Data Mobilisation from GBIF to the EBV Data Portal for IAS of Union Concern

Notebook 02 - Data exploration of the IAS occurrence cube

true

2024-08-20

Introduction

In this notebook we explore the occurrence data of invasive alien species (IAS) of union concern available in GBIF—the Global Biodiversity Information Facility— until mid August 2024. To do this, an IAS occurrence cube was previously created using the occurrence cube software developed by GBIF under the Biodiversity Building Blocks for Policy (B3) project. Details of the data query in GBIF are available at DOI 10.15468/dl.gxk3vh. The cube generation script is also part of this repository.

Note: This series of notebooks is part of the results of Task 3.3 of the Biodiversity Building Blocks for Policy project funded by the European Union's Horizon Europe Research and Innovation Programme (ID No 101059592). Additional notebooks exploring the results and calculating simple metrics are also available in the same repository.

Load library and input data

```r
rm(list=ls())
gc()
```

```
##           used  (Mb) gc trigger  (Mb) max used  (Mb)
## Ncells 2132677 113.9    5842758 312.1 11612665 620.2
## Vcells 4349395  33.2   43064972 328.6 70740088 539.8
```

```r
# load requiered libraries
library(here)
library(ggplot2)
library(dplyr)
library(lubridate)
library(b3gbi)
library(stringr)
```

After loading the necessary libraries, we load the occurrence cube of IAS obtained previously through the GBIF API.

```
# Load IAS occurrence cube obtained through GBIF

# File name from the JSON query
occcube <- "0077925-240506114902167"

# Load occurrence cube using b3gbi
cin <- process_cube(here(paste0("output/datacubes/csv/ias/", occcube,".csv")))

# Load species taxonomy resulting from the match between the updates IAS list of union concern and the
tax <- read.csv(here("input/data/ias/taxonomy/List87IAS_EU_match_gbif_synonyms.csv"))
```

As some of the scientific names of the IAS of union are considered synonyms by the GBIF backbone taxonomy, we will use the `acceptedUsageKey` for the synonyms and the `key` for the accepted scientific names. To do this, we will fill in the `acceptedUsageKey` column with the `key` for accepted names, and keep the `acceptedUsageKey` for synonyms. Thus, the accepted keys for all species in the list will appear in the `acceptedUsageKey` column.

```
# Merge in one column `key` of accepted names and `acceptedUsageKey` of synonyms
tax <- tax %>%
  mutate(acceptedUsageKey = coalesce(acceptedUsageKey, key))

# Write CSV with `acceptedUsageKey` for all species to be used in the JSON query
write.csv(tax, here("input/data/ias/taxonomy/List87IAS_EU_match_gbif_synonyms_acceptedUsageKeys.csv"))
```

Data Analysis

```
cdata <- cin[["data"]]
# rename columns
colnames(cdata)[colnames(cdata) == "order"] <- "order_"
colnames(cdata)[colnames(cdata) == "taxonKey"] <- "acceptedUsageKey"

# Aggregate occurrences at species level
cag <- cdata %>%
  group_by(acceptedUsageKey) %>%
  summarize(totalOcc = sum(obs))

 # Sort in ascending order
cag <- cag[order(cag$totalOcc), ]
```

```
# Rename columns for joining occurence cube with GBIF Backbone taxonomy
xout <- merge(x=cag, y=tax[,c("scientificName", "acceptedUsageKey", "kingdom", "phylum", "class", "orde
xout <- xout[order(xout$totalOcc), ]

write.csv(xout, here("output/summary_data/csv/ias/summary_ias_totalOccurrences.csv"), quote = FALSE)
```

Calculate Total Number of Occurrence

```r
# # Find what species have no records in GBIF
noocc <- anti_join(tax, cag, by = "acceptedUsageKey")
write.csv(noocc, here("output/summary_data/csv/ias/ias_noOccurrences.csv"), quote = FALSE)
print(noocc)
```

Identify IAS without Records in GBIF

```
##    occurrenceId
## 1            NA
## 2            NA
## 3            NA
## 4            NA
## 5            NA
## 6            NA
## 7            NA
## 8            NA
## 9            NA
## 10           NA
##                                                                             verbatimScientificName
## 1                                                                      Channa argus (Cantor, 1842)
## 2                                  Cortaderia selloana subsp. jubata (Lemoine) Testoni & Villamil
## 3                                                              Limnoperna fortunei (Dunker, 1857)
## 4                                                                 Morone americana (Gmelin, 1789)
## 5                                                              Plotosus lineatus (Thunberg, 1787)
## 6  Pueraria montana (Lour.) Merr. var. lobata (Willd.) Maesen & S.M.Almeida ex Sanjappa & Predeep
## 7                                                           Solenopsis geminata (Fabricius, 1804)
## 8                                                                 Solenopsis richteri Forel, 1909
## 9                                                               Urva auropunctata (Hodgson, 1836)
## 10                                                    Vespa velutina nigrithorax du Buysson, 1905
##                                                                       scientificName
## 1                                                        Channa argus (Cantor, 1842)
## 2                                                  Cortaderia jubata (Lemoine) Stapf
## 3                                                 Limnoperna fortunei (Dunker, 1857)
## 4                                                    Morone americana (Gmelin, 1789)
## 5                                                 Plotosus lineatus (Thunberg, 1787)
## 6  Pueraria montana var. lobata (Willd.) Maesen & S.M.Almeida ex Sanjappa & Predeep
## 7                                              Solenopsis geminata (Fabricius, 1804)
## 8                                                    Solenopsis richteri Forel, 1909
## 9                          Herpestes javanicus subsp. auropunctatus (Hodgson, 1836)
## 10                                        Vespa velutina nigrithorax Buysson, 1905
##         key matchType confidence   status       rank  kingdom       phylum
## 1   4284921     EXACT         99 ACCEPTED    SPECIES Animalia     Chordata
## 2   9355348     EXACT        100  SYNONYM SUBSPECIES  Plantae Tracheophyta
## 3   5855350     EXACT         99 ACCEPTED    SPECIES Animalia     Mollusca
## 4   2394604     EXACT         99 ACCEPTED    SPECIES Animalia     Chordata
## 5   7965247     EXACT         99 ACCEPTED    SPECIES Animalia     Chordata
## 6   2977647     EXACT        100 ACCEPTED    VARIETY  Plantae Tracheophyta
## 7   5035187     EXACT         99 ACCEPTED    SPECIES Animalia   Arthropoda
## 8   5035017     EXACT        100 ACCEPTED    SPECIES Animalia   Arthropoda
## 9  10504616     EXACT         99  SYNONYM    SPECIES Animalia     Chordata
## 10  6247411     EXACT        100 ACCEPTED SUBSPECIES Animalia   Arthropoda
##            class        order      family      genus             species
## 1                 Perciformes   Channidae     Channa        Channa argus
## 2     Liliopsida       Poales     Poaceae Cortaderia   Cortaderia jubata
## 3       Bivalvia     Mytilida   Mytilidae Limnoperna Limnoperna fortunei
## 4                 Perciformes   Moronidae     Morone    Morone americana
## 5                Siluriformes  Plotosidae   Plotosus   Plotosus lineatus
## 6  Magnoliopsida      Fabales    Fabaceae   Pueraria    Pueraria montana
## 7        Insecta  Hymenoptera  Formicidae Solenopsis Solenopsis geminata
## 8        Insecta  Hymenoptera  Formicidae Solenopsis Solenopsis richteri
## 9       Mammalia    Carnivora Herpestidae  Herpestes Herpestes javanicus
## 10       Insecta  Hymenoptera    Vespidae      Vespa      Vespa velutina
##                 canonicalName
## 1                Channa argus
## 2  Cortaderia selloana jubata
## 3         Limnoperna fortunei
## 4            Morone americana
## 5           Plotosus lineatus
## 6     Pueraria montana lobata
## 7         Solenopsis geminata
## 8         Solenopsis richteri
## 9           Urva auropunctata
## 10 Vespa velutina nigrithorax
##                                              authorship usageKey
## 1                                       (Cantor, 1842)   4284921
## 2                         (Lemoine) Testoni & Villamil   9355348
## 3                                       (Dunker, 1857)   5855350
## 4                                       (Gmelin, 1789)   2394604
## 5                                     (Thunberg, 1787)   7965247
## 6  (Willd.) Maesen & S.M.Almeida ex Sanjappa & Predeep   2977647
## 7                                    (Fabricius, 1804)   5035187
## 8                                          Forel, 1909   5035017
## 9                                      (Hodgson, 1836)  10504616
## 10                                       Buysson, 1905   6247411
##    acceptedUsageKey
## 1           4284921
## 2           2704521
## 3           5855350
## 4           2394604
## 5           7965247
## 6           2977647
## 7           5035187
## 8           5035017
## 9           6164088
## 10          6247411
```

Data Exploration

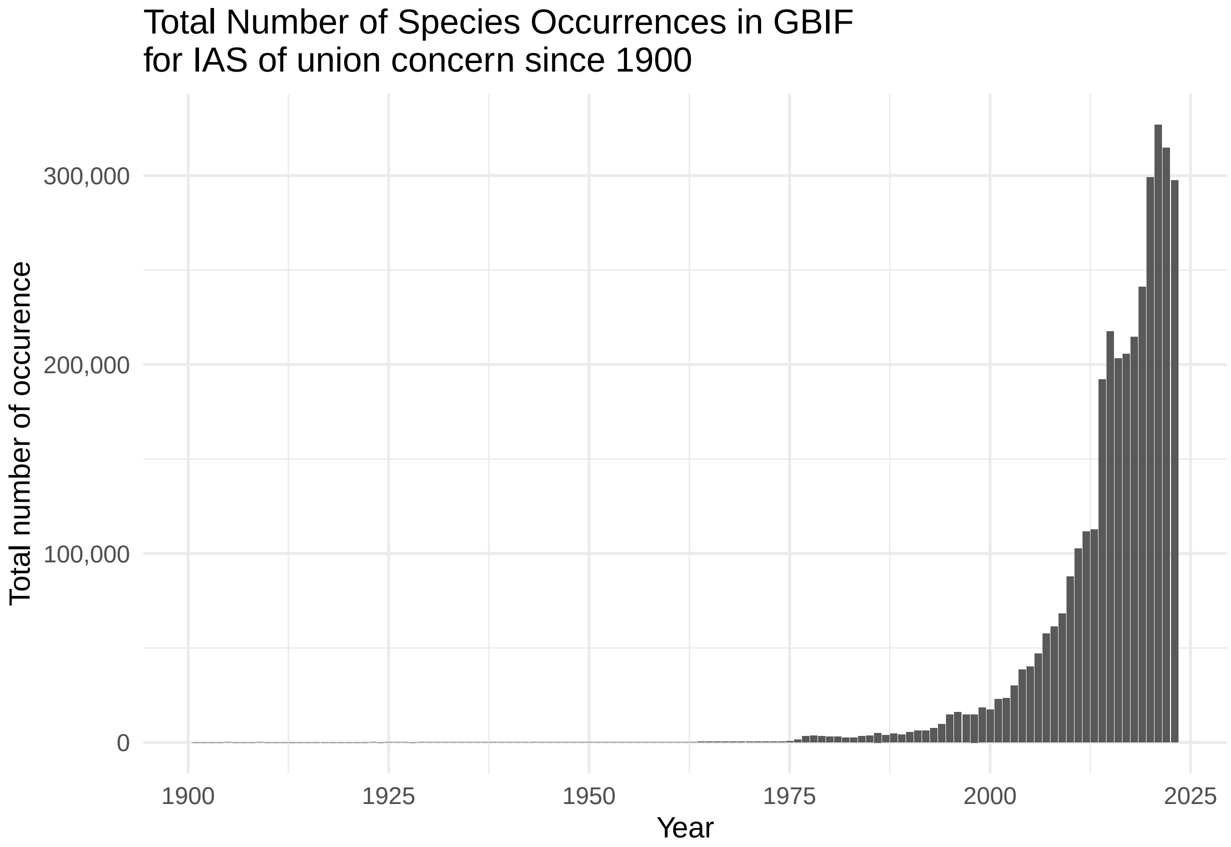
We will explore data available since 1900. To do this, first we split the 'yearMonth' column into 'year' and 'month'.

```
# Convert date from character to numeric
todates <- as.data.frame(str_split(cdata$yearMonth, "-", simplify = TRUE))
colnames(todates) <- c("year", "month")
```

```
# Add year and month columns separate to the initial data
cdata$year <- todates$year
cdata$month <- todates$month

# Filter data starting from 1900
cdata2 <- cdata %>%
  filter(year > 1900)

# Group data by year
cag_year <- cdata2 %>%
    group_by(year) %>%
  summarize(totalOcc = sum(obs))

# Group data by month
cag_month <- cdata %>%
    group_by(month) %>%
  summarize(totalOcc = sum(obs))
```

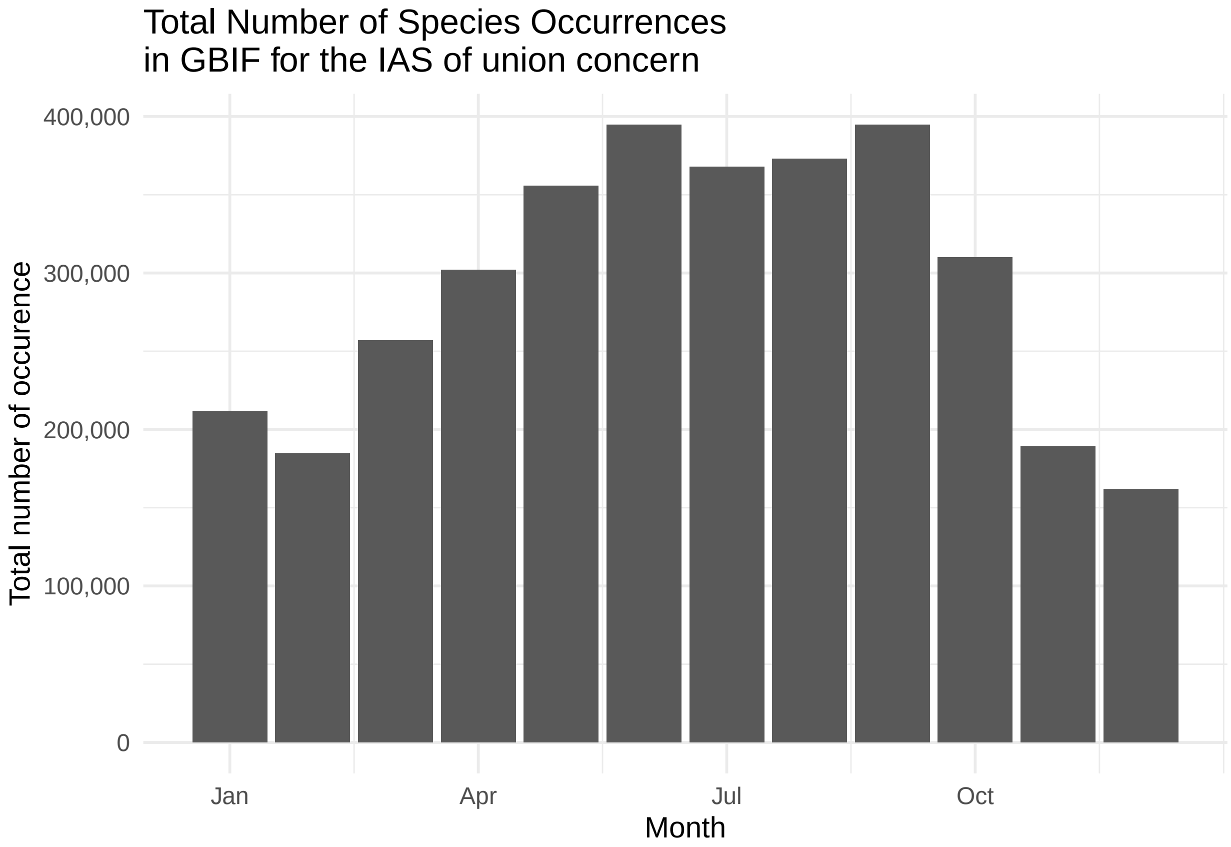
In the following code chunk we plot the GBIF occurrences available since 1900 for the IAS of union concern.

```
theme_update(plot.title = element_text(hjust = 0.5))
ggplot(cag_year, aes(x = as.numeric(year), y = totalOcc)) +
  geom_bar(stat = "identity") +
  scale_y_continuous(label = scales::label_comma(accuracy = 1)) +
  ggtitle("Total Number of Species Occurrences in GBIF \nfor IAS of union concern since 1900") +
  labs(x = "Year",
       y = "Total number of occurence") +
       guides(color = guide_legend(override.aes = list(size = 5))) +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_minimal()
```



Now we plot the all GBIF occurrences per month for the IAS of union concern.

```r
ggplot(cag_month, aes(x = as.numeric(month), y = totalOcc)) +
  geom_bar(stat = "identity") +
  scale_x_continuous(breaks = c(1,4,7,10), labels = c("Jan", "Apr", "Jul", "Oct")) +
  scale_y_continuous(label = scales::label_comma(accuracy = 1)) +
  ggtitle("Total Number of Species Occurrences \nin GBIF for the IAS of union concern") +
  labs(x = "Month",
       y = "Total number of occurence") +
       guides(color = guide_legend(override.aes = list(size = 5))) +
  theme_minimal()
```

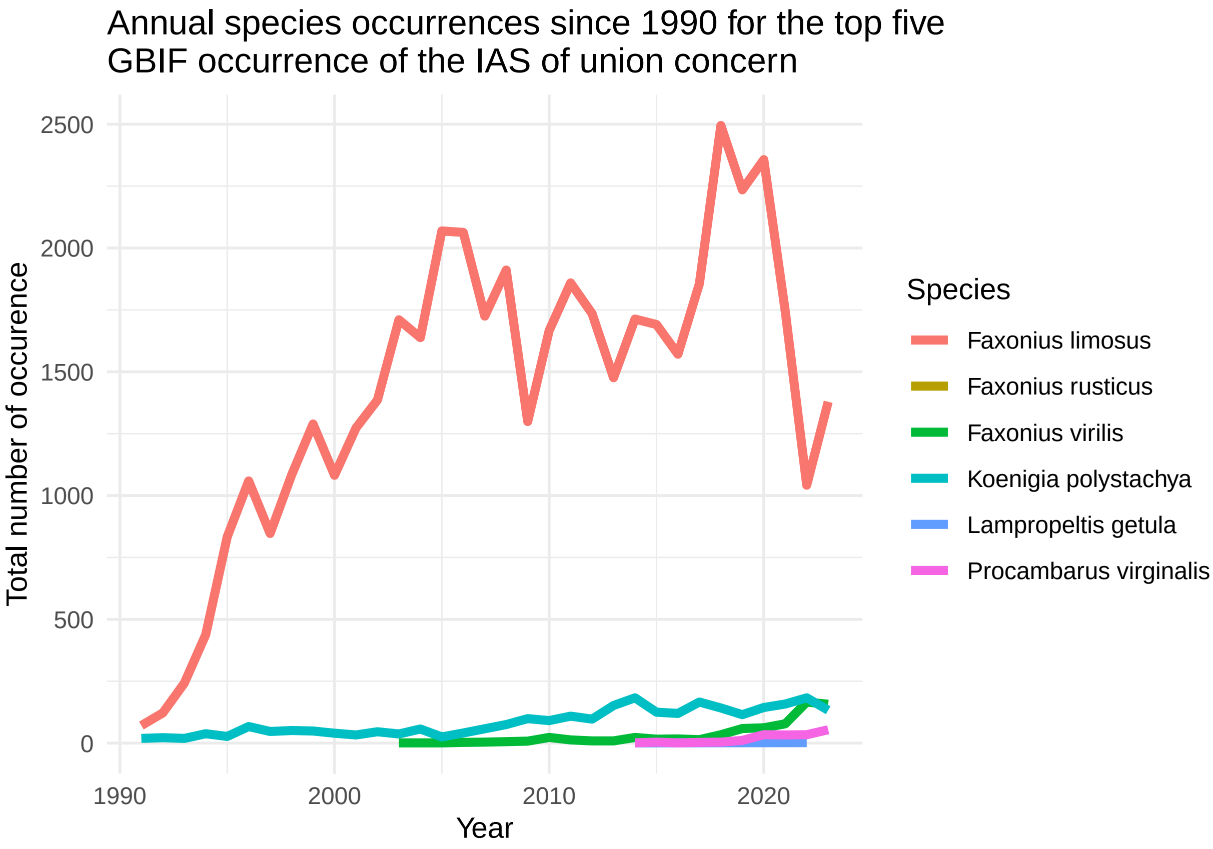


Our last plot shows the increasing number of records since 2000 for the five species with the highest records.

```
# Aggregate occurrences at species level
cag_sp <- cdata %>%
  group_by(acceptedUsageKey) %>%
  summarize(totalOcc = sum(obs))

# Select the five species with more records
cag_sel <- cag_sp[(dim(cag_sp)[1]-5):dim(cag_sp)[1],]

# Subset data set
cag_top5 <- cdata %>%
  inner_join(cag_sel[,c("acceptedUsageKey")], by = "acceptedUsageKey") %>%
  filter(year > 1990)

# Aggregate occurrences by year
cag_top5year <- cag_top5 %>%
  group_by(year, scientificName) %>%
  summarize(totalOcc = sum(obs), scientificName = first(scientificName))
```

Plot the five species with more records in GBIF since 1990.

```
ggplot(cag_top5year, aes(x = as.numeric(year), y = totalOcc, color = scientificName, name = "Species"))
  geom_line(linewidth = 1.6) +
  guides(color=guide_legend(ncol=1, title = "Species")) +
  ggtitle("Annual species occurrences since 1990 for the top five \nGBIF occurrence of the IAS of union
  labs(x = "Year",
       y = "Total number of occurence") +
  theme_minimal()
```


Annual species occurrences since 1990 for the top five
GBIF occurrence of the IAS of union concern
Total number of occurence
Year
Species
Faxonius limosus
Faxonius rusticus
Faxonius virilis
Koenigia polystachya
Lampropeltis getula
Procambarus virginalis