Exploration of 2016-2017 CITES Poaching Database

https://github.com/emilyrwood/CITES_Poaching_Wood

Emily Wood

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Poaching remains a considerable threat to biodiversity across the planet. However, it often is overshadowed by other environmental issues in the policy realm. This is especially true for species that aren't charismatic mega-fauna. I'm undertaking this research project to better understand the trends in poaching over time and understand the regions where the most risk is occuring. This could help target poaching hotspots and help policy makers make more informed decisions when it comes to mitigating poaching risk for sesitive species.

Main Questions:

Is there any correlation between class, term, appendix number and poaching amount?

Sub Questions:

What is the distribution of amounts found exported and imported for all species included in the database?

Are there differences between Appendix Levels for the amount exported vs. imported?

What are the most commonly poached classes between Appendix Levels?

What are the top 3 Classes for species poached in Appendix I?

What are the dominant good types (terms) for the top 3 classes poached in Appendix I?

Which Counties are the biggest Importer and Exporters of CITES species?

2 Dataset Information

The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) is an international agreement between governments. This group studies poaching trends and advises the UN on necessary action to mitigate this risk. This CITES dataset contains records on every international import or export conducted with species from the CITES monitoring list in 2016 and 2017. It contains columns identifying the species, the import and export countries, and the amount and characteristics of the goods being traded (which range from live animals to skins and cadavers). The 'Term' colum described the type of good the poached species takes when encountered. Examples for this column include live, trophies, tusks, ect.

Notably, this dataset has a column named, 'Appendix' which sorts the occurrences into three groups. Appendix I species are those whose poaching directly threatens them with extinction whose trade threatens them with extinction. There are around 1200 such species in this Appendix. Appendix II species are those who do not directly face extinction from poaching, but would experience detrimental impacts regardless. This group makes up the majority of the data set with around 21000 species. Appendix III animals are considered controls because their export and import requires permits. There are around 170 in this dataset. I will be strongly focused on Appendix I in my analysis.

```
# Processing Goal: data cleaning, integration, transformation, and reduction
# Replace codes with labels for 'Purpose' column.

CITESPoaching2016$Purpose <- ifelse(CITESPoaching2016$Purpose == "B", "Breeding",
    ifelse(CITESPoaching2016$Purpose == "E", "Educational", ifelse(CITESPoaching2016$Pur
        "G", "Garden", ifelse(CITESPoaching2016$Purpose == "H", "Hunting", ifelse(CITESP
        "L", "Law", ifelse(CITESPoaching2016$Purpose == "M", "Medical", ifelse(CITESPoac
        "R", "Reintroduction", ifelse(CITESPoaching2016$Purpose == "P", "Personal",
        ifelse(CITESPoaching2016$Purpose == "Q", "Circus", ifelse(CITESPoaching2016$Purp
            "S", "Scientific", ifelse(CITESPoaching2016$Purpose == "T", "Commercial",
            ifelse(CITESPoaching2016$Purpose == "Z", "Zoo", "Unknown")))))))))))

# rename some of the columns

colnames(CITESPoaching2016)[which(names(CITESPoaching2016) == "Importer reported quantit

colnames(CITESPoaching2016)[which(names(CITESPoaching2016) == "Exporter reported quantit

colnames(CITESPoaching2016)[which(names(CITESPoaching2016) == "App.")] <- "Appendix"
```

```
# wrangling Goal: cleaning the raw dataset into a format compatible

CITESPoaching2016$Reported_quantity <- CITESPoaching2016$Importer_reported_quantity +
    CITESPoaching2016$Exporter_reported_quantity
```

```
# Limit dataset to poaching occurrence measured by count not by weight.  I'm
# choosing count because the majority of cases are recorded in count not
# weight.  Select relevant columns

CITESPoaching2016["Unit"][is.na(CITESPoaching2016["Unit"])] <- "Count"

CITES2016_Processed <- CITESPoaching2016 %>%
    filter(Unit == "Count") %>%
    select(Year, Appendix, Class, Genus, Importer, Exporter, Importer_reported_quantity,
        Exporter_reported_quantity, Term, Purpose, Reported_quantity) %>%
    group_by(Appendix)

# Save the processed dataframe to processed folder
getwd()
```

```
## [1] "/home/guest/EDA_2022/CITES_Poaching_Wood"
```

```
write.csv(CITES2016_Processed, row.names = FALSE, file = "./Data_Processed/CITES2016_Pro
```

Wrangling the Dataset:

To prepare this dataset I first replaced the Codes for the 'Purpose' column with text to match the other columns such as the 'Term' column. I then renamed certain columns to get rid of abbreviations and spaces. Next I set the NAs in both the importer and exporter quantity columns to 0 so that they could be combined into a total reported quantity column. After this column was created I noted that the Unit column combined individual occurrence counts and those measured by weight in grams and kilograms. After viewing the amounts of each type, I made the decision to remove occ-urrences measured in weight because there were significantly less when compared to those measured in counts. To do this I set The NAs in the Unit column to "Count" and filtered for it when finishing my processed dataframe. As part of this process I also selected the variables I needed which included Year, Appendix, Class, Genus, Importer, Exporter, Importer_reported_quantity, Exporter_reported_quantity, Term, Purpose, Reported_quantity. Last, I grouped the dataframe by Appendix number. Table 1 depicts the variables within my processed dataframe.

Variables	Units	Ranges
Year		2016-2017
Appendix		I,II,III
Class		Actinopteri, Anthozoa, Aves,... Reptilia
Genus		Alligator, Aloe, Alveopora, ... Ursus
Importer	Country Code	AD, AE, AU,... ZW
Exporter	Country Code	AD, AE, AR,... ZZ
Importer_reported_quantity	Count	1,2,3, ... 19524978
Exporter_reported_quantity	Count	1,2,3, ... 21543618
Term		baleen, bark, bodies, ... wood product

Variables	Units	Ranges
Purpose		Breeding, Circus, Commercial, ... Zoo
Reported Quantity	Count	1, 2, 3, ... 21543639

Table 1. Processed Data Frame

```
Export.districution <- ggplot(CITESPoaching2016, aes(x = Exporter_reported_quantity)) +
    geom_histogram(binwidth = 50, colour = "blue", fill = "light blue", alpha = 0.3) +
    xlim(0, 1500) + ylim(0, 2500) + labs(title = "Export Distribution")


Import.distribution <- ggplot(CITESPoaching2016, aes(x = Importer_reported_quantity)) +
    geom_histogram(binwidth = 50, colour = "dark green", fill = "light green", alpha = 0
    xlim(0, 1500) + ylim(0, 2500) + labs(title = "Import Distribution")


Graphnew <- ggplot(CITESPoaching2016, aes(x = Exporter_reported_quantity, y = Importer_r
    color = Appendix, alpha = 0.3)) + geom_point() + xlim(0, 4e+06) + ylim(0, 4e+06)


Appendix1 <- ggplot(subset(CITESPoaching2016, Appendix == "I"), aes(x = Exporter_reporte
    y = Importer_reported_quantity, color = Appendix, alpha = 0.3)) + geom_point()


# Remove Null Values and combine Export and Import amounts into a total
# reported quantity column. Previous graphs should not contain the zeros.

CITESPoaching2016["Importer_reported_quantity"][is.na(CITESPoaching2016["Importer_report

CITESPoaching2016["Exporter_reported_quantity"][is.na(CITESPoaching2016["Exporter_report


# Visualizing countries with greater number of exports and imports Found
# journal to visualize countries
# https://journal.r-project.org/archive/2011-1/RJournal_2011-1_South.pdf
# install.packages('rworldmap')
library(rworldmap)

# Importer Countries
CITES_Importer_withmap <- joinCountryData2Map(CITES2016_Processed, joinCode = "ISO_A2",
    nameJoinColumn = "Importer")
```

```
## 59647 codes from your data successfully matched countries in the map
## 1112 codes from your data failed to match with a country code in the map
## 41 codes from the map weren't represented in your data
```

```
mapDevice()  #create world map shaped window
mapCountryData(CITES_Importer_withmap, nameColumnToPlot = "Importer")
```

```
## using catMethod='categorical' for non numeric data in mapCountryData
```

```
Importerplot1 <- ggplot(CITES2016_Processed, aes(x = Importer)) + geom_bar(binwidth = 50
    colour = "blue", fill = "light blue", alpha = 0.3)
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
# Difficult to interpret. Wrangle new dataset and create bar graph with only
# top 10 import countries:

# table(CITES2016_Processed$Importer)

ImporterData <- CITES2016_Processed %>%
    group_by(Importer) %>%
    summarise(Importer, count = n()) %>%
    filter(Importer == "US" | Importer == "JP" | Importer == "DE" | Importer == "FR" |
        Importer == "HK" | Importer == "CH" | Importer == "CN" | Importer == "SG" |
        Importer == "AE" | Importer == "CA")
```

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

```
Importerplot2 <- ggplot(ImporterData, aes(x = Importer)) + geom_bar(binwidth = 50,
    colour = "purple", fill = "lavender", alpha = 0.3)
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
# Exporter Countries
CITES_Exporter_withmap <- joinCountryData2Map(CITES2016_Processed, joinCode = "ISO_A2",
    nameJoinColumn = "Exporter")
```

```
## 60130 codes from your data successfully matched countries in the map
## 629 codes from your data failed to match with a country code in the map
## 44 codes from the map weren't represented in your data
```

```
mapDevice()  #create world map shaped window
mapCountryData(CITES_Exporter_withmap, nameColumnToPlot = "Exporter")
```

```
## using catMethod='categorical' for non numeric data in mapCountryData
```

```
Exporterplot1 <- ggplot(CITES2016_Processed, aes(x = Exporter)) + geom_bar(binwidth = 50
    colour = "blue", fill = "light blue", alpha = 0.3)
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
# Difficult to interpret. Wrangle new dataset and create bar graph with only
# top 10 export countries:

# table(CITES2016_Processed$Exporter)

ExporterData <- CITES2016_Processed %>%
```

```
    group_by(Exporter) %>%
    summarise(Exporter, count = n()) %>%
    filter(Exporter == "NL" | Exporter == "ID" | Exporter == "IT" | Exporter == "US" |
        Exporter == "FR" | Exporter == "DE" | Exporter == "EC" | Exporter == "AU" |
        Exporter == "TH" | Exporter == "ZA")
```

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

```
Exporterplot2 <- ggplot(ExporterData, aes(x = Exporter)) + geom_bar(binwidth = 50,
    colour = "blue", fill = "light blue", alpha = 0.3)
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
# Which Classes get poached most often?

Classplot1 <- ggplot(CITES2016_Processed, aes(x = Class)) + geom_bar(aes(fill = Class))
    labs(title = "Classes of poaching occurances 2016", x = "Class", y = "Count") +
    theme(legend.position = "none") + coord_flip()



# This has a lot of NAs perhaps they didn't record always the Class Faceted by
# appendix

Classplot2 <- ggplot(CITES2016_Processed, aes(x = Class)) + geom_bar(aes(fill = Class))
    labs(title = "Classes of poaching occurances 2016", x = "Class", y = "Count") +
    coord_flip() + facet_grid(rows = "Appendix") + theme(axis.text = element_text(size =

# create a facet dataset for Appendix I

AppendixI_Facet <- CITES2016_Processed %>%
    filter(Appendix == "I")


# A closer look at Mammalia

Mammalia_terms <- ggplot(subset(AppendixI_Facet, Class %in% "Mammalia"), aes(x = Term))
    geom_bar(aes(fill = Term)) + coord_flip() + theme(axis.text = element_text(size = 5)
    legend.key.size = unit(0.2, "cm"), legend.key.height = unit(0.2, "cm"), legend.key.w
        "cm"), legend.title = element_text(size = 3), legend.text = element_text(size =


# A closer look at Aves

Aves_terms <- ggplot(subset(AppendixI_Facet, Class %in% "Aves"), aes(x = Term)) +
```

```
    geom_bar(aes(fill = Term)) + coord_flip() + theme(axis.text = element_text(size = 5)
    legend.key.size = unit(0.2, "cm"), legend.key.height = unit(0.2, "cm"), legend.key.w
        "cm"), legend.title = element_text(size = 3), legend.text = element_text(size =


# A closer look at Reptilia

Reptilia_terms <- ggplot(subset(AppendixI_Facet, Class %in% "Reptilia"), aes(x = Term))
    geom_bar(aes(fill = Term)) + coord_flip() + theme(axis.text = element_text(size = 5)
    legend.key.size = unit(0.2, "cm"), legend.key.height = unit(0.2, "cm"), legend.key.w
        "cm"), legend.title = element_text(size = 3), legend.text = element_text(size =


# Why arent plants represented in Class? monocotyledons or dicotyledons not
# depicted so potentiall in NAs

AppendixI_Facet["Class"][is.na(AppendixI_Facet["Class"])] <- "Not_recorded"

NA_terms <- ggplot(subset(AppendixI_Facet, Class %in% "Not_recorded"), aes(x = Term)) +
    geom_bar(aes(fill = Term)) + coord_flip() + theme(axis.text = element_text(size = 5)
    legend.key.size = unit(0.2, "cm"), legend.key.height = unit(0.2, "cm"), legend.key.w
        "cm"), legend.title = element_text(size = 3), legend.text = element_text(size =
```

3 Exploratory Analysis

First I explored the distribution of Import and export quantity. Figure 1 and Figure 2 depict these distributions. The majority of occurrence fall in the sigle digits. Frown these distributions you can also see subtle jumps around whole numbers. From this, I beleive that some rounding has occured, especially with counts around 500 and 100.

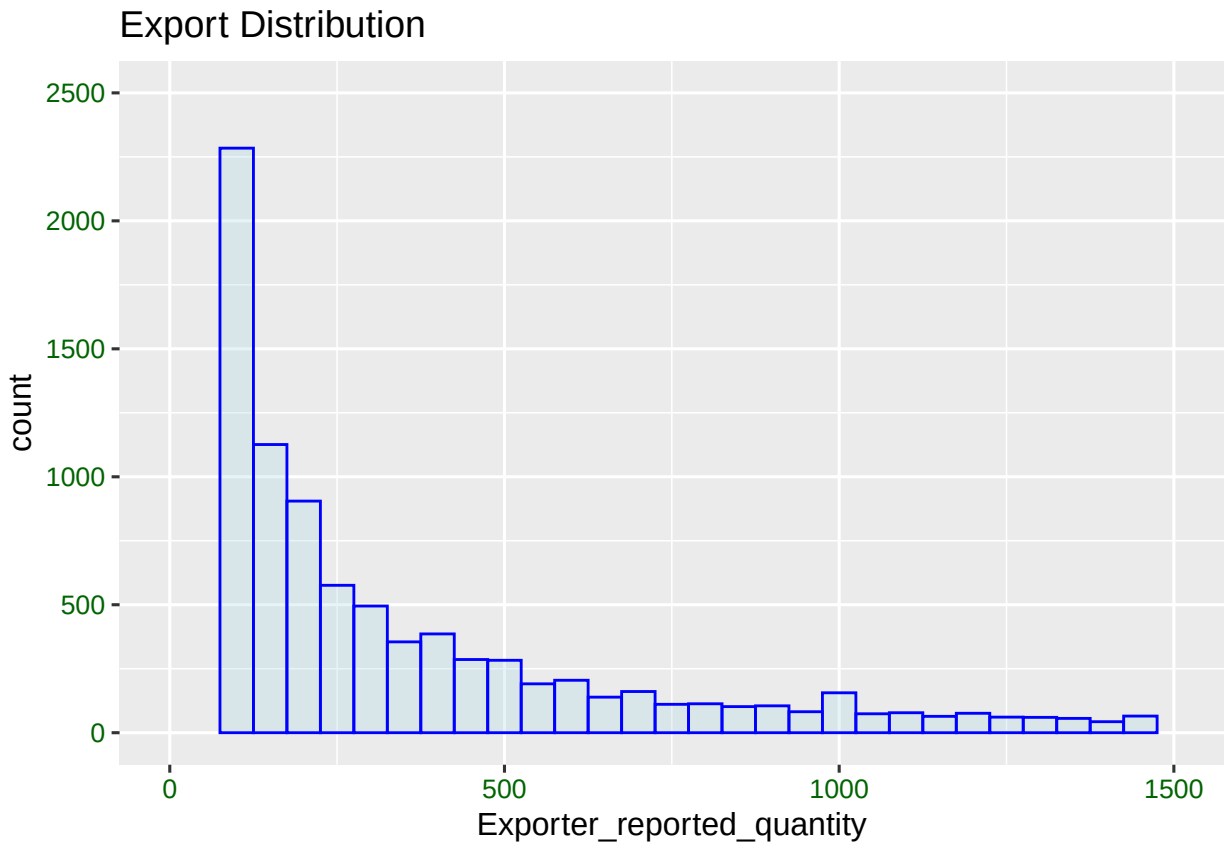

Figure 1. Export Distribution

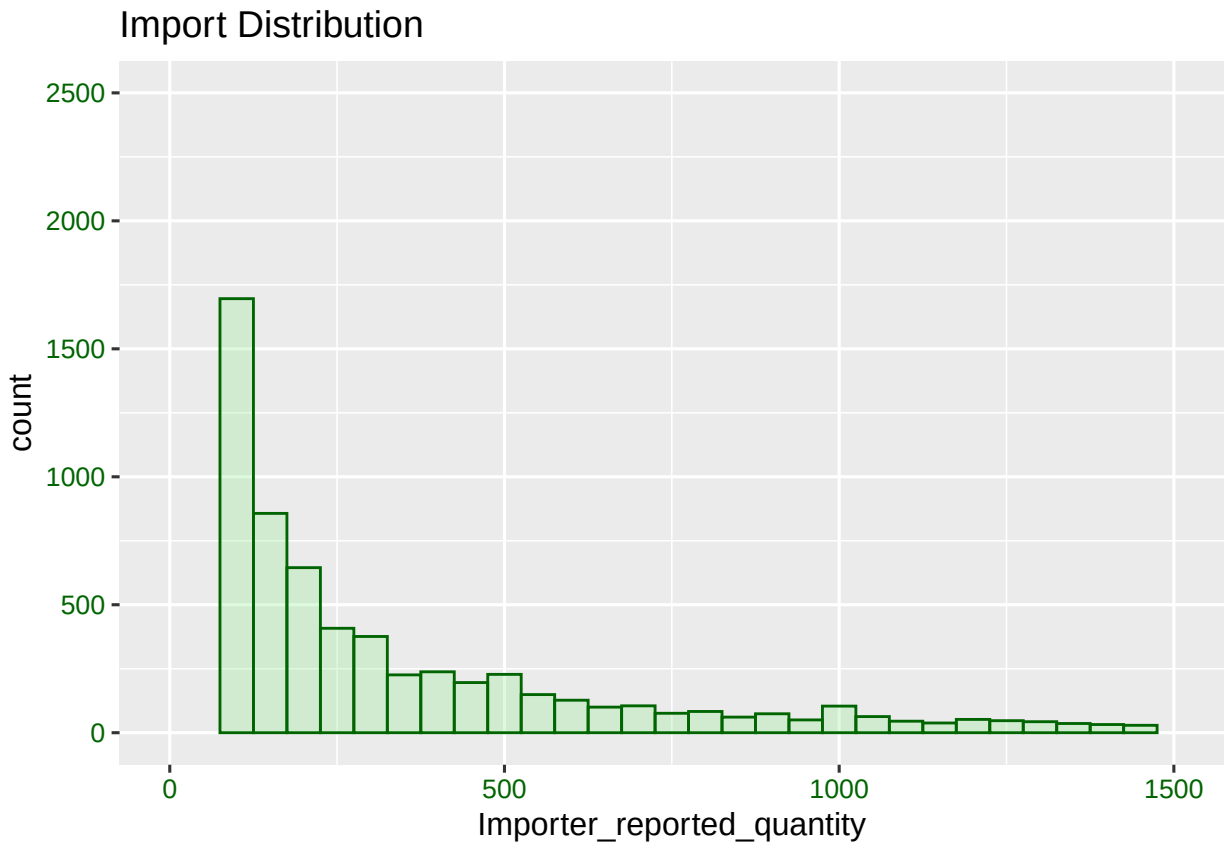

Figure 2. Import Distribution

Next I visualized export quantity and compared to import quantity. I broke this up by appendix levels which is visualized in Figure 3. Unfortunately, Appendix I was not easily viewing when combines with Appendix II and Appendix III so I also visualized that separately in Figure 4.

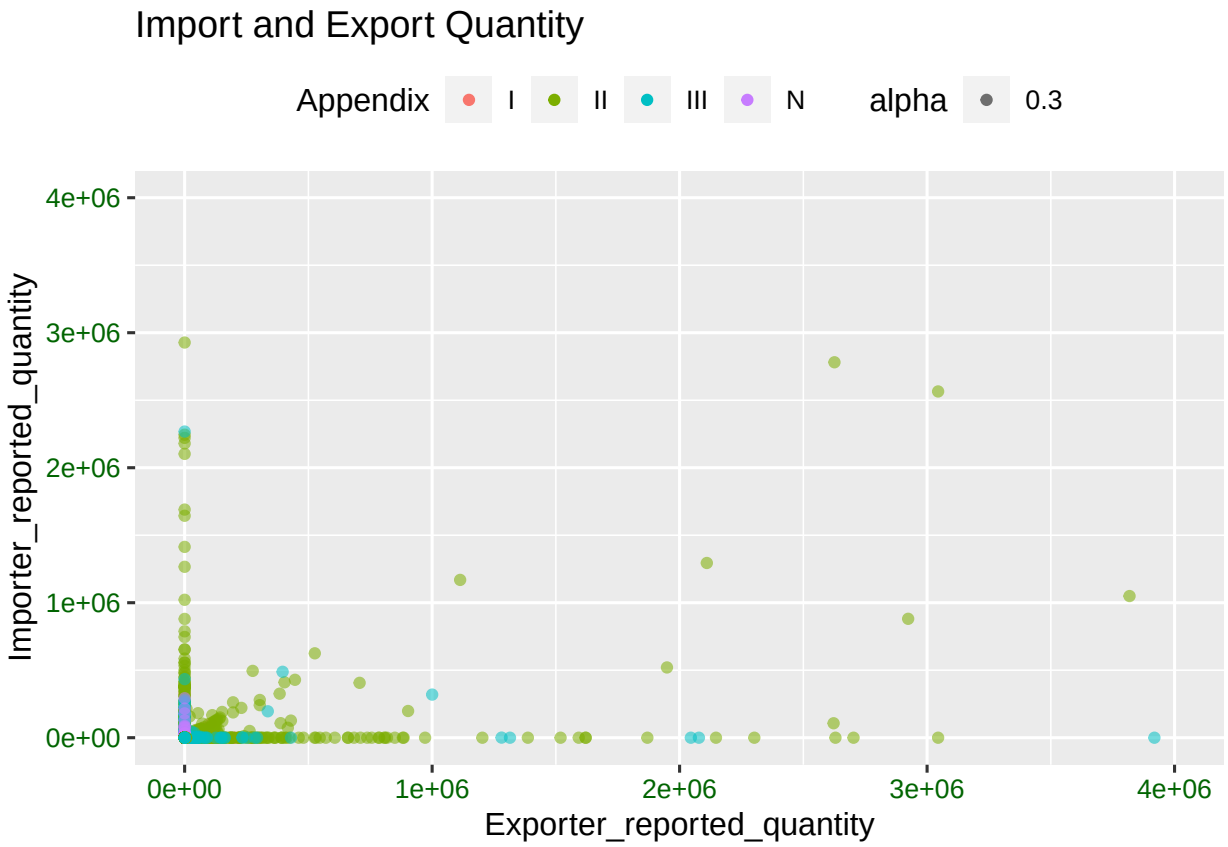

Figure 3. Export and Import Quantity

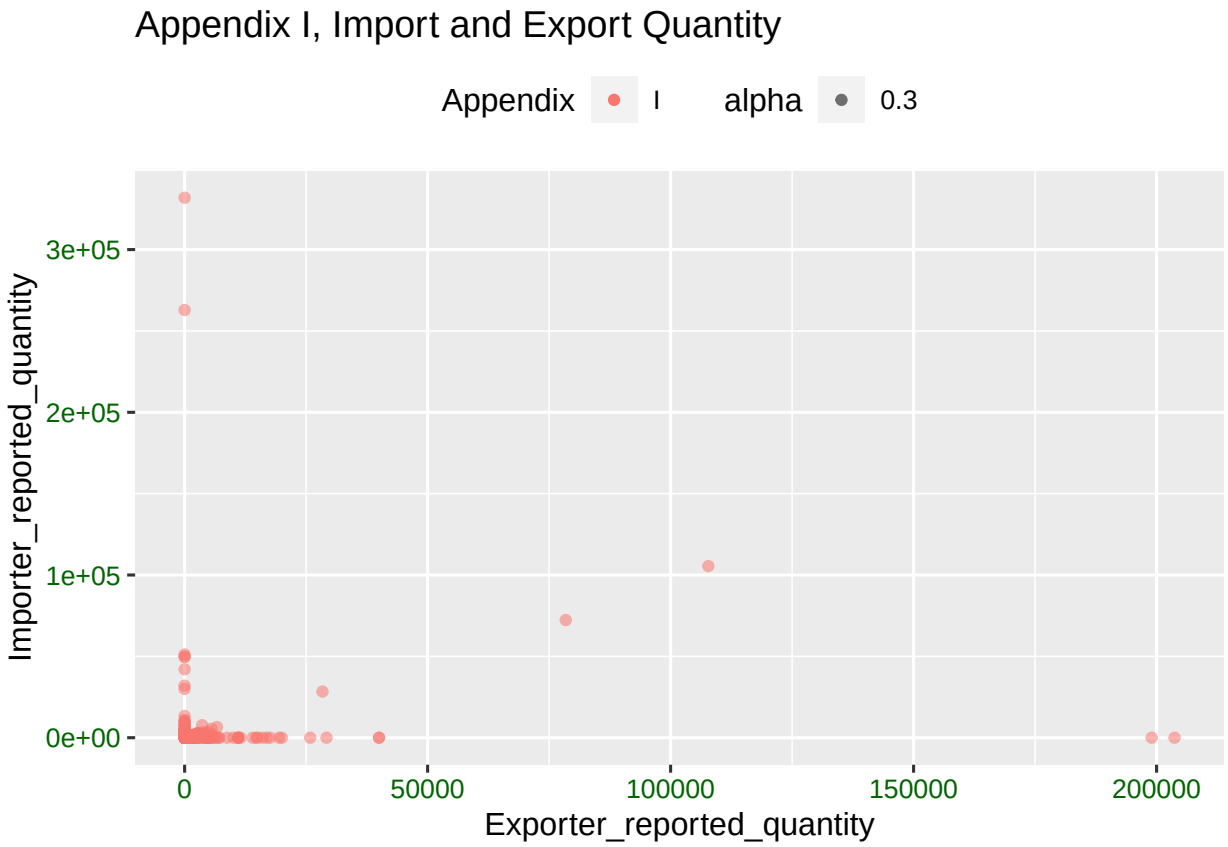

Figure 4. Export and Import Quantity for Appendix I

Next, I visualized the Countries with the most imports and Exports. The countries with the higher Imports are visualized in Figure 5. For clarity I also plotted the top ten countries with the most Imports in Figure 6. I did the same process to visualize exports in Figure 7 and Figure 8.

Interestingly, we see that the top importers and exporters are different. The top importer in terms of count is the United stated and the top exporter in terms of count is the Netherlands. Its important to note that this could be do to the way they screen their importans and exports. As there is no accompanied standardization protocol its hard to say how accurate this is.

```
## 59647 codes from your data successfully matched countries in the map
## 1112 codes from your data failed to match with a country code in the map
## 41 codes from the map weren't represented in your data
```

Figure 5. Importer World Map (In Separate Document)

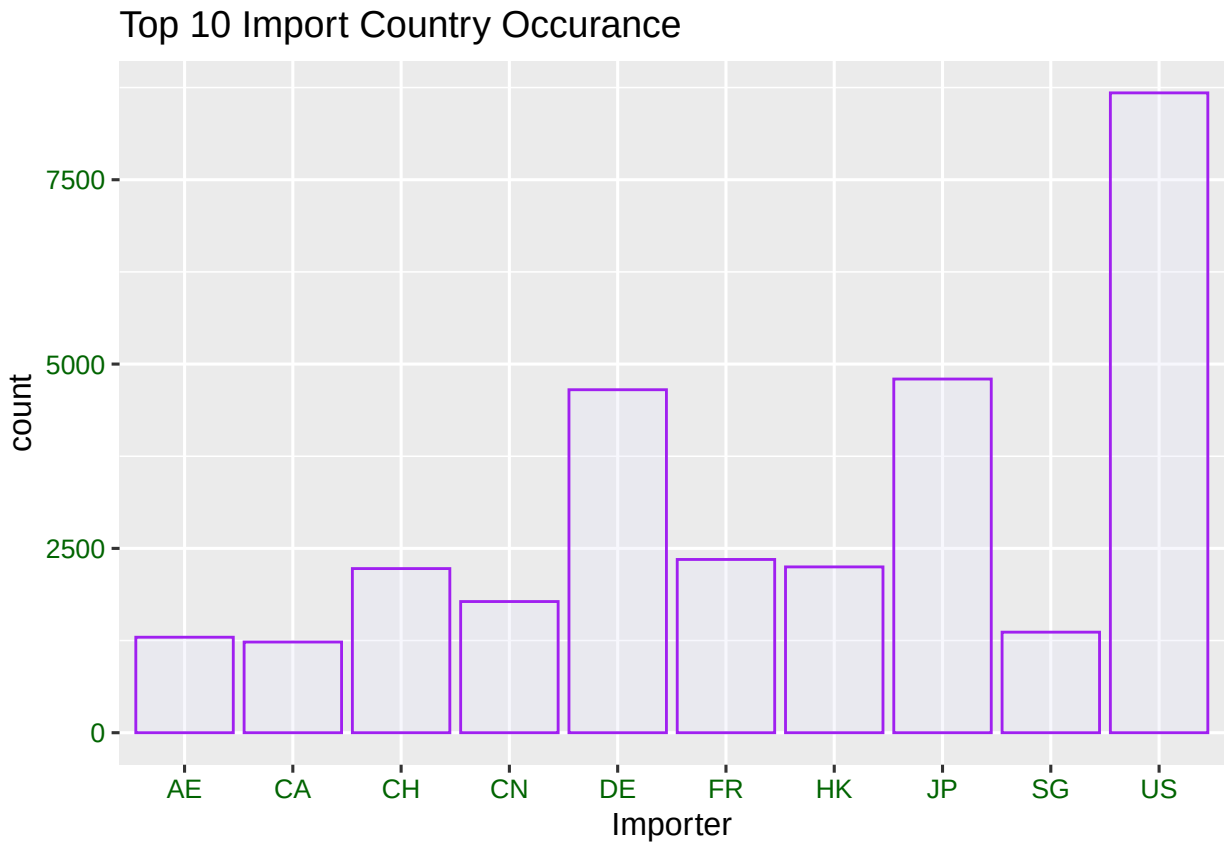

Figure 6. Importer Country Plot Top 10

```
## 60130 codes from your data successfully matched countries in the map
## 629 codes from your data failed to match with a country code in the map
## 44 codes from the map weren't represented in your data
```

Figure 7. Exporter World Map (In Separate Document)

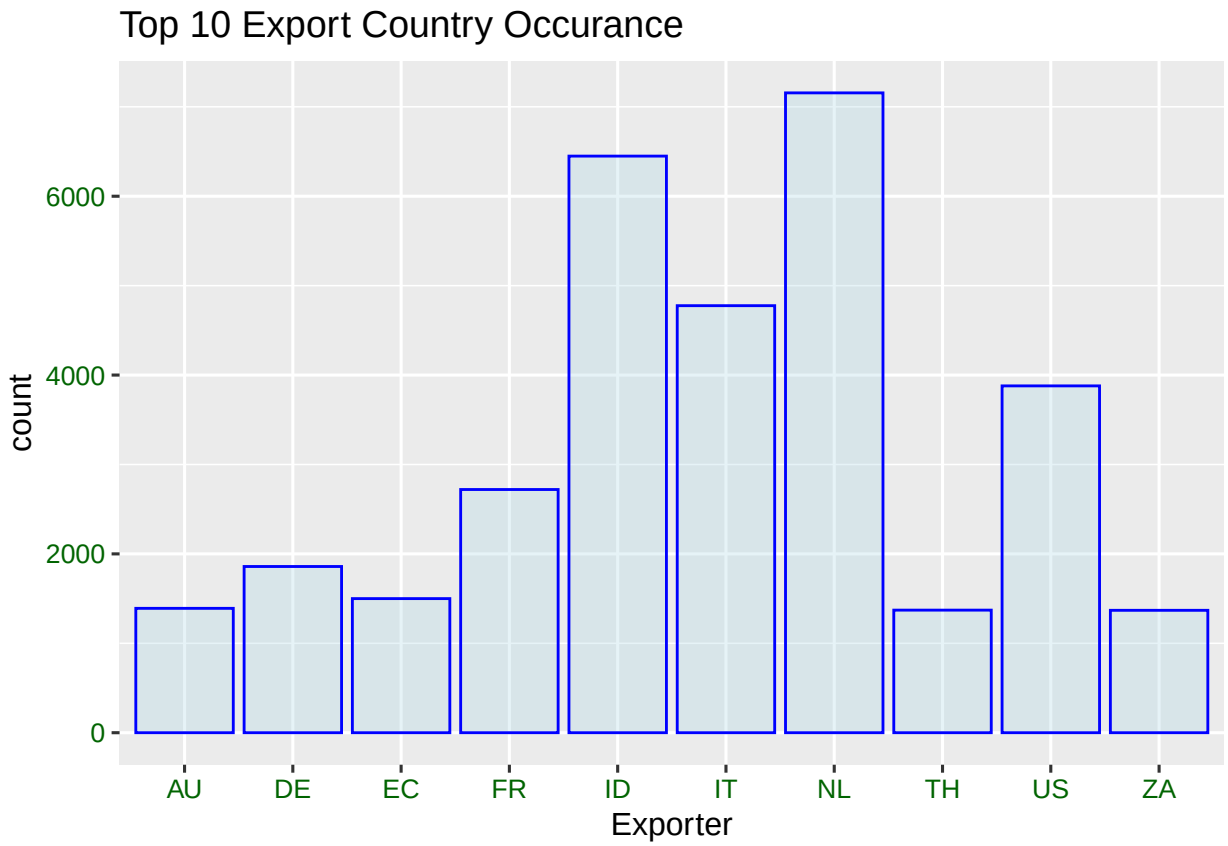

Figure 8. Exporter Country Plot Top 10

Next, I explored the 'Class' column. I wanted to know which taxonomic classes were poached most often. Figure 9 depicts the count for Classes for the entire dataset. Ass Appendix I contains the subset of species threatened with extinction by poaching I decided to subset this visually in Figure 10. From this graph I noted that the top three classes were Mamilia, Aves, and Reptilia. This was interesting to me as plants are high risk for poaching because of their transportability. It is then That I noticed that neither plant Class was represented in this data set so I decided to explore the NA Class.

Figure 11. shows the breakdown of the Mamilia class by type of good for Appendix I species. From this information we see that the most common good type is ivory carvings followed by live animals. In figure 12 we see the same breakdown for Appendix I Aves. In this figure, we see that the trade for this class is dominated by live animals. Figure 13 depicts the breakdown for Reptilia in Appendix I. From this graph we see the most commonly traded good is small leather products followed by skins. To explore my hunch that Appendix I plants were hidden in the NAs, I assigned them a variable and explored the good type breakdown in Figure 14. Here, I saw that the NAs were dominated by live animals followed by Seeds. From this exploration it is clear that the plant class is embedded in the Class NAs, but there is not enough information to conclusively assign them to the plant Classes. More investigation of the dataframe is needed and this is outside the scope of my project.

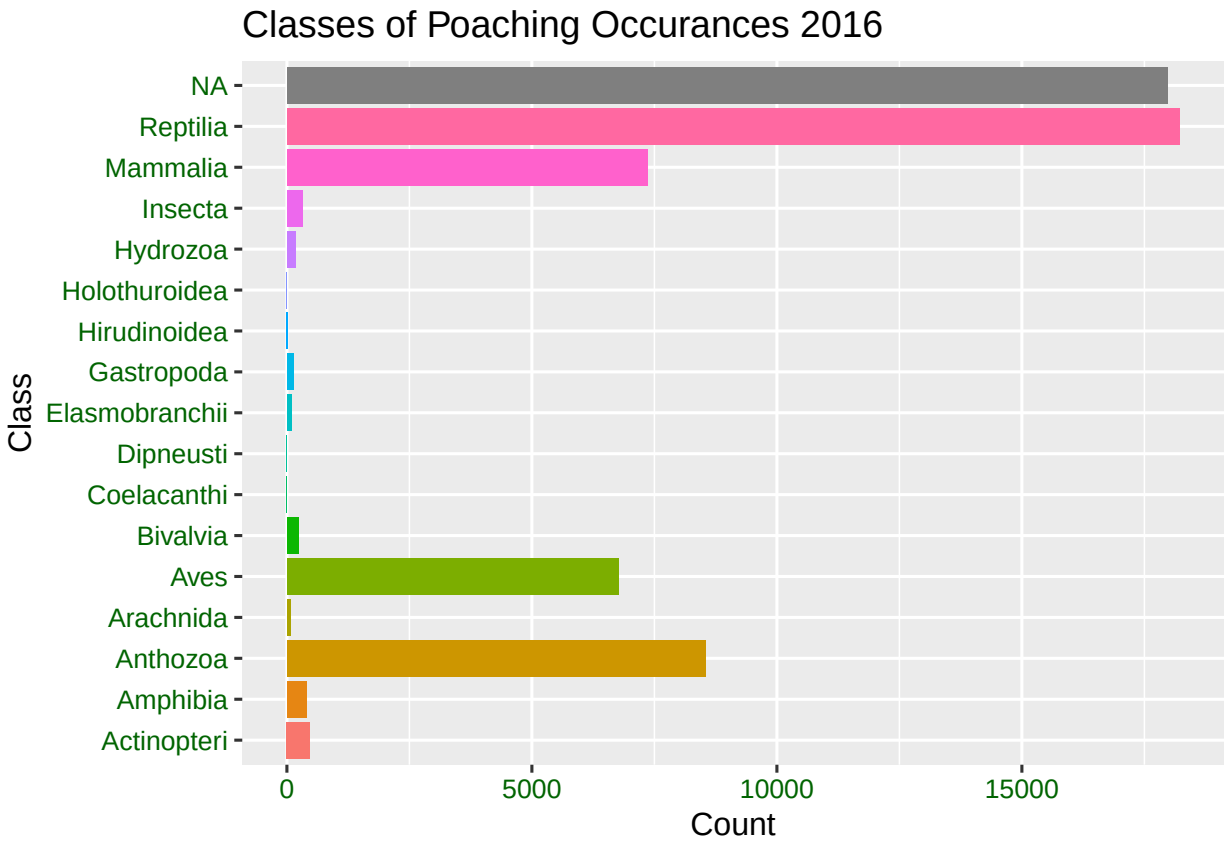

Figure 9. Classes of Poaching Occurrences 2016

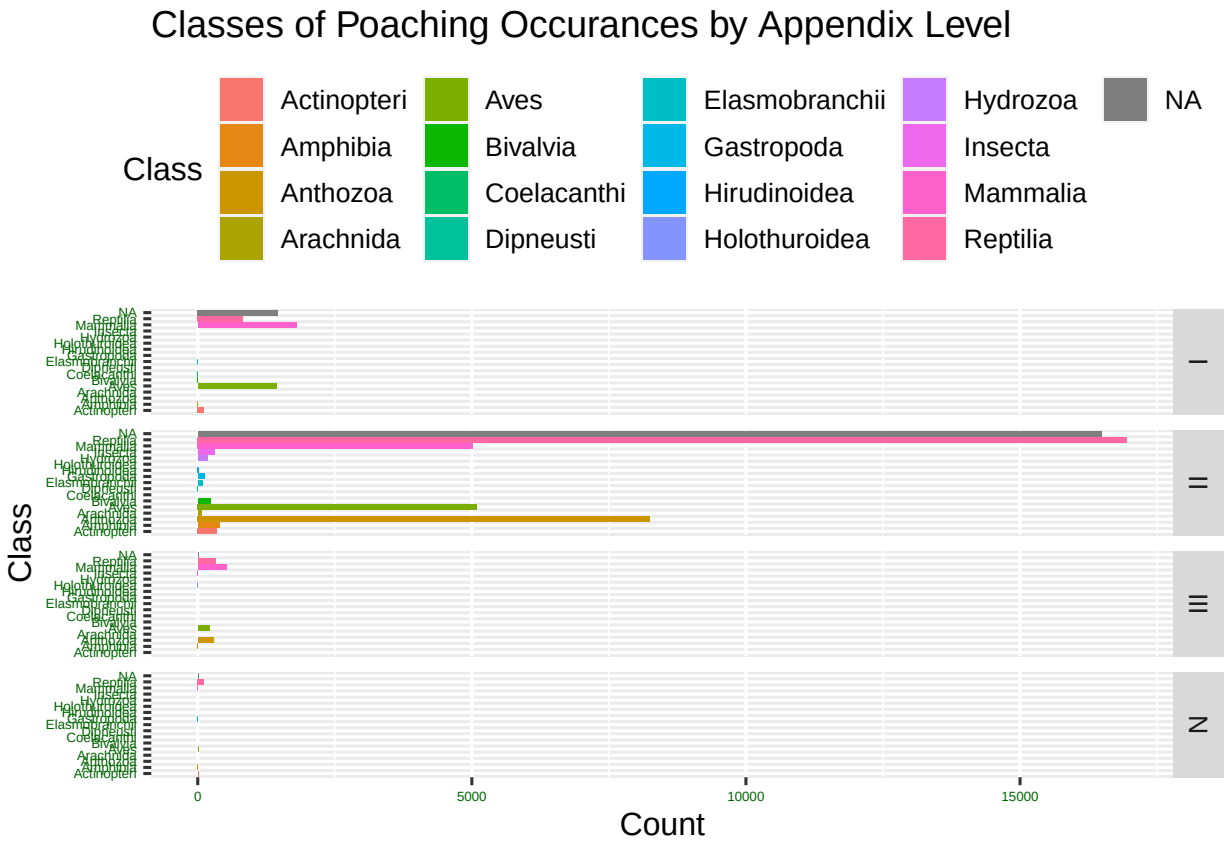

Figure 10. Classes of Poaching Occurrences per Appendix Level

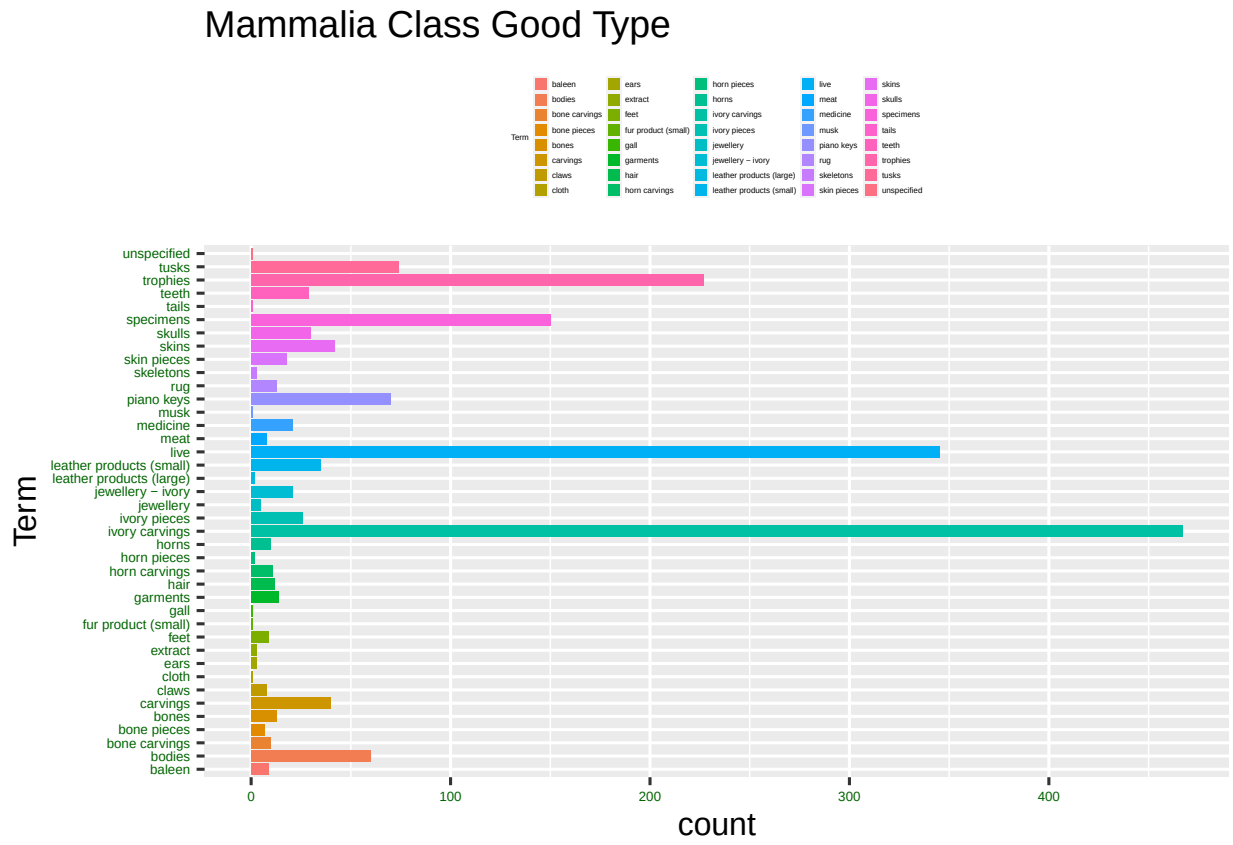

Figure 11. Mammalia Class Good Type

Figure 12. Aves Class Good Type

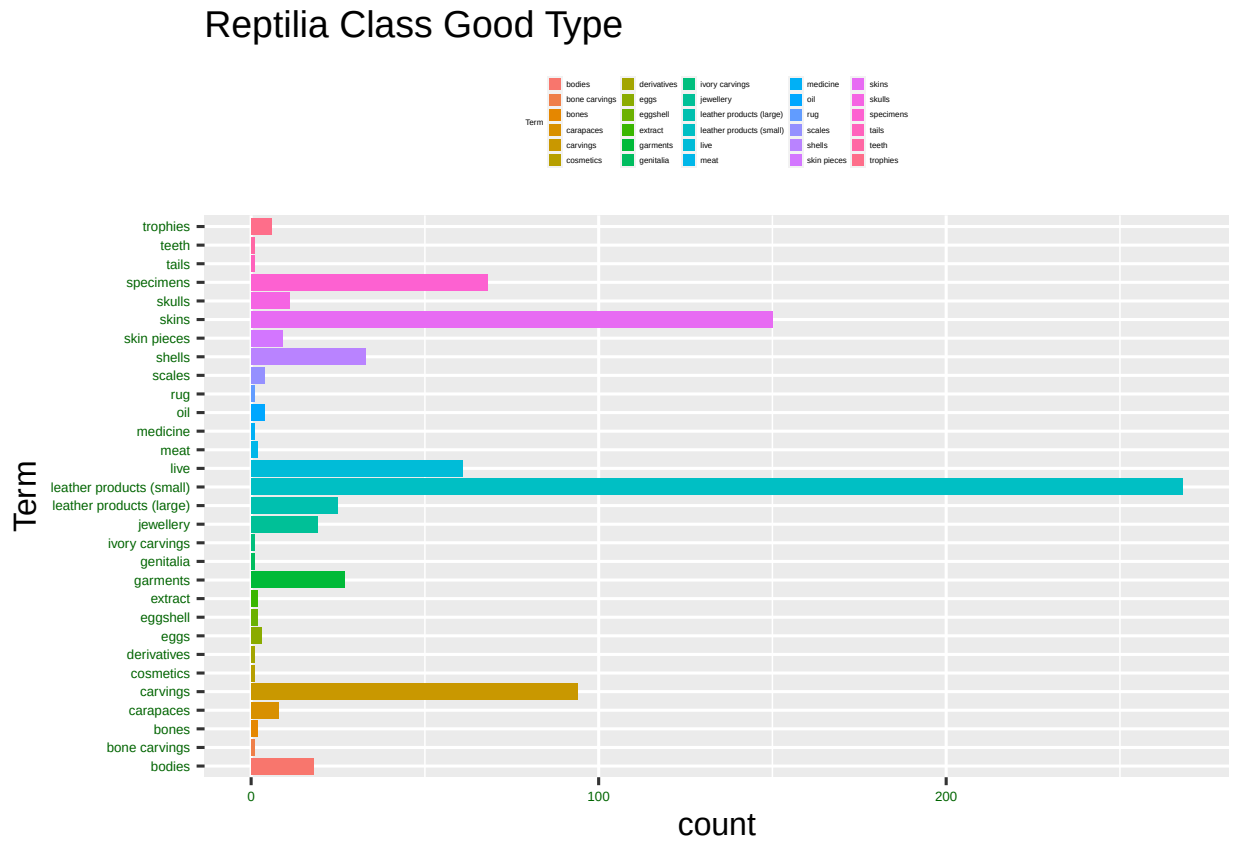

Figure 13. Reptilia Class Good Type

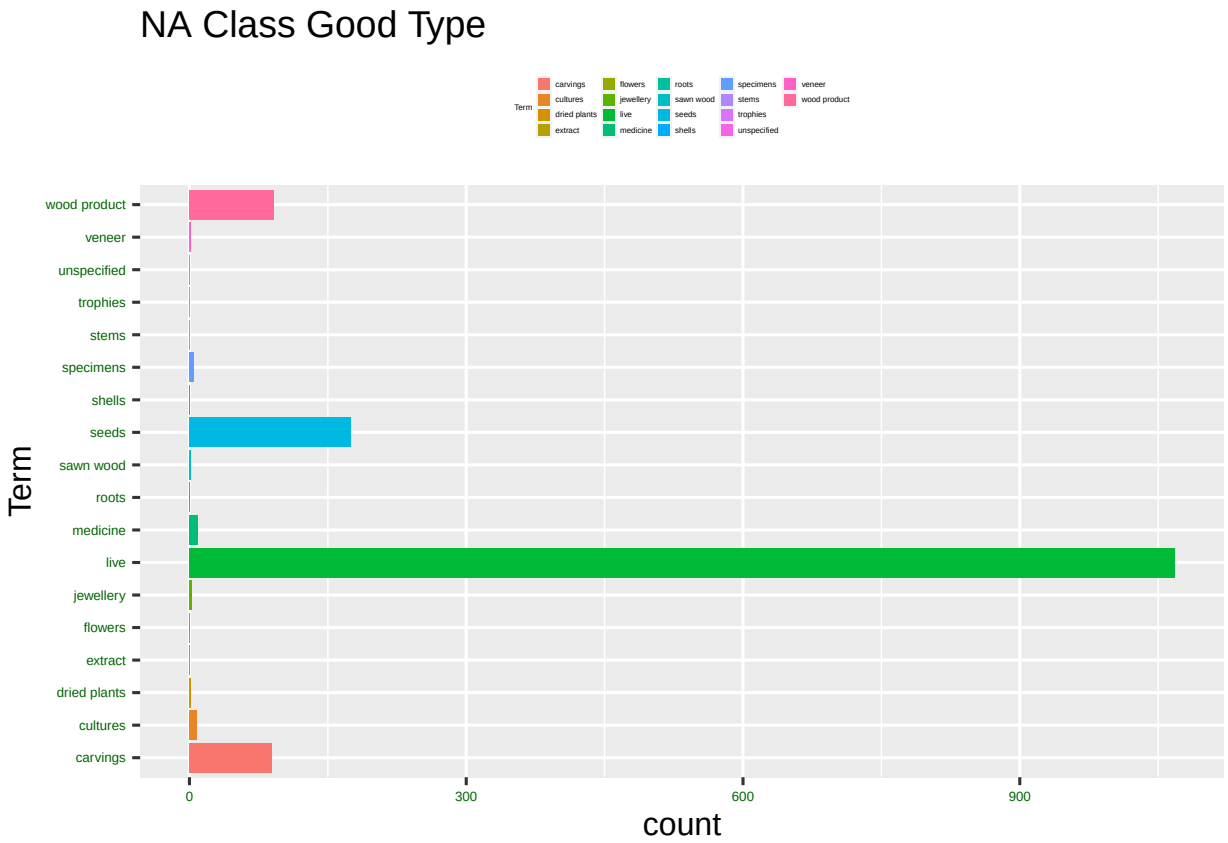

Figure 14. NA Class Good Type

4 Analysis

4.1 Question 1: Is there any correlation between class, term, appendix number and total poaching amount?

For my main analysis, I wanted to know if the mean of poaching amount changes according to levels of Class, Term, and Appendix. These last three variables are categorical. Because of this, I decided to use two-way ANOVAs to understand the relationship between these variables. I also wanted to see if the combination of these variables affected the mean of total poaching amount.

1. For my first model I wanted to know how Poaching Quanitity is influenced by Class and Term only within the Appendix I subset.

H: There is no difference in the means of Class in Appendix I H: There is no difference in means of Term in Appendix I H: There is no interaction between Class and Term in Appendix I

Null Hypothesis: The alternative hypothesis for cases 1 and 2 is: the means are not equal.

Null Hypothesis: The alternative hypothesis for case 3 is: there is an interaction between Class and Term for Appendix I

The summary of our two-way ANOVA (Table. 2) shows us a statistically significant p-value for Class (DF= 7, P-value = <2e-16). This means that for Class we reject the null hypothesis that the mean across our groups is different. The p-value for term is also statistically significant (DF= 52, P-value = <0.256). In this case we also reject the null hypothesis. This model only gives us part of the picture so I also used a two-way LM.

```
##               Df    Sum Sq   Mean Sq F value   Pr(>F)
## Class          6 4.467e+09 744583316   9.911 1.37e-10 ***
## Term          26 3.259e+09 125332949   1.668   0.0198 *
## Residuals    853 6.408e+10  75125682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 4768 observations deleted due to missingness
```

Table 2. Two Way ANOVA Summary

This model gives us a detailed summary of the groups within class which are significant. The results can be viewed below in Table 3.The results of this model show high chance of variability (Residual standard error: 4170 on 4132 degrees of freedom). This means that the actual recorded count can deviate from the true regression line by approximately 4170 occurrences. We also see that our multiple R squared is low (Multiple R-squared: 0.06678) and therefore our model does not explain the variance recorded quantity well.

```
##
## Call:
## lm(formula = Reported_quantity ~ Class + Term, data = AppendixI_Facet)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11288    -54    -40    -11 201942
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1.128e+04  6.391e+03   1.764    0.078 .
## ClassAmphibia                  -1.169e+04  9.028e+03  -1.295    0.196
## ClassAves                      -1.123e+04  1.542e+03  -7.282 7.46e-13 ***
## ClassCoelacanthi               -1.124e+04  9.390e+03  -1.197    0.232
## ClassMammalia                  -1.127e+04  1.812e+03  -6.220 7.76e-10 ***
## ClassNot_recorded              -1.123e+04  1.521e+03  -7.386 3.59e-13 ***
## ClassReptilia                  -1.087e+04  2.602e+03  -4.175 3.28e-05 ***
## Termbodies                     -3.633e+01  7.032e+03  -0.005    0.996
## Termbones                       2.100e+01  1.062e+04   0.002    0.998
## Termcarvings                   -2.632e+02  6.612e+03  -0.040    0.968
## Termcloth                       2.000e+00  1.062e+04   0.000    1.000
## Termcosmetics                   5.633e+04  1.083e+04   5.204 2.45e-07 ***
## Termeggs                        7.105e+01  7.359e+03   0.010    0.992
## Termfeathers                   -3.855e+01  1.069e+04  -0.004    0.997
## Termfeet                       -2.000e+00  8.668e+03   0.000    1.000
## Termgarments                   -2.610e+02  7.397e+03  -0.035    0.972
## Termhair                        2.260e+02  1.062e+04   0.021    0.983
## Termhorn carvings              -2.000e+00  1.062e+04   0.000    1.000
## Termivory carvings              2.679e+01  6.248e+03   0.004    0.997
## Termivory pieces                1.570e-11  1.062e+04   0.000    1.000
## Termleather products (large)   -4.078e+02  1.083e+04  -0.038    0.970
## Termleather products (small)   -3.057e+02  6.644e+03  -0.046    0.963
## Termlive                        1.360e+01  6.230e+03   0.002    0.998
## Termoil                         1.063e+04  1.083e+04   0.982    0.326
## Termpiano keys                  2.540e+02  7.912e+03   0.032    0.974
## Termskin pieces                 1.765e+03  1.083e+04   0.163    0.871
## Termskins                       3.111e+02  6.609e+03   0.047    0.962
## Termskulls                      5.000e+00  8.668e+03   0.001    1.000
## Termspecimens                   4.168e+02  6.420e+03   0.065    0.948
## Termteeth                       1.880e+02  8.668e+03   0.022    0.983
## Termtrophies                    8.776e+00  6.253e+03   0.001    0.999
## Termtusks                       1.830e+01  6.292e+03   0.003    0.998
## Termwood product               -1.015e+01  6.776e+03  -0.001    0.999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8668 on 853 degrees of freedom
##   (4768 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.1076, Adjusted R-squared:  0.07412
## F-statistic: 3.214 on 32 and 853 DF,  p-value: 9.76e-09
```

Table 3. Two Way LM Summary 2

The Last thing I did for this model was check to see if there was a relationship between the dependant variables. The results (Table 4) show a no significant interaction between the variables (P-Value = 0.954).

```
##              Df    Sum Sq   Mean Sq F value  Pr(>F)
## Class         6 4.467e+09 744583316   9.830 1.7e-10 ***
## Term         26 3.259e+09 125332949   1.655  0.0215 *
## Class:Term    7 1.259e+06    179886   0.002  1.0000
## Residuals   846 6.408e+10  75745801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 4768 observations deleted due to missingness
```

Table 4. Two Way ANOVA comparing Class and Term

```
Poaching.2way <- aov(data = AppendixI_Facet, Reported_quantity ~ Class + Term)

# P-value is small. Reject our null hypothesis that the mean across our groups
# is different.  Not enough information to tell us which groups are differing

Poaching.2way2lm <- lm(data = AppendixI_Facet, Reported_quantity ~ Class + Term)

# means are different. Reject null

# TukeyHSD(Poaching.2way)

# check interaction between the variables

Poaching.2way3 <- aov(data = AppendixI_Facet, Reported_quantity ~ Class * Term)

# Interaction is not significant between variables
```

2. For my second model I wanted to know how Poaching Quantity is influenced by Class and Appendix for the entire dataframe

H: There is no difference in the means of Class H: There is no difference in means of Appendix H: There is no interaction between Class and Appendix

Null Hypothesis: The alternative hypothesis for cases 1 and 2 is: the means are not equal.

Null Hypothesis: The alternative hypothesis for case 3 is: there is an interaction between Class and Appendix

The summary of our two-way ANOVA (Table. 5) shows us a statistically significant p-value

for Class (DF= 7, P-value = <2e-16). This means that for Class we reject the null hypothesis that the mean across our groups is different. The p-value for Appendix is also statistically significant (DF= 3, P-value = <2e-16). In this case we can also reject the null hypothesis.

```
##               Df    Sum Sq   Mean Sq F value Pr(>F)
## Class         13 1.463e+10 1.125e+09    1.61 0.0745 .
## Appendix       2 8.712e+10 4.356e+10   62.35 <2e-16 ***
## Residuals   6785 4.740e+12 6.986e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 53958 observations deleted due to missingness
```

Table 5. Model 2 - Two Way ANOVA Summary

This model gives us a detailed summary of the groups within class which are significant. The results can be viewed below in Table 6.The results of this model show high chance of variability (Residual standard error: 17690 on 42772 degrees of freedom). This means that the actual recorded count can deviate from the true regression line by approximately 17690 occurrences. Again, we also see that our multiple R squared is low (Multiple R-squared: 0.01144) and therefore our model does not explain all the variance in reported quantity.

```
##
## Call:
## lm(formula = Reported_quantity ~ Class + Appendix, data = CITES2016_Processed)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -31209   -2443   -2022    -761 1288663
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)            13105       3274   4.002 6.35e-05 ***
## ClassAmphibia         -13713       4232  -3.240 0.001202 **
## ClassAnthozoa         -11972       3380  -3.542 0.000400 ***
## ClassArachnida        -13015      10520  -1.237 0.216079
## ClassAves             -11703       3350  -3.493 0.000480 ***
## ClassBivalvia         -13445       5845  -2.300 0.021456 *
## ClassCoelacanthi      -13103      26632  -0.492 0.622749
## ClassElasmobranchii   -13968      12272  -1.138 0.255043
## ClassGastropoda         2839       8312   0.342 0.732683
## ClassHirudinoidea     -11188      18978  -0.590 0.555539
## ClassHydrozoa         -13835       7049  -1.963 0.049712 *
## ClassInsecta          -13527       5138  -2.633 0.008492 **
## ClassMammalia         -14989       3374  -4.442 9.03e-06 ***
## ClassReptilia         -11495       3322  -3.461 0.000542 ***
## AppendixII               883       1215   0.727 0.467556
## AppendixIII            29602       2777  10.659  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26430 on 6785 degrees of freedom
##   (53958 observations deleted due to missingness)
## Multiple R-squared:  0.02101,    Adjusted R-squared:  0.01885
## F-statistic:  9.71 on 15 and 6785 DF,  p-value: < 2.2e-16
```

Table 6. Model 2 - Two Way LM Summary 2

I decided to check the interaction between Class and Appendix using a new ANOVA model. In this case, the P-value was statistically significant (DF= 16, P-value = <2e-16). Because of this we accept the Null hypothesis that there is an interaction between Class and Appendix. This relationship between Class and Appendix when compared to reported quantity can be visualized in Figure 15.

```
##                   Df    Sum Sq   Mean Sq F value Pr(>F)
## Class             13 1.463e+10 1.125e+09   1.748 0.0453 *
## Appendix           2 8.712e+10 4.356e+10  67.690 <2e-16 ***
## Class:Appendix     8 3.787e+11 4.734e+10  73.564 <2e-16 ***
## Residuals       6777 4.361e+12 6.435e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 53958 observations deleted due to missingness
```

Table 7. Model 2 - Two Way ANOVA Comparing Class and Appendix

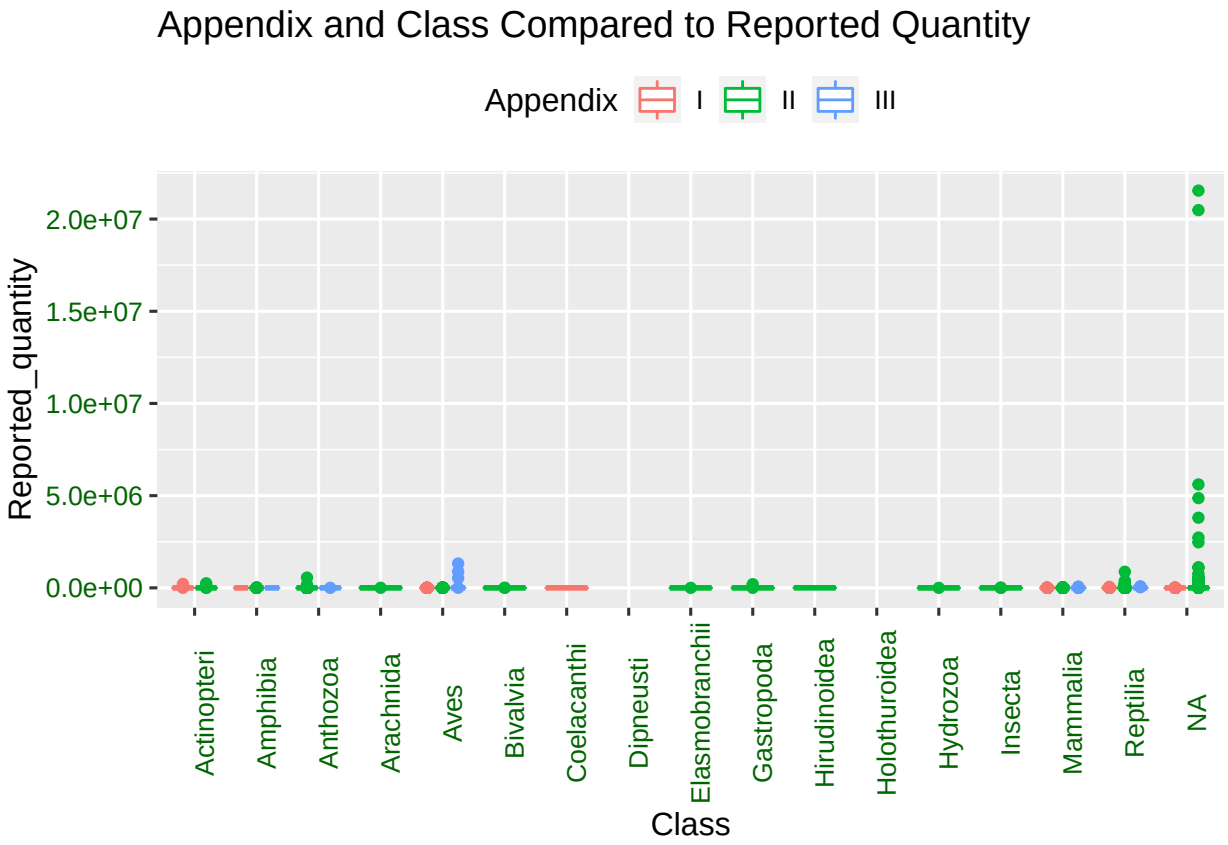

Figure 15. Appendix and Class Compared to Reported Quantity

```
# 2. How Does Reported Poaching Quantity vary over Class and Appendix for
# entire data

Poaching.2way.test <- aov(data = CITES2016_Processed, Reported_quantity ~ Class +
    Appendix)

# P-value is small. Reject our null hypothesis that the mean across our groups
# is different.  Not enough information to tell us which groups are differing

Poaching.2way2lm.test <- lm(data = CITES2016_Processed, Reported_quantity ~ Class +
    Appendix)

# means are different. Reject null

TukeyHSD(Poaching.2way.test)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Reported_quantity ~ Class + Appendix, data = CITES2016_Processed)
##
```

```
## $Class
##                                       diff         lwr         upr     p adj
## Amphibia-Actinopteri          -12944.04086  -26971.327   1083.2455 0.1070969
## Anthozoa-Actinopteri          -11235.95598  -22350.741   -121.1712 0.0445431
## Arachnida-Actinopteri         -12527.11301  -47748.305  22694.0794 0.9959387
## Aves-Actinopteri              -10987.64711  -22136.056    160.7616 0.0580844
## Bivalvia-Actinopteri          -12957.77015  -32436.582   6521.0413 0.6061167
## Coelacanthi-Actinopteri       -13497.97015 -102826.805  75830.8651 0.9999998
## Elasmobranchii-Actinopteri    -13480.77015  -54588.024  27626.4833 0.9981976
## Gastropoda-Actinopteri          3326.77985  -24467.802  31121.3612 1.0000000
## Hirudinoidea-Actinopteri      -10699.97015  -74327.749  52927.8086 0.9999991
## Hydrozoa-Actinopteri          -13347.69237  -36887.907  10192.5226 0.8257038
## Insecta-Actinopteri           -13039.12570  -30129.043   4050.7919 0.3669473
## Mammalia-Actinopteri          -12916.68103  -24167.268  -1666.0943 0.0088847
## Reptilia-Actinopteri          -10809.15892  -21752.623    134.3051 0.0568147
## Anthozoa-Amphibia               1708.08488   -7544.368  10960.5378 0.9999973
## Arachnida-Amphibia               416.92785  -34261.622  35095.4779 1.0000000
## Aves-Amphibia                   1956.39374   -7336.424  11249.2114 0.9999869
## Bivalvia-Amphibia                -13.72929  -18493.273  18465.8145 1.0000000
## Coelacanthi-Amphibia            -553.92929  -89670.203  88562.3444 1.0000000
## Elasmobranchii-Amphibia         -536.72929  -41180.003  40106.5448 1.0000000
## Gastropoda-Amphibia            16270.82071  -10832.836  43374.4772 0.7591836
## Hirudinoidea-Amphibia           2244.07071  -61084.941  65573.0820 1.0000000
## Hydrozoa-Amphibia               -403.65152  -23123.932  22316.6294 1.0000000
## Insecta-Amphibia                 -95.08485  -16036.689  15846.5188 1.0000000
## Mammalia-Amphibia                 27.35982   -9387.795   9442.5149 1.0000000
## Reptilia-Amphibia               2134.88194   -6911.048  11180.8119 0.9999505
## Arachnida-Anthozoa             -1291.15703  -34897.340  32315.0257 1.0000000
## Aves-Anthozoa                    248.30886   -3375.245   3871.8624 1.0000000
## Bivalvia-Anthozoa              -1721.81417  -18100.679  14657.0509 1.0000000
## Coelacanthi-Anthozoa           -2262.01417  -90966.489  86442.4610 1.0000000
## Elasmobranchii-Anthozoa        -2244.81417  -41977.035  37487.4071 1.0000000
## Gastropoda-Anthozoa            14562.73583  -11154.610  40280.0817 0.8270863
## Hirudinoidea-Anthozoa            535.98583  -62212.220  63284.1918 1.0000000
## Hydrozoa-Anthozoa              -2111.73640  -23158.942  18935.4688 1.0000000
## Insecta-Anthozoa               -1803.16973  -15253.389  11647.0499 0.9999999
## Mammalia-Anthozoa              -1680.72506   -5607.406   2245.9562 0.9777372
## Reptilia-Anthozoa                426.79706   -2506.210   3359.8038 0.9999999
## Aves-Arachnida                  1539.46589  -32077.852  35156.7842 1.0000000
## Bivalvia-Arachnida              -430.65714  -37649.754  36788.4397 1.0000000
## Coelacanthi-Arachnida           -970.85714  -95762.618  93820.9034 1.0000000
## Elasmobranchii-Arachnida        -953.65714  -52873.243  50965.9284 1.0000000
## Gastropoda-Arachnida           15853.89286  -26316.901  58024.6864 0.9930787
## Hirudinoidea-Arachnida          1827.14286  -69266.678  72920.9633 1.0000000
## Hydrozoa-Arachnida              -820.57937  -40317.146  38675.9875 1.0000000
```

```
## Insecta-Arachnida             -512.01270  -36538.425  35514.3993 1.0000000
## Mammalia-Arachnida            -389.56803  -34040.909  33261.7734 1.0000000
## Reptilia-Arachnida            1717.95409  -31831.956  35267.8644 1.0000000
## Bivalvia-Aves                -1970.12304  -18371.824  14431.5780 1.0000000
## Coelacanthi-Aves             -2510.32304  -91219.018  86198.3716 1.0000000
## Elasmobranchii-Aves          -2493.12304  -42234.763  37248.5174 1.0000000
## Gastropoda-Aves              14314.42696  -11417.469  40046.3226 0.8447708
## Hirudinoidea-Aves              287.67696  -62466.494  63041.8476 1.0000000
## Hydrozoa-Aves                -2360.04526  -23425.026  18704.9357 1.0000000
## Insecta-Aves                 -2051.47859  -15529.497  11426.5399 0.9999997
## Mammalia-Aves                -1929.03392   -5949.905   2091.8369 0.9440651
## Reptilia-Aves                  178.48820   -2879.469   3236.4458 1.0000000
## Coelacanthi-Bivalvia          -540.20000  -90675.485  89595.0849 1.0000000
## Elasmobranchii-Bivalvia       -523.00000  -43354.473  42308.4728 1.0000000
## Gastropoda-Bivalvia          16284.55000  -14001.875  46570.9749 0.8748846
## Hirudinoidea-Bivalvia         2257.80000  -62497.300  67012.9002 1.0000000
## Hydrozoa-Bivalvia             -389.92222  -26826.081  26046.2368 1.0000000
## Insecta-Bivalvia               -81.35556  -20980.974  20818.2632 1.0000000
## Mammalia-Bivalvia               41.08912  -16430.234  16512.4121 1.0000000
## Reptilia-Bivalvia             2148.61123  -14114.481  18411.7039 1.0000000
## Elasmobranchii-Coelacanthi      17.20000  -97115.450  97149.8504 1.0000000
## Gastropoda-Coelacanthi        16824.75000  -75465.467 109114.9674 0.9999977
## Hirudinoidea-Coelacanthi      2798.00000 -105799.605 111395.6045 1.0000000
## Hydrozoa-Coelacanthi           150.27778  -90949.048  91249.6038 1.0000000
## Insecta-Coelacanthi            458.84444  -89190.532  90108.2213 1.0000000
## Mammalia-Coelacanthi           581.28912  -88140.305  89302.8828 1.0000000
## Reptilia-Coelacanthi          2688.81123  -85994.360  91371.9826 1.0000000
## Gastropoda-Elasmobranchii    16807.55000  -30390.434  64005.5343 0.9958896
## Hirudinoidea-Elasmobranchii   2780.80000  -71405.487  76967.0871 1.0000000
## Hydrozoa-Elasmobranchii        133.07778  -44691.611  44957.7662 1.0000000
## Insecta-Elasmobranchii         441.64444  -41357.593  42240.8819 1.0000000
## Mammalia-Elasmobranchii        564.08912  -39206.335  40334.5137 1.0000000
## Reptilia-Elasmobranchii       2671.61123  -37013.025  42356.2477 1.0000000
## Hirudinoidea-Gastropoda     -14026.75000  -81749.255  53695.7549 0.9999892
## Hydrozoa-Gastropoda         -16674.47222  -49719.671  16370.7265 0.9188871
## Insecta-Gastropoda          -16365.90556  -45174.042  12442.2308 0.8237102
## Mammalia-Gastropoda         -16243.46088  -42019.790   9532.8682 0.6916709
## Reptilia-Gastropoda         -14135.93877  -39779.707  11507.8293 0.8533184
## Hydrozoa-Hirudinoidea        -2647.72222  -68738.120  63442.6752 1.0000000
## Insecta-Hirudinoidea         -2339.15556  -66416.173  61737.8621 1.0000000
## Mammalia-Hirudinoidea        -2216.71088  -64989.114  60555.6924 1.0000000
## Reptilia-Hirudinoidea         -109.18877  -62827.275  62608.8973 1.0000000
## Insecta-Hydrozoa               308.56667  -24420.196  25037.3290 1.0000000
## Mammalia-Hydrozoa              431.01134  -20688.224  21550.2468 1.0000000
## Reptilia-Hydrozoa             2538.53346  -18418.704  23495.7711 1.0000000
```

```
## Mammalia-Insecta               122.44467  -13440.212  13685.1018 1.0000000
## Reptilia-Insecta              2229.96679  -11079.029  15538.9627 0.9999992
## Reptilia-Mammalia             2107.52212   -1304.192   5519.2358 0.7205625
##
## $Appendix
##             diff       lwr       upr     p adj
## II-I    1229.498 -1405.787  3864.783 0.5179994
## III-I  29383.421 22913.602 35853.240 0.0000000
## III-II 28153.923 22139.310 34168.535 0.0000000
```

```
# check interaction between the variables

Poaching.2way3 <- aov(data = CITES2016_Processed, Reported_quantity ~ Class * Appendix)

# significant because the pvalue is <2e-16

# TukeyHSD(Poaching.2way3)

Poaching.interaction <- with(CITES2016_Processed, interaction(Class, Appendix))

Poaching.anova.2way5 <- aov(data = CITES2016_Processed, Reported_quantity ~ Poaching.int

Poaching.groups <- HSD.test(Poaching.anova.2way5, "Poaching.interaction", group = TRUE)

Poaching.anova.plot <- ggplot(CITES2016_Processed, aes(y = Reported_quantity, x = Class,
    color = Appendix)) + geom_boxplot() + theme(axis.text.x = element_text(angle = 90))
    labs(title = "Appendix and Class compared to Reported Quantity")
```

5 Summary and Conclusions

I found that there was a correlation between Class, Term, APpendix and Reported Quantity. However, I also found that neither of my models was explaining the variability in our response variable well. More data analysis is needed to find what else is missing from these models to better explain this variance.

I was also able to visually compare export and import quantity, as well as identify and highlight the countries with the highest import and export amounts. This was accomplished through a bar graphs and a maps located in an additional document.

Another goal of mine was to explore Appendix I specifically. This Appendix contains species in which poaching can directly lead to extinction. I was able to identify the top three classes in the Appendix and explore the good types that are most commonly traded. I also had time to explore the case of missing plant classes and identified them within the NAs.

In conclusion, I have increased the understanding of the CITES 2016 Poaching dataframe. This information can be used to inform policies surrounding poaching, wildlife trafficking and trade internationally.