

Tipología y Ciclo de vida de los datos

Práctica 2 - Limpieza - Análisis de datos



Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres

José Ramón Martínez-Carbonell Martín
Miguel Ángel Pérez García

Apartados

1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?

La globalización en la que nos encontramos inmersos en pleno año 2021, hace que el comercio haya crecido tanto en volumen como en posibilidades de intercambio internacional. Un aspecto importante de este comercio es el que se realiza con animales. El CITES (<https://trade.cites.org/>) es la organización que se encarga de monitorear y registrar este intercambio. No forma parte de la ONU como tal, pero esta sí que está pendiente de los informes realizados por la IUCN, entidad que está por encima de CITES. Esta organización elabora anualmente un informe en el que comenta el tipo de especies que se intercambian, así como diversos datos asociados a ellas. Este es el dataset elegido, debido a que creemos importante poder extraer conclusiones acerca del comportamiento del comercio con este tipo de elementos.

A continuación mostramos una serie de preguntas que intentaremos responder con nuestro análisis de este dataset:

- ¿Qué tipo de especie son las importadas en los reportes?
- ¿Qué país es el que más exportaciones reporta?
- ¿Existe alguna región o país especialmente activa con algún tipo de animal? Esto nos podría llevar a deducir que hay que poner especial vigilancia con este animal en ese país, dado que el mercado negro de animales puede estar interesado.
- ¿El comercio de animales con fines de creación de pieles, tiene alguna relación con algún determinado país/área?.
- ¿Se prioriza el comercio de animales como producto (pieles, alimento...) frente al comercio de animales vivos?.
- ¿Cuáles es el animal más demandado cuyo comercio amenaza con la desaparición?

Cabe destacar que estas no son las únicas preguntas que se responderán, dado que la exploración de los datos nos llevará a indagar en otras cuestiones que no hemos planteado inicialmente.

Seguidamente se detallarán los atributos con las que contamos inicialmente en el dataset:

Campo	Descripción
year	Año en el que tiene lugar la importación o exportación. En este caso nos centraremos en el año 2020 y 2021.
app.	Abreviación de apéndice, el cual puede ser de tipo I, II o III. Indica el grado de protección de la especie, siendo las de tipo I las más protegidas.
taxon	Identificación de la especie por medio de su taxonomía.
class	Identificación taxonómica de la especie por medio de su clase.
order	Identificación taxonómica de la especie por medio de su orden.
family	Identificación taxonómica de la especie por medio de su familia.
genus	Identificación de la especie por su nombre científico.
importer	País que importa la especie.
exporter	País que exporta la especie.
origin	País de origen de la especie.
imported reported quantity	Número de especies importadas de este tipo. Si es una exportación este campo estará a cero.
exporter reported quantity	Número de especies exportadas de este tipo. Si es una importación este campo estará a cero.
term	Concepto del intercambio. Esta variable informa de si se ha realizado el intercambio con un fin de obtener una piel, de que el animal viva, de trofeo....
unit	Cantidad de animales que se comercia. Si no aparece nada es que el animal se ha traspasado por completo.
purpose	Propósito con el cual se realiza la importación/exportación.
source	Variable que explica cómo se obtuvo la especie para su intercambio, es decir, si procedía del medio natural, si

	había sido criada en cautividad, si era un espécimen de zoo etc etc.
--	--

2. Integración y selección de los datos de interés a analizar.

El conjunto de datos que se presenta posee una serie de elementos que hemos comentado en el apartado anterior. Tras una exploración nos damos cuenta de que la variable **Origin** no nos interesa, dado que la misma está ausente en múltiples registros y además no nos aporta información de gran valor. Tampoco podemos inferir de un modo lógico este valor ausente, por lo que creemos necesaria eliminarla del dataset. Cabe destacar que el comando para este paso aparecerá en la siguiente sección.

Por otro lado, vemos que hay una columna que indica el número de importaciones y otra que nos informa del número de exportaciones, por lo que vamos a optar por unificarla en una única columna. Esto lo hacemos porque cuando se trata de una importación, ese registro en la columna de exportación no tiene ningún dato y viceversa. Además, añadimos una nueva columna para diferenciar si el reporte registrado se trata de una importación o una exportación. Este tratamiento se realiza con el siguiente fragmento de código:

```
```${r}
#Declaramos las nuevas columnas que utilizaremos
auxNumber <- c()
auxType <- c()
#Según el si el campo de importación/exportación está vacío o no,
#vamos creando el vector según convenga
for (i in 1:nrow(cites)) {
 if(!is.na(cites$Importer.reported.quantity[i])){
 auxNumber[i] <- cites$Importer.reported.quantity[i]
 auxType[i] <- "Imported"
 }else if(!is.na(cites$Exporter.reported.quantity[i])){
 auxNumber[i] <- cites$Exporter.reported.quantity[i]
 auxType[i] <- "Exported"
 }
}
#Asignamos las nuevas columnas al dataset
cites$Exportation.type <- auxType
cites$Exportation.number <- auxNumber
````
```

El resto de columnas del dataset sí que nos interesan. A continuación pasamos a explicar y detallar cómo tratamos los nulos y/o vacíos.

- Caso de unit -> cuando sea nulo se pondrá no aplica.
- Class -> meter valor ausente.
- Origin -> meter valor ausente
- Purpose y source se tratarán de la misma manera.

3. Limpieza de los datos.

3.1. ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos?

La mayoría de nuestros datos son categóricos. Los que son numéricos son las columnas *Importer reported quantity* y *Exporter reported quantity* (las cuales fusionaremos en *Exportation Number*). Podríamos eliminar aquellos registros que tengan un valor ausente, aunque perderíamos capacidad de análisis y datos, así que seguiremos otra estrategia:

Los datos del dataset que son numéricos no presentan ceros como valor de ausencia, aunque sí se detectan valores vacíos o N/A.

```
sapply(cites, function(x) sum(is.na(x))) # NA
```

| | | |
|----|----------------------------|----------------------------|
| ## | Year | App. |
| ## | 0 | 0 |
| ## | Taxon | Class |
| ## | 0 | 0 |
| ## | Order | Family |
| ## | 0 | 0 |
| ## | Genus | Importer |
| ## | 0 | 0 |
| ## | Exporter | Origin |
| ## | 0 | 0 |
| ## | Importer.reported.quantity | Exporter.reported.quantity |
| ## | 769 | 144 |
| ## | Term | Unit |
| ## | 0 | 0 |
| ## | Purpose | Source |
| ## | 0 | 0 |

Los datos que son categóricos presentan como valor de ausencia una cadena vacía.

```
sapply(cites, function(x) sum(x=="")) #Si es vacio
```

```
##           Year           App.
##           0             0
##           Taxon          Class
##           0             415
##           Order          Family
##           0             0
##           Genus           Importer
##           7             0
##           Exporter        Origin
##           0             844
## Importer.reported.quantity Exporter.reported.quantity
##           NA             NA
##           Term            Unit
##           0             753
##           Purpose          Source
##           33             34
```

```
sapply(cites, function(x) sum(x==" ")) # Si no es vacio, aunque tiene un espacio
```

```
##           Year           App.
##           0             0
##           Taxon          Class
##           0             0
##           Order          Family
##           0             0
##           Genus           Importer
##           0             0
##           Exporter        Origin
##           0             0
## Importer.reported.quantity Exporter.reported.quantity
##           NA             NA
##           Term            Unit
##           0             0
##           Purpose          Source
##           0             0
```

Para el caso de los datos numéricos, dada su naturaleza original del conjunto de datos y como se ha comentado anteriormente se fusionan las columnas. Para el caso de los atributos no numéricos, se reemplazará el valor vacío por un “Not applied”.

```
#sustición de cadena vacía por "Not applied"
for (i in 1:nrow(cites)) {
  for(j in 1:ncol(cites)){
    if((cites[i,j]=="")){
      cites[i,j] <- "Not applied"
    }
  }
}
```

3.2. Identificación y tratamiento de valores extremos.

En nuestro dataset existen algunos outliers, como por ejemplo para el campo Importer/Exporter reported quantity cuyo valor es de 194000 unidades, que se contempla como válido porque en este caso son plantas que se han criado artificialmente y que han distribuido en la naturaleza.

Como son valores legítimos aunque se alejan de la distribución normal, no se eliminarán de nuestro conjunto.

```
cites[duplicated(cites),]
```

```
##      Year App.      Taxon  Class  Order  Family  Genus
## 34 2020      I Pteronura brasiliensis Mammalia Carnivora Mustelidae Pteronura
##      Importer Exporter Origin Importer.reported.quantity
## 34      CO      US                      1
##      Exporter.reported.quantity Term Unit Purpose Source
## 34                      NA live          Z      F
```

Nuestro dataset contiene un intercambio duplicado, que procederemos a eliminarlo porque no tenemos una fecha del intercambio. Si tuviéramos una fecha podemos saber si son dos transacciones diferentes.

Para ir terminando, borramos las columnas Origin, Importer reported quantity y Exporter reported quantity. Además eliminaremos también las columnas Order, Family y Genus, ya que hacen alusión a términos científicos que además no nos van aportar valor. Para identificar a un animal usaremos su “nombre”, que viene dado por la columna *taxon* y su clase, que vendría a ser la clasificación de mamíferos, aves, reptiles...

```
` ``{r}
#Eliminamos las columnas que ya no nos interesan
cites$Origin <- NULL
cites$Importer.reported.quantity <- NULL
cites$Exporter.reported.quantity <- NULL
cites$Family <- NULL
cites$Order <- NULL
cites$Genus <- NULL
` ``
```

4. Análisis de los datos.

4.1. Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar).

Realizaremos un análisis de cada uno de los registros teniendo en cuenta principalmente la especie, los países importadores/exportadores, el objetivo del intercambio, la cantidad de animal traspasado y el propósito. Pretendemos, en resumen, cruzar la especie y los países implicados con otros datos, para así poder extraer conclusiones que nos lleven a entender el comportamiento del traspaso de animales.

4.2. Comprobación de la normalidad y homogeneidad de la varianza.

Nuestro dataset no contiene muchos datos numéricos, por lo que la comprobación de la normalidad y homogeneidad de la varianza no será muy extenso.

Para la comprobación de la normalidad utilizaremos la que viene en los paquetes básicos de R, la de Shapiro-Wilk. Para ello, trabajaremos con un nivel de significancia de 0.5. Con ello, las dos hipótesis que tenemos que es que:

H0: La muestra proviene de una distribución normal.

H1: La muestra no proviene de una distribución normal.

Si $p < \text{Alfa}$ se rechaza H0

Si $p \geq \text{Alfa}$ no se rechaza H0

Hacemos esta prueba con la variable *Exportation.Number*.


```
```{r}
#Prueba de shapiro-wilk
shapiro.test(cites$Exportation.number)
```
```

Shapiro-Wilk normality test

```
data:  cites$Exportation.number
W = 0.12117, p-value < 2.2e-16
```

Vemos que el valor de P es muy pequeño, con lo que podemos concluir que no sigue una distribución normal el parámetro que hemos estudiado. Esto coincide con lo que podíamos ver a simple vista en el boxplot, dado que la cantidad de animales importados/exportados no sigue un patrón.

Para la homogeneidad de la varianza necesitaríamos poblaciones en las que tengamos cierta seguridad de que la distribución es normal, sin embargo en nuestro caso hemos visto que no lo es. Para ello utilizaremos test que se fijen sobre todo en la mediana, como puede ser el Test de Levene. Como H0 tenemos que las varianzas de todos los grupos son iguales y como H1 al menos una varianza distinta entre todos los grupos.

```
```{r}
leveneTest(y = cites$Exportation.number, group = cites$Importer, center = "median")
```
```

```
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  77  0.3588      1
      834
```

Vemos que el $\Pr(>F)$ es mayor que 0.5, con lo que se acepta la hipótesis nula de que la varianza de todos los grupos es la misma. POr otro lado, si agrupamos por importación/exportación, vemos que ese coeficiente baja pero no llega a ser menor que 0.5.

```

```{r}
leveneTest(y = cites$Exportation.number, group = cites$Exportation.type, center = "median")
```

```

cites\$Exportation.type coerced to factor. Levene's Test for Homogeneity of Variance (center = "median")

| | Df | F value | Pr(>F) |
|-------|-----|---------|---------|
| group | 1 | 3.142 | 0.07664 |
| | 910 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.3. Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.

En primer lugar utilizaremos la prueba de chi-cuadrado para deducir la relación entre algunas variables, en concreto veremos la relación entre el país importador/exportador y el tipo de animal y su cantidad.

```

```{r}
table2 <- table(cites$Importer, cites$Exportation.number)
chisq.test(table2)
```

```

Chi-squared approximation may be incorrect
Pearson's Chi-squared test

data: table2
X-squared = 27071, df = 20328, p-value < 2.2e-16

```

```{r}
table2 <- table(cites$Exporter, cites$Exportation.number)
chisq.test(table2)
```

```

Chi-squared approximation may be incorrect
Pearson's Chi-squared test

data: table2
X-squared = 12685, df = 11352, p-value < 2.2e-16

```
```{r}  
table2 <- table(cites$Exporter, cites$Taxon)
chisq.test(table2)
```
```

Chi-squared approximation may be incorrect
Pearson's Chi-squared test

data: table2
X-squared = 27715, df = 13932, p-value < 2.2e-16

```
```{r}  
table1 <- table(cites$Importer, cites$Taxon)
chisq.test(table1)
```
```

Chi-squared approximation may be incorrect
Pearson's Chi-squared test

data: table1
X-squared = 26104, df = 24948, p-value = 1.709e-07

Vemos que en ningún caso obtenemos un nivel de significancia superior a 0.05, con lo que no podemos determinar ningún tipo de relación entre estas variables.

La siguiente técnica que vamos a aplicar consiste en una regresión lineal para ver si el número de sujetos que se intercambia tiene una relación directa con ser una exportación o una importación.

```
``{r}
model <- lm(cites$Exportation.number ~ cites$Exportation.type)
summary(model)
``
```

Call:

```
lm(formula = cites$Exportation.number ~ cites$Exportation.type)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|-------|-------|--------|------|--------|
| -1532 | -1504 | -1382 | -116 | 192468 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------------------------|----------|------------|---------|--------------|
| (Intercept) | 1532.0 | 311.4 | 4.919 | 1.03e-06 *** |
| cites\$Exportation.typeImported | -1414.7 | 786.5 | -1.799 | 0.0724 . |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8636 on 910 degrees of freedom

Multiple R-squared: 0.003543, Adjusted R-squared: 0.002448

F-statistic: 3.236 on 1 and 910 DF, p-value: 0.07238

De aquí podemos sacar varios datos interesantes. En primer lugar, si nos fijamos en los residuos, vemos que ni la mediana está cercana a 0 y que ni el primer y tercer cuartil son simétricos. Esto ocurre porque la variable que estamos estudiando no sigue una distribución normal. Vemos también que R-squared nos dice que el modelo explica el 0,3% de la varianza de la variable dependiente, un registro muy bajo. Podemos concluir por tanto que no hay una relación directa entre el tipo de intercambio y la calidad del mismo.

Por otro lado, si miramos la relación entre el número de animales intercambiados y su fin, sí que conseguimos mejorar algo ese R-squared y llegar a que el modelo explica el 20% de la varianza de la variable dependiente, pero aun así sigue siendo un porcentaje bastante bajo.

Residual standard error: 7892 on 884 degrees of freedom
Multiple R-squared: 0.1915, Adjusted R-squared: 0.1668
F-statistic: 7.755 on 27 and 884 DF, p-value: < 2.2e-16

Por último vamos a usar ANOVA para determinar si una variable aporta significancia al modelo. Tendremos las siguientes hipótesis:

H0: la variable de la Term no aporta información para el modelo

HA: la variable de la Term sí aporta información para el modelo

```
```{r}
model <- lm(cites$Exportation.number ~ cites$Term + cites$Purpose)
anova(model)
```
```

Analysis of Variance Table

Response: cites\$Exportation.number

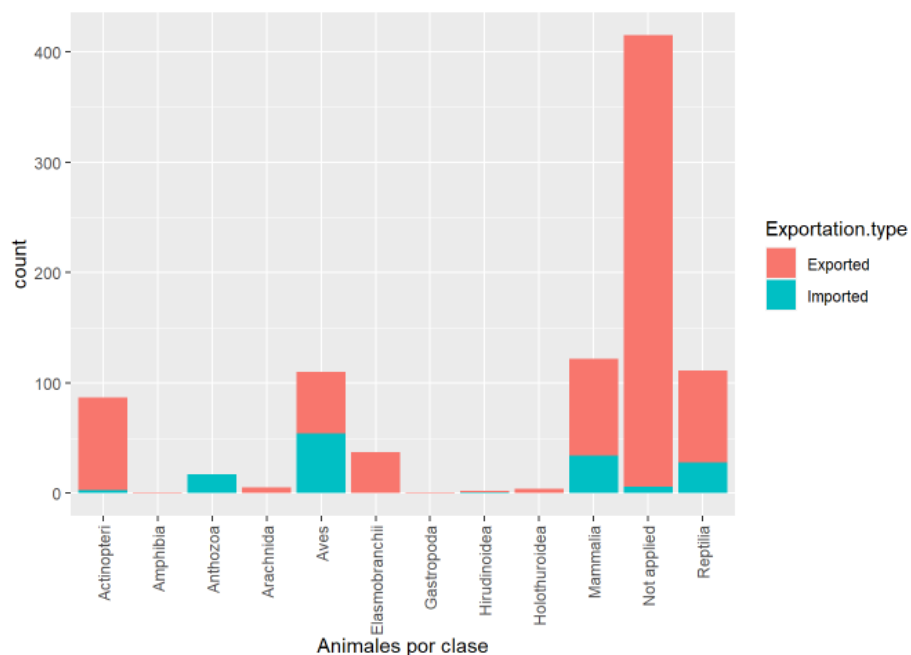
| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|----------------|-----|------------|-----------|---------|------------|
| cites\$Term | 27 | 1.3042e+10 | 483051136 | 7.7115 | <2e-16 *** |
| cites\$Purpose | 10 | 3.1786e+08 | 31786438 | 0.5074 | 0.8856 |
| Residuals | 874 | 5.4748e+10 | 62640291 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Vemos que el valor p asociado a Term es muy bajo, inferior a un nivel usual del 5%, por lo que la Variable Term no aporta información al modelo.

5. Representación de los resultados a partir de tablas y gráficas.

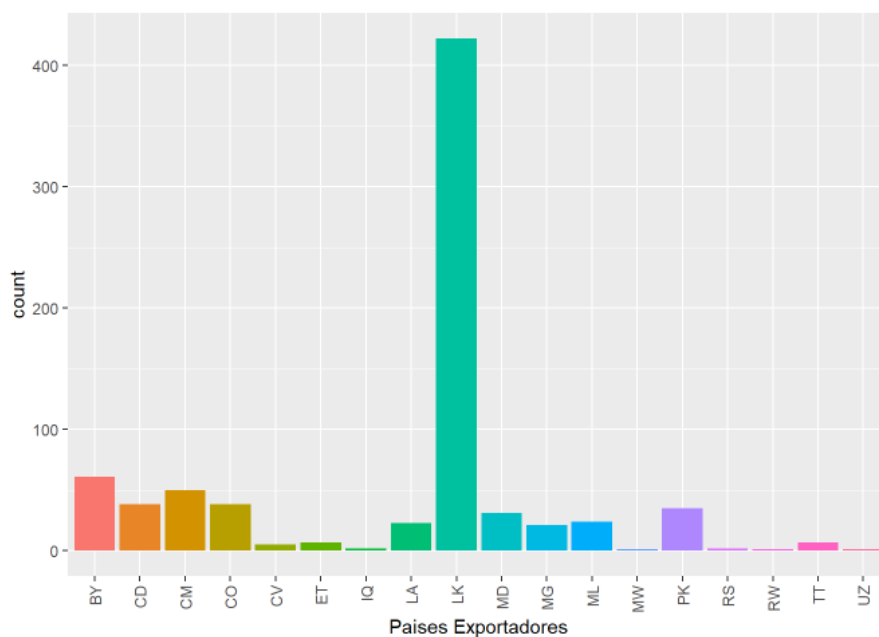
¿Qué tipo de especie son las intercambiadas?



¿Qué país es el que más exportaciones reporta?

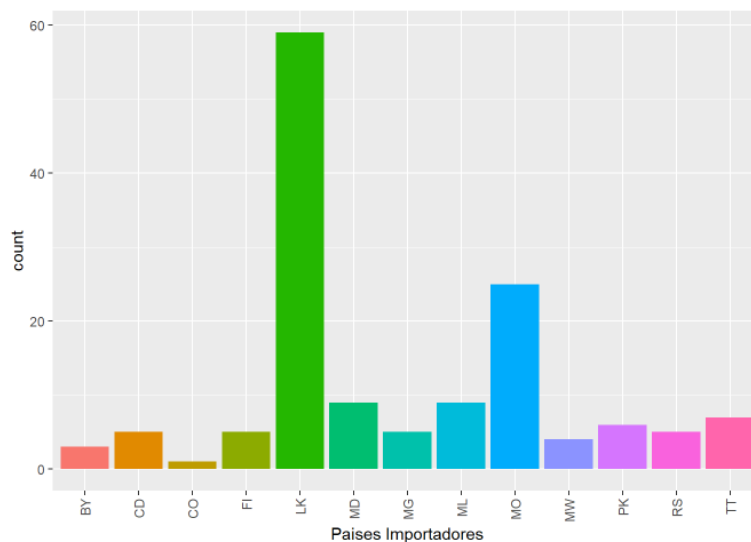
```

#Cantidad de intercambios de exportación por países
ggplot(data=subset(cites, Exportation.type == "Exported"), aes(x=Exporter, fill=Exporter)) + geom_bar() +
  xlab("Países Exportadores") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1), legend.position = "none")
  
```



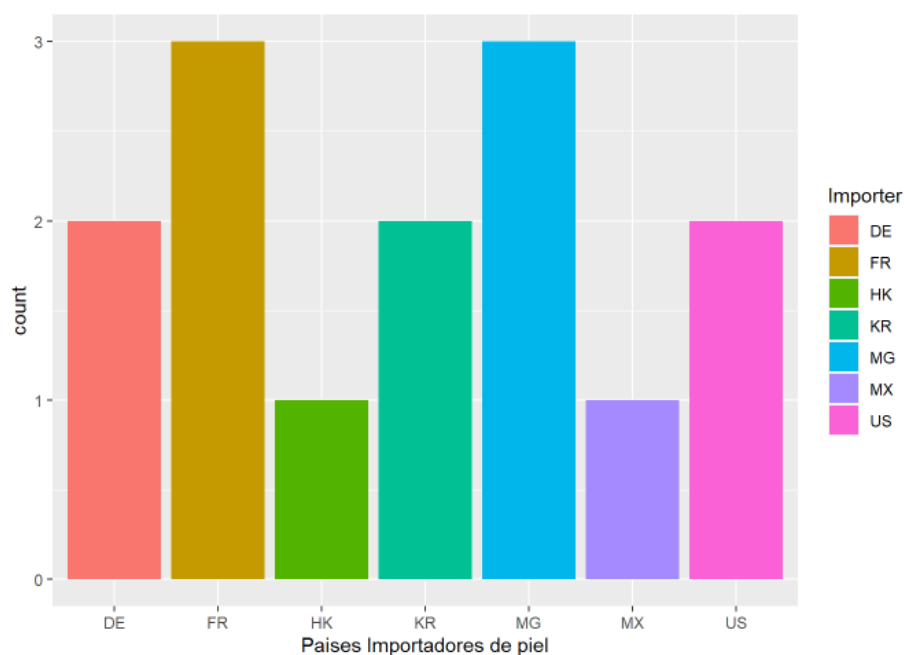
¿Qué país es más activo reportando importaciones?

```
#Cantidad de intercambios de importación por países
ggplot(data=subset(cites, Exportation.type == "Imported"),aes(x=Importer, fill=Importer)) + geom_bar() +
  xlab("Países Importadores") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1), legend.position = "none")
```

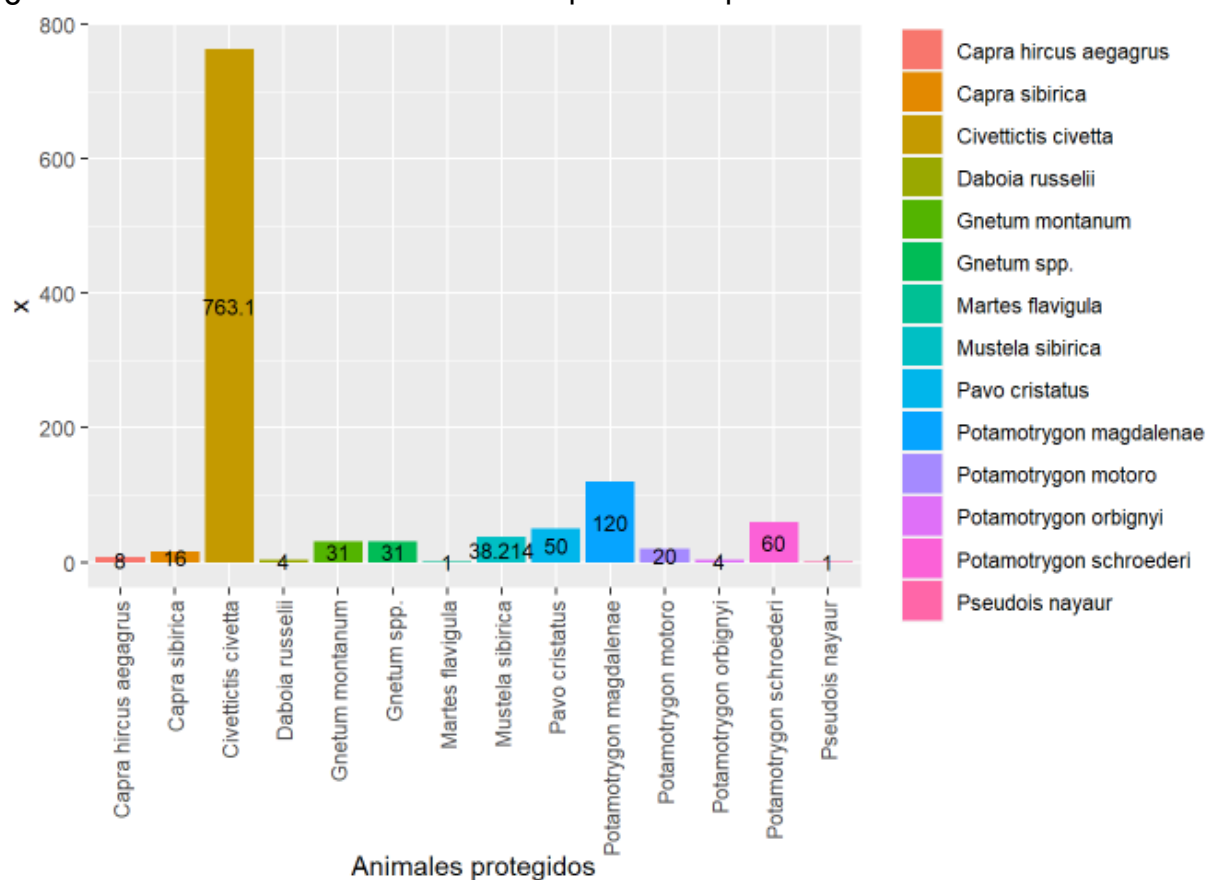


¿El comercio de animales con fines de creación de pieles, tiene alguna relación con algún determinado país/área?

```
#Países/importaciones de piel
ggplot(data=subset(cites, Term == "skins" | Term == "skin pieces"),aes(x=Importer, fill=Importer)) + geom_bar() +
  xlab("Países Importadores de piel")
```



¿Cuáles es el animal más demandado que menos protección tiene?



6. Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?

El apartado de análisis estadístico nos ha servido para comprobar que los registros entre sí son poco coherentes. Esto se debe al tipo de problema y dominio al que nos enfrentamos, en el que el reporte de un intercambio no tiene que guardar relación con el siguiente. No hemos visto una relación clara entre las distintas variables categóricas ni con el número de sujetos intercambiados.

Por otro lado, las conclusiones las hemos podido obtener a partir de un análisis visual con distintos gráficos que hemos ido creando, dando lugar a diversas respuestas que comentaremos a continuación.

¿Qué tipo de especie son las intercambiadas?

Mayoritariamente los intercambios no informan la clase. Sin embargo, los registros que sí se informan; las clases predominantes son Mammalia, Reptilia y Aves.

En cuanto al número de intercambios, en general se registran en el CITES más exportaciones que importaciones, como se aprecia en el gráfico de barras.

¿Qué país es el que más exportaciones/importaciones reporta?

En el gráfico del apartado anterior vemos claramente que Sri Lanka destaca entre el resto, teniendo él solo más que la suma del resto de países. Esto nos lleva a pensar que en este país las autoridades cumplen con el reporte del intercambio de especies.

¿El comercio de animales con fines de creación de pieles, tiene alguna relación con algún determinado país/área?

Vemos que no se comercia mucho con estos fines, o al menos no queda registrado. Sin embargo, claramente se ve que los países a los que se destina tienen cierto nivel económico, como EEUU o Francia.

¿Cuál es el animal más demandado que menos protección tiene?

Vemos que dentro de las especies de tipo III, aquellas que están menos, hay una que destaca sobre todas ellas, a la cual las autoridades deberían de prestar especial atención. Concretamente se trata del *Civettictis Civetta*.

Como conclusión final podemos dictaminar que el registro de animales no se hace del mismo en todos los lugares del mundo. Muchos registros están incompletos, seguramente por fallo humano. Las autoridades pertinentes deberían controlar más este tipo de intercambio por la materia con la que se comercia.

7. Código: Hay que adjuntar el código, preferiblemente en R, con el que se ha realizado la limpieza, análisis y representación de los datos. Si lo preferís, también podéis trabajar en Python

URL al repositorio: <https://github.com/miguel-a-ngel/DataCleaning>

```
---
title: "Data Cleaning and Analysis"
output:
  html_document:
    highlight: default
    number_sections: yes
    theme: cosmo
    toc: yes
    toc_depth: 2
    includes:
      in_header: header.html
  word_document: default
  pdf_document:
    highlight: zenburn
    toc: yes
---
```

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
print(getwd())
library(car)
library(dplyr)
library(carData)
library(ggplot2)
```
```

Asignatura:

Tipología y Ciclo de vida de los datos

Autores:

José Ramón Martínez-Carbonell Martín

Miguel Ángel Pérez García

```
# Exploración
```

```
```{r}
#Captura
cites <- read.csv("cites_2021_original.csv", header = TRUE)
head(cites)
summary(cites)
```

```{r}
#Detección de nulos/vacíos
#sapply(cites, function(x) sum(is.na(x))) # NA
sapply(cites, function(x) sum(x=="")) #Si es vacío
#sapply(cites, function(x) sum(x==" ")) # Si no es vacío, aunque
tiene un espacio
```
```

```
# Limpieza
```

```
```{r}
#Declaramos las nuevas columnas que utilizaremos
auxNumber <- c()
auxType <- c()
#Según el si el campo de importación/exportación está vacío o no,
#vamos creando el vector según convenga
for (i in 1:nrow(cites)) {
 if(!is.na(cites$Importer.reported.quantity[i])){
 auxNumber[i] <- cites$Importer.reported.quantity[i]
 auxType[i] <- "Imported"
 }else if(!is.na(cites$Exporter.reported.quantity[i])){
 auxNumber[i] <- cites$Exporter.reported.quantity[i]
 auxType[i] <- "Exported"
 }
}
#Asignamos las nuevas columnas al dataset
cites$Exportation.type <- auxType
cites$Exportation.number <- auxNumber
```
```

```
```{r}
#Creación de boxplot para detección de outliers
boxplot(cites$Exportation.number)
boxplot.stats(cites$Exportation.number)$out
```
```

```
```{r}
#Duplicados
cites[duplicated(cites),]
dim(cites)
```
```

```
```{r}
#Eliminar duplicados
cites = unique(cites)
dim(cites)
```
```

```
```{r}
#Eliminamos las columnas que ya no nos interesan
cites$Origin <- NULL
cites$Importer.reported.quantity <- NULL
cites$Exporter.reported.quantity <- NULL
cites$Family <- NULL
cites$Order <- NULL
cites$Genus <- NULL
```
```

```
```{r}
#Añadir nivel Not applied a Unit, Class, Source and Genus
levels(cites$Unit) <- c(levels(cites$Unit), "Not applied")
levels(cites$Class) <- c(levels(cites$Class), "Not applied")
levels(cites$Source) <- c(levels(cites$Source), "Not applied")
levels(cites$Purpose) <- c(levels(cites$Purpose), "Not applied")
```
```

```
```{r}
#sustitución de cadena vacía por "Not applied"
for (i in 1:nrow(cites)) {
 for(j in 1:ncol(cites)){
 if((cites[i,j]=="")){
 cites[i,j] <- "Not applied"
 }
 }
}
```
```

```
` `{r}
#Guardamos el dataset transformado
write.csv(cites, "cites_2021_limpio.csv")
` `{r}

# Análisis

` `{r}
#Prueba de shapiro-wilk
shapiro.test(cites$Exportation.number)
` `{r}

` `{r}
install.packages("car")
` `{r}

` `{r}
leveneTest(y = cites$Exportation.number, group = cites$Importer,
center = "median")
` `{r}

` `{r}
leveneTest(y = cites$Exportation.number, group =
cites$Exportation.type, center = "median")
` `{r}

` `{r}
table1 <- table(cites$Importer, cites$Taxon)
chisq.test(table1)
` `{r}

` `{r}
table2 <- table(cites$Exporter, cites$Taxon)
chisq.test(table2)
` `{r}

` `{r}
table2 <- table(cites$Exporter, cites$Exportation.number)
chisq.test(table2)
` `{r}

` `{r}
table2 <- table(cites$Importer, cites$Exportation.number)
chisq.test(table2)
` `{r}

` `{r}
model <- lm(cites$Exportation.number ~ cites$Exportation.type)
```

```
summary(model)
```

```{r}
model <- lm(cites$Exportation.number ~ cites$Term)
summary(model)
```

```{r}
model <- lm(cites$Exportation.number ~ cites$Term + cites$Purpose)
anova(model)
```

Visualización

```{r}
#Clases intercambio
ggplot(data=cites,aes(x=Class, fill=Exportation.type)) +
geom_bar() + xlab("Animales por clase") + labs(fill =
"Exportation.type") + theme(axis.text.x = element_text(angle = 90,
vjust = 0.5, hjust=1))

#Cantidad de intercambios de exportación por paises
ggplot(data=subset(cites, Exportation.type ==
"Exported"),aes(x=Exporter, fill=Exporter)) + geom_bar() +
xlab("Países Exportadores") +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5,
hjust=1), legend.position = "none")

#Cantidad de intercambios de importación por paises
ggplot(data=subset(cites, Exportation.type ==
"Imported"),aes(x=Importer, fill=Importer)) + geom_bar() +
xlab("Países Importadores") +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5,
hjust=1), legend.position = "none")

#Países/importaciones_de_piel
ggplot(data=subset(cites, Term == "skins" | Term == "skin
pieces"),aes(x=Importer, fill=Importer)) + geom_bar() +
xlab("Países Importadores de piel")

tempA <- subset(cites, App. == "III")
tempB <- select(tempA,Taxon,Exportation.number)
```

```
animals_count = aggregate(x = tempB$Exportation.number,

                           by = list(Taxon = tempB$Taxon),
                           FUN = sum)

#Animales protegidos
ggplot(data=animals_count, aes(x=Taxon, y=x, fill=Taxon, label=x))
+ geom_bar(stat='identity') + xlab("Animales protegidos") +
labs(fill = "Taxon") + theme(axis.text.x = element_text(angle =
90, vjust = 0.5, hjust=1)) + geom_text(aes(label = x), position =
position_stack(0.5), size=3)

```
```

## Contribuciones

| Contribuciones           | Firma               |
|--------------------------|---------------------|
| Análisis e investigación | José R. y Miguel A. |
| Redacción                | José R. y Miguel A. |
| Desarrollo               | José R. y Miguel A. |
| Pruebas                  | José R. y Miguel A. |
| Referencias              | José R. y Miguel A. |

## Referencias

1. Kaggle. (2018). *CITES Wildlife Trade Database*. Recuperado de <https://www.kaggle.com/cites/cites-wildlife-trade-database>
2. StackOverflow. (2020). *StackOverflow*. Recuperado de <https://es.stackoverflow.com/>
3. Wikipedia (2021). *Wikipedia Viverrinae*. Recuperado de: <https://es.wikipedia.org/wiki/Viverrinae>