Can scale-dependant geodiversity improve species distribution models in a Neotropical biodiversity hotspot?

Beth Gerstner

2023-01-02

Overview: This script obtains the occurrence records for species of interest in Colombia. It extracts records over the BioModelos API. Records are then counted. Certain species could not be downloaded over the API and had to be obtained manually through the BioModelos website. These are then cleaned and merged into the full dataset.

\*note currently this API does not seem to be working.

Data output: csv file of occurrence records for all species.

Date: 10/1/22 Modified: 1/1/23

Load packages

library(spocc)  
library(spThin)  
library(dismo)  
library(rgeos)  
library(ENMeval)  
library(wallace)  
library(dplyr)   
library(plyr)   
library(readr)   
library(data.table)  
library(janitor)

Read in species list of interest. These are species that have validated models on BioModelos. We use these occurrence records because they have been vetted by experts at BioModelos. Also, we would like to make comparisons with BioModelos models as a way of validating our method of making distribution maps.

scientific\_names\_mammals <- read.csv("C:/Users/bgers/Desktop/final\_species\_list.csv")  
bioKey <- "5NbMdjylEPEN1cBCIsX9dl:3vbVvDAXQdjf40NeYHfN1g"

Download occurrence records from BioModelos

all.mammals.biomod <- data.frame()  
for(i in 1:nrow(scientific\_names\_mammals)){  
 species.i <- scientific\_names\_mammals[i,"species"]  
 bioModelos <- occs\_biomodelos(  
 spN = species.i,  
 bioKey = bioKey)  
 occs\_i <- bioModelos$cleaned  
 occs\_i <- occs\_i[,c("scientific\_name","latitude","longitude","year","state\_province","record\_type", "catalog\_number", "institution\_code")]  
 all.mammals.biomod= rbind(all.mammals.biomod, occs\_i)  
}

Split species list because one species Cheracebus lucifer was causing the loop to throw errors.

all.mammals.1 <- data.frame()  
for(i in 27:32){  
 species.i <- scientfic\_names\_mammals[i,"species"]  
 bioModelos <- occs\_biomodelos(  
 spN = species.i,  
 bioKey = bioKey)  
 occs\_i <- bioModelos$cleaned  
 occs\_i <- occs\_i[,c("scientific\_name","latitude","longitude","year","state\_province","record\_type", "catalog\_number", "institution\_code")]  
 all.mammals.1= rbind(all.mammals, occs\_i)  
}

Join list for species downloaded over API

all.mammals.full <- rbind(all.mammals, all.mammals.1)  
setwd("C:/Users/bgers/Desktop/MSU/Zarnetske\_Lab/Data/Chapter\_1/candidate\_species\_dataset")  
  
#write to csv  
write.csv(all.mammals.full,path="all\_mammals\_biomodelos.csv")

Count number of records per species

species\_num <- all.mammals %>%  
 count(scientific\_name, sort = TRUE)

Downloaded certain species individually on the BioModelos website because of errors and pulled them in all together

single\_data <- list.files(path = "C:/Users/bgers/Desktop/MSU/Zarnetske\_Lab/Data/Chapter\_1/candidate\_species\_dataset/biomodelos\_singles",   
 pattern = "\*.csv", full.names = TRUE) %>%   
 lapply(read\_csv) %>%   
 bind\_rows

## Rows: 19 Columns: 28  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (16): suggestedStateProvince, acceptedNameUsage, basisOfRecord, catalogN...  
## dbl (9): taxID, day, environmentalOutlier, month, year, reported, longitude...  
## lgl (3): url, collectionCode, updated  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 49 Columns: 28  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (16): suggestedStateProvince, acceptedNameUsage, basisOfRecord, catalogN...  
## dbl (9): taxID, day, environmentalOutlier, month, year, reported, longitude...  
## lgl (3): url, collectionCode, updated  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 38 Columns: 28  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (16): suggestedStateProvince, acceptedNameUsage, basisOfRecord, catalogN...  
## dbl (8): taxID, day, environmentalOutlier, month, year, longitude, latitude...  
## lgl (4): url, collectionCode, reported, updated  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 27 Columns: 28  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (16): suggestedStateProvince, acceptedNameUsage, basisOfRecord, catalogN...  
## dbl (8): taxID, day, environmentalOutlier, month, year, longitude, latitude...  
## lgl (4): url, collectionCode, reported, updated  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 26 Columns: 28  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (17): suggestedStateProvince, acceptedNameUsage, suggestedCounty, collec...  
## dbl (8): taxID, day, environmentalOutlier, month, year, longitude, latitude...  
## lgl (3): url, reported, updated  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#Convert to dataframe  
single\_data <- as.data.frame(single\_data)  
  
#Subset columns  
single\_data\_clean <- single\_data[,c("species","latitude","longitude","year", "stateProvince","basisOfRecord","catalogNumber",  
 "institutionCode")]

Read in API dataset if haven’t run above *all.mammals <- read.csv(“C:/Users/bgers/Desktop/MSU/Zarnetske\_Lab/Data/Chapter\_1/candidate\_species\_dataset/all\_mammals\_biomodelos\_2022.csv”)*

Rename multiple columns for old to new and bind all data

setnames(single\_data\_clean, old = c('species','stateProvince','basisOfRecord','catalogNumber','institutionCode'),   
 new = c('scientific\_name','state\_province','record\_type','catalog\_number','institution\_code'))  
names(single\_data\_clean)  
  
#bind all biomodelos datasets together  
full\_biomodelos\_data <- rbind(all.mammals,single\_data\_clean)  
write.csv(full\_biomodelos\_data, "full\_biomodelos\_data.csv")

Count records for each species

species\_num\_final <- full\_biomodelos\_data %>%  
 count(scientific\_name)   
  
tabyl(full\_biomodelos\_data, scientific\_name)