

# An Open Science Framework for Research on Cyanobacteria in Lakes and Ponds

US EPA, Region 7

Jeff Hollister, Farnaz Nojavan, Betty Kreakie, Stephen Shivers, and  
Bryan Milstead

2017-10-11

Lenexa, KS

# Twitter?



hashtag: #AAG2017

me: @jhollist

Who, what, why, and how?

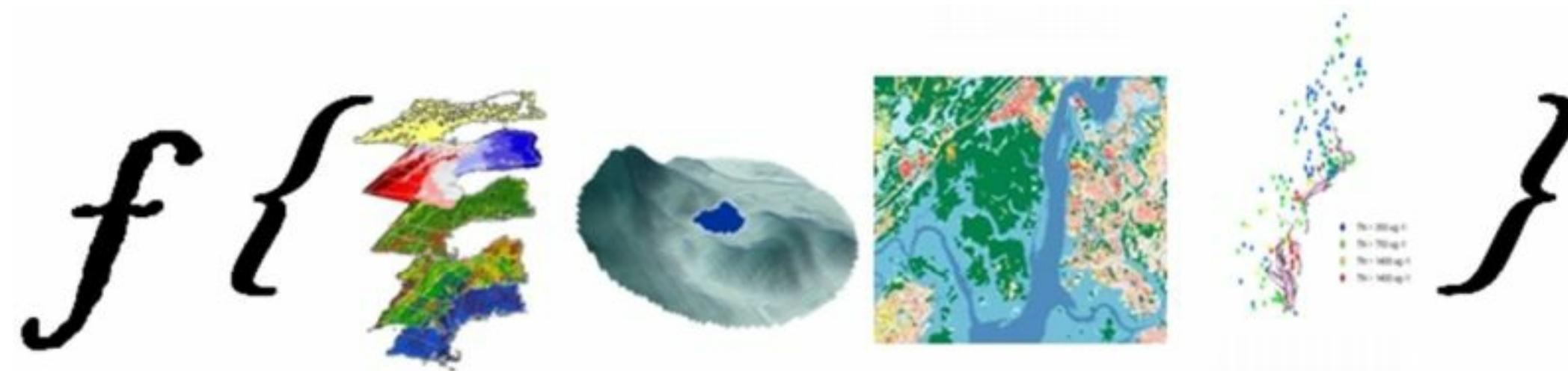
# Who are we?

- Ecologists
- Computational focus
  - Enough to be dangerous
- 3 FTE
  - Myself
  - Betty Kreakie
  - Bryan Milstead
- 2 Post-docs
  - Farnaz Nojavan
  - Stephen Shivers



# What do we do?

- Apply computational approaches to understand water quality impacts in lakes
  - Modelling (Not this talk!, but see Farnaz's talk)
- Open Science



# What is open science?

- Access to materials
- Reproducible/ Repeatable
- The Web!
- A process, not a state



# Why open science?

- Often required
  - Government/Funders/Journals
- Benefits researchers
  - [McIernan et al. \(2016\) How open science helps researchers succeed](#)
- Improves quality
  - [The classic example: Reinhart and Rogoff](#)
- Benefits to society
  - ["Sharing of Data Leads to Progress on Alzheimer's"](#)



# How are we open?

- R package development
- Visualization
- Sharing and collaborating
- Publishing
- Open data (not in this talk)



# Packages

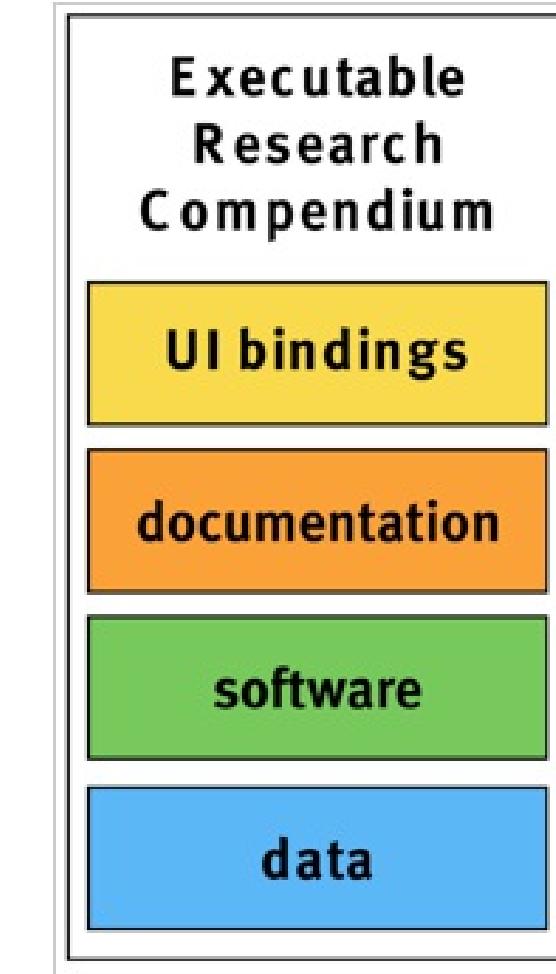
# Why Packages

- Useful structure
- Infrastructure for sharing
  - GitHub
  - CRAN
- We are an R shop!



# Research Compendia

- Define
- Origins
  - [Gentleman and Lang \(2004\)](#)
- Part of
  - Reproducible Research
  - Literate Programming (ala Donald Knuth)
- ROpenSci efforts
  - [rrrpkgs](#)
  - [ROpenSci unconf 2017 discussion](#)



from Nüst, Konkol, et al (2017),  
<https://doi.org/10.1045/january2017-nuest>

# Packages as Research Compendia

- R, Data, and Vignettes folders
- Other examples
  - [Carl Boettiger's template](#)
  - [Ben Marwick](#)
- Our examples
  - <https://github.com/usepa/LakeTrophicModelling>
  - <https://github.com/usepa/Microcystinchla>
- GitHub and Zenodo (Archive)

The image shows two side-by-side browser windows. The left window is a GitHub repository page for 'USEPA / Microcystinchla'. It displays a list of 91 commits, 2 branches, 2 releases, and 3 contributors. The right window is a Zenodo archive page for the same package. It shows a summary of the data, including the title 'Associations between chlorophyll a and various microcystin health advisory concentrations' by Hollister, Jeffrey W. & Kreakie, Betty J., and a detailed abstract about microcystin health impacts and advisory levels. Both pages include links to download files like 'Microcystinchla 2.0.tar.gz'.

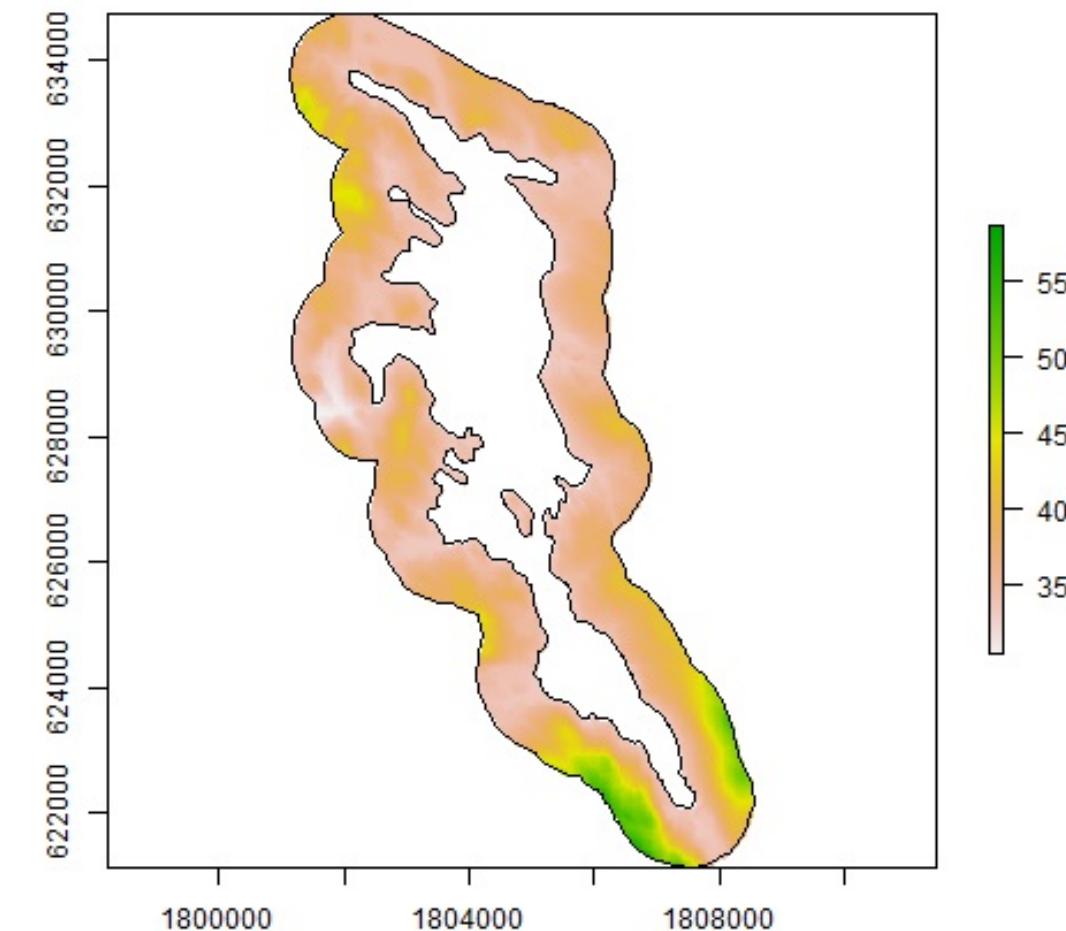
# Packages to solve common problems

- lakemorpho
- elevatr
- goatscape (in development)



# lakemorpho

- Lake morphometry metrics in R
- Version 1.0
  - August 2014
- Version 1.1.0
  - December 2016
- sf support to be added
- [National Lake Morphometry](#)
- [Hollister and Milstead \(2010\)](#)
- [Hollister \*et. al.\* \(2011\)](#)
- [Hollister and Stachelek \(2017\)](#)



Package URL: <https://cran.r-project.org/package=lakemorpho>

lister.com:8787



Ecology Divis Altmetric it! People Plus MightyText Setting up Logitech The Master Ice Crea ORD Application Pro

New Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Environment History Git

Import Dataset

Global Environment

## Values

exampleElev	Large RasterLayer (111930 elements, 885.9 Kb)
exampleLake	Formal class SpatialPolygonsDataFrame
inputLM	Large lakeMorpho (6 elements, 1.5 Mb)

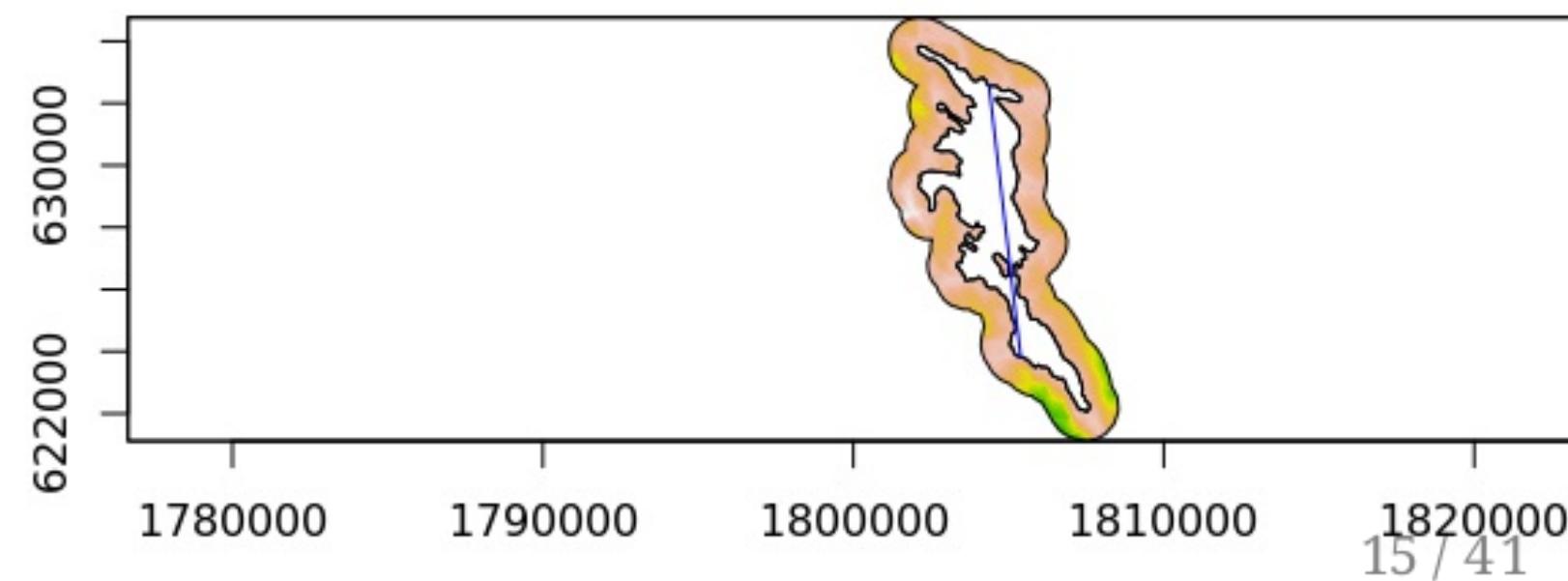
```
rectFactor = 0.553)
rectFactor = 0.553)
ointDens = 100, addLine = TRUE)
```

line

R Script

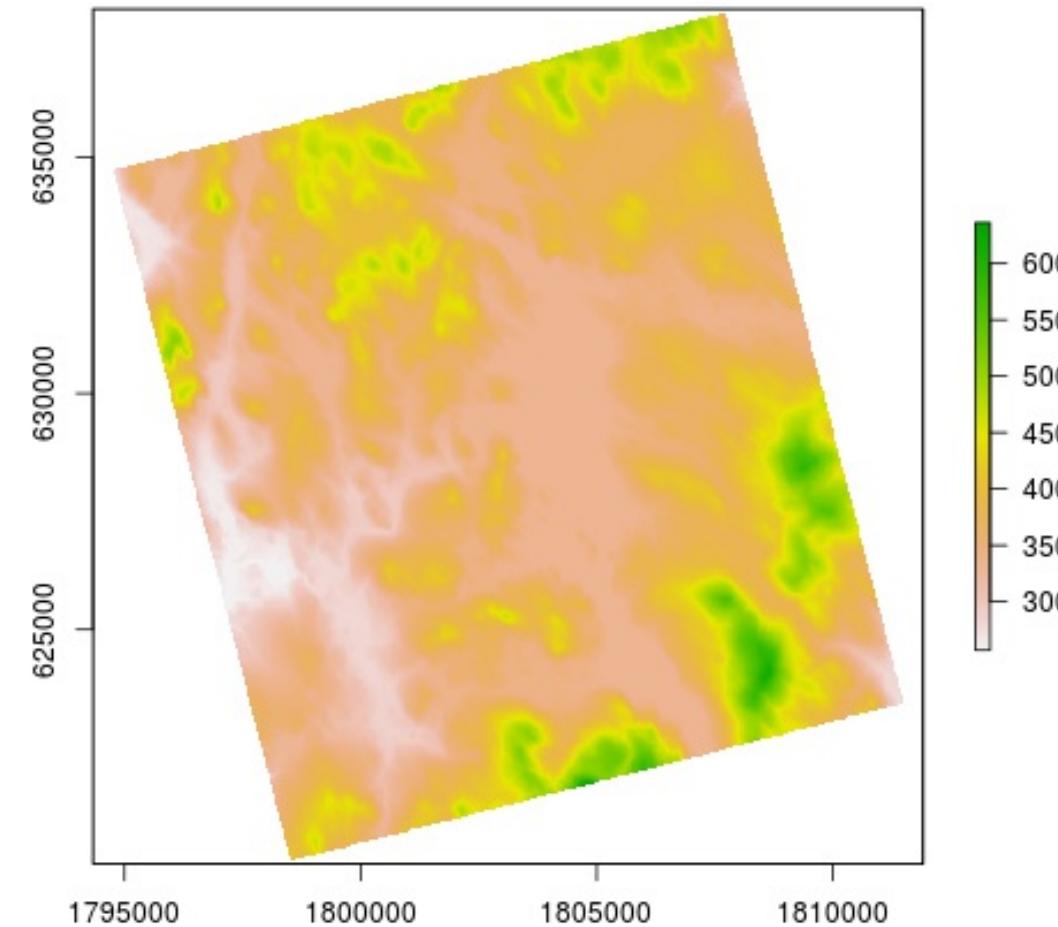
```
"inputLM"
rectFactor = 0.553)
rectFactor = 0.553)
ointDens = 100, addLine = TRUE)
```

lakemorpho::demo



# elevatr

- Access elevation data in R
  - Mapzen
  - AWS
  - USGS
- Version 0.1.1
  - January 2017
- Version 0.1.3
  - March 2017
- Will be paired with `lakemorpho`
- `sf` support to be added



Package URL: <https://cran.r-project.org/package=elevatr>

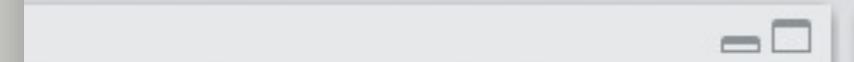
lister.com:8787



Ecology Divis Altmetric it! People Plus MightyText Setting up Logitech The Master Ice Cream ORD Application Pro

File Plots Session Build Debug Profile Tools Help

Go to file/function Addins



Environment History

Import Dataset

Global Environment

## Data

pt\_df 5 obs. of 2 variables

## Values

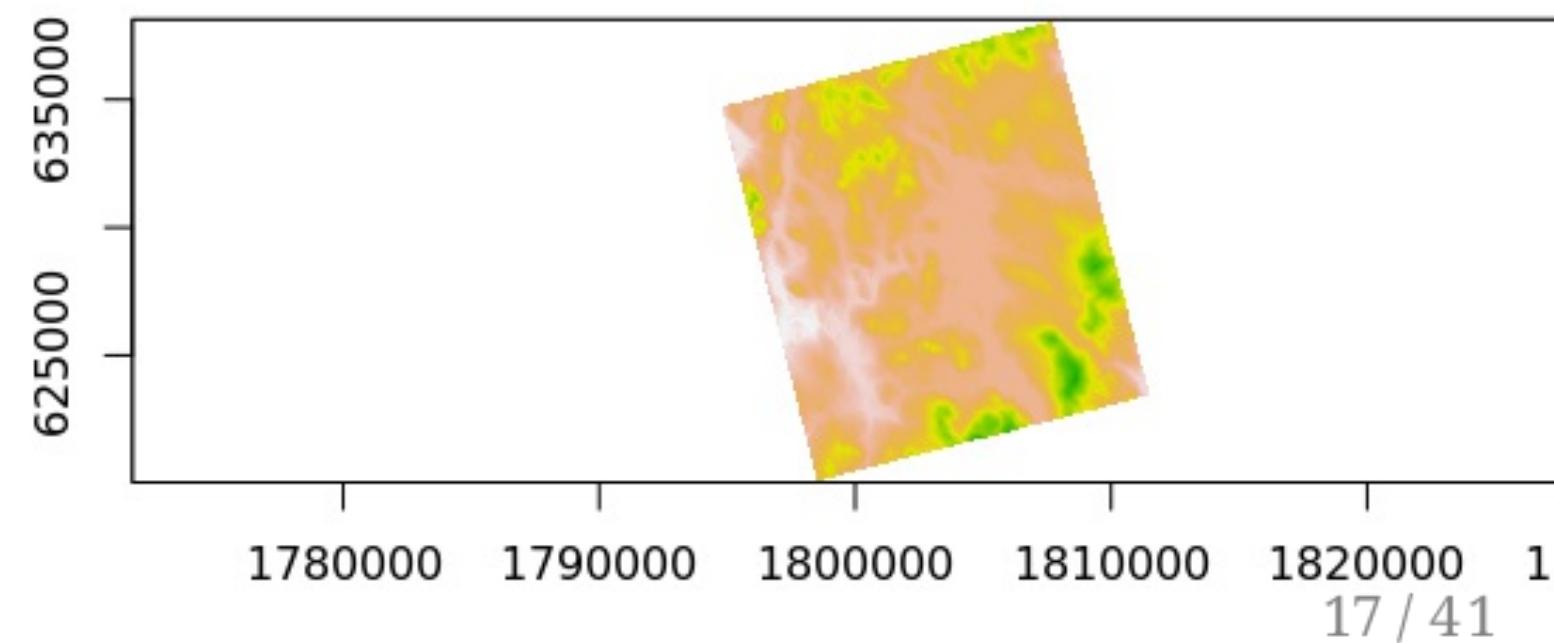
lake Formal class SpatialPolygonsDataFrame

DEM

raster(lake, z = 12, src = "aws")

R Script

```
514 (nrow, ncol, ncell)
())
, 620036.4, 638140.2 (xmin, xmax, ymi
=20 +lat_2=60 +lat_0=40 +lon_0=-96 +x_0
+no_defs +ellps=GRS80 +towgs84=0,0,0
.n, max)
```



# goatscape

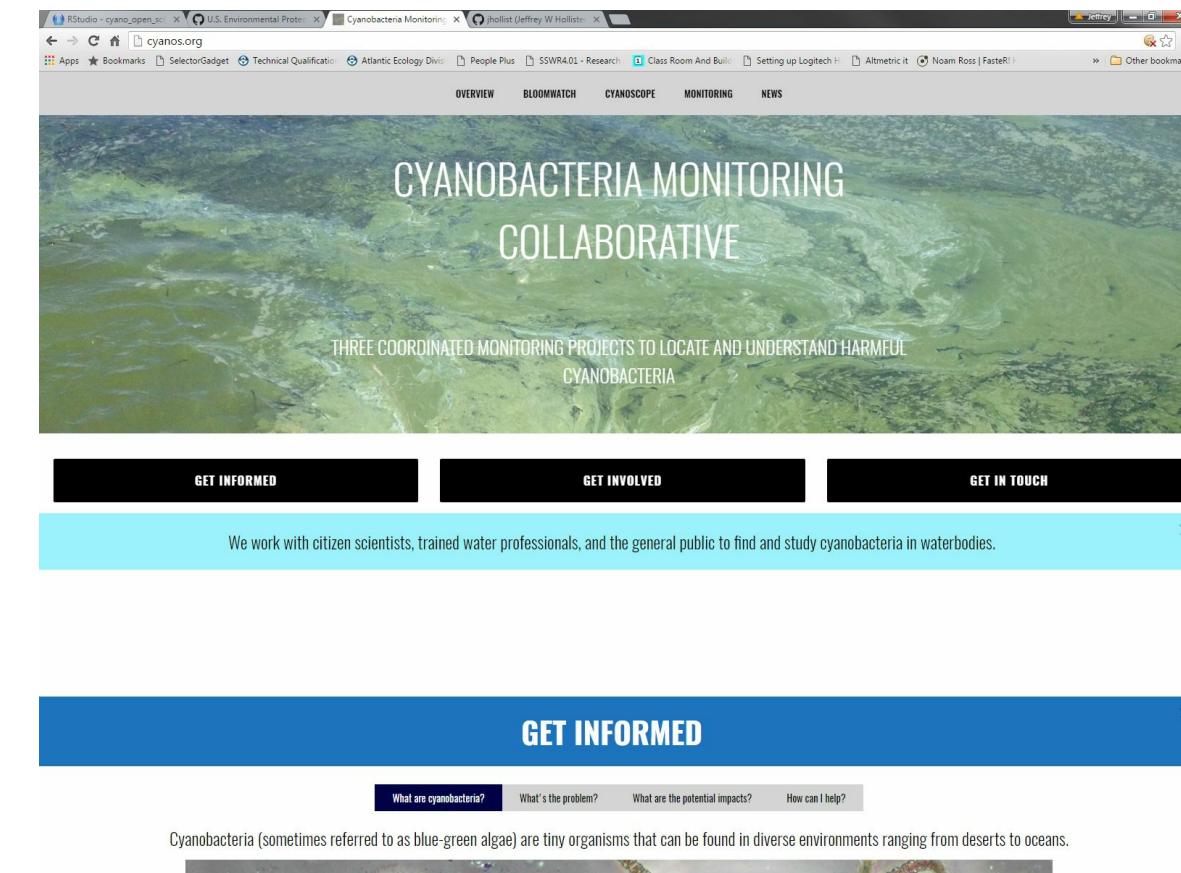
- New effort with Bryan Milstead
- What's in a name?
- Summarizes ancillary data for a user-defined landscape polygon
  - Census (via `censusapi`)
  - Landscover
  - Impervious
- Accepts arbitrary spatial data for the landscape
- Based on `sf` and `tidy` by design
- <https://github.com/usepa/goatscape>



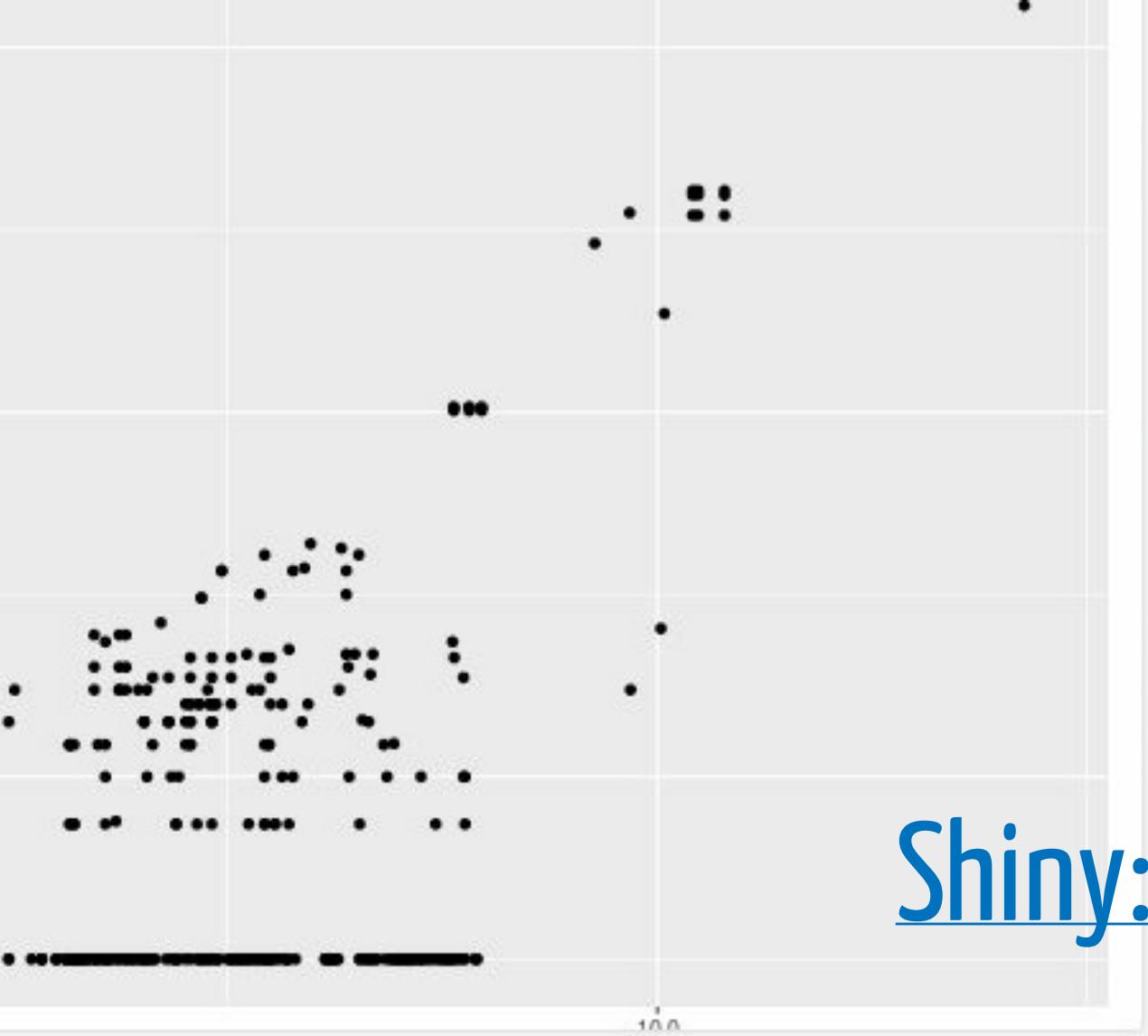
# Data Visualization

# Shiny: Cyanobacteria Monitoring Collaborative

- Started in 2013
  - New England Region Cyanobacteria Monitoring Workgroup
- Three Projects
  - bloomWatch
  - cyanoScope
  - Monitoring
- DataViz with Shiny

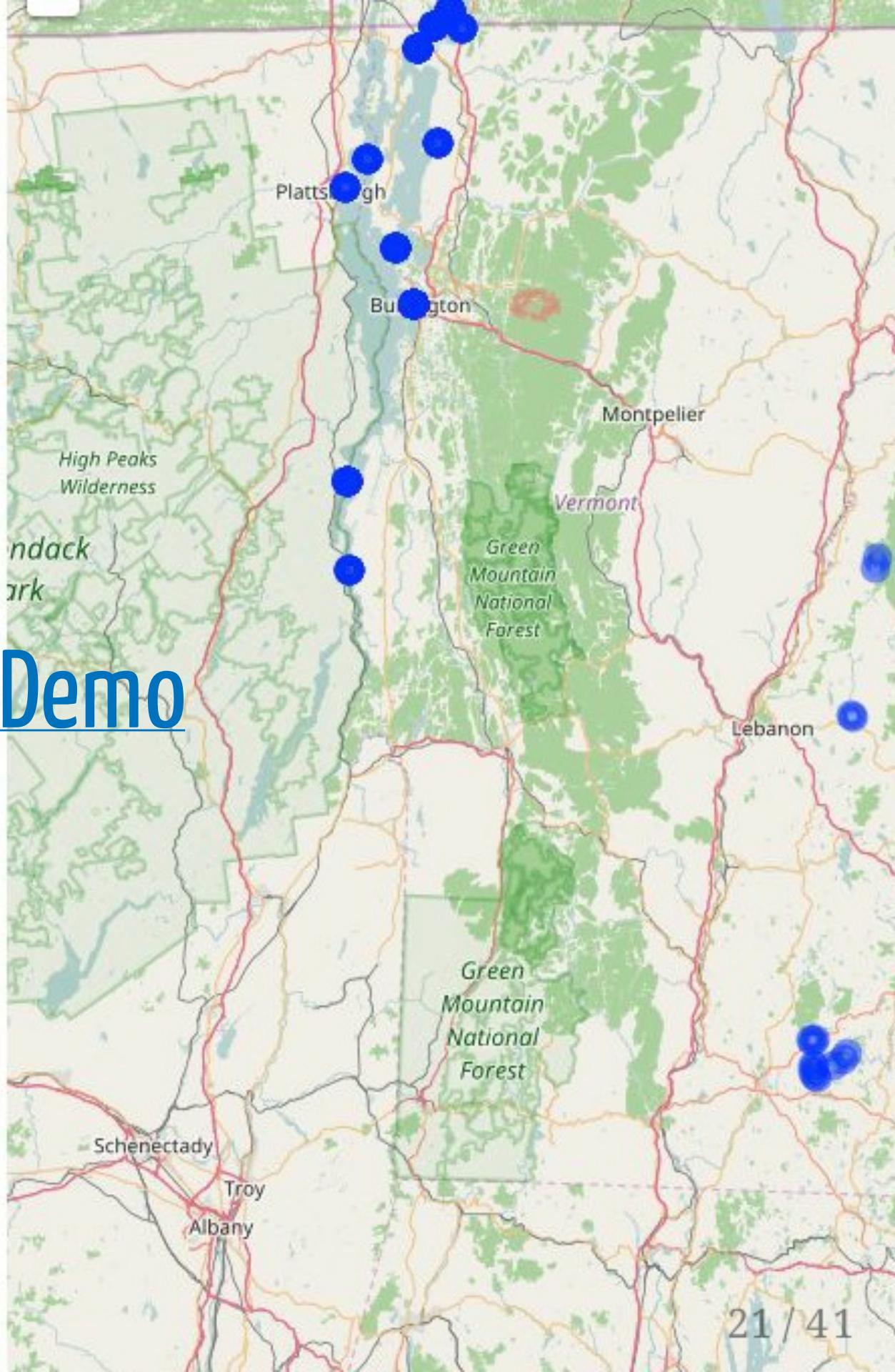


Project URL: <http://cyanos.org>



Date	Chlorophyll	Phycocyanin
2014-09-03	71.37	16998.17
2014-07-10	2.18	0.10
2014-07-17	2.44	1.52
2014-08-08	3.17	0.10
2014-08-08	3.57	0.10
2014-08-08	3.22	0.10

## Shiny: Demo



# Sharing and Collaborating

# GitHub

- What is it?
- How do we use it?





# Hollister

Open Science at

al Protection A...

.gov

com



### quickmapr

An R package for quickly mapping and navigating spatial data

● R ★ 44 ⚡ 6

### elevatr

An R package for accessing elevation dat

● R ★ 33 ⚡ 4

### rmd\_word\_manuscript

rmd to docx: draft manuscript

● TeX ★ 17

# GitHub: Demo

### USEPA/lakemorpho

ORD lakemorpho

● R ★ 8 ⚡ 7

### ropensci/lawn

turf.js R client

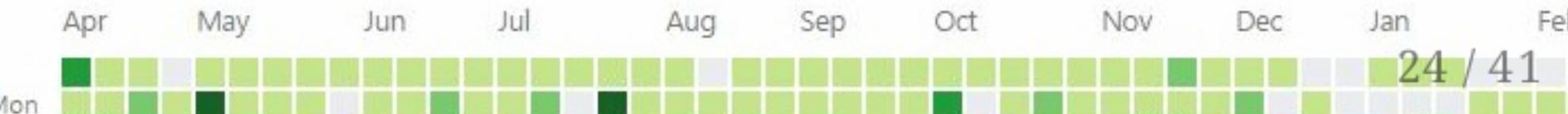
● R ★ 42 ⚡ 8

### manuscriptPackage

Template for writing manuscripts as an R

● R ★ 30 ⚡ 6

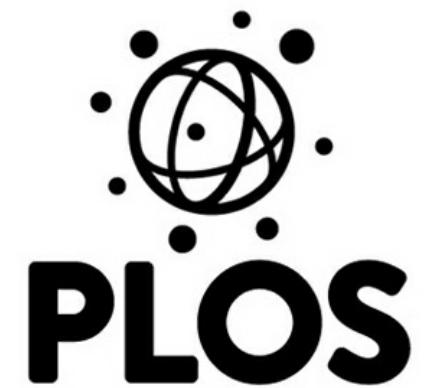
1,876 contributions in the last year



# Open Access

# Publishing

- Preprints
  - [Hollister et al. \(2016\) PeerJ Preprints](#)
- Open first
  - [Milstead et al. \(2013\) PLoS One](#)
  - [Hollister and Kreakie \(2016\) F1000Research](#)
- Money where our mouth(s) is(are)
  - [Kreakie et al. \(2015\) LakeLines](#)



# Research efforts

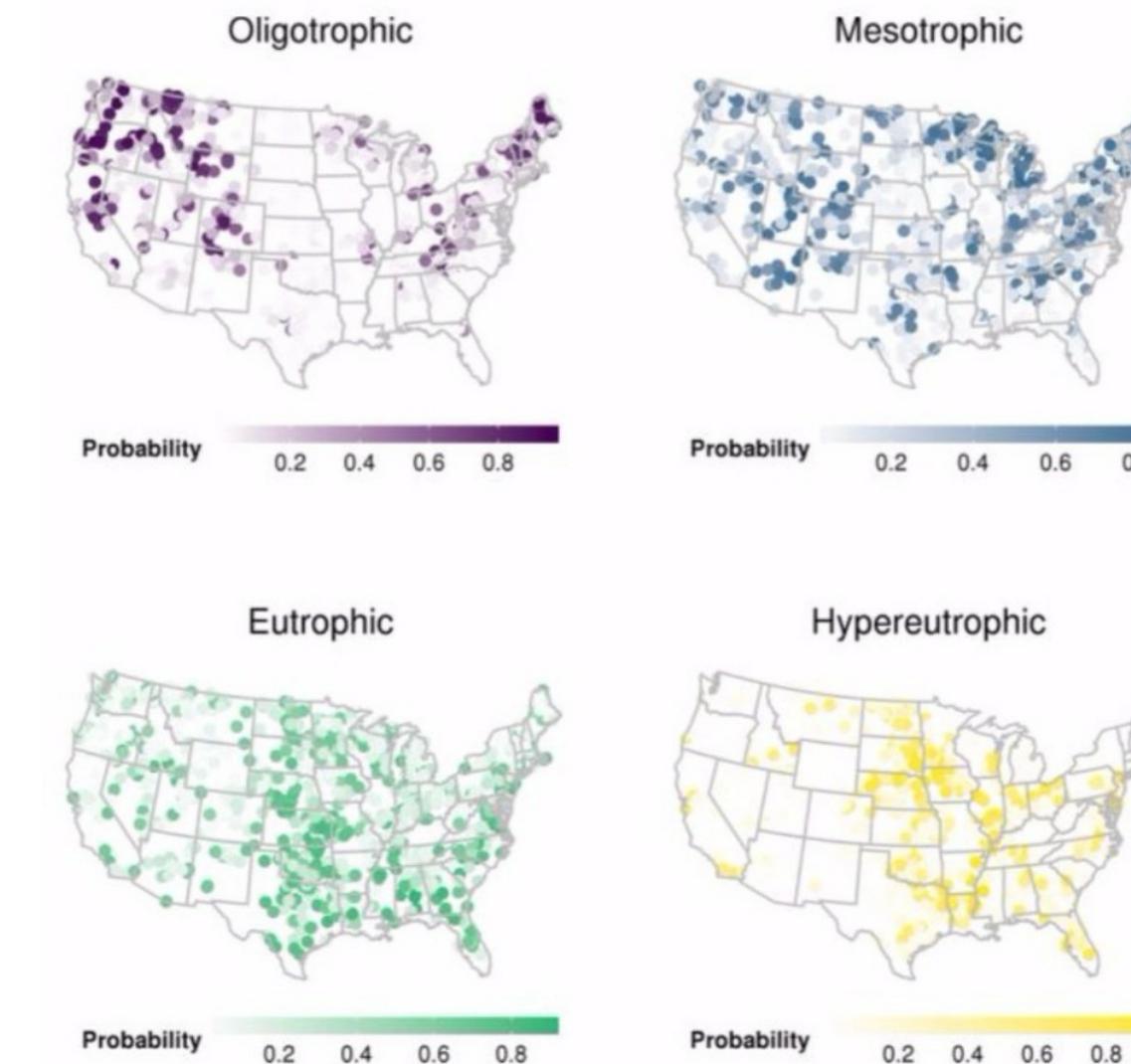
# Models and field research

- Random forest models of trophic state and chlorophyll **a**
- Re-thinking the Lake Trophic State Index
- Chlorophyll *a* and microcystin
- Temporal and spatial dynamics of cyanobacteria blooms
- New work
  - Lake photic zone temperature
  - Phytoplankton community analysis



# Random forest models of trophic State and chlorophyll a

- National
- Data
  - National Lakes Assessment
  - Landcover
- randomForest package
- Variable selection
- All variables (water quality and GIS)
  - 68.7% Total Accuracy
- GIS only variables
  - 49% Total Accuracy
- But ...

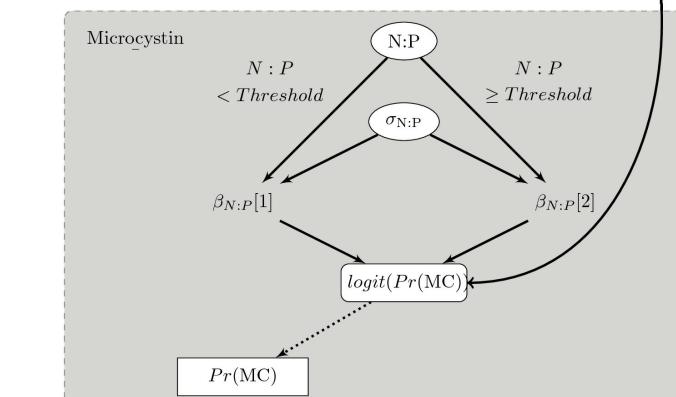
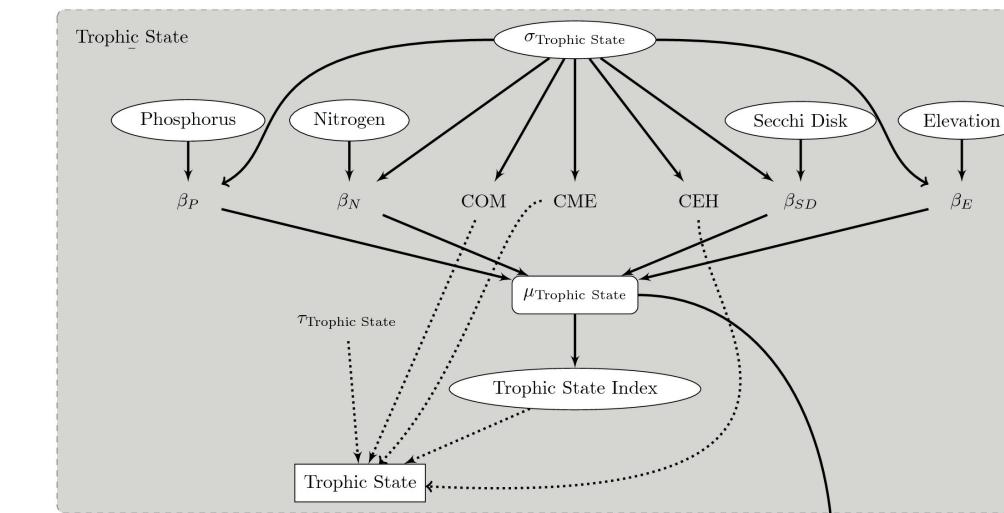


# Random forest models of trophic State and chlorophyll a

- How is it open and reproducible?
  - [GitHub](#)
  - [10.5281/zenodo.40271](#)
  - [PeerJ Pre-print](#)
  - [Ecosphere \(OA\)](#)

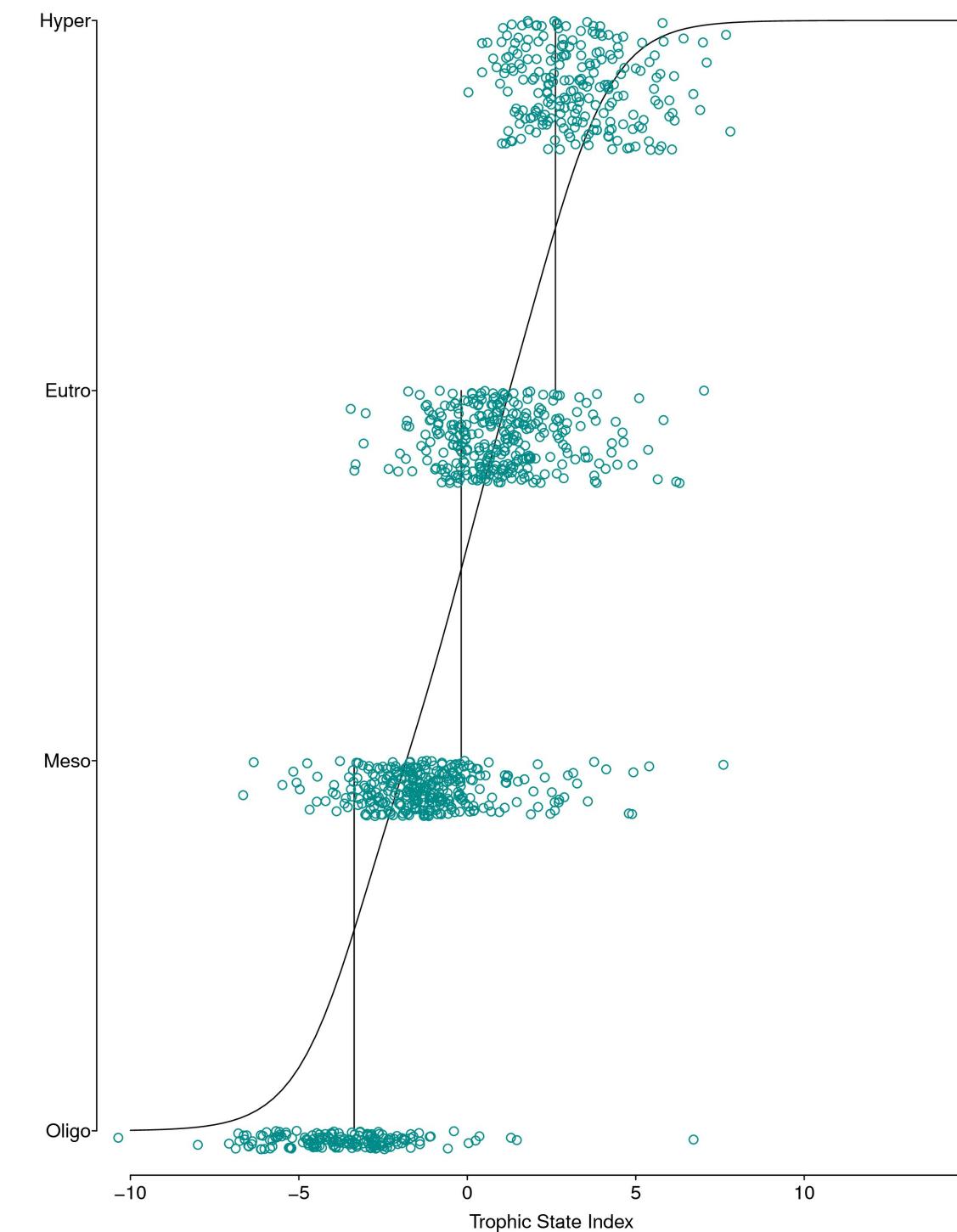
# Re-thinking the Lake Trophic State Index

- Led by Farnaz Nojavan
- Hierarchical model
  - Nitrogen and Phosphorus
  - POLR: Revised Trophic State Index
- Total Accuracy
  - 0.6
- Balanced Accuracy
  - 0.68 to 0.78



# Re-thinking the Lake Trophic State Index

- Hierarchical model
  - Nitrogen and Phosphorus
  - POLR: Revised Trophic State Index
- Total Accuracy
  - 0.6
- Balanced Accuracy
  - 0.68 to 0.78

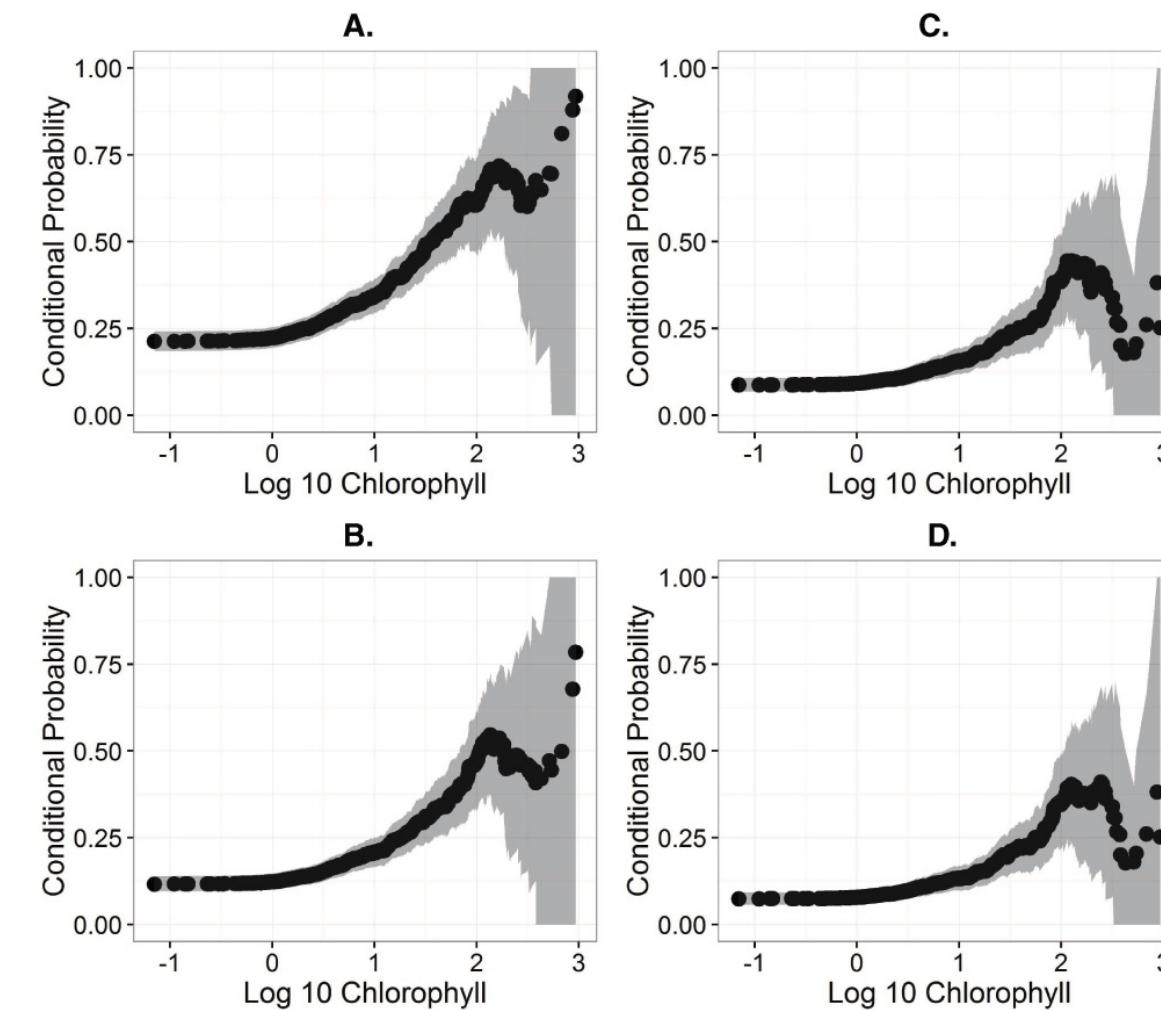


# Re-thinking the Lake Trophic State Index

- How is it open and reproducible?
  - [GitHub](#)
  - [10.5281/zenodo.556175](#)
  - OA (when published)

# Chlorophyll a and microcystin

- National
- Diagnostic tool
- Probability
  - Exceeding microcystin advisory
  - Given chlorophyll *a* concentration



# Chlorophyll a and microcystin

- The numbers!

Cond. Probability	USEPA Child (0.3 µg/L)	WHO Drink (1 µg/L)	USEPA Adult (1.6 µg/L)	WHO Recreational (2 µg/L)
0.1	0.07	0.07	0.07	1
0.2	0.07	4	12	17
0.3	3	17	32	45
0.4	11	37	68	77
0.5	23	68	84	104
0.6	39	97	115	185
0.7	66	126	871	871
0.8	116	271	871	871
0.9	170	516	871	871

# Chlorophyll a and microcystin

- How is it open?
  - [GitHub](#)
  - [Zenodo](#)
  - [F1000Research](#)
    - Pre-print and peer-reviewed in one!

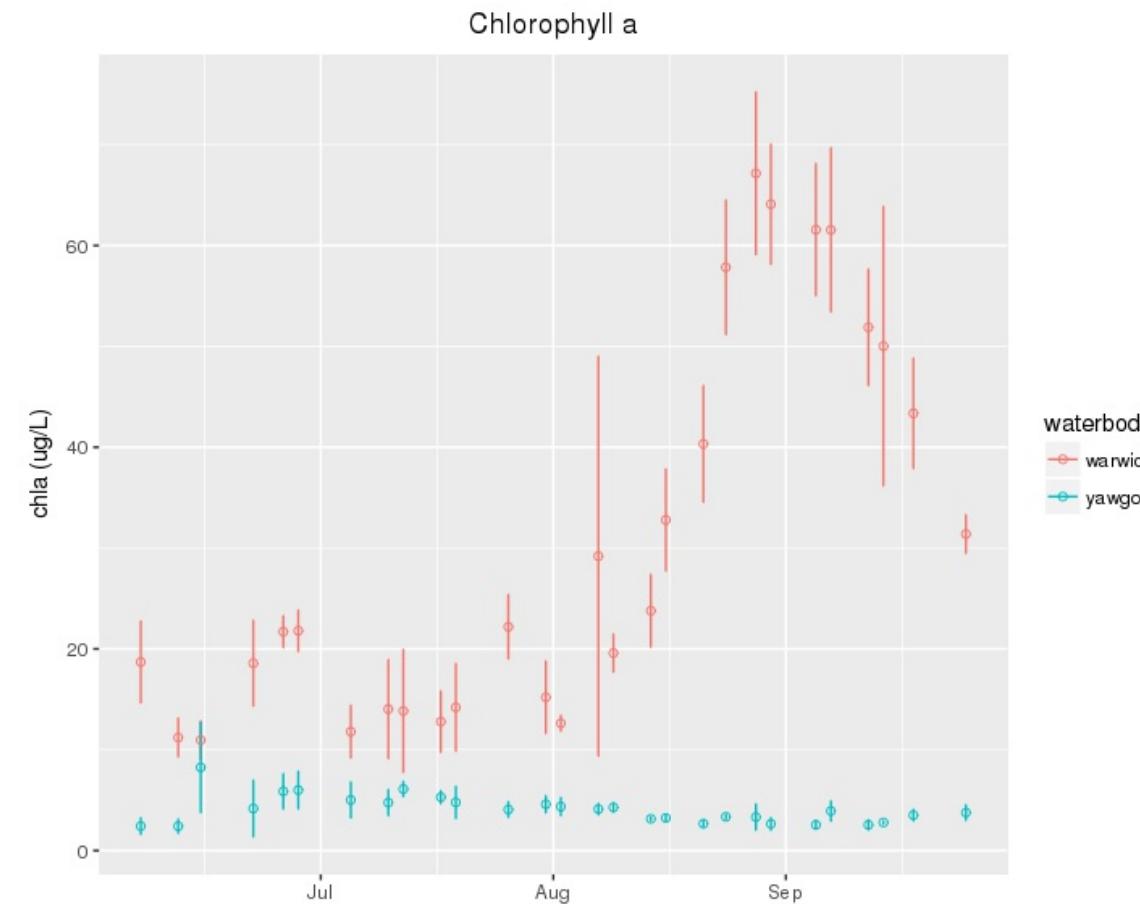
# Temporal and spatial dynamics of cyanobacteria blooms

- Led by Stephen Shivers
- Rhode Island
- Field effort
- 2 ponds
  - Yawgoo Pond (the nice wooded site)
  - Warwick Pond (gritty and (somewhat) urban site)
- Twice weekly
- Seven sampling locations in each

need maps

# Temporal and spatial dynamics of cyanobacteria blooms

- Measurements
  - Chlorophyll *a*
  - Phycocyanin
  - Microcystin
  - Turbidity
  - Physical profiles
  - Secchi
  - Plankton
  - Nutrients



# Temporal and spatial dynamics of cyanobacteria blooms

- How will it be open?
  - [Private \(for now\) GitHub](#)
  - Zenodo
  - Open Access publications
  - Data publication?

image of repo

# New work

- Lake photic zone temperature
- Phytoplankton community analysis

# Thanks!

## Jeff Hollister

US EPA  
Atlantic Ecology Division  
Narragansett, RI  
email: [hollister.jeff@epa.gov](mailto:hollister.jeff@epa.gov)  
twitter: [@jhollist](https://twitter.com/jhollist)  
github: [jhollist](https://github.com/jhollist)

Slides created via the R package [xaringan](#).