

Ignorance scores

Where and when is data enough?

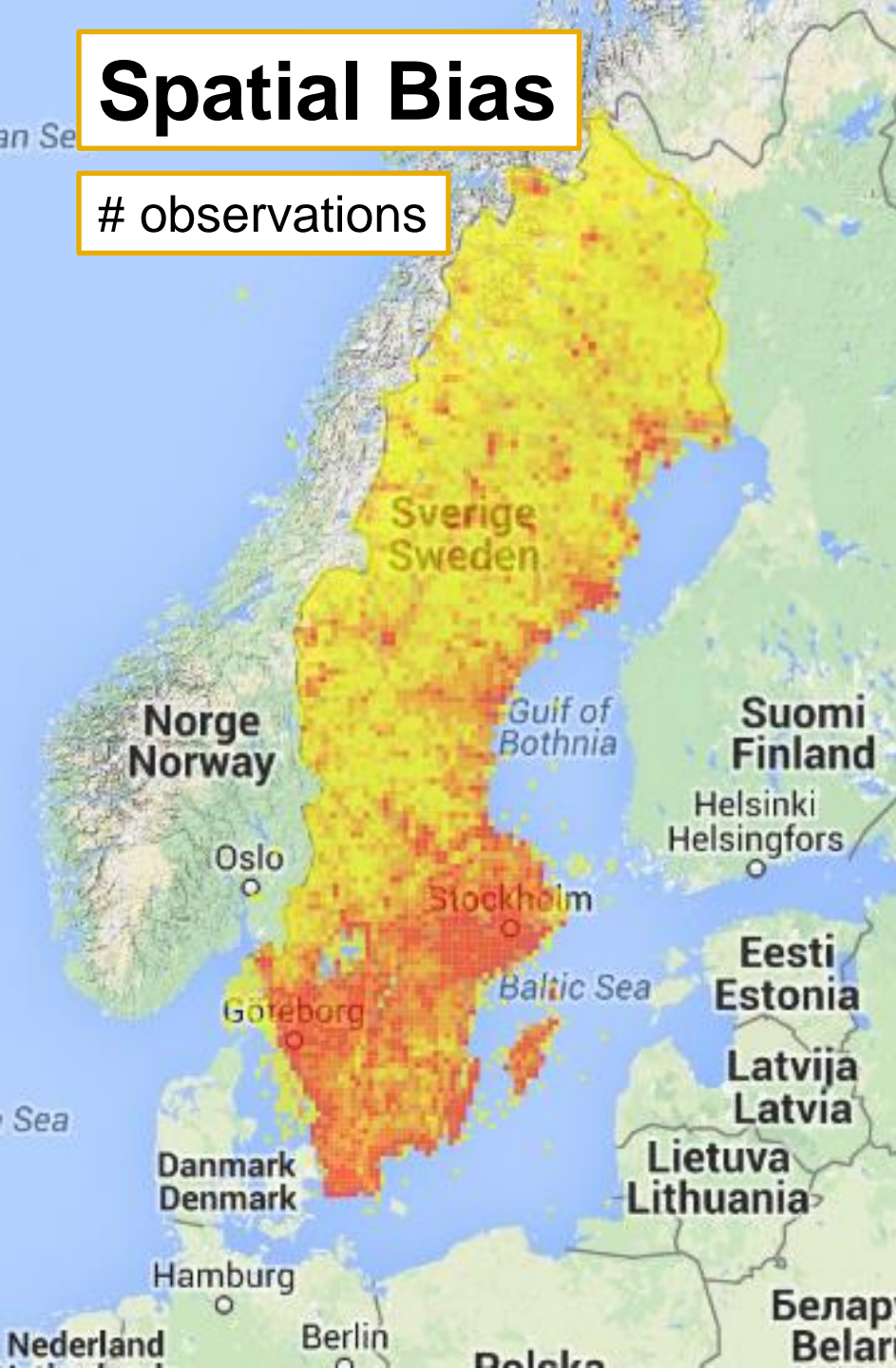
Alejandro **Ruete**, PhD
Alejandro@greensway.se

**“The greatest enemy of knowledge
is not IGNORANCE;
it is the illusion of knowledge”**

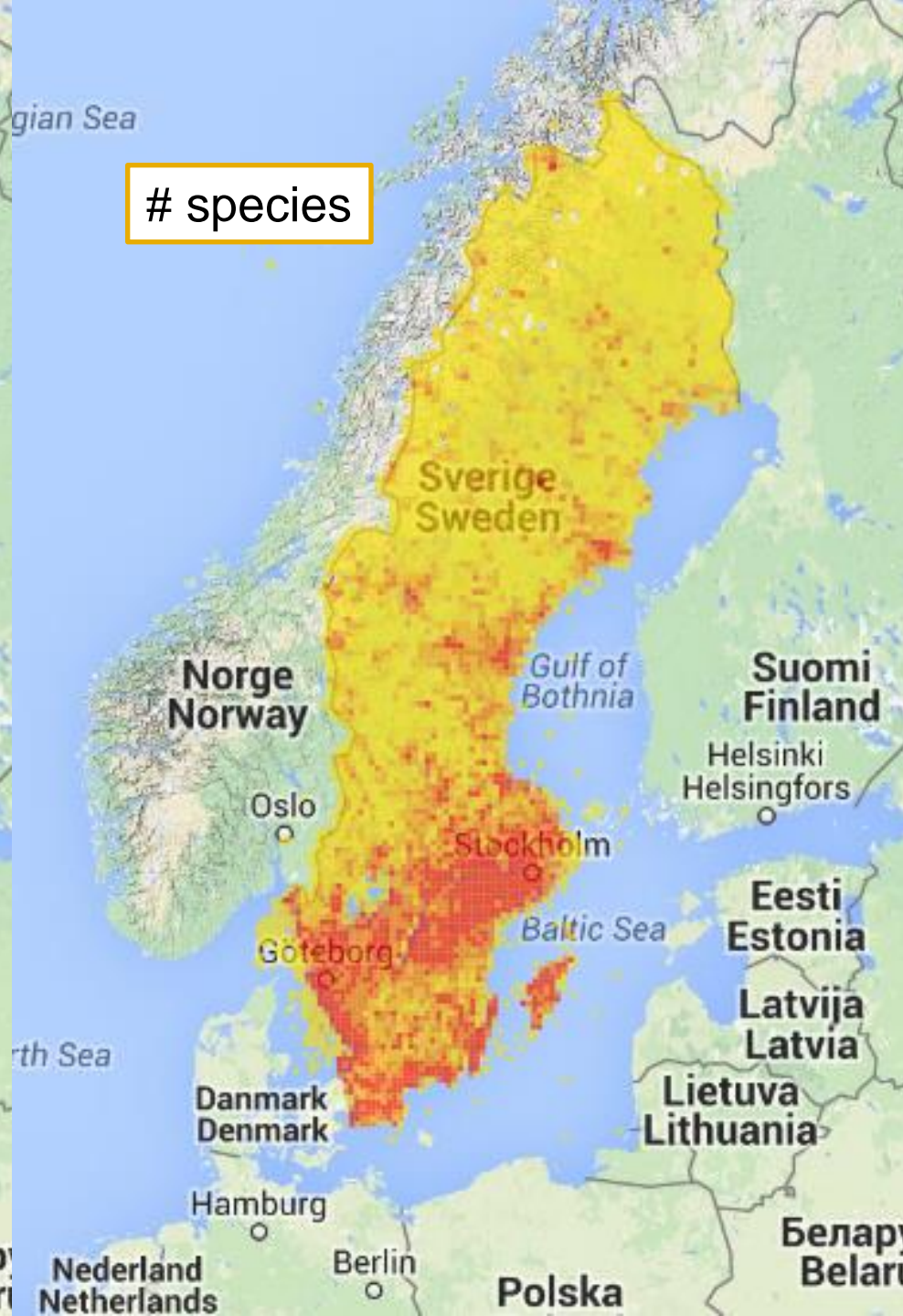
Daniel J. Boorstin

Spatial Bias

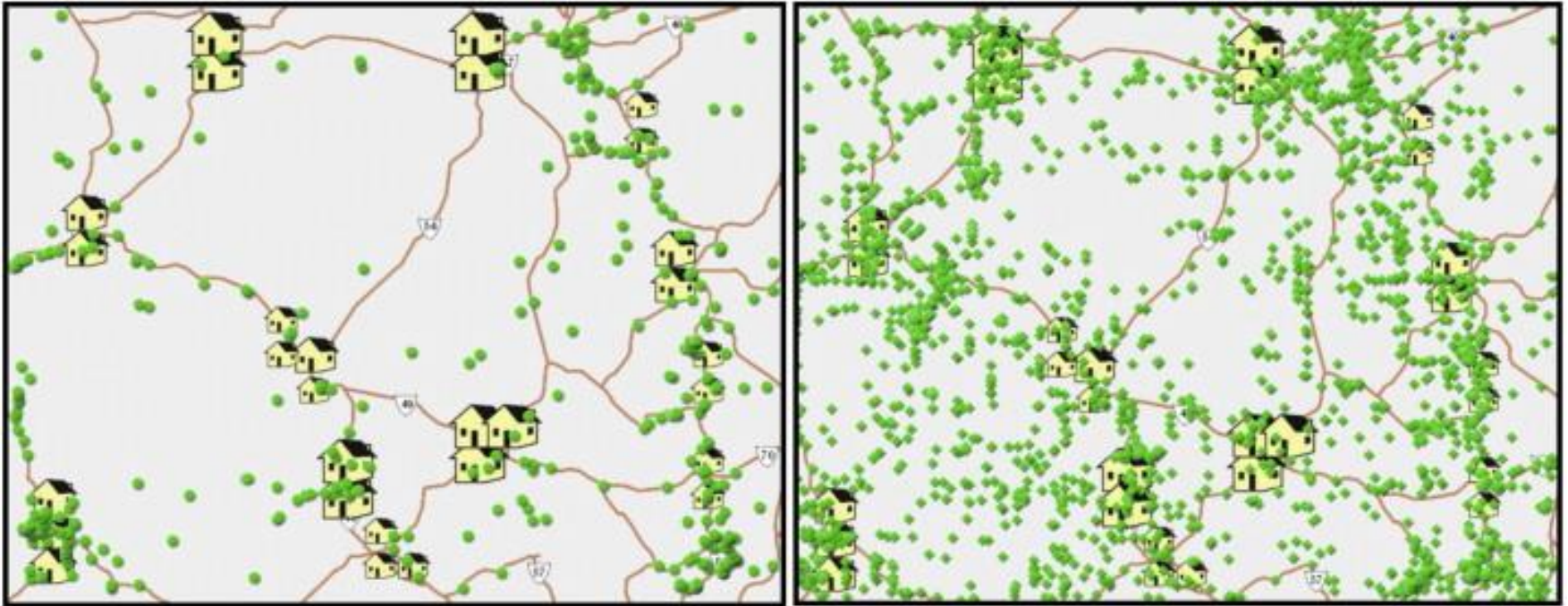
observations



species



Non-homogeneous sampling



Presence-only data for butterflies (left) and mammals (right), central Mexico.

Fernández & Nakamura. **Estimation of spatial sampling effort based on presence-only data and accessibility**

Ecological Modelling, Volume 299, 2015, 147–155. <http://dx.doi.org/10.1016/j.ecolmodel.2014.12.017>



Ignorance Maps

One solution (out of many)

alejandroruete.github.io/IgnoranceMaps

Biodiversity Data Journal 3: e5361 (2015)

Exploring ignorance in space and time

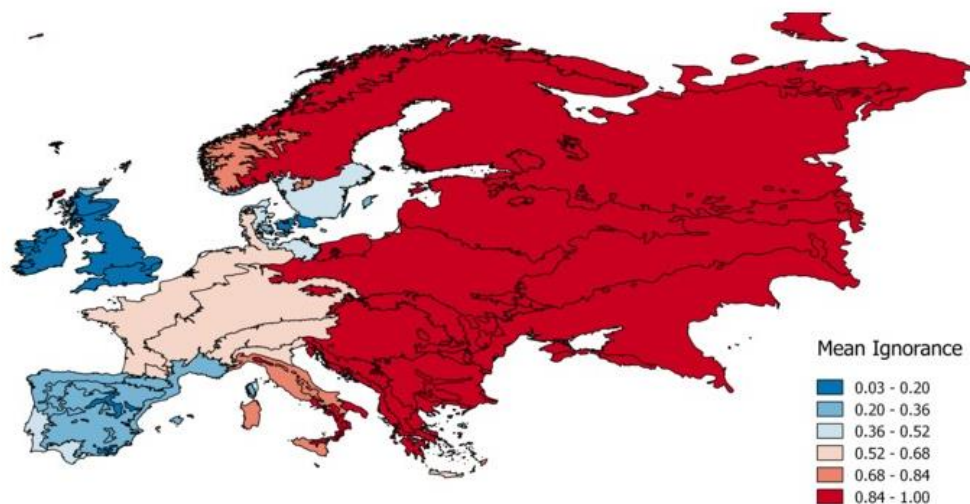
Where and when is data enough? A simple algorithm for fast implementations and comparable results.

♥ Liked | 6

💬 Comment

STORY

UPDATES



Ignorance of GBIF data on European Amphibians summarized per ecoregion

Edit project

SUBMITTED TO



2016 GBIF Ebbe Nielsen Challenge

WINNER First Prize

CREATED BY

Describe your contribution

E.g., I worked on the backend and cleaned up

Save

Cancel



Alejandro Ruete

Conservation biologist and Population ecologist. Analyst. Passionate photographer.

Approach

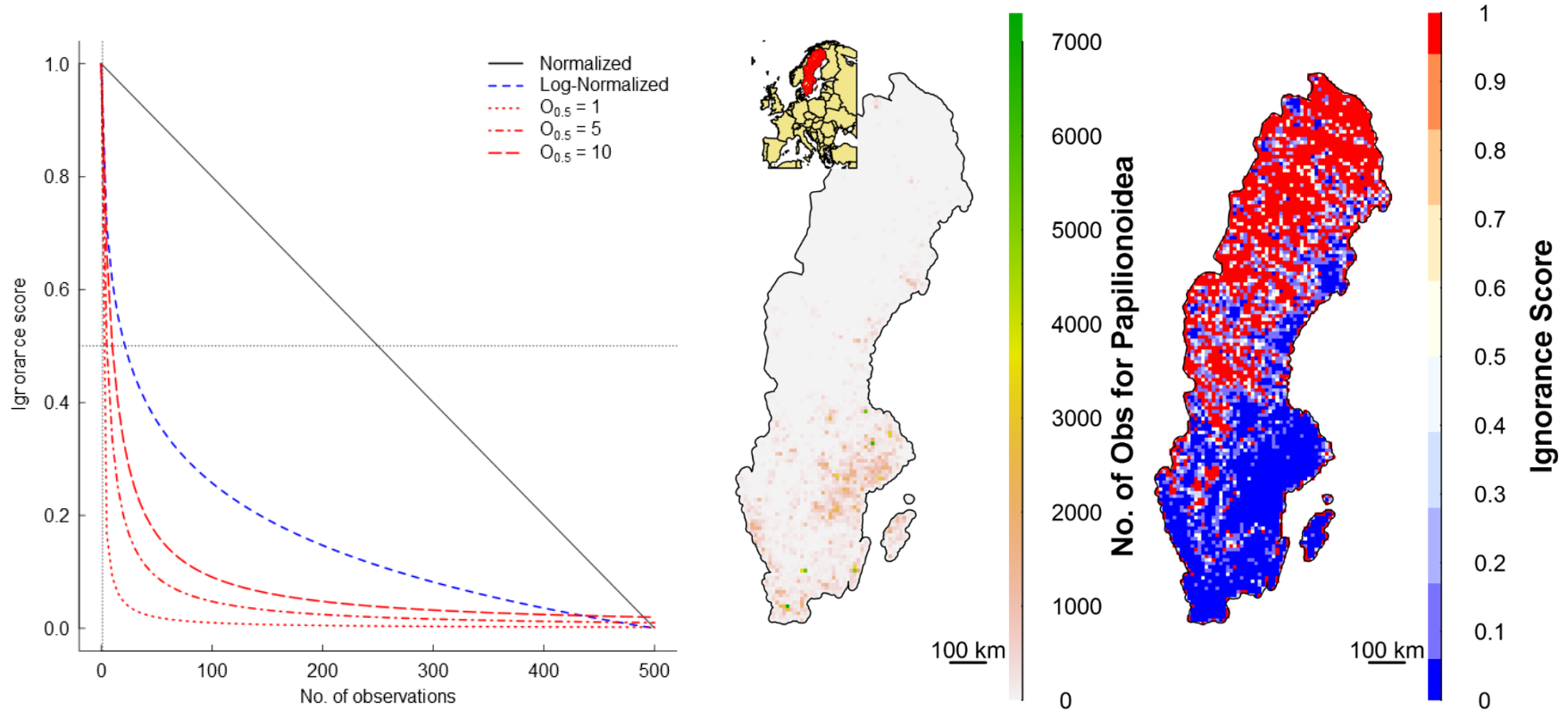
- Based “mainly” on presence-only data
- It is not a model, it is a TRANSFORMATION
- Report of the spatial distribution of sampling effort (or lack of it)
- Least possible number of assumptions
- Generality: inform users to analyse raw data
- Scalability
- Comparability
- Low computational requisites

A few assumptions are needed

- Observers are assumed to be fond of or specialist on one or more taxonomic groups (e.g. family, order) and follow roughly the same methods
- REFERENCE TAXONOMIC GROUPS will share similar bias, and are surrogate for sampling effort (Phillips et al. 2009, Ponder et al. 2001).
- The lack of reports of any species from the RTG at a particular location is likely due to a lack of observers, rather than to the total absence of species.

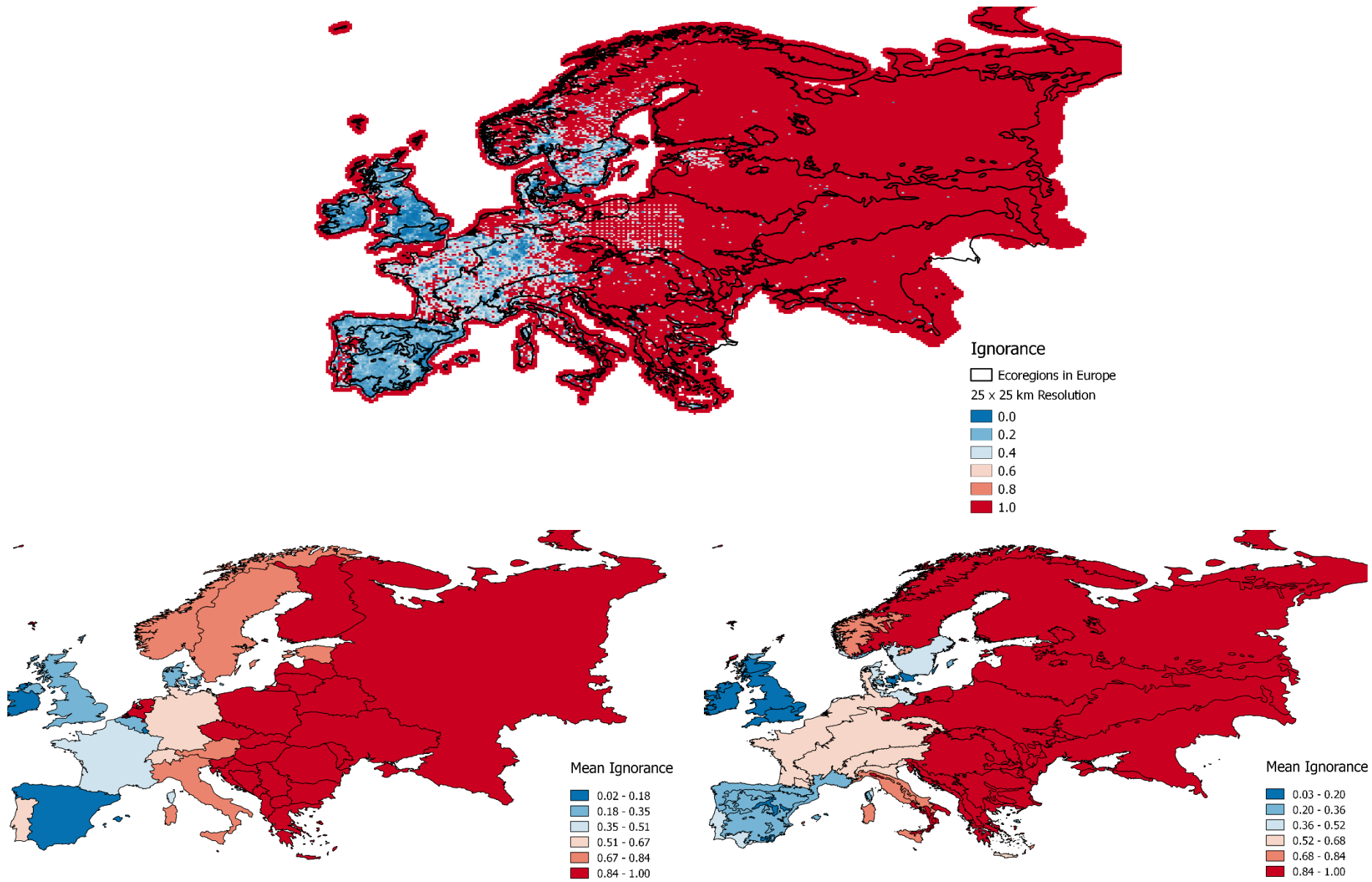
Algorithms

Transformation of the number of observations (N) per space-time unit into a scale of knowledge



Let's take a look

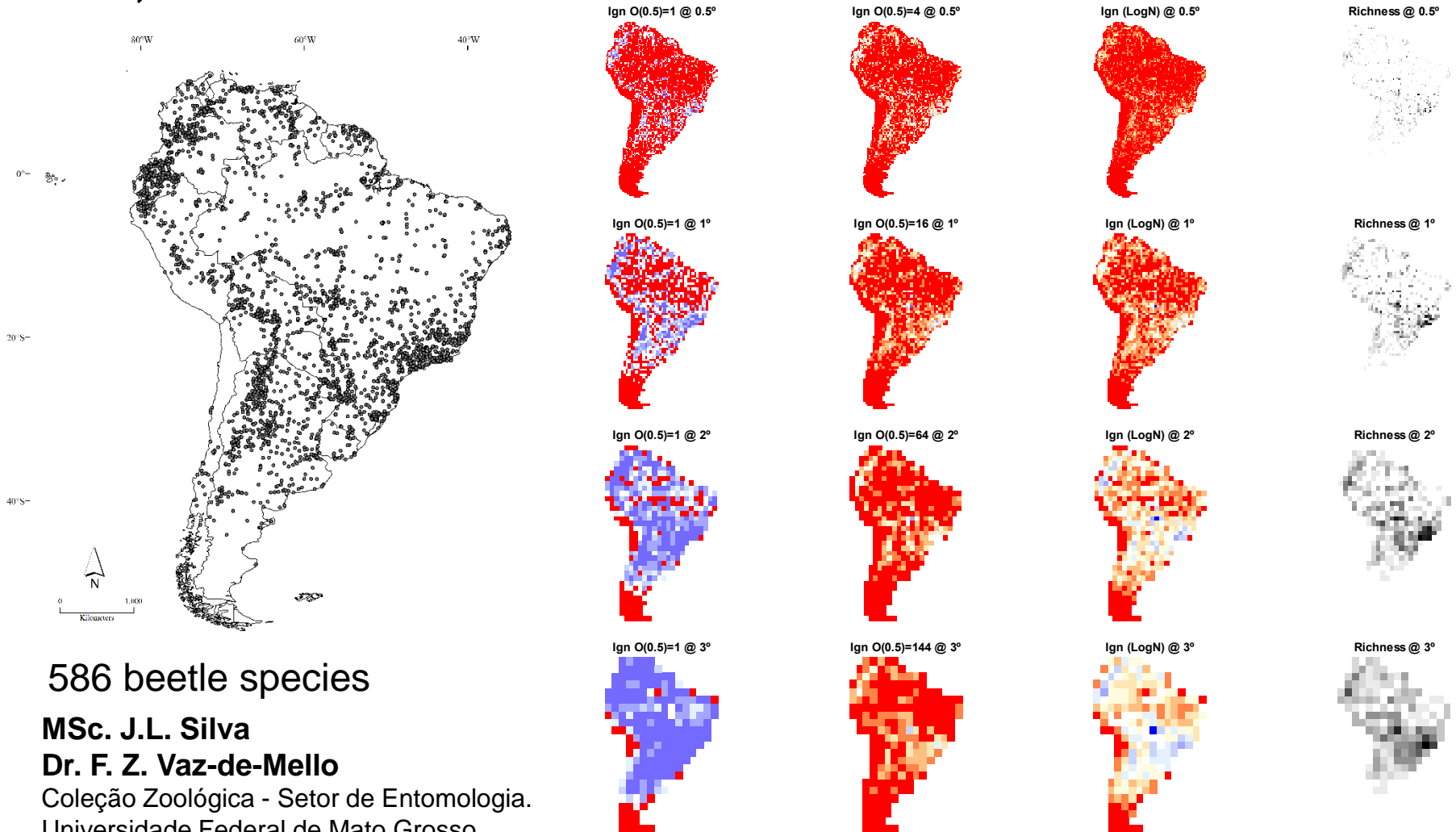
Let's take another look



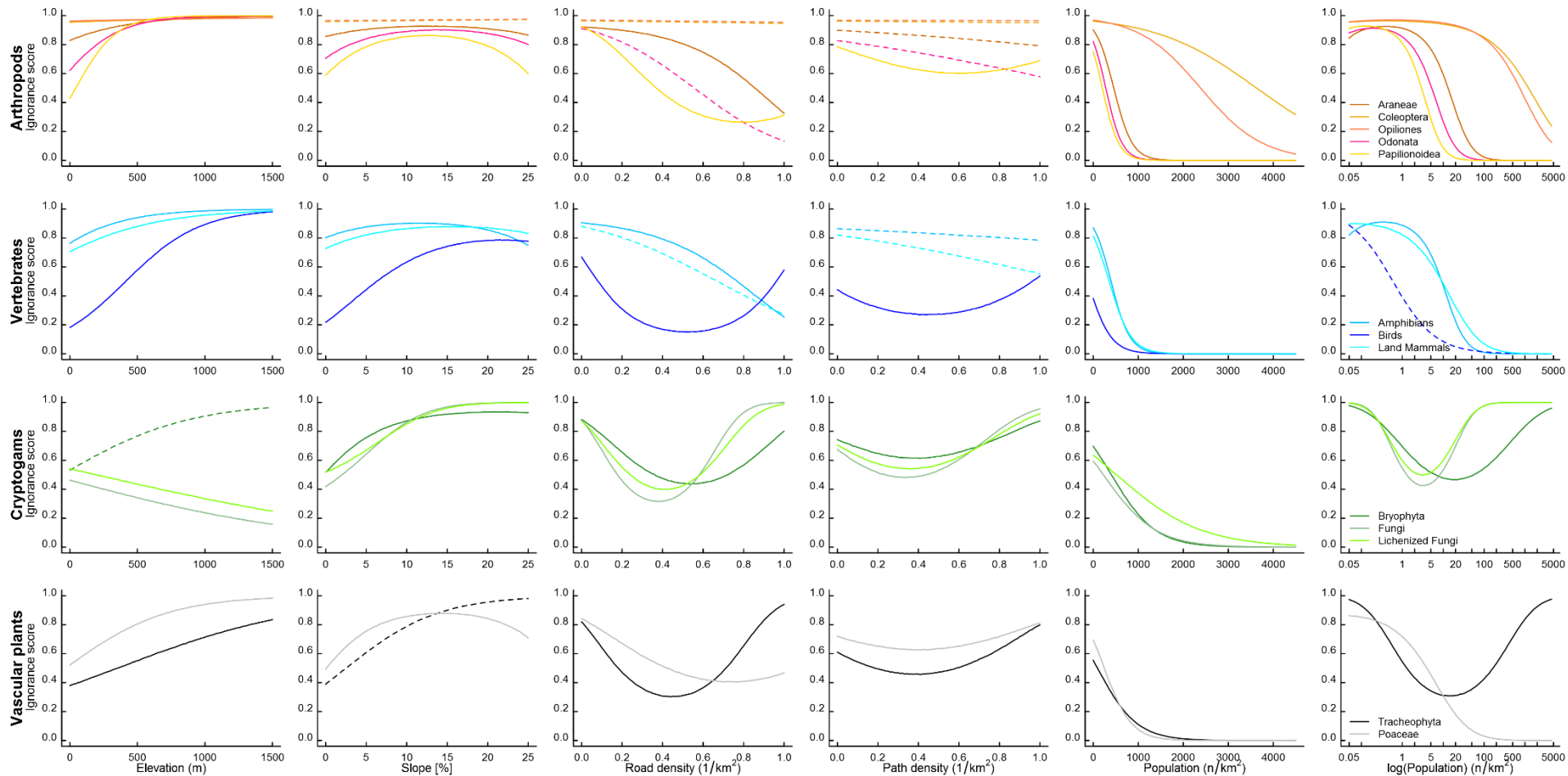
Different ways to summarize pixel values into polygons (e.g. countries vs ecorregions)

Scalability

($O_{0.5}$ can be scaled to match the raster resolution)

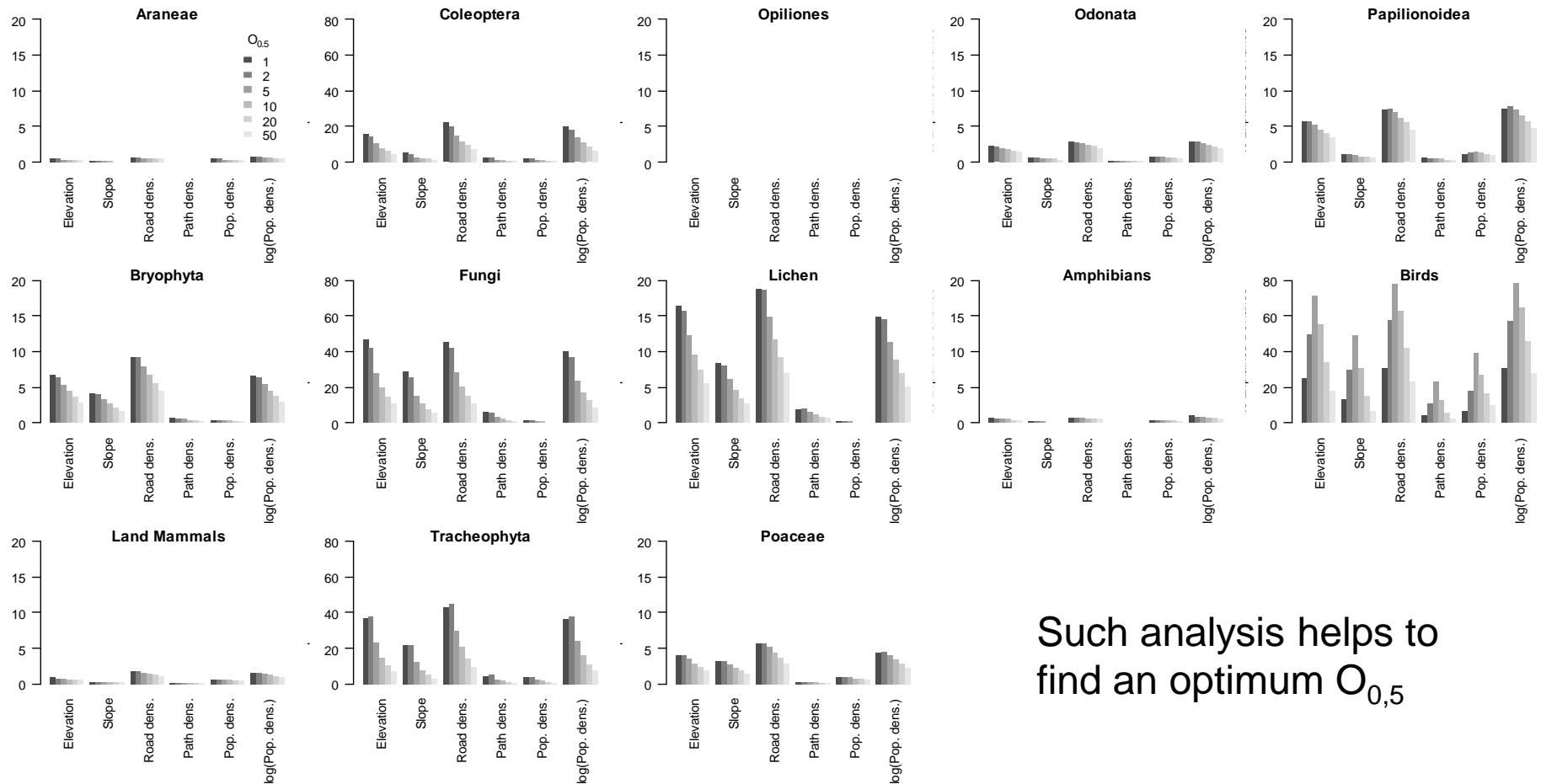


Where do observers go?



Mair & Ruete. 2016. **Explaining spatial variation in the recording effort of citizen science data across multiple taxa** PLoS ONE 11(1): e0147796.

Explained Deviance (%) per RTG per variable



Such analysis helps to find an optimum $O_{0.5}$

Applications

1. Consultants performing environmental impact assessments (e.g. ignorance maps as precautionary statements)
2. Observers (e.g. interested in under-sampled locations)
3. Researchers (this is the juicy part!)

Applications: Species distributions

- generate pseudo-absences
- mask out areas of high uncertainty from other raster layers derived from the raw data
- accurate assessment of species richness (ongoing work)
- ignorance maps as confidence or "All-in-one" bias layers for background sampling (e.g. MaxEnt)

**Knowing what you ignore
you already know a lot**

Thank you for your attention

Questions?

