Data on swift parrot populations near Toowoomba and Geelong

Bhavya Bhargava

March 22, 2024

Importing the packages that were utilized and citing them

```
#install.packages(c("galah", "tidyverse"))
library(tidyverse)
library(galah)
#packages cited
citation("tidyverse")
```

```
## To cite package 'tidyverse' in publications use:
##
##   Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R,
##   Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller
##   E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V,
##   Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to
##   the tidyverse." _Journal of Open Source Software_, *4*(43), 1686.
##   doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {Welcome to the {tidyverse}},
##     author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Luc
y D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henr
y and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bach
e and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spin
u and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani},
##     year = {2019},
##     journal = {Journal of Open Source Software},
##     volume = {4},
##     number = {43},
##     pages = {1686},
##     doi = {10.21105/joss.01686},
##   }
```

```
citation("galah")
```

```
## To cite galah in publications use:
##
##   Westgate M, Stevenson M, Kellie D, Newman P (2024). _galah:
##   Biodiversity Data from the GBIF Node Network_. R package version
##   2.0.1, <https://CRAN.R-project.org/package=galah>.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {galah: Biodiversity Data from the GBIF Node Network},
##     author = {Martin Westgate and Matilda Stevenson and Dax Kellie and Peggy Newman},
##     year = {2024},
##     note = {R package version 2.0.1},
##     url = {https://CRAN.R-project.org/package=galah},
##   }
```

Downloading the data and filtering it according to the coordinates

The data for this analysis is being retrieved from the Atlas of Living Australia (ALA) website using their ‘galah’ package, a powerful tool designed for this purpose. Below are of the steps taken to perform data acquisition, including comments on crucial decisions I made along the way:

1. Starting off I configured the ‘galah’ session with my email address and specifying the reason for the data download (As taught in tutorial). This is an essential requirement for maintaining the integrity and sustainability of the ALA database.
2. For checking the relevant occurrences, the search parameters were pre-defined (According to requirements).
3. Stored the required location coordinates into respective variables named according to the locations. Here instead of using the tibbles, I explicitly used the coordinates of both the locations as it provides better flexibility.
4. To search for occurrences while keeping code duplication to a minimum, a function was developed which took the coordinates of the location along with the species name and start date of the search as input and churned out the occurrences as per the ALA database.
5. As the data which was received had some irrelevant fields with repetitive data, some filtering was done to remove those and a few fields were even renamed to make data as relevant as possible to the search and further analysis if required.
6. Further the data for the two locations was saved int two separate RDAs for easy location specific analysis later.

Below is the code for the above statement:

```
# Configuring the 'galah' query
galah_config(email = "bbha0006@student.monash.edu",
             download_reason_id = 10,
             verbose = TRUE)

# Specifying the species name and the start date for checking occurrences
species_name <- "Lathamus discolor"
start_date <- "2000-01-01"

# Creating two variables to store the latitude and longitude for the two locations
toowoomba_coords <- c(-27.5538, 151.7956)
geelong_coords <- c(-38.0220, 144.4768)

# Creating a function to minimize code duplication for searching the occurrences
search_occurrences <- function(coords, species, start_date) {
  galah_call() %>%
    galah_identify(species) %>%
    galah_filter(year >= as.numeric(format(as.Date(start_date), "%Y")),
                 decimalLatitude >= coords[1] - 1, decimalLatitude <= coords[1] + 1,
                 decimalLongitude >= coords[2] - 1, decimalLongitude <= coords[2] + 1) %>%
    atlas_occurrences()
}

# Fetching occurrences for Toowoomba
toowoomba_occurrences <- search_occurrences(toowoomba_coords, species_name, start_date) %>%
  select(decimalLatitude, decimalLongitude, eventDate, dataResourceName) %>% # storing only r
equired fields
  rename(Latitude = decimalLatitude, Longitude = decimalLongitude)

# Fetching occurrences for Geelong
geelong_occurrences <- search_occurrences(geelong_coords, species_name, start_date) %>%
  select(decimalLatitude, decimalLongitude, eventDate, dataResourceName) %>% # storing only r
equired fields
  rename(Latitude = decimalLatitude, Longitude = decimalLongitude)

# Data saved into two seperate .rda files for easy analysis
save(toowoomba_occurrences, file = "data/toowoomba_occurrences.rda")
save(geelong_occurrences, file = "data/geelong_occurrences.rda")
```

Data sets description

The data for both the locations is in files `toowoomba_occurrences.rda` and `geelong_occurrences.rda` respectively in the `data` directory.

They contain the following variables:

- Latitude and Longitude-> Contains the coordinates of the exact location where an observation had occurred.
- eventDate-> Indicates the date (and sometimes the time) of the observation or occurrence. This field can prove useful in providing insights into the seasonal movement of the species around the area.
- dataResourceName-> Contains the source of the occurrence data. This field identifies the organization, database, or project that provided the data. This information can prove important for understanding the context of the data, its reliability, and the scope of its observations.

Dataset population and limitations analysis

Our dataset contains observations from reputable sources like `eBird Australia` and `NSW BioNet Atlas`. It aims to represent the species' occurrences around the specified locations in detail. By including data spanning multiple years and from multiple sources, we've aimed to capture a representative sample of the presence of the species, taking into consideration the yearly variations. However, limitations such as potential sampling biases caused by more active bird watching seasons and the usual inaccuracy observer reporting accuracy could affect the dataset's completeness and reliability. Despite these challenges, this dataset provides valuable insights into the specie's distribution and it's activity patterns in the selected areas, helping us to make informed decisions regarding the suitability of these locations for a new storage facility based on their occurrences.

Data Licensing

The licensing situation for data shared via the Atlas of Living Australia (ALA), and presumably for data accessed through the galah package, encourages the use of Creative Commons Australia licenses, with a preference for the Creative Commons Attribution (CC BY) license.
(https://support.ala.org.au/support/solutions/articles/6000197133-what-licence-should-i-use-
(https://support.ala.org.au/support/solutions/articles/6000197133-what-licence-should-i-use-))

To understand the connectivity and observe the location with the least interactions

```
# Loading the subset for analysis(Perhaps changing the below path to just "data/`place_name`_
occurrences.rda" might help.)
load(here::here("C:/Users/user/Desktop/ETC5512/assignment1/data/toowoomba_occurrences.rda"))
load(here::here("C:/Users/user/Desktop/ETC5512/assignment1/data/geelong_occurrences.rda"))

# Function to count the number of occurrences
count_occurences <- function(location_data){
  nrow(location_data)
}

# Presenting Results for Comparison
cat("Number of Swift Parrot encounters in Toowoomba: ", count_occurences(toowoomba_occurrence
s),"\n")
```

```
## Number of Swift Parrot encounters in Toowoomba:  143
```

```
cat("Number of Swift Parrot encounters in Geelong: ", count_occurences(geelong_occurrences),
"\n")
```

```
## Number of Swift Parrot encounters in Geelong:  3343
```

As can be seen from the drastic difference in the number of occurrences of Swift Parrots between both the locations, Toowoomba might prove to be the most suitable location for the storage facility.