

# Better science in less time using data science tools

Julia Stewart Lowndes, PhD // lowndes @nceas.ucsb.edu  
November 11, 2016 // WSN, Monterey, California

# Trying to do reproducible science

`data_final_v2.xls`

`Re: FWD: data question`

# Actually doing reproducible science

We struggled to reproduce our own work.

Data science principles and tools have changed how we do science:

- reproducibility
- collaboration
- communication

# Data science

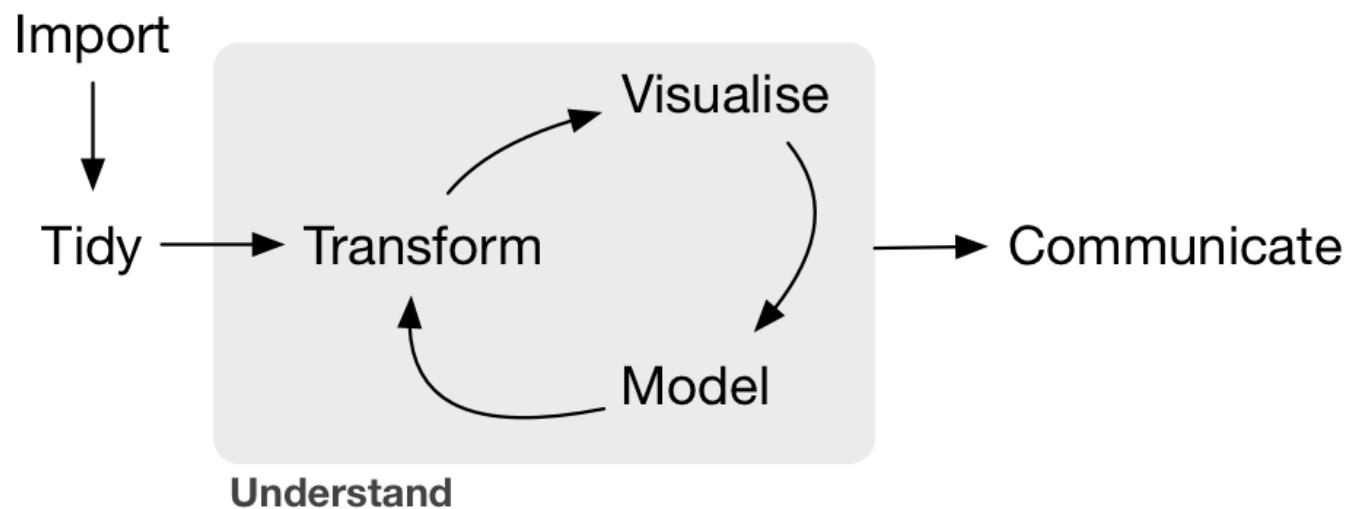
A discipline "to turn raw data into understanding, insight, and knowledge"

---

# Data science

A discipline "to turn raw data into understanding, insight, and knowledge"

---



# Science and data science

Science:

- when is X most abundant?
- what is Y's habitat preferences?



# Science and data science

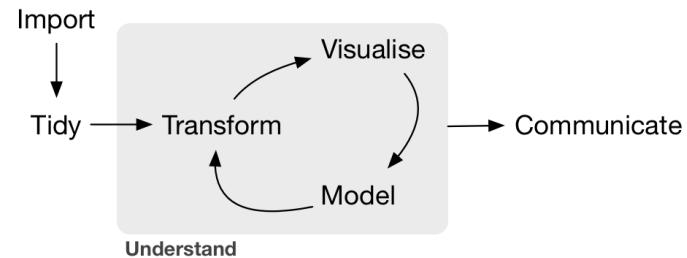
## Science:

- when is X most abundant?
- what is Y's habitat preferences?



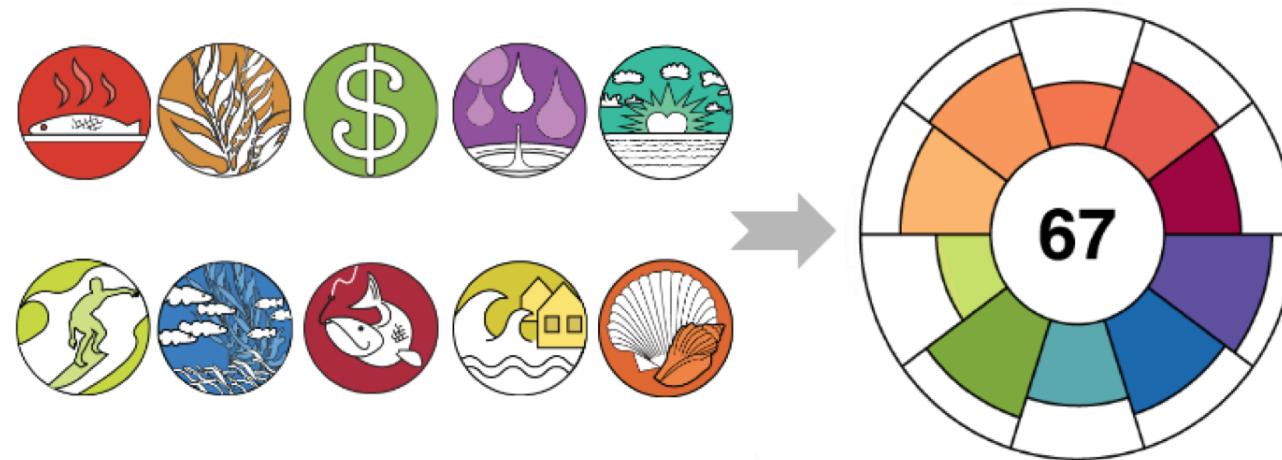
## Data science:

- how do I import my data?
- how do I subset the years I want?



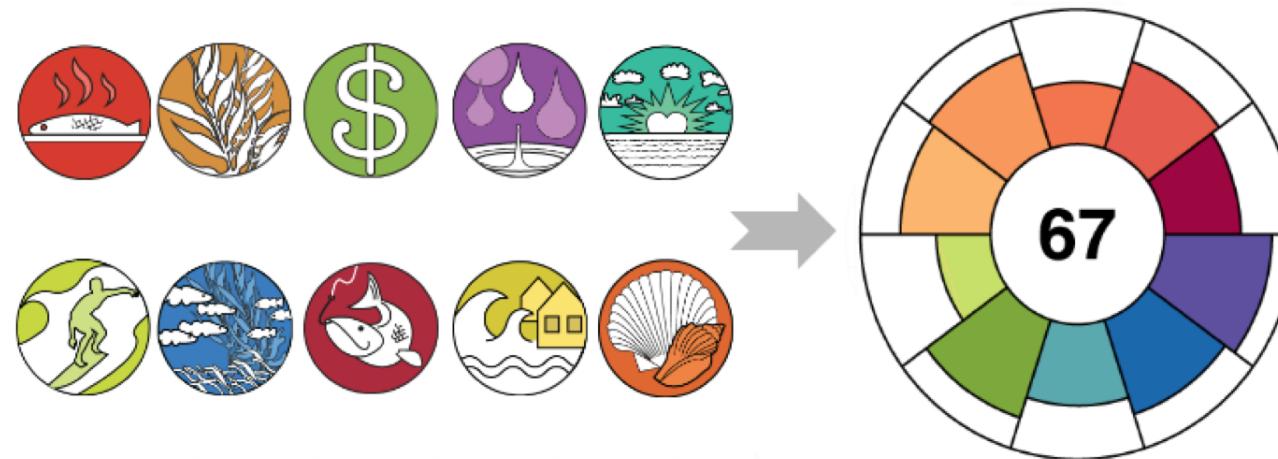
# Ocean Health Index

a method to score benefits that oceans provide to people



# Ocean Health Index

a method to score benefits that oceans provide to people

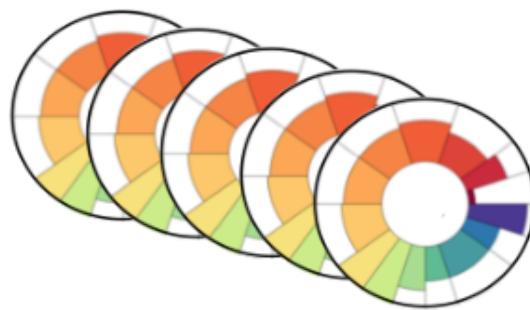


- we struggled to reproduce our own work
- we had focused only on scientific methods, not data prep (= data science)

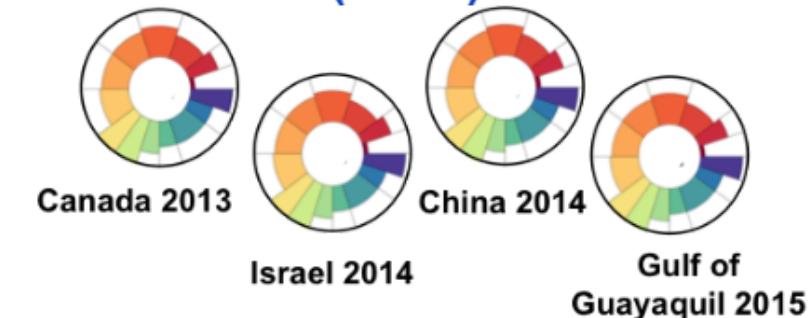
# Ocean Health Index

a method to score benefits that oceans provide to people

## Global Assessments 2012, 2013, 2014, 2015, 2016



## Independent Assessments (OHI+)



~ 15 additional in development

;

;

# Data science tools



**programming software**  
coding, graphics



**R environment; packages**  
code editor, visualization,  
communication, support



**version control system**  
track changes

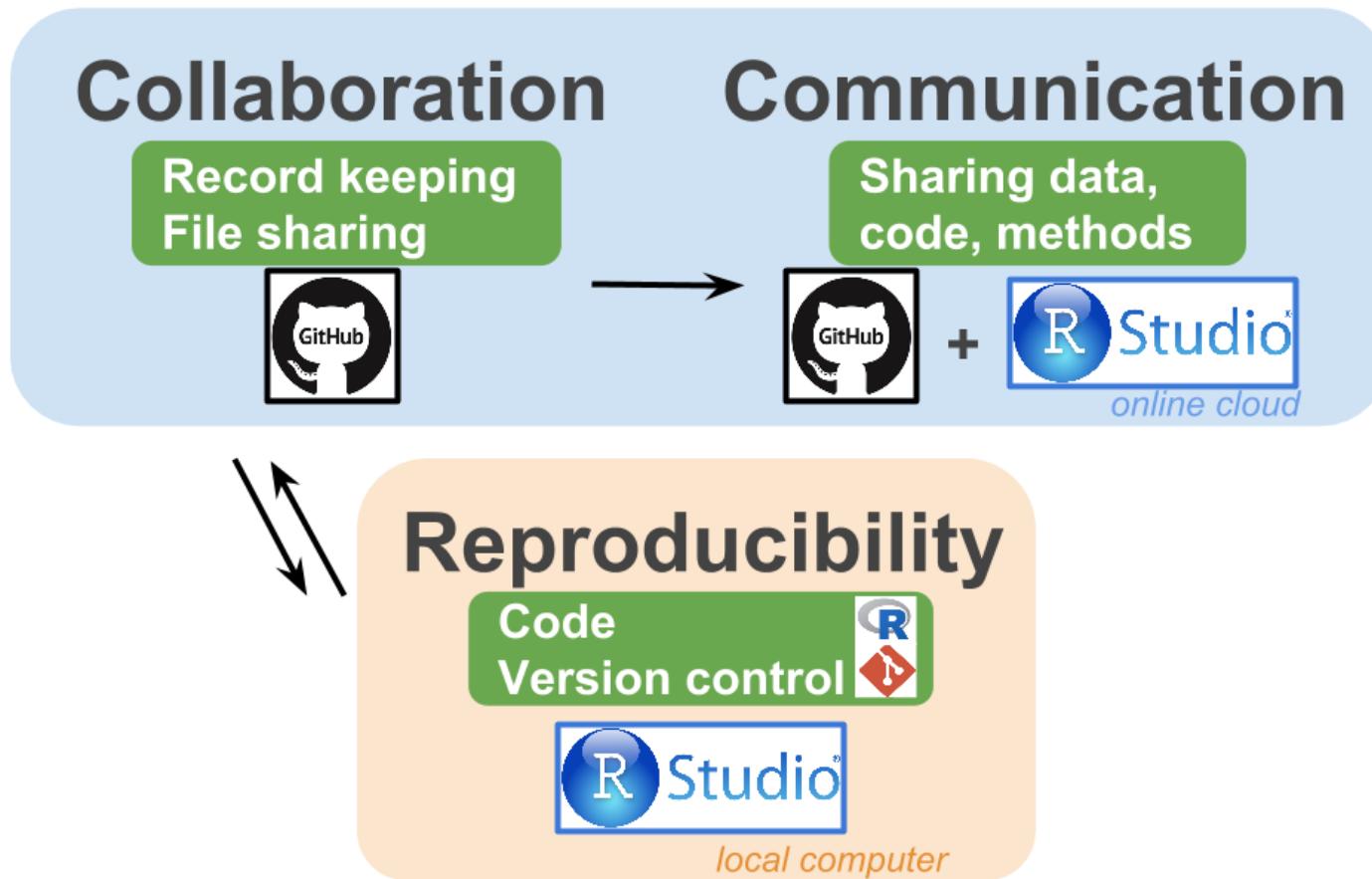


**online platform**  
organization, collaboration

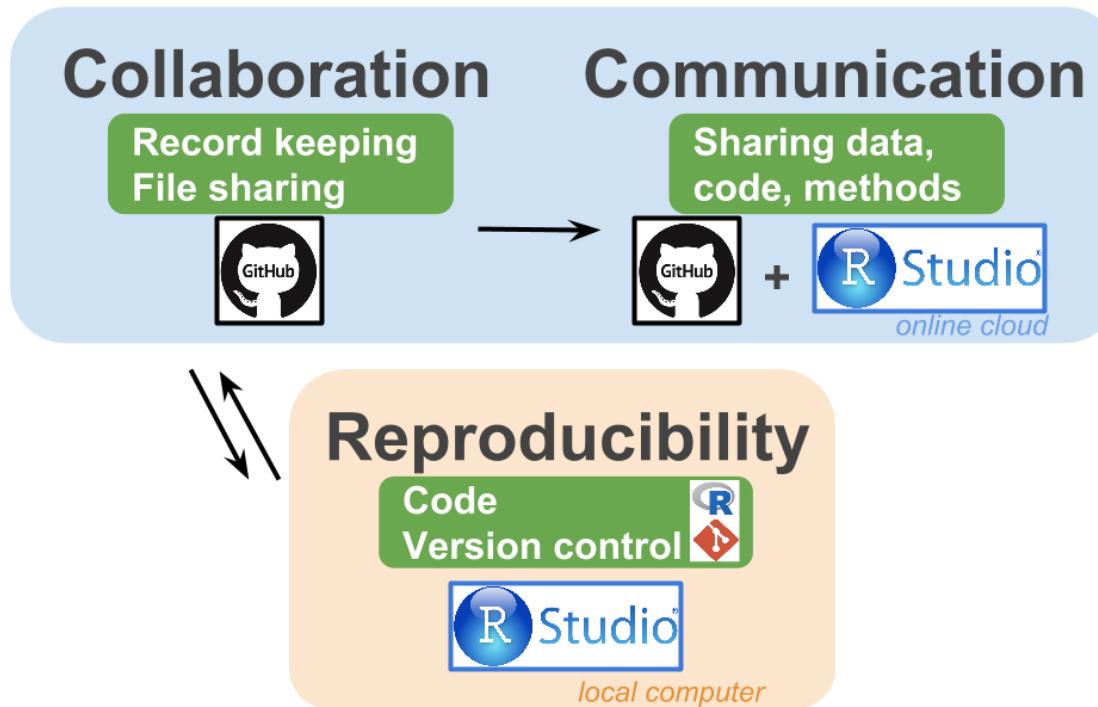
# Doing reproducible science



# Doing reproducible science



# Doing reproducible science



this talk: [jules32.github.io/opensci-talk/short](https://jules32.github.io/opensci-talk/short)

# Doing reproducible science

Better science in less time

## 1. Learn to code

- in [R](#)
- with [RStudio](#)

## 2. Use version control

- [git](#)
- with [GitHub](#)
- through [RStudio](#)

# Doing reproducible science

Better science in less time

## 1. Learn to code

- in [R](#)
- with [RStudio](#)

## 2. Use version control

- [git](#)
- with [GitHub](#)
- through [RStudio](#)

## 3. Learn in an intentional way

- feel empowered (~~in a panic~~)
- think ahead (~~for a single purpose~~)
- with a community (~~in isolation~~)

# Great resources

Free online books, trainings, and webinars:

- [ohi-science.org/news: Resources for R and Data Science](https://ohi-science.org/news/resources-for-r-and-data-science)

Recent academic publications:

- "How open science helps researchers succeed" ( [\\_\\_\\_\\_\\_](#) )
- "Best Practices for Scientific Computing" ( [\\_\\_\\_\\_\\_](#) )
- "Good Enough Practices for Scientific Computing" ( [\\_\\_\\_\\_\\_](#) )
- "Practical Computing for Biologists" ( [\\_\\_\\_\\_\\_](#) )

# Thank you

to the OHI team, colleagues, #rstats community



[ohi-science.org](http://ohi-science.org)

email: [lowndes @nceas.ucsb.edu](mailto:lowndes@nceas.ucsb.edu)

twitter: [@juliesquid](https://twitter.com/juliesquid)

website: [jules32.github.io](https://jules32.github.io)

talk url: [jules32.github.io/opensci-talk/short](https://jules32.github.io/opensci-talk/short)

NCEAS is hiring: [nceas.ucsb.edu/positionsopen](https://nceas.ucsb.edu/positionsopen)

