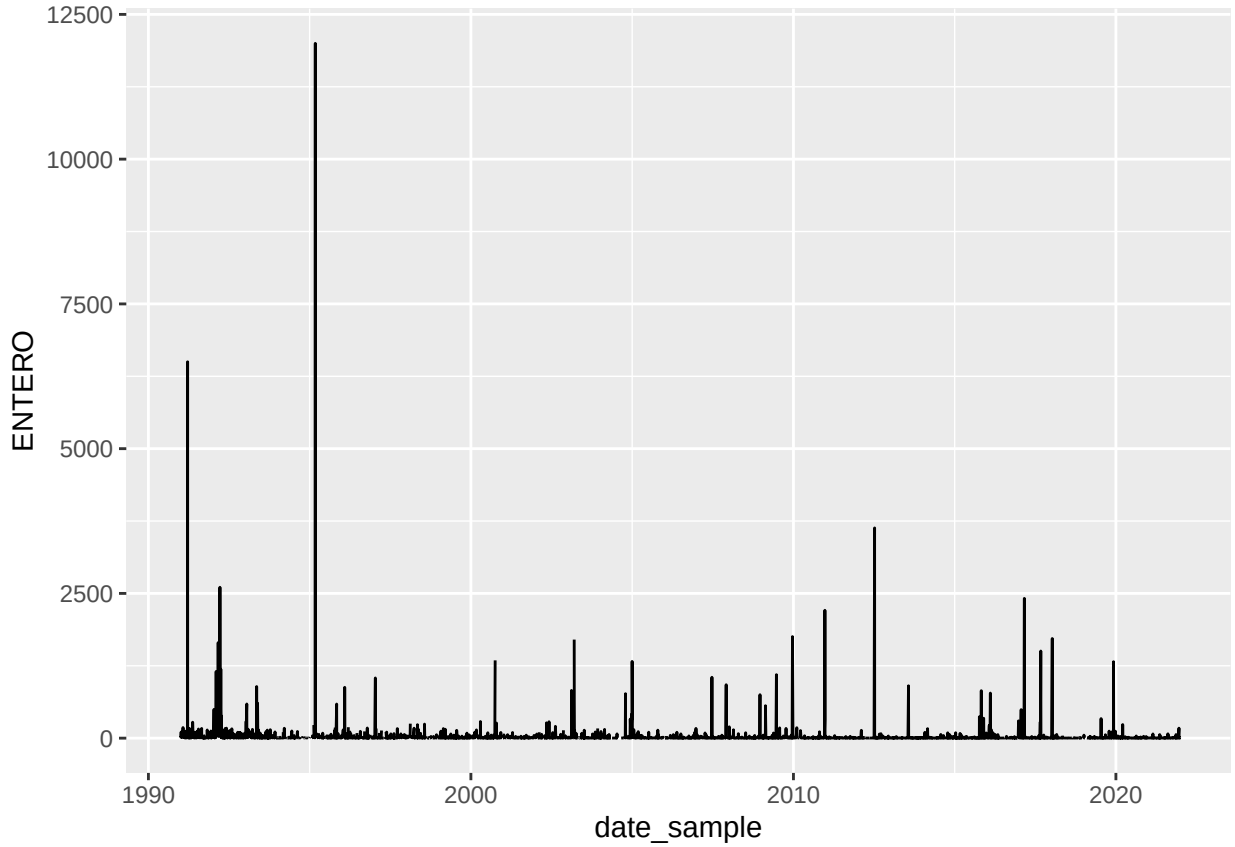

```
owt_df02_gb09_mrgd01a <- owt_df02_gb09_mrgd[owt_df02_gb09_mrgd$project == "PLOO", ]
#plot(owt_df02_gb09_mrgd01a$ENTERO,
#     xlab = "Time",
#     ylab = "ENTERO Levels",
#     main = "ENTERO Levels Over Time: PLOO")
#grid()
print(acf(owt_df02_gb09_mrgd01a$ENTERO, pl=TRUE, na.action = na.pass))
```



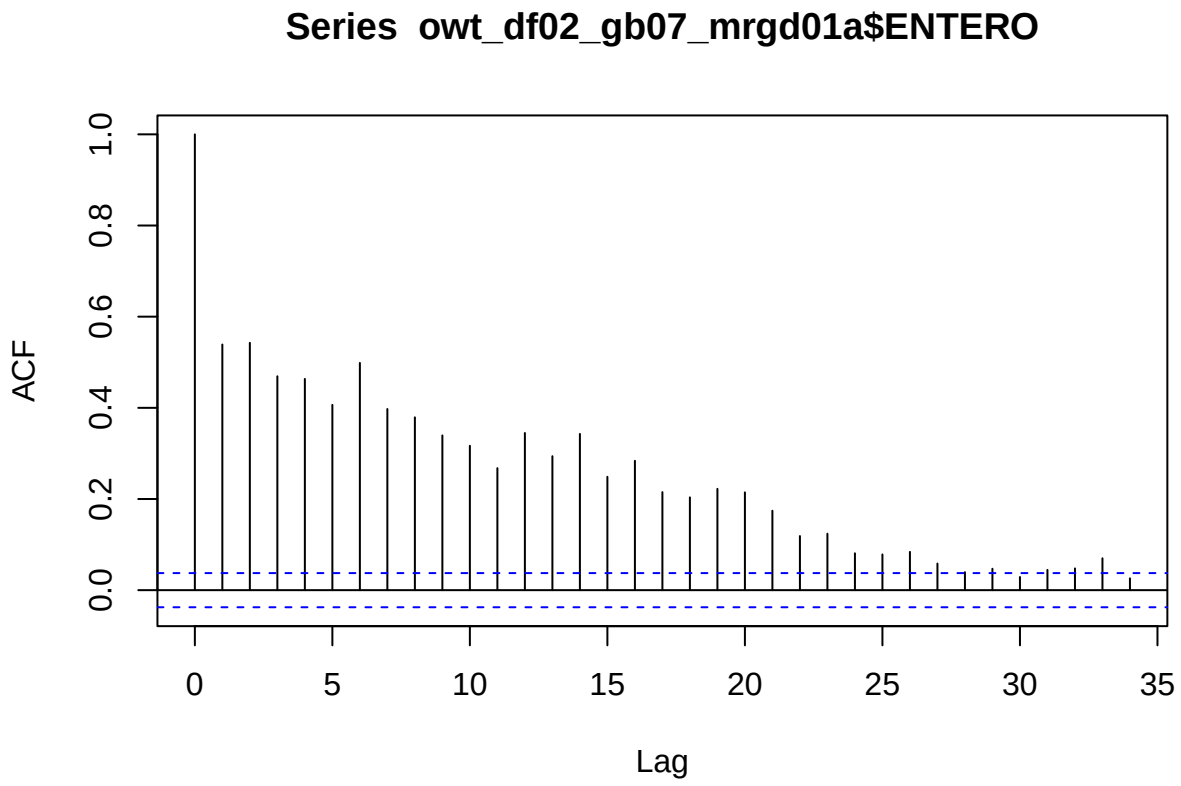

```
##
## Autocorrelations of series 'owt_df02_gb09_mrgd01a$ENTERO', by lag
##
##     0     1     2     3     4     5     6     7     8     9    10    11    12
## 1.000 0.067 0.055 0.048 0.039 0.047 0.039 0.041 0.050 0.043 0.043 0.035 0.033
##    13    14    15    16    17    18    19    20    21    22    23    24    25
## 0.036 0.035 0.028 0.026 0.035 0.030 0.035 0.045 0.024 0.044 0.038 0.027 0.030
##    26    27    28    29    30    31    32    33    34    35
## 0.032 0.021 0.024 0.019 0.023 0.019 0.016 0.020 0.020 0.018
```

```
ggplot(owt_df02_gb09_mrgd01a, aes(x=date_sample, y=ENTERO, group = 1)) +
  geom_line()
```

```
## Warning: Removed 1 row containing missing values (`geom_line()`).
```

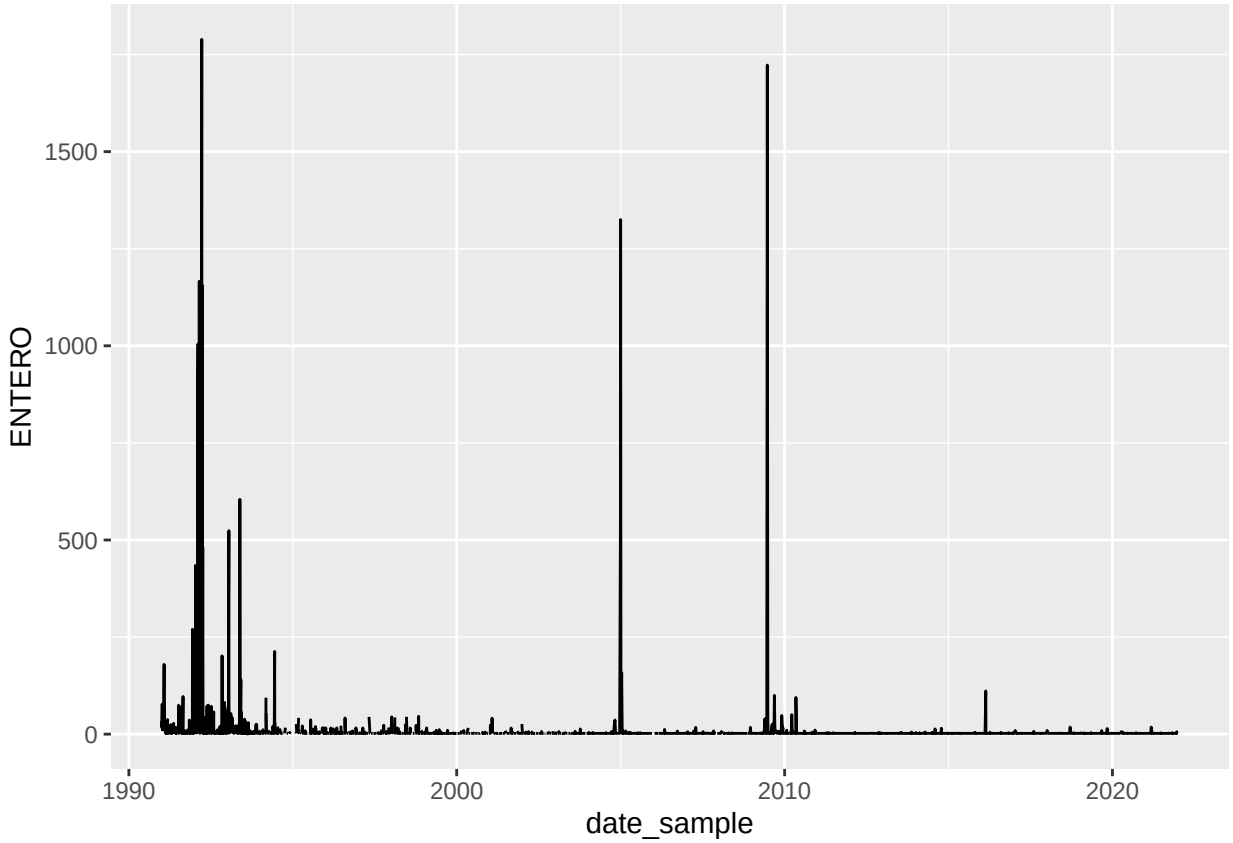



```
#scale_x_discrete(guide = guide_axis(check.overlap = TRUE))

owt_df02_gb07_mrgd01a <- owt_df02_gb07_mrgd[owt_df02_gb07_mrgd$project == "PLOO" &
↪  owt_df02_gb07_mrgd$depth_m_bin == "[0,8)", ]
#plot(owt_df02_gb07_mrgd01a$ENTERO,
#     xlab = "Time",
#     ylab = "ENTERO Levels",
#     type = "o",
#     main = "ENTERO Levels Over Time: PLOO - [0,8)")
#grid()
print(acf(owt_df02_gb07_mrgd01a$ENTERO, pl=TRUE, na.action = na.pass))
```




```
##
## Autocorrelations of series 'owt_df02_gb07_mrgd01a$ENTERO', by lag
##
##     0     1     2     3     4     5     6     7     8     9    10    11    12
## 1.000 0.539 0.543 0.469 0.463 0.407 0.499 0.397 0.379 0.339 0.317 0.268 0.345
##    13    14    15    16    17    18    19    20    21    22    23    24    25
## 0.294 0.343 0.249 0.284 0.215 0.204 0.222 0.215 0.174 0.119 0.124 0.081 0.078
##    26    27    28    29    30    31    32    33    34
## 0.084 0.058 0.040 0.047 0.029 0.044 0.048 0.070 0.026
```

```
ggplot(owt_df02_gb07_mrgd01a, aes(x=date_sample, y=ENTERO, group = 1)) +
  geom_line()
```


ENTERO
1500
1000
500
0
1990
2000
2010
2020
date_sample