AI Naturalists in the digital undergrowth

Tom August – UKCEH 14th October 2020

UK Centre for Ecology & Hydrology

Natural Historians

Naturalists have been recording wildlife for 100's of years

Millions of data points are generated every year

Data is used to support key studies into biodiversity loss

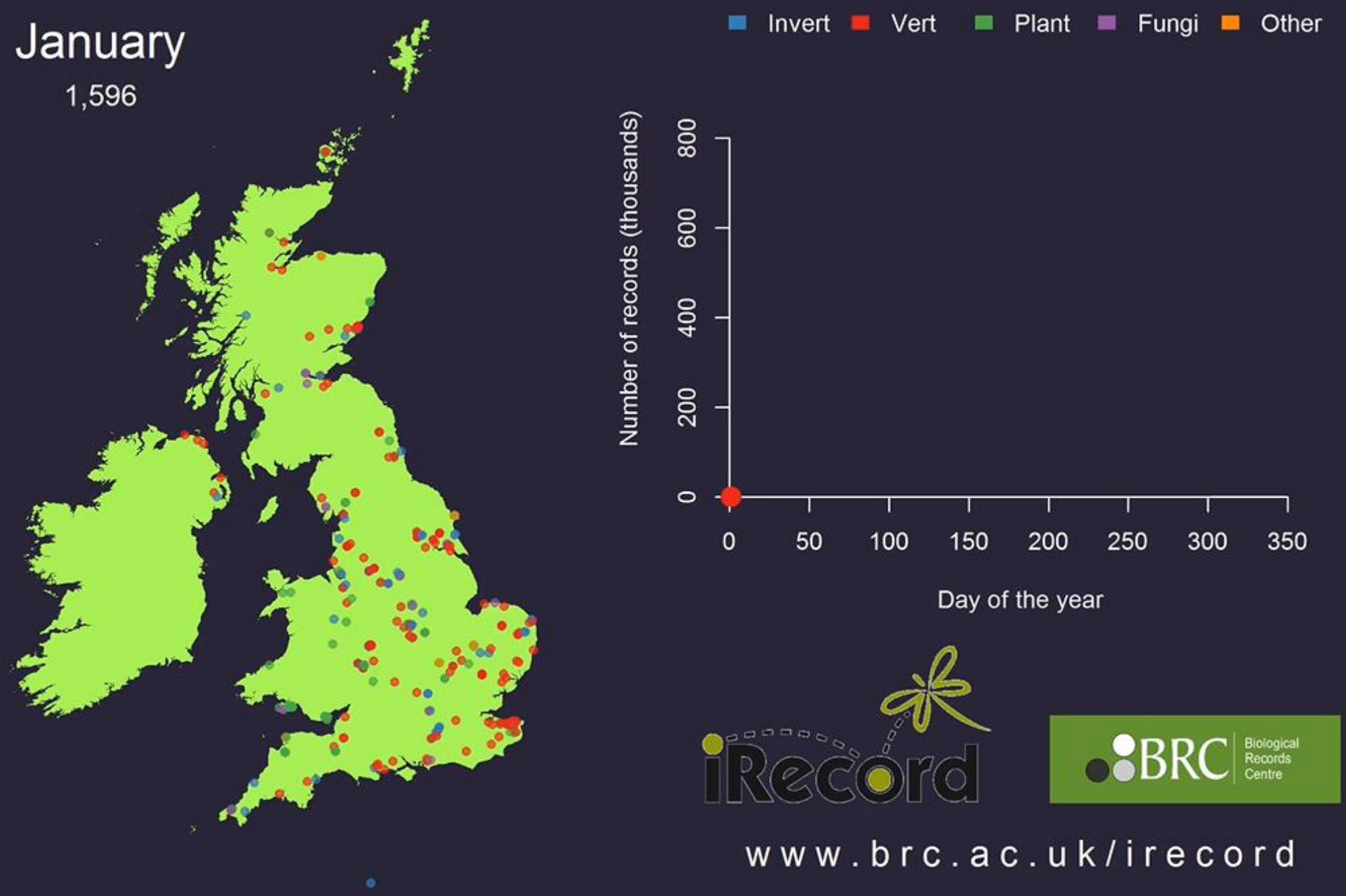
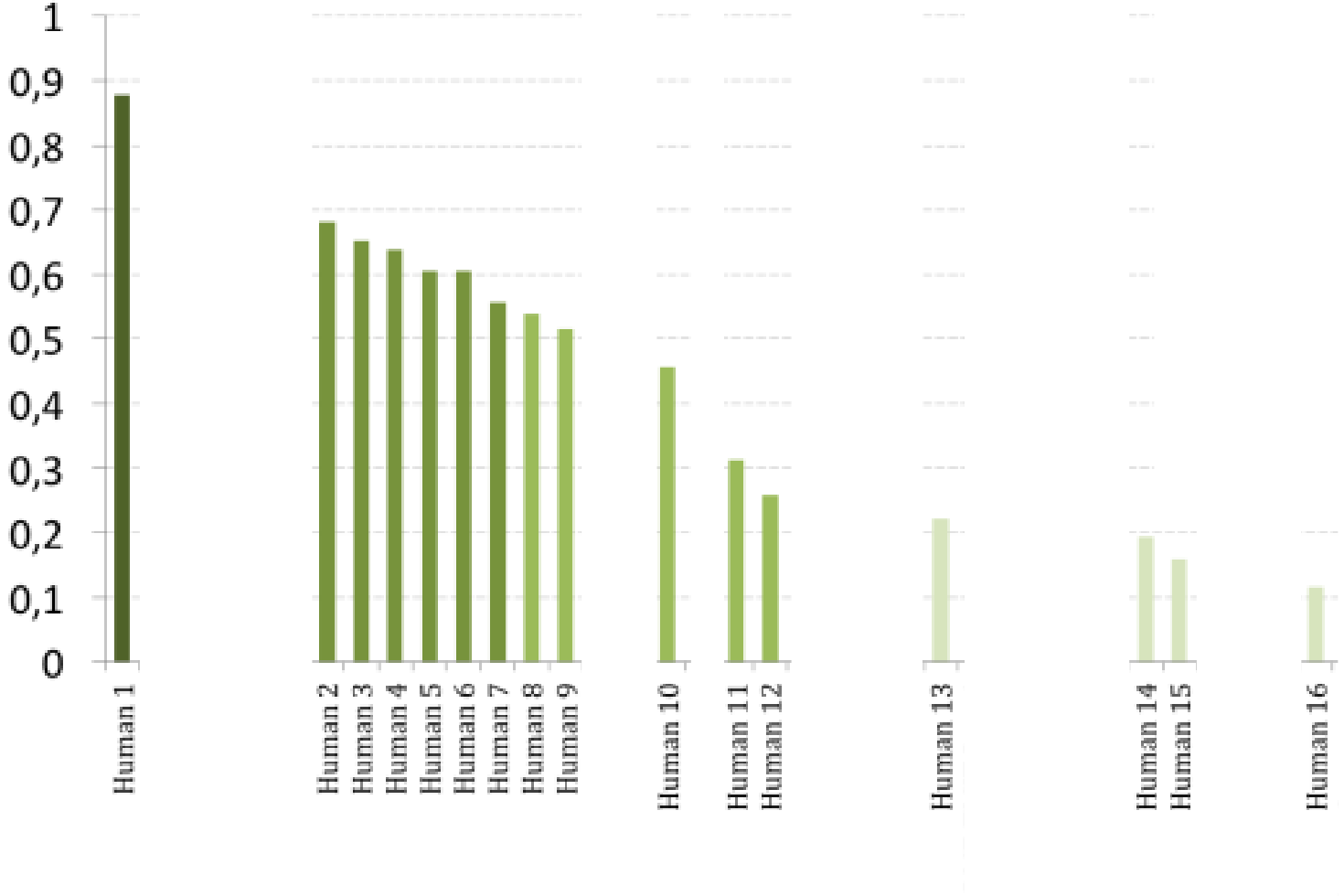

January
1,596
Invert
Vert
Plant
Fungi
Other
Number of records (thousands)
0
200
400
600
800
0
50
100
150
200
250
300
350
Day of the year
iRecord
BRC
Biological
Records
Centre
www.brc.ac.uk/irecord

Image classifiers can match the accuracy of experts

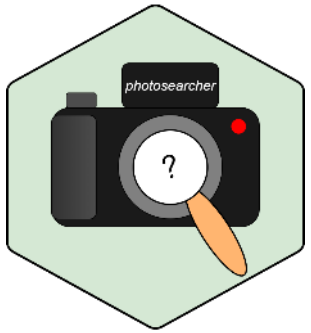


Pierre Bonnet et al. 2018. Plant Identification: Experts vs. Machines in the Era of Deep Learning: Deep learning techniques challenge flora experts.


flickr

The idea
An AI naturalist workflow

flickr → Pl@ntNet

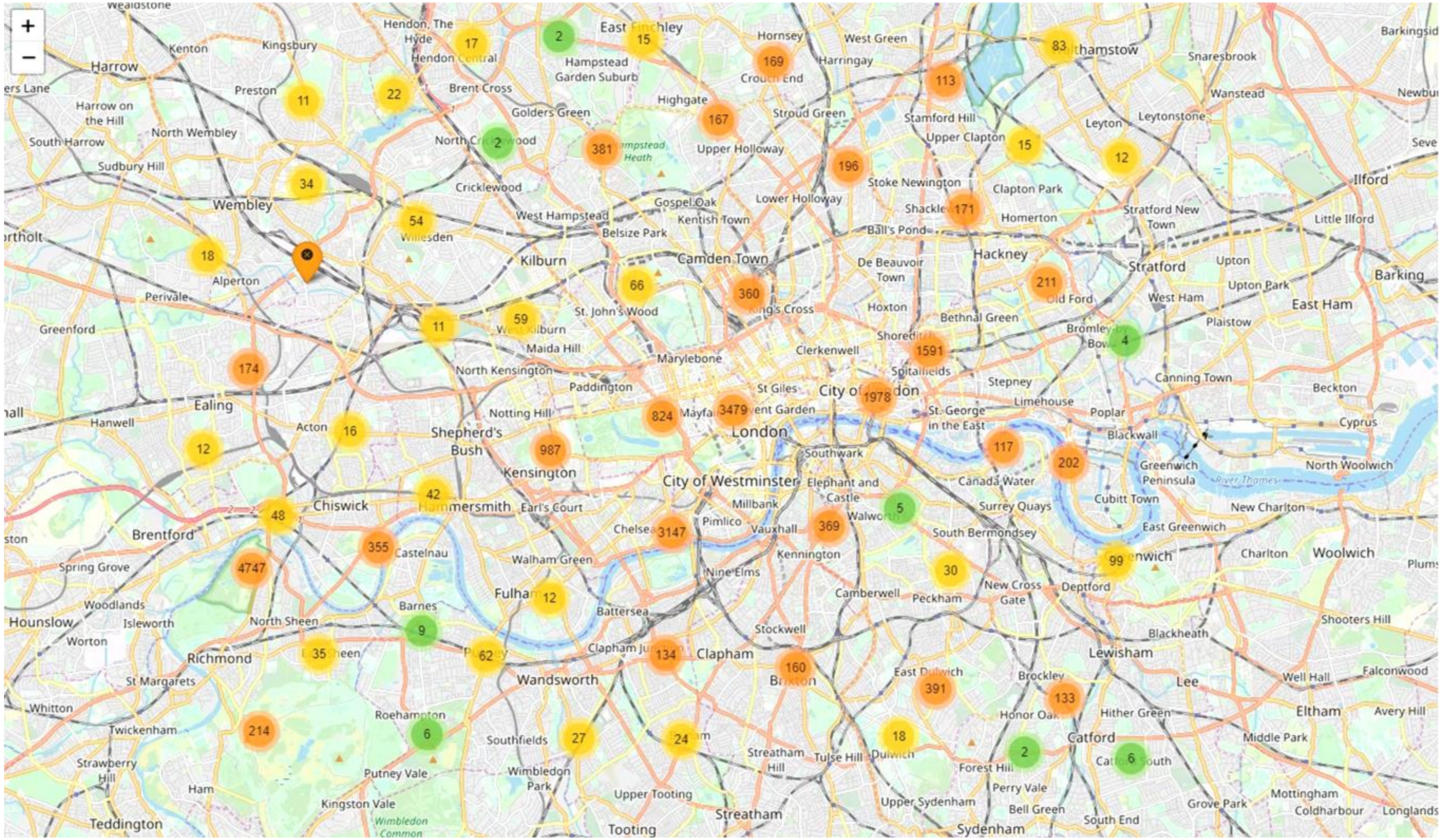


→ R

`plantnet`

github.com/BiologicalRecordsCentre/plantnet
github.com/ropensci/photosearcher

AI Naturalist: Day Out in London



August, et al. 2019. Zenodo. http://doi.org/10.5281/zenodo.3514685

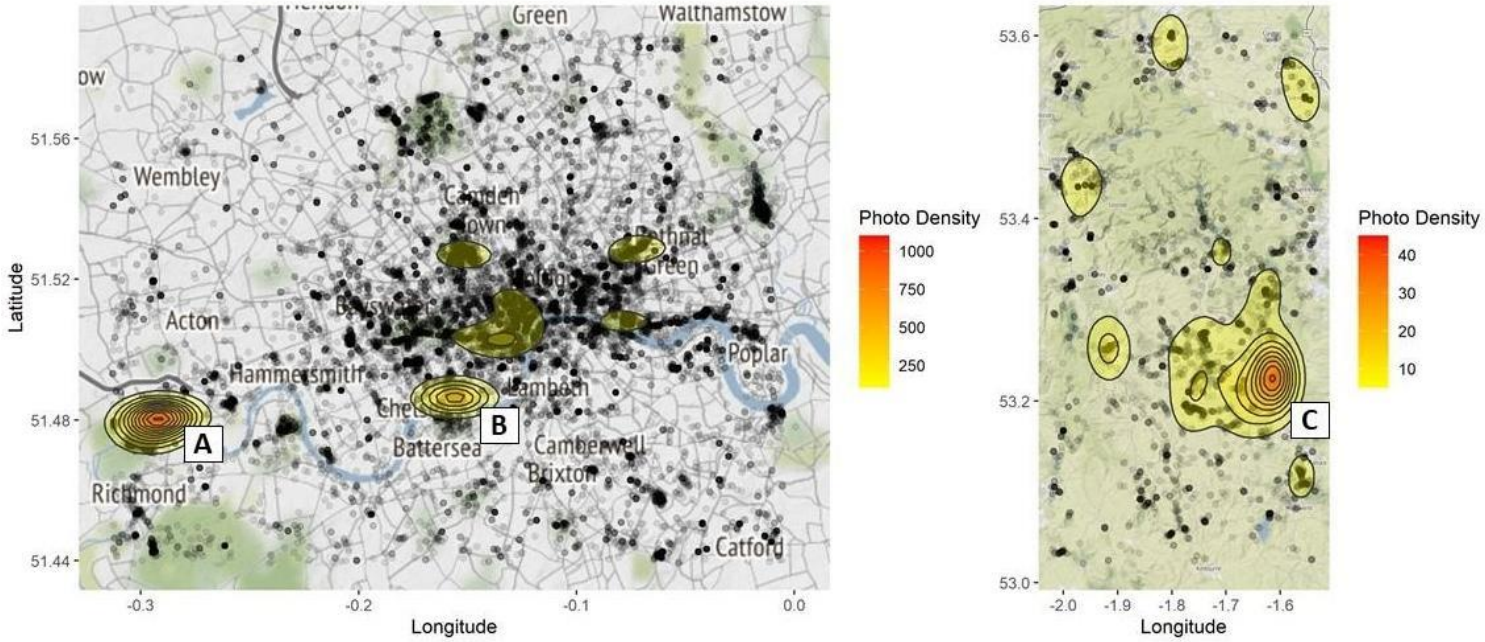
Helianthus annus
Score: 0.23

Challenges
Data are spatially biased

Images are biased by population density

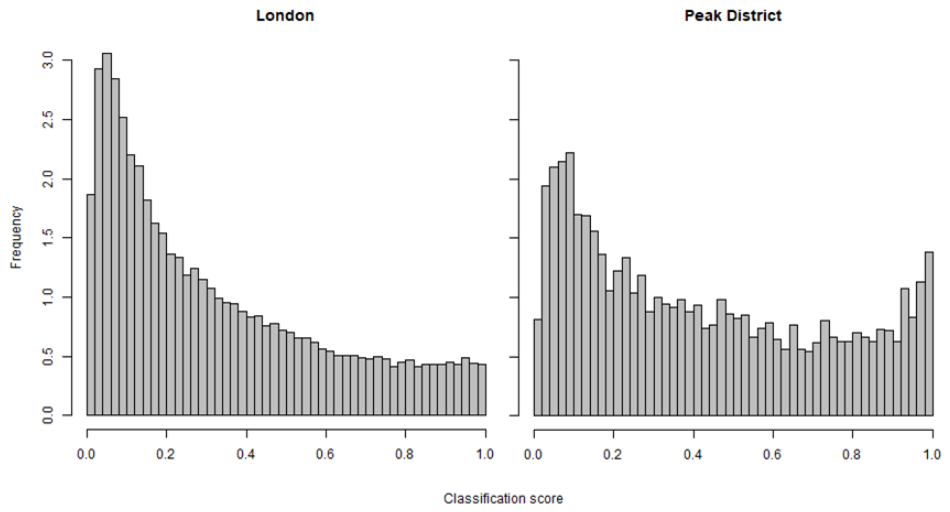
Images are biased by sites of interest

Both biases exist in existing natural history data



August et al. 2020. Patterns. https://doi.org/10.1016/j.patter.2020.100116

Challenges
Images are sub-optimal



Many of the images collected are not of a single species

Many images are of ornamental species

A mismatch results between model train data, and these data

August et al. 2020. Patterns. https://doi.org/10.1016/j.patter.2020.100116

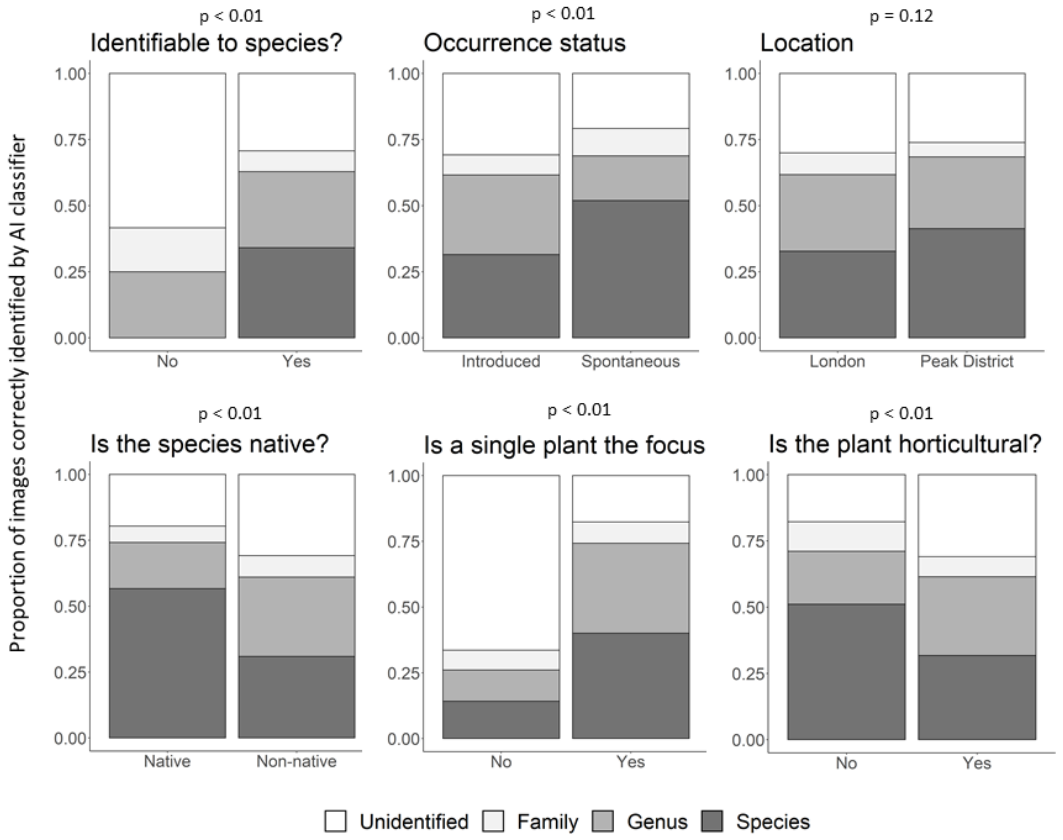
Challenges
Data are spatially biased

Performance is best for:

Naturally occurring

Native plants

Where a single plant is the focus of the image



August et al. 2020. Patterns. https://doi.org/10.1016/j.patter.2020.100116

Solutions
Critical thinking and data exploration

Does the spatial distribution of images fit your needs?

Can you filter images before classification?

What is the appropriate taxonomic resolution for your study?

What reporting biases exist in your dataset?

Do reporting biases change over space or time?

How will you propagate uncertainty in classifications?

Is the dataset used to train your AI naturalist a good match to the images being classified?

Have you adequately documented your dataset?

August et al. 2020. Patterns. https://doi.org/10.1016/j.patter.2020.100116

Thank You