ALA ACT Reptile Records (02-2023)

Audrey Miranda Prasetya

2023-03-01

#loading necessary packages  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 1.0.0   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.5.0   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(viridis)

## Loading required package: viridisLite

.csv file downloaded from [Atlas of Living Australia](https://biocache.ala.org.au/occurrences/search?q=lsid%3Ahttps%3A%2F%2Fbiodiversity.org.au%2Fafd%2Ftaxa%2F682e1228-5b3c-45ff-833b-550efd40c399&qualityProfile=ALA&fq=state%3A%22Australian+Capital+Territory%22) on 28 February 2023. Full occurrence records with Full Darwin Core download format was selected.

#read file  
act.rept <- read.csv("dwc-reptiles-act.csv")  
  
#check columns and select to interested   
head(act.rept)  
colnames(act.rept)  
  
#selected columns of interest, mainly those with columns of that has most data  
act.df <- act.rept %>% select(species, genus, family, order, vernacularName, sex,   
 individualCount, basisOfRecord, occurrenceStatus, provenance,   
 year, month, habitat)

A quick check and view of the data to determine whether to keep some variables.

#quick check and explore  
summary(act.df)   
str(act.df) #some "empty" cells, be aware  
  
act.df %>% count(species) #remove records with no species name and count no of. records  
  
act.df <- act.df %>% filter(species != "") #remove records with no species names as it is less informative  
  
act.df %>% count(family, genus)  
  
act.df %>% count(sex) #very little of the records actually have sex recorded - remove  
act.df %>% count(individualCount) #more than half has individual count records  
act.df %>% count(basisOfRecord) #all has basis of record   
act.df %>% count(occurrenceStatus) #no 'absent' type data - remove   
act.df %>% count(provenance) #20% from published dataset, many not recorded info  
act.df %>% count(year) #check for NAs when use  
act.df %>% count(month) #check for NAs when use  
act.df %>% count(habitat) #inconsistent usage of terms, might be difficult to use for this quick analysis - remove

Final decision on 10 selected variables to use for visualization.

act.df <- act.df %>% select(-sex, -occurrenceStatus, -habitat)

Changing and refactoring of variables for consistency and plotting.

#change meaningful variables to factor  
act.df$family <- as.factor(act.df$family)  
act.df$basisOfRecord <- factor(act.df$basisOfRecord,  
 labels = c("Human Observation", "Material Sample", "Observation", "Occurrence", "Preserved Specimen"))  
act.df$month <- as.factor(act.df$month)  
act.df$provenance <- factor(act.df$provenance,   
 levels = c("Published dataset", "Individual sightings", ""),   
 labels = c("Published", "Sightings", "Unknown"))  
  
#adding a season factor to visualise  
act.df <- act.df %>% mutate(season = case\_when(month %in% c(12, 1, 2) ~ "summer",  
 month %in% c(3, 4, 5) ~ "autumn",  
 month %in% c(6, 7, 8) ~ "winter",  
 month %in% c(9, 10, 11) ~ "spring"))  
act.df$season <- factor(act.df$season)  
  
str(act.df) #check

## 'data.frame': 10326 obs. of 11 variables:  
## $ species : chr "Lampropholis delicata" "Lampropholis guichenoti" "Lampropholis delicata" "Intellagama lesueurii" ...  
## $ genus : chr "Lampropholis" "Lampropholis" "Lampropholis" "Intellagama" ...  
## $ family : Factor w/ 11 levels "Agamidae","Chelidae",..: 9 9 9 1 1 1 1 9 1 9 ...  
## $ order : chr "Squamata" "Squamata" "Squamata" "Squamata" ...  
## $ vernacularName : chr "Dark-flecked Garden Sunskink" "Pale-flecked Garden Sunskink" "Dark-flecked Garden Sunskink" "Eastern Water Dragon" ...  
## $ individualCount: int 1 1 NA 1 3 1 4 2 1 2 ...  
## $ basisOfRecord : Factor w/ 5 levels "Human Observation",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ provenance : Factor w/ 3 levels "Published","Sightings",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ year : int 2014 2013 2014 2014 2014 2014 2013 2014 2014 2014 ...  
## $ month : Factor w/ 12 levels "1","2","3","4",..: 3 3 3 3 3 3 9 3 3 3 ...  
## $ season : Factor w/ 4 levels "autumn","spring",..: 1 1 1 1 1 1 2 1 1 1 ...

head(act.df)

## species genus family order  
## 1 Lampropholis delicata Lampropholis Scincidae Squamata  
## 2 Lampropholis guichenoti Lampropholis Scincidae Squamata  
## 3 Lampropholis delicata Lampropholis Scincidae Squamata  
## 4 Intellagama lesueurii Intellagama Agamidae Squamata  
## 5 Intellagama lesueurii Intellagama Agamidae Squamata  
## 6 Intellagama lesueurii Intellagama Agamidae Squamata  
## vernacularName individualCount basisOfRecord provenance year  
## 1 Dark-flecked Garden Sunskink 1 Observation Published 2014  
## 2 Pale-flecked Garden Sunskink 1 Observation Published 2013  
## 3 Dark-flecked Garden Sunskink NA Observation Published 2014  
## 4 Eastern Water Dragon 1 Observation Published 2014  
## 5 Eastern Water Dragon 3 Observation Published 2014  
## 6 Eastern Water Dragon 1 Observation Published 2014  
## month season  
## 1 3 autumn  
## 2 3 autumn  
## 3 3 autumn  
## 4 3 autumn  
## 5 3 autumn  
## 6 3 autumn

## Temporal Trends in ALA’s ACT Reptilia Records

### Type Trends

How do the type of records change over the years?

First, do we see an increase in no of records?

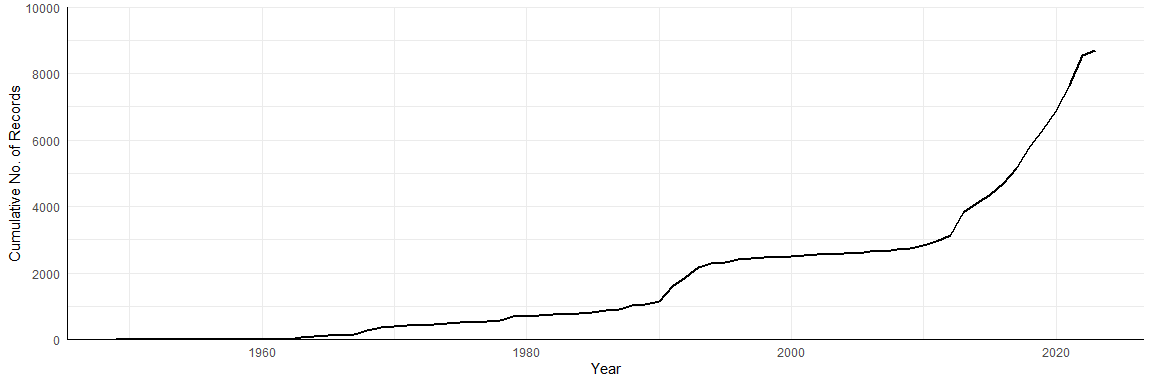
act.year <- act.df %>% filter(year!="") #remove records without years  
head(act.year)

## species genus family order  
## 1 Lampropholis delicata Lampropholis Scincidae Squamata  
## 2 Lampropholis guichenoti Lampropholis Scincidae Squamata  
## 3 Lampropholis delicata Lampropholis Scincidae Squamata  
## 4 Intellagama lesueurii Intellagama Agamidae Squamata  
## 5 Intellagama lesueurii Intellagama Agamidae Squamata  
## 6 Intellagama lesueurii Intellagama Agamidae Squamata  
## vernacularName individualCount basisOfRecord provenance year  
## 1 Dark-flecked Garden Sunskink 1 Observation Published 2014  
## 2 Pale-flecked Garden Sunskink 1 Observation Published 2013  
## 3 Dark-flecked Garden Sunskink NA Observation Published 2014  
## 4 Eastern Water Dragon 1 Observation Published 2014  
## 5 Eastern Water Dragon 3 Observation Published 2014  
## 6 Eastern Water Dragon 1 Observation Published 2014  
## month season  
## 1 3 autumn  
## 2 3 autumn  
## 3 3 autumn  
## 4 3 autumn  
## 5 3 autumn  
## 6 3 autumn

#cumulative no of records  
act.y.sum <- act.year %>% group\_by(year) %>% summarize(n = n())  
act.y.sum$cumulative <- cumsum(act.y.sum$n)  
head(act.y.sum)

## # A tibble: 6 × 3  
## year n cumulative  
## <int> <int> <int>  
## 1 1949 1 1  
## 2 1954 2 3  
## 3 1957 2 5  
## 4 1960 3 8  
## 5 1961 1 9  
## 6 1962 2 11

ggplot(act.y.sum, aes(x = year, y = cumulative)) +  
 geom\_line(linewidth = 1) +  
 scale\_y\_continuous(limits = c(0, 10000), expand = c(0,0), breaks = c(0, 2000, 4000, 6000, 8000, 10000)) +  
 xlab("Year") +  
 ylab("Cumulative No. of Records") +  
 theme\_minimal() +  
 theme(legend.position = "none", axis.line = element\_line(colour = "black"))

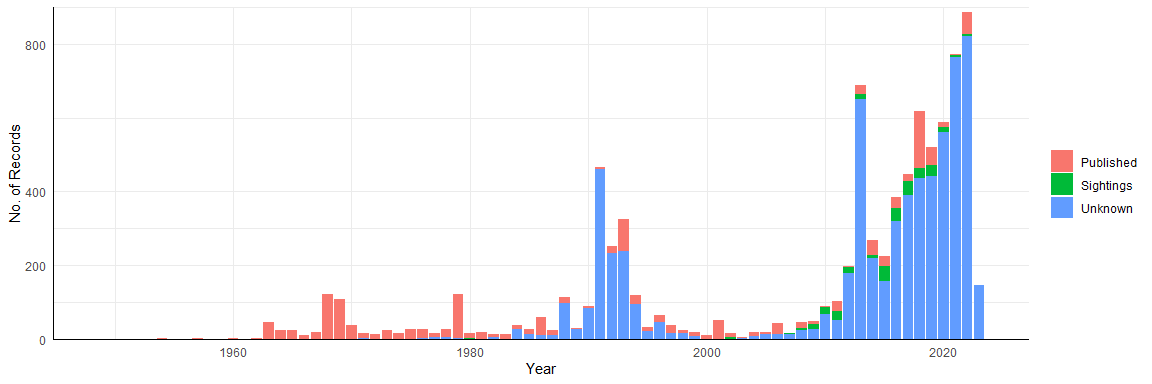


**Figure 1** Cumulative no of reptiles recorded in the Australian Capital Territory from the Atlas of Living Australia.

We can see from Figure 1 that there is a steeper curve post 2010 indicating an increase in the number of records in reptiles. There is also a slight peak at the year 1990.

Do the type of data recorded differ between years?

ggplot(act.year, aes(x = year, fill = provenance)) +  
 geom\_bar(stat="count") +  
 scale\_y\_continuous(limits = c(0, 900), expand = c(0,0), breaks = c(0, 200, 400, 800)) +  
 xlab("Year") +  
 ylab("No. of Records") +  
 labs(fill = "") +  
 theme\_minimal() +  
 theme(legend.position = "right", axis.line = element\_line(colour = "black"))

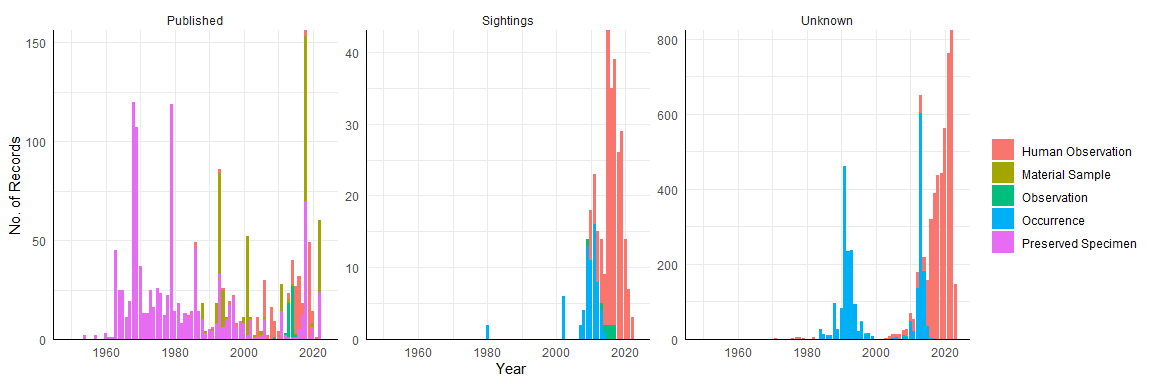


**Figure 2** Number of reptiles recorded in the Australian Capital Territory from the Atlas of Living Australia. Colours highlight provenance of record.

Here in Figure 2 we can see an interesting temporal trend. Early records pre-1990s are mostly published dataset. Only more recently do we see an increase in ‘sightings’ and records with ‘unknown’ provenance.

Can we also see a trend in the basis of record?

ggplot(act.year, aes(x = year, fill = basisOfRecord)) +  
 geom\_bar(stat="count") +   
 facet\_wrap(~provenance, scales = "free\_y") +  
 scale\_y\_continuous(expand = c(0,0)) +  
 xlab("Year") +  
 ylab("No. of Records") +  
 labs(fill = "") +  
 theme\_minimal() +  
 theme(legend.position = "right", axis.line = element\_line(colour = "black"))



**Figure 3** Number of reptiles recorded in the Australian Capital Territory from the Atlas of Living Australia. Records are split based on provenance and colours highlight basis of record.

We get an even clearer view of this in this graph. Most published dataset records are preserved specimens. Very little is actually based on observation/occurences. Human observation samples are mainly found after 2010s, which likely corresponds to the increase in citizen science records. Thus, only more recent papers have started including observation data. Although, only a very small portion of our dataset has been used in published datasets!

### Taxon Trends

Are there changes differences in the taxa recorded over the years? We’ll focus on family-level taxa as it is a manageable category for visualisation.

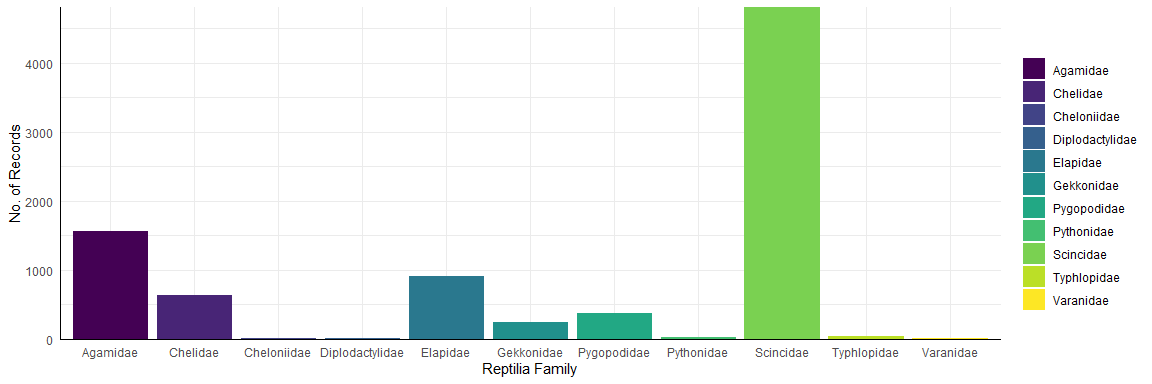
#setting up the dataframe  
fam.count <- act.df %>% filter(year != "") %>% count(family, year)  
head(fam.count)

## family year n  
## 1 Agamidae 1954 1  
## 2 Agamidae 1964 1  
## 3 Agamidae 1965 1  
## 4 Agamidae 1966 1  
## 5 Agamidae 1967 1  
## 6 Agamidae 1968 1

f.col <- viridis(length(unique(fam.count$family))) #colors for visualisation

How many records for each reptile family?

ggplot(fam.count, aes(x = family, y = n, fill = family)) +  
 geom\_bar(stat = "identity") +  
 scale\_fill\_manual(values = f.col) +  
 scale\_y\_continuous(expand = c(0,0)) +  
 xlab("Reptilia Family") +  
 ylab("No. of Records") +  
 labs(fill = "") +  
 theme\_minimal() +  
 theme(legend.position = "right", axis.line = element\_line(colour = "black"))

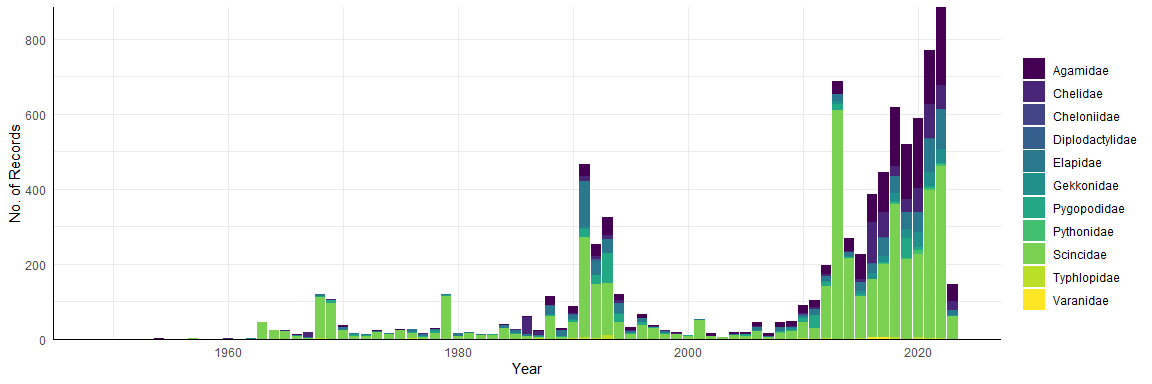


**Figure 4** Number of reptiles recorded in the Australian Capital Territory from the Atlas of Living Australia, categorised by family-level taxonomy.

Most records are overwhelmingly from skinks, followed by agamids.

Does the trend change over time for which family is recorded?

ggplot(act.year, aes(x = year, fill = family)) +  
 geom\_bar(stat="count") +  
 scale\_fill\_manual(values = f.col) +  
 scale\_y\_continuous(expand = c(0,0)) +  
 xlab("Year") +  
 ylab("No. of Records") +  
 labs(fill = "") +  
 theme\_minimal() +  
 theme(legend.position = "right", axis.line = element\_line(colour = "black"), )



**Figure 5** Number of reptiles recorded in the Australian Capital Territory from the Atlas of Living Australia over the years, categorised by family-level taxonomy.

There are consistently a lot of skinks recorded in the ACT!

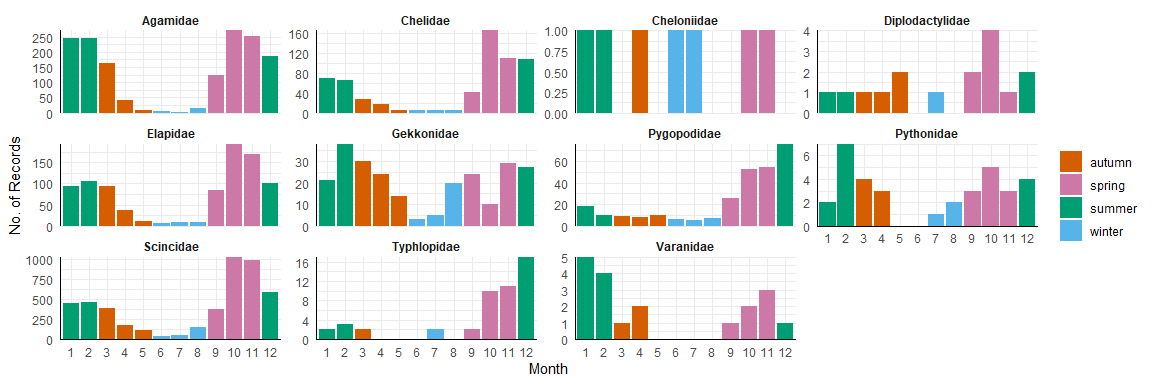
Is there a monthly trend of recorded taxa?

#setting up dataframe and colors  
fam.month <- act.df %>% filter(year != "", month != "") %>% count(family, month, year, season)  
head(fam.month)

## family month year season n  
## 1 Agamidae 1 1970 summer 1  
## 2 Agamidae 1 1975 summer 1  
## 3 Agamidae 1 1986 summer 1  
## 4 Agamidae 1 1988 summer 6  
## 5 Agamidae 1 1991 summer 6  
## 6 Agamidae 1 1992 summer 7

s.col <- c("#D55E00", "#CC79A7", "#009E73","#56B4E9")

ggplot(fam.month, aes(x = month, y = n, fill = season)) +  
 geom\_bar(stat = "identity") +  
 facet\_wrap(~family, scales ="free\_y", ncol=4) +  
 scale\_fill\_manual(values = s.col) +  
 scale\_y\_continuous(expand = c(0,0)) +  
 labs(fill = "") +  
 xlab("Month") +  
 ylab("No. of Records") +  
 theme\_minimal() +  
 theme(axis.line = element\_line(colour = "black"), strip.text = element\_text(face = "bold"))



**Figure 6** Number of reptiles recorded in the Australian Capital Territory from the Atlas of Living Australia over the months, categorised by family-level taxonomy. Colors represent rough seasonal boundaries.

Mostreptile are recorded in the spring and summer and least during the winter which is expected! Trend is less so for sea turtles (Cheloniidae) although we only have one individual recording - this is quite an enigma as the ACT does not have any bodies of water!

I thought that this might have been a museum specimen, but it seems the basis of record is human observation, and from a published study.

act.df %>% filter(family == "Cheloniidae")

## species genus family order vernacularName  
## 1 Natator depressus Natator Cheloniidae Testudines Flatback Turtle  
## 2 Chelonia mydas Chelonia Cheloniidae Testudines Green Turtle  
## 3 Chelonia mydas Chelonia Cheloniidae Testudines Green Turtle  
## 4 Chelonia mydas Chelonia Cheloniidae Testudines Green Turtle  
## 5 Eretmochelys imbricata Eretmochelys Cheloniidae Testudines Hawksbill Turtle  
## 6 Chelonia mydas Chelonia Cheloniidae Testudines Green Turtle  
## 7 Chelonia mydas Chelonia Cheloniidae Testudines Green Turtle  
## individualCount basisOfRecord provenance year month season  
## 1 NA Human Observation Published 2010 6 winter  
## 2 1 Human Observation Published 2010 1 summer  
## 3 1 Human Observation Published 2010 10 spring  
## 4 NA Human Observation Published 2008 7 winter  
## 5 1 Human Observation Published 2010 2 summer  
## 6 1 Human Observation Published 2009 11 spring  
## 7 NA Human Observation Published 2007 4 autumn

Other than that, most reptiles are recorded in months Oct-Dec. Some are present for more towards the spring such as the agamids (Agamidae), skinks (Scincidae), elapids (Elapidae), and freshwater turtles(Cheliidae). Others are found more towards the summer including monitor lizards (Varanidae), geckos (Gekkonidae), and pythons (Pythonidae).

Finally, do we also tend to see more individuals (rather than just no of records?)

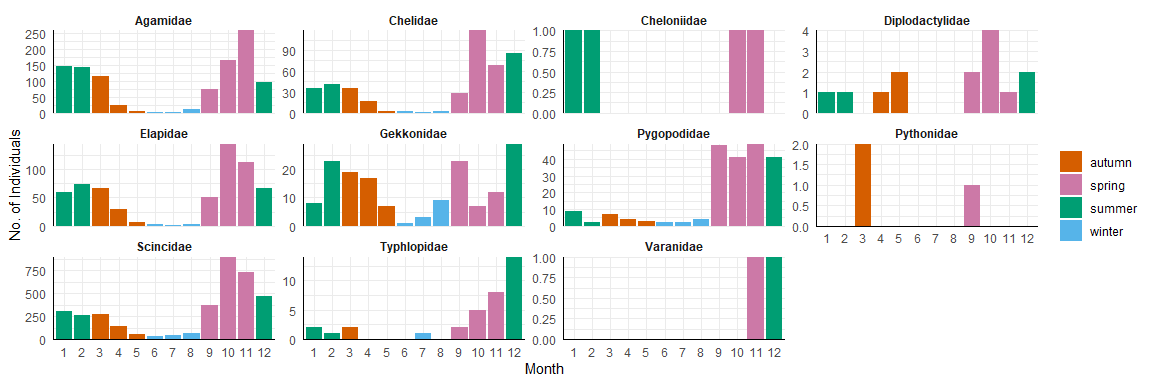
#since we have some data on individual counts for some records, we will sum it up per family per month  
fam.ind <- act.df %>%   
 filter(individualCount > 0, month !="") %>%   
 group\_by(family, month, season) %>%   
 summarize(summed =sum(individualCount))

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

head(fam.ind)

## # A tibble: 6 × 4  
## # Groups: family, month [6]  
## family month season summed  
## <fct> <fct> <fct> <int>  
## 1 Agamidae 1 summer 147  
## 2 Agamidae 2 summer 142  
## 3 Agamidae 3 autumn 116  
## 4 Agamidae 4 autumn 25  
## 5 Agamidae 5 autumn 5  
## 6 Agamidae 6 winter 1

ggplot(fam.ind, aes(x = month, y = summed, fill = season)) +  
 geom\_bar(stat = "identity") +  
 facet\_wrap(~family, scales = "free\_y", ncol=4) +  
 scale\_fill\_manual(values = s.col) +  
 scale\_y\_continuous(expand = c(0,0)) +  
 labs(fill = "") +  
 xlab("Month") +  
 ylab("No. of Individuals") +  
 theme\_minimal() +  
 theme(axis.line = element\_line(colour = "black"), strip.text = element\_text(face = "bold"))



**Figure 7** Number of individual reptiles recorded in the Australian Capital Territory from the Atlas of Living Australia over the months, categorised by family-level taxonomy. Colors represent rough seasonal boundaries.

Here in Figure 7 we show similar patterns to Figure 6. In other words, yes we do see not just more records, but more individuals recorded during the spring-summer months.