



OCEANS OF DATA & REPRODUCIBLE RESEARCH

bit.ly/2017-01-RepRes

Ben Best, PhD (EcoQuants)
2017-01-25 at UCSB Library's Collaboratory
Reproducible Data Research: It's Complicated

Outline

- Reproducible Research?
 - Then
 - Now
- Case Studies
 - Ocean Health Index
 - Marine Geospatial Ecology Tools
- [Data Science] Resources

Reproducible Research (RR)

But on January 10th the stars appeared in the following position with regard to Jupiter; there were two only, and both on the east side



of Jupiter, the third, as I thought, being hidden by the planet.

- Galileo Galilei (1610) *Sidereus nuncius*



RR with Jupyter Notebooks

Jupyter Notebook interface showing the Lorenz system:

Exploring the Lorenz System

In this Notebook we explore the [Lorenz system](#) of differential equations:

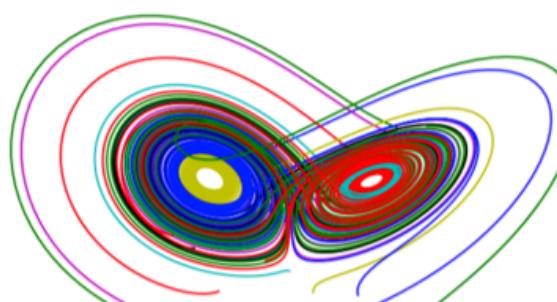
$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= -\beta z + xy\end{aligned}$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ , β , ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

In [7]:

```
interact(Lorenz, N=fixed(10), angle=(0.,360.),
          sigma=(0.0,50.0),beta=(0.,5), rho=(0.0,50.0))
```

angle: 308.2
max_time: 12
 σ : 10
 β : 2.6
 ρ : 28



RR with R and RStudio

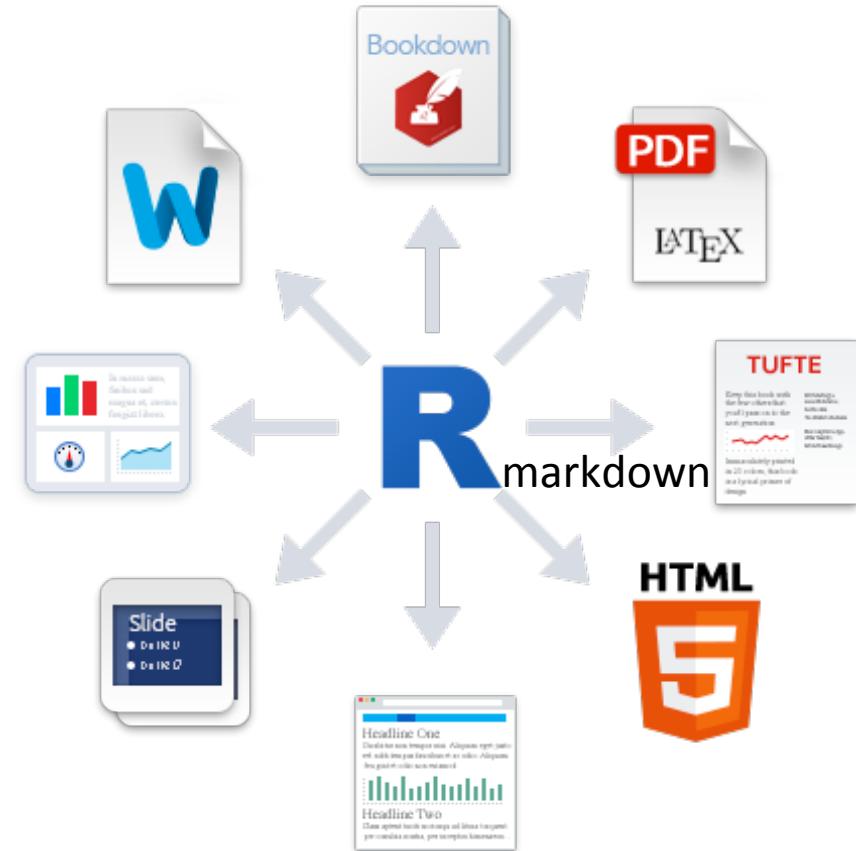
Reproducible Research with R and RStudio

Root

Research Project

Data	Analysis	Presentation
MainData.csv	MainAnalysis.R	Article
GatherSource	ResultsFigures	Paper.Rnw
Makefile	Figure1.R	Main.bib
MergeData.R	Figure2.R	Packages.bib
Gather1.R	Figure3.R	figure
Gather2.R		Figure1.pdf
Gather3.R		Figure2.pdf
		Figure3.pdf

Christopher Gandrud



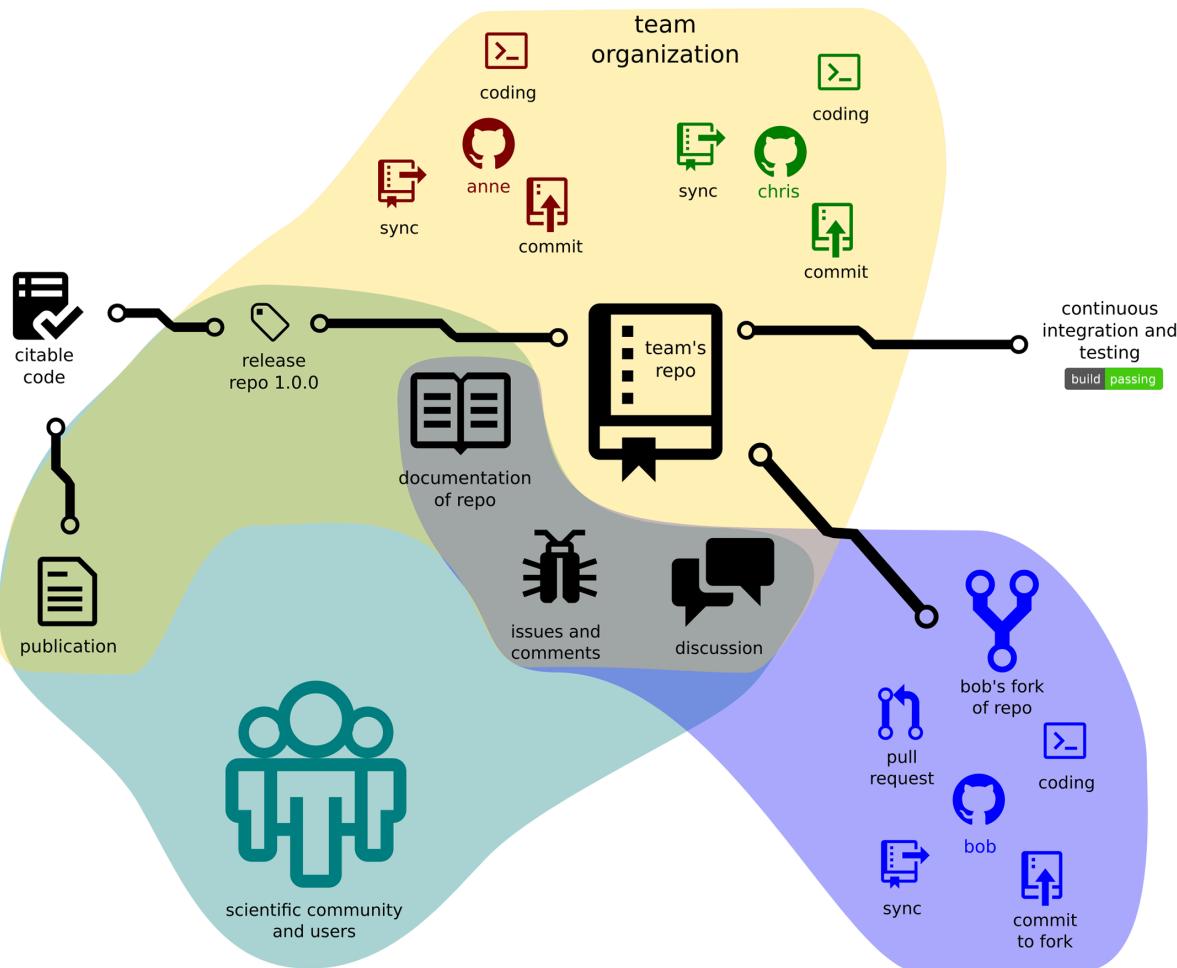
christophergandrud.github.io/RepResR-RStudio

rmarkdown.rstudio.com

What is R Markdown? video(1 min)

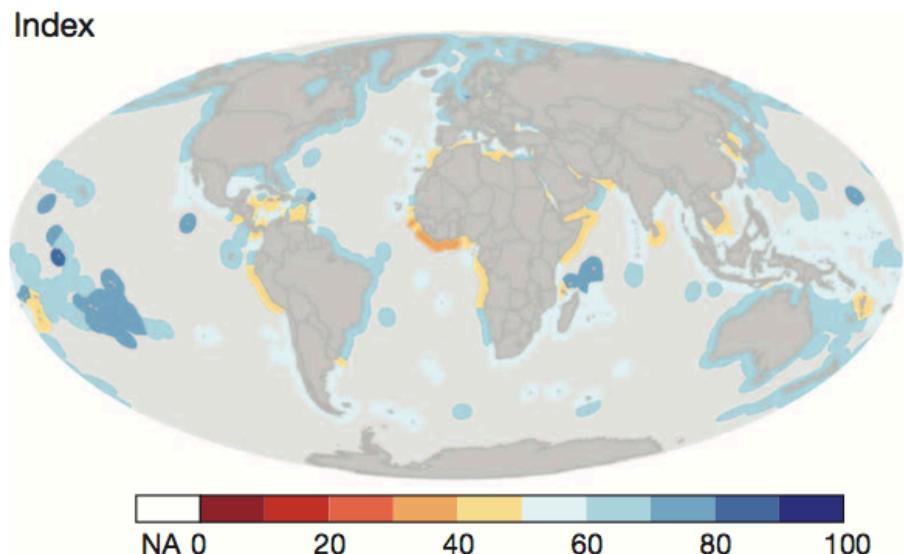
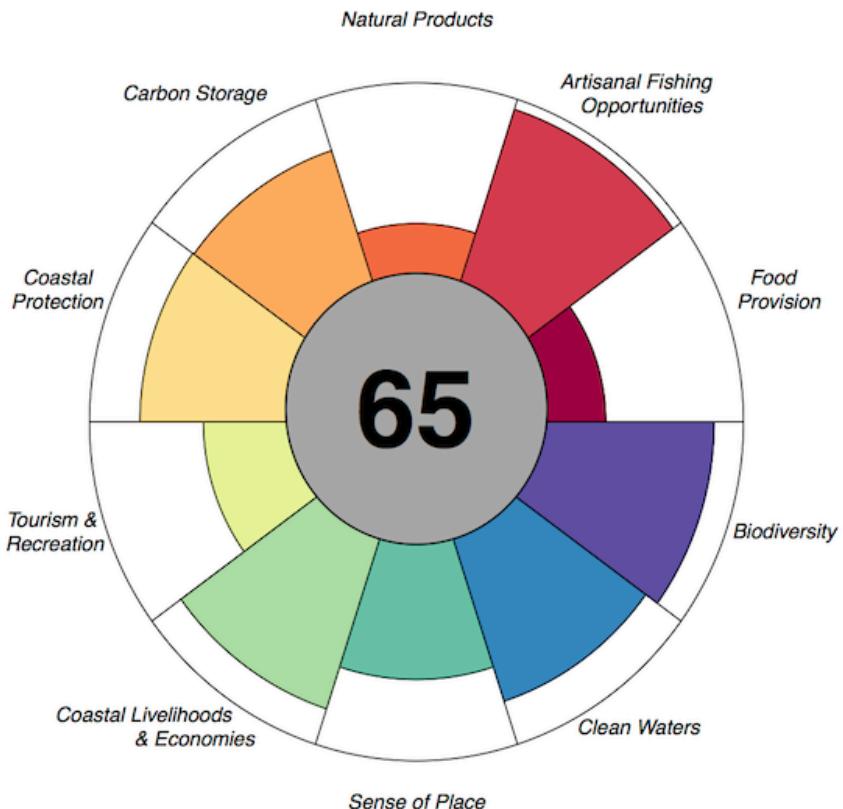
vimeo.com/178485416

Version (Git) & Share (Github)



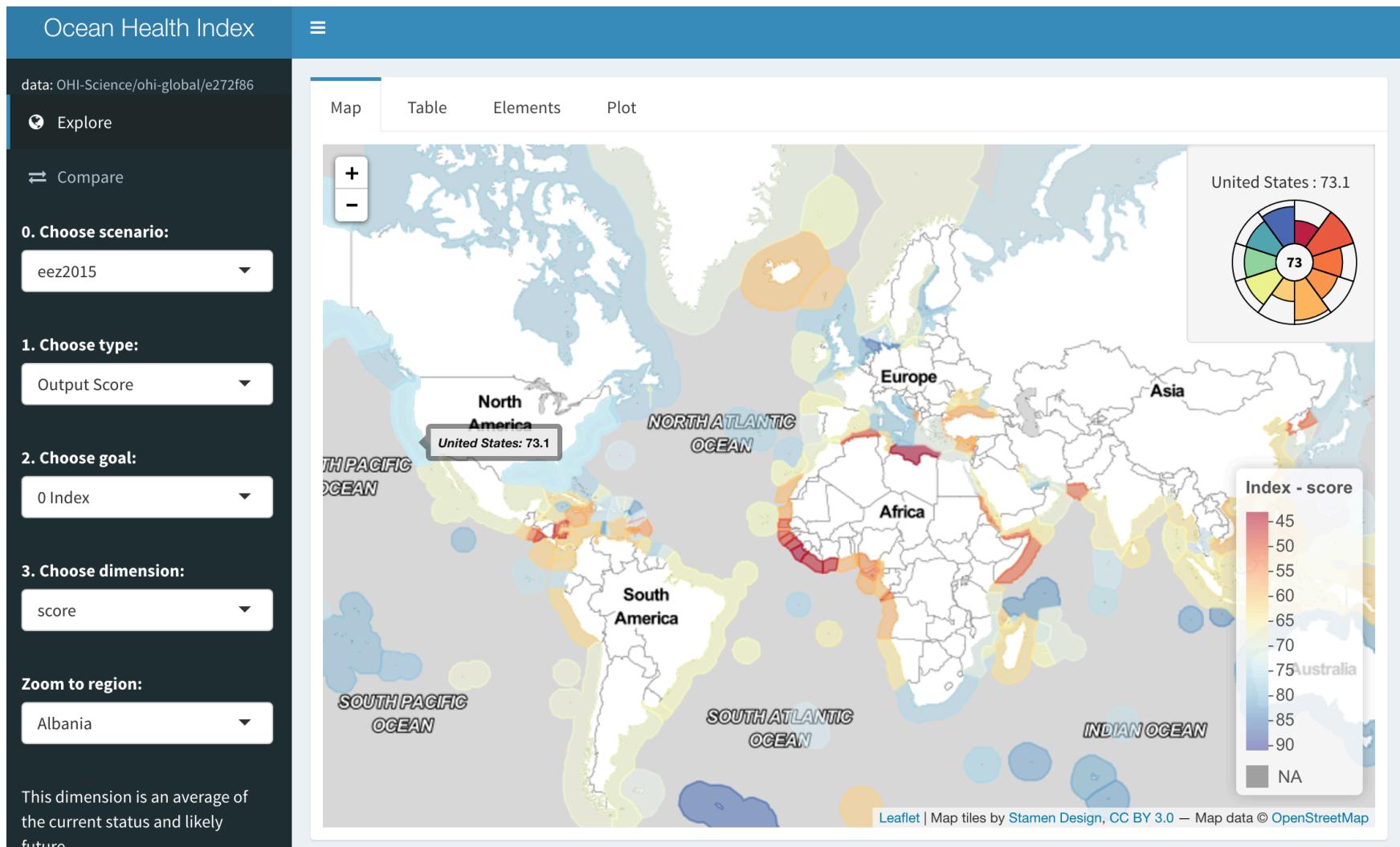
Ocean Health Index (OHI)

"A healthy ocean sustainably delivers a range of benefits to people now and in the future."

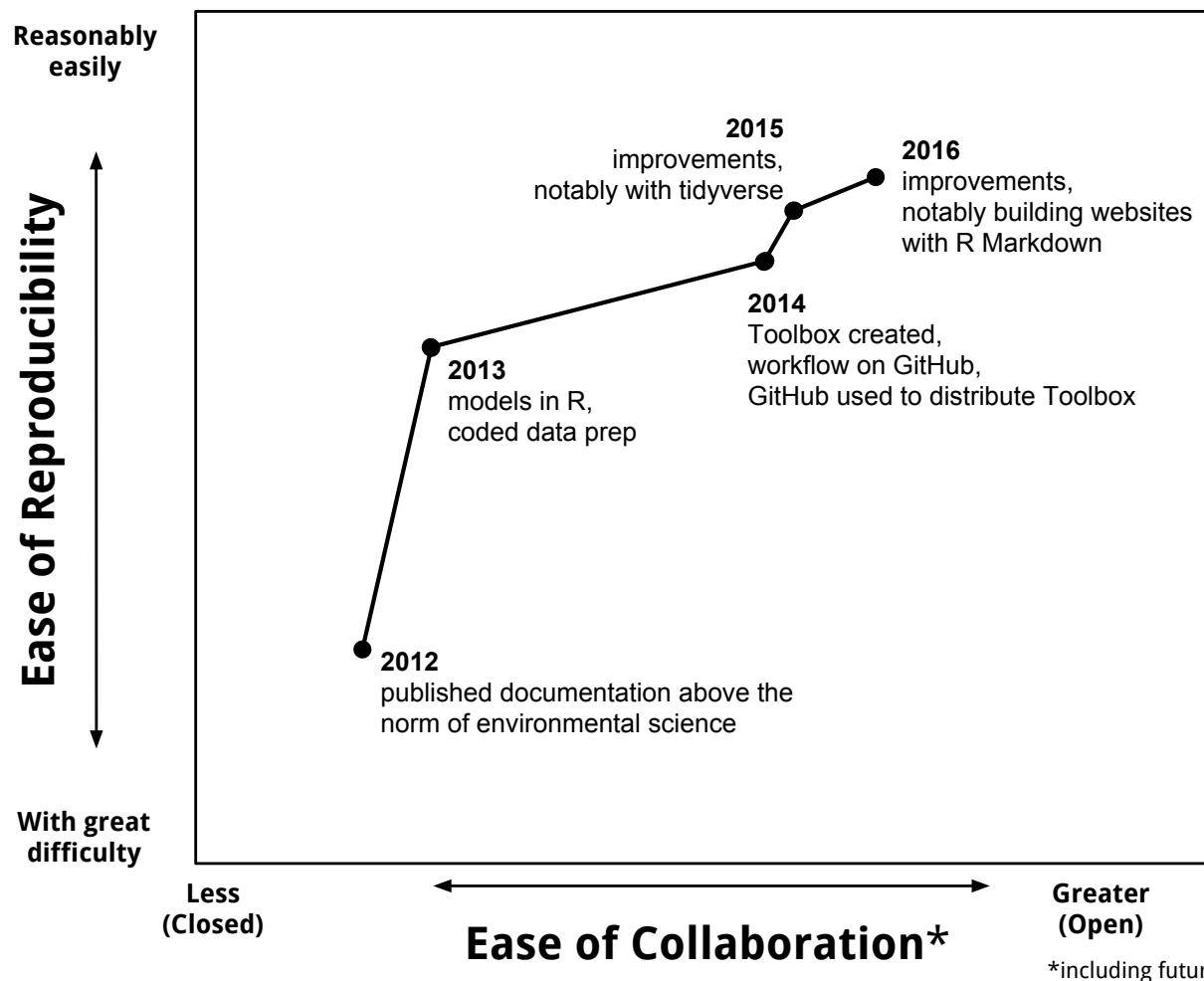


Halpern et al (2012) [An index to assess the health and benefits of the global ocean](#). *Nature*

OHI App (R Shiny)

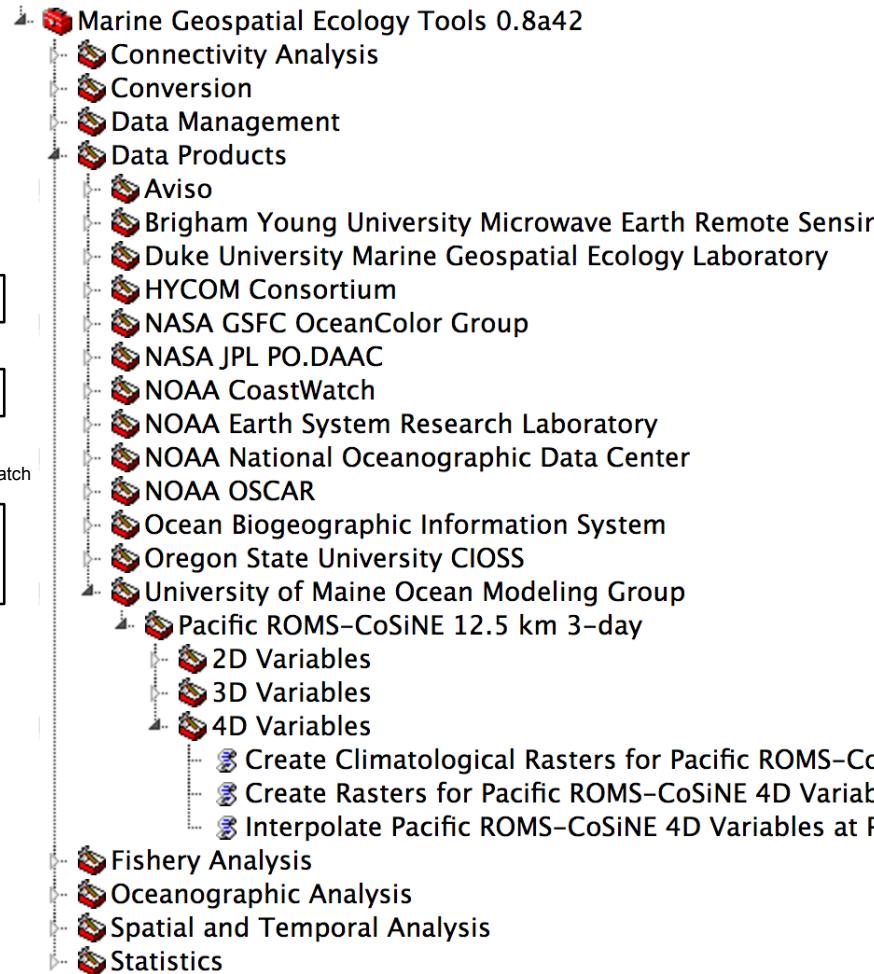
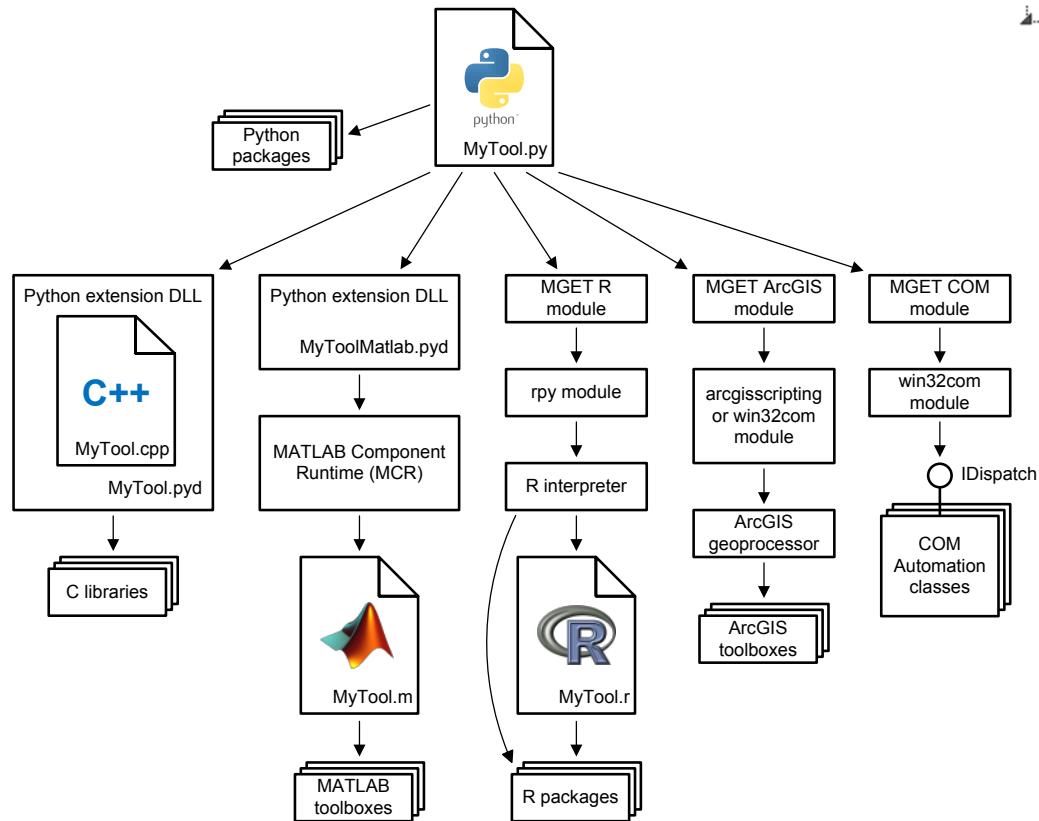


OHI Growth into RR



Lowndes et al (in revision) Better science in less time using open data science tools. *Nature Ecology & Evolution*

Marine Geospatial Ecology Tools



Roberts et al (2010)

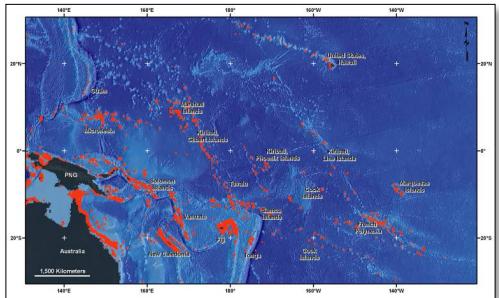
[Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. Env Mdl & Sw](#)

mgel.env.duke.edu/mget

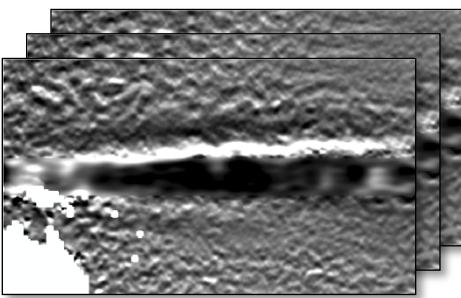
Model larval connectivity



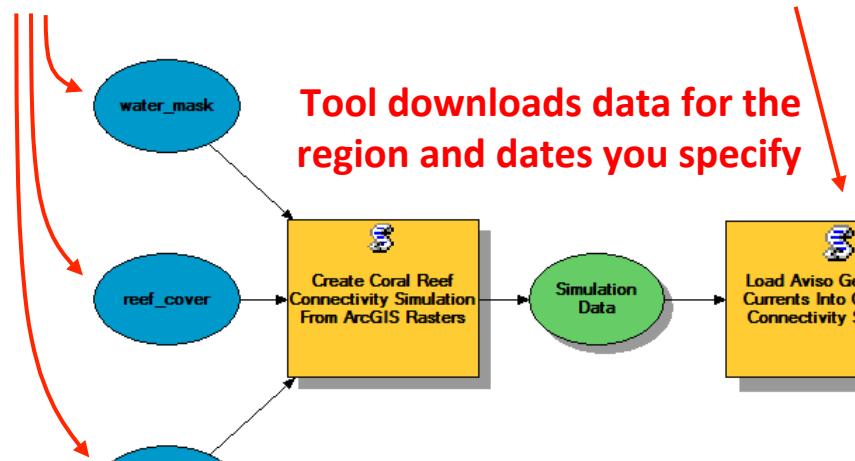
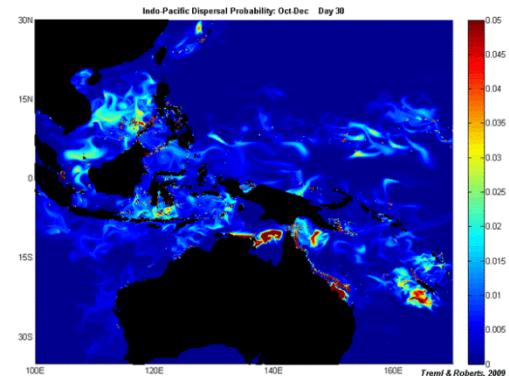
Habitat patches



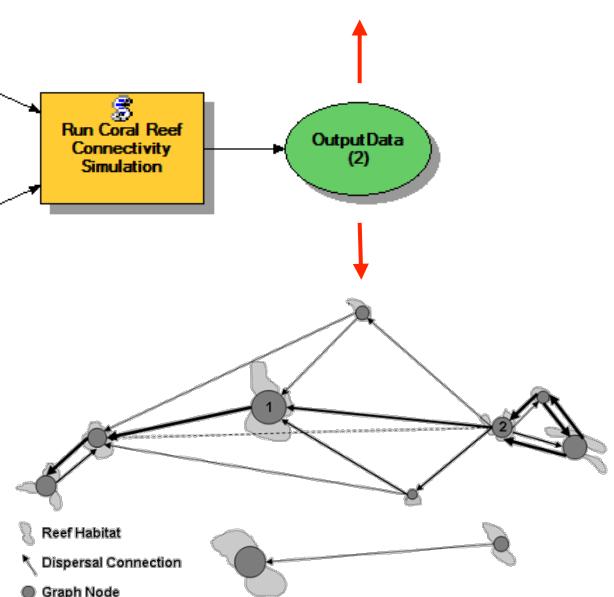
Ocean currents data



Larval density rasters



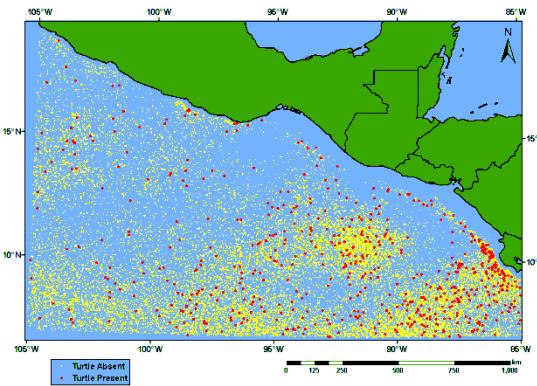
Edge list feature class representing dispersal network



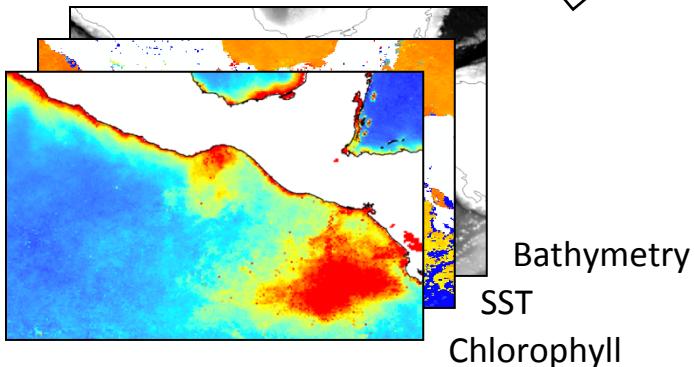
Model species habitat



Point observations of species

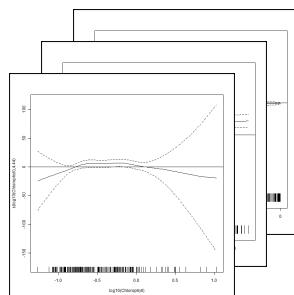


Gridded environmental data

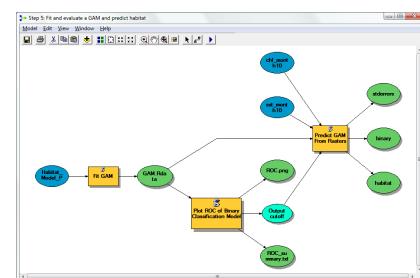
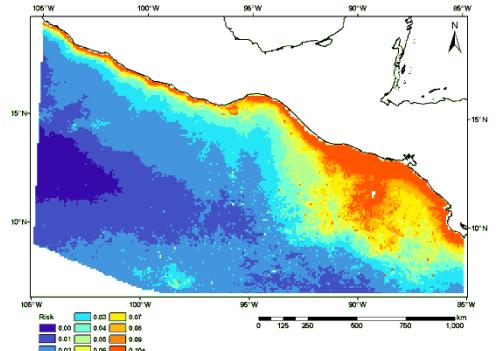


Predictive model

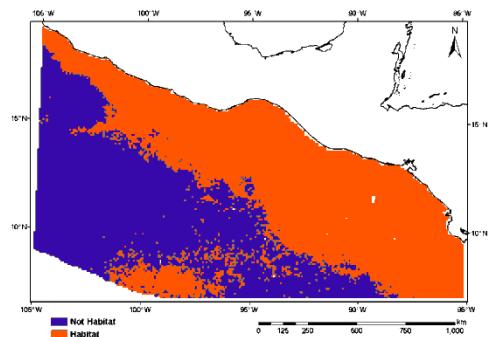
$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$



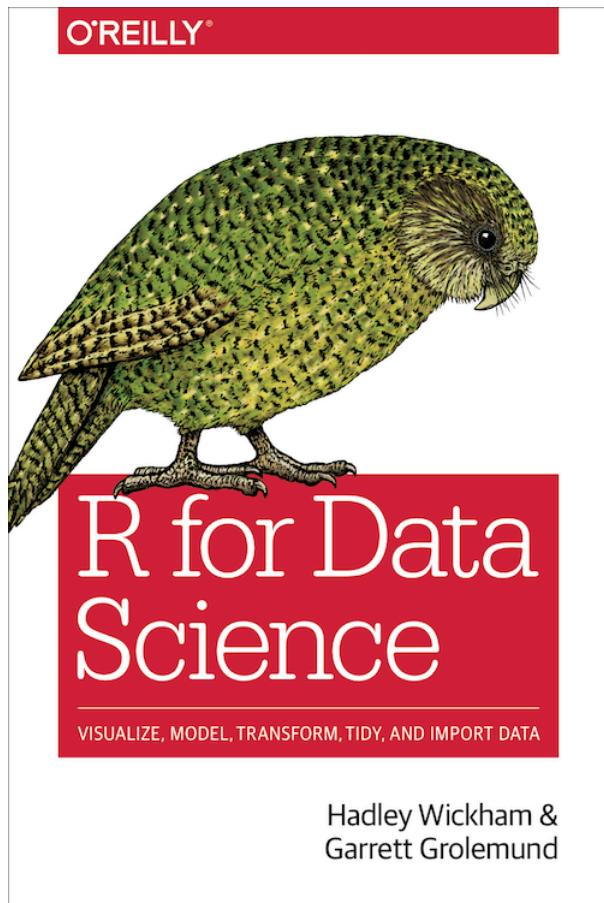
Probability of occurrence predicted from environmental covariates



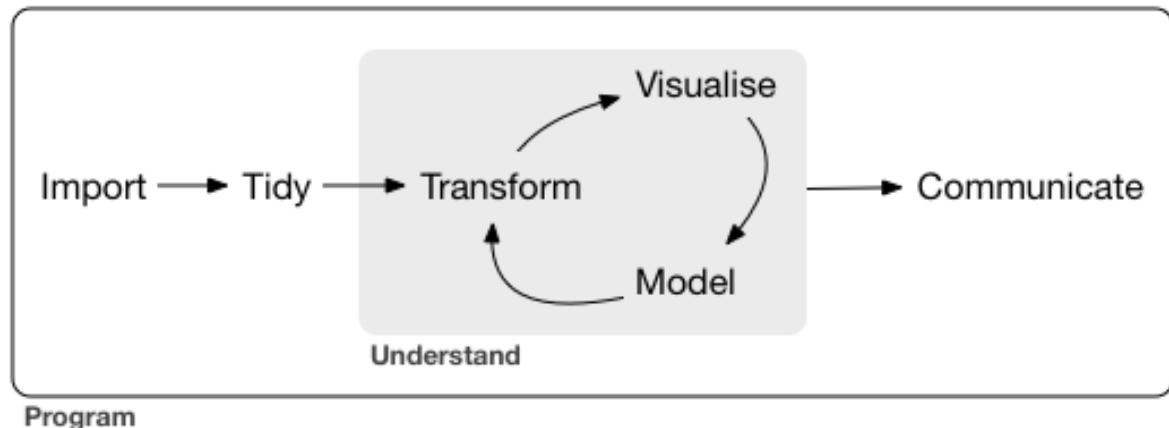
Binary classification



R for Data Science



r4ds.had.co.nz



tidyverse.org

“Data scientists, according to interviews and expert estimates, spend from **50 percent to 80 percent** of their time mired in the mundane labor of collecting and **preparing data**, before it can be explored for useful information.” - [NYTimes \(2014\)](https://www.nytimes.com/2014/02/16/science/the-data-scientist-as-millennial.html)

MAKE DATA GREAT AGAIN

"BIG" DATA IS SO SAD.
WITH ME, YOU WILL GET
"YUGE" DATA, AND WE WILL
MAKE DATA GREAT AGAIN!



KDNUGGETS.COM

User Groups

There are a lot of user groups in SB area:



[Santa Barbara R Users Group](#): community R



[Santa Barbara Data Science](#): community big data



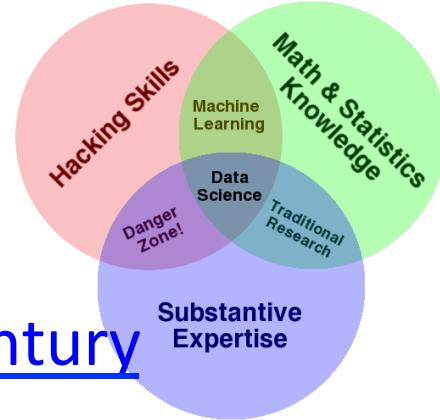
[Data Science at UCSB](#): undergrad seminars



[UCSB Data Science](#): grad learning



[Eco Data Science](#): Bren/NCEAS grad learning

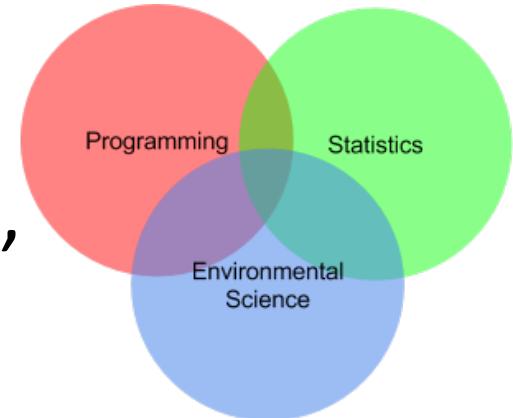


Data Science Testimonials

- [Data Scientist: The Sexiest Job of the 21st Century](#)
(Harvard Business Review, 2012)
- Data scientists are the **magicians** of the Big Data era...
190,000 projected **shortage** in data scientists by 2018.
[Universities Offer Courses in a Hot New Field: Data Science](#) (NY Times, 2013)
- [Data scientists] responsible for a team of 1-3 earn [a median salary of] **\$140,000** and those responsible for a team of 10 or more earn **\$232,500**. By contrast, the mean average annual income for a lawyer in America was \$131,990 in 2013, while doctors earned \$183,940.
[Tell Your Kids to Be Data Scientists, Not Doctors](#)
(Wired, 2014)

Data Science Resources

- [BenBestPhD.com](#) : Teaching :
[Environmental Informatics](#) course,
[Software Carpentry](#) workshop
- [OHI-science.org/resources/links](#)
- [NCEAS.ucsb.edu/ecoinfo](#)



presentation: bit.ly/2017-01-RepRes
email: ben@ecoquants.com

Thank you!