

# 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?

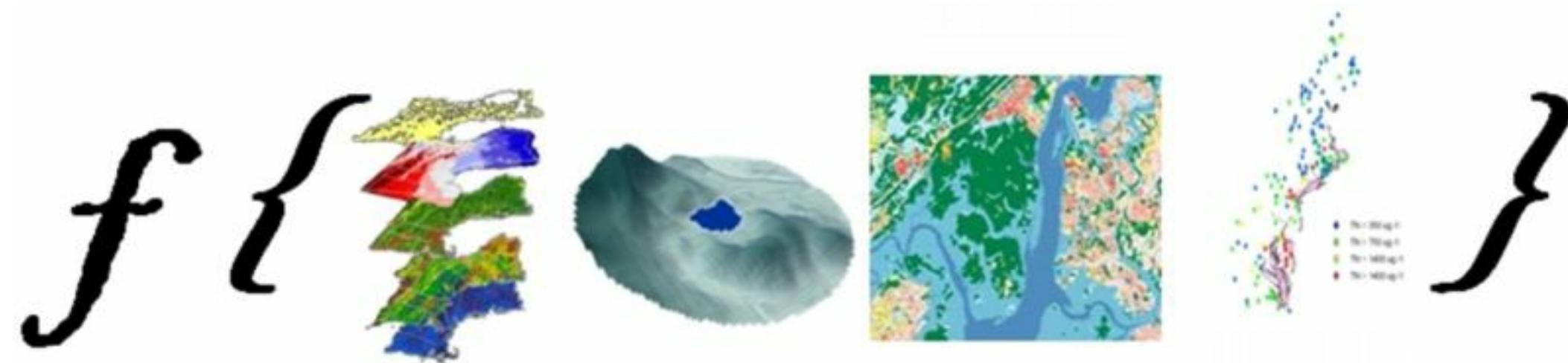
# Who are we?

- 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
- 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
  - Research compendia
  - Tooling for common problems
- Visualization
- Sharing and collaborating
- Publishing
- Apply to our research efforts



# R Packages

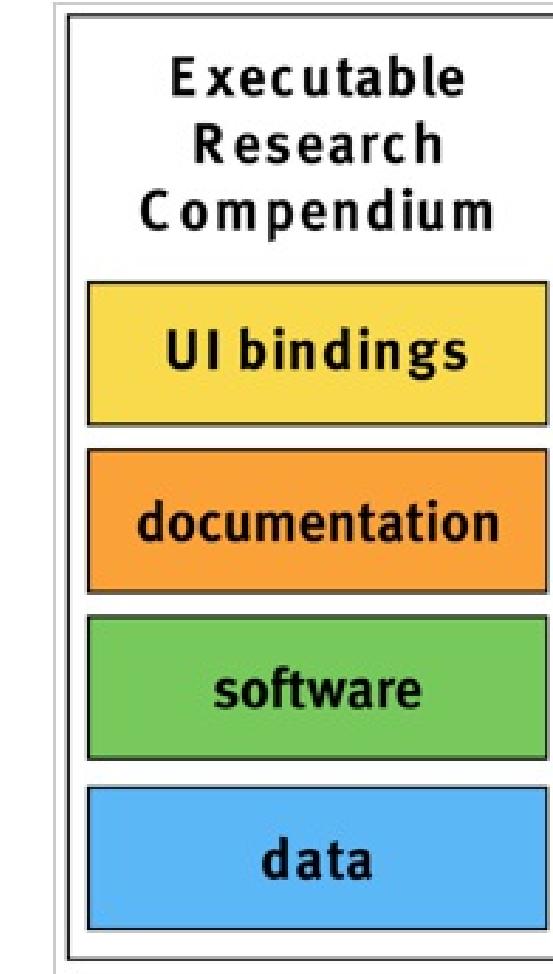
# Why R Packages

- 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'.

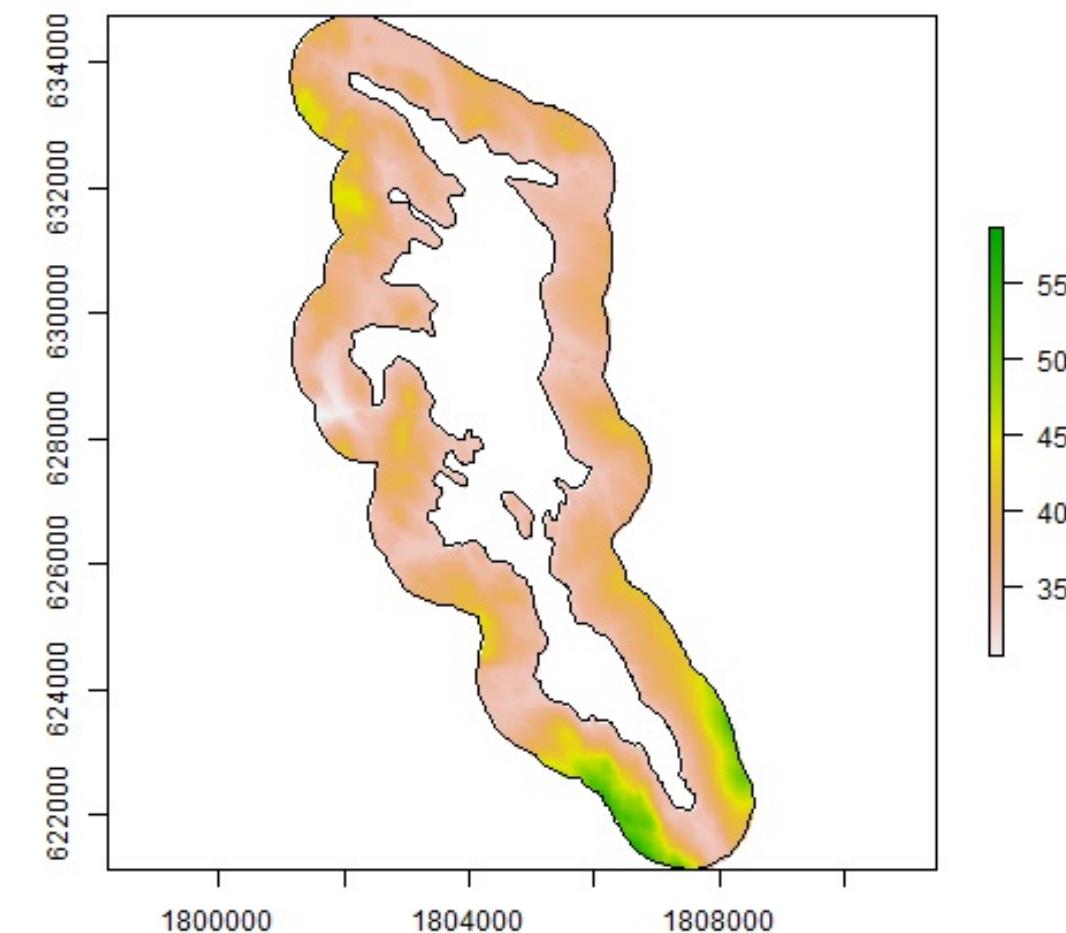
# Packages to solve common problems

- 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)

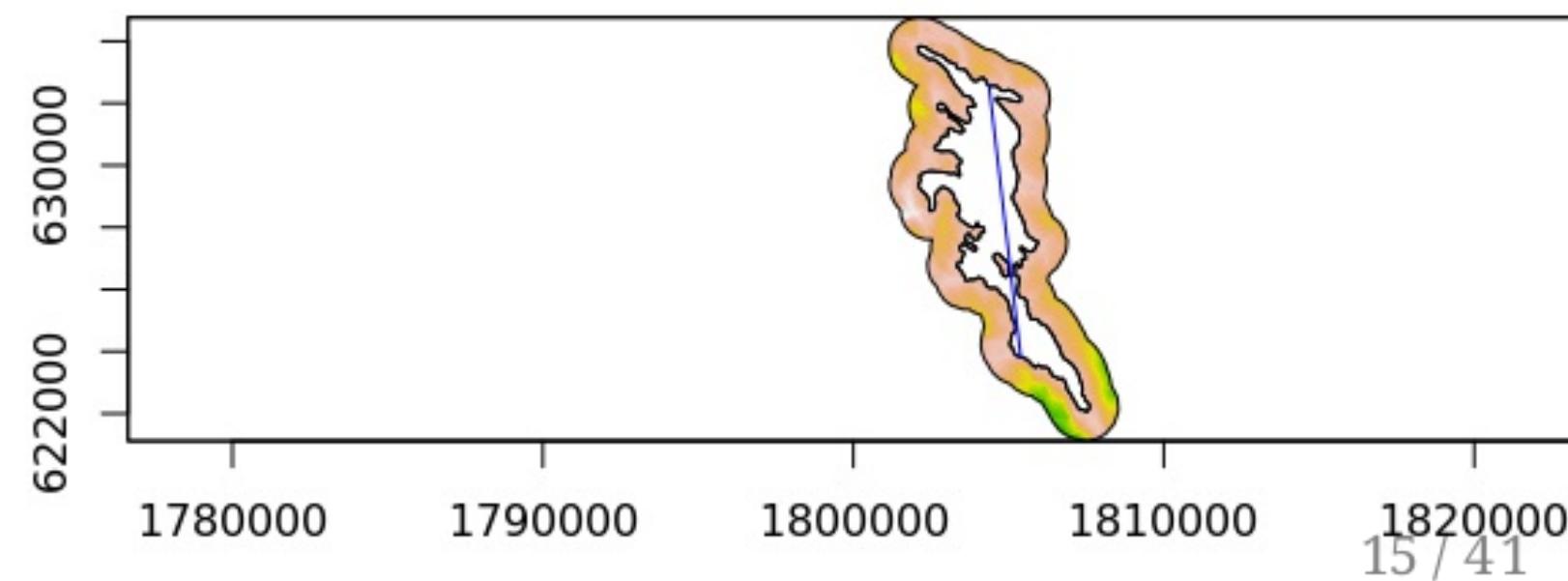
```
rectFactor = 0.553)
rectFactor = 0.553)
ointDens = 100, addLine = TRUE)
```

line

R Script

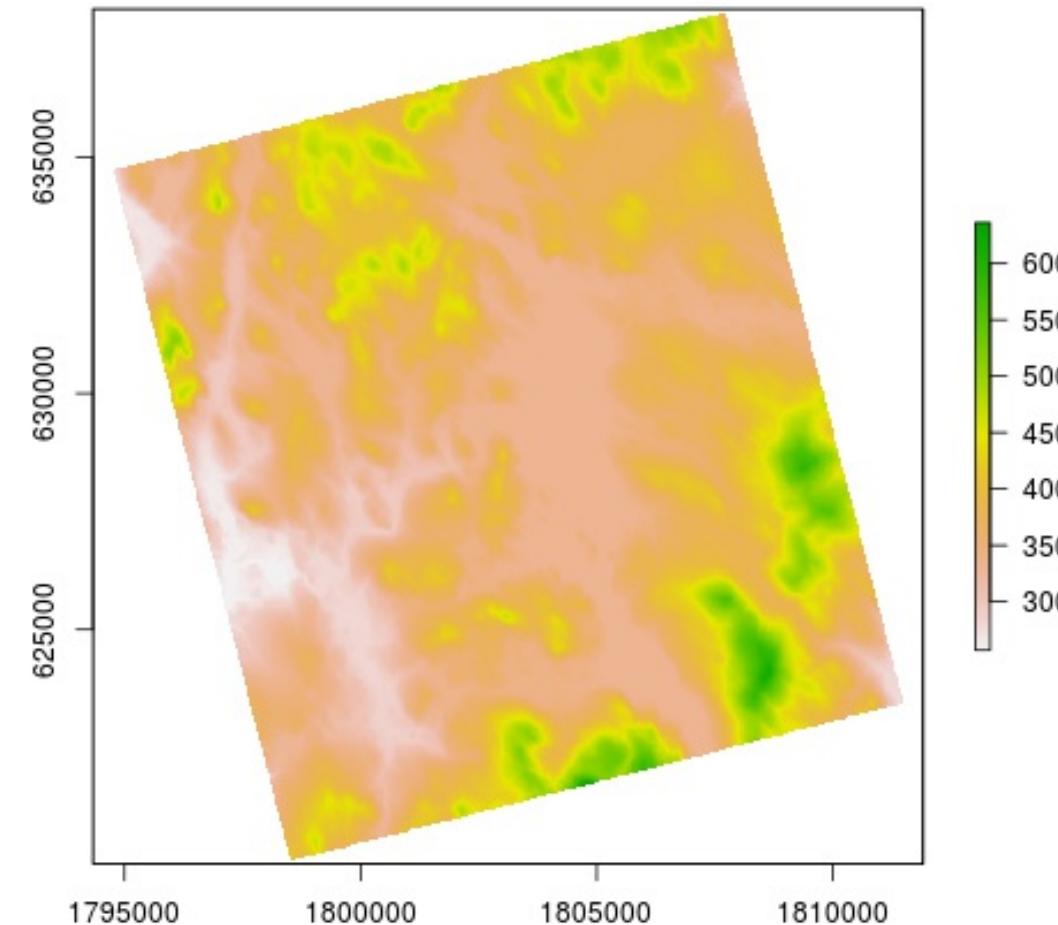
```
"inputLM"
rectFactor = 0.553)
rectFactor = 0.553)
ointDens = 100, addLine = TRUE)
```

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>

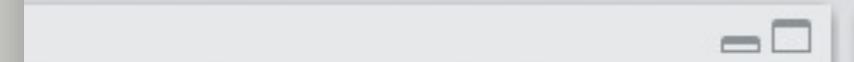
lister.com:8787



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

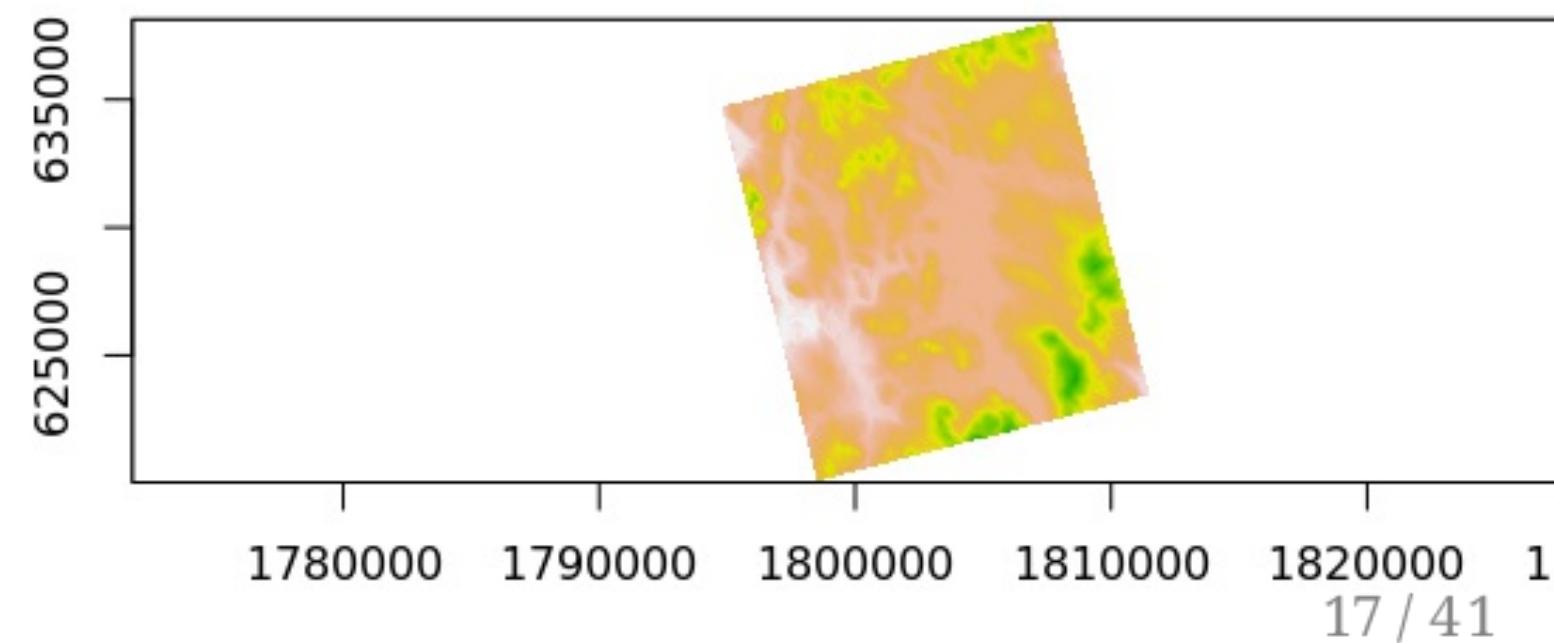
lake Formal class SpatialPolygonsDataFrame

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

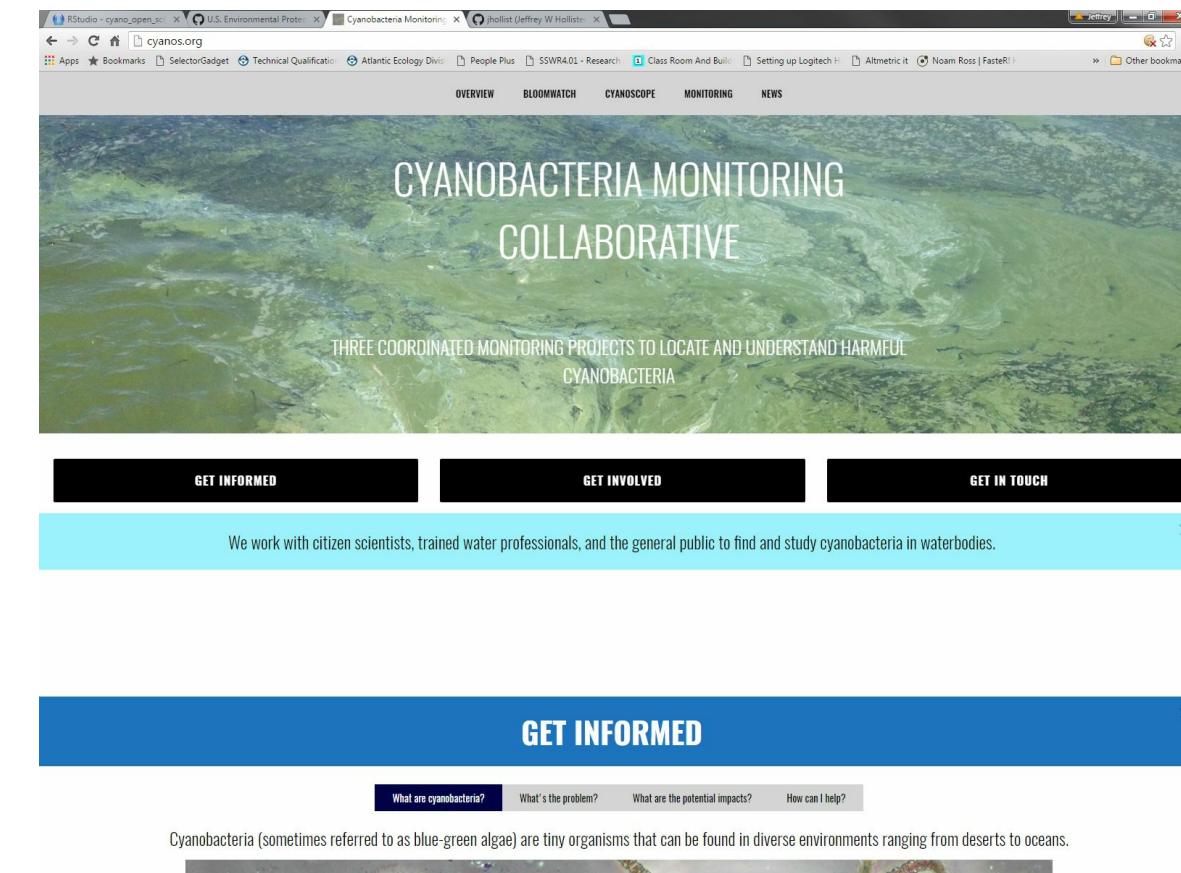
- New effort with Bryan Milstead
- What's in a name?
- Summarizes ancillary data for a user-define landscape polygon
  - Census (via `censusapi`)
  - Landcover
  - Impervious
- Accepts arbitrary spatial data for the landscape
- Based on `sf` and `tidy` by design
- <https://github.com/usepa/goatscape>



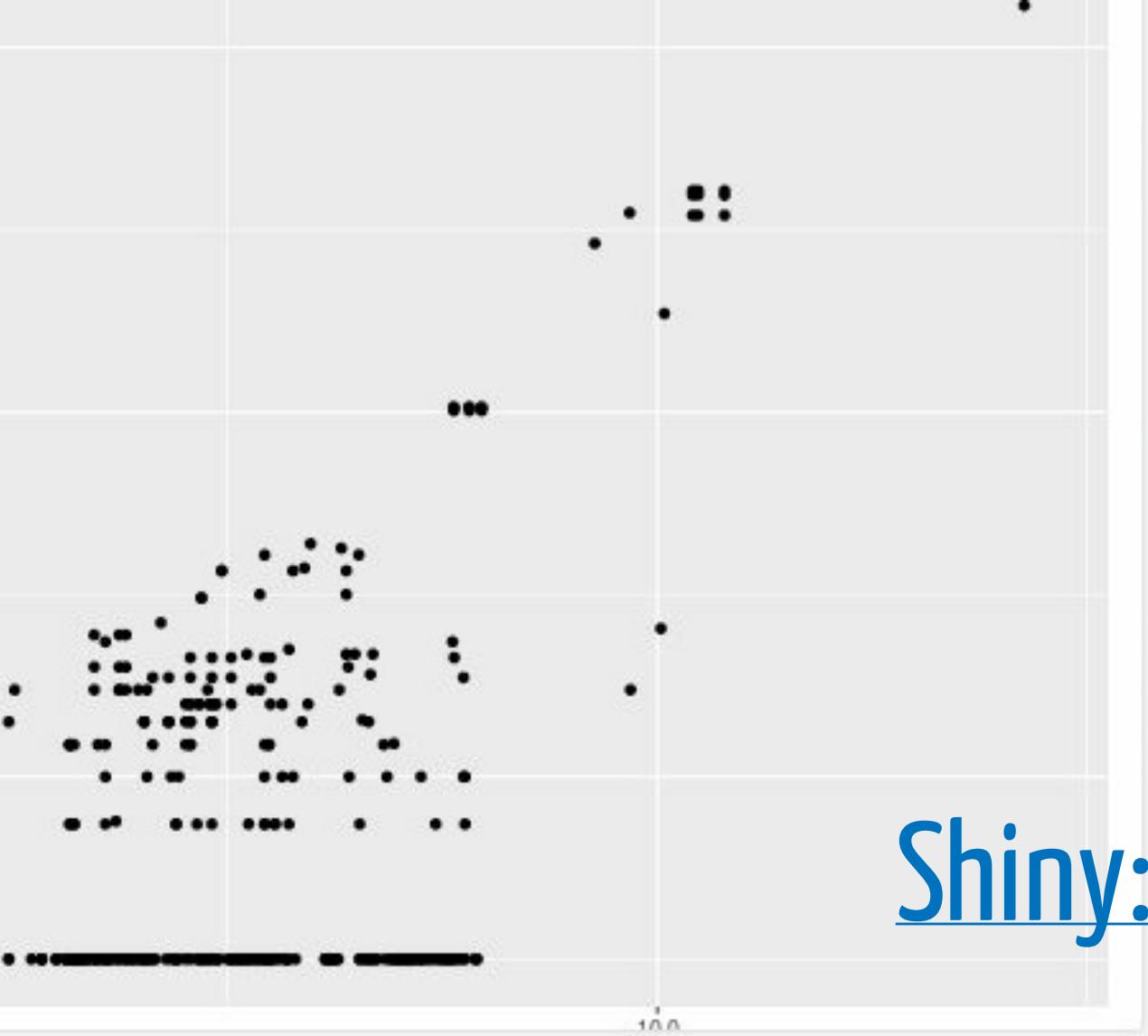
# Data Visualization

# Shiny: Cyanobacteria Monitoring Collaborative

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

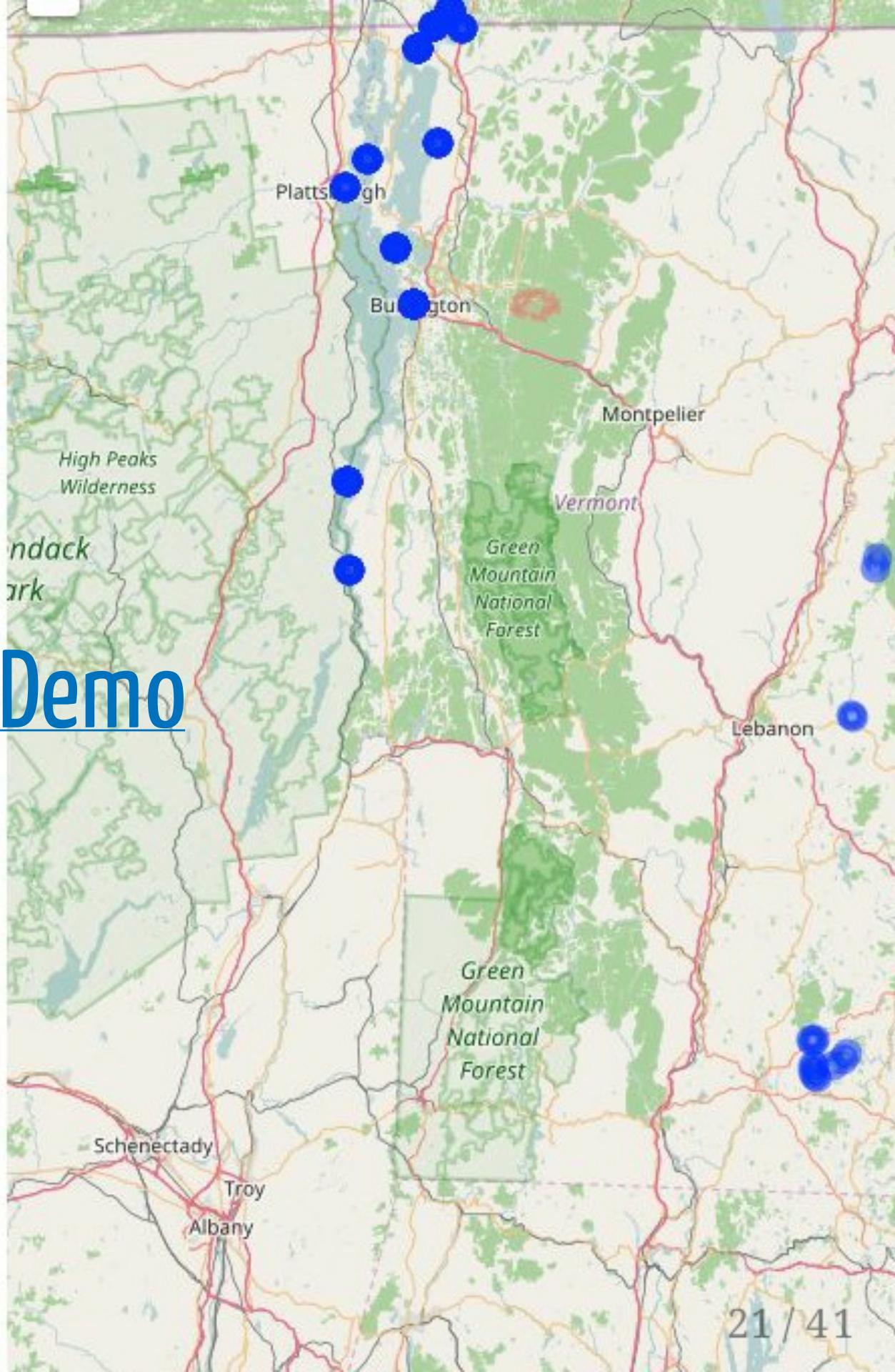


Project URL: <http://cyanos.org>



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?



## Pinned repositories

Customize yo

### 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

## Hollister

Open Science at

al Protection A...

.gov

.com

### rmd\_word\_manuscript

rmd to docx: draft manuscript

● TeX ⭐ 17

## GitHub: Demo

### ropensci/lawn

turf.js R client

● R ⭐ 42 ⚡ 8

### USEPA/lakemorpho

ORD lakemorpho

● R ⭐ 8 ⚡ 7

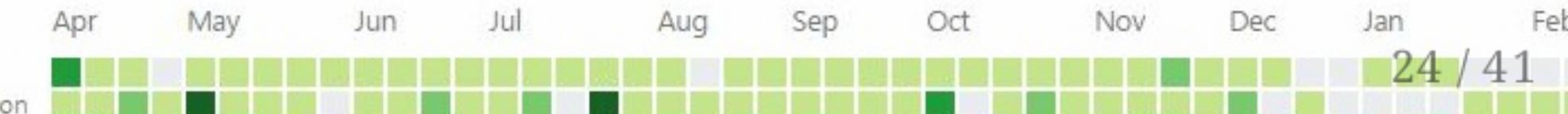
### manuscriptPackage

Template for writing manuscripts as an R

● R ⭐ 30 ⚡ 6



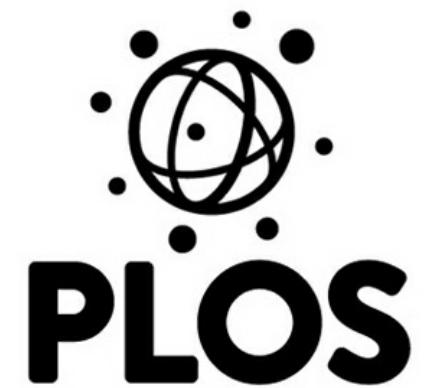
1,876 contributions in the last year



# 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](#)



# Open Science Research

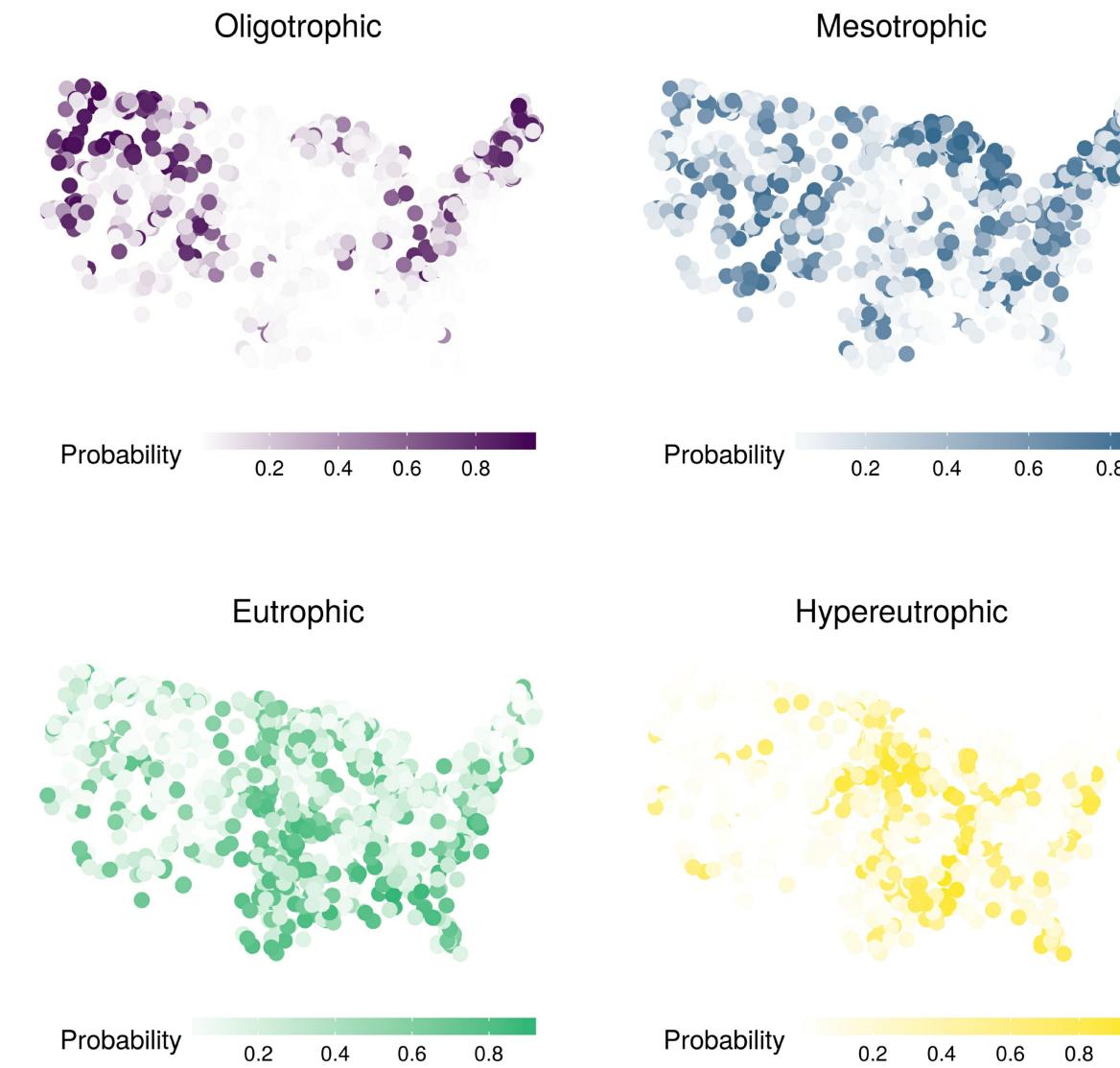
# Models and field research

- 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



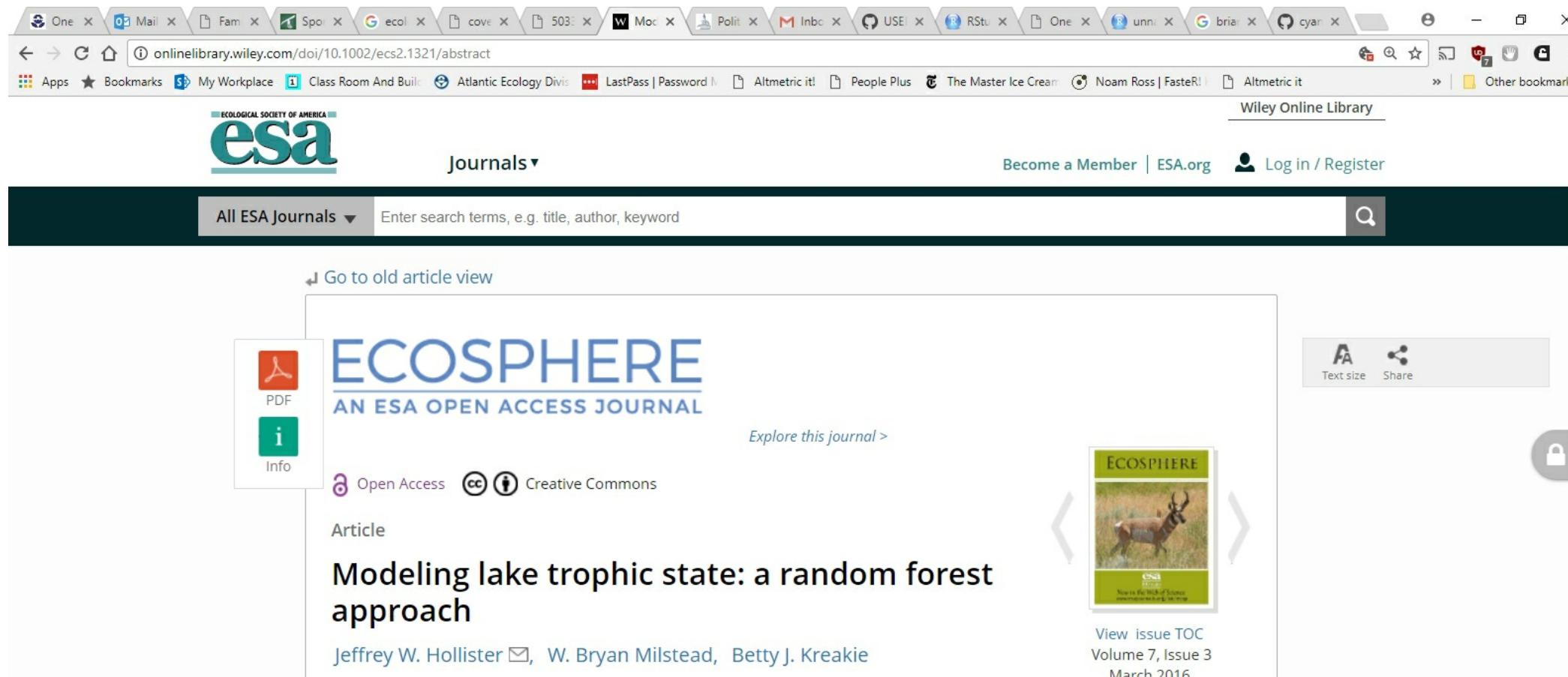
# Random forest models of trophic state and chlorophyll a

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



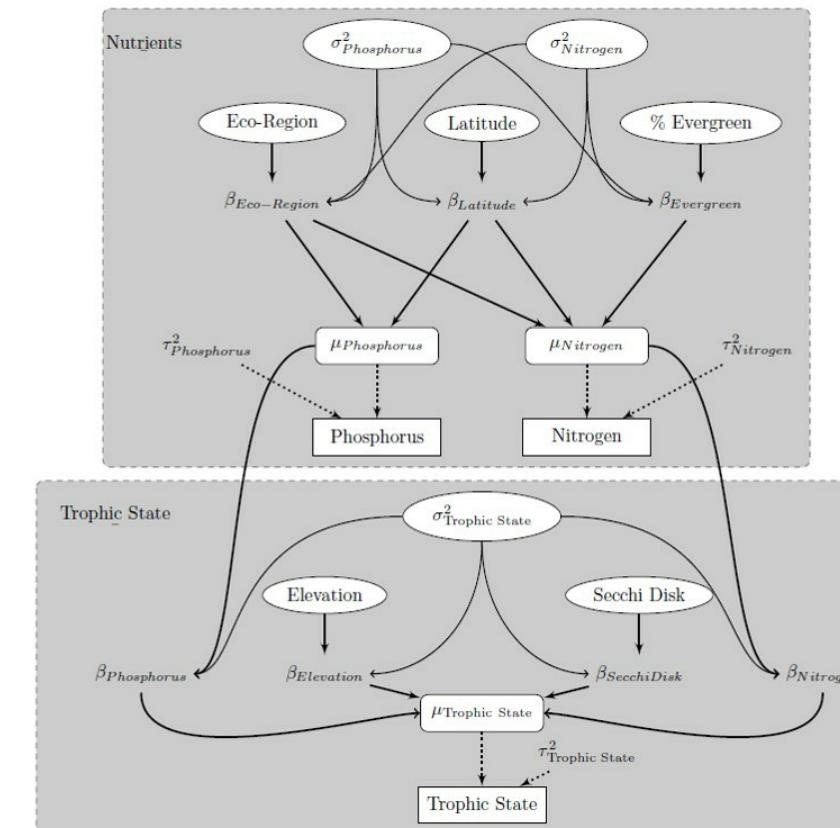
# Random forest models of trophic State and chlorophyll a

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



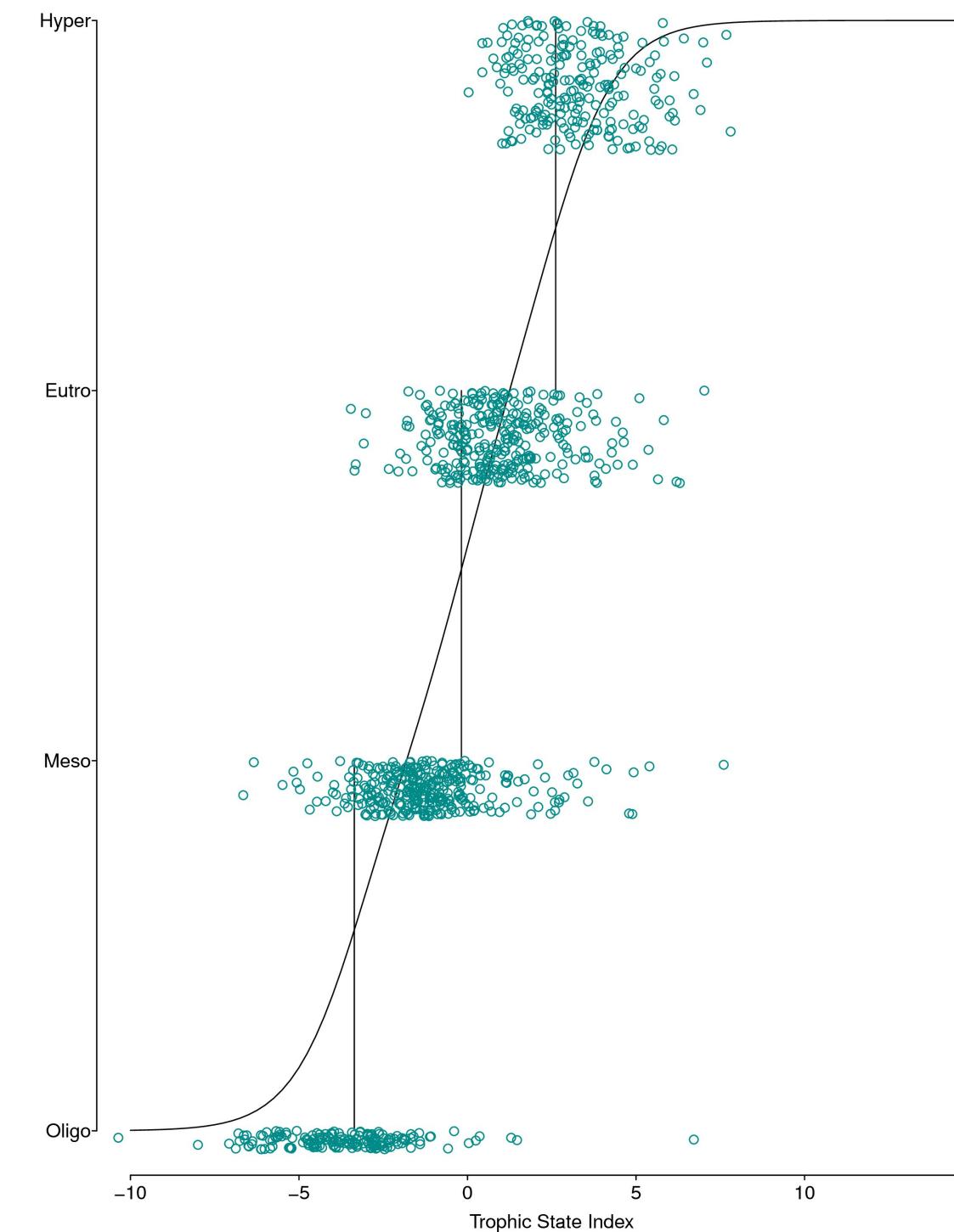
# Re-thinking the Lake Trophic State Index

- 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)

The screenshot shows a web browser window with multiple tabs open at the top. The active tab displays the homepage of the journal "Ecological Modelling". The page features the Elsevier logo, a search bar, and a menu icon. Below the header, there's a navigation bar with links to "Home", "Journals", and "Ecological Modelling". A thumbnail image of the journal cover is shown, along with the ISSN number "0304-3800". To the right, the journal title "Ecological Modelling" is prominently displayed, followed by its description as an "International Journal on Ecological Modelling and Systems Ecology". Below this, there are links for "Supports Open Access", "Editor-in-Chief: Brian D. Fath", and "View Editorial Board". At the bottom left, there are buttons for "Submit Your Paper" and "View Articles". A large block of text describes the journal's focus on mathematical models, systems analysis, ecological processes, and sustainable resource management. The bottom right corner of the slide indicates "33 / 41".

One EPA Workplace × Mail - HollisterJeff@ × E Ecological Modelling × Associations between × USEPA/region7\_r × RStudio - cyano\_ope × An Open Science Fr × LakeTrophicModellir ×

Secure | <https://www.journals.elsevier.com/ecological-modelling>

Apps Bookmarks My Workplace Class Room And Build Atlantic Ecology Divis LastPass | Password M Altmetric it! People Plus The Master Ice Cream Noam Ross | FasteR! Altmetric it! Other bookmarks

ELSEVIER

SEARCH MENU

Home > Journals > Ecological Modelling

ISSN: 0304-3800

ECOLOGICAL MODELLING  
An International Journal on  
ECOLOGICAL MODELLING AND  
SYSTEMS ECOLOGY

Editor-in-Chief: Brian D. Fath

Supports Open Access

View Editorial Board

The journal is concerned with the use of **mathematical models** and **systems analysis** for the description of **ecological processes** and for the **sustainable management of resources**. Human activity and well-being are dependent on and integrated with the functioning of ecosystems and the services they provide...

Submit Your Paper

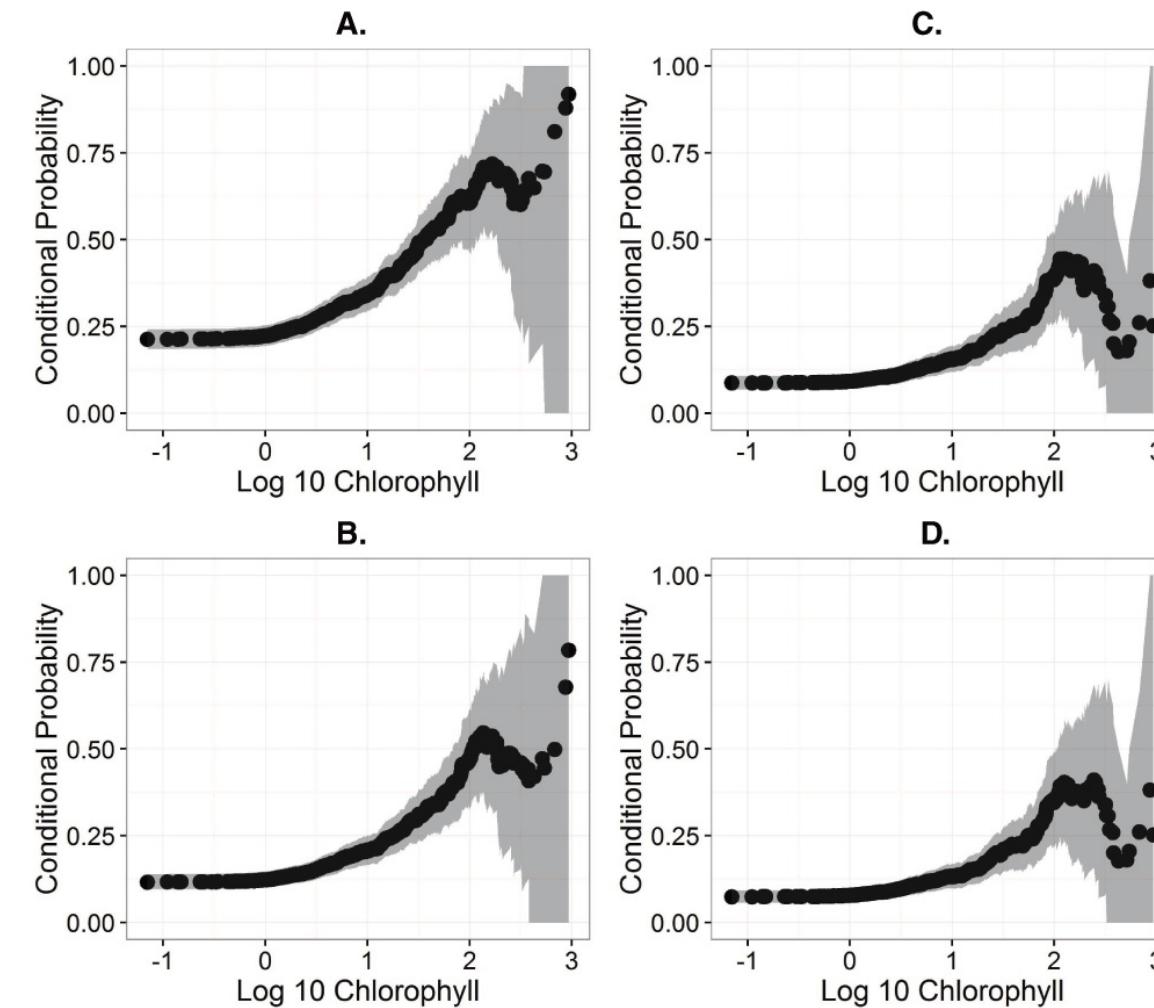
View Articles

Read more

33 / 41

# 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!

The screenshot shows a web browser with multiple tabs open. The active tab is for the F1000Research website, which has an orange header with the text 'F1000Research Open for Science'. Below the header are navigation links for 'BROWSE', 'SUBJECTS' (which is highlighted in red), 'GATEWAYS', 'HOW TO PUBLISH', 'ABOUT', and 'BLOG'. On the right side of the header are 'MY RESEARCH' and 'SIGN IN' buttons. The main content area displays a research article titled 'REVISED Associations between chlorophyll a and various microcystin health advisory concentrations [version 2; referees: 1 approved, 2 approved with reservations]'. The article is authored by Jeffrey W. Hollister and Betty J. Kreakie. It includes sections for 'Abstract' and 'Cyanobacteria harmful algal blooms (cHABs) are associated with a wide range of adverse health effects that stem mostly from the presence of cyanotoxins. To help protect against these impacts, several health advisory levels have been set for some toxins. In particular, one of the more common toxins, microcystin, has several advisory levels set for drinking water and recreational use. However, compared to other water quality measures, field measurements of microcystin are not commonly available due to cost and advanced understanding required to interpret results. Addressing these issues will take time and resources. Thus, there is utility in finding indicators of microcystin that are already widely available, can be estimated quickly and *in situ*, and used as a first defense against high levels of microcystin. Chlorophyll a is commonly measured, can be estimated *in situ*, and has been shown to be positively associated with microcystin. In this paper, we use this association to...'.

On the right side of the article page, there is a sidebar titled 'Open Peer Review' with a section for 'Invited Referees'. It lists three versions of the article:

Version(s)	1	2	3
REVISED	Version 2 published 13 Jun 2016	✓ read report	
Version 1	?	?	✓ read report

Below the sidebar, there are links for 'All reports (4)', 'Responses and comments (3)', and 'Comments on this article'. The sidebar also includes icons for 'Check for updates', 'METRICS' (587 views, 285 downloads), and various sharing options like 'Get PDF', 'Get XML', 'Cite', 'Export', 'Track', 'Email', and 'Share'.

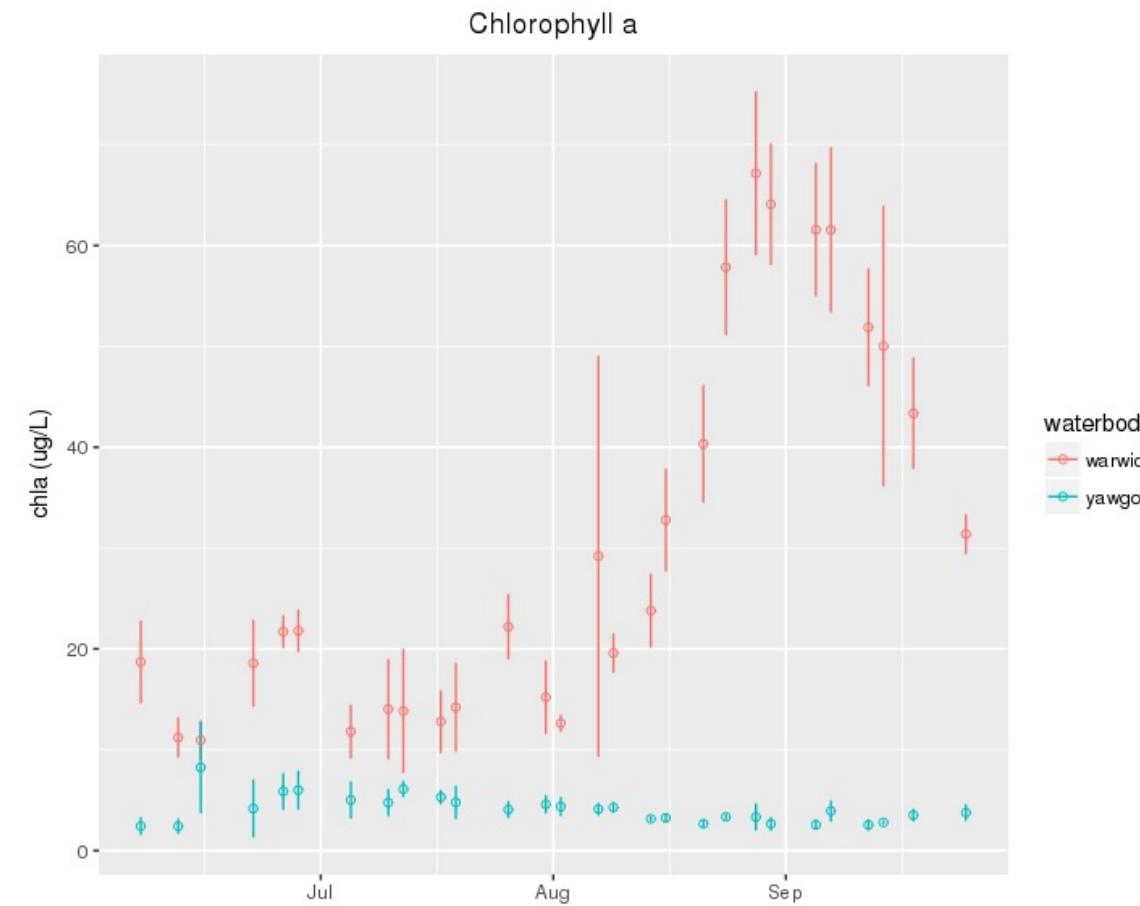
# Temporal and spatial dynamics of cyanobacteria blooms

- 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



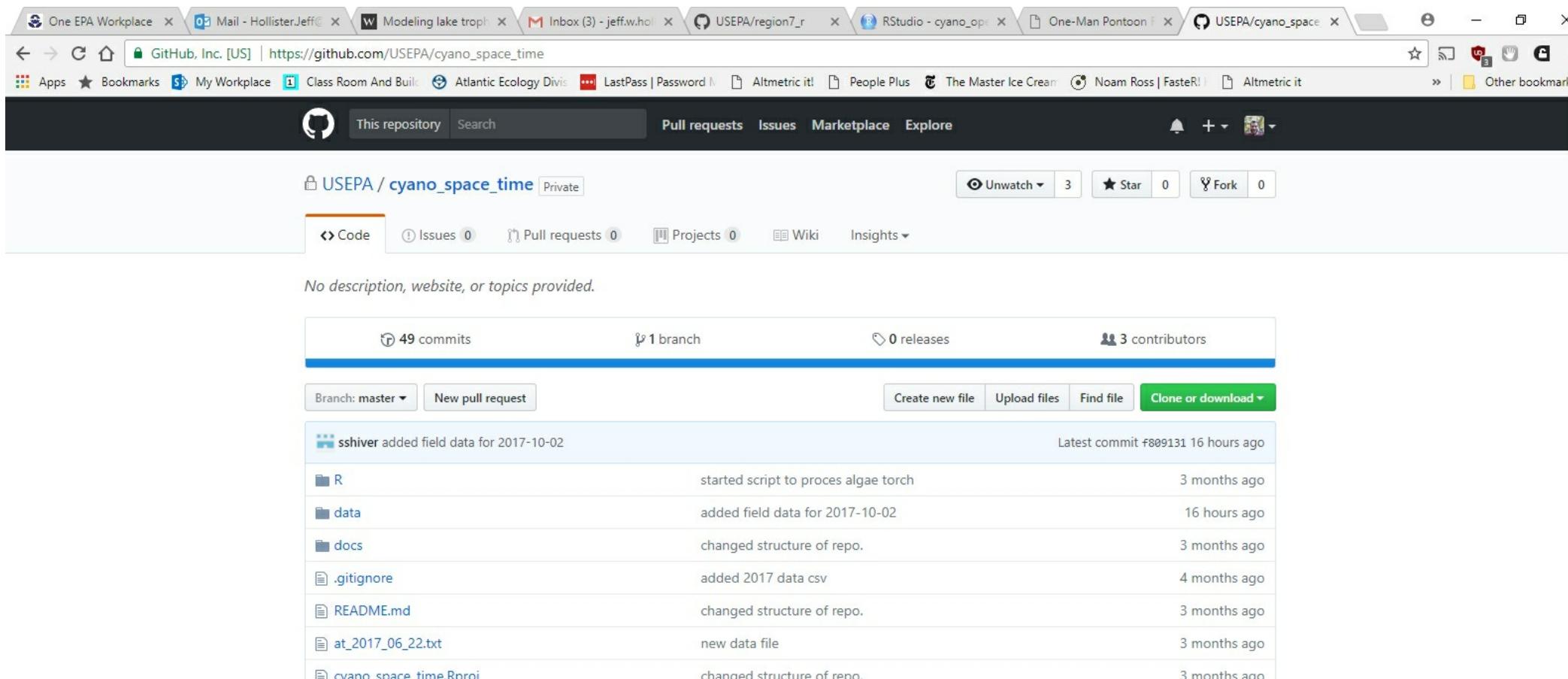
# Temporal and spatial dynamics of cyanobacteria blooms

- 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?



# New work

- Hierarchical Bayes models of microcystin
- 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](#).