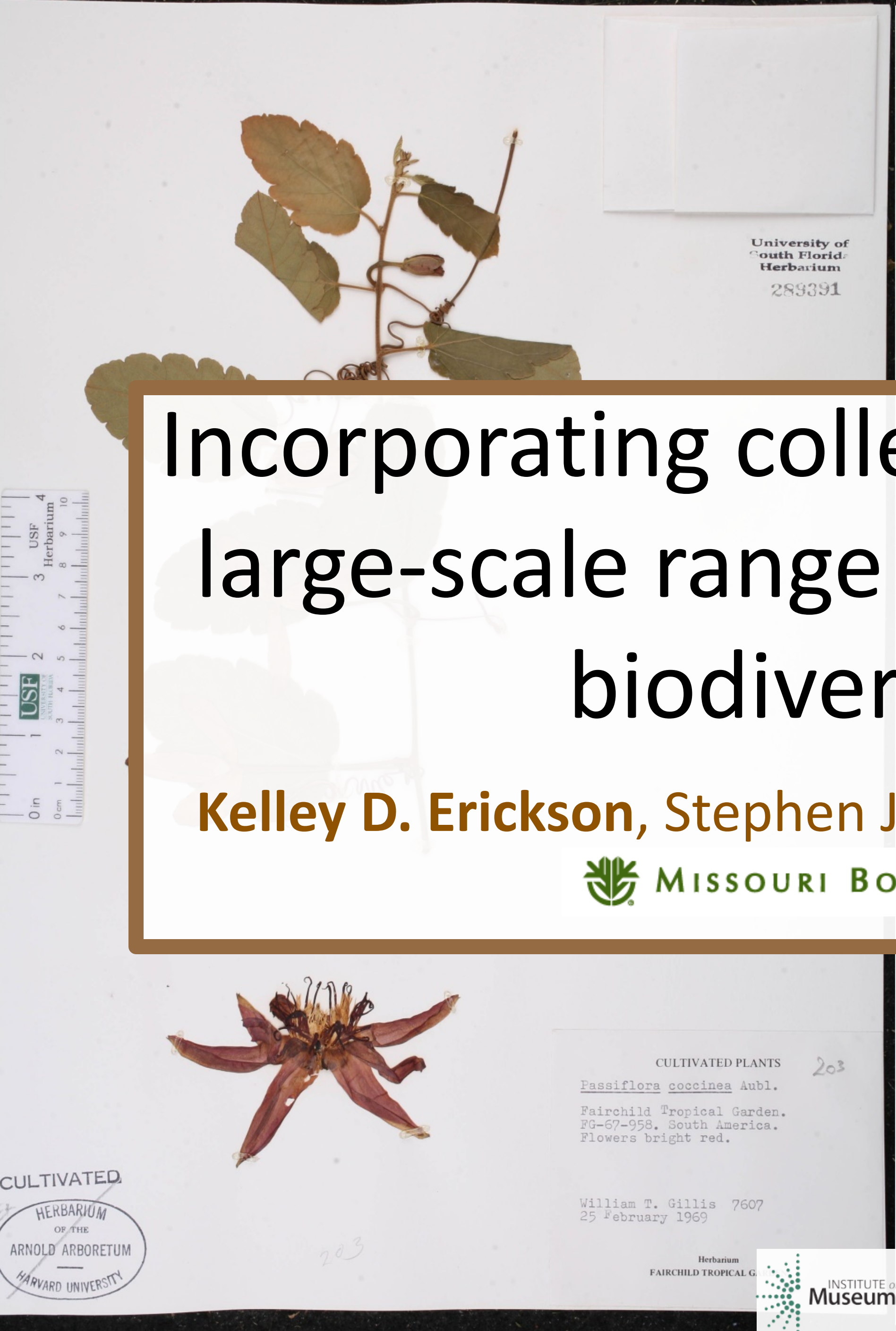
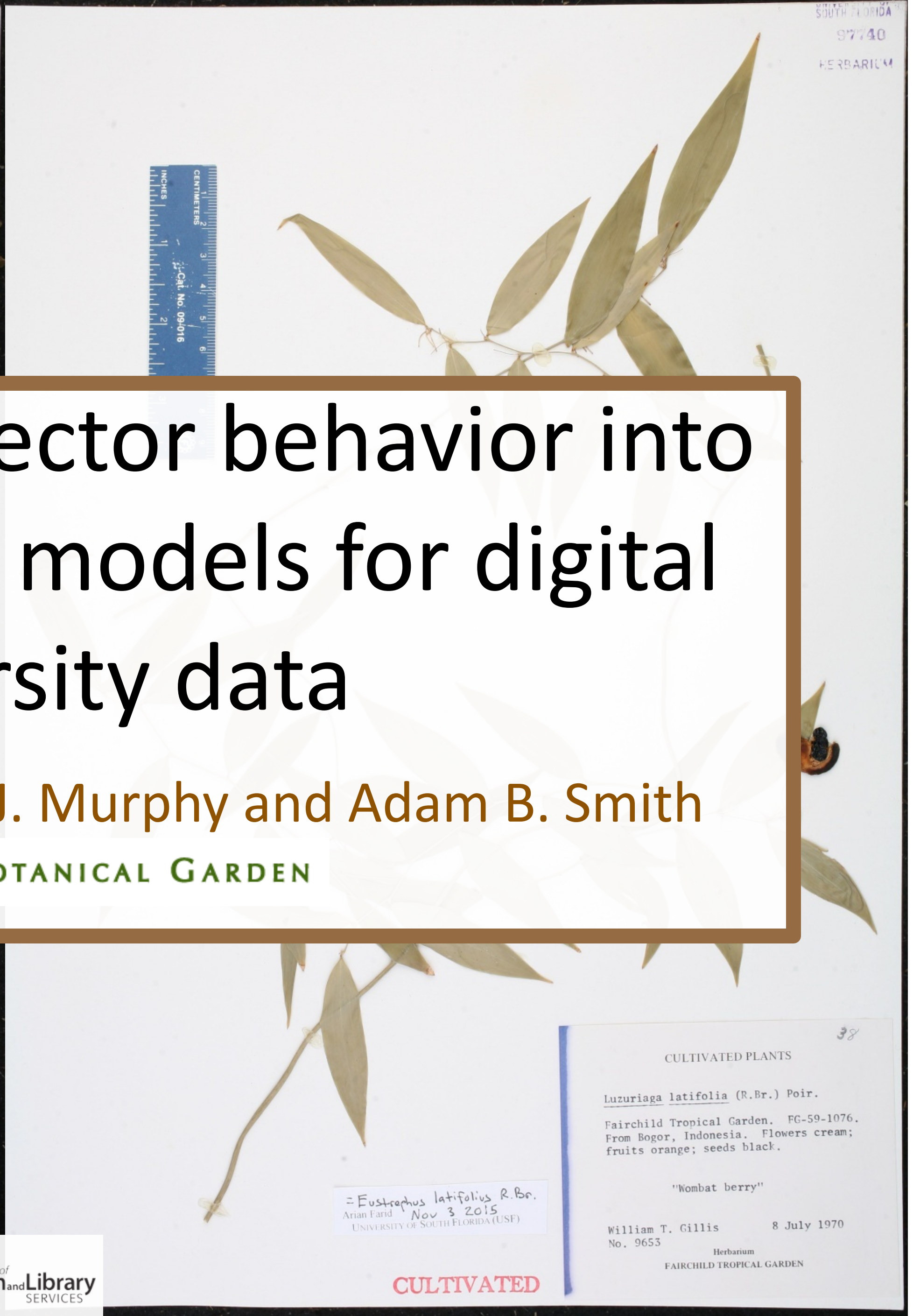

University of
South Florida
Herbarium
289391
USF Herbarium
CULTIVATED PLANTS
203
Passiflora coccinea Aubl.
Fairchild Tropical Garden.
FG-67-958. South America.
Flowers bright red.
William T. Gillis 7607
25 February 1969
Herbarium
FAIRCHILD TROPICAL GARDEN
CULTIVATED
HERBARIUM OF THE ARNOLD ARBORETUM HARVARD UNIVERSITY


UNIVERSITY OF SOUTH FLORIDA
97740
HERBARIUM
INCHES
CENTIMETERS
Cat. No. 09/016
CULTIVATED PLANTS
38
Luzuriaga latifolia (R.Br.) Poir.
Fairchild Tropical Garden. FG-59-1076.
From Bogor, Indonesia. Flowers cream;
fruits orange; seeds black.
"Wombat berry"
William T. Gillis 8 July 1970
No. 9653
Herbarium
FAIRCHILD TROPICAL GARDEN
= Eustrephus latifolius R.Br.
Arian Farid Nov 3 2015
UNIVERSITY OF SOUTH FLORIDA (USF)
CULTIVATED

Incorporating collector behavior into large-scale range models for digital biodiversity data

Kelley D. Erickson, Stephen J. Murphy and Adam B. Smith

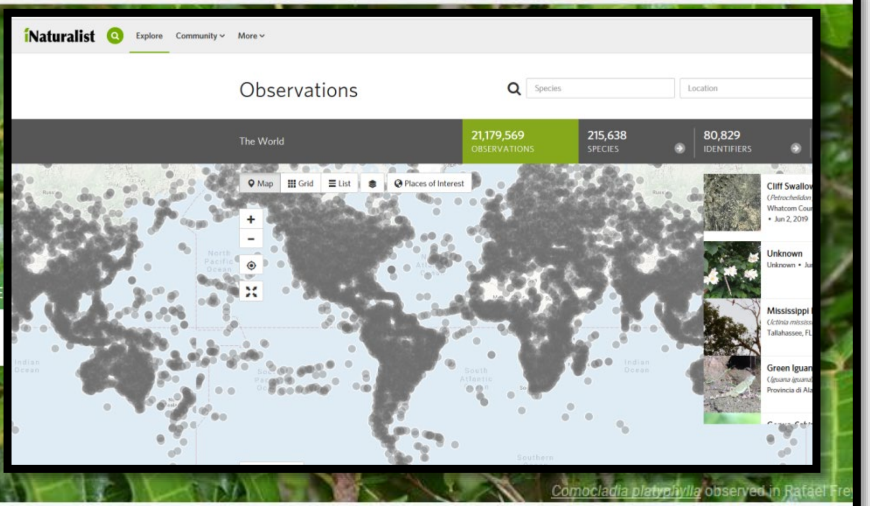
MISSOURI BOTANICAL GARDEN

INSTITUTE of Museum and Library SERVICES

Get data Share Tools Inside GBIF

GBIF | Global Biodiversity Information Facility

Rapidly expanding access to an enormous amount of digital biodiversity data



Observations

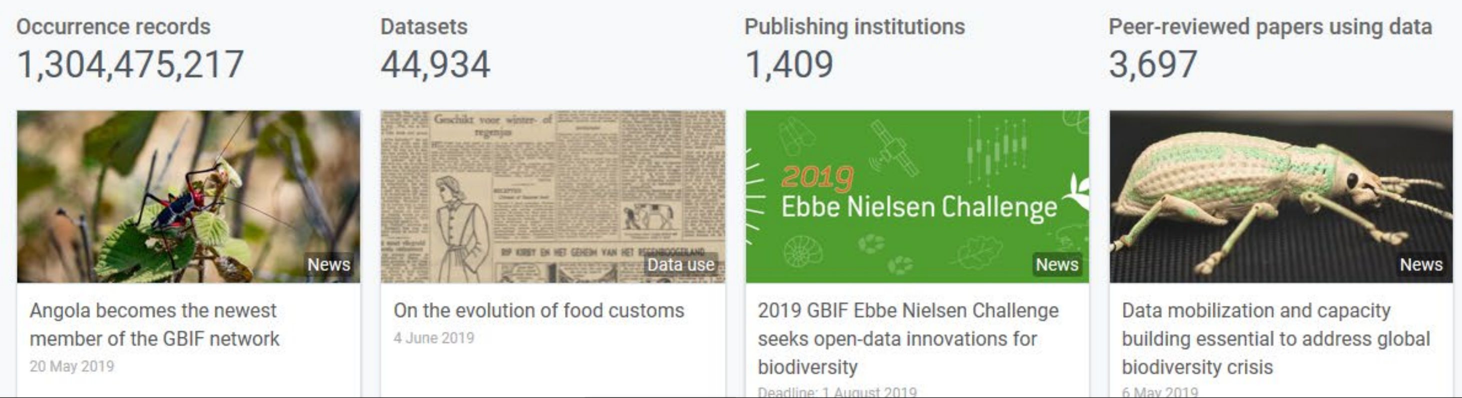
The World

21,179,569 OBSERVATIONS

215,638 SPECIES

80,829 IDENTIFIERS

Comocladia platyphylla observed in Rafael Fre



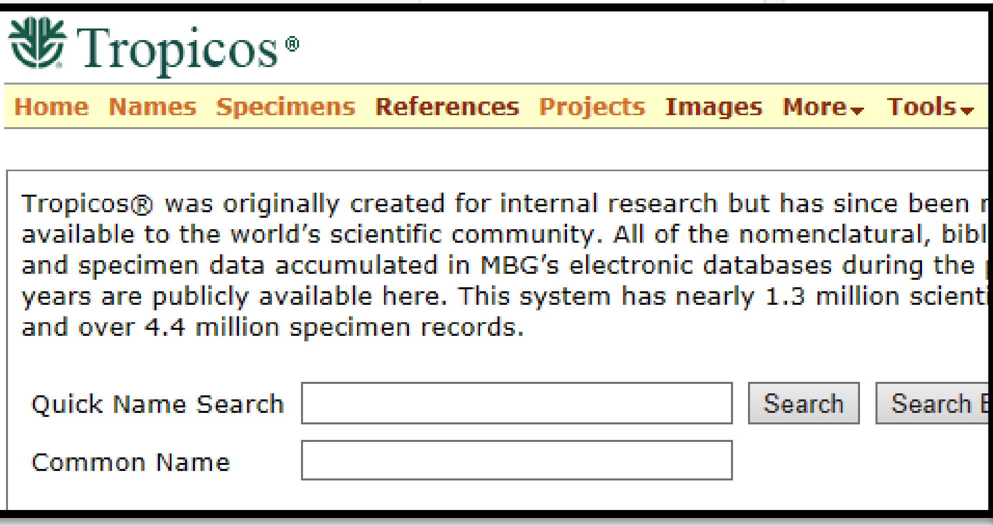
Occurrence records	Datasets	Publishing institutions	Peer-reviewed papers using data
1,304,475,217	44,934	1,409	3,697

Angola becomes the newest member of the GBIF network
20 May 2019

On the evolution of food customs
4 June 2019

2019 GBIF Ebbe Nielsen Challenge seeks open-data innovations for biodiversity
Deadline: 1 August 2019

Data mobilization and capacity building essential to address global biodiversity crisis
6 May 2019



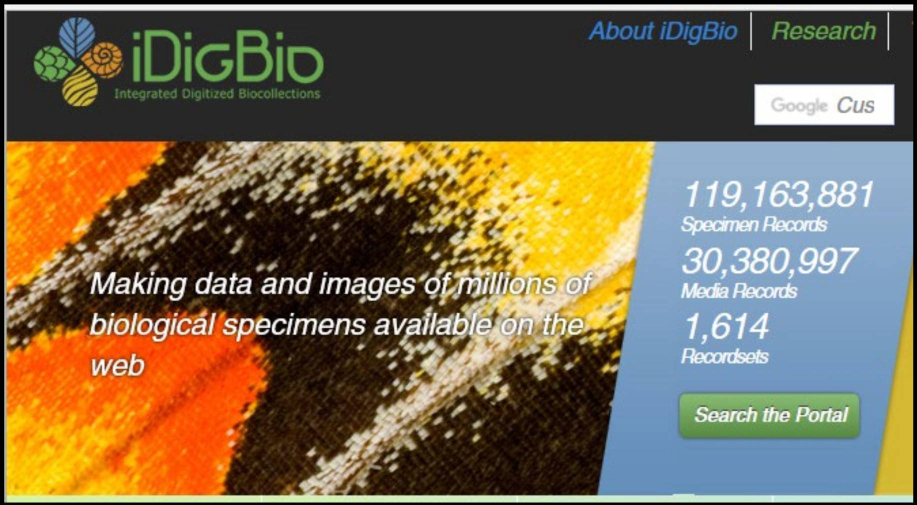
Tropicos®

Home Names Specimens References Projects Images More Tools

Tropicos® was originally created for internal research but has since been r
available to the world's scientific community. All of the nomenclatural, bibl
and specimen data accumulated in MBG's electronic databases during the p
years are publicly available here. This system has nearly 1.3 million scient
and over 4.4 million specimen records.

Quick Name Search □ Search Search E

Common Name □



iDigBio
Integrated Digitized Biocollections

About iDigBio | Research

Google Cus

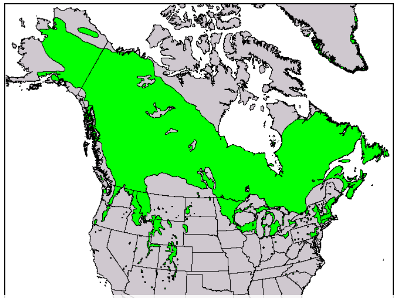
Making data and images of millions of biological specimens available on the web

119,163,881 Specimen Records

30,380,997 Media Records

1,614 Recordsets

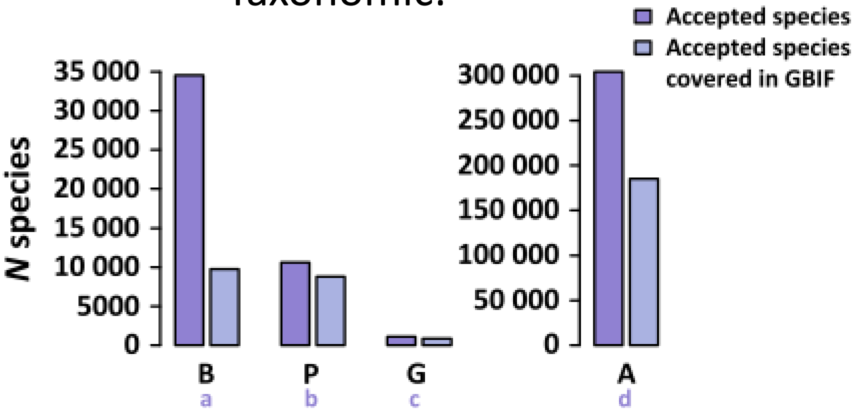
Search the Portal



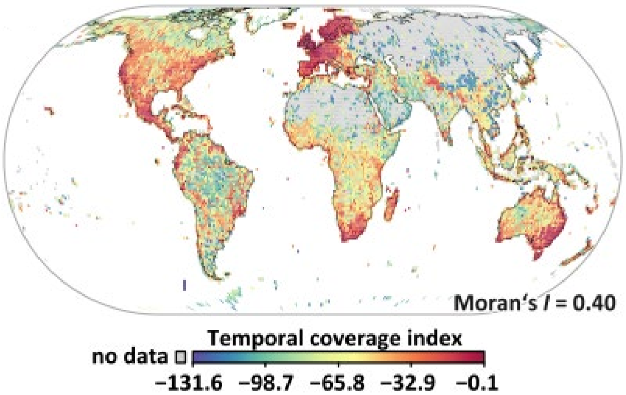
Thompson et al. USGS 1999

Gaps and biases in occurrence data for species ranges

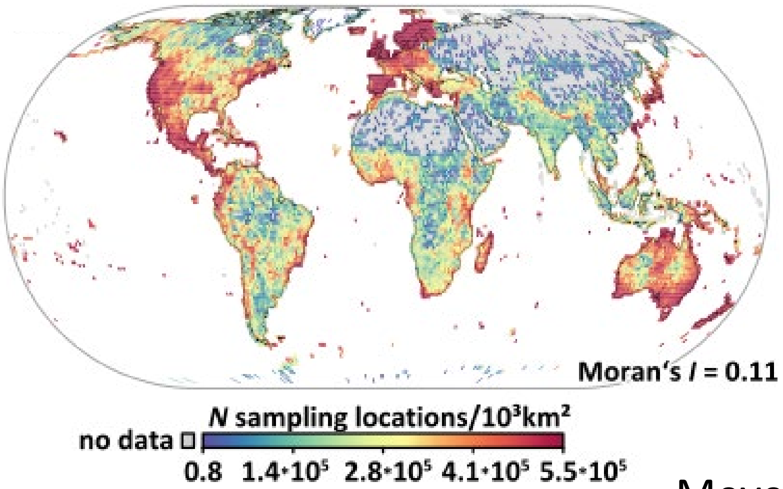
Taxonomic:



Temporal:

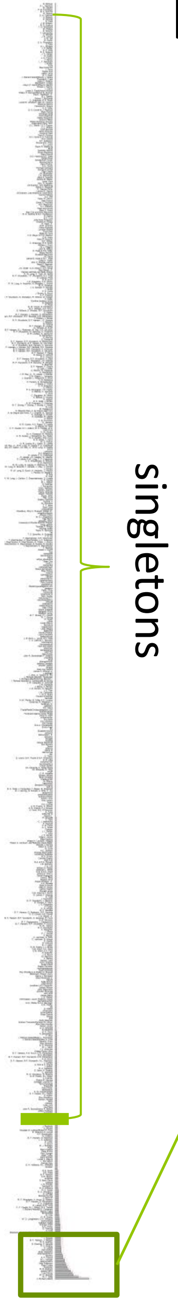


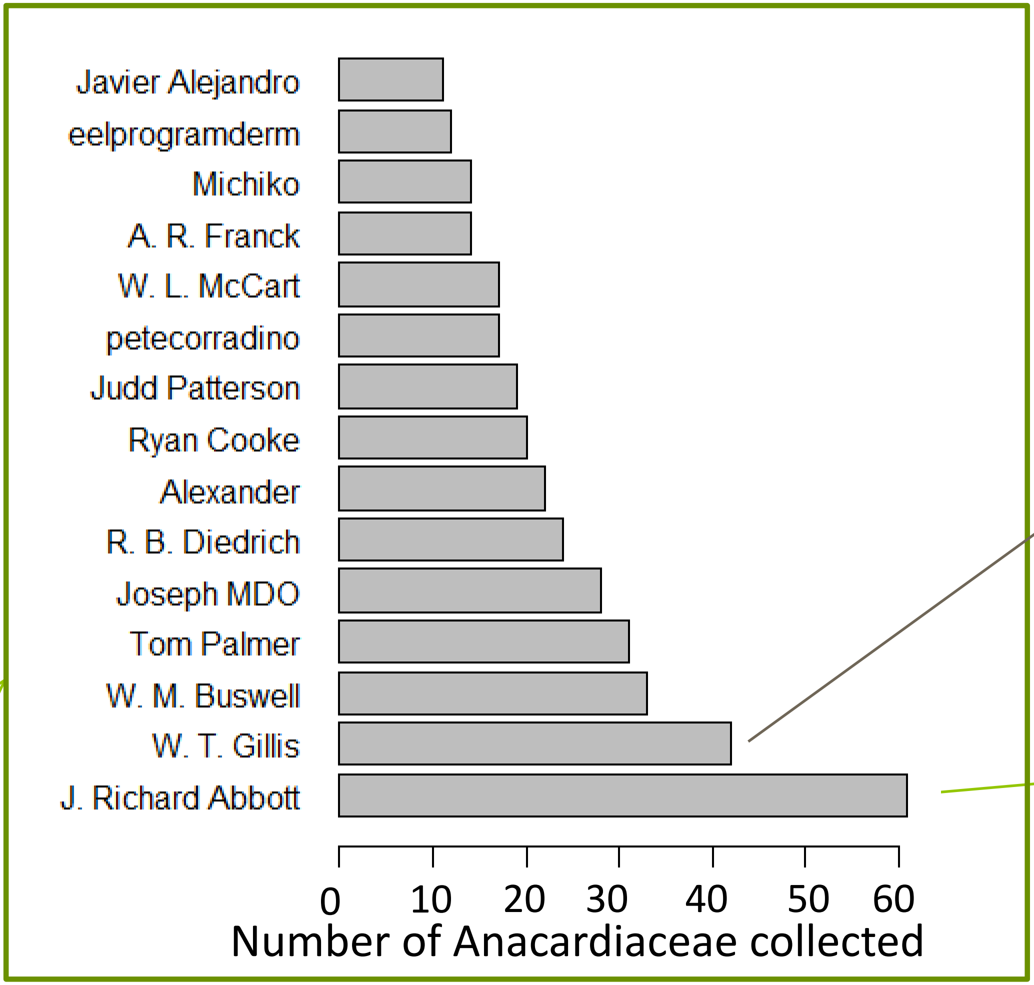
Geographical:



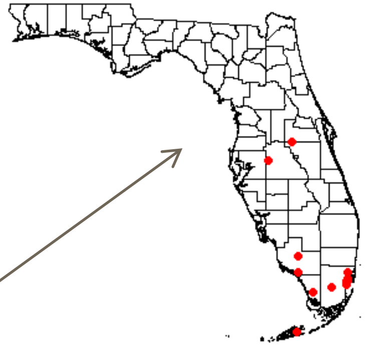
Meyer *et al. Ecology Letters* 2016

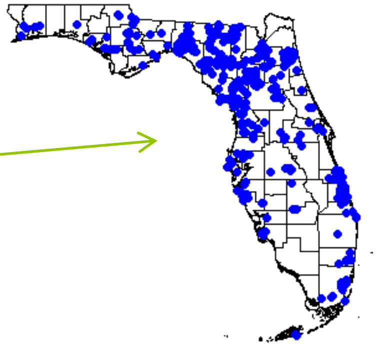
Most collectors collect only once





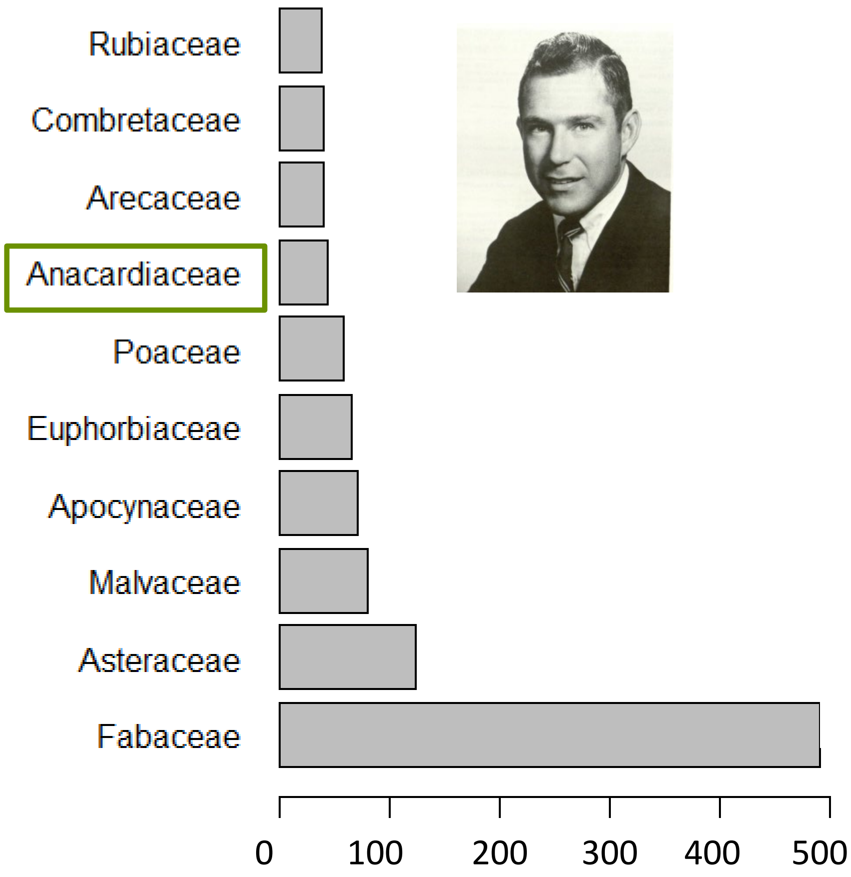
Even those that collect a lot don't collect everywhere:



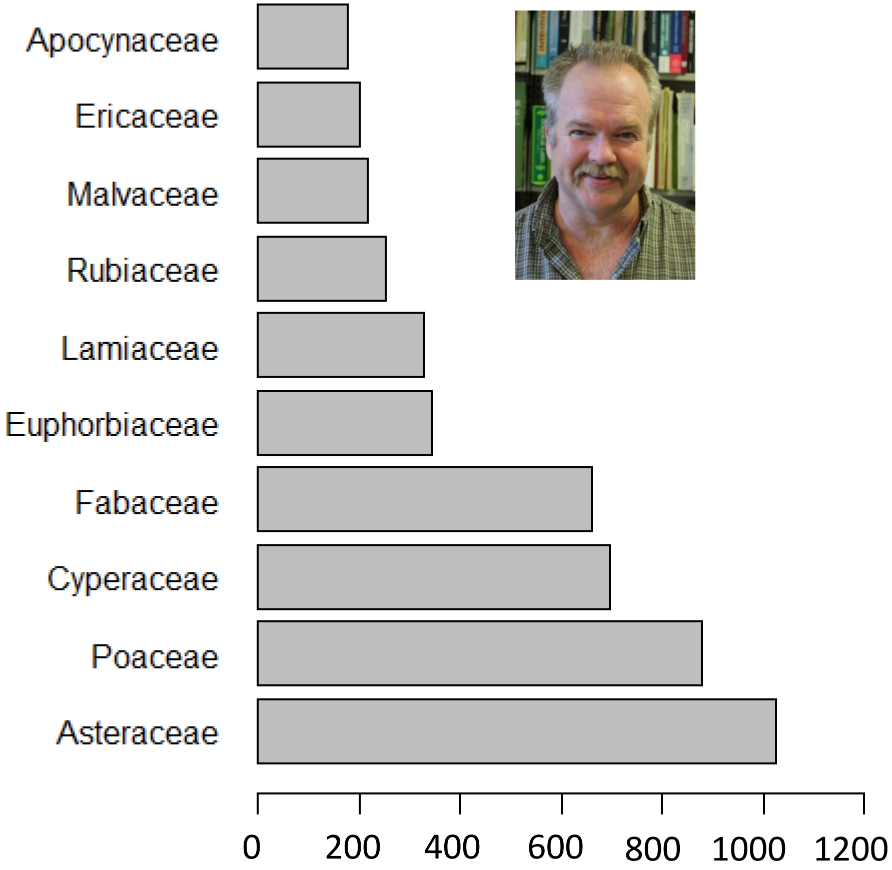


Taxonomic focus varies among collectors: Number of Records by Family

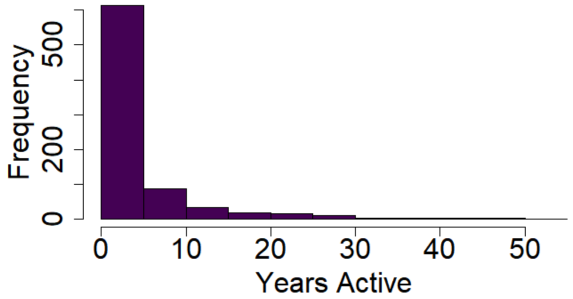
William T. Gillis, Jr.

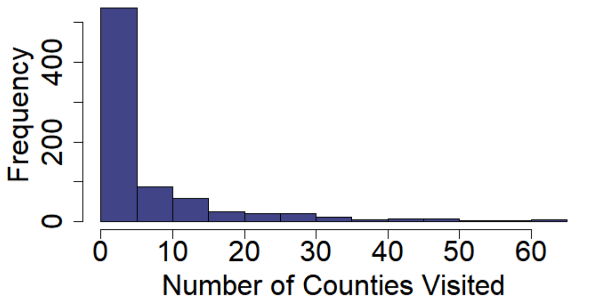


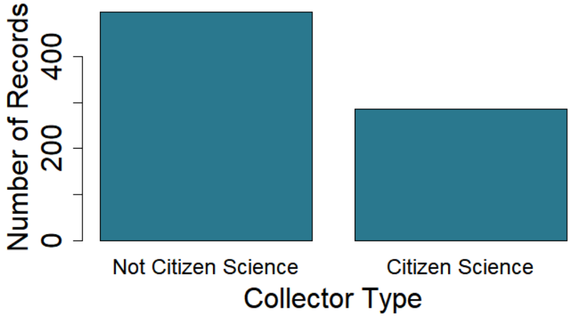
J. Richard Abbott

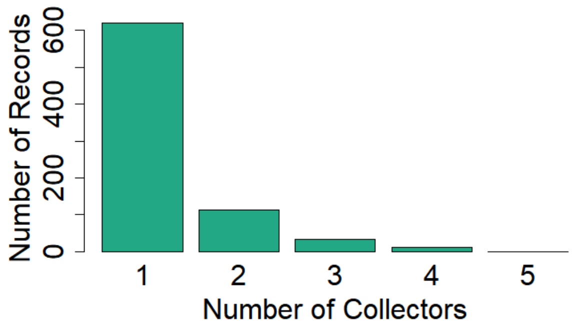


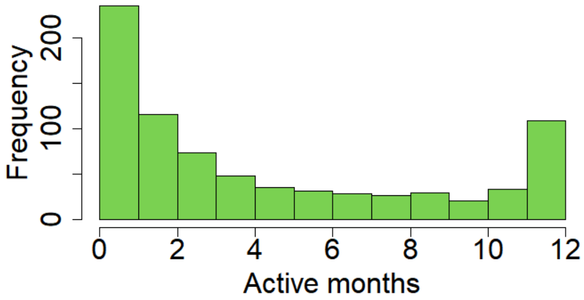
Other ways collectors differ from each other

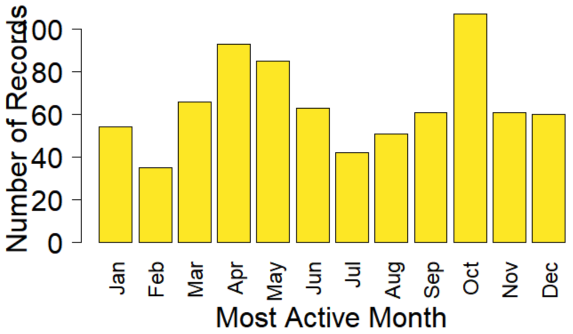




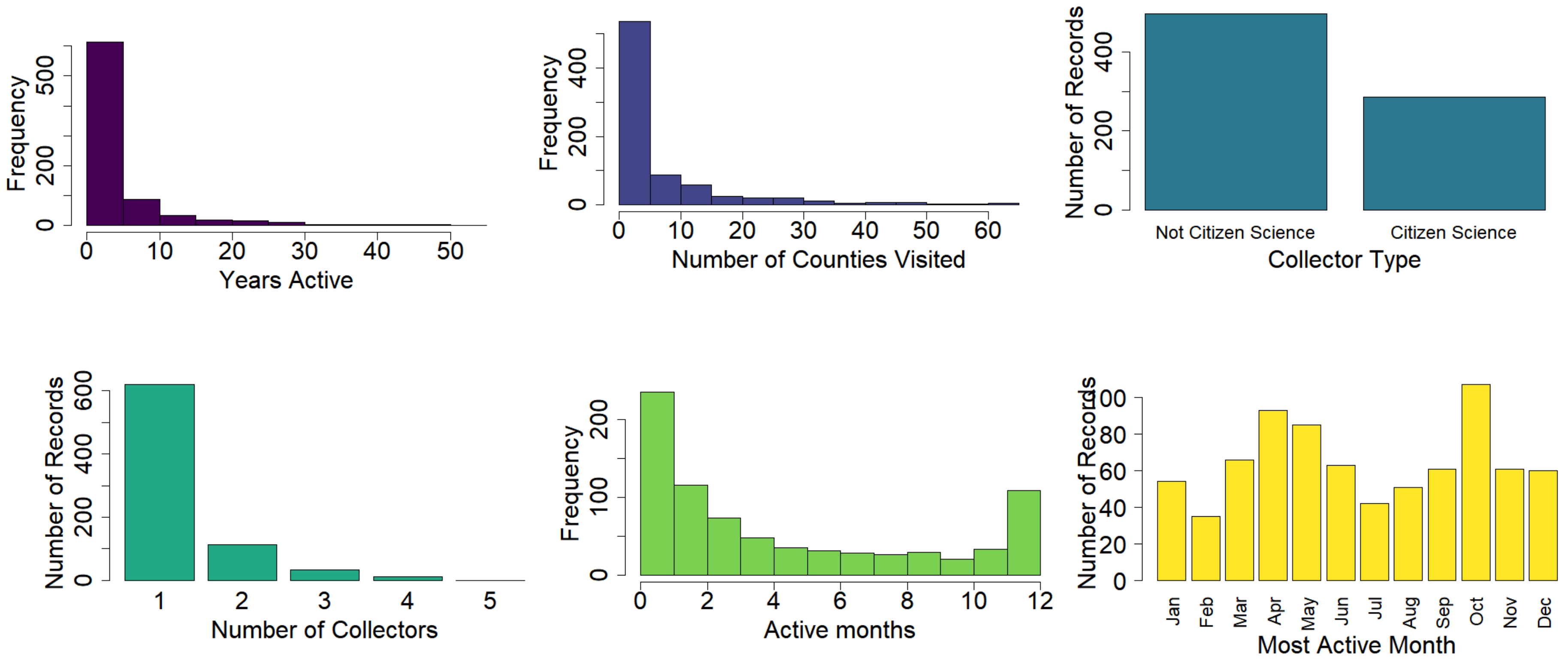








Other ways collectors differ from each other



Differences among collectors give rise to observed biases in occurrence data

Occupancy-detection models

$Z_A = 1$

$Z_B = 1$

$Z_C = 1$

$Z_D = 0$

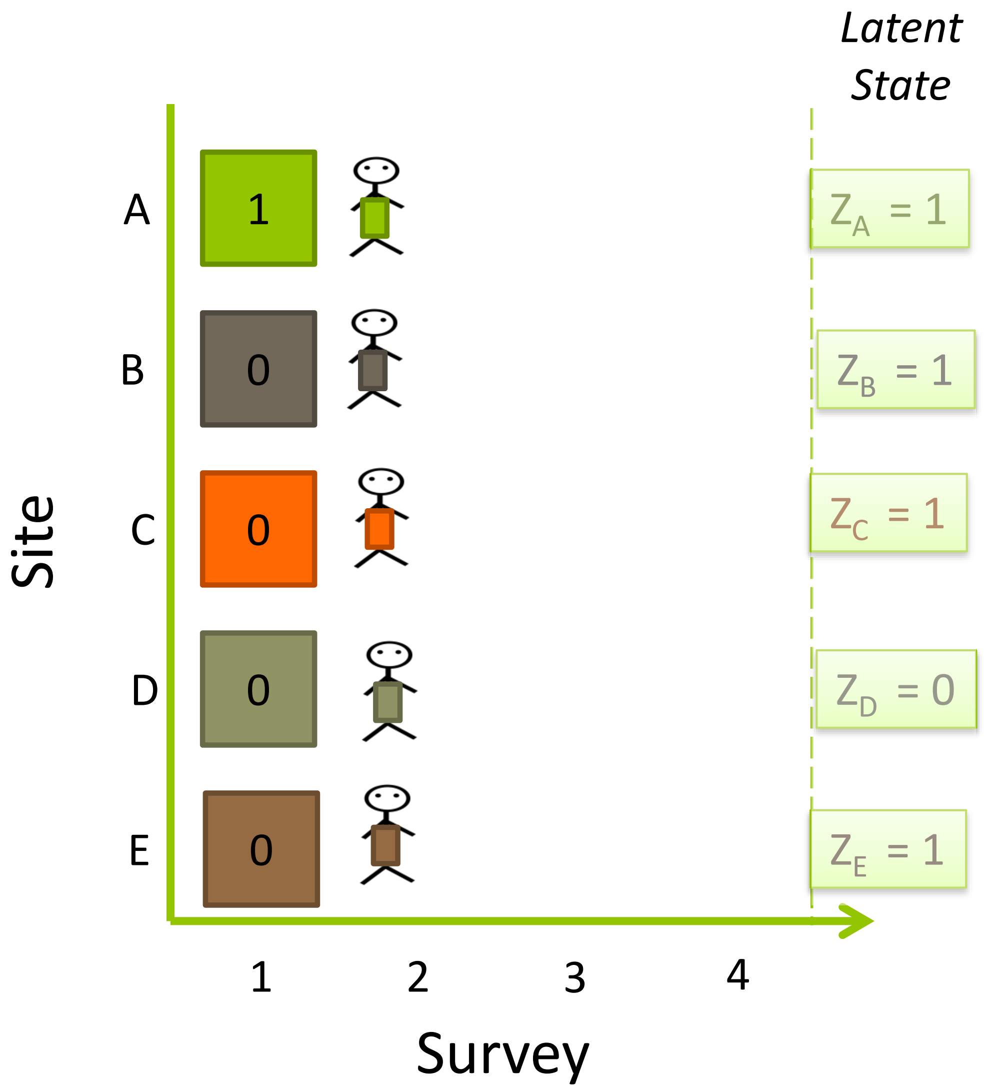
$Z_E = 1$

Ecological Process

$$Z_i \sim Bernoulli\ (\psi_i)$$

MacKenzie et al. *Ecology* 2002

Occupancy-detection models

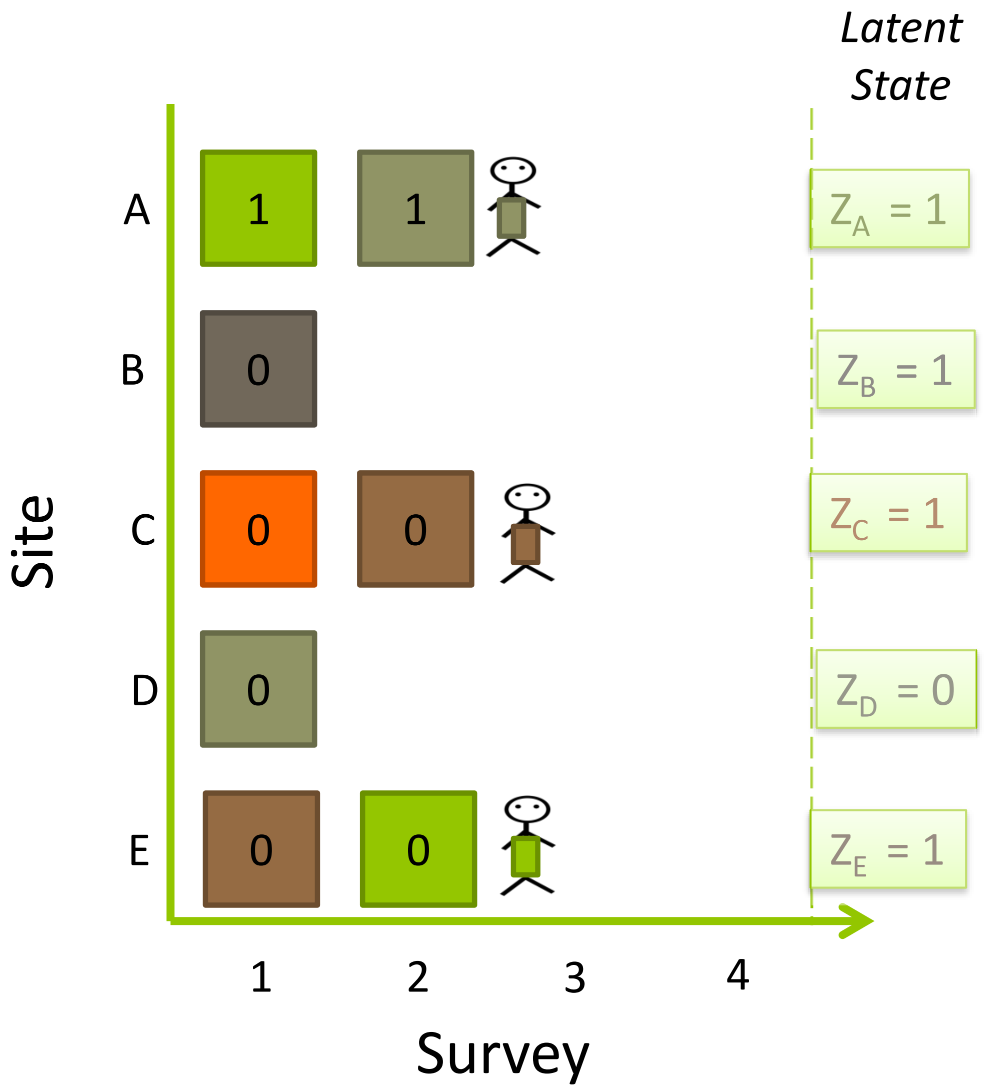


Ecological Process

$$Z_i \sim Bernoulli\ (\psi_i)$$

MacKenzie et al. *Ecology* 2002

Occupancy-detection models



Site	1	2	3	4	*Latent State*
A	1	1			$Z_A = 1$
B	0				$Z_B = 1$
C	0	0			$Z_C = 1$
D	0				$Z_D = 0$
E	0	0			$Z_E = 1$

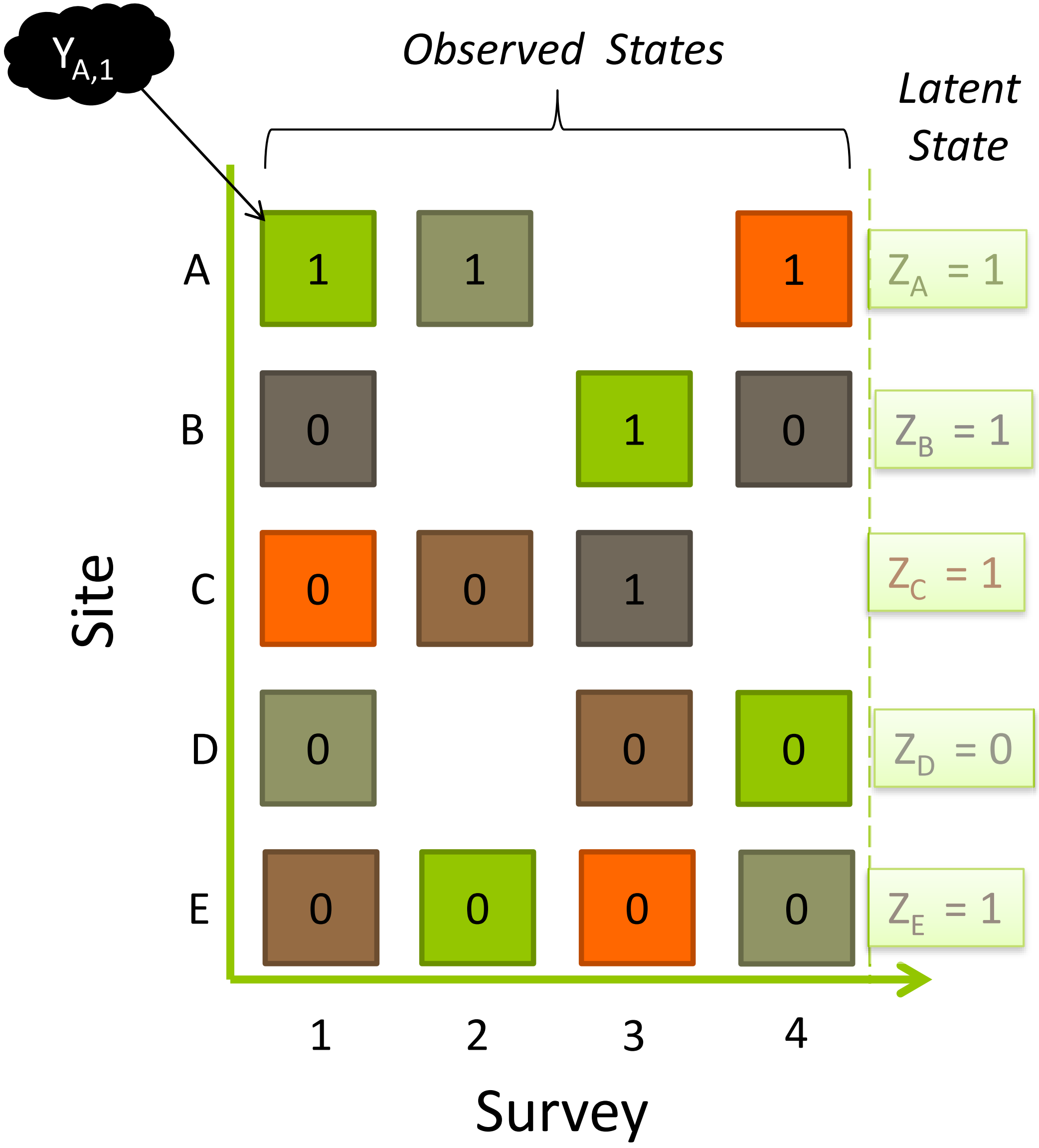
Survey

Ecological Process

$$Z_i \sim Bernoulli\ (\psi_i)$$

MacKenzie et al. *Ecology* 2002

Occupancy-detection models



Site	1	2	3	4	Latent State
A	1	1		1	$Z_A = 1$
B	0		1	0	$Z_B = 1$
C	0	0	1		$Z_C = 1$
D	0		0	0	$Z_D = 0$
E	0	0	0	0	$Z_E = 1$

Ecological Process

$$Z_i \sim Bernoulli\ (\psi_i)$$

Observation Process

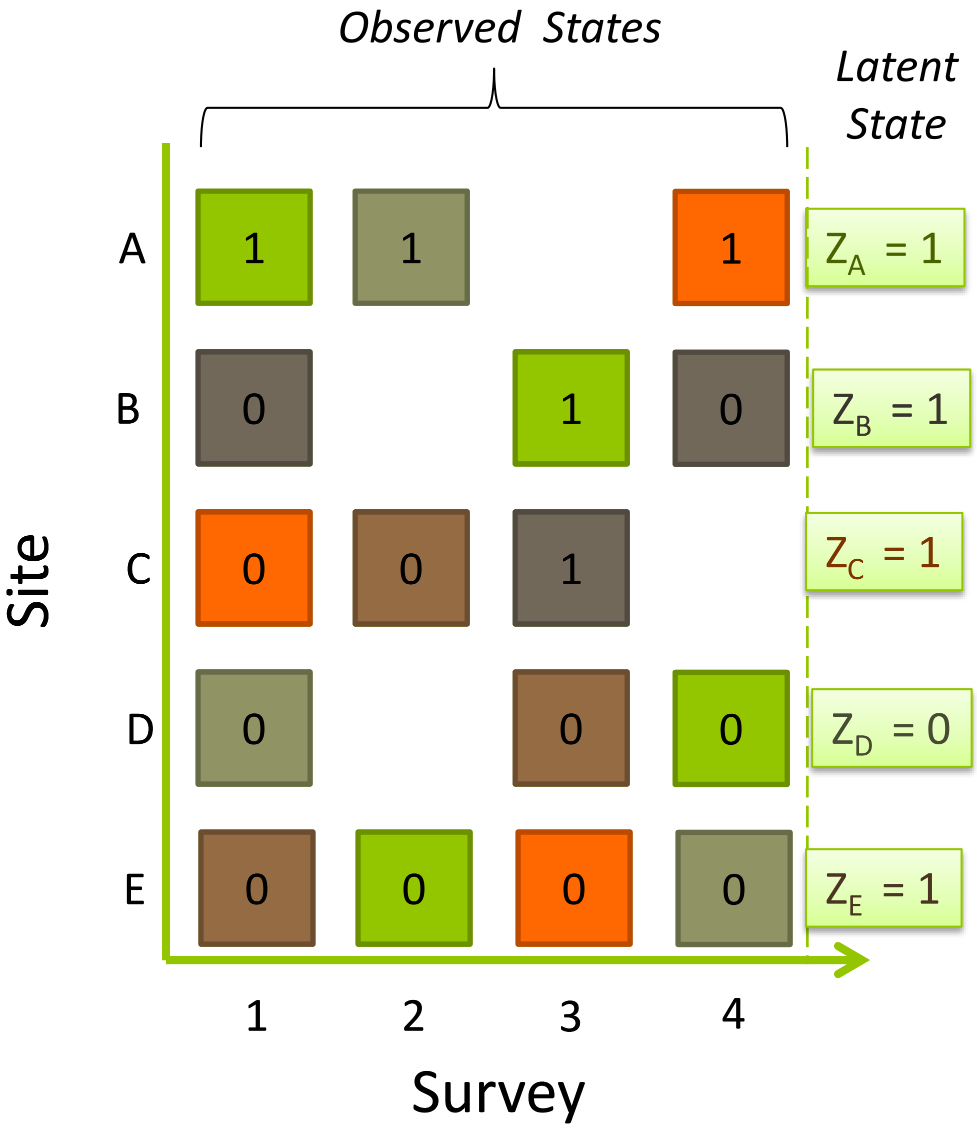
$$Y_{i,t} \mid Z_i \sim Bernoulli\ (p_i \cdot Z_i)$$

Probability of detecting species

Whether species is present at site

MacKenzie et al. *Ecology* 2002

Occupancy-detection models



Observed States

Site	1	2	3	4	Latent State
A	1	1		1	$Z_A = 1$
B	0		1	0	$Z_B = 1$
C	0	0	1		$Z_C = 1$
D	0		0	0	$Z_D = 0$
E	0	0	0	0	$Z_E = 1$

Survey

Ecological Process

$$Z_{i,t} \sim Bernoulli\ (\psi_i)$$

Observation Process

$$Y_{i,t} \mid Z_i \sim Bernoulli\ (p_i \cdot Z_i)$$

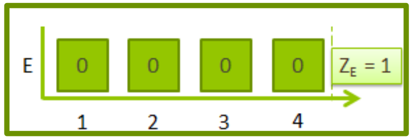
$$\text{logit}(p_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Detectability can depend on characteristics of collectors

MacKenzie et al. *Ecology* 2002

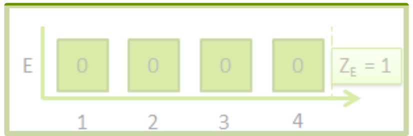
Translating occupancy-detection framework to collections data context

- How do we define a **site**?
  - a county or other defined locality
  - buffer around a given coordinate

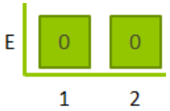


Translating occupancy-detection framework to collections data context

- How do we define a **site**?
  - a county or other defined locality
  - buffer around a given coordinate

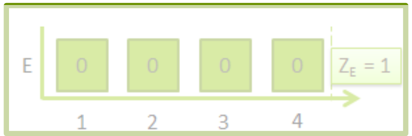


- What is the time scale of the **sampling period?**
  - ***closure assumption:*** *occupancy status of site does not change between sampling periods*

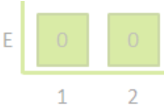


Translating occupancy-detection framework to collections data context

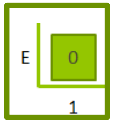
- How do we define a **site**?
  - a county or other defined locality
  - buffer around a given coordinate



- What is the time scale of the **sampling period?**
  - ***closure assumption:*** *occupancy status of site does not change between sampling periods*

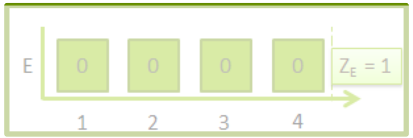


- What counts as a single **survey**?
  - all collections by a **single individual** (or collection group) within a **site** during a single **sampling period**



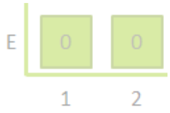
Translating occupancy-detection framework to collections data context

- How do we define a **site**?



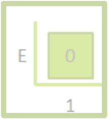
  - a county or other defined locality
  - buffer around a given coordinate

- What is the time scale of the **sampling period?**



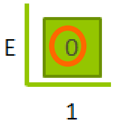
  - ***closure assumption:*** *occupancy status of site does not change between sampling periods*

- What counts as a single **survey**?



  - all collections by a **single individual** (or collection group) within a **site** during a single **sampling period**

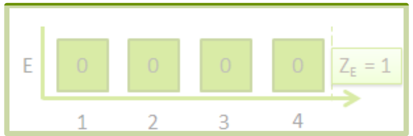
- What is a **non-detection event?**



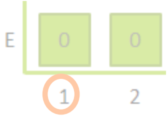
  - **Surveys** by collectors that do not include a single record of the focal species

Translating occupancy-detection framework to collections data context

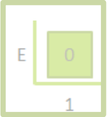
- How do we define a **site**?



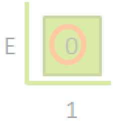
  - a county or other defined locality
  - buffer around a given coordinate
- What is the time scale of the **sampling period?**



  - ***closure assumption:*** *occupancy status of site does not change between sampling periods*
- What counts as a single **survey**?



  - all collections by a **single individual** (or collection group) within a **site** during a single **sampling period**
- What is a **non-detection event?**



  - **Surveys** by collectors that do not include a single record of the focal species

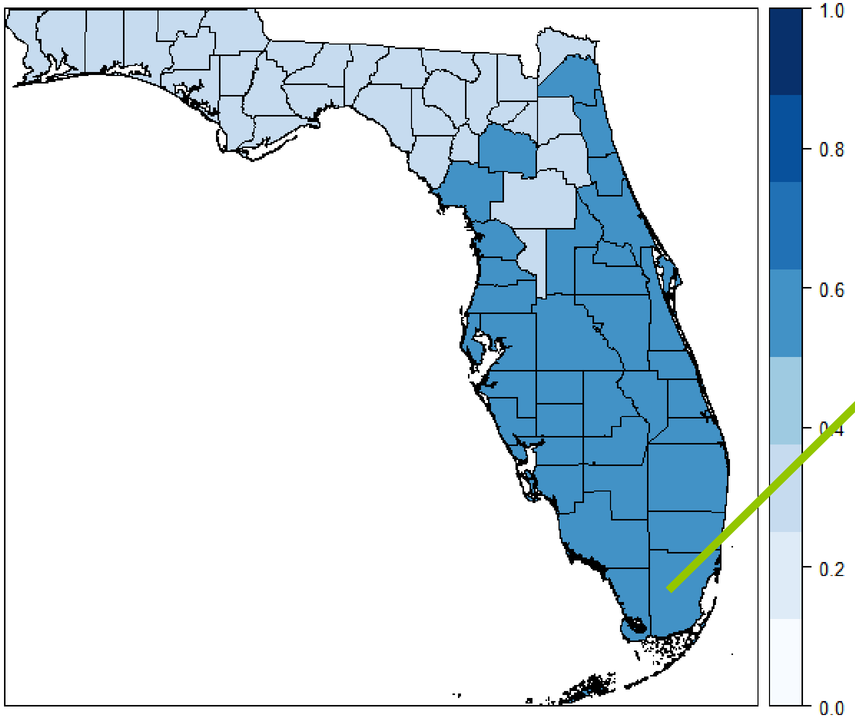
Answers to these questions are study specific!

Case study: Distribution of *Schinus terebinthifolia* in Florida

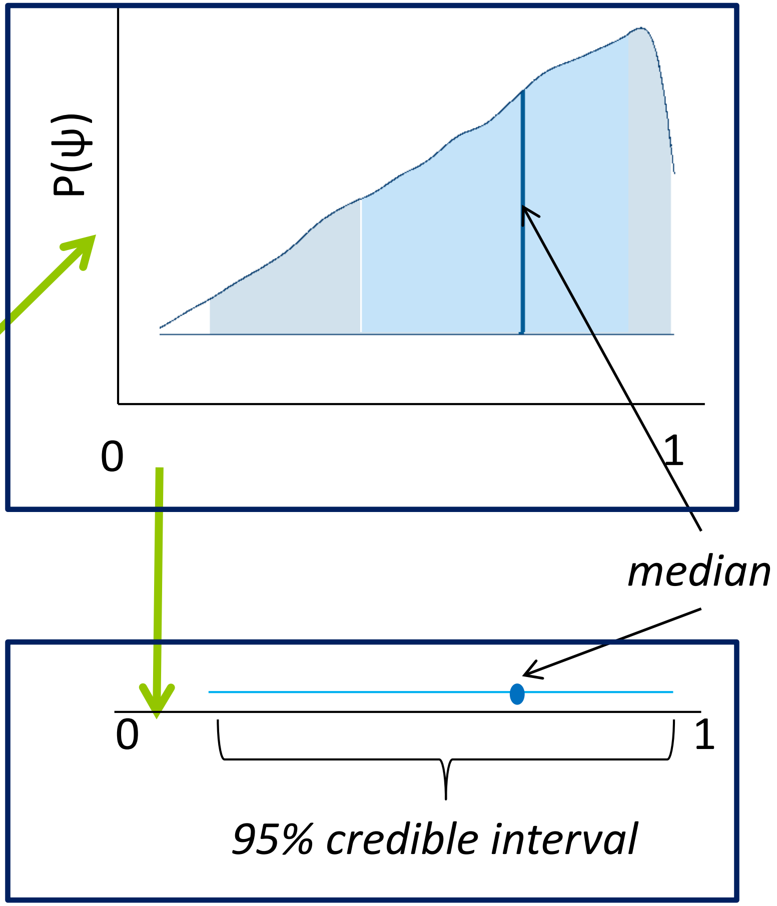
Simple occupancy-detection model: Occupancy

$\text{logit}(p_i) = \varepsilon_i$ ← random effect of county

Median of posterior distribution of ψ



Posterior Distribution of Occupancy for Miami-Dade County



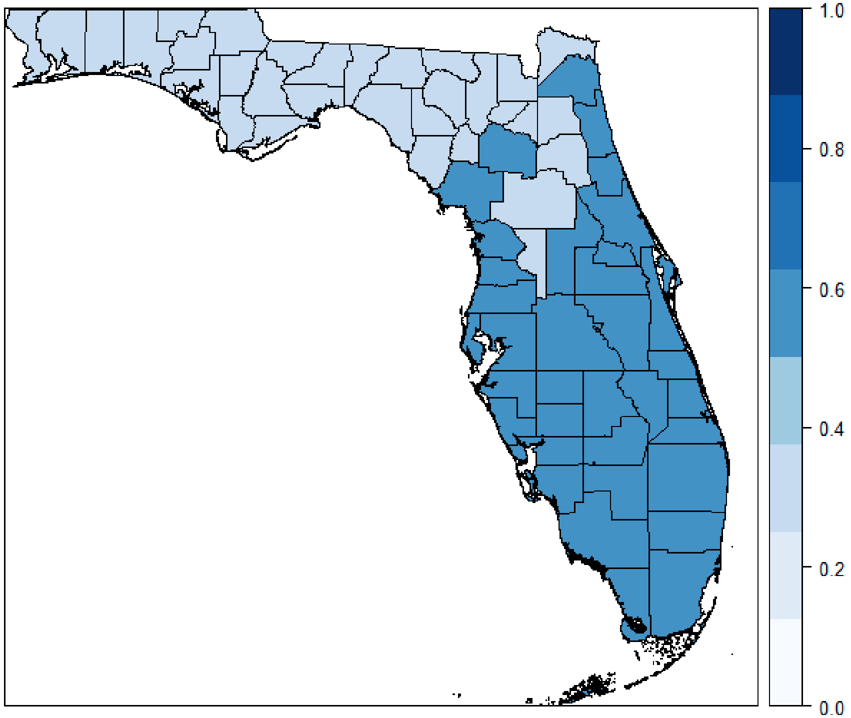
Large uncertainty in estimates of occupancy

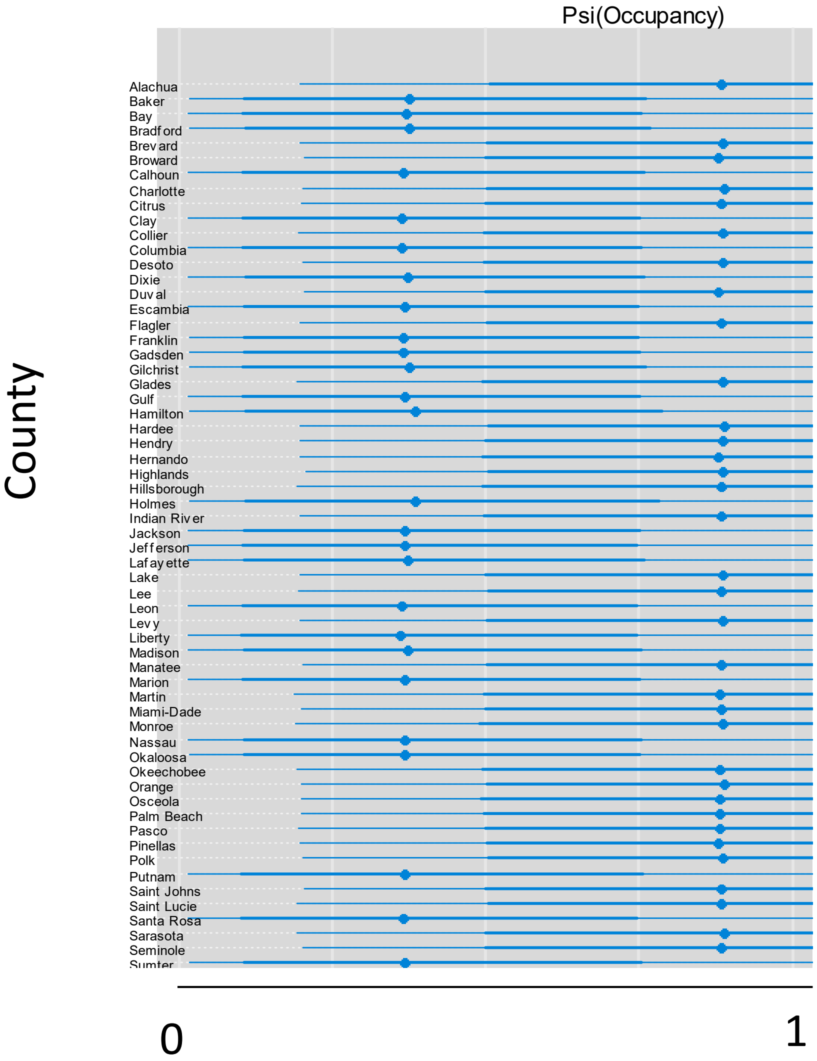
Large uncertainty in estimates of occupancy

$$\text{logit}(p_i) = \varepsilon_i$$

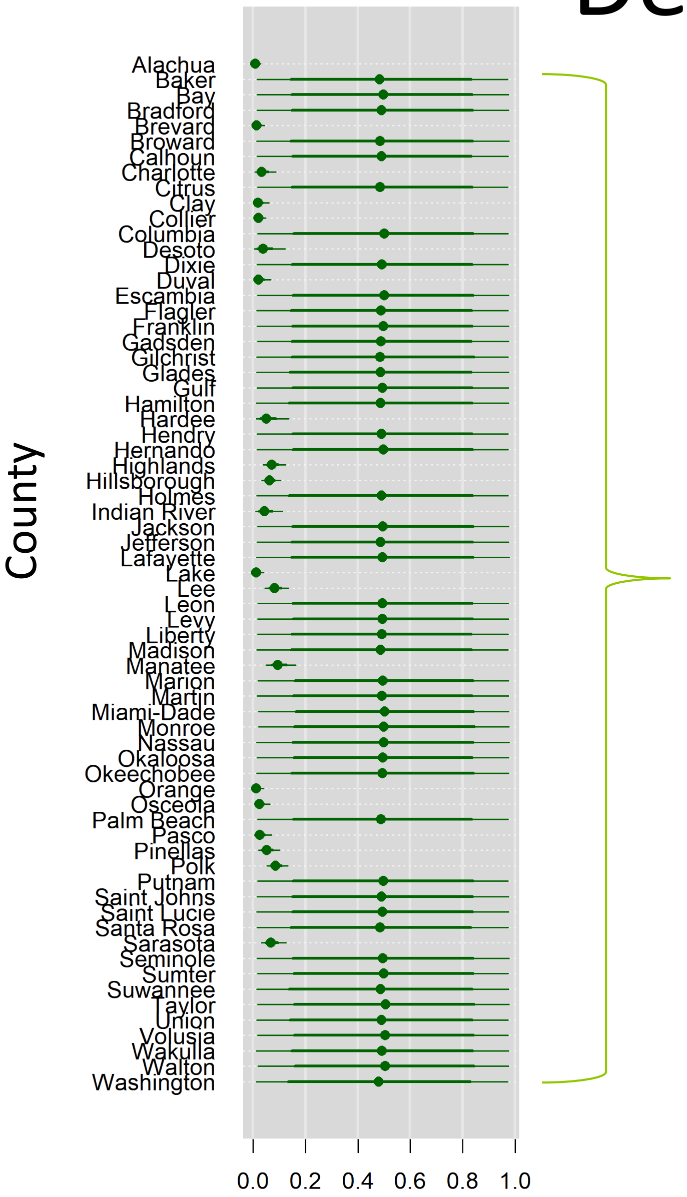
random effect of county

Median of posterior distribution of ψ



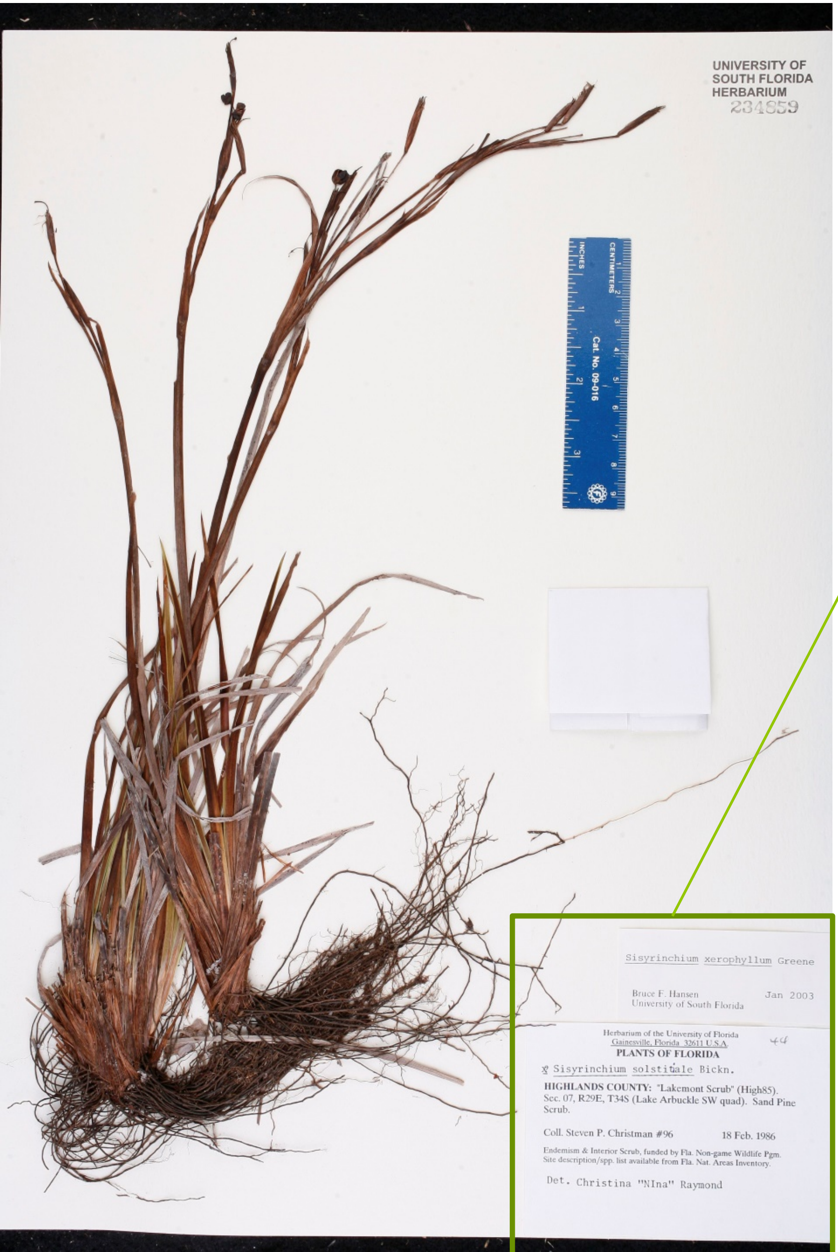


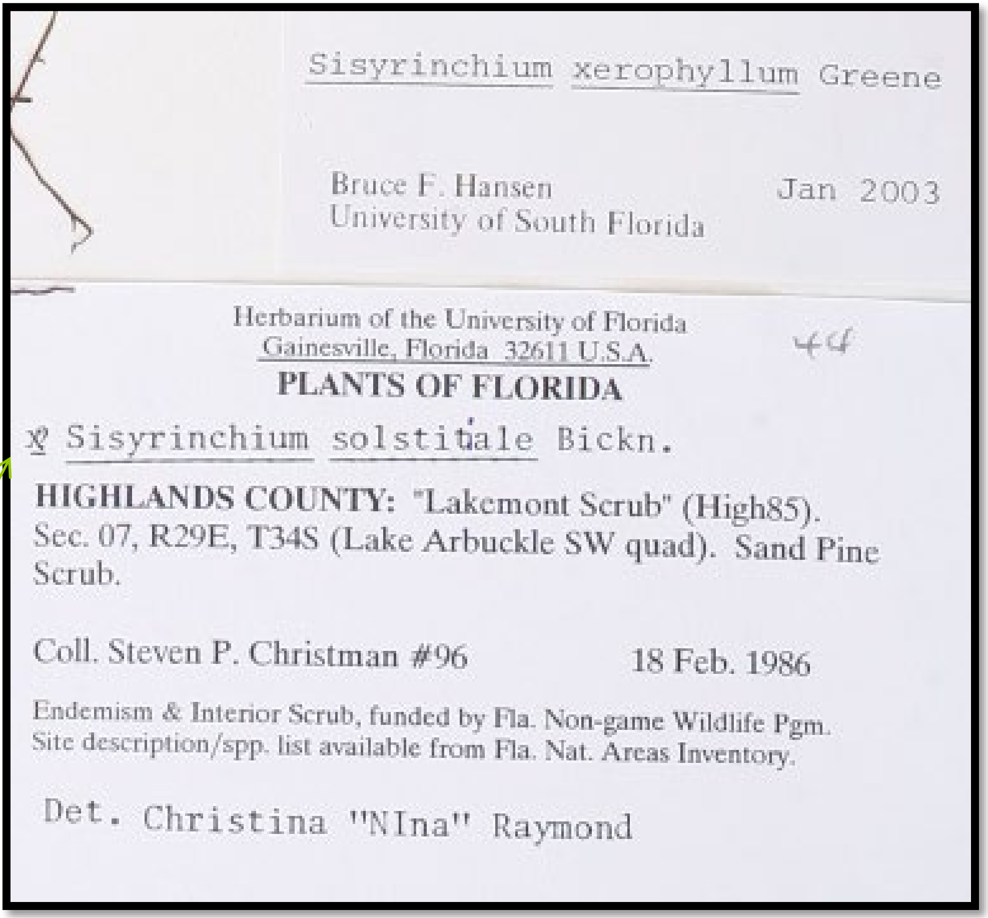
Simple occupancy-detection model: Detectability



Large uncertainty in detectability when ignore collector behavior

Collector data is messy





Sisyrinchium xerophyllum Greene

Bruce F. Hansen
University of South Florida
Jan 2003

Herbarium of the University of Florida
Gainesville, Florida 32611 U.S.A.
44

PLANTS OF FLORIDA

x Sisyrinchium solstitiale Bickn.

HIGHLANDS COUNTY: "Lakemont Scrub" (High85). Sec. 07, R29E, T34S (Lake Arbuckle SW quad). Sand Pine Scrub.

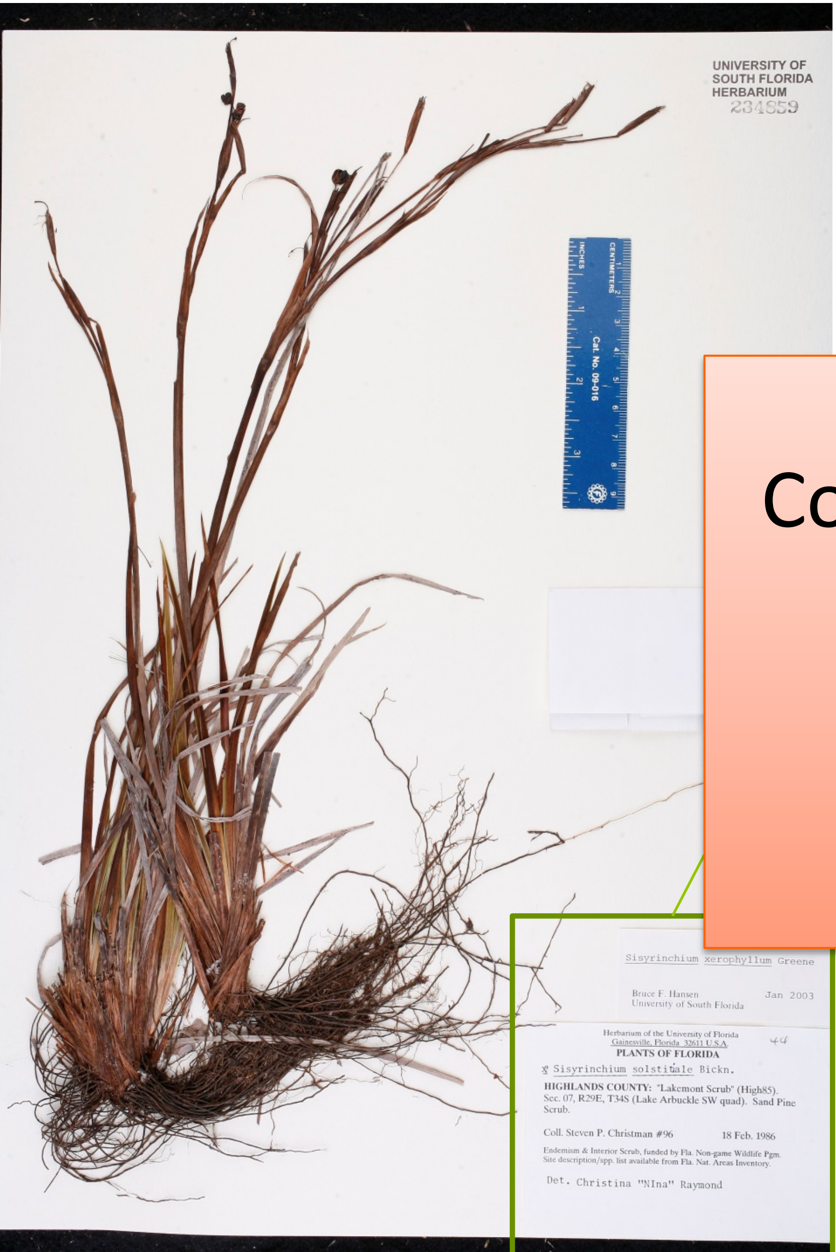
Coll. Steven P. Christman #96 18 Feb. 1986

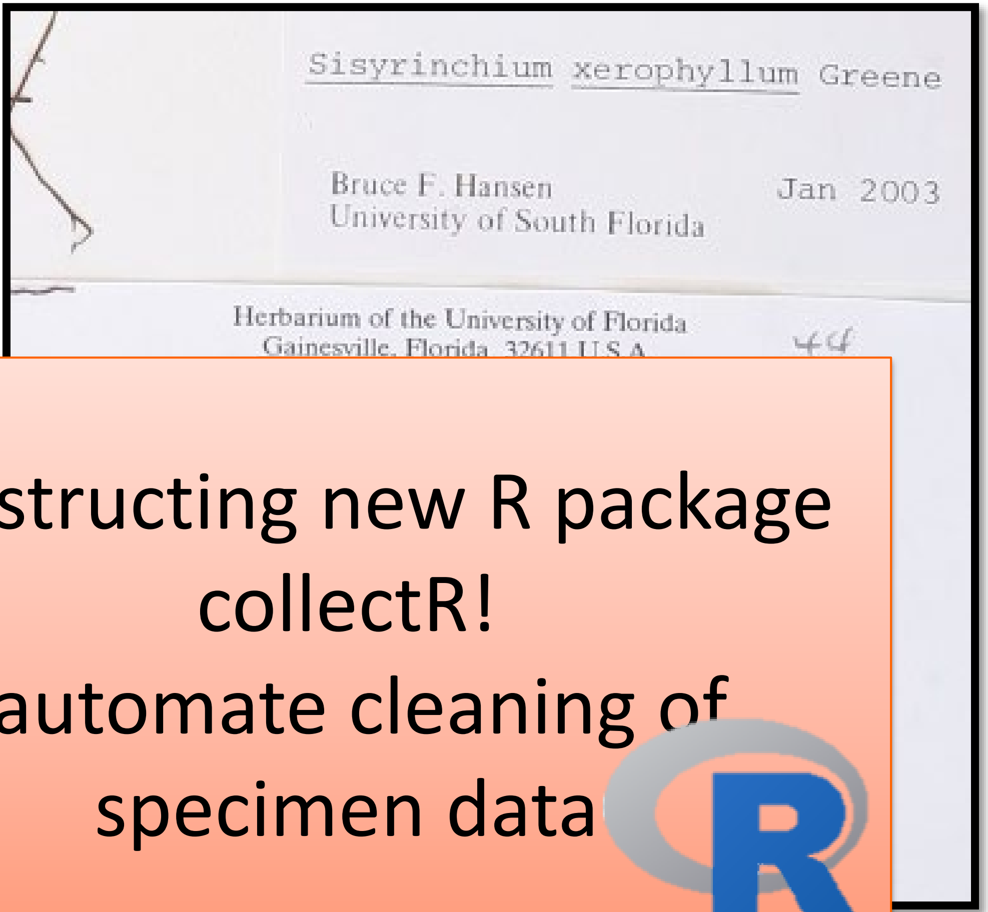
Endemism & Interior Scrub, funded by Fla. Non-game Wildlife Pgm. Site description/spp. list available from Fla. Nat. Areas Inventory.

Det. Christina "NIna" Raymond

L. E. Arnold & Erdman West
Erdman West & Lillian Arnold
L. Arnold & Erdman West
L. E. Arnold, Erdman West
Arnold & West
L. E. Arnold; Erdman West
E. West, L. Arnold

Collector data is messy


UNIVERSITY OF SOUTH FLORIDA HERBARIUM
234859
Cat. No. 09-016
CENTIMETERS
INCHES
Sisyrinchium xerophyllum Greene
Bruce F. Hansen
University of South Florida
Jan 2003
Herbarium of the University of Florida
Gainesville, Florida 32611 U.S.A.
44
PLANTS OF FLORIDA
x Sisyrinchium solstitiale Bickn.
HIGHLANDS COUNTY: "Lakemont Scrub" (High85). Sec. 07, R29E, T34S (Lake Arbuckle SW quad). Sand Pine Scrub.
Coll. Steven P. Christman #96
18 Feb. 1986
Endemism & Interior Scrub, funded by Fla. Non-game Wildlife Pgm. Site description/spp. list available from Fla. Nat. Areas Inventory.
Det. Christina "NIna" Raymond

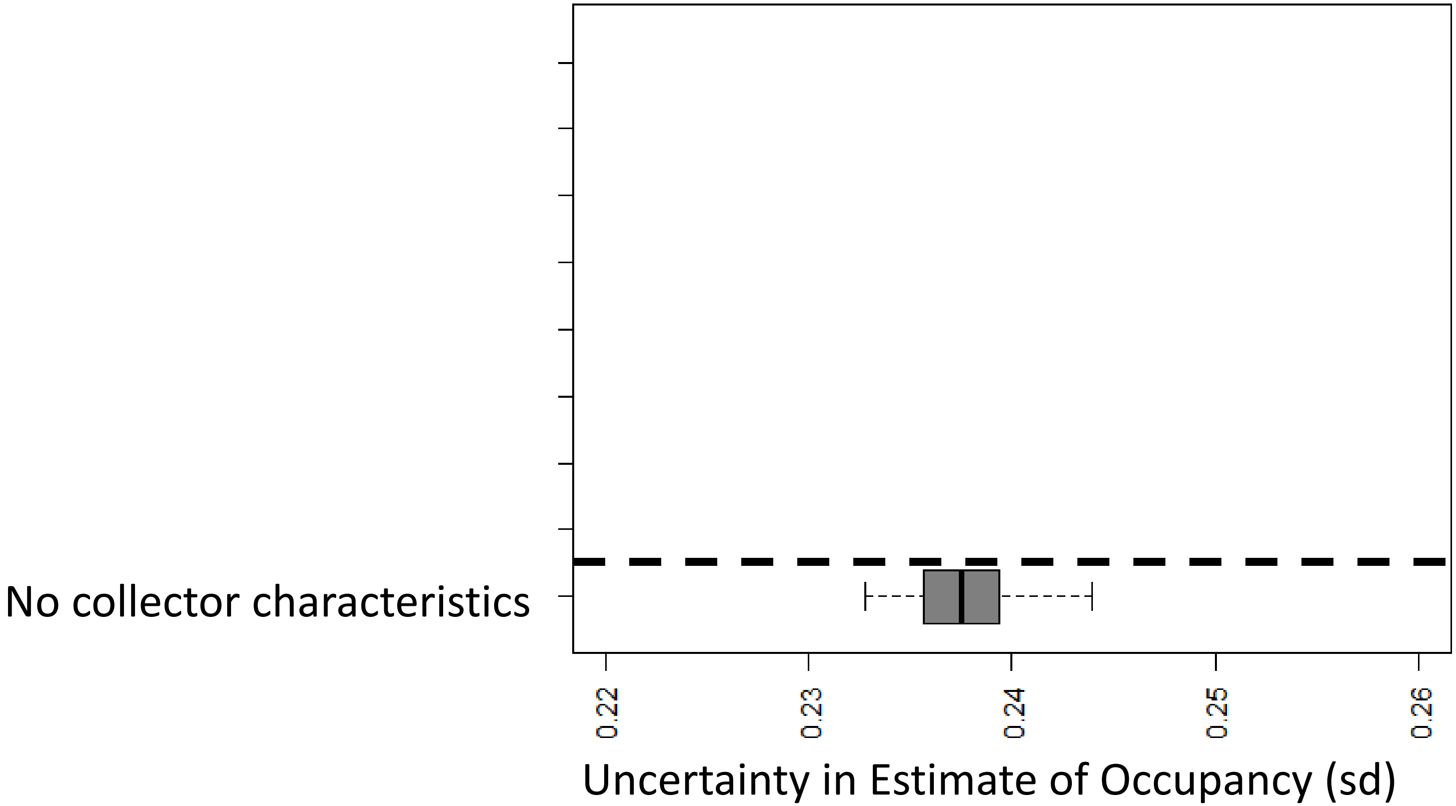

Sisyrinchium xerophyllum Greene
Bruce F. Hansen
University of South Florida
Jan 2003
Herbarium of the University of Florida
Gainesville, Florida 32611 U.S.A.
44

Constructing new R package collectR!

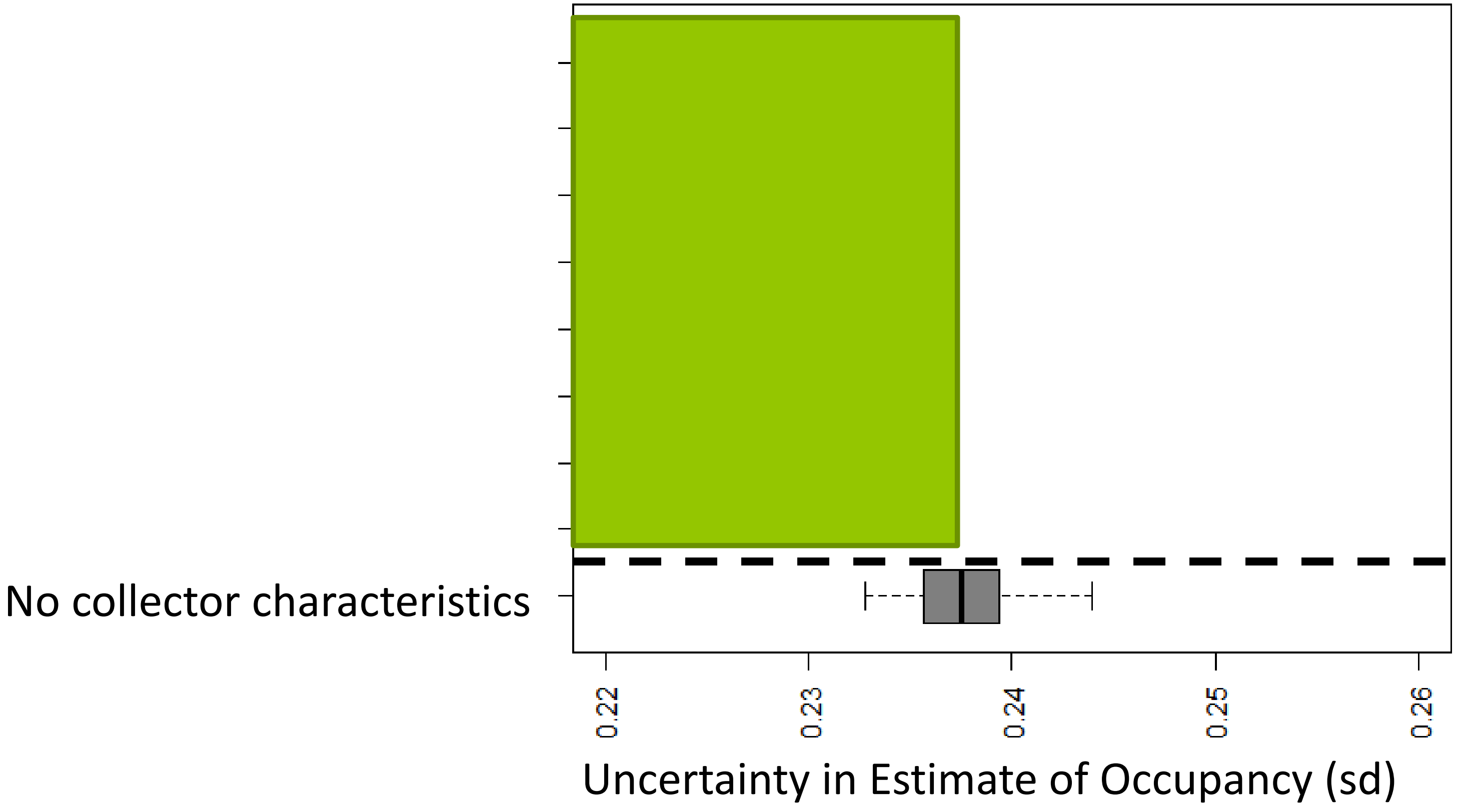
-automate cleaning of specimen data

L. E. Arnold & Erdman West
Erdman West & Lillian Arnold
L. Arnold & Erdman West
L. E. Arnold, Erdman West
Arnold & West
L. E. Arnold; Erdman West
E. West, L. Arnold

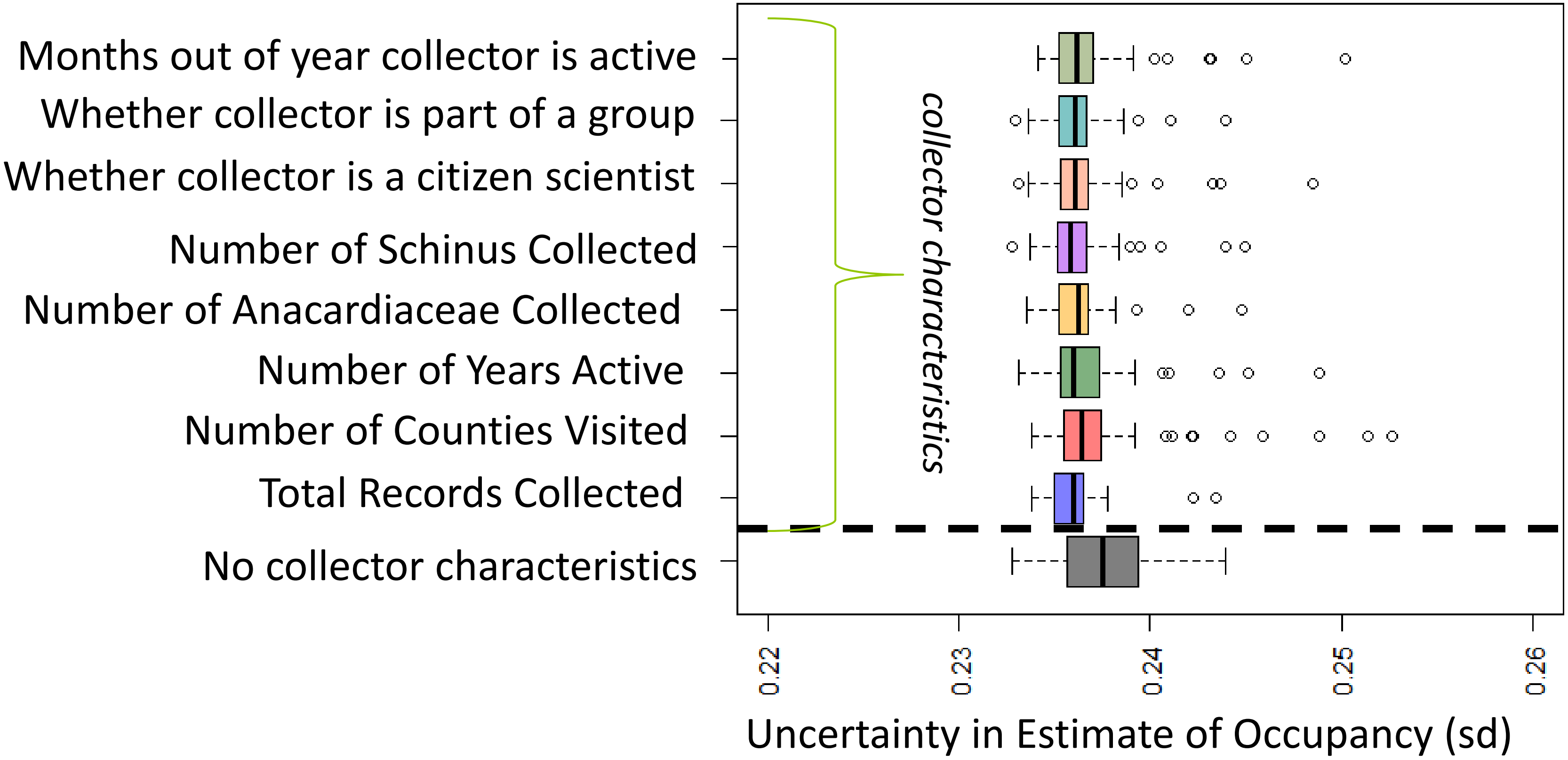
Uncertainty in estimate of occupancy for model without collector behavior



Goal: Shrink uncertainty

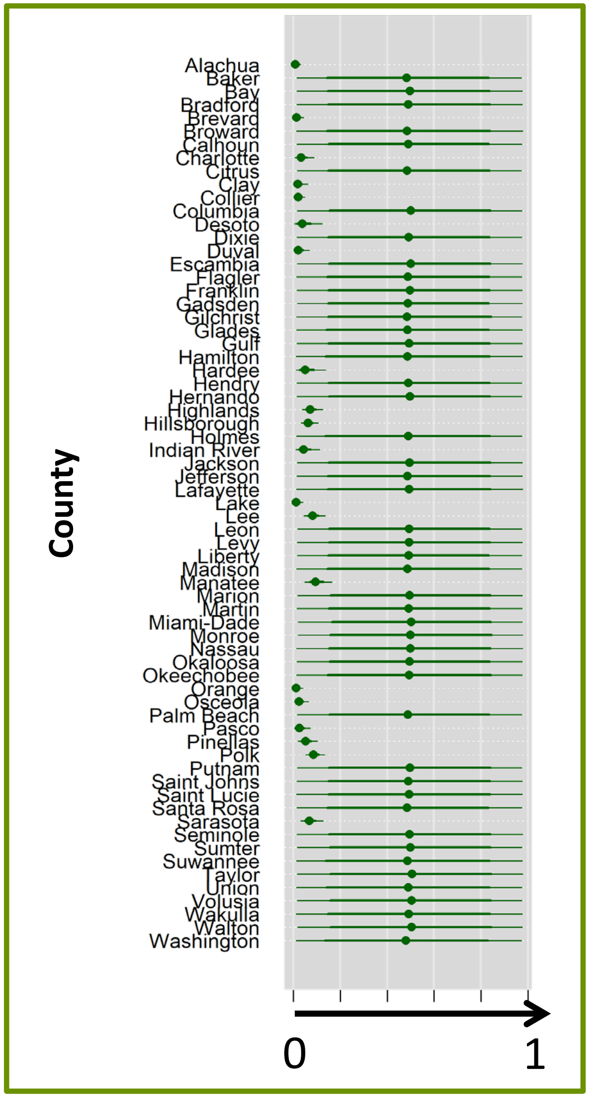


Models that incorporate collector covariates shrink uncertainty in posterior

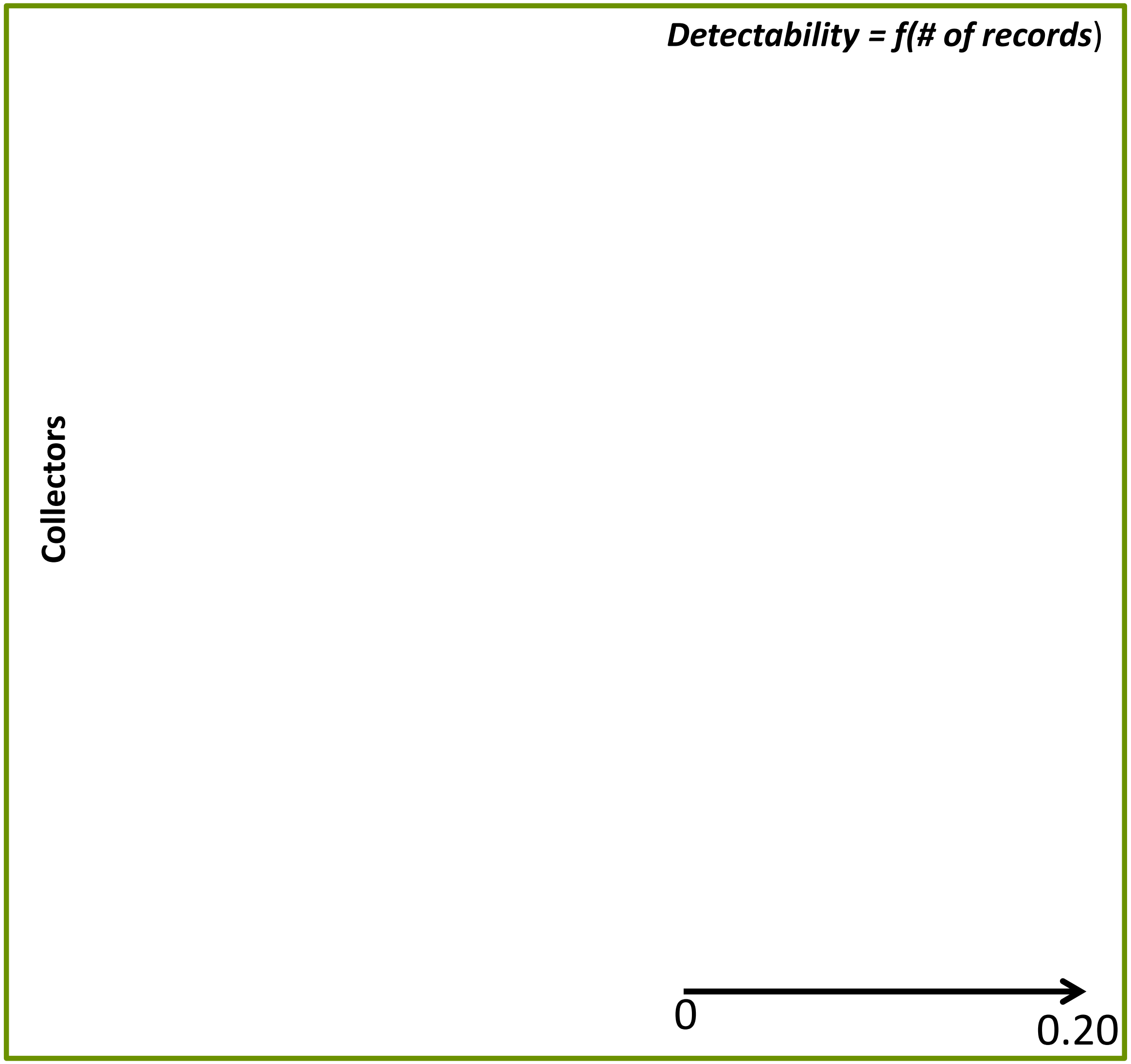


Including collector behavior decreases uncertainty

No collector behavior:



Including collector behavior:



Takeaway

- Accounting for collector behavior improves models
- Standardizing collector name entry important
- Developing new R package -> collectR

Broad-scale efforts to standardize and clean collector covariates are a worthy investment to improve the efficacy of digital biodiversity data for modeling species' ranges.