Environmental Variables & Data Exploration

Sydney Barfus
University of Florida

FLORIDA MUSEUM™

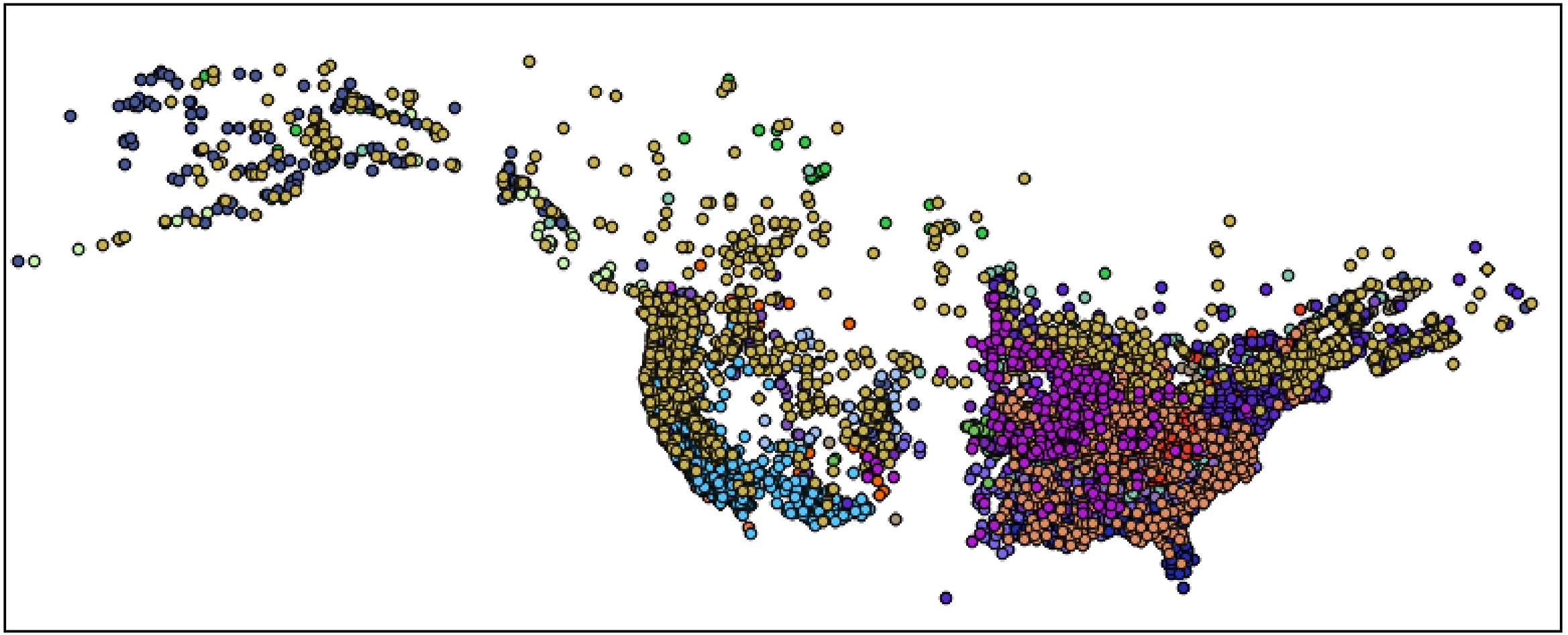
iDigBio
Integrated Digitized Biocollections

NSF

UF | UNIVERSITY of FLORIDA

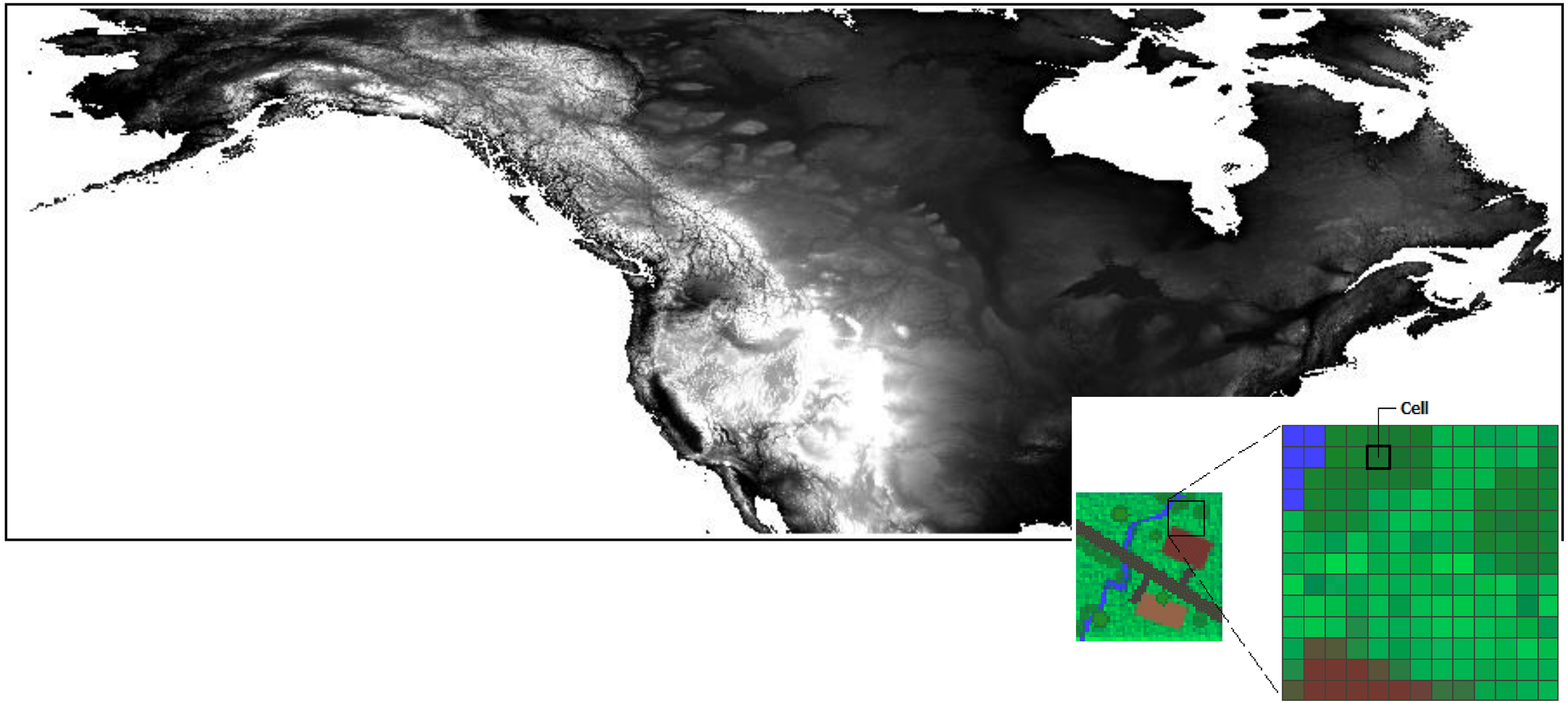
What data are we using?

1. Georeferenced occurrence records



What data are we using?

2. Gridded abiotic data layers (or Raster)



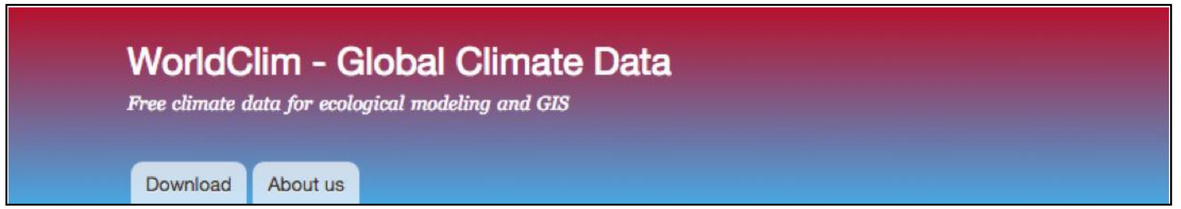
Source: ArcMap

Abiotic data layers

- **What?**
  - Temperature, precipitation, soil, vegetation, land use
- **Where?**
  - Local, USA, North America, Global
- **When?**
  - Past, current, future

Data layer sources

- WorldClim2
- PRISM
- SoilGrids
- Unified North American Soil Map
- USGS
- EPA
- NOAA
- AquaMaps
- PMIP
- And more!



rWBclimate
World

prism
World

weathercan
Canada

getCRUCLdata
World

clifro
New Zealand

chirps
World

WorldClim

www.worldclim.org

- *Global*
- *Past, Current and Future*

- WorldClim2 now available!
- Multiple resolution options

- **19 Bioclimatic variables** derived from monthly temperature and rainfall values

The data is available at the four spatial resolutions, between 30 seconds (~1 km2) to 10 minutes (~340 km2). Each download is a "zip" file containing 12 GeoTiff (.tif) files, one for each month of the year (January is 1; December is 12).

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C)	tmin 10m	tmin 5m	tmin 2.5m	tmin 30s
maximum temperature (°C)	tmax 10m	tmax 5m	tmax 2.5m	tmax 30s
average temperature (°C)	tavg 10m	tavg 5m	tavg 2.5m	tavg 30s
precipitation (mm)	prec 10m	prec 5m	prec 2.5m	prec 30s
solar radiation (kJ m^{-2} day^{-1})	srad 10m	srad 5m	srad 2.5m	srad 30s
wind speed (m s^{-1})	wind 10m	wind 5m	wind 2.5m	wind 30s
water vapor pressure (kPa)	vapr 10m	vapr 5m	vapr 2.5m	vapr 30s

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (×100)
BIO4 = Temperature Seasonality (standard deviation ×100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (BIO5-BIO6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter

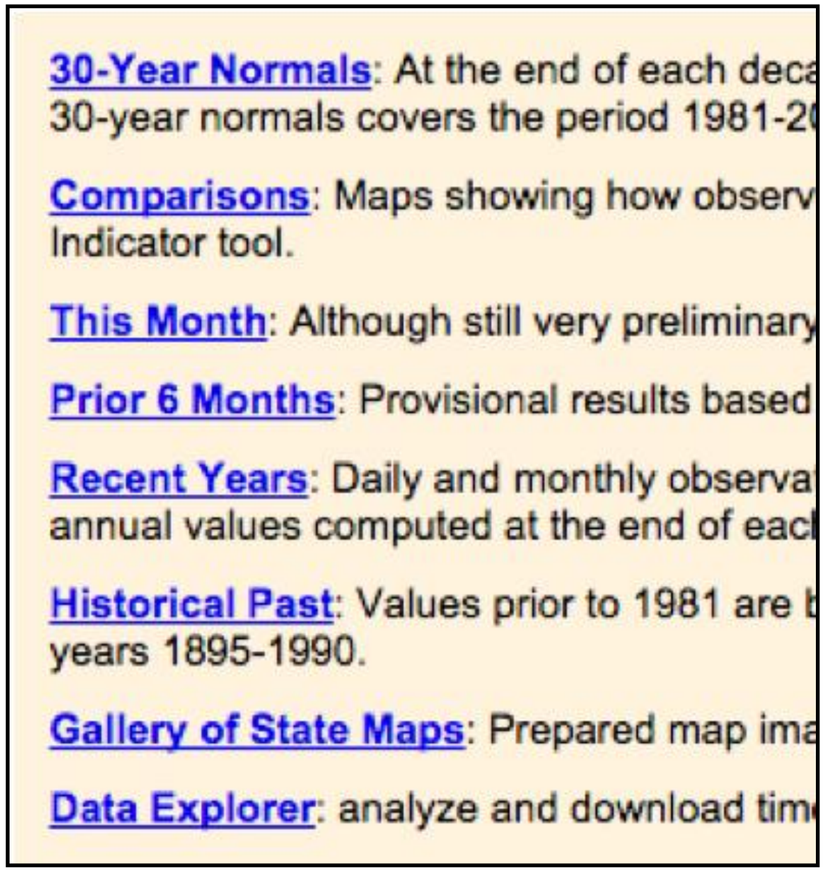
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

PRISM

www.prism.oregonstate.edu

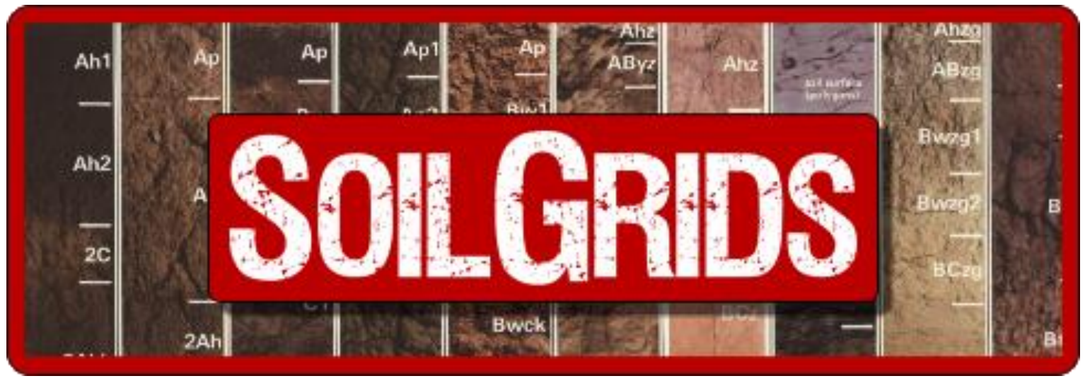
- *US Only*
- *Past and Current*

- More precise than WorldClim
- Data available from 1895 to present
- Lots of data and tools to explore



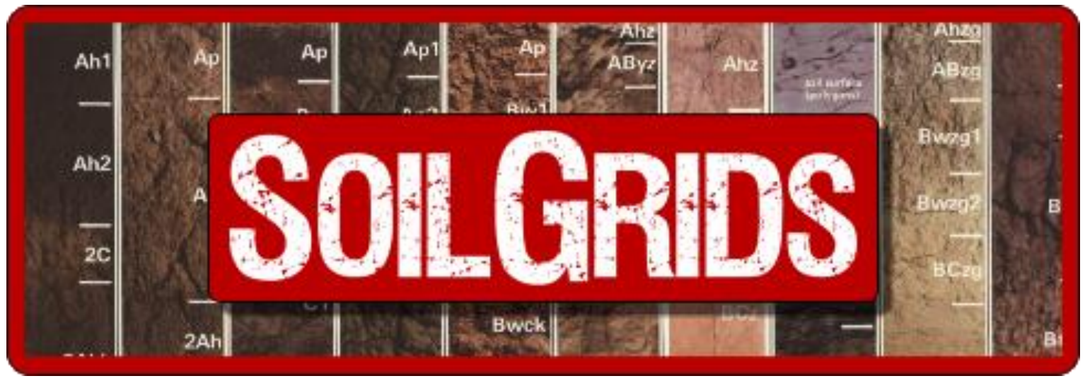
Soil Sources

- **SoilGrids2.0:**
  soilgrids.org
  - *Global*
  - Eight characteristics of topsoil and subsoil
  - 2.5 arc second resolution



Soil Sources

- **SoilGrids2.0:**
  soilgrids.org
  - *Global*
  - Eight characteristics of topsoil and subsoil
  - 2.5 arc second resolution



- **USGS**
  websoilsurvey.sc.egov.usda.gov
  - *US Only*
  - Need to define Area of Interest
  - Big files, Tough to get large areas

EPA

https://www.epa.gov/enviroatlas

- *US Only*
- Selected high-population regions at higher resolution
- Interactive tool and downloadable data

- Air quality, bodies of water, land use, boundaries, housing, and much more

Other Sources

Paleoclimate data

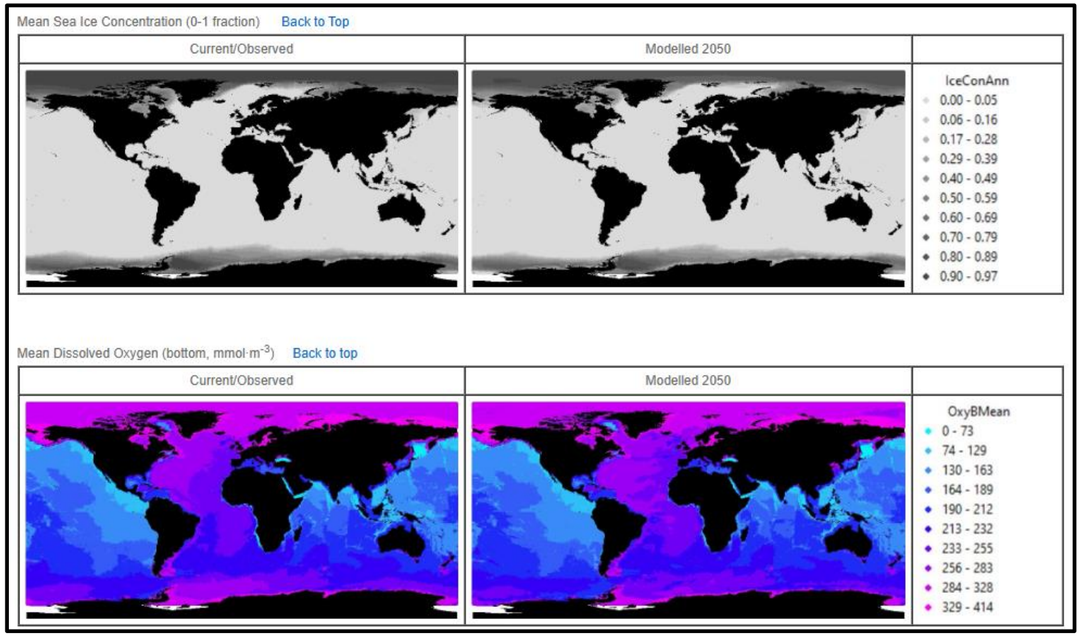
- NOAA/NCEI: ncei.noaa.gov/products/paleoclimatology
- PMIP: pmip.lsce.ipsl.fr
- PaleoClim: paleoclim.org

Aquatic Environments

- AquaMaps: aquamaps.org/main/envt_data.php

Hourly or Daily data:

- OGIMET: ogimet.com
- Daymet: daymet.ornl.gov



Source: AquaMaps

Now What?
Data Exploration

We have:

- Cleaned, georeferenced occurrences
- Environmental variables of interest

Now: examine the data to determine variables of interest, possible issues, and developing trends

★PCA

- **Principal Component Analysis**
- Transforms data into new set of variables, called **Principal Components (PC1, PC2, …)**

- **Loading:** how much each original variable contributes to a PC

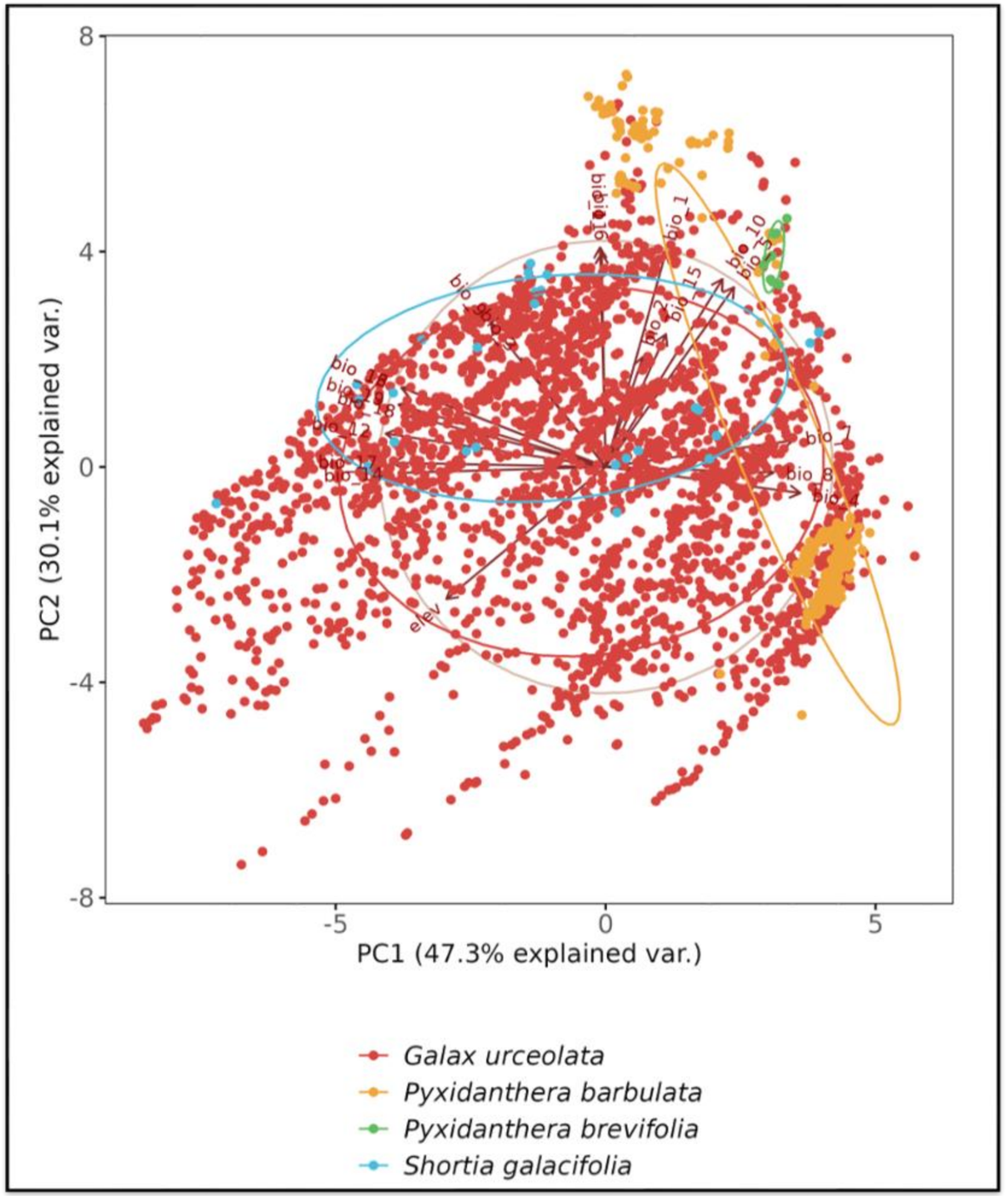
- Can calculate and examine the relative contributions of each climatic variable to each PC
  - Reveals top loading variables

PCA

- *Summarize* many environmental variables into PCs
- *Reveal groupings* that are not obvious in raw data

Plotting the PCA:

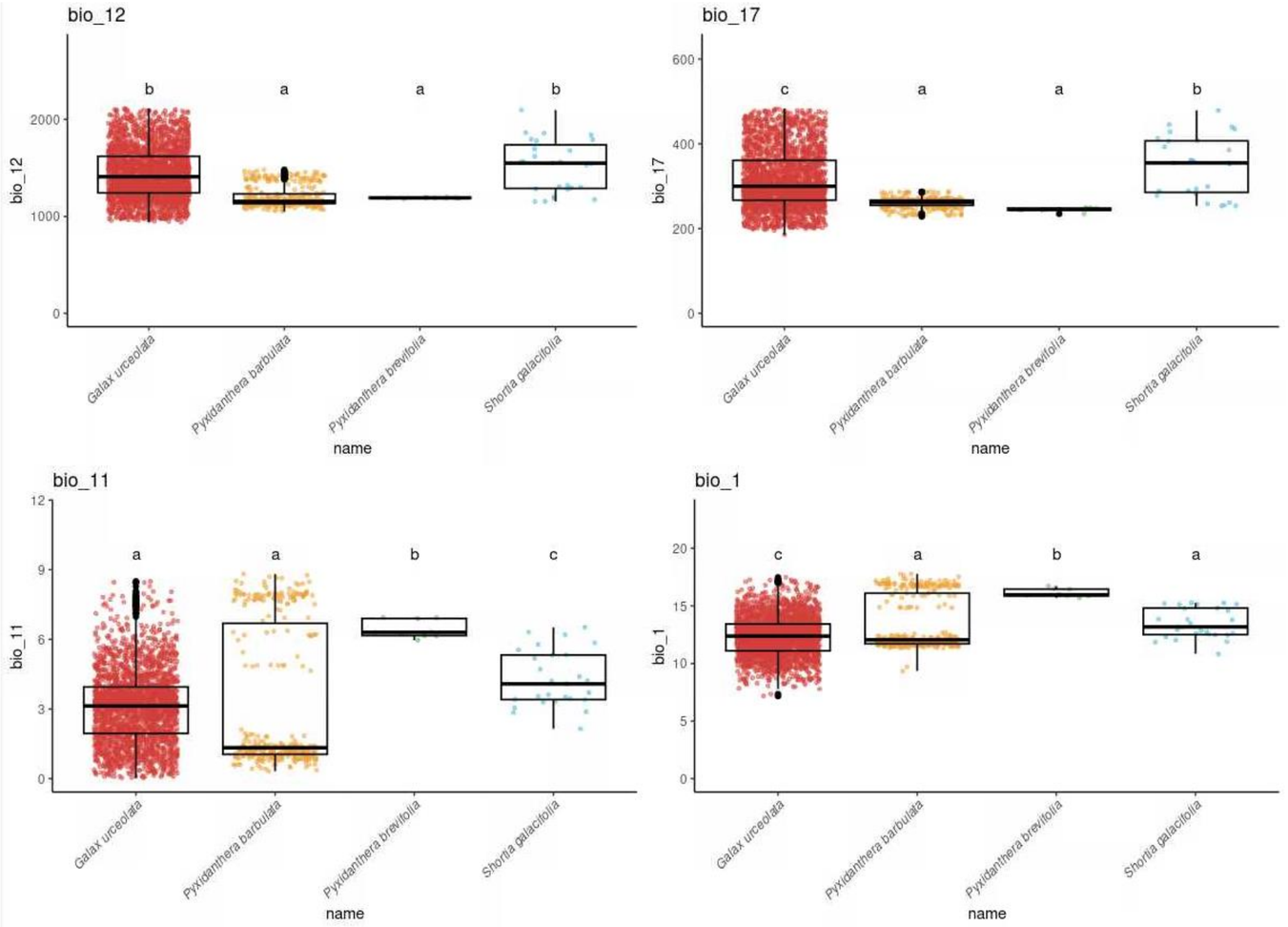
- Points: observations
- Arrows: environmental variables
- Ellipse Groupings: Species



★ANOVA

- Select the ***top two loading variables*** for the ***top two PCs***
- Shows if each species is significantly different

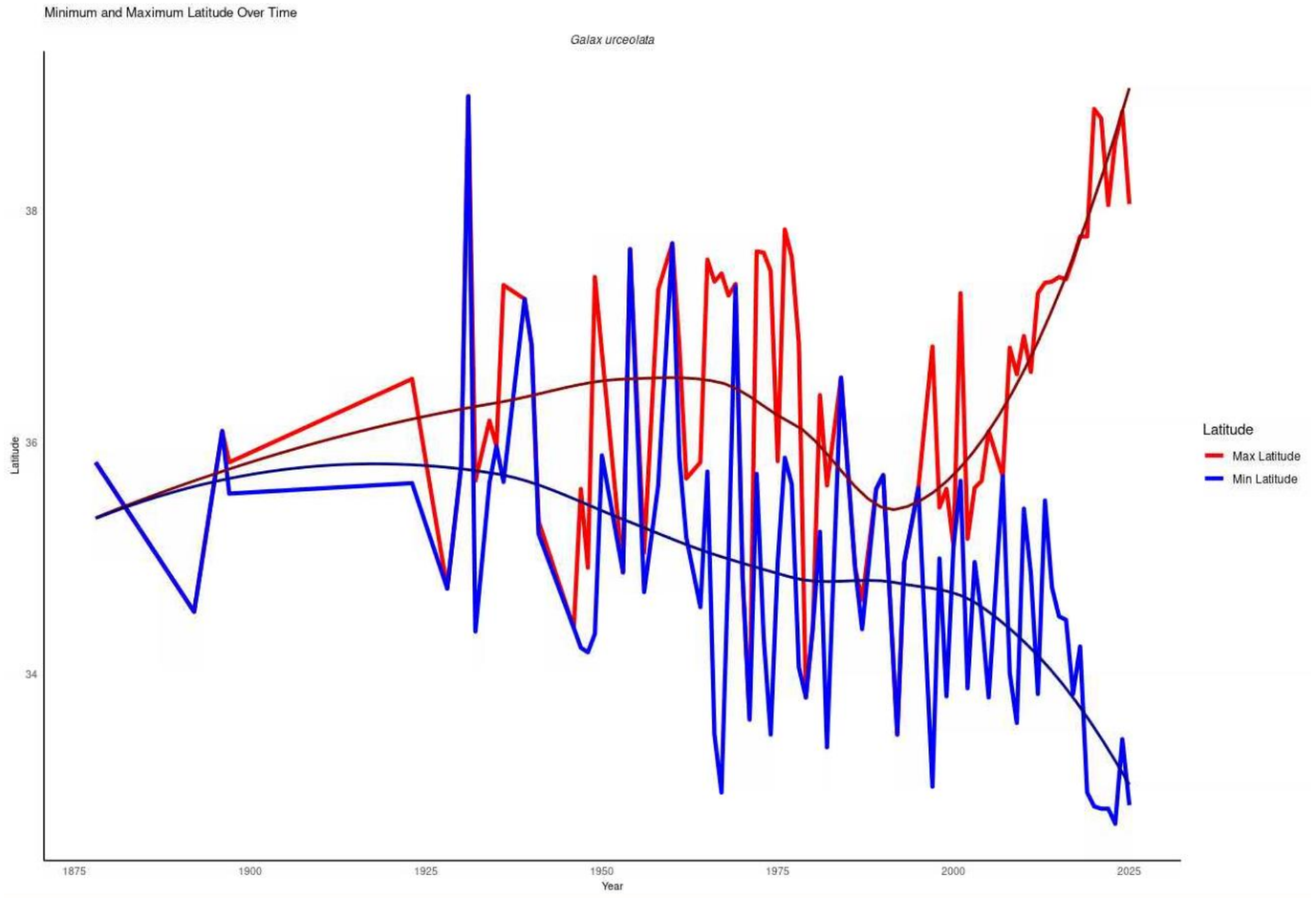
- Add **Tukey HSD test** to determine *which* species are significantly similar/different



Range Shifts Over Time

- Does not rely on environmental variables
- Uses dates of occurrence records to identify historical trends in distribution over time
- ***Latitude vs Time***

- Must consider potential impacts of increases in volume of occurrence records over time



Importance

Data Exploration can reveal:

- Potential issues, biases, and discrepancies in raw data
- Trends that are worth exploring further
- Population/species differences

Data exploration allows us to examine the reliability of data before putting it into the ENM models