

Reptile Analysis ACT

Isaac Towers

2023-02-14

##Introduction

This document was put together to analyse spatial variation in the species and family richness of reptiles across ACT in response to mean annual precipitation. It involves loading packages, accessing and loading data for reptile occurrences from the Atlas of Living Australia (ALA) and spatial analysis tools for rasterising and summarising occurrences records as species richness per grid cell.

##Loading ###Load packages

First, start by loading some packages.

###Load climate data and map of ACT

Bring in climate data. We will use the `Worldclim 2.1` data, which we can obtain using the `geodata` package. We'll use a moderate resolution as this is just an example.

```
precipitation_data_monthly <- geodata::worldclim_country("Australia", "prec", "data", res = 2.5)
precipitation_data_annual <- terra::app(precipitation_data_monthly, fun = sum)
```

Bring in the map of the Australian Capital Territory (ACT). This will be the boundary polygon for the analysis. We start with the `States and Territories - 2021 - Shapefile` available from the ABS (<https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files>), and subset this spatial vector to just the ACT. Before that occurs, we need to define a coordinate reference system (CRS), this CRS will be used consistently throughout to ensure that all of our spatial information is lining up properly. We'll base the target CRS and resolution on the climate data.

```
target_CRS = crs(precipitation_data_annual)
target_res = res(precipitation_data_annual)
```

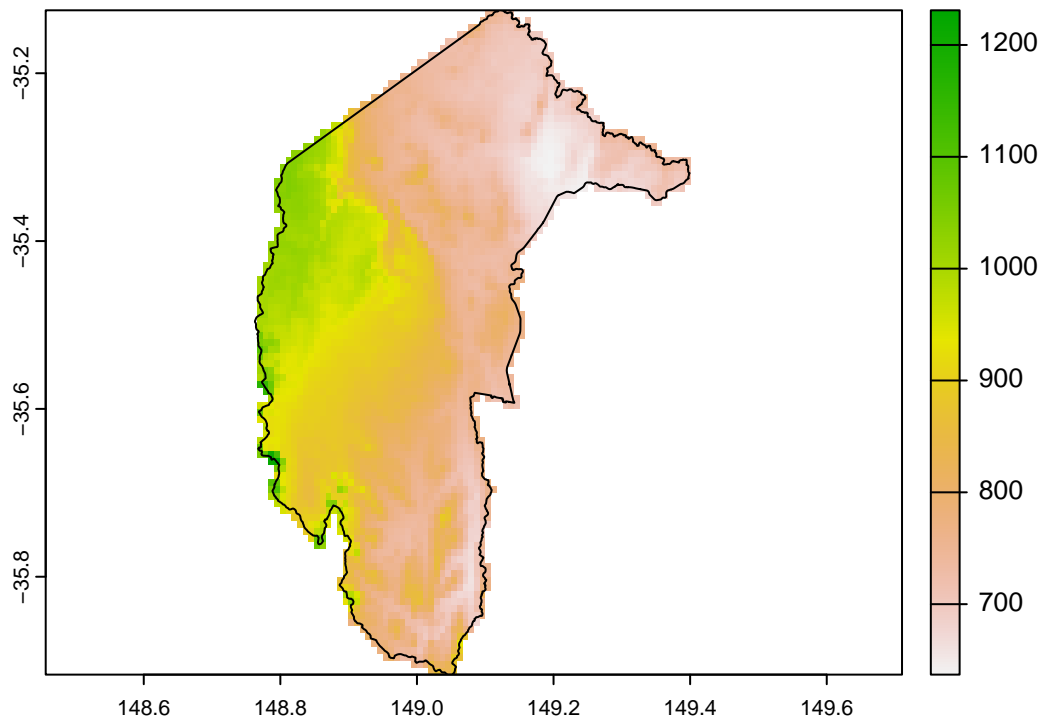
```
australia <- terra::vect("data/STE_2021_AUST_SHP_GDA2020/STE_2021_AUST_GDA2020.shp", crs=target_CRS)
ACT <- terra::subset(australia, australia$STE_NAME21 == "Australian Capital Territory")
```

Crop and mask the climate data to correspond to the boundary of ACT.

```
precipitation_data_annual_ACT <- terra::crop(precipitation_data_annual, ACT, mask = TRUE)
```

Let's just quickly inspect the climate data. In ACT, rainfall is greatest in the west with greater elevation, and declines towards the coast.

```
terra::plot(precipitation_data_annual_ACT)
terra::lines(ACT)
```



```
###Load occurrence data
```

Read in the occurrence data for reptiles. This data was obtained from ALA (<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>).

```
occurrence_data <- read_csv("data/occurrence_data/records-2023-02-14.csv")
```

```
## New names:
```

```
## * 'images' -> 'images...2'
## * 'locality' -> 'locality...87'
## * 'locality' -> 'locality...89'
## * 'verbatimLatitude' -> 'verbatimLatitude...113'
## * 'verbatimLatitude' -> 'verbatimLatitude...114'
## * 'verbatimLongitude' -> 'verbatimLongitude...116'
## * 'verbatimLongitude' -> 'verbatimLongitude...117'
## * 'verbatimCoordinateSystem' -> 'verbatimCoordinateSystem...119'
## * 'verbatimCoordinateSystem' -> 'verbatimCoordinateSystem...120'
## * 'images' -> 'images...218'
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
```

```
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 10520 Columns: 219
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr  (102): dataResourceUid, images...2, raw_recordedBy, dcterms:language, d...
```

```
## dbl  (19): individualCount, organismQuantity, year, month, sensitive_month,...
```

```
## lgl  (95): dcterms:accessRights, informationWithheld, reproductiveCondition...
```

```
## dtm (3): dcterms:modified, eventDate, georeferencedDate
##
## 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.
```

What variables do we have available?

```
names(occurrence_data) %>%
  sort()
```

```
## [1] "acceptedNameUsage"
## [2] "acceptedNameUsageID"
## [3] "associatedMedia"
## [4] "associatedOccurrences"
## [5] "associatedOrganisms"
## [6] "associatedReferences"
## [7] "associatedSequences"
## [8] "associatedTaxa"
## [9] "basisOfRecord"
## [10] "bed"
## [11] "behavior"
## [12] "catalogNumber"
## [13] "class"
## [14] "collectionCode"
## [15] "collectionID"
## [16] "continent"
## [17] "coordinatePrecision"
## [18] "coordinateUncertaintyInMeters"
## [19] "country"
## [20] "countryCode"
## [21] "county"
## [22] "dataGeneralizations"
## [23] "dataResourceUid"
## [24] "datasetID"
## [25] "datasetName"
## [26] "dateIdentified"
## [27] "day"
## [28] "dcterms:accessRights"
## [29] "dcterms:bibliographicCitation"
## [30] "dcterms:language"
## [31] "dcterms:license"
## [32] "dcterms:modified"
## [33] "dcterms:type"
## [34] "decimalLatitude"
## [35] "decimalLongitude"
## [36] "disposition"
## [37] "dynamicProperties"
## [38] "earliestAgeOrLowestStage"
## [39] "earliestEonOrLowestEonothem"
## [40] "earliestEpochOrLowestSeries"
## [41] "earliestEraOrLowestErathem"
## [42] "earliestPeriodOrLowestSystem"
## [43] "endDayOfYear"
```

```

## [44] "establishmentMeans"
## [45] "eventDate"
## [46] "eventID"
## [47] "eventRemarks"
## [48] "eventTime"
## [49] "family"
## [50] "fieldNotes"
## [51] "fieldNumber"
## [52] "footprintSpatialFit"
## [53] "footprintSRS"
## [54] "footprintWKT"
## [55] "formation"
## [56] "genus"
## [57] "geodeticDatum"
## [58] "geologicalContextID"
## [59] "georeferencedBy"
## [60] "georeferencedDate"
## [61] "georeferenceProtocol"
## [62] "georeferenceRemarks"
## [63] "georeferenceSources"
## [64] "georeferenceVerificationStatus"
## [65] "group"
## [66] "habitat"
## [67] "higherClassification"
## [68] "higherGeography"
## [69] "higherGeographyID"
## [70] "highestBiostratigraphicZone"
## [71] "identificationID"
## [72] "identificationQualifier"
## [73] "identificationReferences"
## [74] "identificationRemarks"
## [75] "identificationVerificationStatus"
## [76] "identifiedBy"
## [77] "identifierRole"
## [78] "images...2"
## [79] "images...218"
## [80] "individualCount"
## [81] "informationWithheld"
## [82] "infraspecificEpithet"
## [83] "institutionCode"
## [84] "institutionID"
## [85] "island"
## [86] "islandGroup"
## [87] "kingdom"
## [88] "latestAgeOrHighestStage"
## [89] "latestEonOrHighestEonothem"
## [90] "latestEpochOrHighestSeries"
## [91] "latestEraOrHighestErathem"
## [92] "latestPeriodOrHighestSystem"
## [93] "lifeStage"
## [94] "lithostratigraphicTerms"
## [95] "locality...87"
## [96] "locality...89"
## [97] "locationAccordingTo"

```

```

## [98] "locationID"
## [99] "locationRemarks"
## [100] "lowestBiostratigraphicZone"
## [101] "materialSampleID"
## [102] "maximumDepthInMeters"
## [103] "maximumDistanceAboveSurfaceInMeters"
## [104] "maximumElevationInMeters"
## [105] "measurementAccuracy"
## [106] "measurementDeterminedBy"
## [107] "measurementDeterminedDate"
## [108] "measurementID"
## [109] "measurementMethod"
## [110] "measurementRemarks"
## [111] "measurementType"
## [112] "measurementUnit"
## [113] "measurementValue"
## [114] "member"
## [115] "minimumDepthInMeters"
## [116] "minimumDistanceAboveSurfaceInMeters"
## [117] "minimumElevationInMeters"
## [118] "month"
## [119] "municipality"
## [120] "nameAccordingTo"
## [121] "nameAccordingToID"
## [122] "namePublishedIn"
## [123] "namePublishedInID"
## [124] "namePublishedInYear"
## [125] "nomenclaturalCode"
## [126] "nomenclaturalStatus"
## [127] "occurrenceID"
## [128] "occurrenceRemarks"
## [129] "occurrenceStatus"
## [130] "order"
## [131] "organismID"
## [132] "organismName"
## [133] "organismQuantity"
## [134] "organismQuantityType"
## [135] "organismRemarks"
## [136] "organismScope"
## [137] "originalNameUsage"
## [138] "originalNameUsageID"
## [139] "otherCatalogNumbers"
## [140] "ownerInstitutionCode"
## [141] "parentEventID"
## [142] "parentNameUsage"
## [143] "parentNameUsageID"
## [144] "phylum"
## [145] "pointRadiusSpatialFit"
## [146] "preparations"
## [147] "previousIdentifications"
## [148] "provenance"
## [149] "raw_recordedBy"
## [150] "recordedBy"
## [151] "recordID"

```

```

## [152] "recordNumber"
## [153] "references"
## [154] "relatedResourceID"
## [155] "relationshipAccordingTo"
## [156] "relationshipEstablishedDate"
## [157] "relationshipOfResource"
## [158] "relationshipRemarks"
## [159] "reproductiveCondition"
## [160] "resourceID"
## [161] "resourceRelationshipID"
## [162] "rights"
## [163] "rightsHolder"
## [164] "sampleSizeUnit"
## [165] "sampleSizeValue"
## [166] "samplingEffort"
## [167] "samplingProtocol"
## [168] "scientificName"
## [169] "scientificNameAuthorship"
## [170] "scientificNameID"
## [171] "sensitive_day"
## [172] "sensitive_decimalLatitude"
## [173] "sensitive_decimalLongitude"
## [174] "sensitive_eventDate"
## [175] "sensitive_eventID"
## [176] "sensitive_eventTime"
## [177] "sensitive_footprintWKT"
## [178] "sensitive_locality"
## [179] "sensitive_locationRemarks"
## [180] "sensitive_month"
## [181] "sensitive_verbatimCoordinates"
## [182] "sensitive_verbatimEventDate"
## [183] "sensitive_verbatimLatitude"
## [184] "sensitive_verbatimLocality"
## [185] "sensitive_verbatimLongitude"
## [186] "sex"
## [187] "sounds"
## [188] "source"
## [189] "species"
## [190] "specificEpithet"
## [191] "startDayOfYear"
## [192] "stateProvince"
## [193] "subgenus"
## [194] "taxonConceptID"
## [195] "taxonID"
## [196] "taxonomicStatus"
## [197] "taxonRank"
## [198] "taxonRemarks"
## [199] "typeStatus"
## [200] "verbatimBasisOfRecord"
## [201] "verbatimContinent"
## [202] "verbatimCoordinates"
## [203] "verbatimCoordinateSystem...119"
## [204] "verbatimCoordinateSystem...120"
## [205] "verbatimDepth"

```

```
## [206] "verbatimElevation"
## [207] "verbatimEventDate"
## [208] "verbatimLatitude...113"
## [209] "verbatimLatitude...114"
## [210] "verbatimLocality"
## [211] "verbatimLongitude...116"
## [212] "verbatimLongitude...117"
## [213] "verbatimScientificName"
## [214] "verbatimSRS"
## [215] "verbatimTaxonRank"
## [216] "vernacularName"
## [217] "videos"
## [218] "waterBody"
## [219] "year"
```

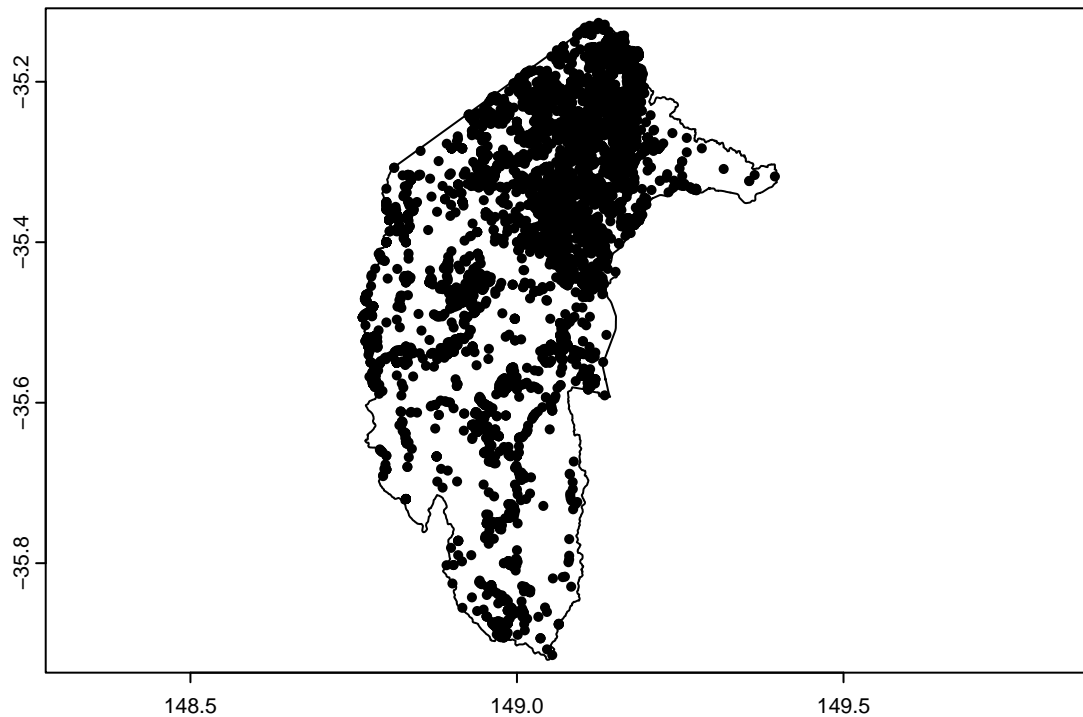
Have a look at the available data. Because this analysis is focussed on species richness throughout the entire period of observations, the only information that we will need is the taxon names, and some way to georeference these taxon names, so that we can obtain a list of the unique taxa occurring in a given location. Based on the `headings.csv` file in `~data\`, we will extract `decimalLongitude`, `decimalLatitude` and `species`. Let's also grab `family` for a follow-up analysis.

```
occurrence_data %>%
  select(decimalLatitude, decimalLongitude, species, family) -> occurrence_data_species_family
```

Now, let's calculate species richness. The scale that we calculate species richness will correspond in this case with the resolution of the climate data. This makes sense if the goal of our analysis is to calculate how species richness varies with broad-scale climate predictors such as MAP, as in this case. To do this we, we first need to convert the occurrence data in to a `SpatVector` using the `terra` package. We can view the distribution of occurrences across the ACT. The greater number of observations are located in the north.

##Conversion and extraction ###Convert occurrence to SpatVector

```
occurrence_data_species_family_vect <- terra::vect(occurrence_data_species_family, geom = c("decimalLongitude", "decimalLatitude"))
terra::plot(ACT)
terra::points(occurrence_data_species_family_vect)
```



Extract values

Ok, with our spatial vector we can now extract the precipitation values associated with each observation. Observations which are nearby each other will have the same precipitation value due to the grid-based dataset.

```
terra::extract(precipitation_data_annual_ACT, occurrence_data_species_family_vect, cells = TRUE, ID = F)
```

We now initialise an empty raster layer, using the resolution and crs of the climate data, and the extent of the ACT. We set all values to NA, as these will now hold richness estimates.

```
richness_raster<-terra::rast(resolution=target_res,
  ext(ACT), # extent
  crs=target_CRS, # CRS used by leaflet
  vals=NA) # make it empty
```

Using `rasterize`, we group nearby observations into a single value per grid cell by some summarizing function. In this case, it is the number of distinct species or families, or the median value of precipitation (remember, there is only one precipitation value per grid cell) which we achieve using `n_distinct`.

```
species_richness_raster <- terra::rasterize(precipitation_per_observation, richness_raster, field = "sp",
```

```
family_richness_raster <- terra::rasterize(precipitation_per_observation, richness_raster, field = "fam",
```

```
precipitation_raster <- terra::rasterize(precipitation_per_observation, richness_raster, field = "sum",
```

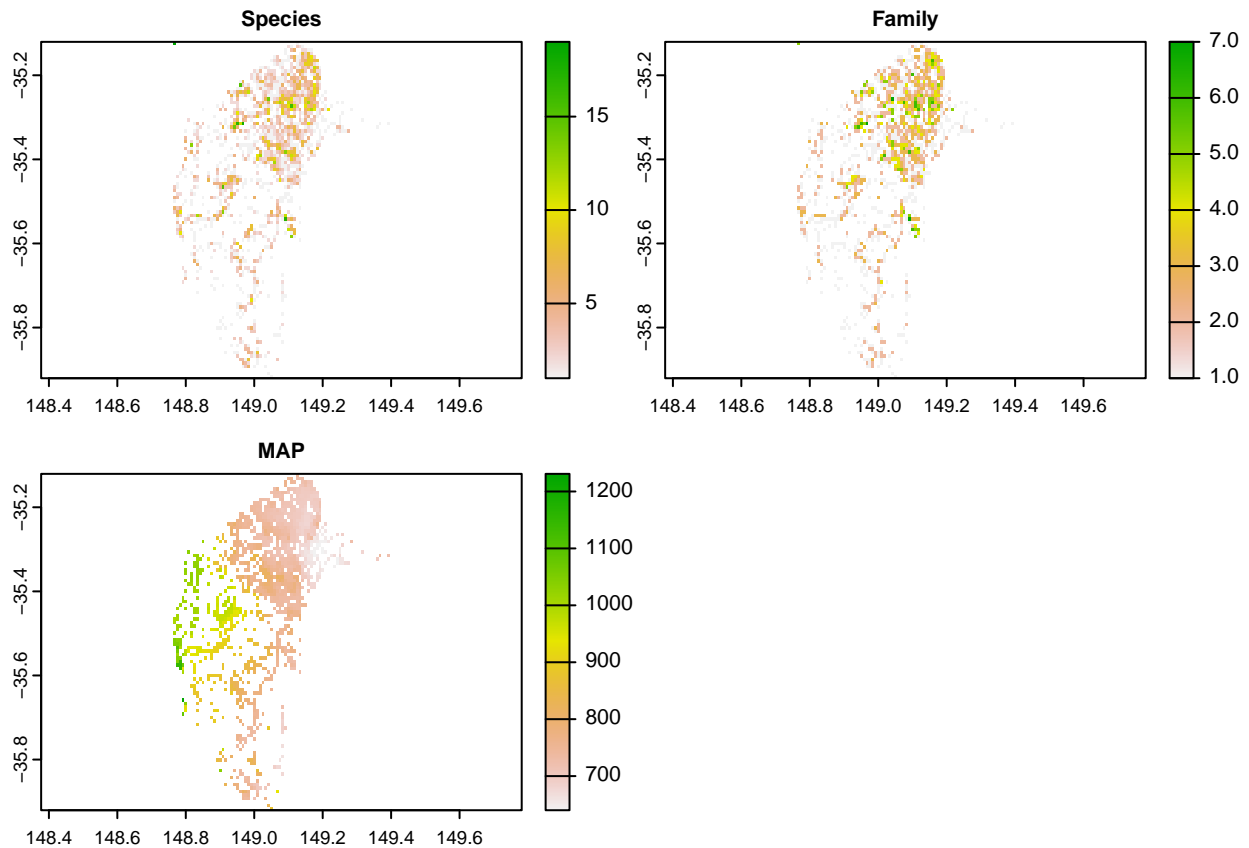
Stack the raster plots so that we can plot easily


```
terra::rast(list(Species = species_richness_raster, Family = family_richness_raster, MAP = precipitation_raster))
```

```
##Plotting of results
```

Ok, lets have a look at the plotted maps. At a pinch, it seems that richness is greater towards the north, where precipitation is lowest. Is there a relationship there?

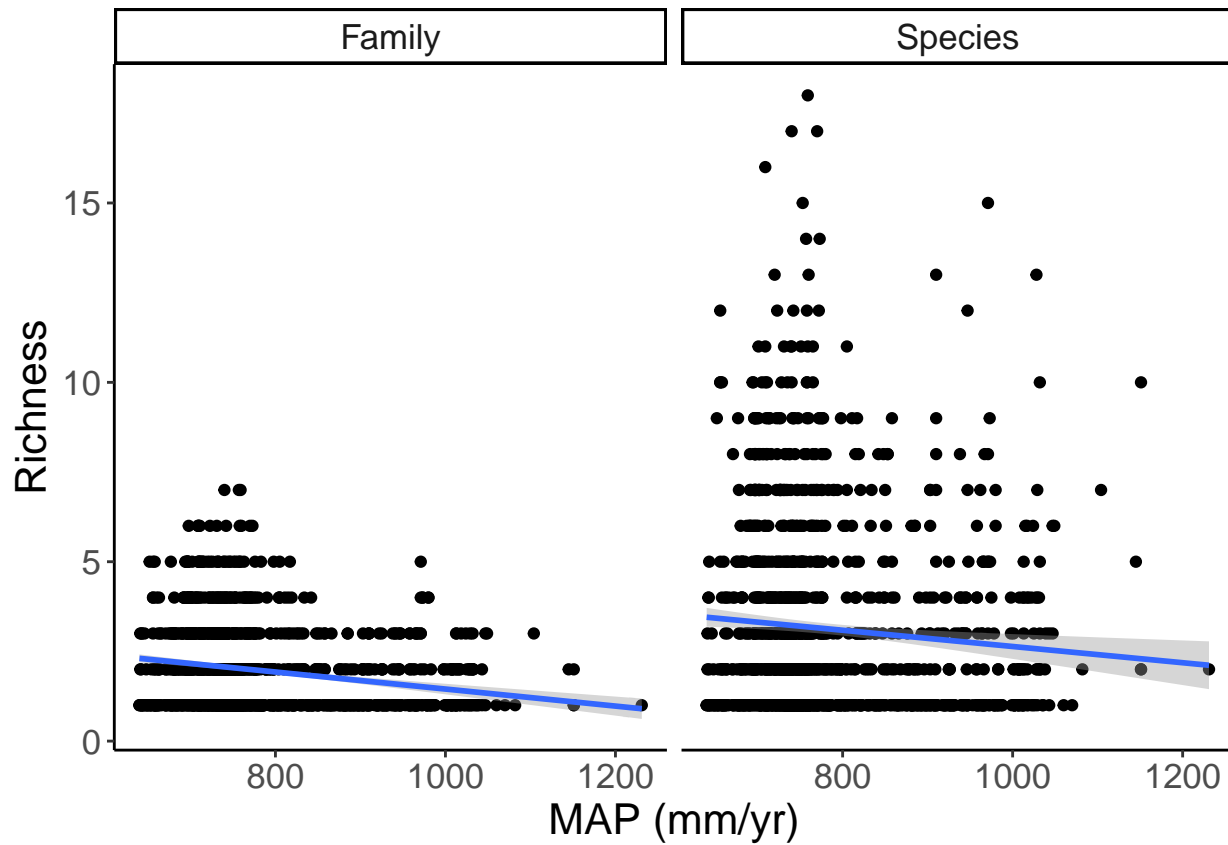
```
terra::plot(output_data)
```



We can be a bit more sure statistically by extracting the data for each grid cell, and plotting it using a scatter plot. A **very** simple analysis indicates that indeed this may be the case, with lower family and species richness at higher MAP values. This trend is roughly equivalent at both taxonomic scales, although of course the intercept value is much lower for family richness.

```
output_data %>%
  as_tibble() %>%
  drop_na() %>%
  pivot_longer(cols = c(Species, Family), names_to = "taxonomic_scale", values_to = "richness") %>%
  ggplot(aes(x = MAP, y = richness)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~taxonomic_scale) +
  theme_classic() +
  labs(x = "MAP (mm/yr)", y = "Richness") +
  theme(text = element_text(size = 16))
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



##Conclusion

There is a lot more that could be done to follow up on this analysis from both a mapping and statistical point of view. For example, the richness data is clearly poisson-distributed, and needs to be treated as such in follow-up work. In addition, we have disregarded cells where there were no observations as NA. This may be somewhat appropriate, because it is more likely that there have not been any observations made in this area, than it is that there are simply no reptiles. Along these lines, we also have a spatially autocorrelated residuals in this analysis, which could be dealt with using an appropriate mixed-effects model.