R script- **Alien ornamental plants on sale in Botswana, Namibia and Zimbabwe: potential sources for plant invasions**

Diana Rodríguez-Cala, John R.U. Wilson, Katharina Dehnen-Schmutz, Jana Fried, Seoleseng O. Tshwenyane, Israel A. Legwaila

```
####to load history every time I re-start

loadhistory(file="C:/Users/rodriguezd/OneDrive - Coventry University/PhD_Coventry/Project_Southern Africa/databases/.Rhistory")

load("C:/Users/rodriguezd/OneDrive - Coventry University/PhD_Coventry/Project_Southern Africa/databases/.RData")

save.image("C:/Users/rodriguezd/OneDrive - Coventry University/PhD_Coventry/Project_Southern Africa/databases/.RData")
savehistory("C:/Users/rodriguezd/OneDrive - Coventry University/PhD_Coventry/Project_Southern Africa/databases/.Rhistory")
#################################################################################################################
setwd("C:/Users/rodriguezd/OneDrive - Coventry University/PhD_Coventry/Project_Southern Africa/databases")


#############################################################
###installing "lcvplants" package to correct names and match species' lists

setwd("C:/Users/rodriguezd/OneDrive - Coventry University/PhD_Coventry/Project_Southern Africa/databases")

install.packages("devtools")
install.packages("vegan")
install.packages("sf")
install.packages("expowo")
install.packages("rWCVP")

devtools::install_github("idiv-biodiversity/LCVP")
devtools::install_github("idiv-biodiversity/lcvplants")
library("lcvplants")
library(dplyr)
library(tidyr)
library("stringi")
library("ggplot2")
library("vegan")
library("readxl")
library("sf")
library("expowo")
library("rWCVP")
##### curating my plants' list dataset
```

```
nursery_plants<-read.csv("C:/Users/rodriguezd/OneDrive - Coventry
University/PhD_Coventry/Project_Southern Africa/databases/Plants in
nurseries_ZIm&Botswana/Confirmed nurseries.csv", header = T)
str(nursery_plants)
#'data.frame':    2333 obs. of  10 variables:
#$ GENUS             : chr  "Abelia" "Abutilon" "Abutilon" "Abutilon" ...
#$ SPECIFIC.NAME     : chr  "grandiflora" "megapotamicum " "striatum" "x hybridum" ...
#$ VARIETY.SUBSPECIES: chr  "" "variegata" "thompsonii" "" ...
#$ COMMON.NAME       : chr  "" "" "" "firebell" ...
#$ FORM              : chr  "pale pink, variegata" "" "" "" ...
#$ Type              : chr  "shrubs for sun" "shrubs for medium sun" "shrubs for medium sun" "shrubs for
medium sun" ...
#$ Nursery           : chr  "Golden Stairs" "Golden Stairs" "Golden Stairs" "Golden Stairs" ...
#$ Sources           : chr  "https://www.goldenstairsnursery.co.zw/" "http://www.goldenstairszim.com"
"" "" ...
#$ Year              : chr  "2021" "2021" "2021" "2021" ...
#$ Country           : chr  "Zimbabwe" "Zimbabwe" "Zimbabwe" "Zimbabwe" ...

#after applying the lcvp search to the vector
#Error: splist should include only valid characters, please check the name(s) at position(s):
#834, 863, 949, 1091, 1176, 1413, 1416, 1436, 1442, 1444, 1504

c(nursery_plants_vector[834], nursery_plants_vector[863], nursery_plants_vector[949],
  nursery_plants_vector[1091], nursery_plants_vector[1176],nursery_plants_vector[1413],
  nursery_plants_vector[1416], nursery_plants_vector[1436], nursery_plants_vector[1442],
  nursery_plants_vector[1444], nursery_plants_vector[1504])
#[1] "Caryopteris \xd7 clandonensis subsp. " "Coprosma \xd7 kirkii  subsp. "
#[3] "Spiraea  \xd7 arguta subsp. "          "Leucanthemum  \xd7 superbum  subsp. "
#[5] "Begonia \xd7 ricinifolia subsp. "      "Citrus \xd7 limon subsp. "
#[7] "Citrus \xd7 aurantiifolia  subsp. "    "Viola \xd7 wittrockiana subsp. "
#[9] "Penstemon \xd7 spectabilis subsp. "    "Petunia \xd7 hybrida subsp. "
#[11] "Bougainvillea \xd7 buttiana subsp. "

###correcting these mistakes from the beginning
nursery_plants[834,2]<-"clandonensis"
nursery_plants[863,2]<-"kirkii"
nursery_plants[949,2]<-"arguta"
nursery_plants[1091,2]<-"superbum"
nursery_plants[1176,2]<-"ricinifolia"
nursery_plants[1413,2]<-"limon"
nursery_plants[1416,2]<-"aurantiifolia"
nursery_plants[1436,2]<-"wittrockiana"
nursery_plants[1442,2]<-"spectabilis"
nursery_plants[1444,2]<-"hybrida"
nursery_plants[1504,2]<-"buttiana"



###deleting columns I don't need
list_lcvp<-nursery_plants[,-c(4:6,8:9)]
str(list_lcvp)
#data.frame':     2333 obs. of  5 variables:
```

```
#$ GENUS             : chr  "Abelia" "Abutilon" "Abutilon" "Abutilon" ...
#$ SPECIFIC.NAME     : chr  "grandiflora" "megapotamicum " "striatum" "x hybridum" ...
#$ VARIETY.SUBSPECIES: chr  "" "variegata" "thompsonii" "" ...
#$ Nursery           : chr  "Golden Stairs" "Golden Stairs" "Golden Stairs" "Golden Stairs" ...
#$ Country           : chr  "Zimbabwe" "Zimbabwe" "Zimbabwe" "Zimbabwe" ...

####Error: splist should include only binomial names, please check the following names:
##' ', 'Coprosma ', 'Eleagnus ', 'Pyracantha ', 'Chlorophytum ', 'Rudbeckia
##'
##' keeping the rows with binomial species only

list_lcvp<-subset.data.frame(x=list_lcvp, SPECIFIC.NAME!="")
str(list_lcvp)
#'data.frame':     2273 obs. of  5 variables:
#$ GENUS             : chr  "Abelia" "Abutilon" "Abutilon" "Abutilon" ...
#$ SPECIFIC.NAME     : chr  "grandiflora" "megapotamicum " "striatum" "x hybridum" ...
#$ VARIETY.SUBSPECIES: chr  "" "variegata" "thompsonii" "" ...
#$ Nursery           : chr  "Golden Stairs" "Golden Stairs" "Golden Stairs" "Golden Stairs" ...
#$ Country           : chr  "Zimbabwe" "Zimbabwe" "Zimbabwe" "Zimbabwe" ....

###### merging columns into a single one by step
### merging genus with specific name
list_lcvp<-unite(data=list_lcvp, "name", GENUS:SPECIFIC.NAME, sep=" ")
str(list_lcvp)
# data.frame':     2273 obs. of  4 variables:
#$ name              : chr  "Abelia grandiflora" "Abutilon megapotamicum " "Abutilon striatum"
"Abutilon x hybridum" ...
#$ VARIETY.SUBSPECIES: chr  "" "variegata" "thompsonii" "" ...
#$ Nursery           : chr  "Golden Stairs" "Golden Stairs" "Golden Stairs" "Golden Stairs" ...
#$ Country           : chr  "Zimbabwe" "Zimbabwe" "Zimbabwe" "Zimbabwe" ...

##### merging the subspecies

list_lcvp<-unite(data=list_lcvp, "name", name:VARIETY.SUBSPECIES, sep=" ")

str(list_lcvp)
#data.frame':     2273 obs. of  3 variables:
#$ name   : chr  "Abelia grandiflora" "Abutilon megapotamicum variegata" "Abutilon
striatumthompsonii" "Abutilon x hybridum" ...
#$ Nursery: chr  "Golden Stairs" "Golden Stairs" "Golden Stairs" "Golden Stairs" ...
#$ Country: chr  "Zimbabwe" "Zimbabwe" "Zimbabwe" "Zimbabwe" ...

str(unique(list_lcvp$name))
##1,835 species
#chr [1:1835] "Abelia grandiflora" "Abutilon megapotamicum variegata" ...

#### contingency table for frequency in nursery
### first factorising

list_lcvp$spp<-factor(list_lcvp$name)
list_lcvp$nurseries<-factor(list_lcvp$Nursery)
```

```
list_lcvp$countries<-factor(list_lcvp$Country)

plant_per_nursery<-list_lcvp[,4:6]
str(plant_per_nursery)
#'data.frame':    2273 obs. of  3 variables:
#$ spp      : Factor w/ 1835 levels "\xd7 Bacurio  delphinatifolius",..: 12 15 16 17 18 19 22 24 27 29 ...
#$ nurseries: Factor w/ 11 levels "Dodhill","Emerald Seedlings",..: 4 4 4 4 4 4 4 4 4 4 ...
#$ countries: Factor w/ 3 levels "Botswana","Namibia",..: 3 3 3 3 3 3 3 3 3 3 ...

spp_data<-ftable(plant_per_nursery, col.vars = c("spp"), row.vars = c("nurseries"))

write.ftable(spp_data, file= "raw data.csv", sep = ",")
#### the frequency numbers were converted to X in the resulted table
#### "Table S2" in Supplementary Information Excel

#####accumulation curve
####
spp_data<-read.csv("plantpernursery.csv", check.names = F)
str(spp_data)
rare<-specaccum(comm=spp_data, method = "collector")
summary(rare)

plot(rare, ci = 0, ci.type = c("line"), ylab="Number of Species", xlab = "Number of nurseries",
xlim=c(1,11))

###nursery names have been deleted from this script and from the resulted
### supplementary materials

#changing to vector for lcvplants package
spp_list<-c(list_lcvp$name)
str(spp_list)
#chr [1:2273] "Abelia grandiflora" "Abutilon megapotamicum variegata" "Abutilon
striatumthompsonii" ...


###standardising names with lcvp
plantlist<-lcvp_search(splist=spp_list, show_correct = T, grammar_check = T, progress_bar = T)

#Error: splist should include only valid characters, please check the name(s) at position(s):
#1701, 1707, 1716, 2147, 2152, 2173, 2175, 2207

c(spp_list[1701], spp_list[1707], spp_list[1716],spp_list[2147], spp_list[2152],spp_list[2173],
  spp_list[2175], spp_list[2207])

#[1] "Citrus  \xd7 aurantium "        "Viola \xd7 wittrockiana"        "\xd7 Bacurio  delphinatifolius"
#[4] "Chrysanthemum \xd7 morifolium"  "Pelargonium \xd7 hybridum"      "Clivia \xd7
cyrtanthiflora"
#[7] "Petunia \xd7 atkinsiana"        "Leucanthemum \xd7 superbum"

spp_list[1701]<-"Citrus x aurantium"
spp_list[1707]<-"Viola x wittrockiana"
```

```
spp_list[1716]<-"Bacurio x delphinatifolius"
spp_list[2147]<-"Chrysanthemum x morifolium"
spp_list[2152]<-"Pelargonium x hybridum"
spp_list[2173]<-"Clivia x cyrtanthiflora"
spp_list[2175]<-"Petunia x atkinsiana"
spp_list[2207]<-"Leucanthemum x superbum"

lcvp_summary(plantlist)
#Species searched: 2273
#Species matched: 2050 (90.19%)
#Species exactly matched: 1863 (81.96%)
#Species fuzzy matched: 187 (8.23%)
#Authors exactly matched: 0 (0%)
#Infracategories exactly matched: 2041 (89.79%)
##### "Table S3" in Supplementary Information Excel

str(plantlist)
#data.frame':     2273 obs. of  15 variables:
#$ Search                   : chr  "Abelia grandiflora" "Abutilon megapotamicum variegata" "Abutilon
striatumthompsonii" "Abutilon x hybridum" ...
#$ global.Id                : int  377 2386 NA 2276 2721 2802 4042 4439 5022 5353 ...
#$ Input.Genus              : chr  "Abelia" "Abutilon" NA "Abutilon" ...
#$ Input.Epitheton          : chr  "grandiflora_x" "megapotamicum" NA "hybridum_x" ...
#$ Rank                     : chr  "species" "species" NA "species" ...
#$ Input.Subspecies.Epitheton: chr  "nil" "nil" NA "nil" ...
#$ Input.Authors            : chr  "(Rovelli ex André) Rehder" "(A.Spreng.) A.St.-Hil. & Naudin" NA
"Voss" ...
#$ Status                   : chr  "accepted" "synonym" NA "unresolved" ...
#$ globalId.of.Output.Taxon  : int  377 211097 NA 2275 1278462 507312 1140256 1278372 1140360
1140396 ...
#$ Output.Taxon             : chr  "Abelia grandiflora_x (Rovelli ex André) Rehder " "Callianthe
megapotamica (A.Spreng.) Dorr" NA "Abutilon hulseanum (Torr. & A.Gray) Torr. ex A.Gray " ...
#$ Family                   : chr  "Linnaeaceae" "Malvaceae" NA "---" ...
#$ Order                    : chr  "Dipsacales" "Malvales" NA "---" ...
#$ Literature               : chr  "" "" NA "" ...
#$ Comments                 : chr  "" "" NA "" ...
#$ Correct                  : logi  TRUE TRUE NA TRUE TRUE TRUE ...

missing<-subset(plantlist, is.na(Status)==T)
str(unique(missing$Search)) #chr [1:207] "Abutilon striatumthompsonii" "Antigonum leptotus"
"Artemesia afra" ...
## 207 names that could not find

unresolved<-subset(plantlist, Status=="unresolved")
str(unique(unresolved$Search)) #chr [1:12] "Abutilon x hybridum" "Begonia hybrida" "Hemerocallis
hybrida" ...
### 12 taxa unresolved

### 219 unresolved or NA taxa

synonyms<-subset.data.frame(plantlist, Status=="synonym")
```

```
str(unique(synonyms$Search)) #chr [1:320] "Abutilon megapotamicum
# 320 synonyms

####writing the result of the taxonomic standardisation
write.csv(plantlist, file = "taxonomic standardisation.csv")##"Table S4" in Supplementary
Information Excel


###removing NA and unresolved and getting the single values
nursery_list<-subset(plantlist, select = c (Output.Taxon, Family, Status))
str(nursery_list)
#data.frame':	2273 obs. of  3 variables:
#$ Output.Taxon: chr  "Abelia grandiflora_x (Rovelli ex André) Rehder " "Callianthe megapotamica
(A.Spreng.) Dorr" NA "Abutilon hulseanum (Torr. & A.Gray) Torr. ex A.Gray " ...
#$ Family      : chr  "Linnaeaceae" "Malvaceae" NA "---" ...
#$ Status      : chr  "accepted" "synonym" NA "unresolved" ...

nursery_list<-subset.data.frame(nursery_list, subset = (Status== "accepted" | Status== "synonym"))
str(nursery_list)
#data.frame':	2035 obs. of  3 variables:
#$ Output.Taxon: chr  "Abelia grandiflora_x (Rovelli ex André) Rehder " "Callianthe megapotamica
(A.Spreng.) Dorr" "Vachellia sieberiana (DC.) Kyal. & Boatwr. " "Faidherbia albida (Delile) A.Chev.
" ...
#$ Family      : chr  "Linnaeaceae" "Malvaceae" "Fabaceae" "Fabaceae" ...
#$ Status      : chr  "accepted" "synonym" "synonym" "synonym" ...

nursery_list<-nursery_list[,-3]

nursery_list<-unique.data.frame(nursery_list)
str(nursery_list)
#'data.frame':	1211 obs. of  3 variables:
#$ Output.Taxon: chr  "Abelia grandiflora_x (Rovelli ex André) Rehder " "Callianthe megapotamica
(A.Spreng.) Dorr" "Vachellia sieberiana (DC.) Kyal. & Boatwr. " "Faidherbia albida (Delile) A.Chev.
" ...
#$ Family      : chr  "Linnaeaceae" "Malvaceae" "Fabaceae" "Fabaceae" ...



###
write.csv(nursery_list, file = "compiled species.csv")## "Table S5" in Supplementary Information
Excel
#### 1,211 species



##### importing the list GIIRS list
GIIRSlist<-read.csv("Subregion_GIIRS_v.(2020).csv", header = T)
str(GIIRSlist)
#'data.frame':	5069 obs. of  18 variables:
#$ id             : int  30543 30544 30545 30546 30547 30548 30549 30550 30551 30552 ...
#$ taxonID        : int  30543 30544 30545 30546 30547 30548 30549 30550 30551 30552 ...
#$ scientificName    : chr  "Acacia saligna (Labill.) H.L.Wendl." "Acanthospermum hispidum DC."
"Achyranthes aspera var. pubescens (Moq.) C.C.Towns." "Acridotheres tristis (Linnaeus, 1766)" ...
```

```
#$ acceptedNameUsage : chr  "" "" "" "" ...
#$ kingdom           : chr  "Plantae" "Plantae" "Plantae" "Animalia" ...
#$ phylum            : chr  "Tracheophyta" "Tracheophyta" "Tracheophyta" "Chordata" ...
#$ class             : chr  "Magnoliopsida" "Magnoliopsida" "Magnoliopsida" "Aves" ...
#$ order             : chr  "Fabales" "Asterales" "Caryophyllales" "Passeriformes" ...
#$ family            : chr  "Fabaceae" "Asteraceae" "Amaranthaceae" "Sturnidae" ...
#$ taxonRank         : chr  "SPECIES" "SPECIES" "VARIETY" "SPECIES" ...
#$ taxonomicStatus   : chr  "ACCEPTED" "ACCEPTED" "ACCEPTED" "ACCEPTED" ...
#$ X                 : chr  "" "" "" "" ...
#$ isInvasive        : chr  "Null" "Invasive" "Invasive" "Null" ...
#$ habitat           : chr  "Terrestrial" "Terrestrial" "Terrestrial" "Terrestrial" ...
#$ countryCode       : chr  "BW" "BW" "BW" "BW" ...
#$ occurrenceStatus  : chr  "Present" "Present" "Present" "Present" ...
#$ establishmentMeans: chr  "Alien" "Alien" "Alien" "Alien" ...
#$ X.1               : int  NA NA NA NA NA NA NA NA NA NA ...

###R is taking the country code NA as NA (missing values)
GIIRSlist$countryCode<-replace_na(GIIRSlist$countryCode, replace = "NA")

###only plants and columns I am interested in
GIIRSlist<-subset.data.frame(GIIRSlist, subset = kingdom== "Plantae", select = c("scientificName",
"isInvasive", "countryCode"))
str(GIIRSlist)
##'data.frame':   4132 obs. of  3 variables:
#$ scientificName: chr  "Acacia saligna (Labill.) H.L.Wendl." "Acanthospermum hispidum DC."
"Achyranthes aspera var. pubescens (Moq.) C.C.Towns." "Agave sisalana Perrine" ...
#$ isInvasive    : chr  "Null" "Invasive" "Invasive" "Null" ...
#$ countryCode   : chr  "BW" "BW" "BW" "BW" ...


###Error: The second list of species name should include only valid characters, please check the
#name(s) at position(s): 130, 408, 1215, 2207, 2332, 2340, 2365, 2481, 2572, 2575, 2611, 2625,
#2729, 2730, 2906, 2912, 2967, 3107, 3196, 3340, 3349, 3366, 3367, 3920

GIIRSlist$scientificName[c(130, 408, 1215, 2207, 2332, 2340, 2365, 2481, 2572, 2575, 2611, 2625,
                           2729, 2730, 2906, 2912, 2967, 3107, 3196, 3340, 3349, 3366, 3367, 3920)]
#"Acalypha amentacea subsp. wilkesiana (M\xfcll.Arg.) Fosberg"
#[2] "Passiflora tripartita var. mollissima (Kunth) Holm-Niels. & P.J\xf8rg."
#[3] "Homalocladium platycladum (F.M\xfcll.) L.H.Bailey"
#[4] "Acalypha wilkesiana M\xfcll.Arg."
#[5] "Anigozanthos flavidus Redout\xe9"
#[6] "Anthurium andraeanum Linden ex Andr\xe9"
#[7] "Asparagopsis taxiformis (Delile) Trevisan de Saint-L\xe9on, 1845"
#[8] "Canna \xd7generalis L. H. Bailey"
#[9] "Citrus \xd7sinensis Pers."
#[10] "Cladophora prolifera (Roth) K\xfctzing, 1843"
#[11] "Cortaderia jubata (Lemoine ex Carri\xe8re) Stapf"
#[12] "Crataegus \xd7lavallei H\xe9rincq ex Lavall\xe9e"
#[13] "Epidendrum \xd7obrienianum Rolfe"
#[14] "Epipremnum aureum (Linden & Andr\xe9) G.S.Bunting"
#[15] "Hymenosporum flavum (Hook.) F.M\xfcll."
```

```
#[16] "Hypochaeris brasiliensis Larra\xf1aga"
#[17] "Lappula heteracantha (Ledeb.) G\xfcrke"
#[18] "Muehlenbeckia platyclados (F.M\xfcll.) Meisn."
#[19] "Paspalum notatum Fl\xfcgg\xe9"
#[20] "Quercus palustris M\xfcnchh."
#[21] "Reynoutria bohemica J. Chrtek & A. Chrtkov\xe1"
#[22] "Rubus \xd7proteus C.H.Stirt."
#[23] "Rubus \xd7proteus C.H.Stirt."
#[24] "Canavalia virosa N\xe1ves ex Fern.-Vill."

GIIRSlist$scientificName[c(130, 408, 1215, 2207, 2332, 2340, 2365, 2481, 2572, 2575, 2611, 2625,
                          2729, 2730, 2906, 2912, 2967, 3107, 3196, 3340, 3349, 3366, 3367, 3920)]<-
  c("Acalypha amentacea subsp. wilkesiana", "Passiflora tripartita var. mollissima",
    "Homalocladium platycladum", "Acalypha wilkesiana", "Anigozanthos flavidus", "Anthurium
andraeanum", "Asparagopsis taxiformis",
    "Canna generalis", "Citrus sinensis", "Cladophora prolifera", "Cortaderia jubata",
    "Crataegus lavallei", "Epidendrum obrienianum", "Epipremnum aureum", "Hymenosporum
flavum",
    "Hypochaeris brasiliensis", "Lappula heteracantha", "Muehlenbeckia platyclados", "Paspalum
notatum",
    "Quercus palustris", "Reynoutria bohemica", "Rubus proteus", "Rubus proteus", "Canavalia
virosa")

##re-running after fixing the error
###what species from the GIIRS list are in the plant list from nurseries

invasive_plants<-lcvp_join(x=GIIRSlist, y=nursery_list,
                          sp_columns=c("scientificName", "Output.Taxon"), type="inner",
                          grammar_check=T, max_distance=10)

str(invasive_plants)
#'data.frame':    694 obs. of  17 variables:
#$ Species.List.1            : chr  "Alternanthera sessilis (L.) DC." "Antigonon leptopus Hook. & Arn."
"Brachychiton acerifolius F.Muell." "Caesalpinia decapetala (Roth) Alston" ...
#$ Species.List.2            : chr  "Alternanthera sessilis (L.) R.Br. ex DC. " "Antigonon leptopus Hook.
& Arn. " "Brachychiton acerifolius (A.Cunn. ex G.Don) Macarthur & C.Moore " "Caesalpinia
decapetala (Roth) Alston " ...
#$ global.Id                : chr  "58012" "86866" "181350" "202606" ...
#$ Input.Epitheton          : chr  "sessilis" "leptopus" "acerifolius" "decapetala" ...
#$ Rank                     : chr  "species" "species" "species" "species" ...
#$ Input.Subspecies.Epitheton: chr  "nil" "nil" "nil" "nil" ...
#$ Input.Authors            : chr  "(L.) R.Br. ex DC." "Hook. & Arn." "(A.Cunn. ex G.Don) Macarthur
& C.Moore" "(Roth) Alston" ...
#$ Status                   : chr  "accepted" "accepted" "accepted" "accepted" ...
#$ globalId.of.Output.Taxon  : chr  "58012" "86866" "181350" "202606" ...
#$ Output.Taxon             : chr  "Alternanthera sessilis (L.) R.Br. ex DC. " "Antigonon leptopus Hook.
& Arn. " "Brachychiton acerifolius (A.Cunn. ex G.Don) Macarthur & C.Moore " "Caesalpinia
decapetala (Roth) Alston " ...
#$ Family                   : chr  "Amaranthaceae" "Polygonaceae" "Sterculiaceae" "Fabaceae" ...
#$ Order                    : chr  "Caryophyllales" "Caryophyllales" "Malvales" "Fabales" ...
#$ Literature               : chr  "" "" "" "SAfr J Bot 2013.89 111-127; " ...
```

```
#$ Comments                 : chr  "" "" "" "" ...
#$ isInvasive               : chr  "Null" "Null" "Null" "Null" ...
#$ countryCode              : chr  "BW" "BW" "BW" "BW" ...
#$ Family.1                 : chr  "Arecaceae" "Arecaceae" "Fabaceae" "Sterculiaceae" ...

##keeping the columns I need
invasive_plants<-subset.data.frame(invasive_plants, select = c("Output.Taxon", "Family",
"isInvasive", "countryCode"))
str(invasive_plants)
#data.frame':     694 obs. of  4 variables:
#$ Output.Taxon: chr  "Alternanthera sessilis (L.) R.Br. ex DC. " "Antigonon leptopus Hook. & Arn.
" "Brachychiton acerifolius (A.Cunn. ex G.Don) Macarthur & C.Moore " "Caesalpinia decapetala
(Roth) Alston " ...
#$ Family      : chr  "Amaranthaceae" "Polygonaceae" "Sterculiaceae" "Fabaceae" ...
#$ isInvasive  : chr  "Null" "Null" "Null" "Null" ...
#$ countryCode : chr  "BW" "BW" "BW" "BW" ...

str(unique(invasive_plants$Output.Taxon)) #289

##### unifying terms

invasive_plants["isInvasive"][invasive_plants["isInvasive"] == "Null"] <- "established"

invasive_plants["isInvasive"][invasive_plants["isInvasive"] == "null"] <- "established"

invasive_plants["isInvasive"][invasive_plants["isInvasive"] == "Invasive"] <- "invasive"

###making it factor
invasive_plants$countryCode<-factor(invasive_plants$countryCode)
invasive_plants$isInvasive<-factor(invasive_plants$isInvasive)
invasive_plants$Output.Taxon<-factor(invasive_plants$Output.Taxon)

####
#'data.frame':     694 obs. of  4 variables:
#$ Output.Taxon: Factor w/ 289 levels "Acalypha wilkesiana Müll.Arg. ",..: 20 25 40 46 50 56 57 58
62 91 ...
#$ Family      : chr  "Amaranthaceae" "Polygonaceae" "Sterculiaceae" "Fabaceae" ...
#$ isInvasive  : Factor w/ 2 levels "established",..: 1 1 1 1 1 1 1 1 1 1 ...
#$ countryCode : Factor w/ 12 levels "AO","BW","CD",..: 2 2 2 2 2 2 2 2 2 2 ...

write.csv(invasive_plants, file= "GRIIS match.csv")## as "Table S6" in Supplementary Information
Excel

(289/1211)*100 ###  24 % of the plants are recorded as established and/or
### invasive in mainland southern Africa

list_match<-unique(invasive_plants$Output.Taxon)

write.csv(list_match, file= "subset GRIIS.csv")##### Table S8 with added info
```

```
############making contingency table
percountry<-table(invasive_plants$isInvasive, invasive_plants$countryCode)
perspp<-table(invasive_plants$Output.Taxon, invasive_plants$isInvasive, invasive_plants$countryCode)

###printing contingency table
write.csv(percountry, file = "numbers per country.csv")### transposed in article (Table 1)
write.csv(perspp, file = "numbers per country_spp.csv")#####Table S7 in Suppl. Materials
##### 16 spp. will be in the article because they are recorded established and/or invasive in 8 countries or more
##############################################################################################

##### characterising established and invasive spp.
est_inv<-read.csv("est&inv.csv", header = T)
str(est_inv)
#'data.frame':	289 obs. of  7 variables:
#$ Output.Taxon         : chr  "Acalypha wilkesiana Müll.Arg. " "Acanthus mollis L. " "Achillea millefolium L. " "Acorus calamus L. " ...
#$ Family               : chr  "Euphorbiaceae" "Acanthaceae" "Asteraceae" "Acoraceae" ...
#$ Form                 : chr  "shrub" "herb" "herb" "rhizomatous geophyte" ...
#$ Broad.form.categories: chr  "shrub" "herb" "herb" "herb" ...
#$ Native_range         : chr  "Bismarck Archipelago to SW. Pacific" "NW. Africa, Sardegna to E. Medit." "Subarctic & Temp. Northern Hemisphere to Guatemala" "Central Asia" ...
#$ Continental_scheme   : chr  "Asia-Tropical, Pacific" "Africa, Europe" "Asia-Temperate, Northern America, Southern America" "Asia-Temperate" ...
#$ Common.names         : chr  "copperleaf, Moore's copper Leaf, Jacob's coat" "bears breeches, wild rhubarb" "white/pink yarrow" "acorus" ...

est_inv$Family<-as.factor(est_inv$Family)
est_inv$Continental_scheme<-as.factor(est_inv$Continental_scheme)
est_inv$Broad.form.categories<-as.factor(est_inv$Broad.form.categories)
str(est_inv)
#$ Family: Factor w/ 86 levels "Acanthaceae",..: 38 1 18 2 39 73 7 3 3 3 ...
#$ Broad.form.categories: Factor w/ 13 levels "climber","epiphyte",..: 9 6
#$ Continental_scheme   : Factor w/ 30 levels "","Africa"


##growth forms

jpeg(filename = "growth forms.jpeg", res=300, quality = 100, height = 1000, width = 1500)

ggplot(growth) + coord_flip() +
  geom_bar(fill="darkblue") + xlab("Growth form") + theme_light() + theme_bw() +
  theme(axis.title.y = element_text(face = "plain", size = 12)) +
  theme(axis.title.x = element_text(face = "plain", size = 12)) + ylab("Number of species") +
  theme(axis.text.x = element_text(size = 10, face = "plain"), axis.text.y = element_text(size = 10, face = "plain"))

dev.off()
```

```
forms<-table(est_inv$Broad.form.categories)
#climber            epiphyte epiphyte-lithophyte           geophyte
#21 (7%)              6                   2                  20 (7%)
#herb             liana         lithophyte              shrub
#32 (11%)          19                  2                  67 (23%)
#shrub-tree           subshrub      subshrub-shrub          succulent
#24 (8%)              15                  6                   9
#tree
#66 (23%)

#### 7 categories as Materials and Methods
### climbers (including lianas): 40
### epiphytes (including lithophytes): 10
### herbs (including geophytes): 52
### shrubs (including subshrubs): 88
### trees: 66
### shrub-trees: 24
### suculents: 9

##total: 289

##### new graph based on 7 categories instead
growth<-cbind.data.frame(c("shrubs", "trees", "shrub-trees", "climbers", "herbs", "epiphytes",
"succulents"),
              c(88, 66,24, 40, 52, 10, 9), stringsAsFactors=TRUE)
colnames(growth)<-c("Growth Form", "Number of taxa")

jpeg(filename = "growth formsv5.jpeg", res=300, quality = 100, height = 1000, width = 1500)
barplot(growth$`Number of taxa`~growth$`Growth Form`, data=growth, col="orange",
ylab="Number of taxa",
        ylim=c(0,100), cex.axis=0.8, cex.names=0.9, las=2, xlab="")
dev.off()


write.csv(forms, file="forms.csv")

######
#by families

families<-table(est_inv$Family)
#Acanthaceae        Acoraceae        Agavaceae        Aizoaceae     Altingiaceae
#8                  1                3                1             1
#Amaranthaceae   Amaryllidaceae    Anacardiaceae        Apiaceae      Apocynaceae
#  5                3               2                   3             12
#Araceae        Araliaceae        Arecaceae Aristolochiaceae     Asparagaceae
#8               4                13               1              8
#Asphodelaceae     Aspleniaceae       Asteraceae    Balsaminaceae      Begoniaceae
#     4               1               14               1              1
#Berberidaceae     Bignoniaceae       Bixaceae     Brassicaceae     Bromeliaceae
#   2                 7               1               2             2
#Cactaceae        Cannaceae   Caprifoliaceae    Casuarinaceae      Clusiaceae
```

```
#1                  2              2              1             1
#Combretaceae    Commelinaceae   Convolvulaceae    Crassulaceae     Cyatheaceae
#    2              5              6              1            1
#Cycadaceae Dennstaedtiaceae    Euphorbiaceae       Fabaceae       Fagaceae
#    1           1             14               15             2
#Haemodoraceae    Heliconiaceae    Hydrangeaceae      Lamiaceae      Lauraceae
#    1              1              1               7             1
#Lythraceae    Malpighiaceae       Malvaceae     Marantaceae  Melastomataceae
#  4             2                 4             1            1
#Meliaceae        Moraceae     Moringaceae       Musaceae       Myrtaceae
#    1               2          2           1           10
#Nephrolepidaceae    Nyctaginaceae       Oleaceae      Onagraceae     Pandanaceae
#2                  2               7          1             1
#Passifloraceae   Phyllanthaceae       Pinaceae      Piperaceae   Plantaginaceae
#1                1                1             2             1
#Plumbaginaceae         Poaceae     Polygalaceae     Polygonaceae    Polypodiaceae
#1                     8             2             3            2
#Portulacaceae      Proteaceae     Pteridaceae      Rhamnaceae       Rosaceae
#1                    3             2             1              4
#Rubiaceae        Rutaceae      Salicaceae      Sapindaceae Scrophulariaceae
#   5              4             2             1             3
# 8              1               8             2                 3
#Solanaceae   Strelitziaceae     Verbenaceae     Viburnaceae       Vitaceae
#Zingiberaceae 4


write.csv(families, file = "families.csv")

##### families above 8 species
library(wordcloud)

jpeg(filename = "families.jpeg", res=300, quality = 100, height = 1000, width = 1500)

wordcloud(est_inv$Family, min.freq = 8, scale=c(2,0.5), colors = brewer.pal(6,"Set3"))

dev.off()

##########regions
range<-table(est_inv$Continental_scheme)
#Africa: 38     Africa, Asia-Temperate: 2  Africa, Asia-Temperate, Asia-Tropical: 2
#Africa, Asia-Temperate, Asia-Tropical, Australasia, Pacific: 2
#Africa, Asia-Temperate, Asia-Tropical, Europe: 1
#Africa, Asia-Temperate, Asia-Tropical, Europe, Northern America: 4
#Africa, Asia-Temperate, Europe: 8
#Africa, Asia-Tropical, Australasia: 2
#Africa, Asia-Tropical, Australasia, Pacific: 1
#Africa, Europe: 2
#Asia-Temperate: 21
#Asia-Temperate, Asia-Tropical: 22
#Asia-Temperate, Asia-Tropical, Australasia: 2
#Asia-Temperate, Asia-Tropical, Australasia, Northern America, Southern America: 1
```

```
#Asia-Temperate, Asia-Tropical, Europe: 1
#Asia-Temperate, Europe: 4
#Asia-Temperate, Northern America, Southern America: 1
#Asia-Tropical: 15
#Asia-Tropical, Australasia: 7
#Asia-Tropical, Australasia, Pacific: 1
#Asia-Tropical, Pacific: 3
#Australasia: 18
#Australasia, Pacific: 1
#Europe: 2
#Northern America: 15
#Northern America, Southern America: 58
#Pacific: 4
#Pacific, Asia-Tropical: 1
#Southern America: 48
#hybrids with no native range: 3

##143 with more than one region, 143 in only one, but latitudinally wide: Southern Africa and Africa

##Regions	Number of species
#Northern America, Southern America	58
#Southern America	48
#Africa	38
#Asia-Temperate, Asia-Tropical	22
#Asia-Temperate	21
#Australasia	18
#Asia-Tropical	15
#Northern America	15

##total 286

write.csv(range, file="ranges.csv")

#######

jpeg(filename = "ranges.jpeg", res=300, quality = 100, height = 1000, width = 2000)

ggplot(data=subset(est_inv, !Continental_scheme==""), aes(Continental_scheme, ..count..)) +
coord_flip() + theme_light() +
  geom_bar(fill="darkorange") + xlab("Botanical regions") + theme_bw() +
  theme(axis.title.y = element_text(face = "plain", size = 8)) +
  theme(axis.title.x = element_text(face = "plain", size = 8)) + ylab("Number of species") +
  theme(axis.text.x = element_text(size = 6, face = "plain"), axis.text.y = element_text(size = 6, face =
"plain"))

dev.off()


#####################################comparing our subset with Glen's compilation
####### dataset accessed from Omer et al. (2021)
(https://datadryad.org/dataset/doi:10.5061/dryad.v6wwpzgwv)
```

```
Glen.native<-read_excel("C:/Users/rodriguezd/OneDrive - Coventry
University/PhD_Coventry/Project_Southern Africa/databases/databases from other articles for the last
appendix/Cultivated_native_taxa_data.xlsx")
Glen.alien<-read_excel("C:/Users/rodriguezd/OneDrive - Coventry
University/PhD_Coventry/Project_Southern Africa/databases/databases from other articles for the last
appendix/Cultivated_alien_taxa_data.xlsx")

str(Glen.alien)
#tibble [5,317 × 70] (S3: tbl_df/tbl/data.frame)
#$ Original names                  : chr [1:5317] "Abelia chinensis R.Br." "Abelia schumannii (Graebn.)
Rehder" "Abelia triflora R.Br." "Abelmoschus esculentus (L.) Moench" ...
#$ TPL names                       : chr [1:5317] "Abelia chinensis" "Abelia parvifolia" "Abelia triflora"
"Abelmoschus esculentus" ...
#$ Family

str(Glen.native)
#tibble [1,595 × 4] (S3: tbl_df/tbl/data.frame)
#$ Original taxon names: chr [1:1595] "Abutilon sonneratianum (Cav.) Sweet" "Acacia ataxacantha
DC." "Acacia caffra Willd." "Acacia galpinii Burtt Davy" ...
#$ TPL names           : chr [1:1595] "Abutilon sonneratianum" "Acacia ataxacantha" "Acacia caffra"
"Acacia galpinii" ...
#$ Family              : chr [1:1595] "Malvaceae" "Leguminosae" "Leguminosae" "Leguminosae" ...
#$ Taxonomic group     : chr [1:1595] "Dicotyledonae" "Dicotyledonae" "Dicotyledonae"
"Dicotyledonae" ...

Glen_complete<-rbind.data.frame(Glen.alien[,2:3], Glen.native[,2:3])
str(Glen_complete)
#tibble [6,912 × 2] (S3: tbl_df/tbl/data.frame)
#$ TPL names: chr [1:6912] "Abelia chinensis" "Abelia parvifolia" "Abelia triflora" "Abelmoschus
esculentus" ...
#$ Family   : chr [1:6912] "Caprifoliaceae" "Caprifoliaceae" "Caprifoliaceae" "Malvaceae" ...

str(nursery_list)
# $ Output.Taxon: chr  "Abelia grandiflora_x (Rovelli ex André) Rehder " "Callianthe megapotamica
(A.Spreng.) Dorr" "Callianthe striata (Dicks. ex Lindl.) Donnell " "Vachellia sieberiana (DC.) Kyal.
& Boatwr. " ...
#$ Family      : chr  "Linnaeaceae" "Malvaceae" "Malvaceae" "Fabaceae" ...

###what species from the Glen list are in the plant list from nurseries

cover<-lcvp_join(x=Glen_complete, y=nursery_list,
                  sp_columns=c("TPL names", "Output.Taxon"), type="inner",
                  grammar_check=T, max_distance=10)
str(cover)
#'data.frame':    680 obs. of  16 variables:
#$ Species.List.1          : chr  "Acalypha wilkesiana" "Acanthus mollis" "Achillea filipendulina"
"Achillea millefolium" ...
#$ Species.List.2          : chr  "Acalypha wilkesiana Müll.Arg. " "Acanthus mollis L. " "Achillea
filipendulina Lam. " "Achillea millefolium L. " ...
```

```
unique(cover$Output.Taxon)## 679

####checking match with the GIIRS subset
str(est_inv)
#'data.frame':	289 obs. of  7 variables:
#$ Output.Taxon         : chr  "Acalypha wilkesiana Müll.Arg. " "Acanthus mollis L. " "Achillea
millefolium L. " "Acorus calamus L. " ...
#$ Family               : Factor w/ 86 levels "Acanthaceae",..: 38 1 18 2 39 73 7 3 3 3 ...
#$ Form                 : chr  "shrub" "herb" "herb" "rhizomatous geophyte" ...
#$ Broad.form.categories: Factor w/ 13 levels "climber","epiphyte",..: 8 5 5 5 13 7 4 12 12 12 ...


cover_est<-lcvp_join(x=Glen_complete, y=est_inv, sp_columns=c("TPL names", "Output.Taxon"),
type="inner",
          grammar_check=T, max_distance=10)
str(cover_est)
#'data.frame':	249 obs. of  21 variables:
#$ Species.List.1           : chr  "Acalypha wilkesiana" "Acanthus mollis" "Achillea millefolium"
"Acorus calamus" ...
#$ Species.List.2           : chr  "Acalypha wilkesiana Müll.Arg. " "Acanthus mollis L. " "Achillea
millefolium L. " "Acorus calamus L. " ...
#$ global.Id                : chr  "8355" "9590" "12876" "18622" ...

unique(cover_est$Output.Taxon)##248
###
########## the script is recorded as Supplementary Material
######
```