ICCB Species data download

Scott Forrest and Charlotte Patterson

2025-06-06

In this script we are downloading data from ALA (with an option for GBIF) for koalas (Phascolarctos cinereus) in the South-East Queensland (SEQ) region. We are then cleaning the data and in other scripts creating a species distribution model for current and future climate scenarios.

Table of contents

Install packages

```
# install.packages("galah")
```

Import packages

```
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(purrr)
library(ggplot2)
library(galah)
```

```
galah: version 2.1.1
i Default node set to ALA (ala.org.au).
i See all supported GBIF nodes with `show_all(atlases)`.
i To change nodes, use e.g. `galah_config(atlas = "GBIF")`.


Attaching package: 'galah'

The following object is masked from 'package:dplyr':

    desc

The following object is masked from 'package:stats':

    filter
```

```
library(terra)
```

```
terra 1.8.50
```

```
library(sf)
```

```
Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```
library(predicts)
library(tidyterra)
```

```
Attaching package: 'tidyterra'

The following object is masked from 'package:stats':

    filter
```

To cite the galah package

```
# Package citation
citation(package = "galah")
```

```
To cite galah in publications use:

  Westgate M, Kellie D, Stevenson M, Newman P (2025). _galah:
  Biodiversity Data from the GBIF Node Network_. R package version
  2.1.1, <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 Dax Kellie and Matilda Stevenson and Peggy Newman},
    year = {2025},
    note = {R package version 2.1.1},
    url = {https://CRAN.R-project.org/package=galah},
  }
```

Load South East Queensland (SEQ) boundary

We made this boundary in the 'ICCB_Environmental_data.qmd' script. Also load local government area polygons (LGAs) just for plotting.

```
SEQ_extent.vect <- vect("Data/Environmental_variables/SEQ_extent.shp")

# Define an sf object as well
SEQ_extent <- st_as_sf(SEQ_extent.vect,
                       coords = c("x", "y"),
```

```
                        crs = 3112)



# Load the study area shapefile
LGA <- st_read("Data/Environmental_variables/Local_Government_Areas.shp")
```

```
Reading layer `Local_Government_Areas' from data source
  `/Users/scottforrest/Library/CloudStorage/OneDrive-QueenslandUniversityofTechnology/PhD -
  using driver `ESRI Shapefile'
Simple feature collection with 78 features and 8 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: 137.9946 ymin: -29.17927 xmax: 153.5519 ymax: -9.087991
Geodetic CRS:  GDA94
```

```
# Check the coordinate reference system (CRS)
st_crs(LGA)
```

```
Coordinate Reference System:
  User input: GDA94
  wkt:
GEOGCRS["GDA94",
    DATUM["Geocentric Datum of Australia 1994",
        ELLIPSOID["GRS 1980",6378137,298.257222101,
            LENGTHUNIT["metre",1]]],
    PRIMEM["Greenwich",0,
        ANGLEUNIT["degree",0.0174532925199433]],
    CS[ellipsoidal,2],
        AXIS["geodetic latitude (Lat)",north,
            ORDER[1],
            ANGLEUNIT["degree",0.0174532925199433]],
        AXIS["geodetic longitude (Lon)",east,
            ORDER[2],
            ANGLEUNIT["degree",0.0174532925199433]],
    USAGE[
        SCOPE["Horizontal component of 3D system."],
        AREA["Australia including Lord Howe Island, Macquarie Island, Ashmore and Cartier Is
        BBOX[-60.55,93.41,-8.47,173.34]],
    ID["EPSG",4283]]
```

```r
# Convert to WGS84
LGA <- LGA %>% st_transform(3112)

# Select local govt. areas for South East Queensland
LGA_SEQ <- LGA %>%
  filter(lga %in% c("Brisbane City",
                    "Moreton Bay City",
                    "Logan City",
                    "Ipswich City",
                    "Redland City",
                    "Scenic Rim Regional",
                    "Somerset Regional",
                    "Lockyer Valley Regional",
                    "Gold Coast City",
                    "Sunshine Coast Regional",
                    "Toowoomba Regional",
                    "Noosa Shire"))
```

Atlas of Living Australia using "galah" package

To access records from the ALA, you will need to be registered.

There are two registration options:

1. If you are affiliated with an Australian institution, you should be able to register for ALA through your institution. Find your institution in the list of institutions when you select 'Login' on the ALA website. Follow the prompts to enter your institution email and password. If your institution is not listed, you will need to register for an ALA account.
2. If you are not affiliated with an Australian institution, you will need to register for an ALA account. You can do this by selecting 'Register' on the ALA website. Follow the prompts to enter your email and password.

```r
galah_config(atlas = "ALA",
             # username = "scott.forrest@hdr.qut.edu.au",
             email = "scott.forrest@hdr.qut.edu.au"
             # password = "Password"
             )
```

Download koala data from ALA

Usually you would check that the CRS for the occurrences is the same as the shapefile. However, we know that the 'galah' package operates in WGS84.

```
koala_occurrences <- galah_call() %>%
  galah_identify("Phascolarctos cinereus") %>%
  galah_filter(
    stateProvince == "Queensland",
    occurrenceStatus == "PRESENT"
  ) %>%
  atlas_occurrences()
```

```
Request for 93456 occurrences placed in queue
Current queue length: 1
```

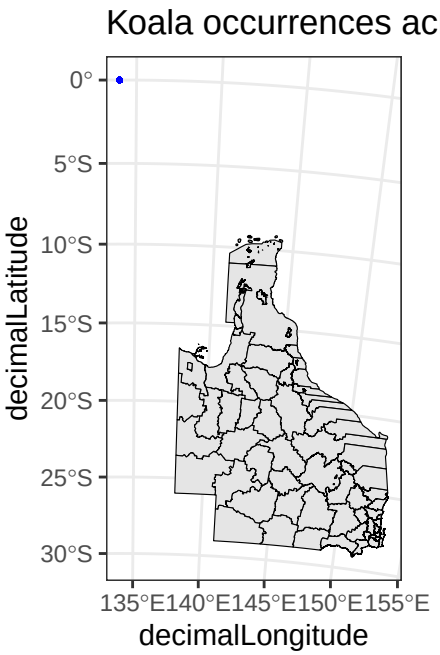
```
-----
```

```
Downloading
```

Clean koala data

```
# Create a map using ggplot2
ggplot() +
  geom_sf(data = LGA, color = "black") +
  geom_point(data = koala_occurrences,
             aes(x = decimalLongitude,
                 y = decimalLatitude),
             color = "blue", size = 0.5) + # Add points for occurrences
  ggtitle("Koala occurrences across Queensland") +             # Add title
  theme_bw()
```

```
Warning: Removed 215 rows containing missing values or values outside the scale range
(`geom_point()`).
```



Filter by SEQ region

```
# Check for missing values in decimalLongitude and decimalLatitude
paste0("Number of NAs in 'longitude' ", sum(is.na(koala_occurrences$decimalLongitude)))
```

```
[1] "Number of NAs in 'longitude' 215"
```

```
paste0("Number of NAs in 'latitude' ", sum(is.na(koala_occurrences$decimalLatitude)))
```

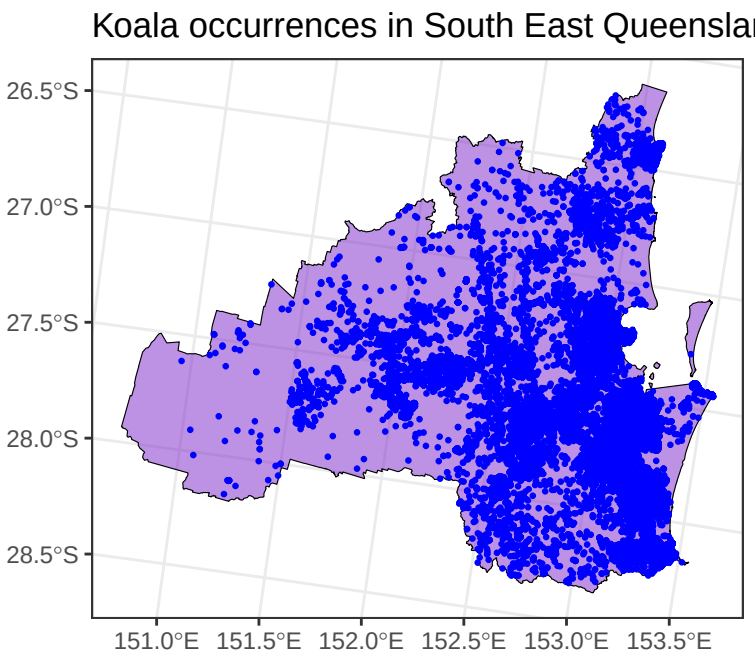
```
[1] "Number of NAs in 'latitude' 215"
```

```
koala_occ_sf <- koala_occurrences %>%
  drop_na(decimalLongitude, decimalLatitude) %>% # remove NA values
  st_as_sf(coords = c("decimalLongitude", "decimalLatitude"),
           crs = 4326) %>%
  st_transform(3112) %>% # Transform to the same CRS as SEQ_extent
  st_intersection(SEQ_extent) %>% # Mask to extent
  distinct() # drop any duplicated records
```

```
Warning: attribute variables are assumed to be spatially constant throughout
all geometries
```

Plot the filtered data

```
# Create a map using ggplot2
ggplot() +
  geom_sf(data = SEQ_extent.vect, fill = "purple3", alpha = 0.5, color = "black", size = 0.2
  geom_sf(data = koala_occ_sf,                              # Add koala presence locations
          aes(geometry = geometry),
              color = "blue", size = 0.5) +                 # Add points for occurrences
  ggtitle("Koala occurrences in South East Queensland") +       # Add title
  theme_bw()
```



Sample background points

We need to load the environmental covariate grid to use as our template grid for creating background points. We generated the environmental covariates in the 'ICCB_Environmental_data.qmd' script.

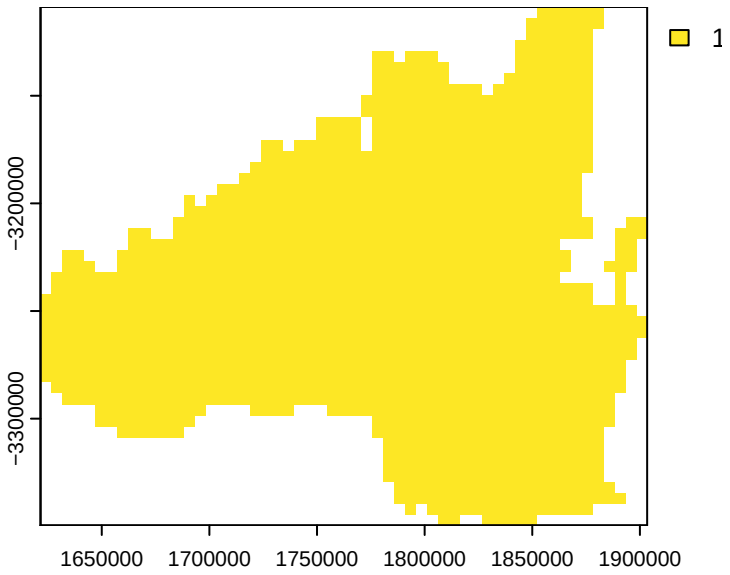
```
covs <- rast("Data/Environmental_variables/current_bioclim.tif")

# Make a blank raster of all 1s of the same dimensions as our covariate grid
domain <- covs[[1]]
values(domain) <- 1
```

```r
# Mask extent by SEQ boundary
domain <- terra::mask(domain, SEQ_extent.vect, updatevalue = NA)

names(domain) <- "SEQ_extent"

plot(domain)
```




```r
# Set the location and number of background points
# Random sampling at first
# The mask means that any NA (not SEQ) locations do not include background points
background <- predicts::backgroundSample(mask = domain,
                                         n = 2500)
```

```
Warning in predicts::backgroundSample(mask = domain, n = 2500): generated
random points = 0.5944 times requested number
```

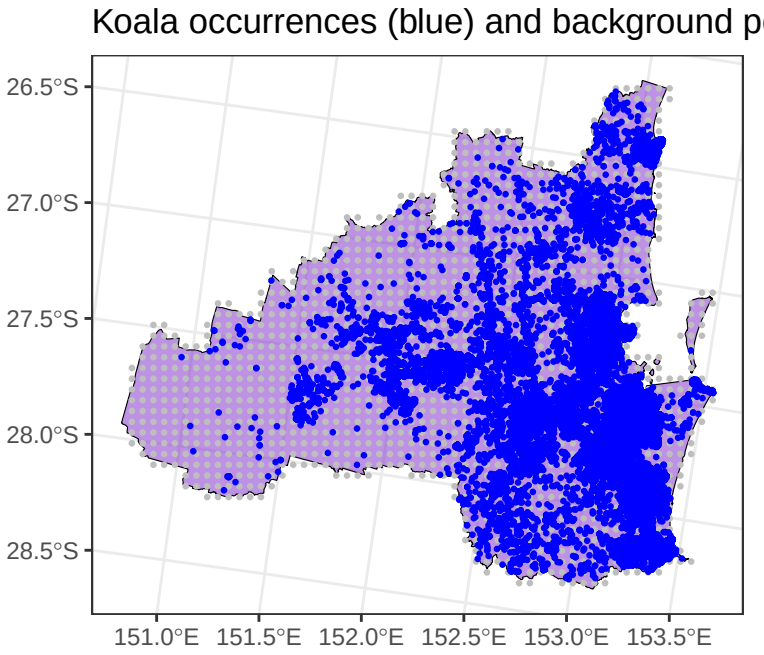
```r
# Convert to terra SpatVector object
background <- terra::vect(background, crs = "EPSG:3112")

# Convert background points (SpatVector) to data frame
background_df <- as.data.frame(geom(background))

koala_occ.vect <- vect(koala_occ_sf)

# Plot the presences (blue) and the background points (grey)
ggplot() +
```

```
  geom_sf(data = SEQ_extent, fill = "purple3", alpha = 0.5, color = "black", size = 0.2) +
  geom_spatvector(data = background,                              # Add koala presence location
             color = "gray", cex = 0.5) +                  # Add points for occurrences
  geom_spatvector(data = koala_occ.vect,
                  aes(geometry = geometry),
                  color = "blue", cex = 0.5) + # Add background points
  ggtitle("Koala occurrences (blue) and background points (grey) in South East Queensland")
  theme_bw()
```



Save koala presences and background points

This ensures quick upload in the future and that these records can be used among models.

```
writeVector(koala_occ.vect, "Data/Biological_records/SEQ_koala_occurrences.shp", overwrite =

writeVector(background, "Data/Biological_records/background_points_2.5k_random.shp", overwri
```