SEACAR Coastal Wetlands Analysis

Last compiled on 29 June, 2022

# Important Notes

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

<https://github.com/FloridaSEACAR/SEACAR_Panzik>

This script is based off of code originally wrtten by Katie May Laumann

# Libraries and Settings

Loads libraries used in the script. Loads the Segoe UI font for use in the figures. 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(hrbrthemes)  
windowsFonts(`Segoe UI` = windowsFont('Segoe UI'))  
options(scipen=999)  
opts\_chunk$set(warning=FALSE, message=FALSE, dpi=200)

# File Import

Imports file that is determined in the SEACAR\_CoastalWetlands\_ReportRender.R script.

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

wetlands <- read.csv(file\_in)  
colnames(wetlands)

## [1] "RowID" "ProgramID"   
## [3] "ProgramName" "ProgramLocationID"   
## [5] "QuadIdentifier" "CommonIdentifier"   
## [7] "SpeciesName" "GenusName"   
## [9] "SpeciesGroup1" "SpeciesGroup2"   
## [11] "CLC\_Num" "LC\_Name"   
## [13] "SampleDate" "Transect\_M"   
## [15] "QuadSize\_m2" "ManagedArea"   
## [17] "Region" "CHIMMP"   
## [19] "Year" "Month"   
## [21] "DataFileName" "MADup"   
## [23] "X.PercentCover.SpeciesComposition\_.." "X.StemDensity\_..m2."   
## [25] "X.Total.CanopyPercentCover.SpeciesComposition\_.." "X.BasalArea\_m2.ha."

# Data Filtering

The processing and filtering that is done to the data is as follows:

1. X.PercentCover.SpeciesComposition\_.. column is renamed perccov
2. Removes data that contains NA values in GenusName, SpeciesName, Month, Year, SpeciesGroup1, and removes invasive species data
3. Excludes data from program 5015 and 651
4. Removes duplicates (MADup==1) and where percent cover is 0
5. Combines genus and species names
6. Makes month and year numeric values
7. Makes a combined year+month column where all months have 2 digits.
8. Summarize data by managed area and remove managed areas with less than 5 years
9. Include only data with more than 5 years

#colnames(wetlands) [23]<-"perccov"  
colnames(wetlands)[colnames(wetlands) == "X.PercentCover.SpeciesComposition\_.."] <- "perccov"  
  
#remove NAs from Genus and Species. This is for spp level analysis.   
#\*\*\*\*\*can do this for group level analysis, but don't need to. All genus NA's are also NA in 'Group' columns  
wetlands<-  
 wetlands%>%  
 filter(GenusName!="NA")  
  
wetlands<-  
 wetlands%>%  
 filter(SpeciesName!="NA")  
  
wetlands<-  
 wetlands%>%  
 filter(Month!="NA")  
  
wetlands<-  
 wetlands%>%  
 filter(Year!="NA")  
  
wetlands<-  
 wetlands%>%  
 filter(SpeciesGroup1!="NA")  
  
wetlands<-  
 wetlands%>%  
 filter(SpeciesGroup1!="Invasive")  
  
#You may decide to remove certain programs, for example here we remove   
#program 5015, which only looks at a specific site, and program 651, at request of SMEs  
wetlands<-  
 wetlands%>%  
 filter(ProgramID!="5015")  
  
wetlands<-  
 wetlands%>%  
 filter(ProgramID!="651")  
  
  
#remove duplicates and percent cover 0s  
wetlands<-  
 wetlands%>%  
 filter(MADup==1)  
  
wetlands<-  
 wetlands%>%  
 filter(perccov!=0)  
  
#bring together Genus and species in one column so we have correct species names  
wetlands$gensp<-paste(wetlands$GenusName, wetlands$SpeciesName, sep=" ")  
  
wetlands$mth<-as.numeric(wetlands$Month)  
  
wetlands$yr<-as.numeric(wetlands$Year)  
  
#make a yearmonth column so we can analyze at that level  
wetlands$yearmonth<-ifelse(wetlands$mth==10|wetlands$mth==11|wetlands$mth==12, (paste(wetlands$yr, wetlands$mth, sep=".")),  
 (paste(wetlands$yr, wetlands$mth, sep=".0")))  
  
#correct misspellings in the dataset  
wetlands$gensp2<-ifelse(wetlands$gensp=="Sarcocornia ambigua","Salicornia ambigua",paste(wetlands$gensp))  
wetlands$gensp<-wetlands$gensp2  
  
wetlands$SpeciesGroup1<-ifelse(wetlands$SpeciesGroup1=="Marsh succulents " ,"Marsh succulents",paste(wetlands$SpeciesGroup1))  
  
#some Managed areas have the same spp counted by different programs in different years-this is not an issue for spp richness  
#but will be important when we move on to % cover  
  
wetlandsMA<-unique(wetlands[,c("Year","gensp","ManagedArea","yearmonth","Month","ProgramID")])  
  
#remove program ID if you determine it does not sig impact spp rich  
#How many Managed Areas do we have?  
wetlandsMA<-  
 wetlandsMA%>%  
 filter(ManagedArea!="NA")  
  
#remove Managed Areas that don't have at least 5 years of data.   
##First, ID which ones have fewer than 5 yrs of data  
nyearsMA<-  
 wetlandsMA%>%  
 group\_by(ManagedArea)%>%  
 summarise(length(unique(Year)))  
colnames(nyearsMA) [2]<-"nyrsMA"  
  
#Store names of Managed areas that have <5 years of data in the object "eliminateMA"   
eliminateMA<-  
 nyearsMA%>%  
 filter(nyrsMA<5)  
  
#eliminate those MAs from Analyses using antijoin  
#use anti\_join to include only programs>=5 yrs of data  
wetlandsMA<-anti\_join(wetlandsMA,eliminateMA,by=c("ManagedArea"))

# Species Richness

Gets species richness 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, and Month.
2. Write species richness to a.csv file in the output directory
   * [Output Files in SEACAR GitHub](https://github.com/FloridaSEACAR/SEACAR_Panzik)

wetlandsspprichReg<-  
 wetlandsMA%>%  
 group\_by(ManagedArea,yearmonth,Year,Month)%>%  
 summarise(length(unique(gensp)))  
  
#"n" is the number of unique spp per Region and Year, aka the species Richness  
##rename that column  
colnames(wetlandsspprichReg) [5]<-"speciesrichness"  
#what do we have?  
wetlandsspprichReg

## # A tibble: 45 × 5  
## # Groups: ManagedArea, yearmonth, Year [45]  
## ManagedArea yearmonth Year Month speciesrichness  
## <chr> <chr> <int> <int> <int>  
## 1 Guana River Marsh 2012.11 2012 11 2  
## 2 Guana River Marsh 2013.08 2013 8 2  
## 3 Guana River Marsh 2014.01 2014 1 2  
## 4 Guana River Marsh 2014.02 2014 2 3  
## 5 Guana River Marsh 2014.03 2014 3 2  
## 6 Guana River Marsh 2014.04 2014 4 2  
## 7 Guana River Marsh 2014.05 2014 5 2  
## 8 Guana River Marsh 2014.06 2014 6 2  
## 9 Guana River Marsh 2014.07 2014 7 2  
## 10 Guana River Marsh 2014.08 2014 8 2  
## # … with 35 more rows

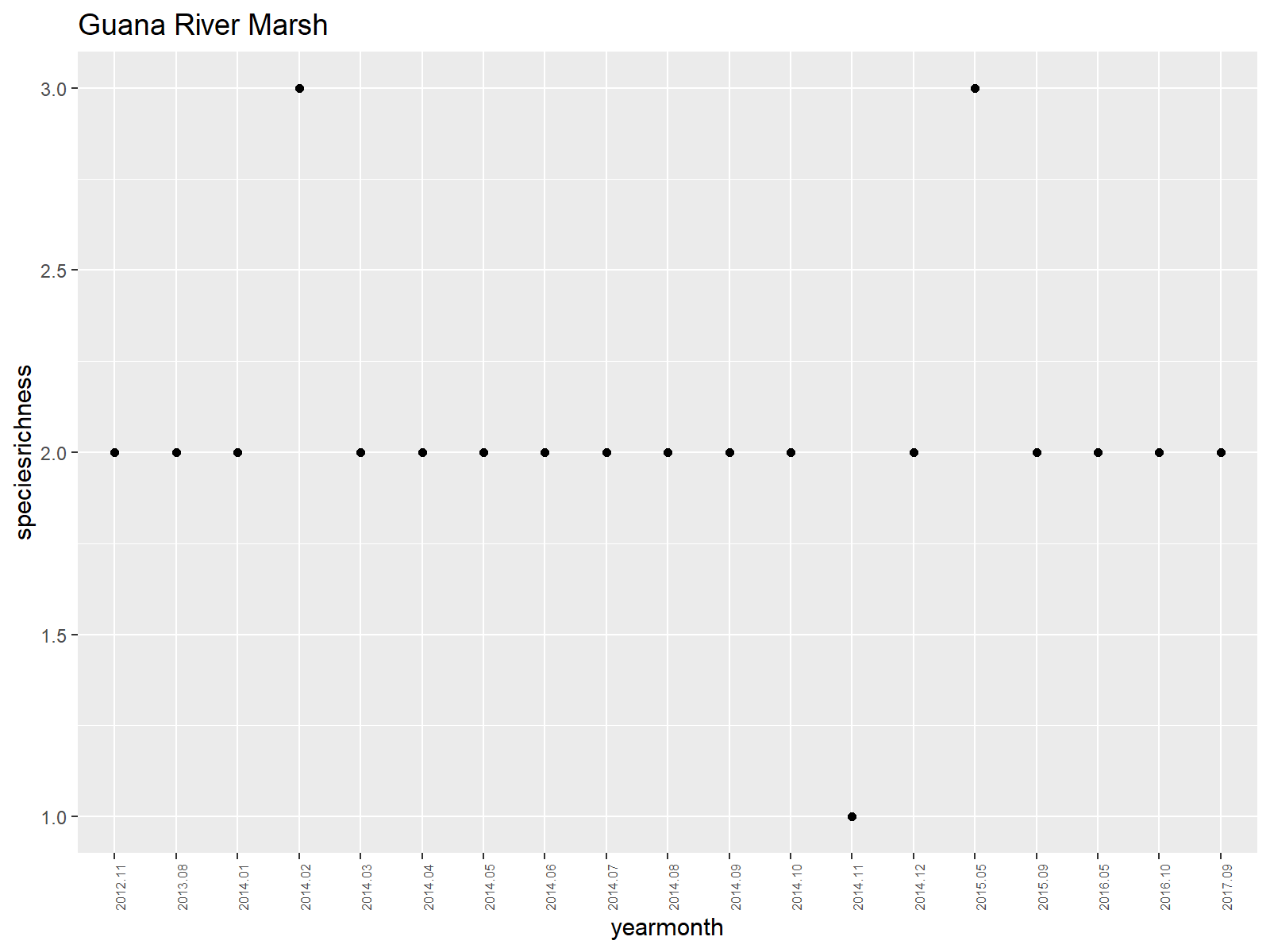
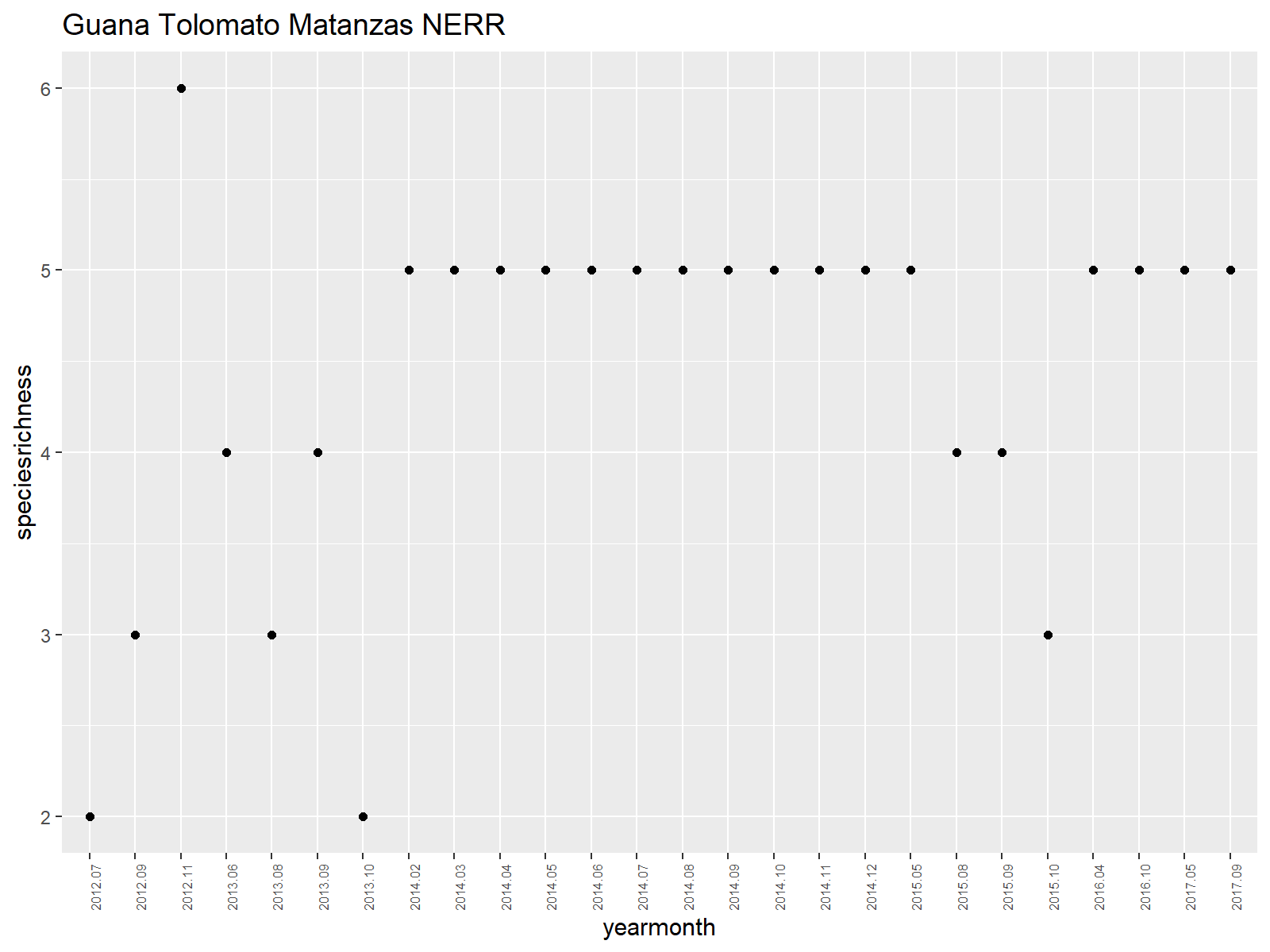
#now we have richness  
write.csv(wetlandsspprichReg,"output/wetlandsspprichReg.csv")

# Appendix I: Species Richness Plots

The plots shown here are the species richness for each managed area with a yearly average.

1. Creates a function that plots scatter plots and histograms for managed areas with enough data.
2. Save figure as png file in output directory

#Plot species richness  
plots\_listyrmth <- lapply(unique(wetlandsspprichReg$ManagedArea), function (i) {  
   
 dat <- filter(wetlandsspprichReg, ManagedArea == i)  
   
 ggplot(data = dat, aes(x = yearmonth, y = speciesrichness)) +  
 geom\_point() +  
 labs(title = i) +  
 scale\_x\_discrete(breaks = wetlandsspprichReg$yearmonth[seq(1, length(wetlandsspprichReg$yearmonth), by = 1)])+  
 theme(  
 axis.text.x = element\_text(angle = 90, hjust = 1,size=6)  
 )  
   
})  
plots\_listyrmth

[[1]]  [[2]] 

#save as "Coastal Wetlands Species Richness by Year/Month"  
png("output/Coastal Wetlands Species Richness by YearMonth, Managed Area.png")  
print(plots\_listyrmth)

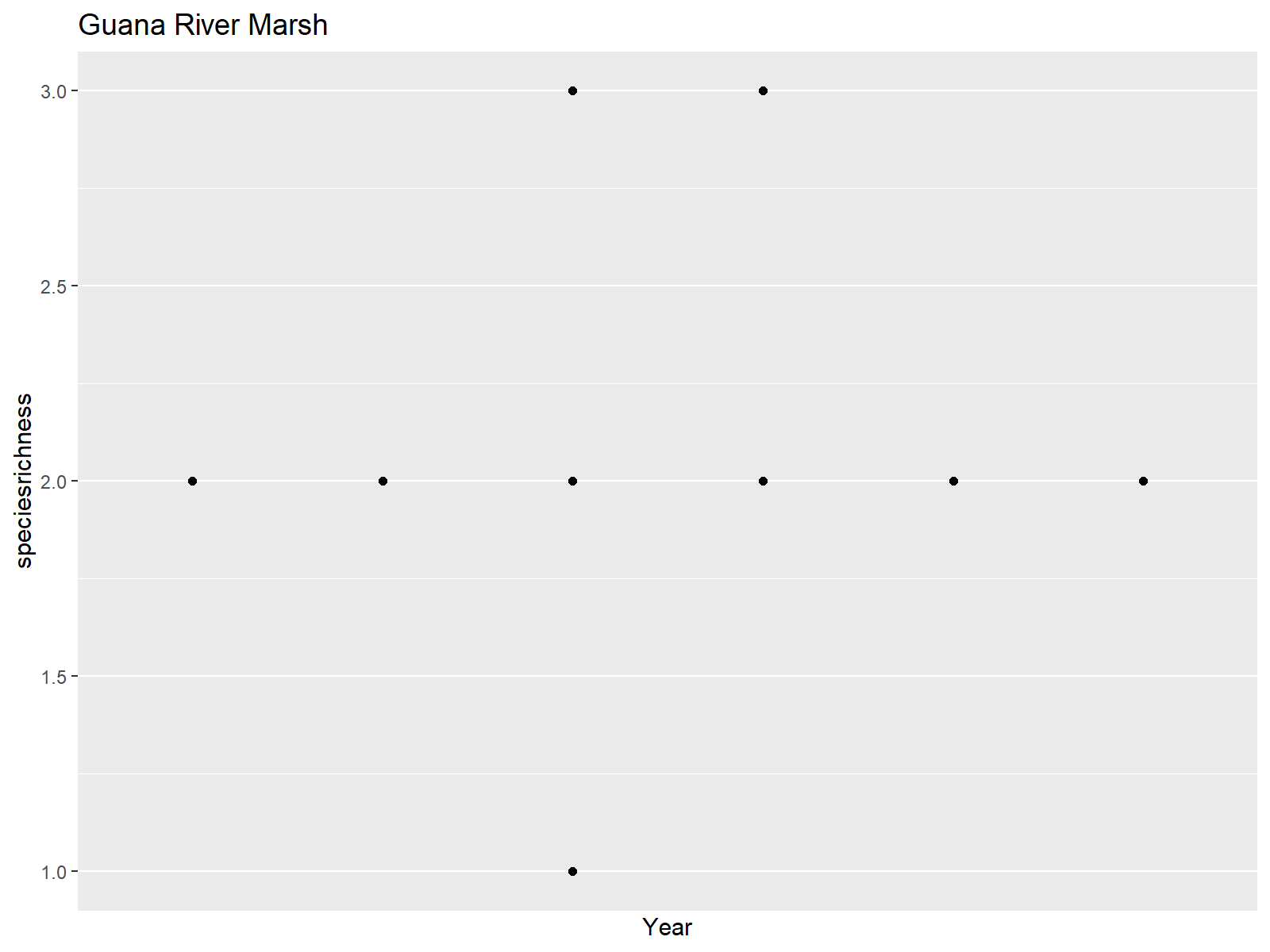
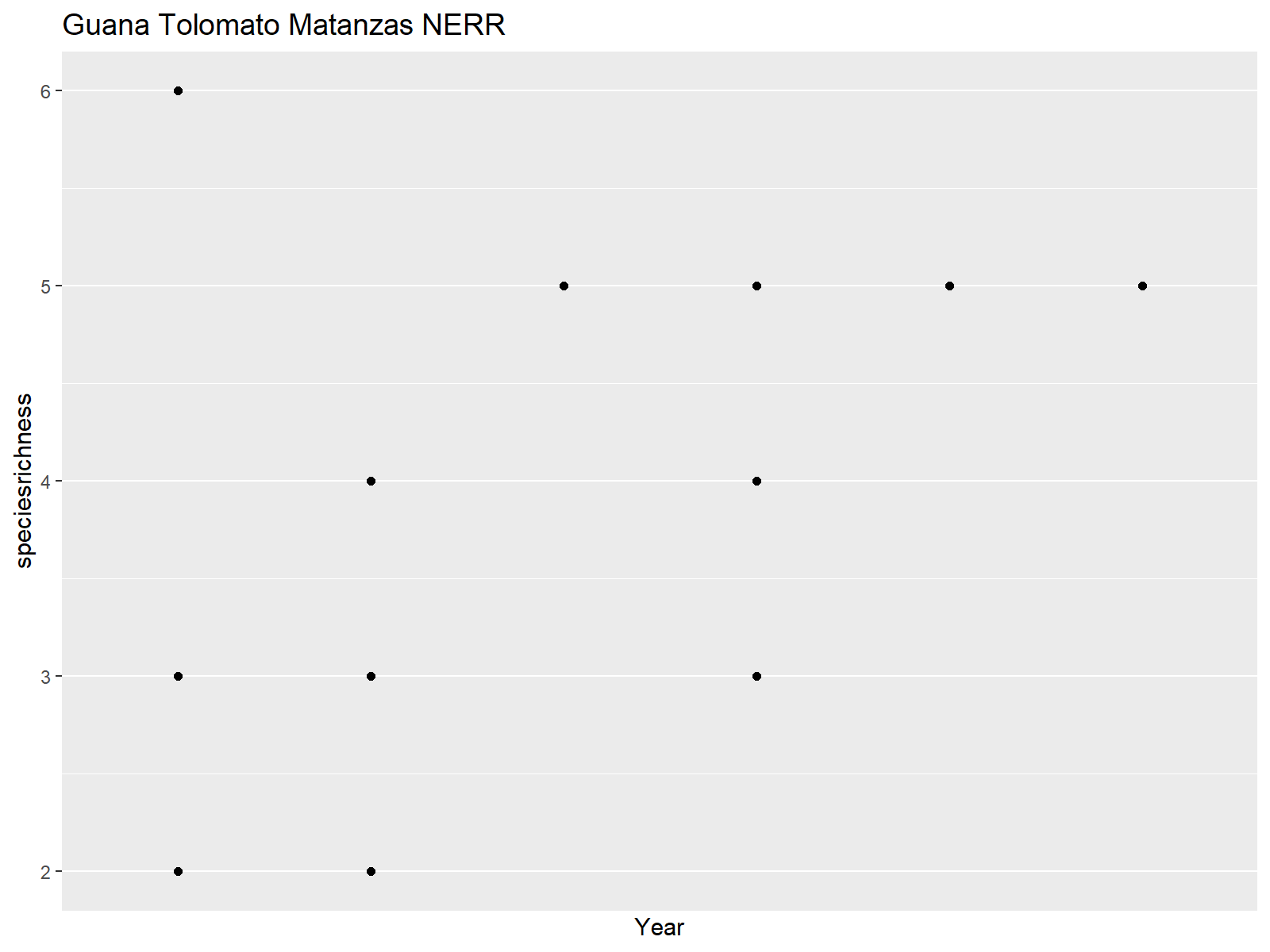
[[1]]

[[2]]

dev.off()

png 2

#and for Year  
plots\_listyr <- lapply(unique(wetlandsspprichReg$ManagedArea), function (i) {  
   
 dat <- filter(wetlandsspprichReg, ManagedArea == i)  
   
 ggplot(data = dat, aes(x = Year, y = speciesrichness)) +  
 geom\_point() +  
 labs(title = i) +  
 scale\_x\_discrete(breaks = wetlandsspprichReg$Year[seq(1, length(wetlandsspprichReg$Year), by = 1)])+  
 theme(  
 axis.text.x = element\_text(angle = 90, hjust = 1,size=6)  
 )  
   
})  
plots\_listyr

[[1]]  [[2]] 

png("output/Coastal Wetlands Species Richness by Year, Managed Area.png")  
print(plots\_listyr)

[[1]]

[[2]]

dev.off()

png 2

#RUN GLMs, one MA at a time  
#start with just one Managed Area: GTMNERR  
unique(wetlandsspprichReg$ManagedArea)

[1] “Guana River Marsh” “Guana Tolomato Matanzas NERR”

GTMNERR<-  
 wetlandsspprichReg%>%  
 filter(ManagedArea=="Guana Tolomato Matanzas NERR")  
colnames(GTMNERR)

[1] “ManagedArea” “yearmonth” “Year” “Month” “speciesrichness”

#  
#See if Month influences spp richness  
GTMNERRMontheffect=glm(GTMNERR$speciesrichness~GTMNERR$Month)  
summary(GTMNERRMontheffect)

Call: glm(formula = GTMNERRMonth)

Deviance Residuals: Min 1Q Median 3Q Max  
-2.4626 -0.3770 0.3663 0.6572 1.8113

Coefficients: Estimate Std. Error t value Pr(>|t|)  
(Intercept) 4.94178 0.61572 8.026 0.0000000298 \*\*\* GTMNERR$Month -0.06846 0.07676 -0.892 0.381  
— Signif. codes: 0 ‘***’ 0.001 ’****’ 0.01 ’*’ 0.05 ‘.’ 0.1 ’ ’ 1

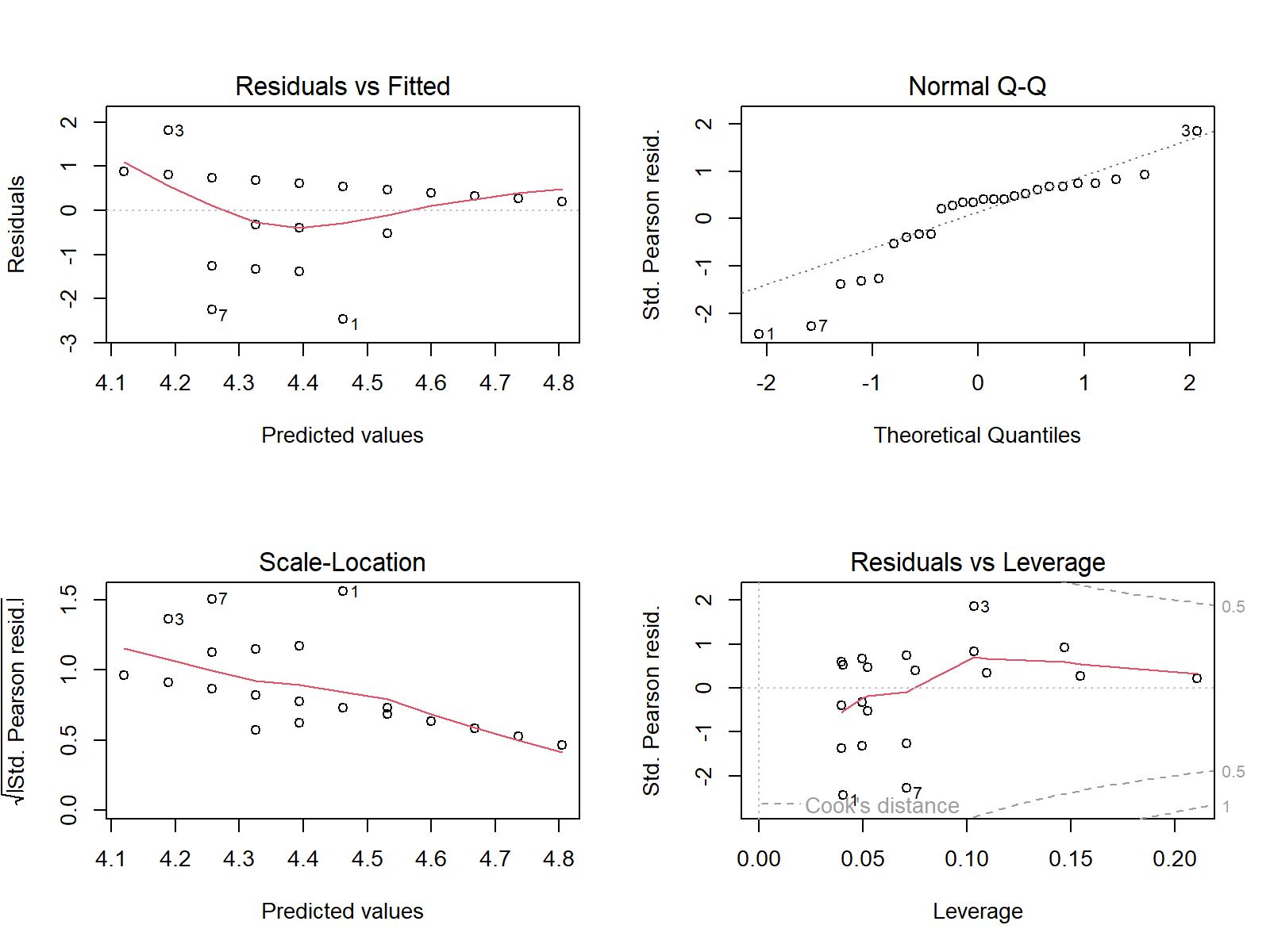
(Dispersion parameter for gaussian family taken to be 1.06254)

Null deviance: 26.346 on 25 degrees of freedom

Residual deviance: 25.501 on 24 degrees of freedom AIC: 79.281

Number of Fisher Scoring iterations: 2

par(mfrow=c(2,2))  
plot(GTMNERRMontheffect)



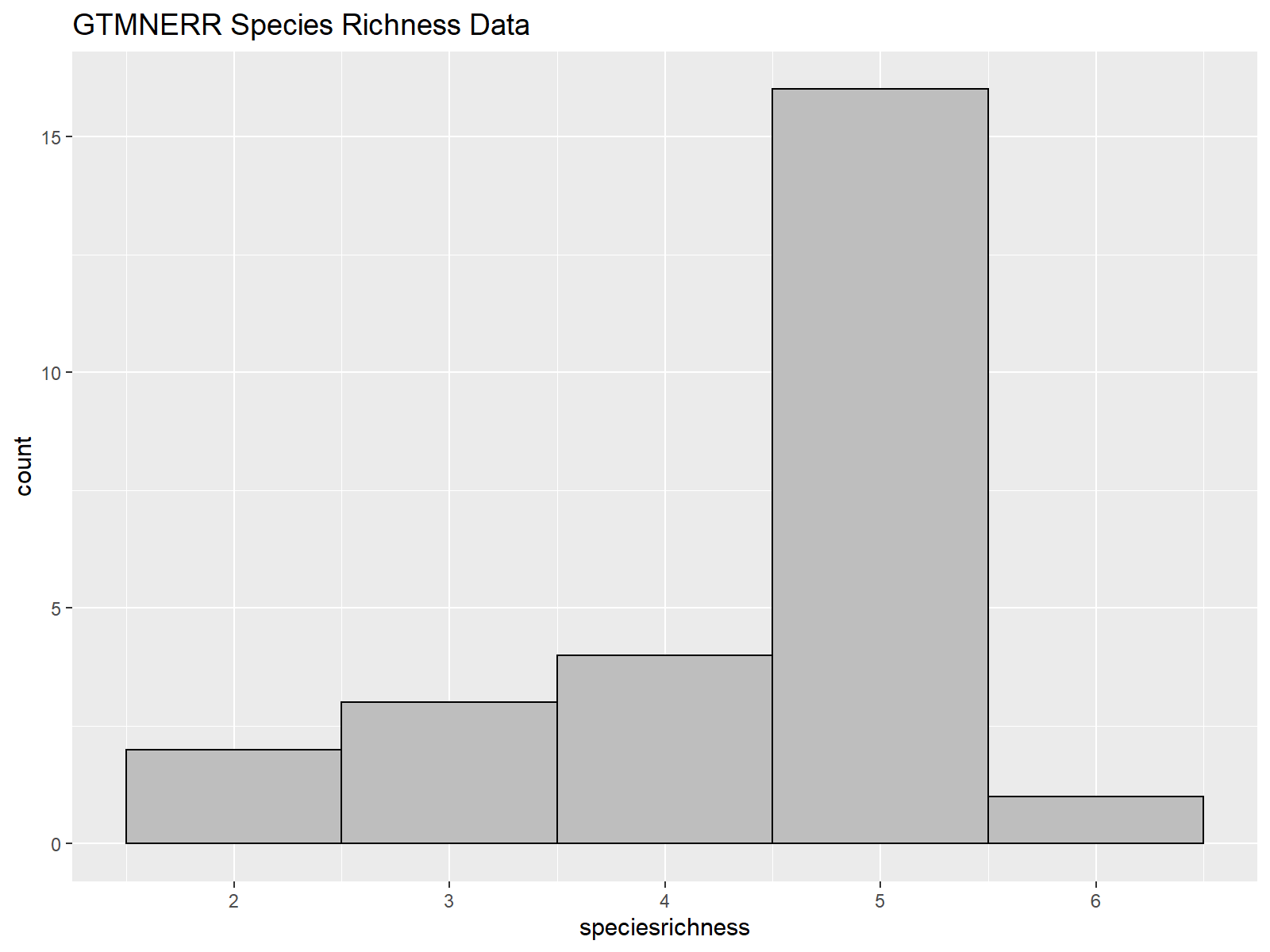
#if month does have an effect, you will likely want to incorporate it into the model. Here, we are  
##modeling species richness over time by year  
  
colnames(GTMNERR)

[1] “ManagedArea” “yearmonth” “Year” “Month” “speciesrichness”

#first let's plot the richness:  
##by yearmonth  
png("output/GTMNERR Coastal Wetlands Species Richness Data.png")  
par(mfrow=c(1,1))  
plot(GTMNERR$yearmonth, GTMNERR$speciesrichness,   
 xlab="Year",ylab="# of species",main="GTMNERR Coastal Wetlands Species Richness by Month & Year")  
plot(GTMNERR$Year, GTMNERR$speciesrichness,   
 xlab="Year",ylab="# of species",main="GTMNERR Coastal Wetlands Species Richness by Year")  
dev.off()

png 2

#histogram of data  
library(ggplot2)  
par(mfrow=c(1,1))  
ggplot(GTMNERR, aes(x=speciesrichness)) +   
 geom\_histogram(binwidth=1,color="black", fill="grey")+  
 ggtitle("GTMNERR Species Richness Data")



ggsave("output/GTMNERR Spp Richness Histogram.png")  
  
#Run the glm- below are samples to run it first by year, then by yearmonth.   
#Following that are sample lines of code to run the same analyses with different families. We are plotting the first example,  
#but you will want to run each family and compare the model outputs.  
#YEAR  
GTMNERRmodelYear=glm(GTMNERR$speciesrichness~GTMNERR$Year)  
summary(GTMNERRmodelYear)

Call: glm(formula = GTMNERRYear)

Deviance Residuals: Min 1Q Median 3Q Max  
-2.0848 -0.5678 0.1822 0.6220 2.2085

Coefficients: Estimate Std. Error t value Pr(>|t|)  
(Intercept) -586.1678 289.1236 -2.027 0.0539 . GTMNERR$Year 0.2932 0.1435 2.043 0.0522 . — Signif. codes: 0 ‘***’ 0.001 ’****’ 0.01 ’*’ 0.05 ‘.’ 0.1 ’ ’ 1

(Dispersion parameter for gaussian family taken to be 0.9351695)

Null deviance: 26.346 on 25 degrees of freedom

Residual deviance: 22.444 on 24 degrees of freedom AIC: 75.961

Number of Fisher Scoring iterations: 2

png("output/GTMNERR Coastal Wetlands Species Richness by Year\_Diagnostics.png")  
par(mfrow=c(2,2))  
plot(GTMNERRmodelYear)  
dev.off()

png 2

#PLOT the model  
GTMNERRmodel<-ggplot(GTMNERR, aes(x = Year, y = speciesrichness) ) +  
 geom\_point() +  
 geom\_smooth(method = "glm", alpha = .1,se=TRUE,color="darkgreen", fill="green") +   
 theme\_ipsum() +  
 ggtitle("GTMNERR Species Richness by Year Fitted Model")  
  
#export file by clicking #Export on the left and saving as:   
#GTMNERR Coastal Wetlands Species Richness by Year\_Model fit  
  
  
#YEARMONTH  
#GTMNERRmodelYearmonth=glm(GTMNERR$speciesrichness~GTMNERR$yearmonth)  
#summary(GTMNERRmodelYearmonth)  
  
#try other families  
#GTMNERR3=glm(GTMNERR$speciesrichness~GTMNERR$Year,family=poisson)  
#summary(NGTMNERR3)  
  
#GTMNERR4=glm(GTMNERR$speciesrichness~GTMNERR$Year,family=Gamma)  
#summary(GTMNERR4)  
  
#GTMNERR5=glm(GTMNERR$speciesrichness~GTMNERR$Year,family=inverse.gaussian)  
#summary(GTMNERR5)  
  
#############################################  
  
#Next MA  
#start with just one Managed Area: GuanaRiverMarsh  
unique(wetlandsspprichReg$ManagedArea)

[1] “Guana River Marsh” “Guana Tolomato Matanzas NERR”

GuanaRiverMarsh<-  
 wetlandsspprichReg%>%  
 filter(ManagedArea=="Guana River Marsh")  
colnames(GuanaRiverMarsh)

[1] “ManagedArea” “yearmonth” “Year” “Month” “speciesrichness”

#  
#See if Month influences spp richness  
GuanaRiverMarshMontheffect=glm(GuanaRiverMarsh$speciesrichness~GuanaRiverMarsh$Month)  
summary(GuanaRiverMarshMontheffect)

Call: glm(formula = GuanaRiverMarshMonth)

Deviance Residuals: Min 1Q Median 3Q Max  
-0.82231 -0.17713 0.00028 0.11855 0.82287

Coefficients: Estimate Std. Error t value Pr(>|t|)  
(Intercept) 2.47281 0.20812 11.881 0.00000000117 \*\* *GuanaRiverMarsh$Month -0.05914 0.02679 -2.208 0.0413*   
— Signif. codes: 0 ‘***’ 0.001 ’****’ 0.01 ’*’ 0.05 ‘.’ 0.1 ’ ’ 1

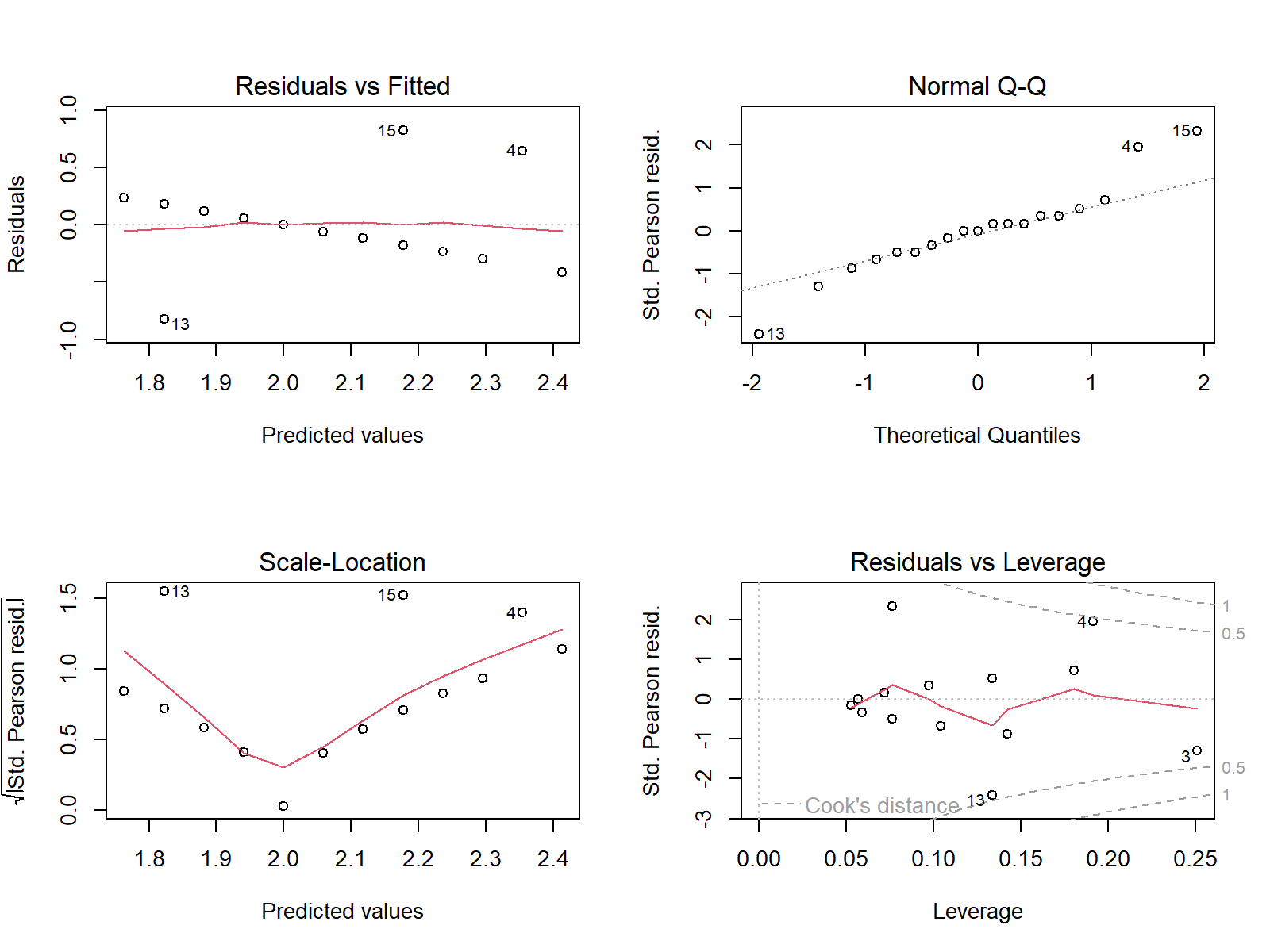
(Dispersion parameter for gaussian family taken to be 0.1347435)

Null deviance: 2.9474 on 18 degrees of freedom

Residual deviance: 2.2906 on 17 degrees of freedom AIC: 19.723

Number of Fisher Scoring iterations: 2

par(mfrow=c(2,2))  
plot(GuanaRiverMarshMontheffect)



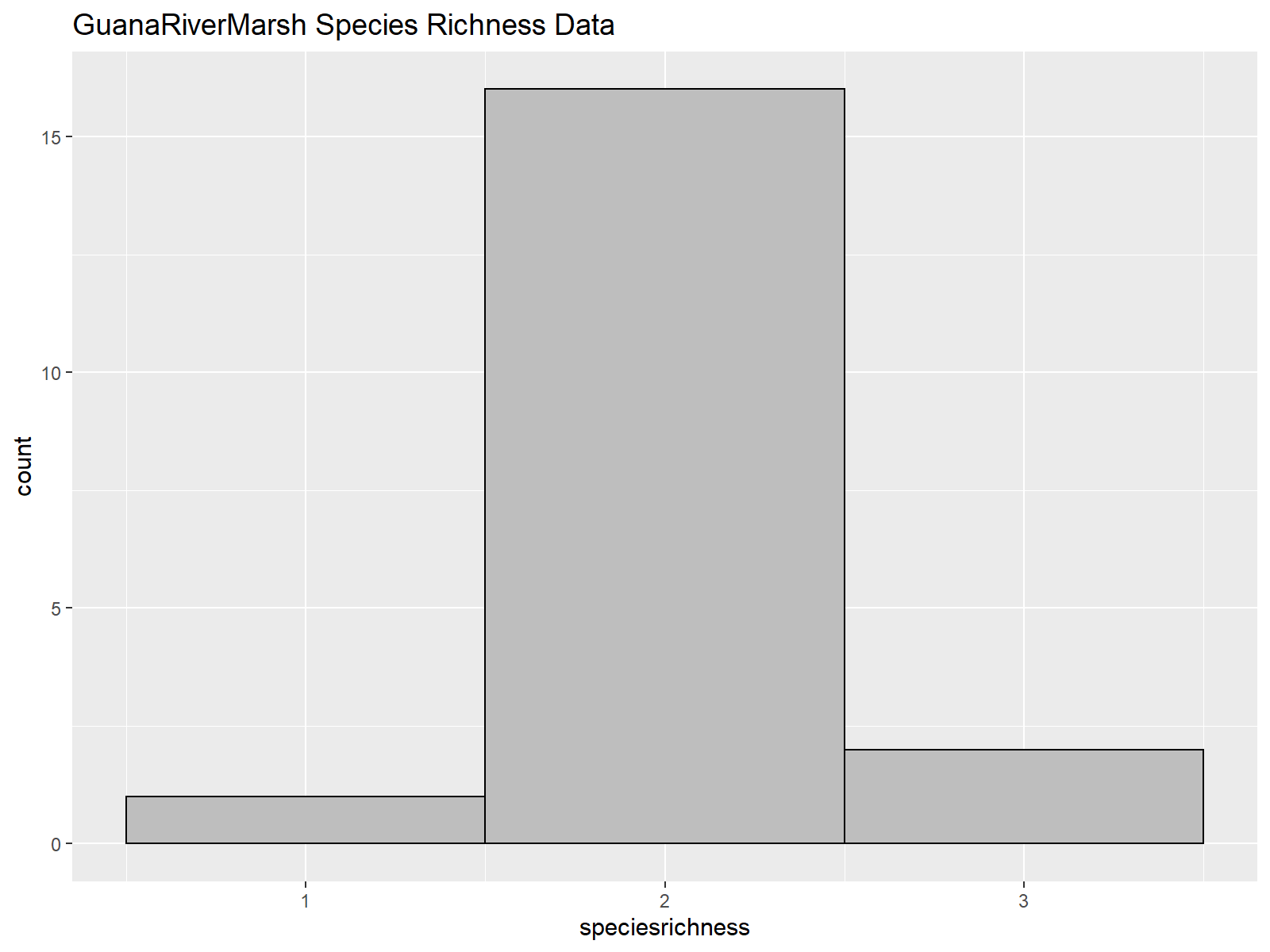
#if month does have an effect, you will likely want to incorporate it into the model. Here, we are  
##modeling species richness over time by year  
  
colnames(GuanaRiverMarsh)

[1] “ManagedArea” “yearmonth” “Year” “Month” “speciesrichness”

#first let's plot the richness:  
##by yearmonth  
png("output/GuanaRiverMarsh Coastal Wetlands Species Richness Data.png")  
par(mfrow=c(1,1))  
plot(GuanaRiverMarsh$yearmonth, GuanaRiverMarsh$speciesrichness,   
 xlab="Year",ylab="# of species",main="GuanaRiverMarsh Coastal Wetlands Species Richness by Month & Year")  
plot(GuanaRiverMarsh$Year, GuanaRiverMarsh$speciesrichness,   
 xlab="Year",ylab="# of species",main="GuanaRiverMarsh Coastal Wetlands Species Richness by Year")  
dev.off()

png 2

#histogram of data  
library(ggplot2)  
par(mfrow=c(1,1))  
ggplot(GuanaRiverMarsh, aes(x=speciesrichness)) +   
 geom\_histogram(binwidth=1,color="black", fill="grey")+  
 ggtitle("GuanaRiverMarsh Species Richness Data")



ggsave("output/GuanaRiverMarsh Spp Richness Histogram.png")  
  
#Run the glm- below are samples to run it first by year, then by yearmonth.   
#Following that are sample lines of code to run the same analyses with different families. We are plotting the first example,  
#but you will want to run each family and compare the model outputs.  
#YEAR  
GuanaRiverMarshmodelYear=glm(GuanaRiverMarsh$speciesrichness~GuanaRiverMarsh$Year)  
summary(GuanaRiverMarshmodelYear)

Call: glm(formula = GuanaRiverMarshYear)

Deviance Residuals: Min 1Q Median 3Q Max  
-1.04286 -0.05833 -0.04286 -0.04286 0.95714

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) -60.29524 177.74877 -0.339 0.739 GuanaRiverMarsh$Year 0.03095 0.08824 0.351 0.730

(Dispersion parameter for gaussian family taken to be 0.1721289)

Null deviance: 2.9474 on 18 degrees of freedom

Residual deviance: 2.9262 on 17 degrees of freedom AIC: 24.376

Number of Fisher Scoring iterations: 2

png("output/GuanaRiverMarsh Coastal Wetlands Species Richness by Year\_Diagnostics.png")  
par(mfrow=c(2,2))  
plot(GuanaRiverMarshmodelYear)  
dev.off()

png 2

#PLOT the model  
GuanaRiverMarshmodel<-ggplot(GuanaRiverMarsh, aes(x = Year, y = speciesrichness) ) +  
 geom\_point() +  
 geom\_smooth(method = "glm", alpha = .1,se=TRUE,color="darkgreen", fill="green") +   
 theme\_ipsum() +  
 ggtitle("GuanaRiverMarsh Species Richness by Year Fitted Model")