

What can predictive mapping tell us about the ecology of vector-borne diseases?

Michelle V. Evans

MPE 2013+ Workshop on Global Change and Vector-borne Disease

Mechanistic

- Based on a known relationship between environmental variables and species physiology
- Requires empirical data on physiological responses to temperature (or other environmental variable)
- Often used to extrapolate in the case of climate change or novel environments

Correlative

- Infers a relationship between data and environmental variables (black box)
- Requires species occurrence or disease incidence data, which is generally widely available
- Not appropriate for extrapolation (depending on who you talk to)

Data-driven predictions of vector-borne diseases

- Strong relationship between environmental variables (temperature, precipitation) and mosquito dynamics
- Remotely sensed data is increasing in both temporal and spatial resolutions all the time
- Disease data can be coarse (country or state level)
- Models can be updated in near real-time as new data comes in

Using bagging to deal with sparse datasets

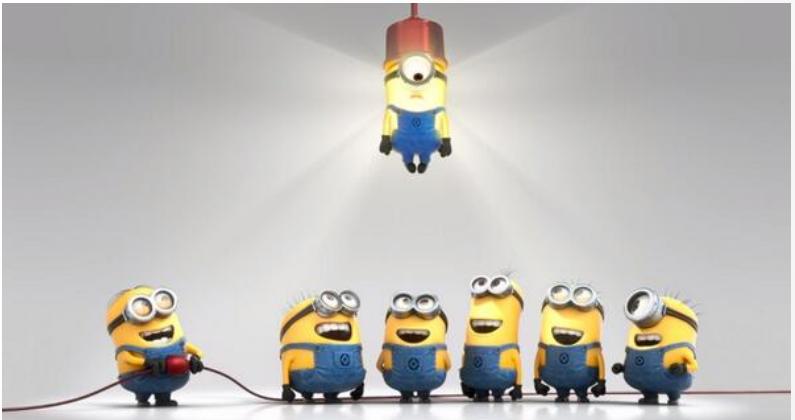
Bagging = bootstrap aggregating

PROS:

Reduces overfitting

Very flexible for non-linear
relationships

Limits bias-variance trade-off



Take your average dataset...

Presence	X₁	X₂	X₃	X₄	X₅
0	1.06	0.1	-2.34	0.27	0.99
1	2.72	-2.09	0.25	-2.86	1.55
1	0.56	1.67	1.83	0.42	-2.86
0	-0.24	0.84	-0.94	2.92	0.04
1	0.02	-0.98	0.34	0.81	0.27

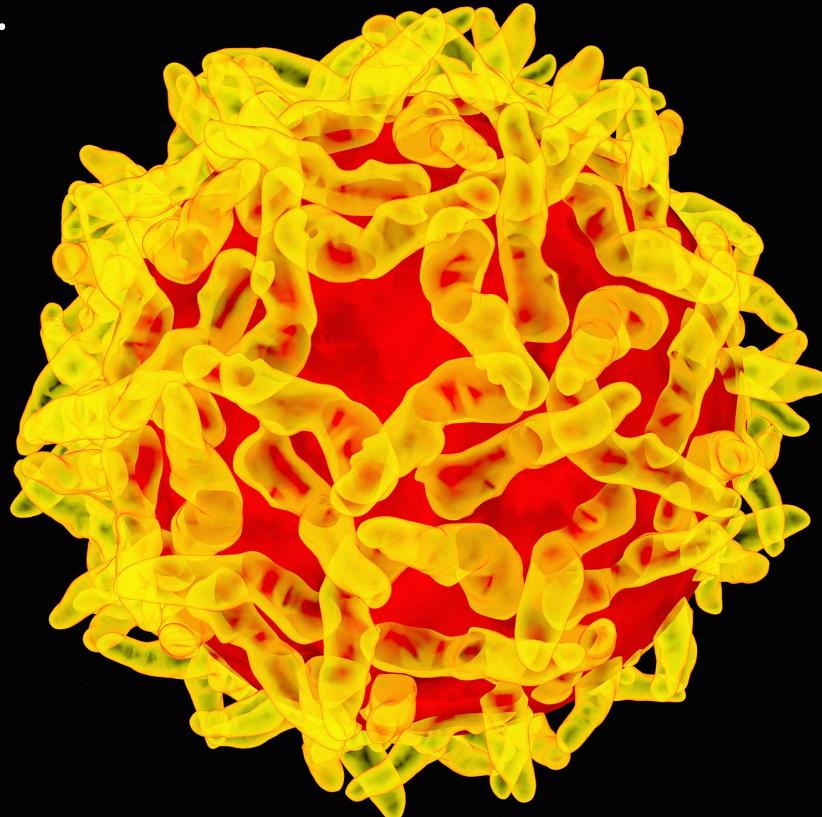
Conventional bagging

Presence	X ₁	X ₂	X ₃	X ₄	X ₅
0	1.06	0.1	-2.34	0.27	0.99
1	2.72	-2.09	0.25	-2.86	1.55
1	0.56	1.67	1.83	0.42	-2.86
0	-0.24	0.84	-0.94	2.92	0.04
1	0.02	-0.98	0.34	0.81	0.27

Range bagging

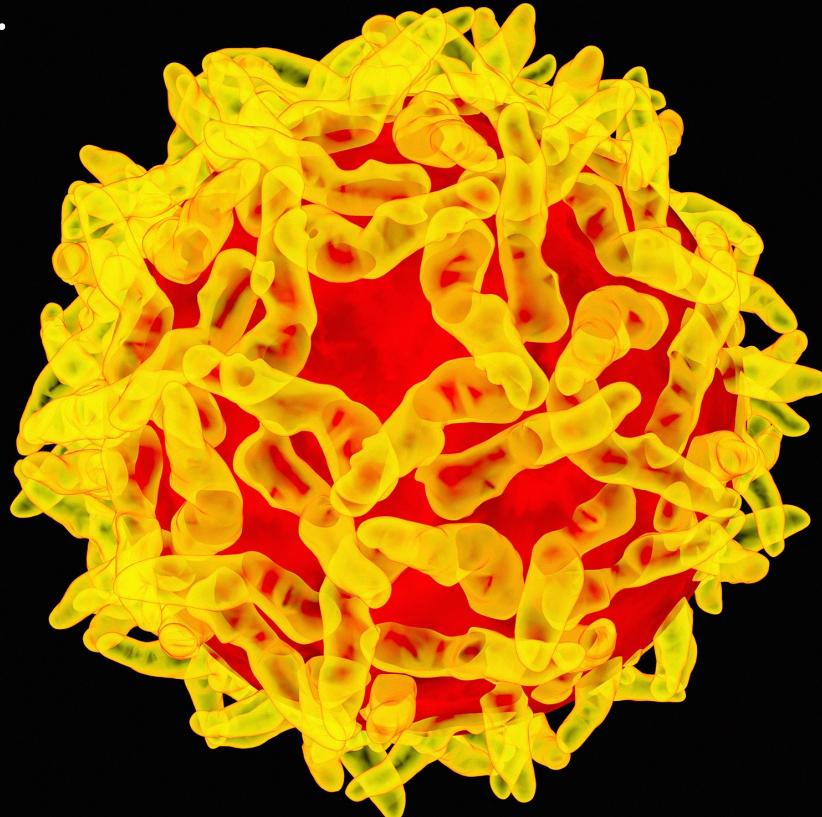
Presence	X ₁	X ₂	X ₃	X ₄	X ₅
0	1.06	0.1	-2.34	0.27	0.99
1	2.72	-2.09	0.25	-2.86	1.55
1	0.56	1.67	1.83	0.42	-2.86
0	-0.24	0.84	-0.94	2.92	0.04
1	0.02	-0.98	0.34	0.81	0.27

Bagging in action...



for yellow fever

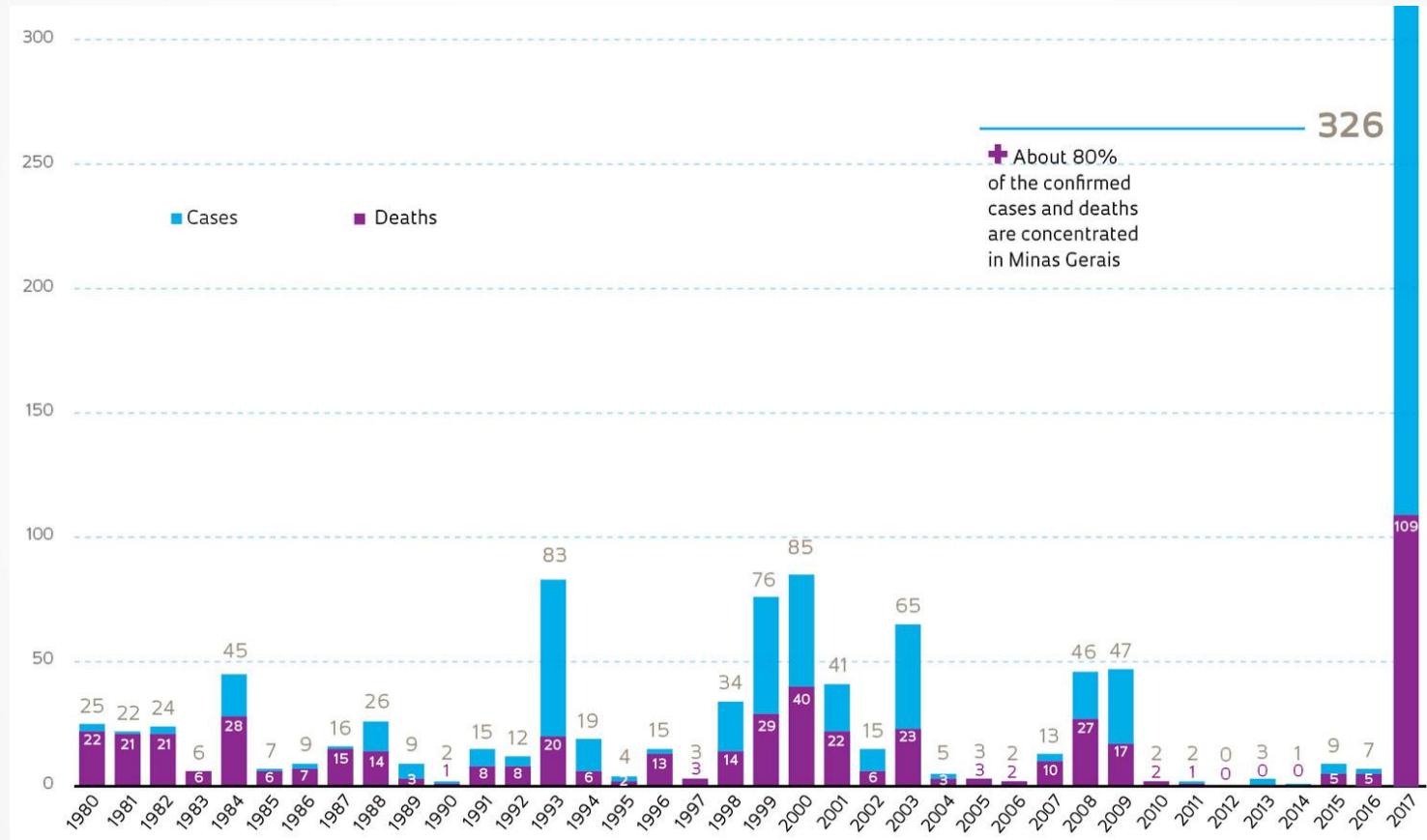
Bagging in action...



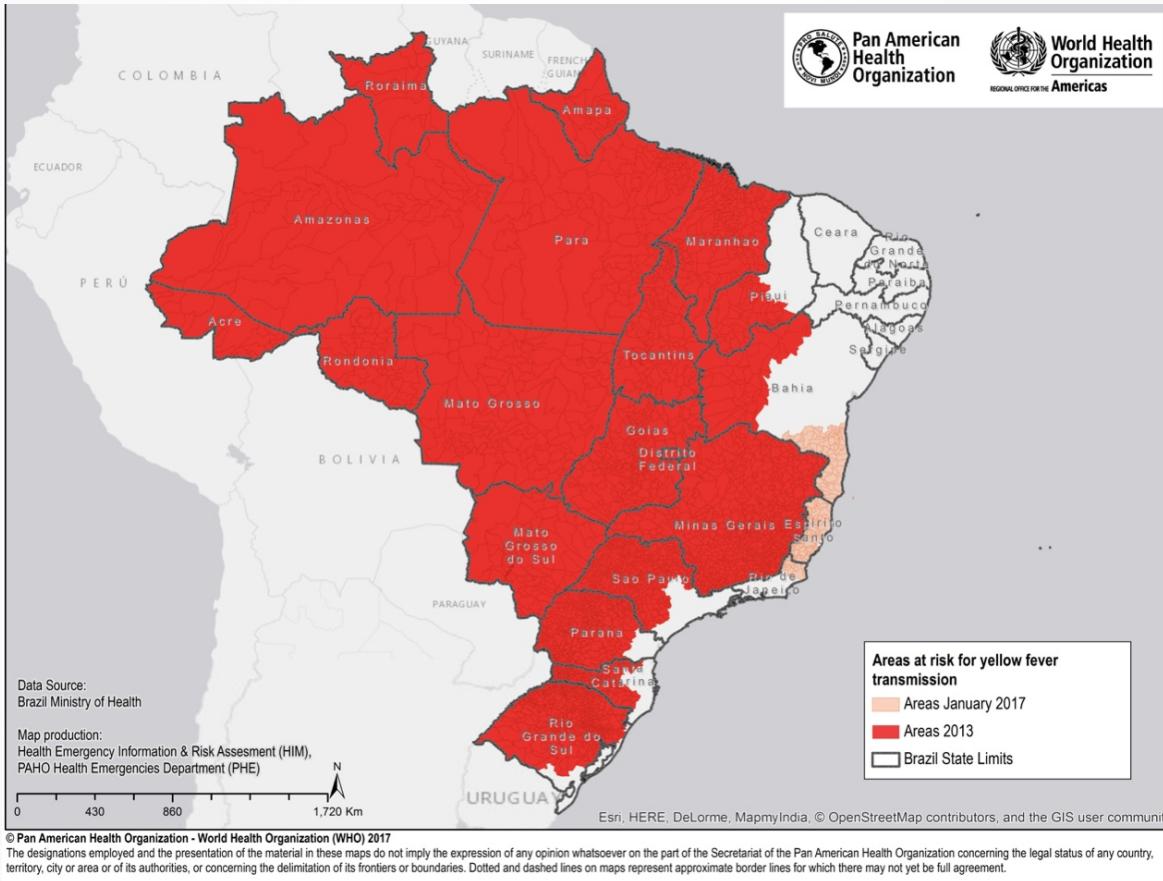
Reni Kaul

for yellow fever

Yellow fever outbreak of 2017 in Brazil



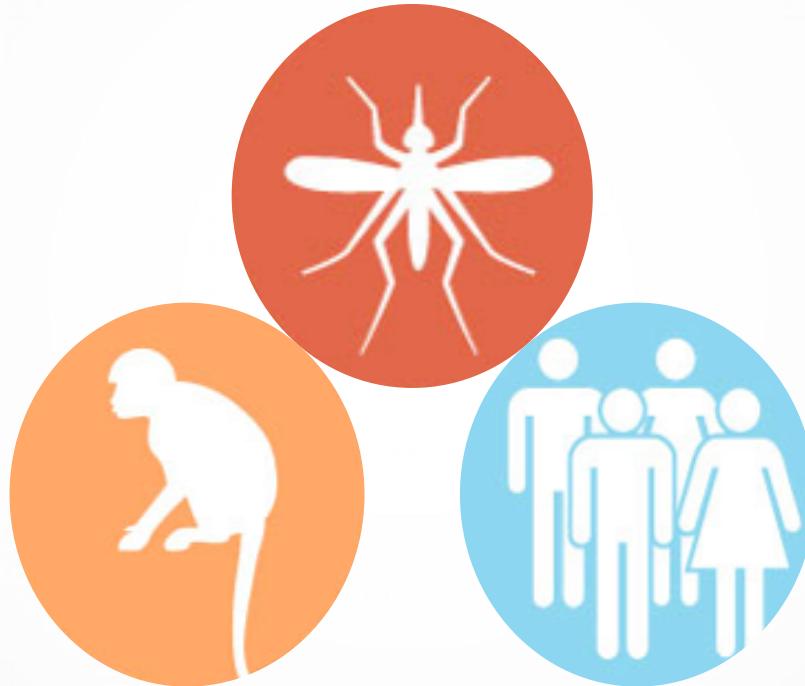
Yellow fever outbreak of 2017 in Brazil



What is the monthly risk of yellow fever
spillover across Brazil?
Can we predict it?

What are the drivers of spillover events?

Covariate creation based on *a priori* knowledge



Covariate creation based on *a priori* knowledge



Temperature
Rainfall
Vector Ranges



NDVI
Primate Ranges



Population Density
Fire Density
Agricultural – NHP Overlap

Covariate creation based on *a priori* knowledge



Temperature
Rainfall
Vector Ranges



NDVI
Primate Ranges



Population Density
Fire Density
Agricultural – NHP Overlap

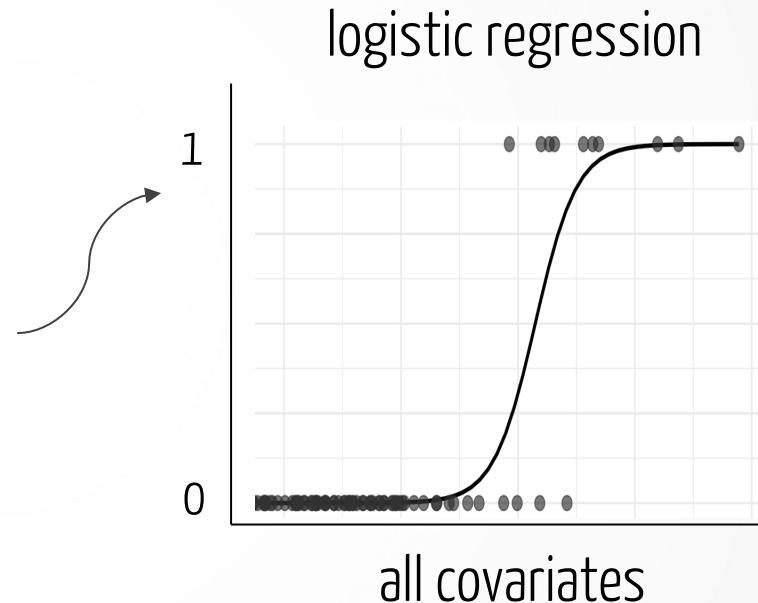
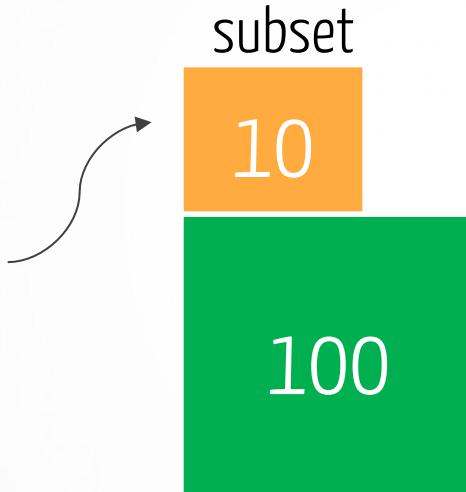
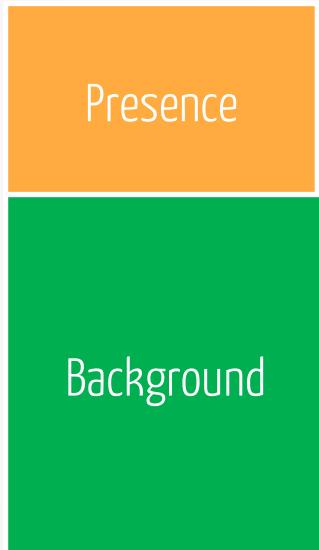
Environmental anomalies that trigger spillover

Temperature
Rainfall
NDVI
Fire Density



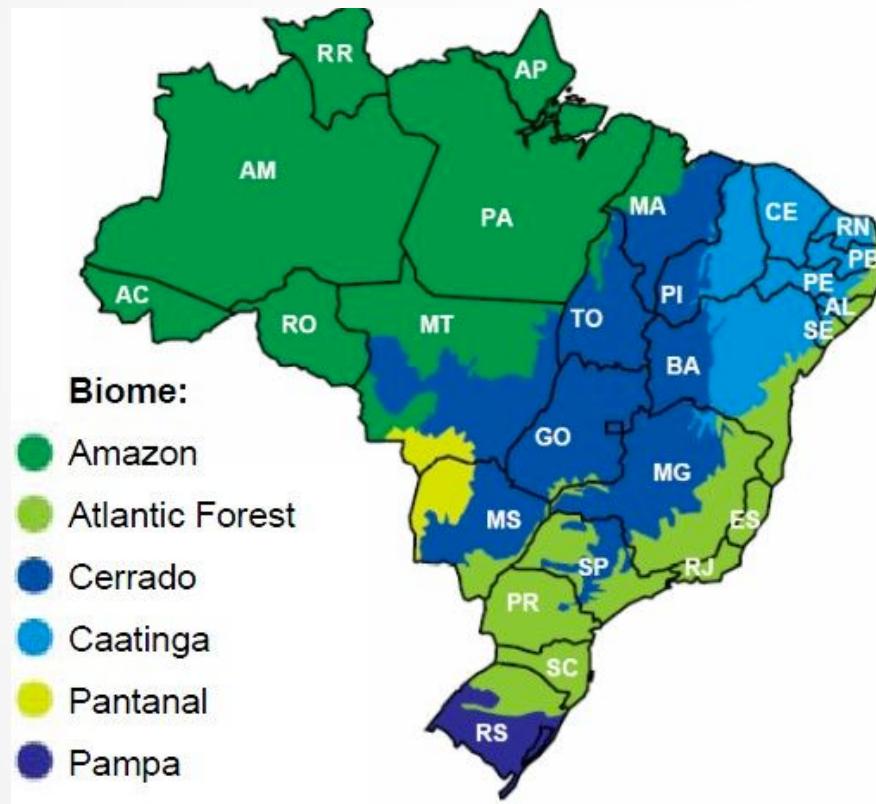
Scaled to maximum for
that calendar month

Bagged logistic regression

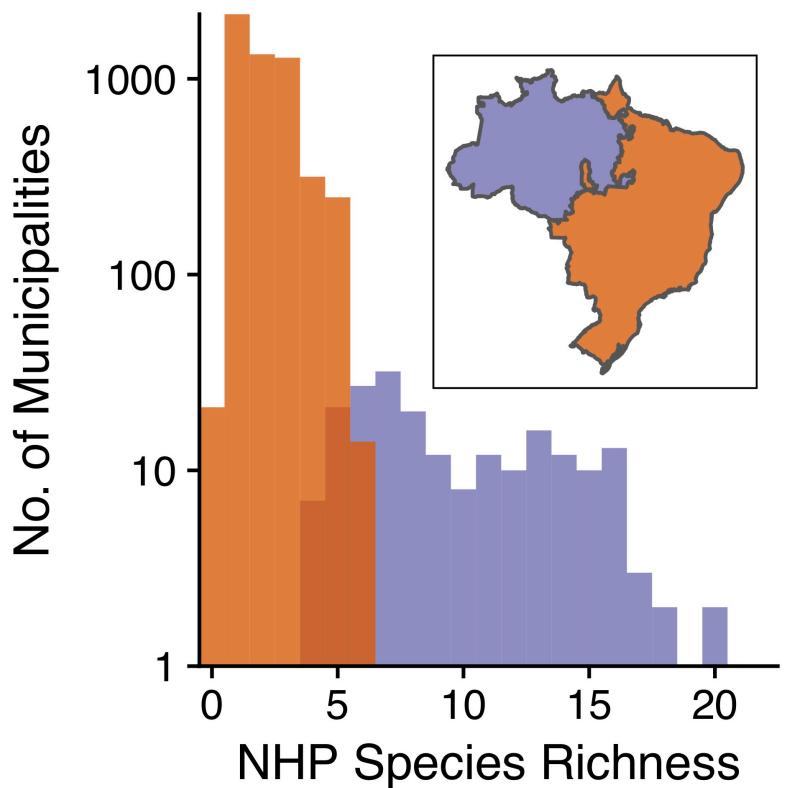
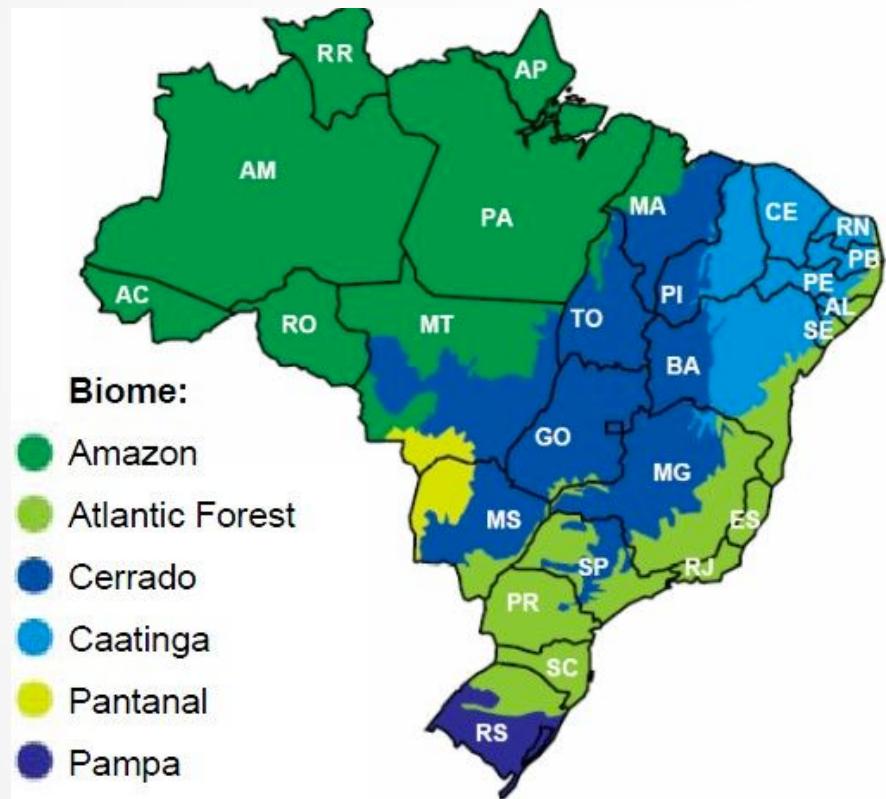


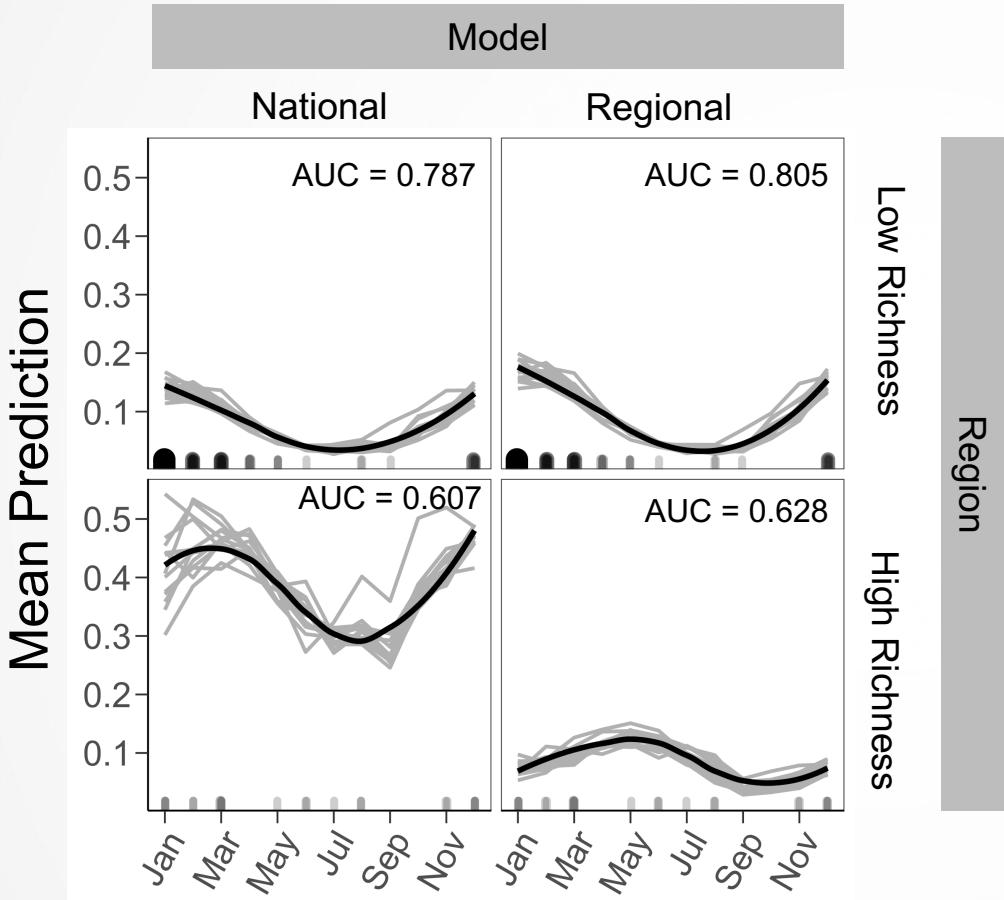
Final model has 500 bags

Multiple models to explain regional processes



Multiple models to explain regional processes

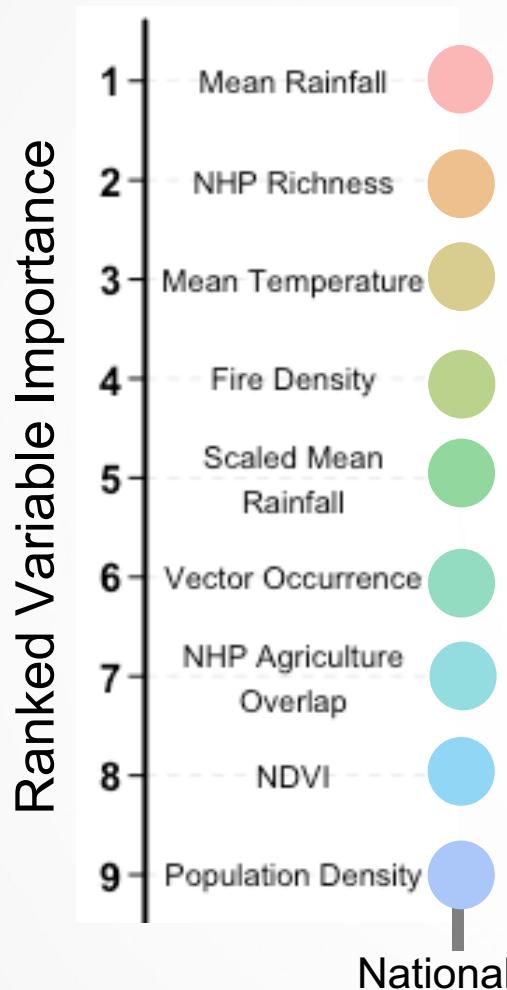


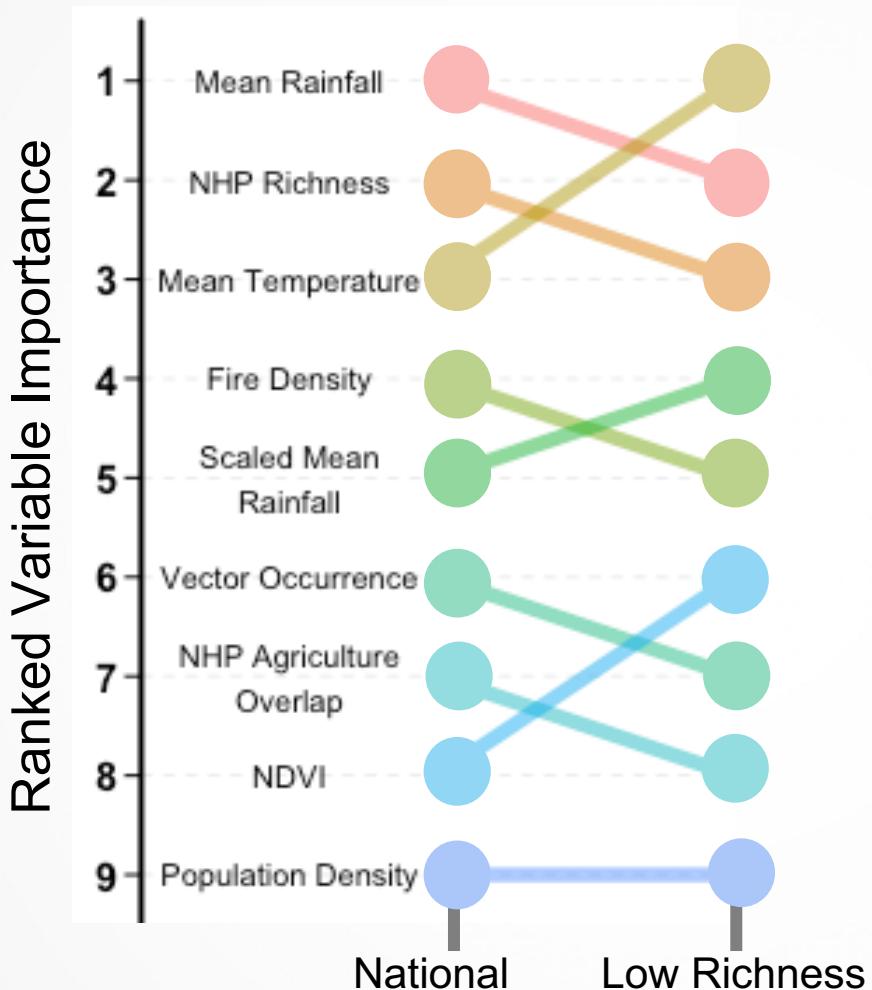


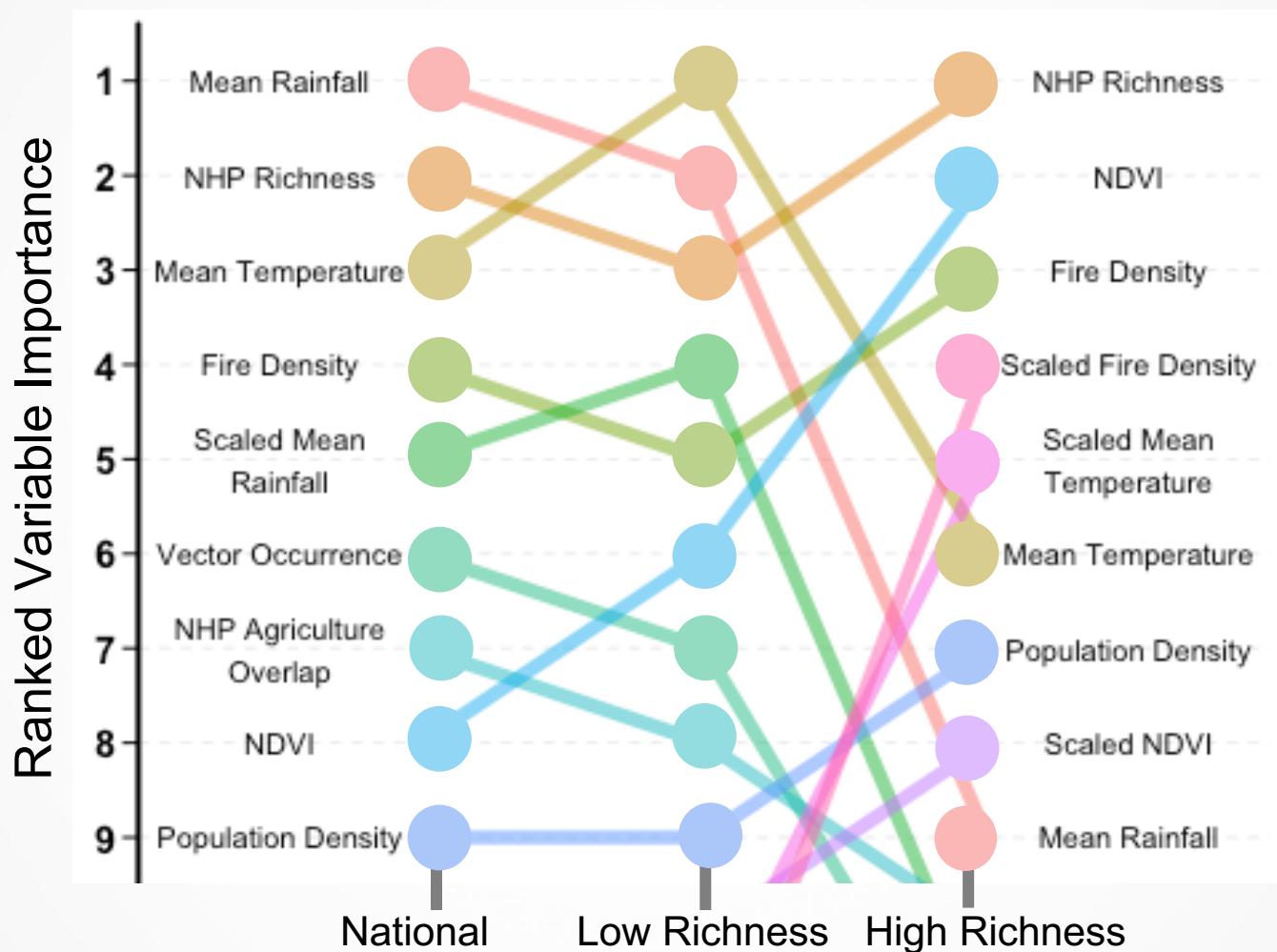
Low richness region nearly identical to national model

Seasonal trends differ across models

Regional model performs marginally better

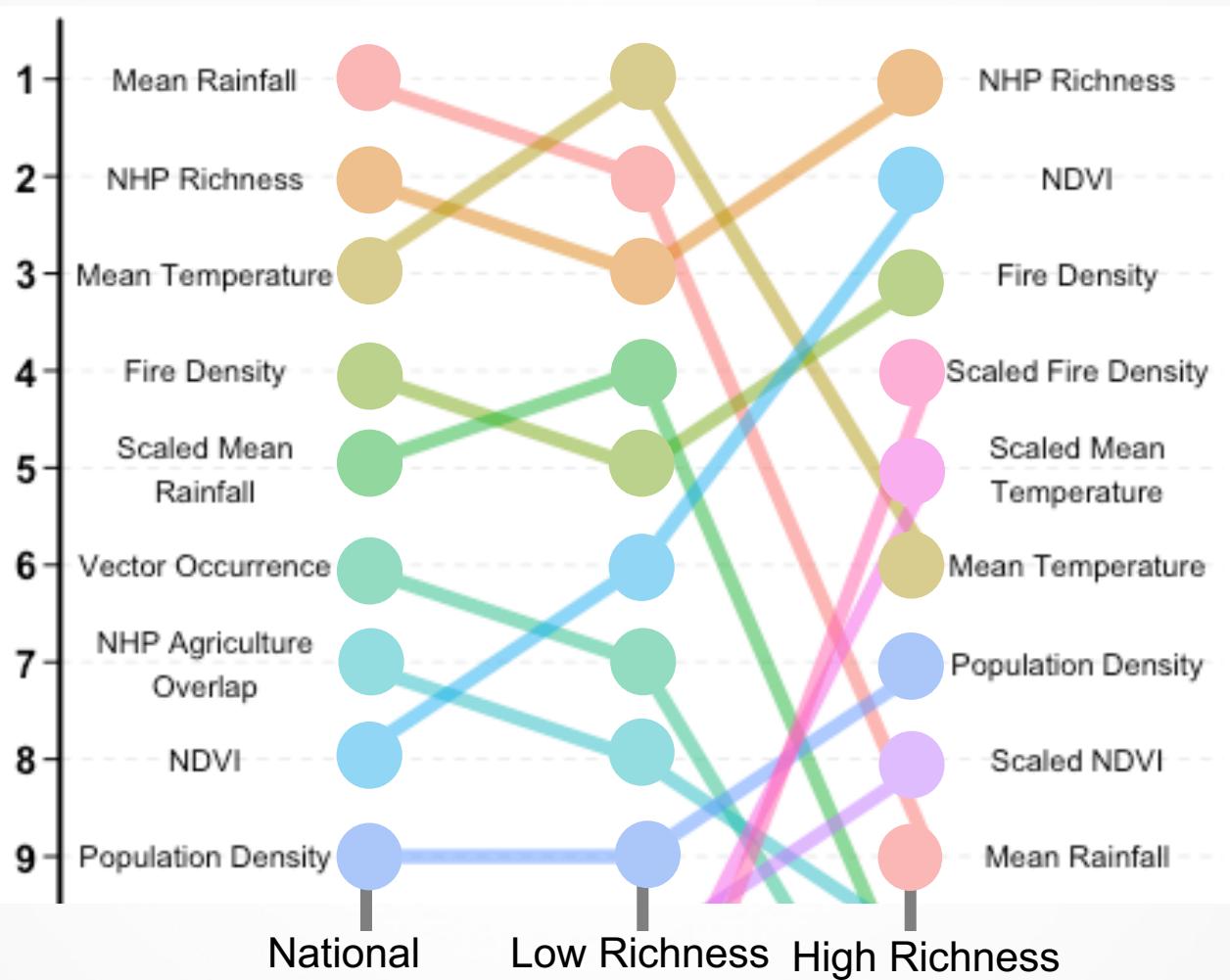






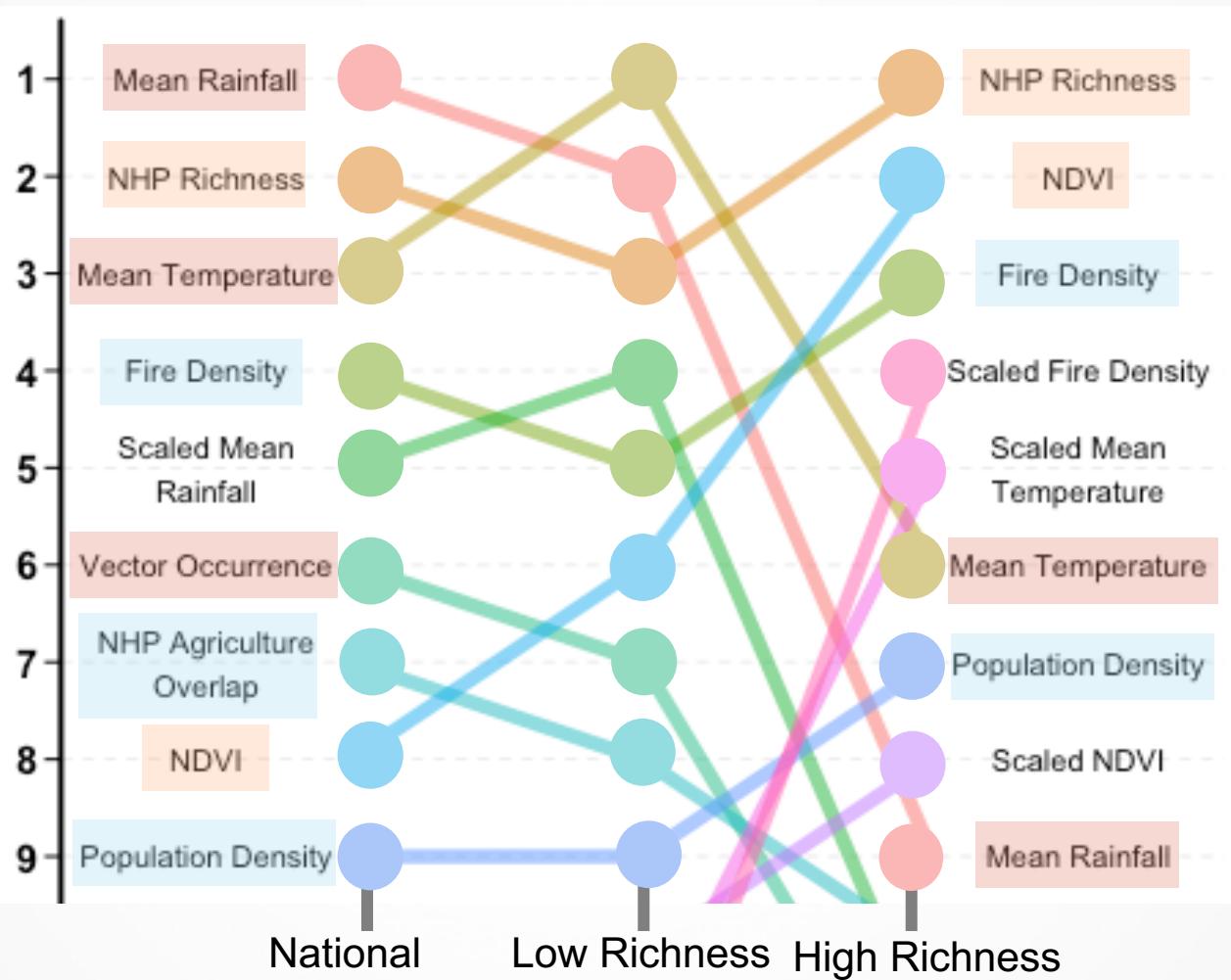


Ranked Variable Importance



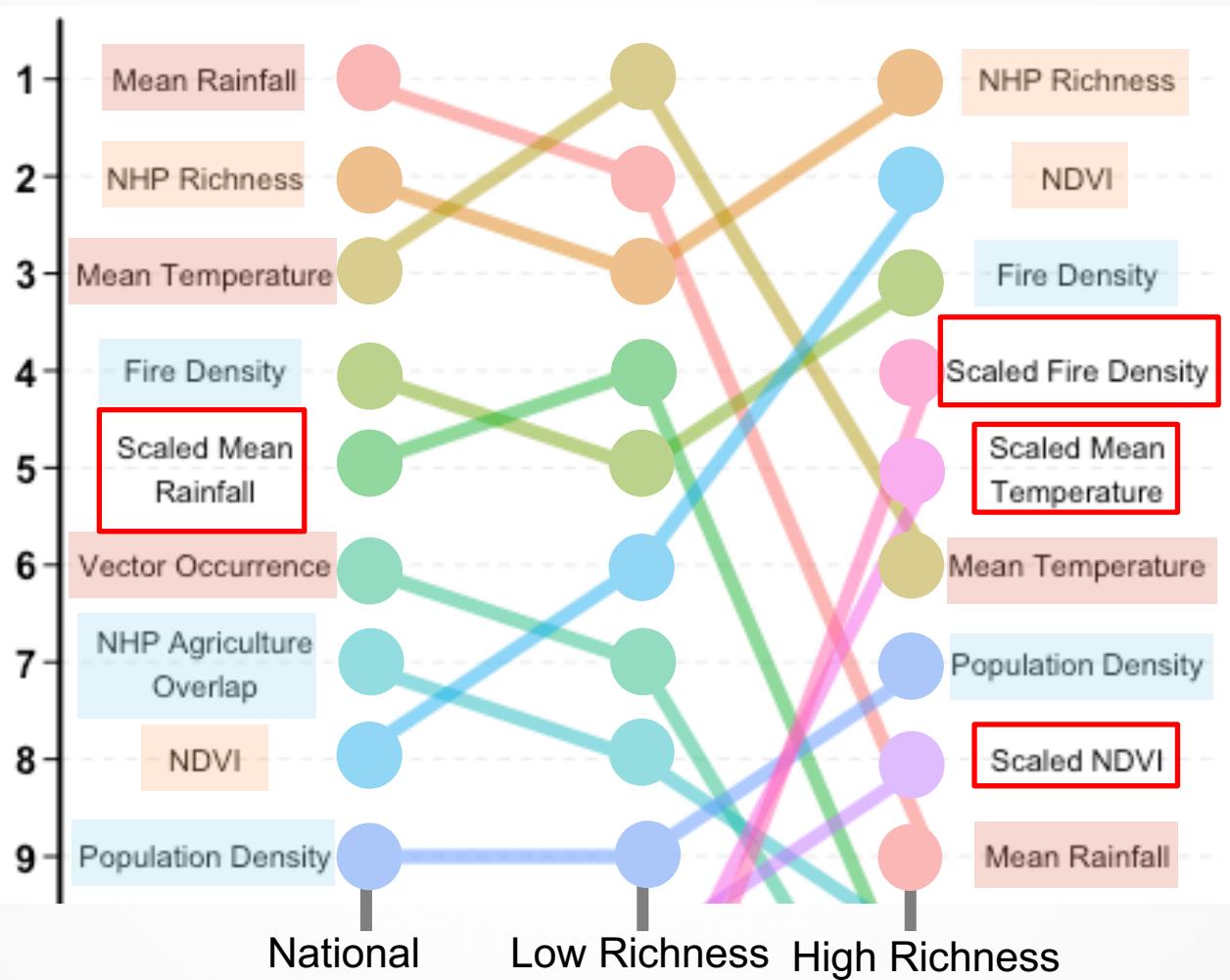


Ranked Variable Importance





Ranked Variable Importance



What did we learn?

- One model may not fit all over a large and varied geographic area
- Drivers of yellow fever differ in the Amazon basin and more populated coastal regions
 - Spillover in the Amazon is ‘triggered’ by encroachment events and environmental anomalies
 - Spillover in Southeastern region is driven by mosquito population dynamics
- Correlative models can work in tandem with empirical work and mechanistic models, highlighting future areas of study

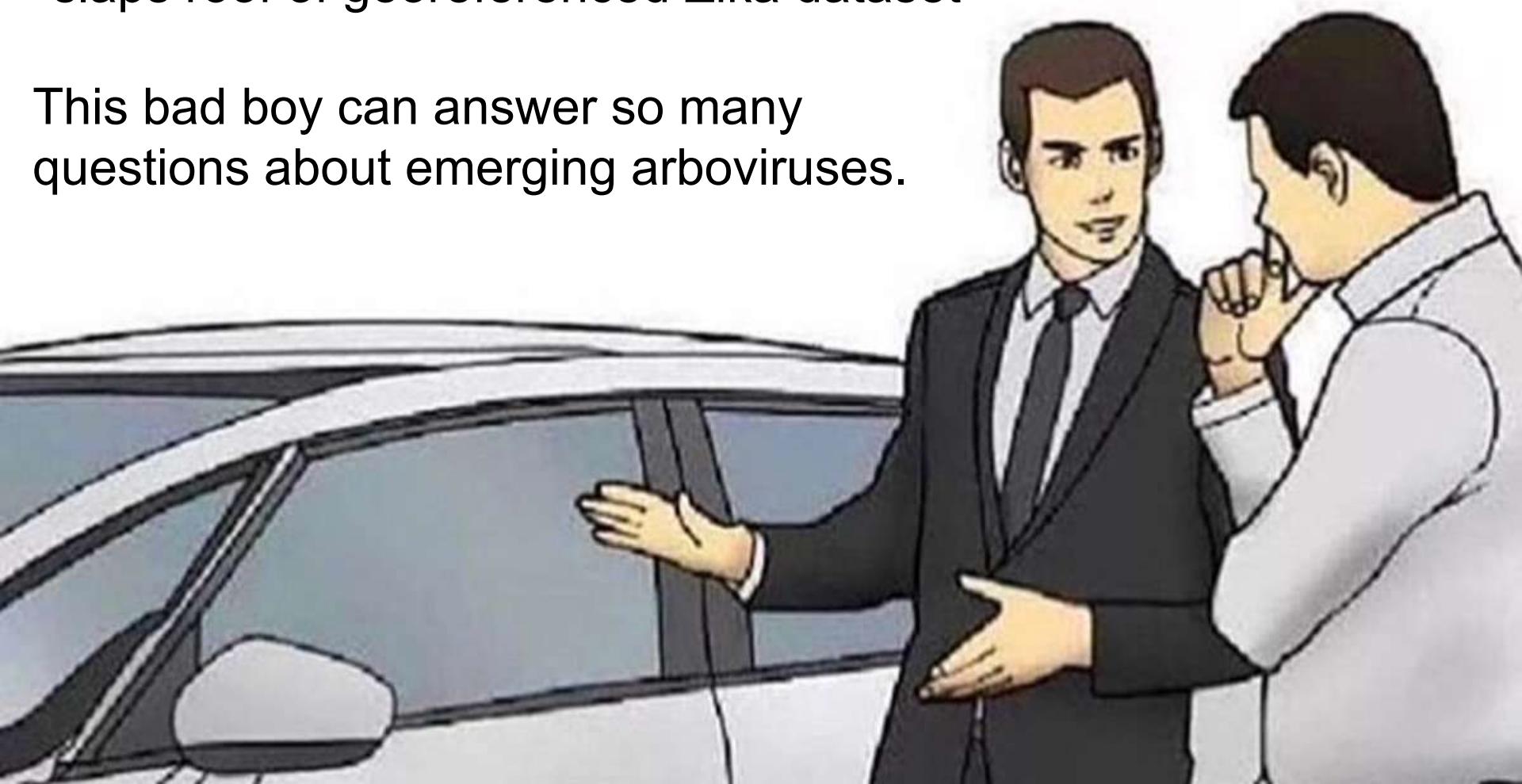
I got 99 problems and data are all of them...

- Data wasn't easily available (relatively), and only at municipality level
- Missing fine-scale data on vaccination
- Issue of scale-mismatch between environmental and disease data (temporal and spatial)

These can be even worse for other diseases in countries with less health reporting infrastructure

slaps roof of georeferenced Zika dataset

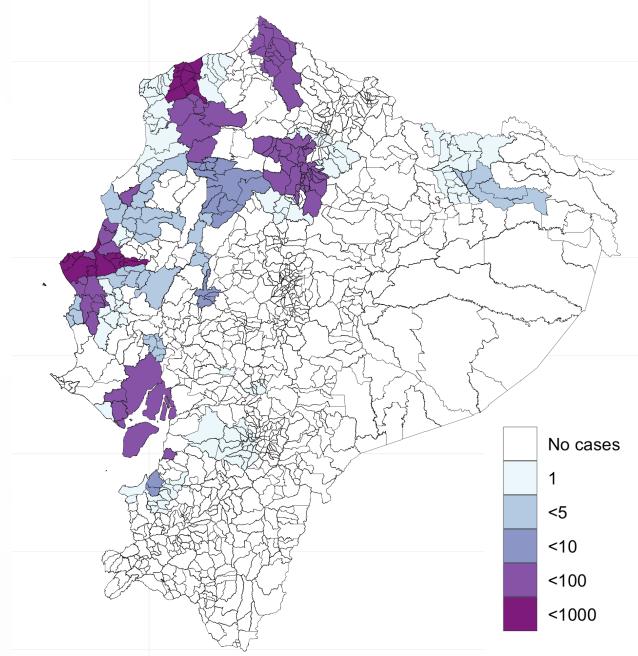
This bad boy can answer so many questions about emerging arboviruses.



Crowd-sourced Zika data for S. and Central America

- Organized by CDC researchers
- Data collected from Ministry of Health and PAHO documents
- Date range: 2016 – Present
- Fifteen countries
- Currently in the process of georeferencing
- One issue: lack of standardization across countries

Cumulative cases in Ecuador



Check it out: github.com/cdcepi/zika