SEACAR Nekton Analysis

Last compiled on 02 March, 2023

Contents

Important Notes

These scripts were created by J.E. Panzik for SEACAR.

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

https://github.com/FloridaSEACAR/SEACAR_Panzik

This markdown file is designed to be compiled by SEACAR_Nekton_ReportRender.R.

Libraries and Settings

Loads libraries used in the script. The inclusion of `scipen` option limits how frequently R defaults to scientific notation. Sets default settings for displaying warning and messages in created document, and sets figure dpi.

```
library(knitr)
library(data.table)
library(dplyr)
library(lubridate)
library(ggplot2)
library(scales)
library(tidyr)
library(gridExtra)
#library(tidyverse)
library(ggpubr)
```

```
library(scales)
options(scipen=999)
opts_chunk$set(warning=FALSE, message=FALSE, dpi=200)
```

File Import

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

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

The script then gets the name of the parameter as it appears in the data file and units of the parameter.

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

# Determine what parameter is being used
parameter <- unique(data$ParameterName)
if(parameter=="Presence"){
      parameter <- "Species Richness"
      param_name <- "SpeciesRichness"
}
unit <- unique(data$ParameterUnits)
```

Data Filtering

Documentation on database filtering is provided here: SEACAR Documentation- Analysis Filters and Calculations.docx

The filtering that is performed by the script at this point removes rows that are missing values for `ResultValue` and `EffortCorrection_100m2`, and removes any `EffortCorrection_100m2` that is 0 because it will cause an infinite number when determining Species Richness.

A group of unique `ManagedAreaName`, `ProgramID`, `ProgramName`, `ProgramLocationID`, `SampleDate`, and `GearSize_m` are being considered a "reference" for measurement. For each "reference", the number of observed species is summed and then divided by the `EffortCorrection_100m2`to determine the Species Richness per 100 square meters.

The `ManagedAreaName` values from the data are actually shortened versions, and are merged with the full versions. The species richness data is then written to a file. And the list of Managed Areas with observations is stored.

```
data <- data[!is.na(data$EffortCorrection_100m2),]
data <- data[data$EffortCorrection_100m2!=0,]
data <- data[!is.na(data$ResultValue),]
```

```
data <- data %>%
   group_by(ManagedAreaName, ProgramID, ProgramName, ProgramLocationID,
            SampleDate, GearType, GearSize_m) %>%
   summarise(ParameterName=parameter,
             Year=unique(Year), Month=unique(Month),
             N_Species=sum(ResultValue),
             EffortCorrection_100m2=unique(EffortCorrection_100m2),
             SpeciesRichness=N_Species/unique(EffortCorrection_100m2))

#setnames(data, c("ManagedAreaName"), c("ShortName"))

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

#data$ShortName <- NULL

fwrite(data, paste0(out_dir,"/Nekton_", param_name, "_UsedData.txt"), sep="|")

data$SampleDate <- as.Date(data$SampleDate)

MA_Include <- unique(data$ManagedAreaName[!is.na(data$N_Species)])
MA_Include <- MA_Include[order(MA_Include)]
n <- length(MA_Include)
```

Managed Area Statistics

Gets summary statistics for each managed area. Uses piping from dplyr package to feed into subsequent steps. The following steps are performed:

1. Group data that have the same `ManagedAreaName`, `Year`, `Month`, `GearType`, and `GearSize_m`.
   - Second summary statistics do not use the `Month` grouping and are only for `ManagedAreaName` and `Year`.
   - Third summary statistics do not use `Year` grouping and are only for `ManagedAreaName` and `Month`
2. For each group, provide the following information: Parameter Name (ParameterName), Number of Entries (N_Data), Lowest Value (Min), Largest Value (Max), Median, Mean, Standard Deviation, and a list of all Program IDs included in these measurements.
3. Sort the data in ascending (A to Z and 0 to 9) order based on `ManagedAreaName` then `Year` then `Month`
4. Write summary stats to a pipe-delimited .txt file in the output directory
   - Output Files in SEACAR GitHub

```
MA_YM_Stats <- data %>%
   group_by(AreaID, ManagedAreaName, Year, Month, GearType, GearSize_m) %>%
   summarize(ParameterName=parameter,
             N_Data=length(na.omit(SpeciesRichness)),
             Min=min(SpeciesRichness),
             Max=max(SpeciesRichness),
             Median=median(SpeciesRichness),
             Mean=mean(SpeciesRichness),
```

```
            StandardDeviation=sd(SpeciesRichness),
            ProgramIDs=paste(sort(unique(ProgramID), decreasing=FALSE),
                             collapse=', '))
MA_YM_Stats <- as.data.table(MA_YM_Stats[order(MA_YM_Stats$ManagedAreaName,
                                               MA_YM_Stats$Year,
                                               MA_YM_Stats$Month,
                                               MA_YM_Stats$GearSize_m), ])
fwrite(MA_YM_Stats, paste0(out_dir,"/Nekton_", param_name,
                           "_ManagedArea_YearMonth_Stats.txt"), sep="|")
rm(MA_YM_Stats)

MA_Y_Stats <- data %>%
   group_by(AreaID, ManagedAreaName, Year, GearType, GearSize_m) %>%
   summarize(ParameterName=parameter,
            N_Data=length(na.omit(SpeciesRichness)),
            Min=min(SpeciesRichness),
            Max=max(SpeciesRichness),
            Median=median(SpeciesRichness),
            Mean=mean(SpeciesRichness),
            StandardDeviation=sd(SpeciesRichness),
            ProgramIDs=paste(sort(unique(ProgramID), decreasing=FALSE),
                             collapse=', '))
MA_Y_Stats <- as.data.table(MA_Y_Stats[order(MA_Y_Stats$ManagedAreaName,
                                             MA_Y_Stats$Year,
                                             MA_Y_Stats$GearSize_m), ])
fwrite(MA_Y_Stats, paste0(out_dir,"/Nekton_", param_name,
                          "_ManagedArea_Year_Stats.txt"), sep="|")

MA_M_Stats <- data %>%
   group_by(AreaID, ManagedAreaName, Month, GearType, GearSize_m) %>%
   summarize(ParameterName=parameter,
            N_Data=length(na.omit(SpeciesRichness)),
            Min=min(SpeciesRichness),
            Max=max(SpeciesRichness),
            Median=median(SpeciesRichness),
            Mean=mean(SpeciesRichness),
            StandardDeviation=sd(SpeciesRichness),
            ProgramIDs=paste(sort(unique(ProgramID), decreasing=FALSE),
                             collapse=', '))
MA_M_Stats <- as.data.table(MA_M_Stats[order(MA_M_Stats$ManagedAreaName,
                                             MA_M_Stats$Month,
                                             MA_M_Stats$GearSize_m), ])
fwrite(MA_M_Stats, paste0(out_dir,"/Nekton_", param_name,
                          "_ManagedArea_Month_Stats.txt"), sep="|")
rm(MA_M_Stats)

MA_Ov_Stats <- data %>%
   group_by(AreaID, ManagedAreaName, GearType, GearSize_m) %>%
   summarize(ParameterName=parameter,
            N_Years=length(unique(na.omit(Year))),
            EarliestYear=min(Year),
            LatestYear=max(Year),
            N_Data=length(na.omit(SpeciesRichness)),
```

```
            Min=min(SpeciesRichness),
            Max=max(SpeciesRichness),
            Median=median(SpeciesRichness),
            Mean=mean(SpeciesRichness),
            StandardDeviation=sd(SpeciesRichness),
            ProgramIDs=paste(sort(unique(ProgramID), decreasing=FALSE),
                             collapse=', '))
MA_Ov_Stats <- as.data.table(MA_Ov_Stats[order(MA_Ov_Stats$ManagedAreaName,
                                               MA_Ov_Stats$GearSize_m), ])
MA_Ov_Stats$Year_MinRichness <- NA
MA_Ov_Stats$Year_MaxRichness <- NA

for(m in 1:nrow(MA_Ov_Stats)){
   ma <- MA_Ov_Stats$ManagedAreaName[m]
   gear <- MA_Ov_Stats$GearType[m]
   size <- MA_Ov_Stats$GearSize_m[m]
   if(MA_Ov_Stats$N_Data[m]==0){
      next
   }
   ds <- MA_Y_Stats[MA_Y_Stats$ManagedAreaName==ma &
                       MA_Y_Stats$GearType==gear &
                       MA_Y_Stats$GearSize_m==size,]
   min <- min(ds$Mean)
   max <- max(ds$Mean)
   year_min <- ds$Year[ds$Mean==min]
   year_max <- ds$Year[ds$Mean==max]
   MA_Ov_Stats$Year_MinRichness[m] <- year_min
   MA_Ov_Stats$Year_MaxRichness[m] <- year_max
}
MA_Ov_Stats$ProgramIDs <- gsub("", NA, MA_Ov_Stats$ProgramIDs)
fwrite(MA_Ov_Stats, paste0(out_dir,"/Nekton_", param_name,
                           "_ManagedArea_Overall_Stats.txt"), sep="|")
MA_Ov_Stats <- MA_Ov_Stats[!is.na(MA_Ov_Stats$EarliestYear), ]
```

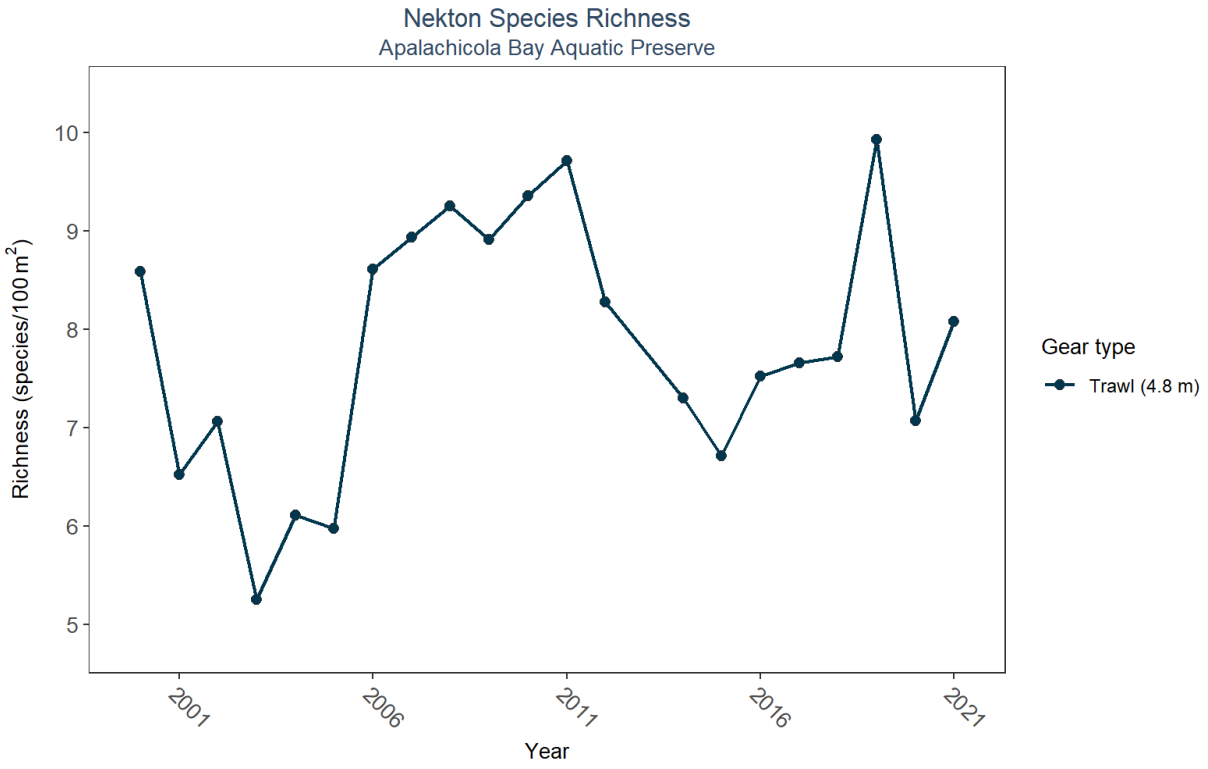
Appendix I: Managed Area Species Richness

The plots shown here are the species richness for each managed area with a yearly average, separated by gear size.

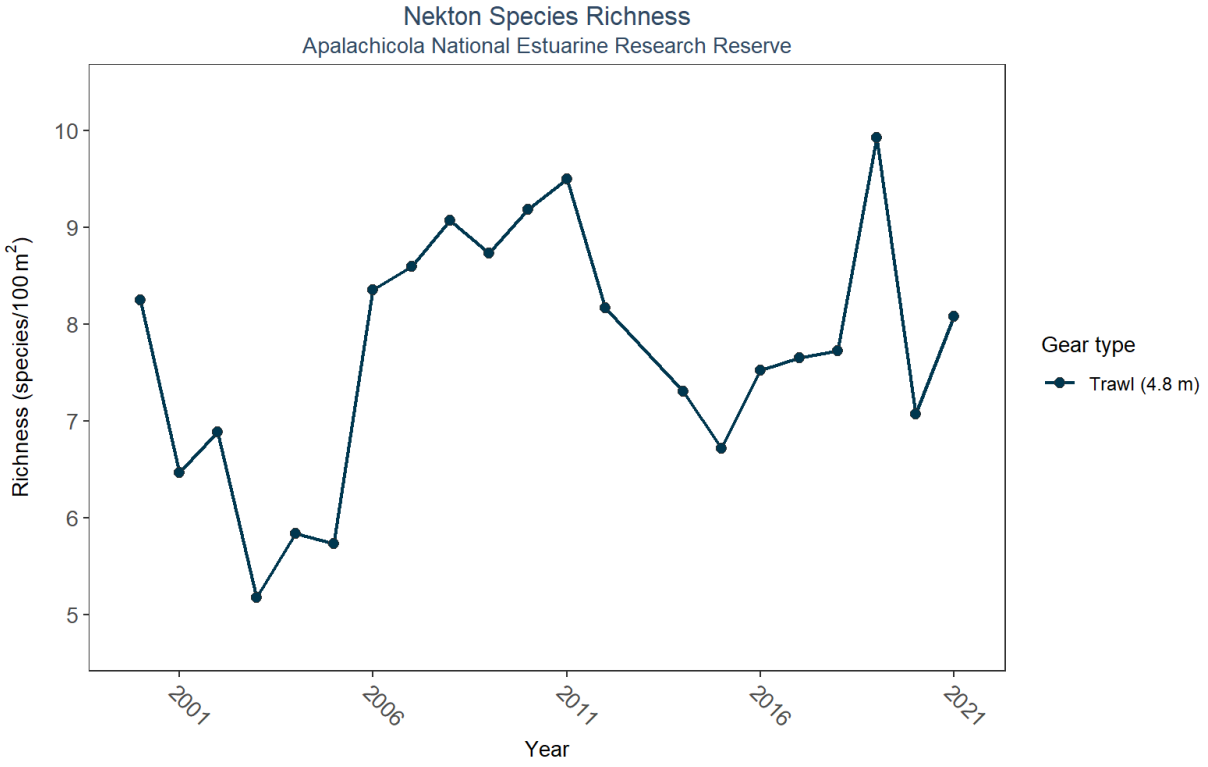
1. Set common plot theme.
2. Determine the earliest and latest year of the data to create x-axis scale and intervals
3. determine the upper and lower limit of the plot for better y-axis labels
4. Determines what gear types are present and adjusts legend entries
5. Add the plot line
6. Set the plot type as a point plot with the size of the points
7. Create the title, x-axis, y-axis, and color fill labels
8. Set the y and x limits
9. Apply common plot theme
10. Create file name to save figure
11. Save figure as png file

```
plot_theme <- theme_bw() +
      theme(panel.grid.major = element_blank(),
            panel.grid.minor = element_blank(),
            text=element_text(family="Arial"),
            #title=element_text(face="bold"),
            plot.title=element_text(hjust=0.5, size=12, color="#314963"),
            plot.subtitle=element_text(hjust=0.5, size=10, color="#314963"),
            legend.title=element_text(size=10),
            legend.text.align = 0,
            axis.title.x = element_text(size=10, margin = margin(t = 5, r = 0,
                                                                 b = 10, l = 0)),
            axis.title.y = element_text(size=10, margin = margin(t = 0, r = 10,
                                                                 b = 0, l = 0)),
            axis.text=element_text(size=10),
            axis.text.x=element_text(angle = -45, hjust = 0))

gear_colors <- c("Trawl (4.8 m)"="#00374f",
                 "Trawl (6.1 m)"="#007c99",
                 "Seine (183 m)"="#00c9db")

gear_shapes <- c("Trawl (4.8 m)"=21,
                 "Trawl (6.1 m)"=22,
                 "Seine (183 m)"=24)

if(n==0){
      print("There are no monitoring locations that qualify.")
} else {
      for (i in 1:n) {
            plot_data <- MA_Y_Stats[MA_Y_Stats$ManagedAreaName==MA_Include[i]]
            plot_data$GearType_Plot <- paste0(plot_data$GearType, " (",
                                              plot_data$GearSize_m, " m)")
            t_max <- max(MA_Ov_Stats$LatestYear[MA_Ov_Stats$ManagedAreaName==
                                                  MA_Include[i]])
            t_min <- min(MA_Ov_Stats$EarliestYear[MA_Ov_Stats$ManagedAreaName==
                                                    MA_Include[i]])
            t <- t_max-t_min


            if(t>=30){
                  brk <- -10
            }else if(t<30 & t>=10){
                  brk <- -5
            }else if(t<10 & t>=4){
                  brk <- -2
            }else if(t<4){
                  brk <- -1
            }

            y_range <- max(plot_data$Mean) - min(plot_data$Mean)

            y_min <- if(min(plot_data$Mean)-(0.1*y_range)<0){
               y_min <- 0
            } else {
```

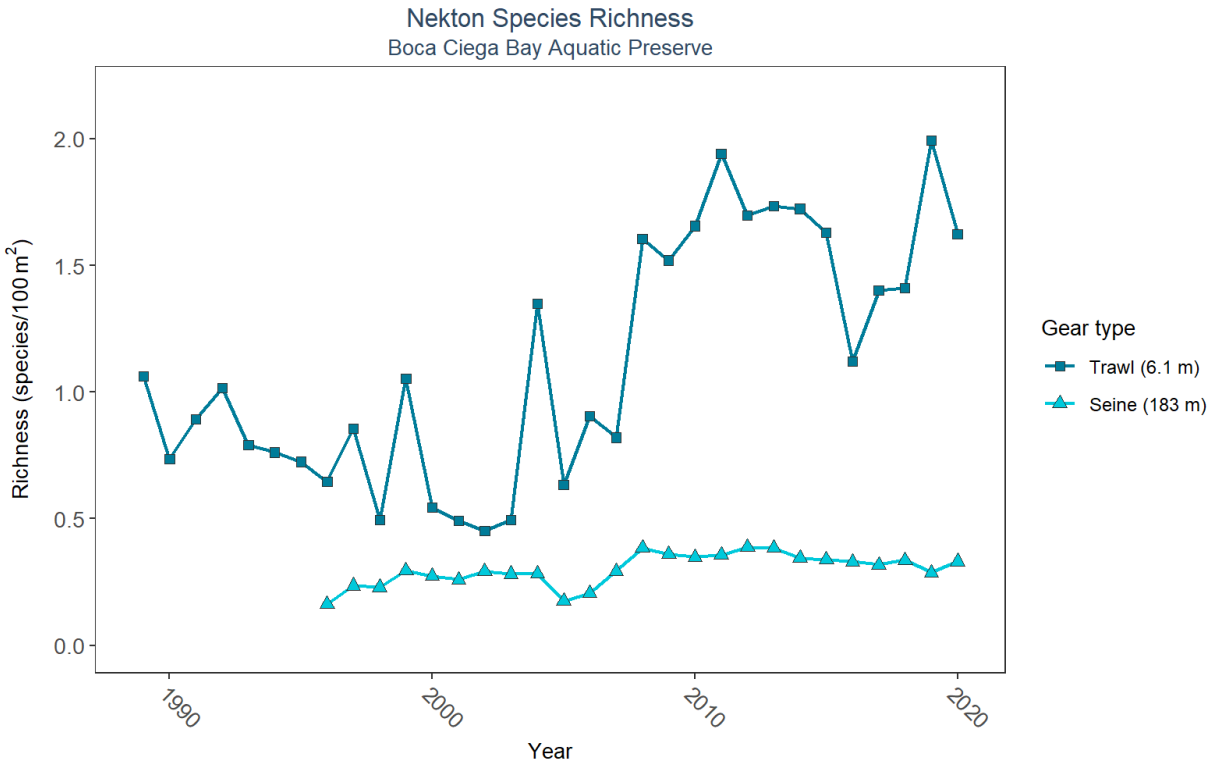
```
            y_min <- min(plot_data$Mean)-(0.1*y_range)
        }

        y_max <- max(plot_data$Mean)+(0.1*y_range)

        gear_colors_plot <- gear_colors[unique(plot_data$GearType_Plot)]
        gear_shapes_plot <- gear_shapes[unique(plot_data$GearType_Plot)]

        p1 <- ggplot(data=plot_data, group=as.factor(GearType_Plot)) +
          geom_line(aes(x=Year, y=Mean, color=as.factor(GearType_Plot)),
                    size=0.75, alpha=1) +
          geom_point(aes(x=Year, y=Mean, fill=as.factor(GearType_Plot),
                         shape=as.factor(GearType_Plot)), size=2,
                     color="#333333", alpha=1) +
          labs(title="Nekton Species Richness",
               subtitle=MA_Include[i],
               x="Year", y=bquote('Richness (species/100'*~m^{2}*')'),
               fill="Gear type", color="Gear type", shape="Gear type") +
          scale_x_continuous(limits=c(t_min-0.25, t_max+0.25),
                             breaks=seq(t_max, t_min, brk)) +
          scale_y_continuous(limits=c(y_min, y_max),
                             breaks=pretty_breaks(n=5)) +
          scale_fill_manual(values=gear_colors_plot) +
          scale_color_manual(values=gear_colors_plot) +
          scale_shape_manual(values=gear_shapes_plot) +
          plot_theme

        outname <- paste0("Nekton_", gsub(" ", "", MA_Include[i]), "_",
                          param_name, ".png")
        png(paste0(out_dir, "/Figures/", outname),
            width = 8,
            height = 4,
            units = "in",
            res = 200)
        #rm(plot_data)

        print(p1)
        dev.off()


        ResultTable <- MA_Ov_Stats[MA_Ov_Stats$ManagedAreaName==MA_Include[i],]
        ResultTable <- ResultTable[,-c("AreaID", "ManagedAreaName",
                                       "ProgramIDs", "GearType_Plot")]
        t1 <- ggtexttable(ResultTable, rows = NULL,
                          theme=ttheme(base_size=7))
        print(ggarrange(p1, t1, ncol=1, heights=c(0.85, 0.15)))

        cat("\n \n \n \n")
    }
}
```



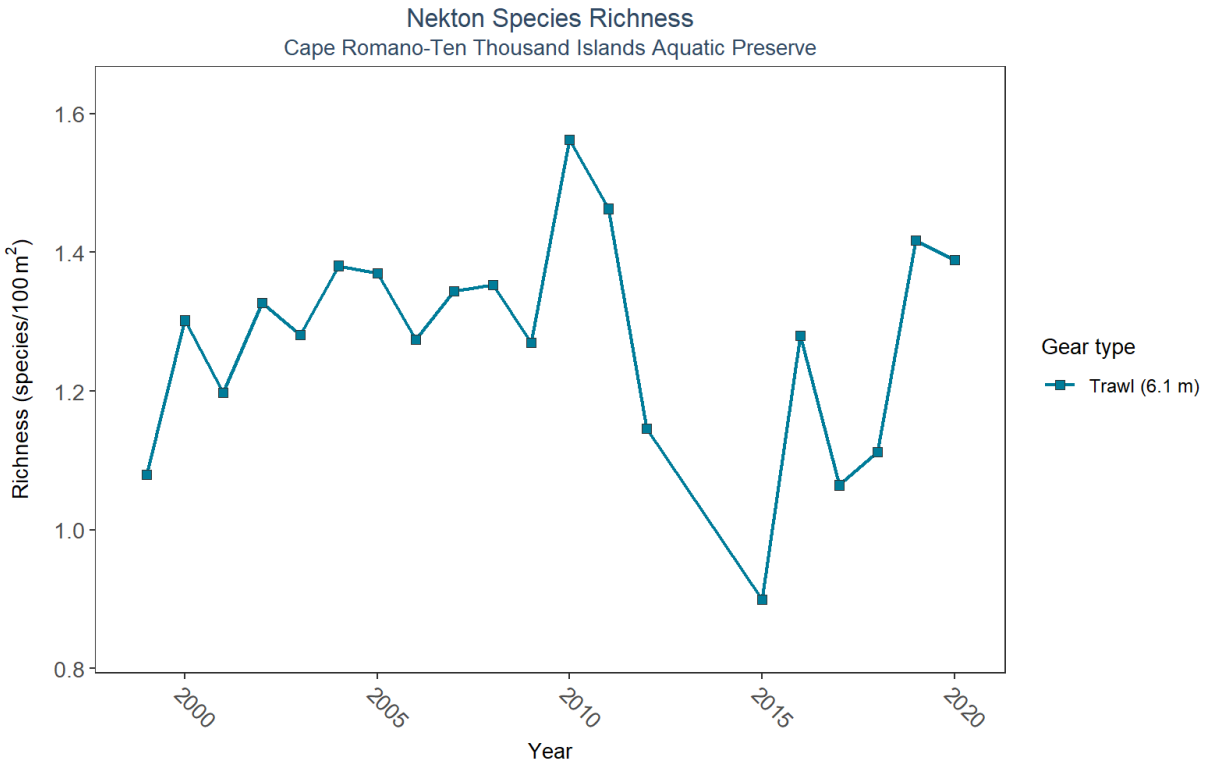
e_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
	Species Richness	21	2000	2021	1664	0.1851852	24.81481	7.777778	7.790687	3.621108	2003



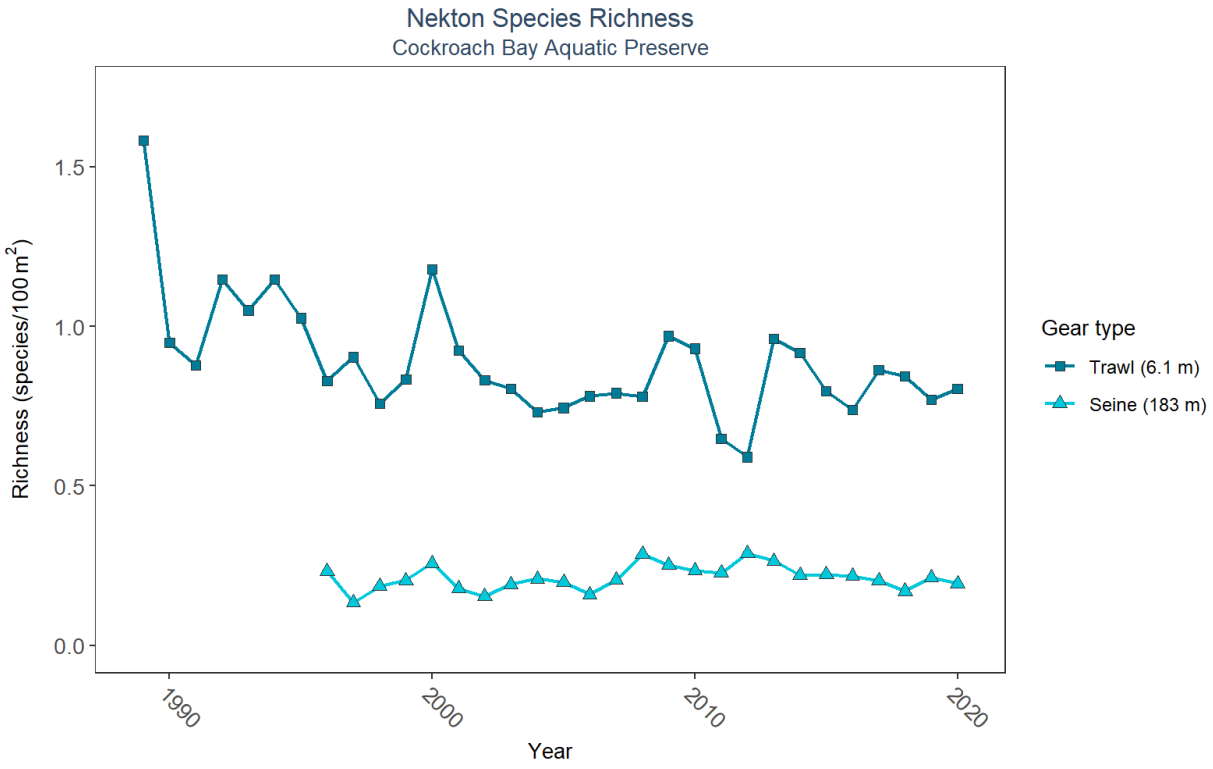
ze_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
)	Species Richness	21	2000	2021	1931	0.1851852	24.81481	7.407407	7.62424	3.724814	2003



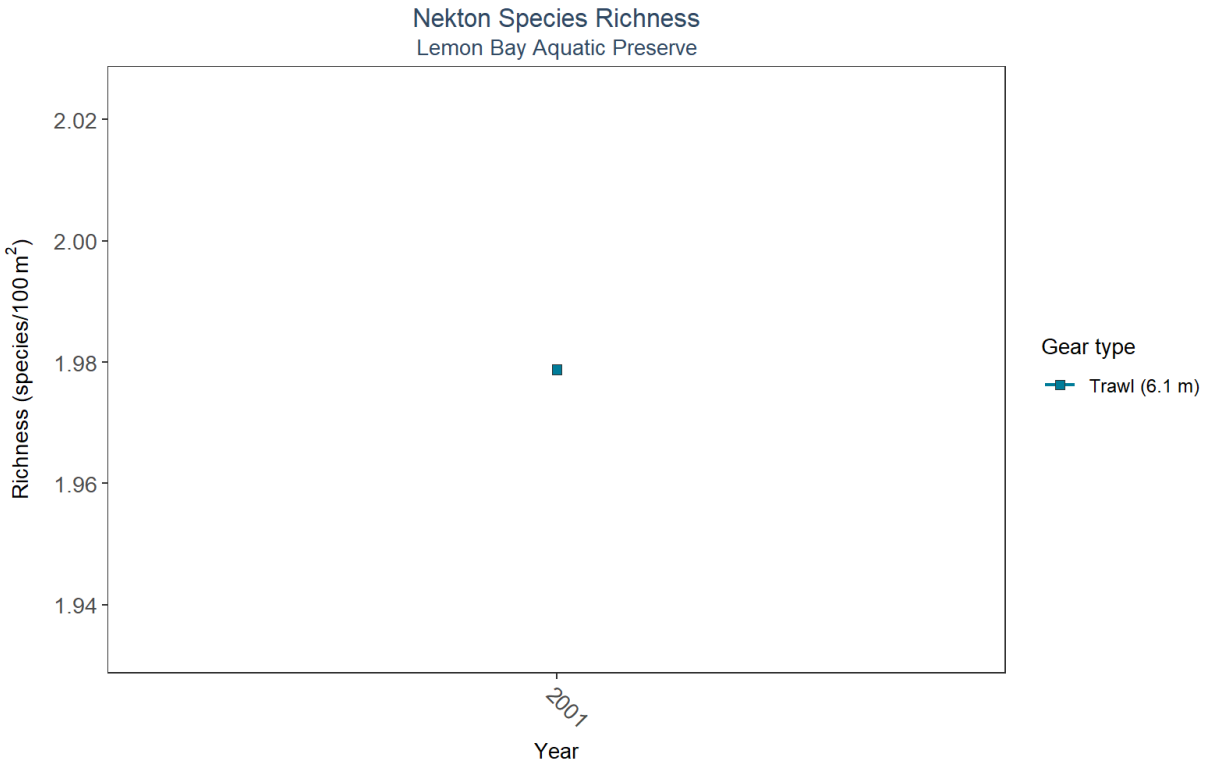
ize_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
.1	Species Richness	32	1989	2020	868	0	4.5871560	1.0118726	1.3286098	0.9235100	2002
3.0	Species Richness	25	1996	2020	997	0	0.9223301	0.2912621	0.3142437	0.1651043	1996



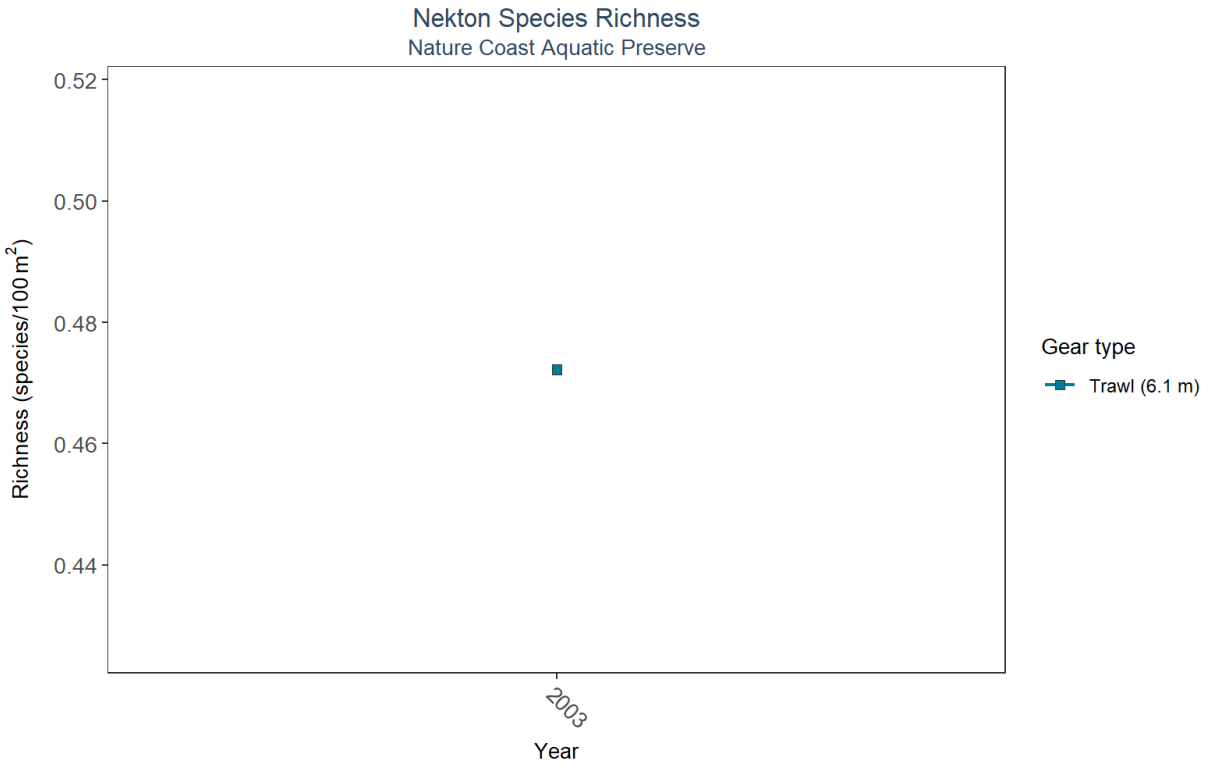
Size_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness	Y
6.1	Species Richness	20	1999	2020	2555	0	3.372909	1.349164	1.305283	0.5310301	2015	



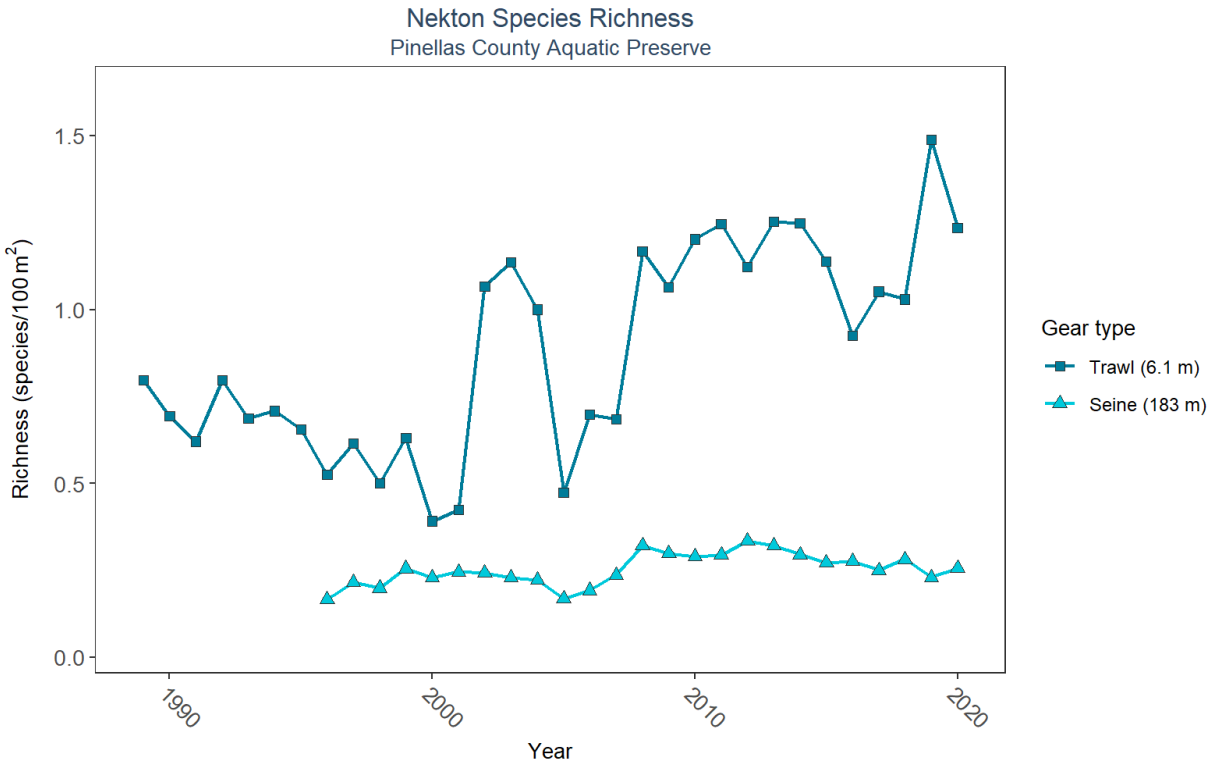
ize_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
.1	Species Richness	32	1989	2020	1666	0	3.747676	0.7495353	0.8377281	0.5409541	2012
3.0	Species Richness	25	1996	2020	476	0	0.776699	0.1941748	0.2139594	0.1190414	1997



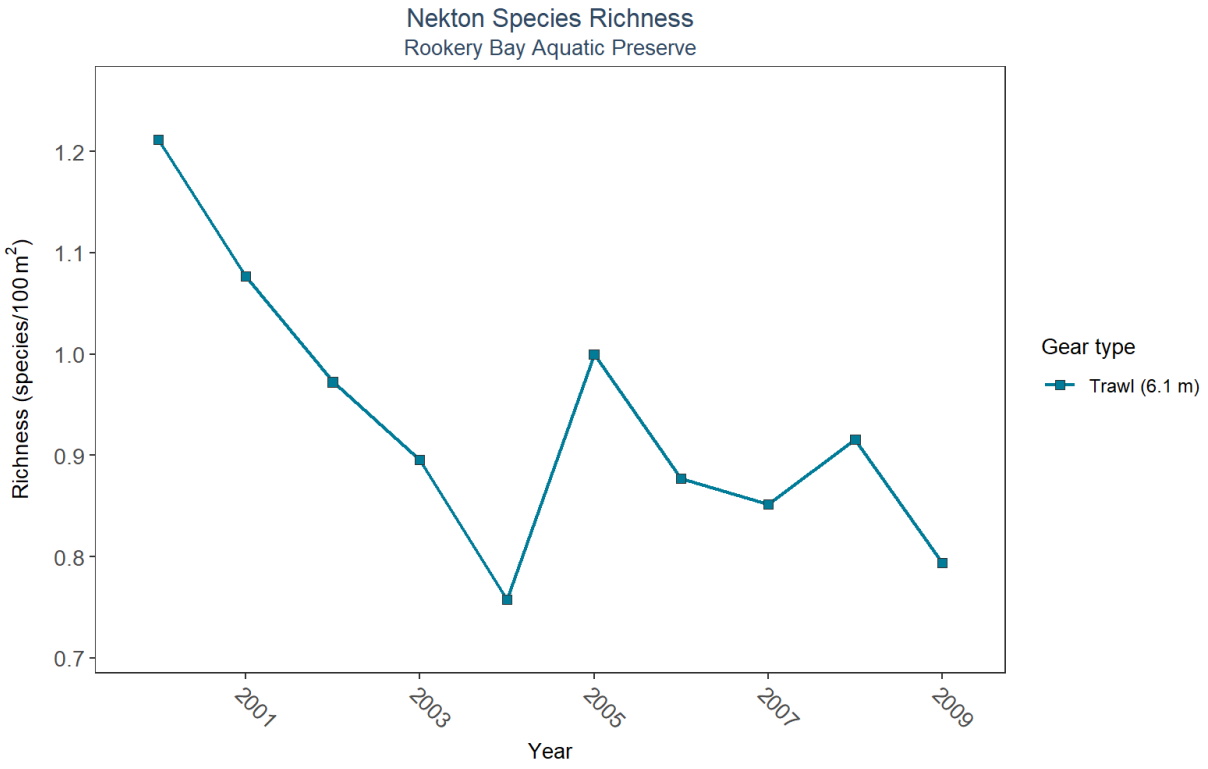
ize_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
1	Species Richness	1	2001	2001	3	1.48408	2.428494	2.023745	1.978773	0.4738107	2001



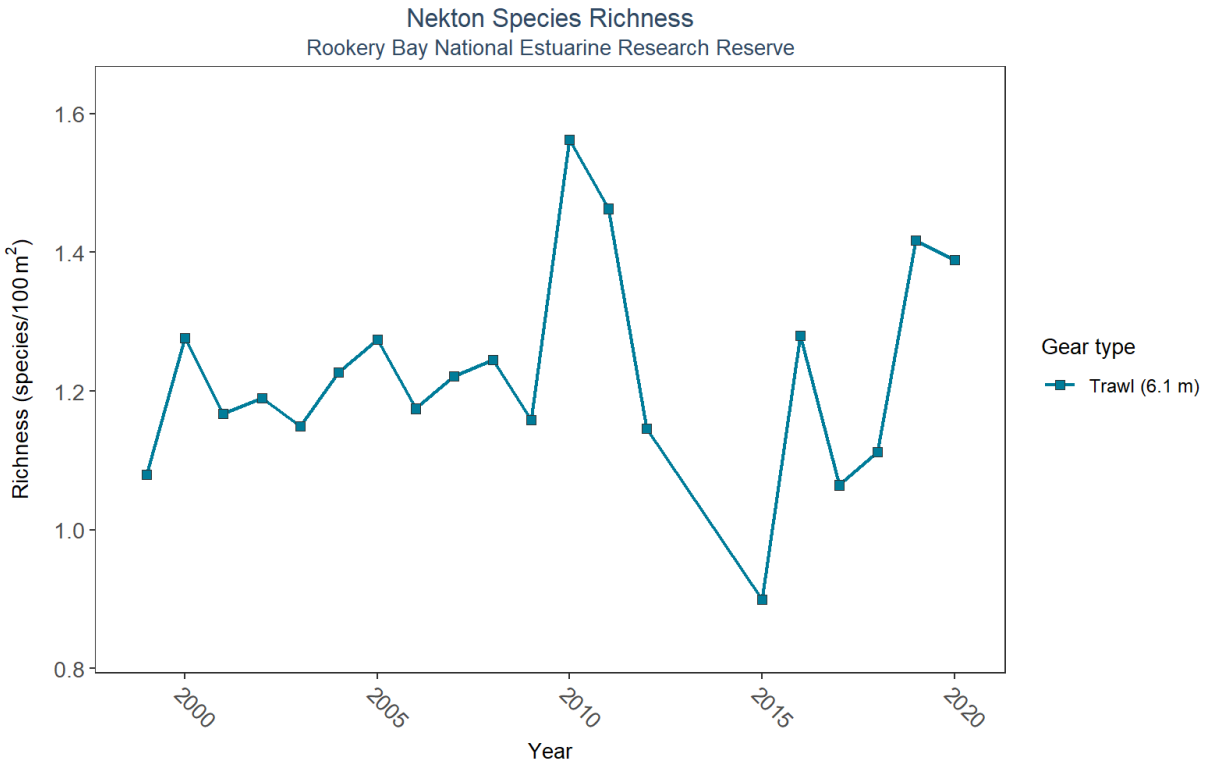
ize_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
1	Species Richness	1	2003	2003	2	0	0.9444145	0.4722072	0.4722072	0.6678019	2003



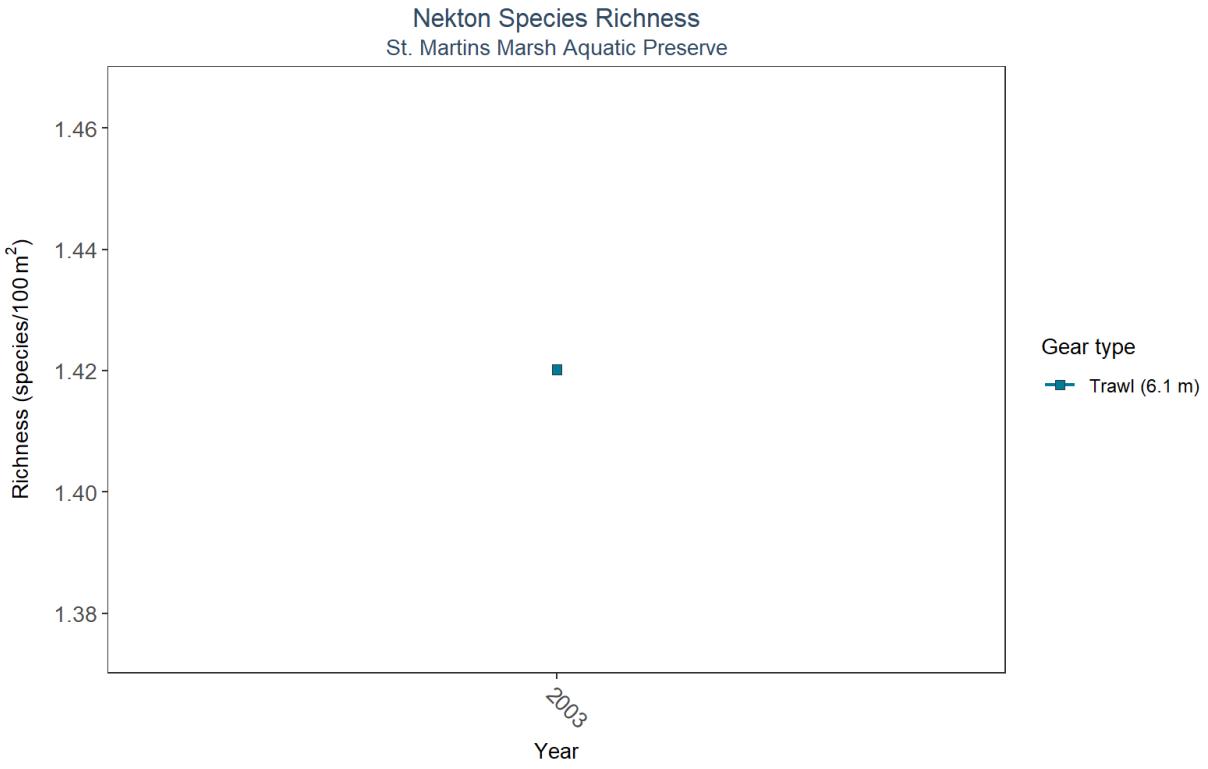
ize_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
.1	Species Richness	32	1989	2020	2903	0	6.2061522	0.7100861	0.9508383	0.7793933	2000
3.0	Species Richness	25	1996	2020	2621	0	0.9223301	0.2184466	0.2576835	0.1548152	1996



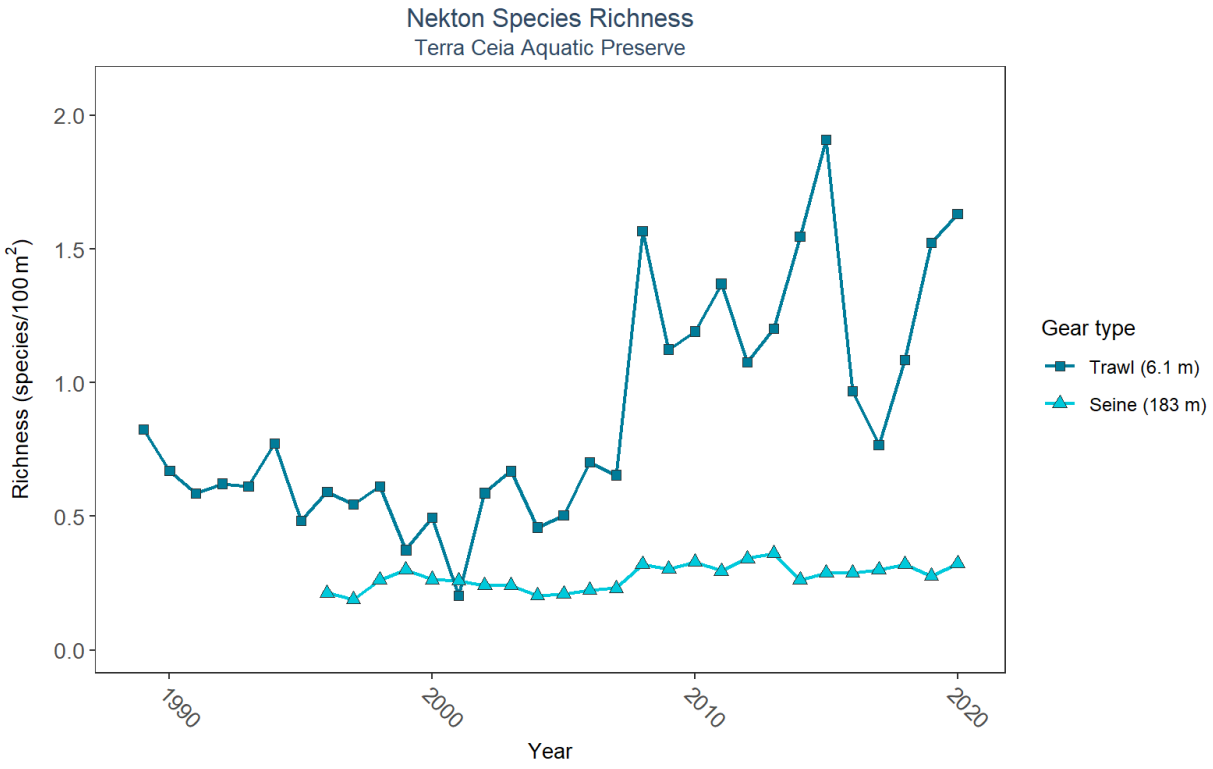
ize_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness	Y
.1	Species Richness	10	2000	2009	535	0	2.563411	0.9444145	0.9368491	0.4215923	2004	



Size_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness	Y
6.1	Species Richness	20	1999	2020	3098	0	3.372909	1.214247	1.240812	0.5320749	2015	



ze_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
	Species Richness	1	2003	2003	1	1.420172	1.420172	1.420172	1.420172	NA	2003



ize_m	ParameterName	N_Years	EarliestYear	LatestYear	N_Data	Min	Max	Median	Mean	StandardDeviation	Year_MinRichness
.1	Species Richness	32	1989	2020	709	0	4.2161360	0.7420399	1.0343022	0.8365280	2001
3.0	Species Richness	25	1996	2020	921	0	0.7524272	0.2669903	0.2750809	0.1325194	1997