Data visualization for ecological analysis and restoration

S. Kyle McKay, Ph.D., P.E.
Ecosystem Management and Restoration Research Program
(EMRRP) Webinar Series
February 2020

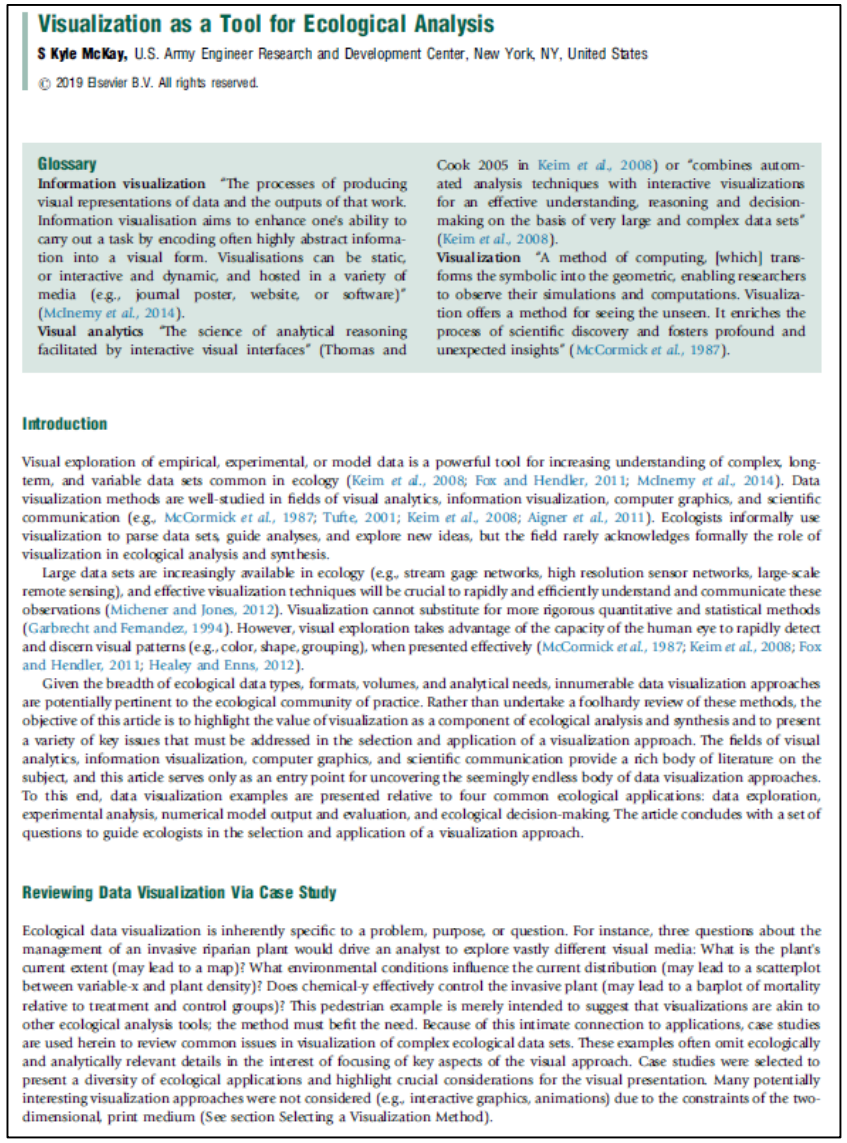
US Army Corps of Engineers

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Overview

- What is data visualization?
  - Making the case for data-based story-telling
- Stories about data viz
  - Using case studies to make broader observations
- Selecting data viz methods
  - A few potentially useful thoughts and snarky observations

Want to skip the webinar?



Visualization as a Tool for Ecological Analysis

S Kyle McKay, U.S. Army Engineer Research and Development Center, New York, NY, United States

Glossary

Information visualization "The processes of producing visual representations of data and the outputs of that work. Information visualisation aims to enhance one's ability to carry out a task by encoding often highly abstract information into a visual form. Visualisations can be static, or interactive and dynamic, and hosted in a variety of media (e.g., journal poster, website, or software)" (McInemy *et al.*, 2014).

Visual analytics "The science of analytical reasoning facilitated by interactive visual interfaces" (Thomas and Cook 2005 in Keim *et al.*, 2008) or "combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision-making on the basis of very large and complex data sets" (Keim *et al.*, 2008).

Visualization "A method of computing, [which] transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights" (McCormick *et al.*, 1987).

Introduction

Visual exploration of empirical, experimental, or model data is a powerful tool for increasing understanding of complex, long-term, and variable data sets common in ecology (Keim *et al.*, 2008; Fox and Hendler, 2011; McInemy *et al.*, 2014). Data visualization methods are well-studied in fields of visual analytics, information visualization, computer graphics, and scientific communication (e.g., McCormick *et al.*, 1987; Tufte, 2001; Keim *et al.*, 2008; Aigner *et al.*, 2011). Ecologists informally use visualization to parse data sets, guide analyses, and explore new ideas, but the field rarely acknowledges formally the role of visualization in ecological analysis and synthesis.

Large data sets are increasingly available in ecology (e.g., stream gage networks, high resolution sensor networks, large-scale remote sensing), and effective visualization techniques will be crucial to rapidly and efficiently understand and communicate these observations (Michener and Jones, 2012). Visualization cannot substitute for more rigorous quantitative and statistical methods (Garbrecht and Fernandez, 1994). However, visual exploration takes advantage of the capacity of the human eye to rapidly detect and discern visual patterns (e.g., color, shape, grouping), when presented effectively (McCormick *et al.*, 1987; Keim *et al.*, 2008; Fox and Hendler, 2011; Healey and Enns, 2012).

Given the breadth of ecological data types, formats, volumes, and analytical needs, innumerable data visualization approaches are potentially pertinent to the ecological community of practice. Rather than undertake a foolhardy review of these methods, the objective of this article is to highlight the value of visualization as a component of ecological analysis and synthesis and to present a variety of key issues that must be addressed in the selection and application of a visualization approach. The fields of visual analytics, information visualization, computer graphics, and scientific communication provide a rich body of literature on the subject, and this article serves only as an entry point for uncovering the seemingly endless body of data visualization approaches. To this end, data visualization examples are presented relative to four common ecological applications: data exploration, experimental analysis, numerical model output and evaluation, and ecological decision-making. The article concludes with a set of questions to guide ecologists in the selection and application of a visualization approach.

Reviewing Data Visualization Via Case Study

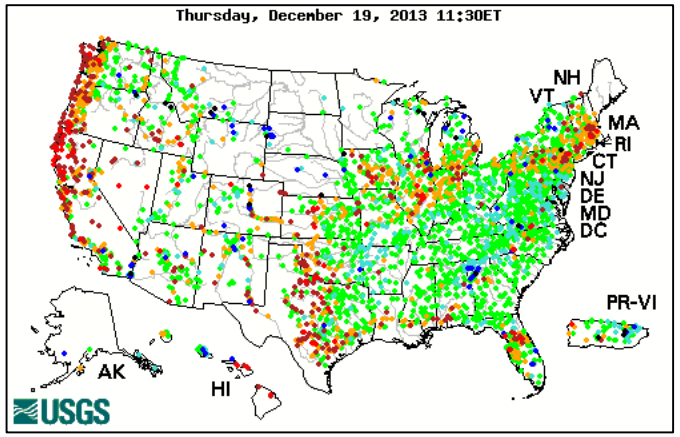
Ecological data visualization is inherently specific to a problem, purpose, or question. For instance, three questions about the management of an invasive riparian plant would drive an analyst to explore vastly different visual media: What is the plant's current extent (may lead to a map)? What environmental conditions influence the current distribution (may lead to a scatterplot between variable-x and plant density)? Does chemical-y effectively control the invasive plant (may lead to a barplot of mortality relative to treatment and control groups)? This pedestrian example is merely intended to suggest that visualizations are akin to other ecological analysis tools; the method must befit the need. Because of this intimate connection to applications, case studies are used herein to review common issues in visualization of complex ecological data sets. These examples often omit ecologically and analytically relevant details in the interest of focusing of key aspects of the visual approach. Case studies were selected to present a diversity of ecological applications and highlight crucial considerations for the visual presentation. Many potentially interesting visualization approaches were not considered (e.g., interactive graphics, animations) due to the constraints of the two-dimensional, print medium (See section Selecting a Visualization Method).

What is data visualization?

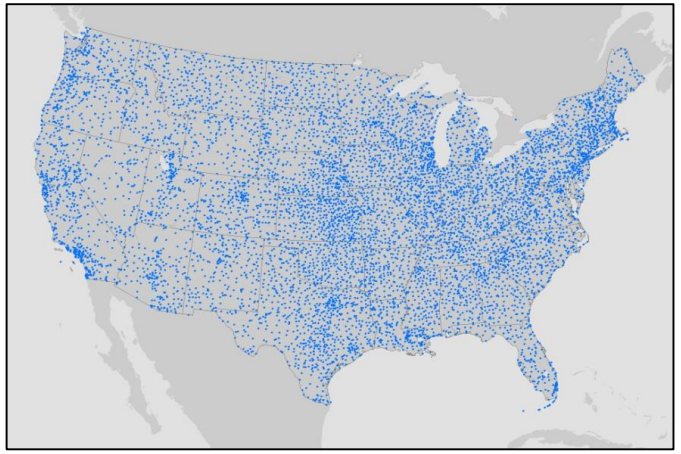
“Big data” isn’t a new phenomenon

NSF’s Long-Term Ecological Research (LTER) Stations

USGS Streamflow Gage Network



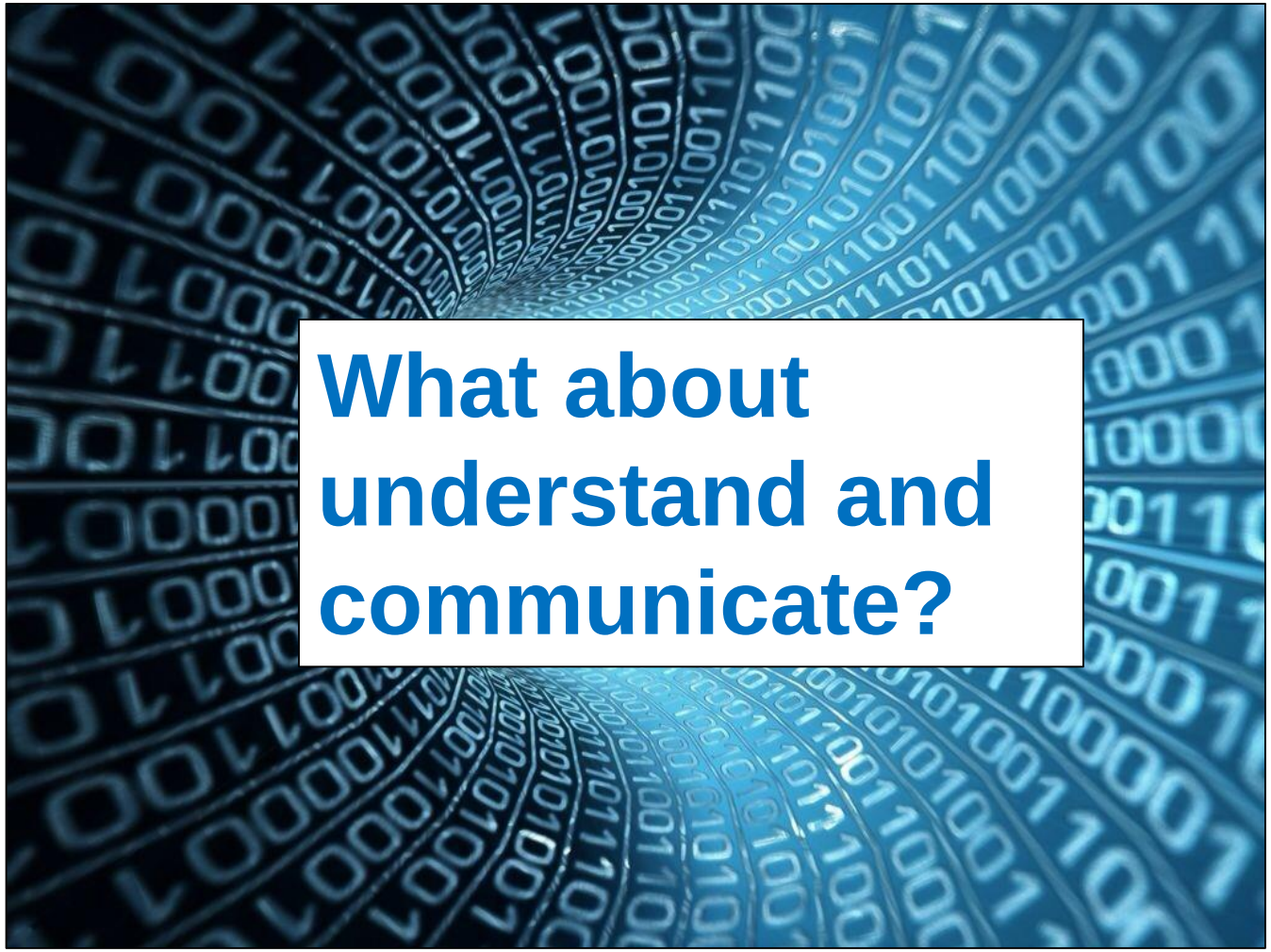
Weather Stations



“Palermo Stone” used to record Nile River flood (~5,000 years old)

We can collect data on many processes at small intervals, but what are we going to do with it?

- Funding
- Collection
- Verification
- Storage
- Curation
- Sharing
- Analysis
- …



Data visualization: a rich discipline of its own!

- **Information visualization**: “the processes of producing visual representations of data and the outputs of that work. *Information visualisation aims to enhance one’s ability to carry out a task* by encoding often highly abstract information into a visual form. Visualisations can be static, or interactive and dynamic, and hosted in a variety of media (e.g., journal poster, website, or software).” (McInerny et al. 2015)

- **Visualization**: “a method of computing, [which] transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. *Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights*.” (McCormick et al. 1987)

- **Visual analytics**: “the science of analytical reasoning facilitated by interactive visual interfaces” (Thomas and Cook 2005 in Keim et al. 2008) or “combines automated analysis techniques with interactive visualizations for an *effective understanding, reasoning and decision making on the basis of very large and complex data sets*.” (Keim et al. 2008)

Why is visualization useful?

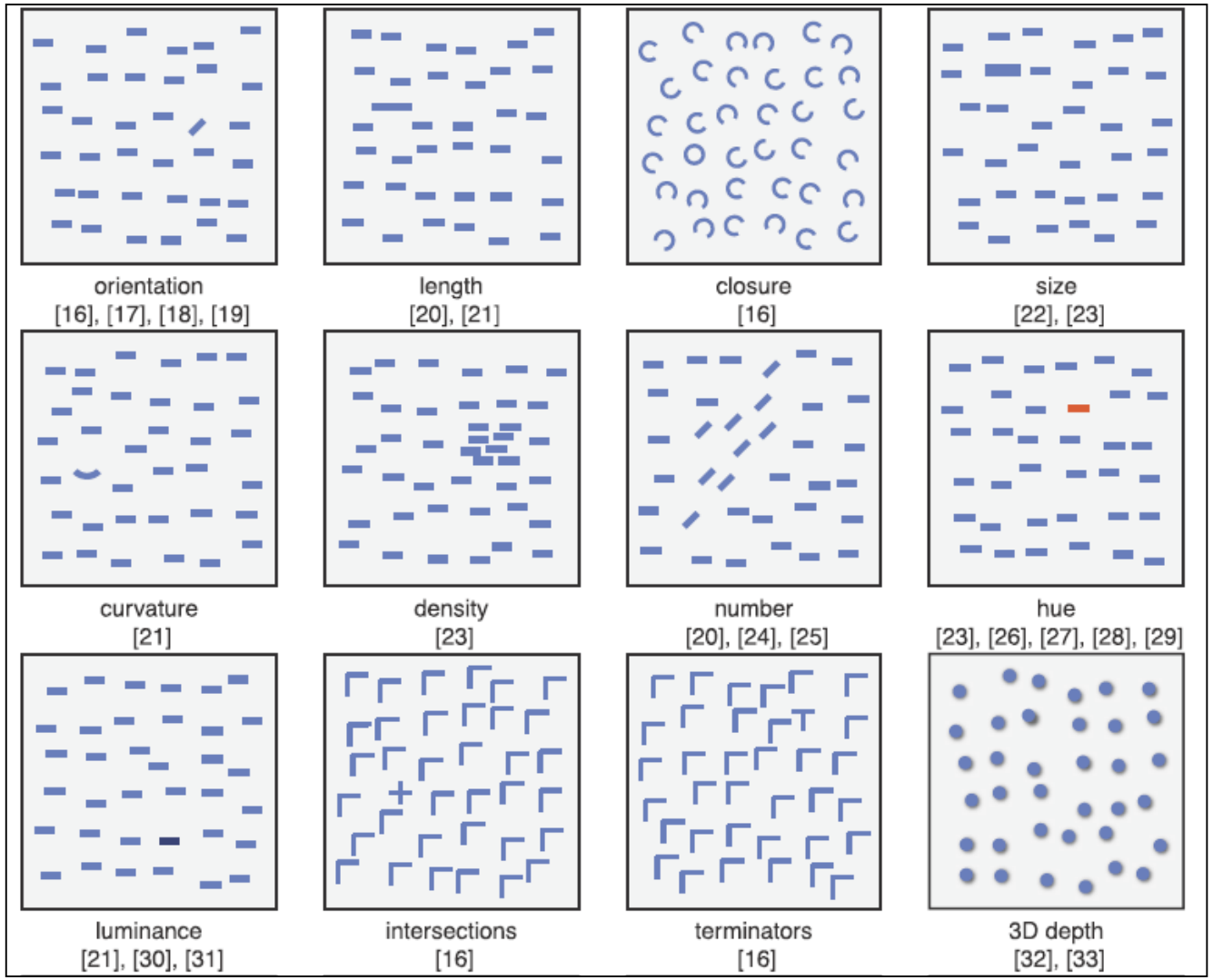
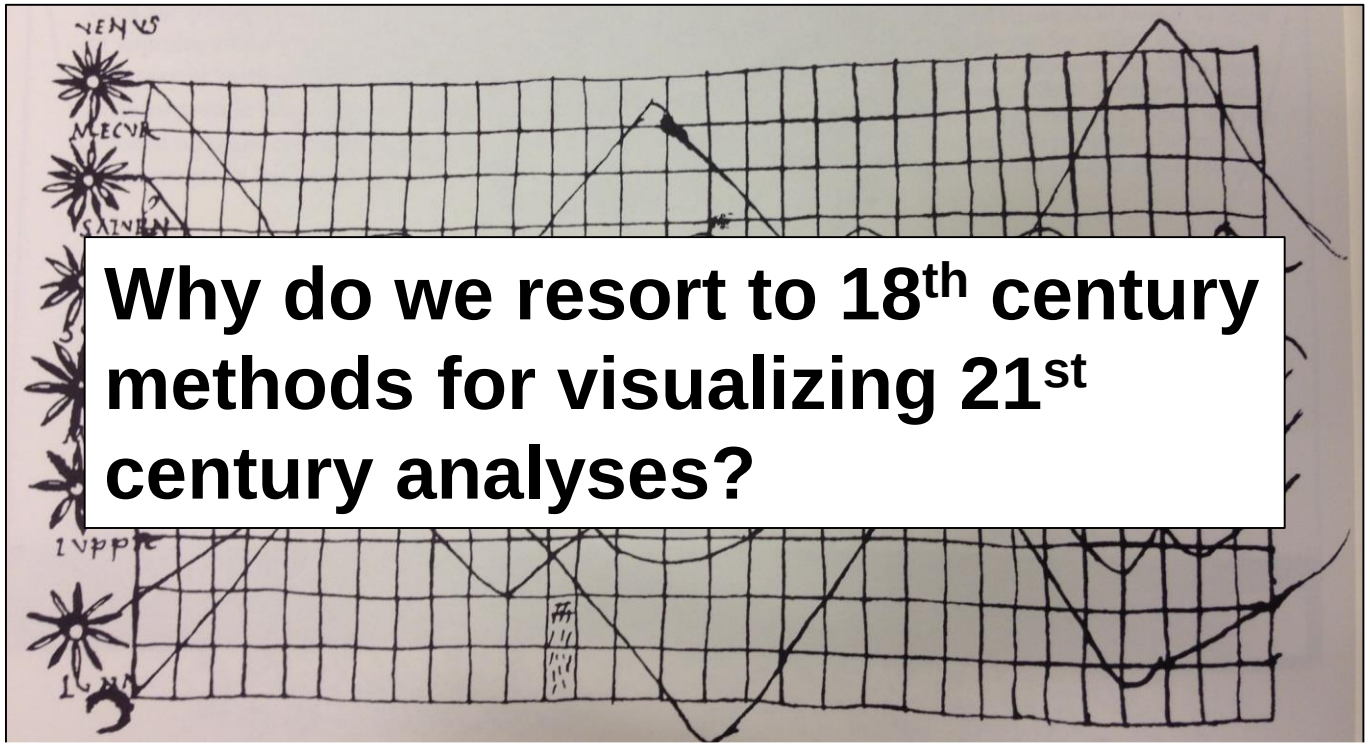
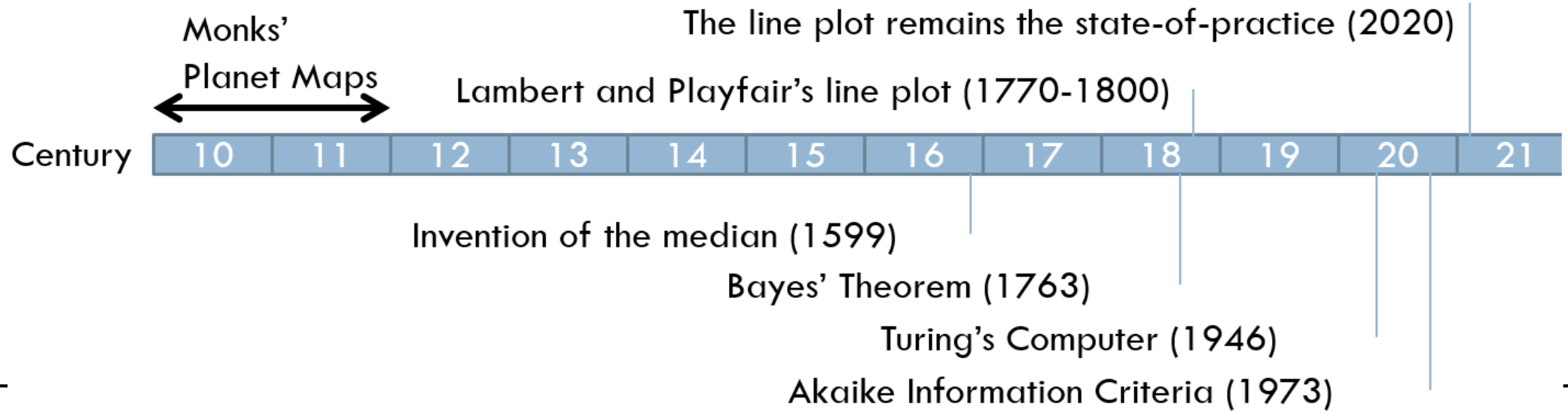


Figure: Healey and Enns. 2012. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics, 18* (7), 1170–1188.



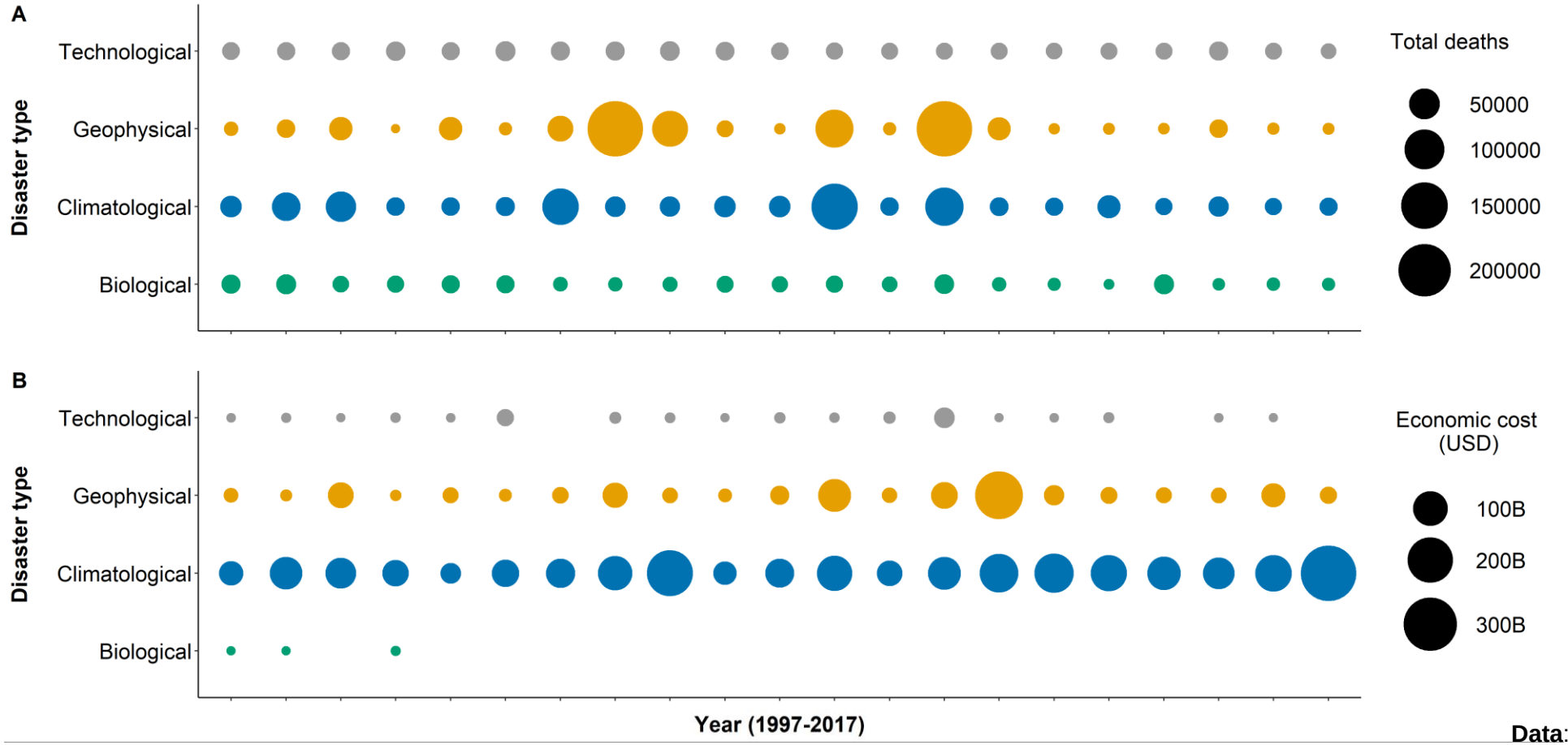
Why do we resort to 18th century methods for visualizing 21st century analyses?



Stories about data viz

- *Problem statement*: disaster management
- *Data exploration*: long-term shrimp population data
- *Data analysis*: butterfly flight trials
- *Ecological model development*: oyster restoration
- *Management alternatives*: environmental flows
- *Stakeholder opinions*: urban stream restoration in Georgia
- *Informing decision-making*: multi-criteria urban restoration in NYC

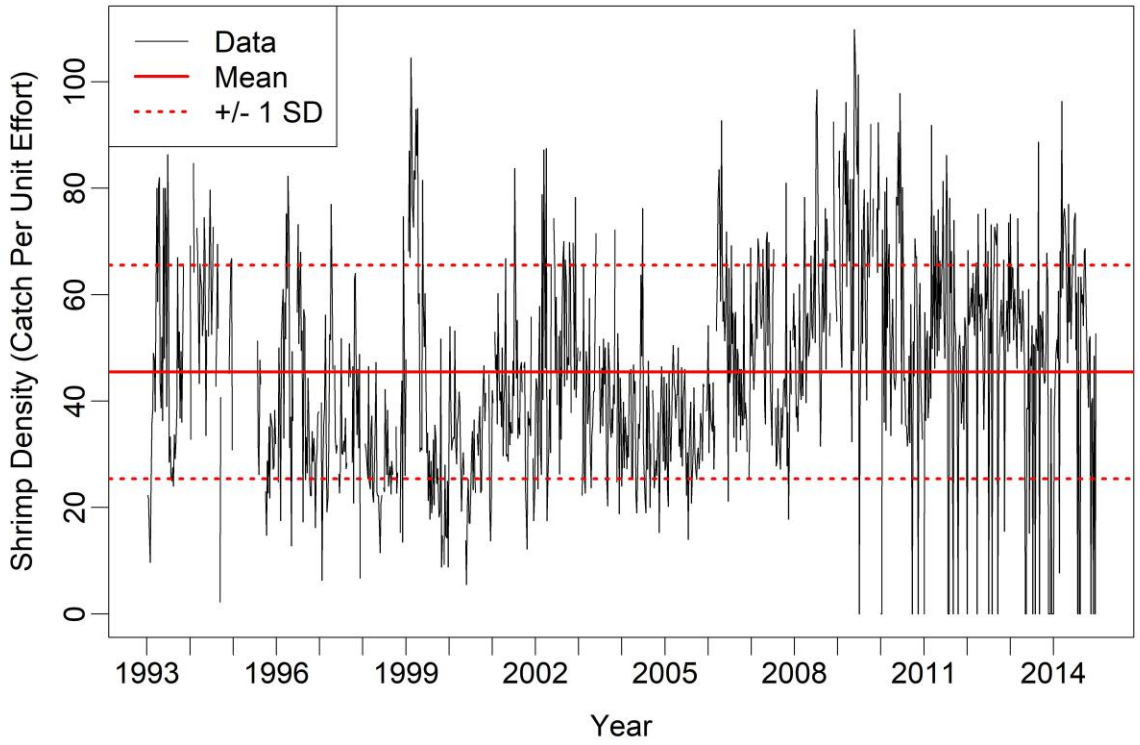
Problem Statement:
Disaster Management

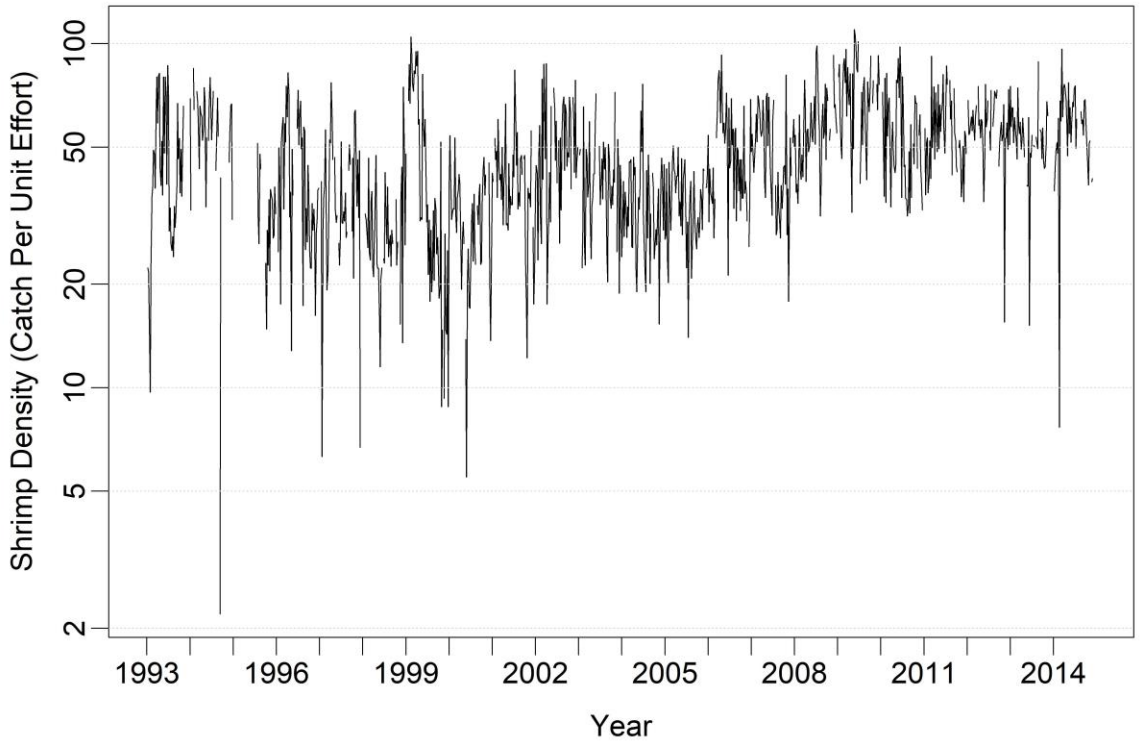


Data: EM-DAT (www.emdat.be)
Figure: McKay, Hernández-Abrams, et al. Environmental scientists' role in disaster management. *In preparation*.

Data Exploration:

Long-term shrimp sampling in Puerto Rico

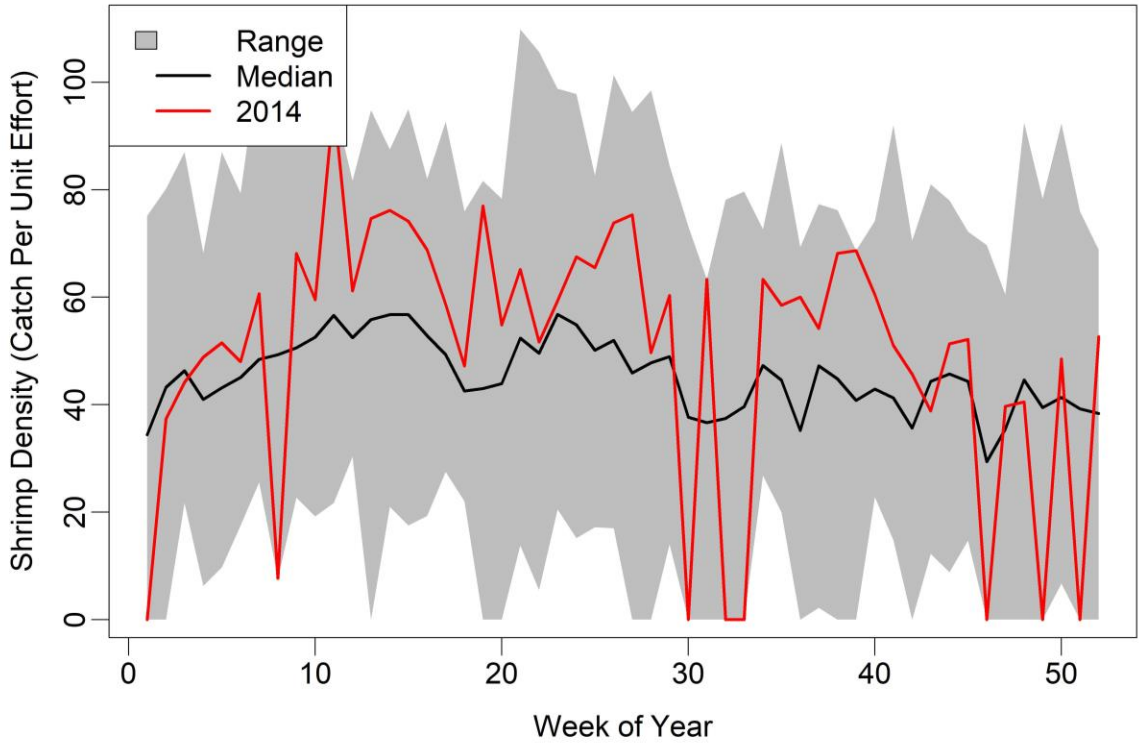


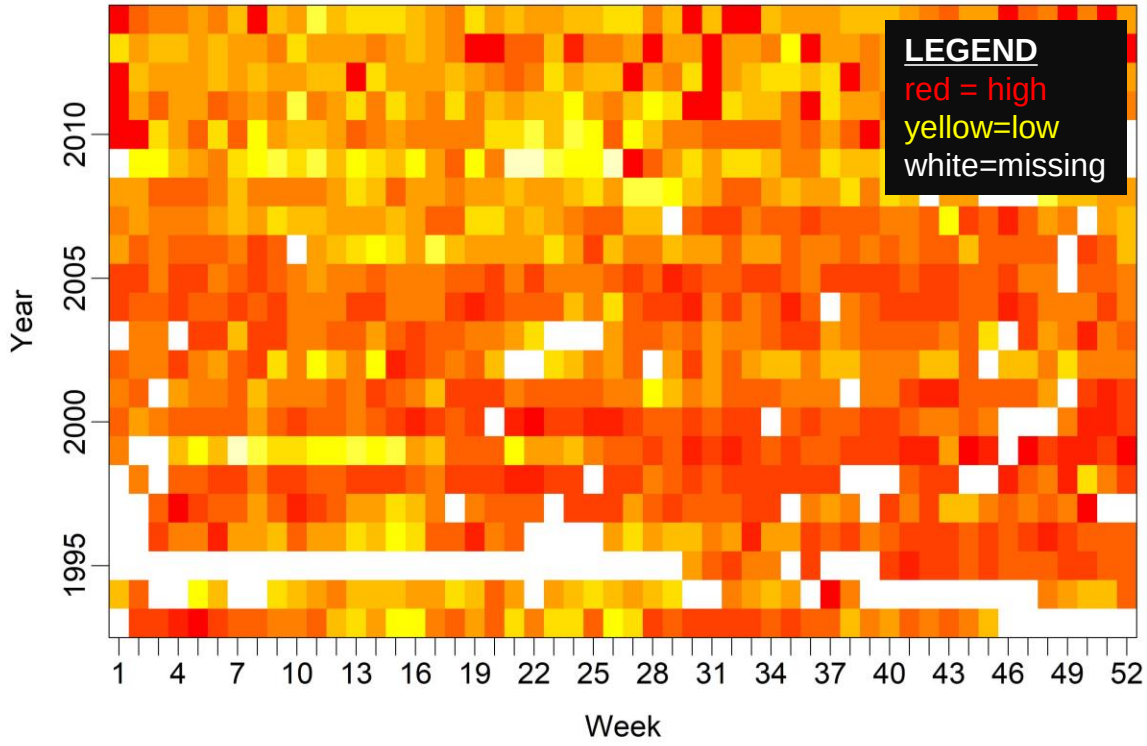


Data: Weekly freshwater shrimp (*Atya lanipes*) trapping densities collected at pool-0 of the Luquillo Long-Term Ecological Research (LTER-LUQ) site in Puerto Rico

Data Exploration:

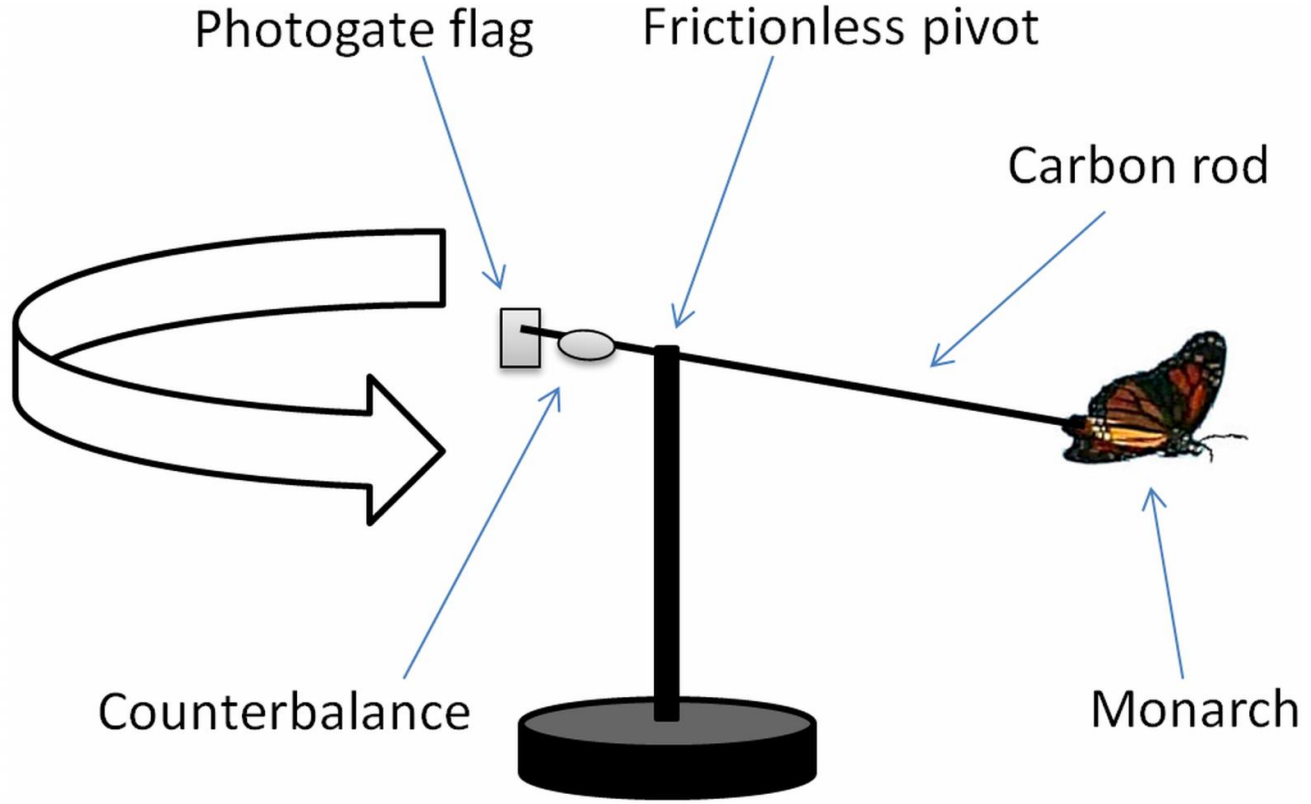
Long-term shrimp sampling in Puerto Rico





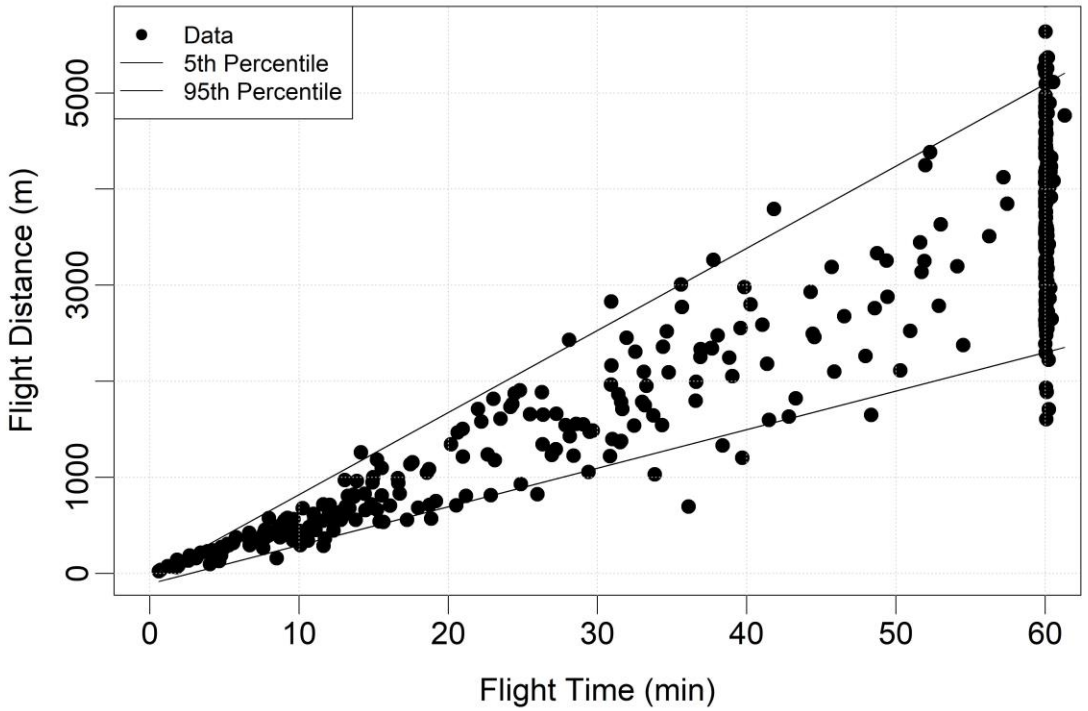
Data: Weekly freshwater shrimp (*Atya lanipes*) trapping densities collected at pool-0 of the Luquillo Long-Term Ecological Research (LTER-LUQ) site in Puerto Rico

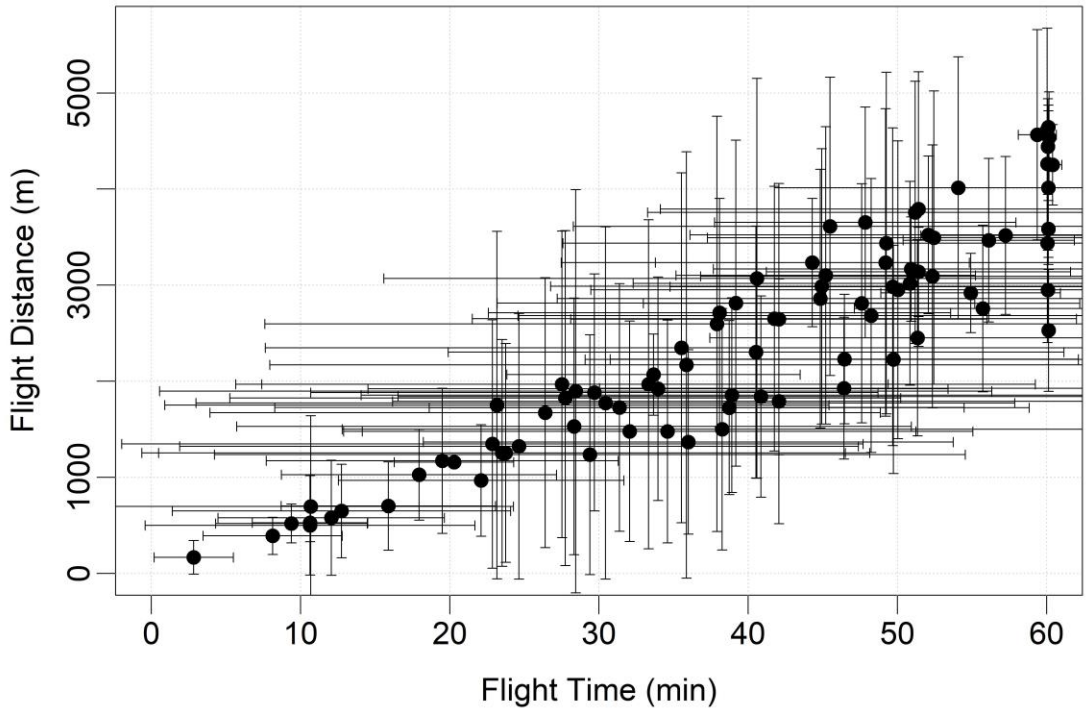
Data Analysis:
Butterfly flight trials



Schematic: Davis, Chi, Bradley, and Altizer. 2012. The redder the better: Wing color predicts flight performance in Monarch butterflies. PLoS ONE 7(7): e41323.

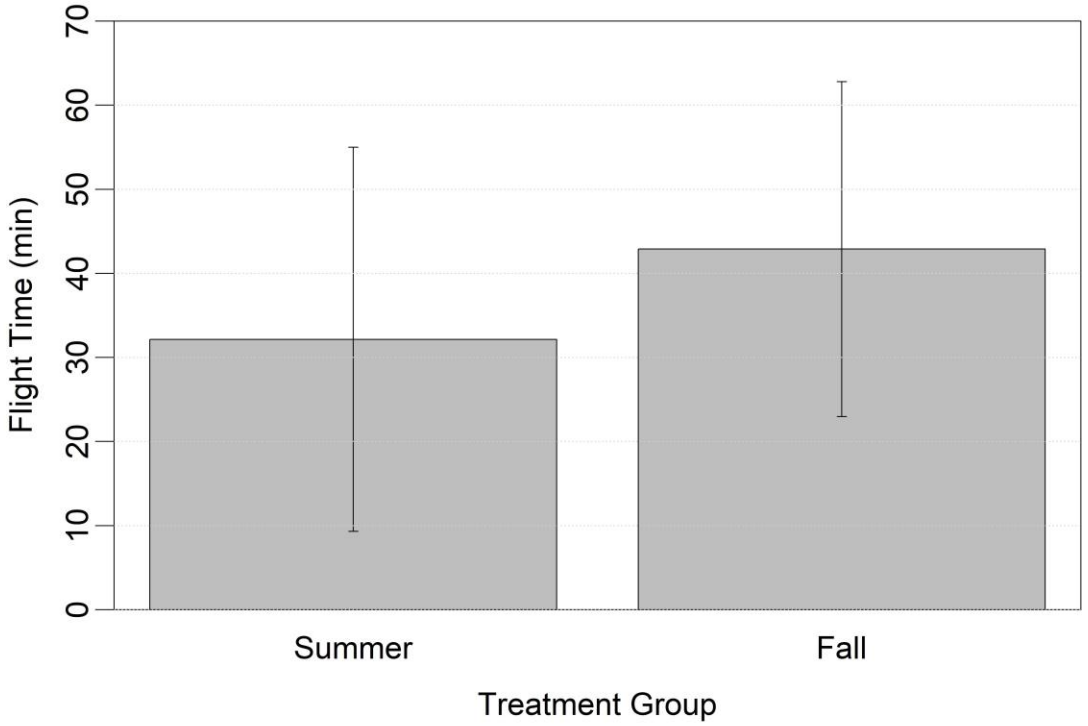
Data Analysis:
Butterfly flight trials

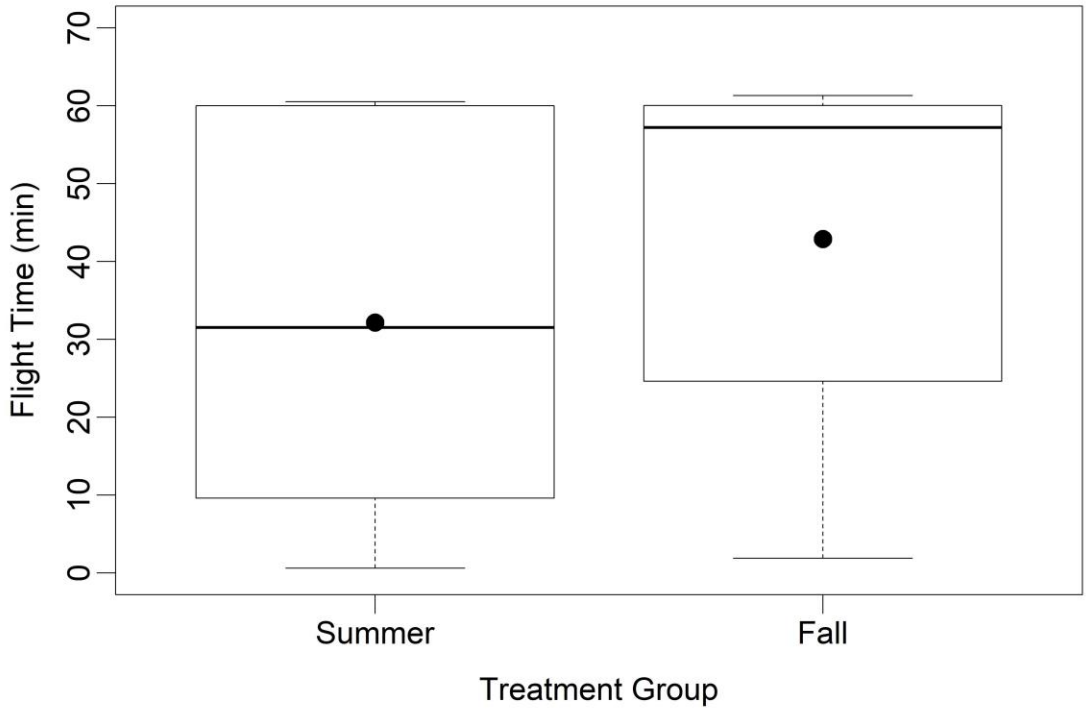




Data: McKay A.F., Ezenwa V.O., and Altizer S. 2017. Unravelling the costs of flight for immune defenses in the migratory monarch butterfly. Integrative and Comparative Biology, 56 (2), 278–289.

Data Analysis:
Butterfly flight trials

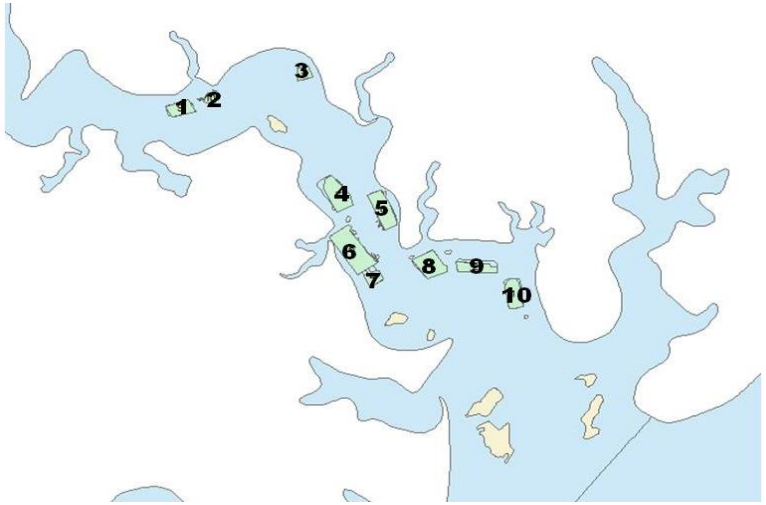




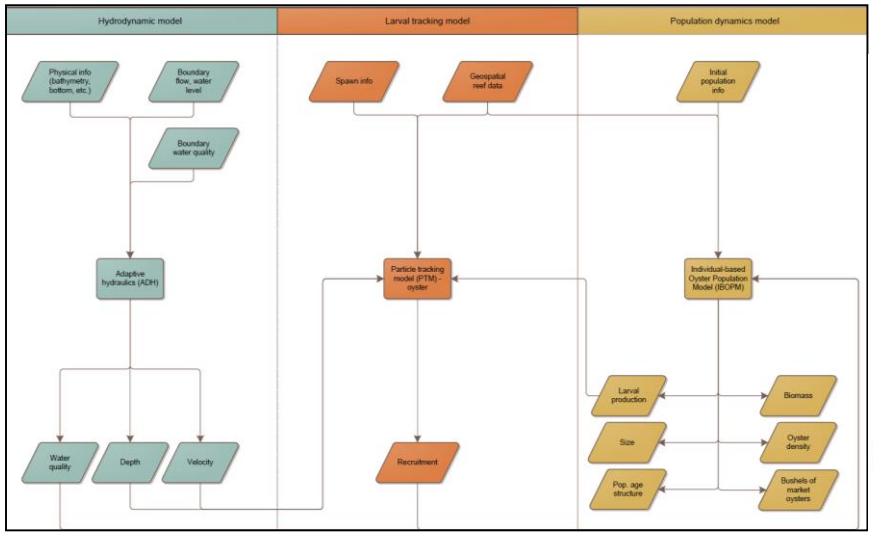
Data: McKay A.F., Ezenwa V.O., and Altizer S. 2017. Unravelling the costs of flight for immune defenses in the migratory monarch butterfly. Integrative and Comparative Biology, 56 (2), 278–289.

Model Development:
Oyster Restoration

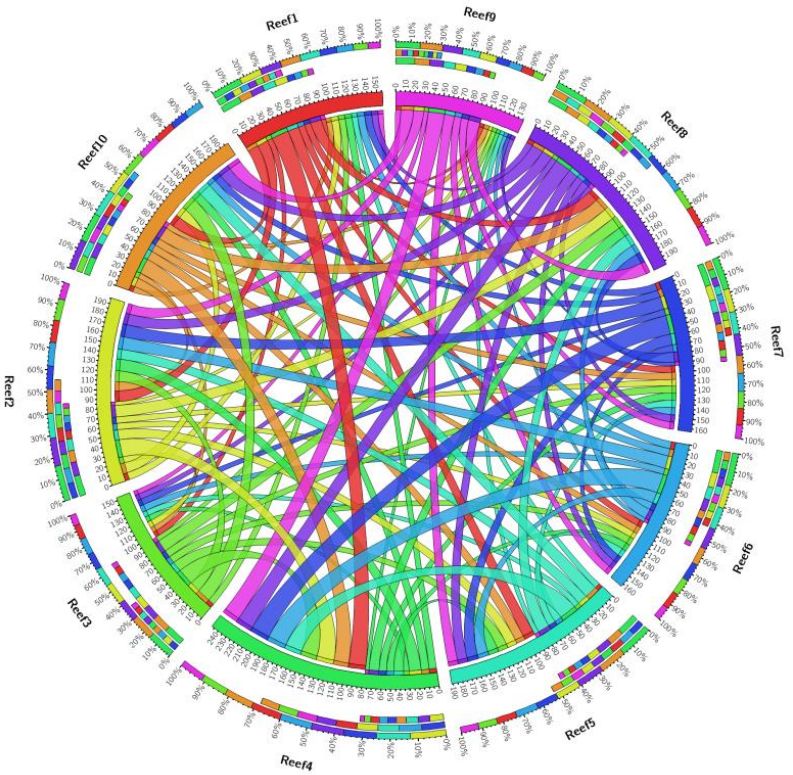
Model Purpose



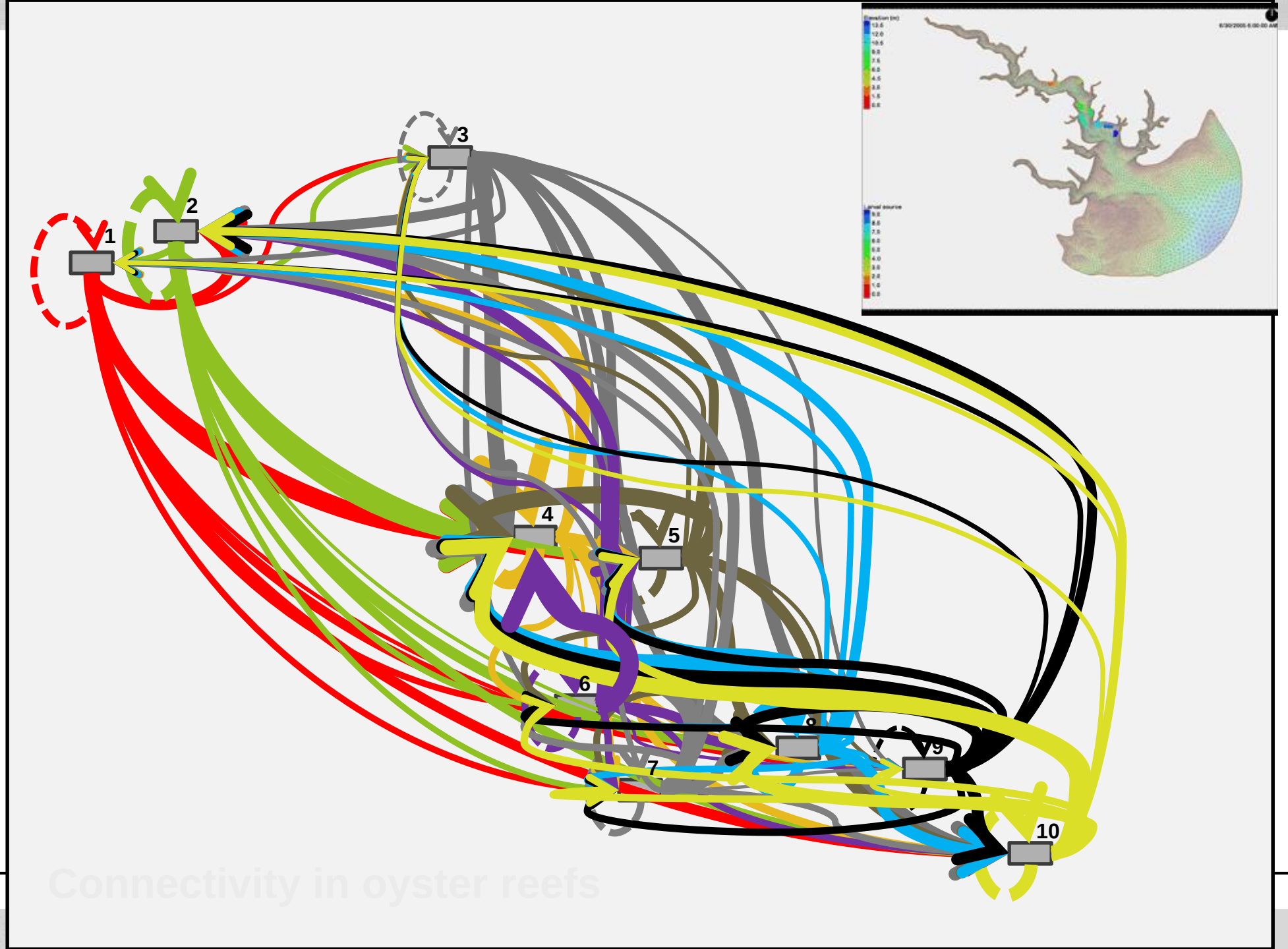
Modeling Process



Model Outcomes



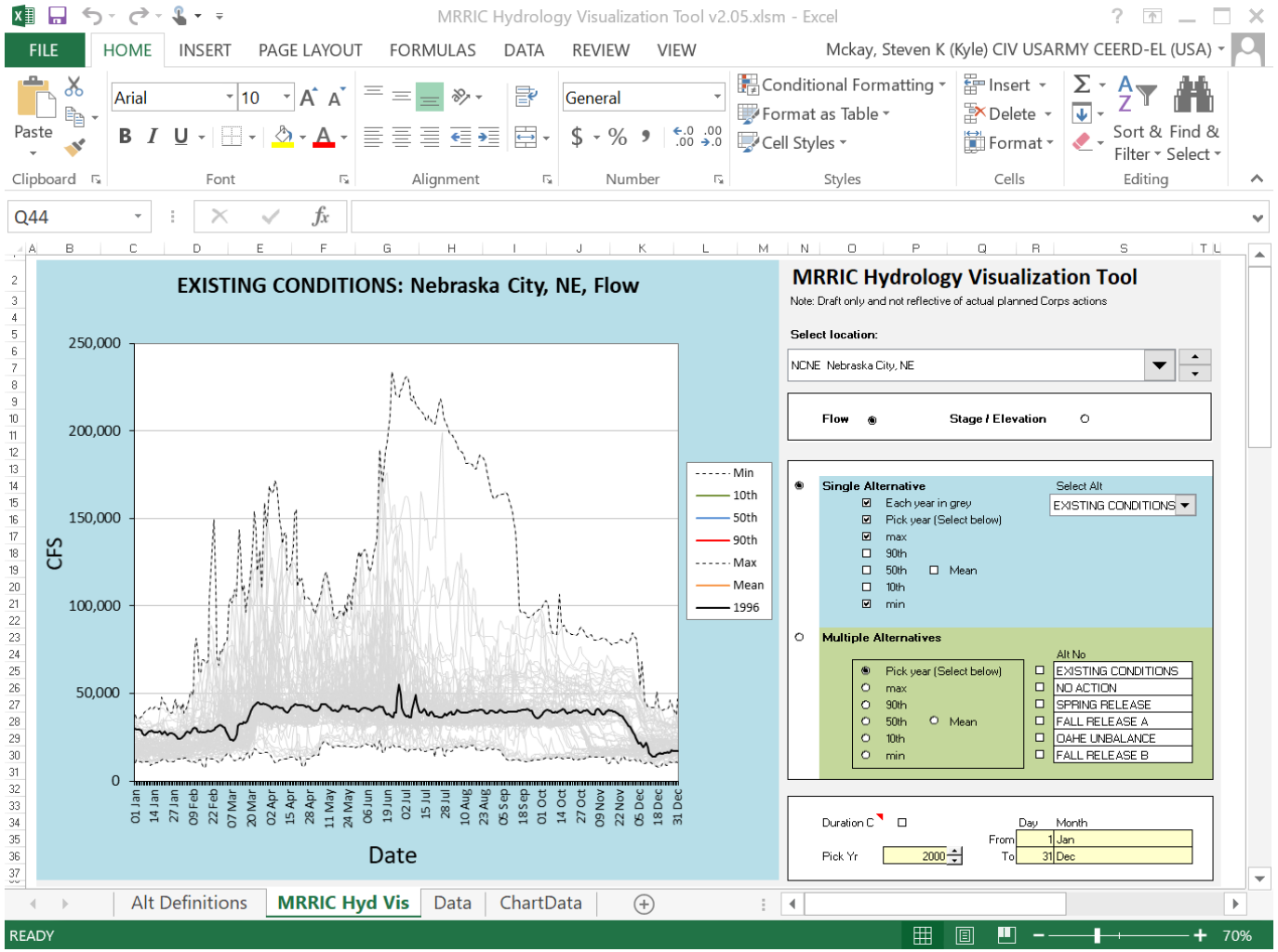
Data and Figures: Kjelland M.E., Piercy C.D., Lackey T., and Swannack T.M. 2015. An integrated modeling approach for elucidating the effects of different management strategies on Chesapeake Bay oyster metapopulation dynamics. Ecological Modelling, 308, 45-62.
POC: Todd Swannack



Connectivity in oyster reefs

Management Alternatives:

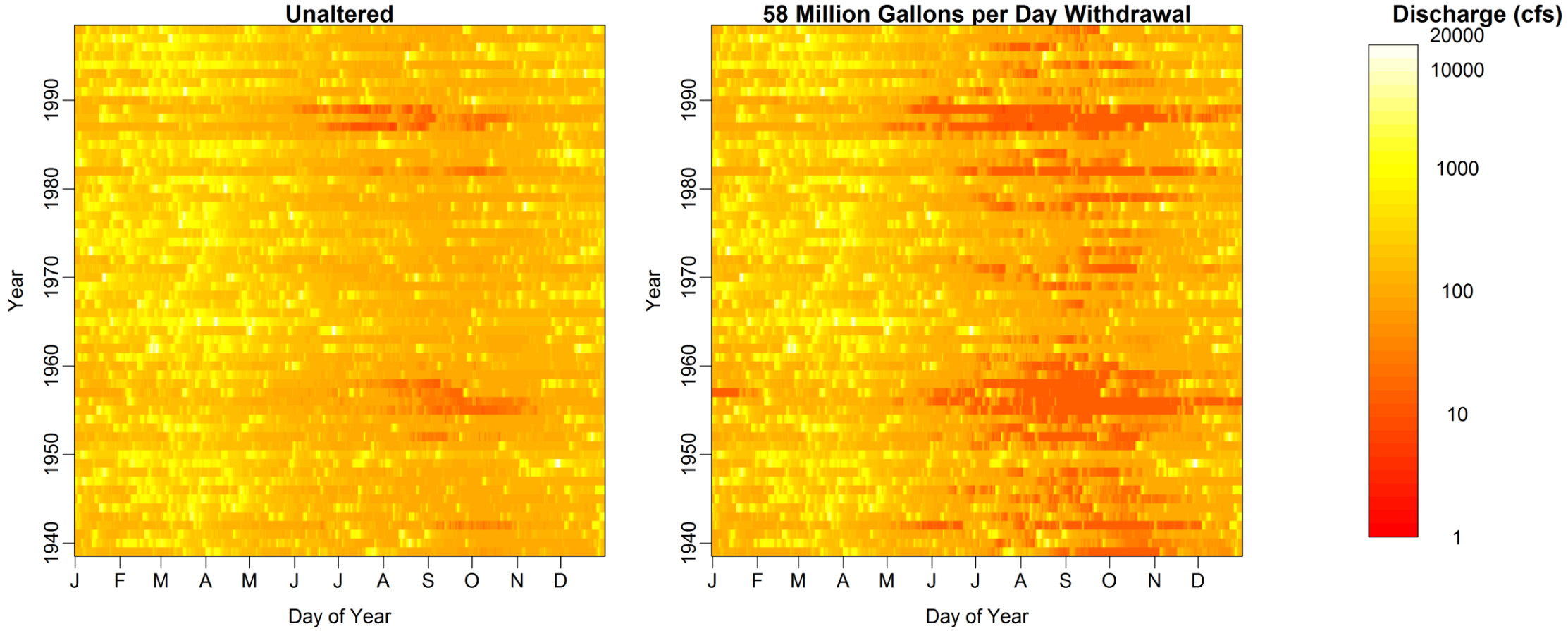
Missouri River flow management



Tool: Long, G. 2010. Brief guide to using the MRRIC hydrology visualization tool. Compass Resource Management Ltd.
POC: Jeff Tripe (NWK)

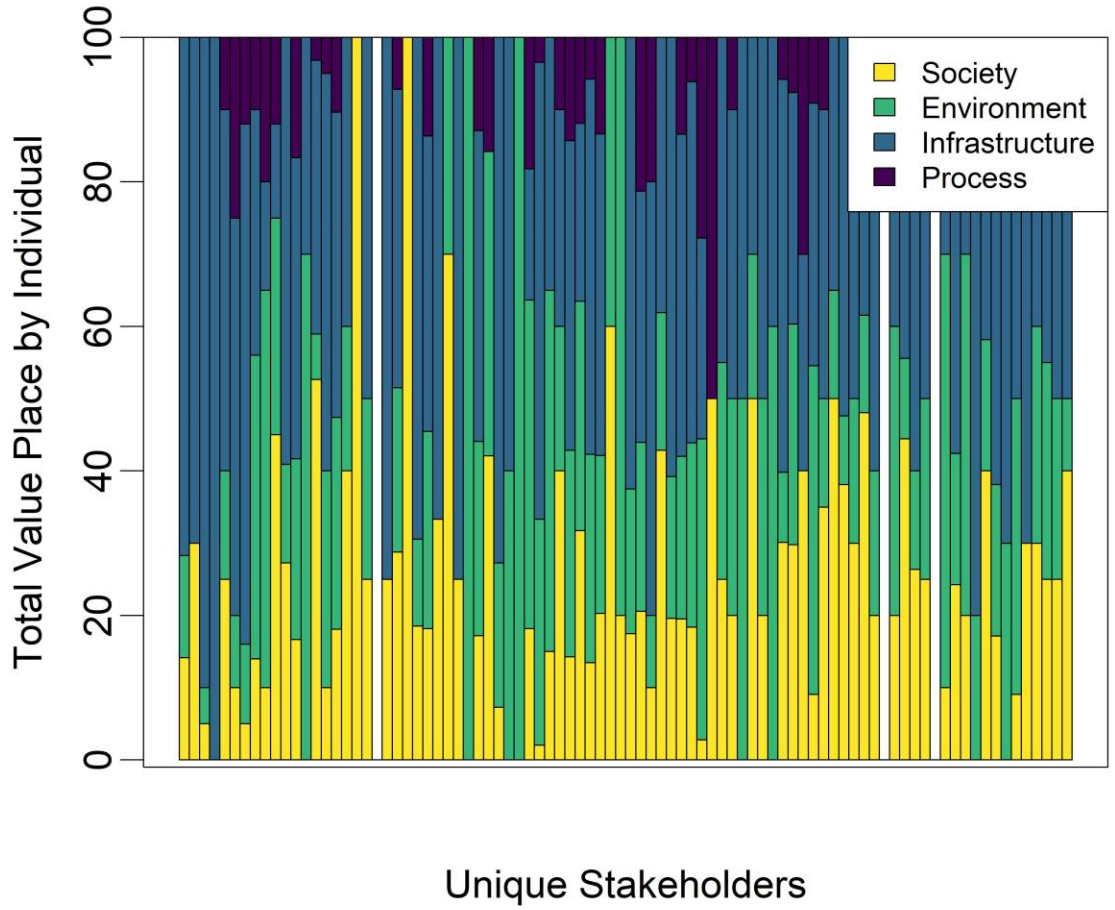
Management Alternatives:

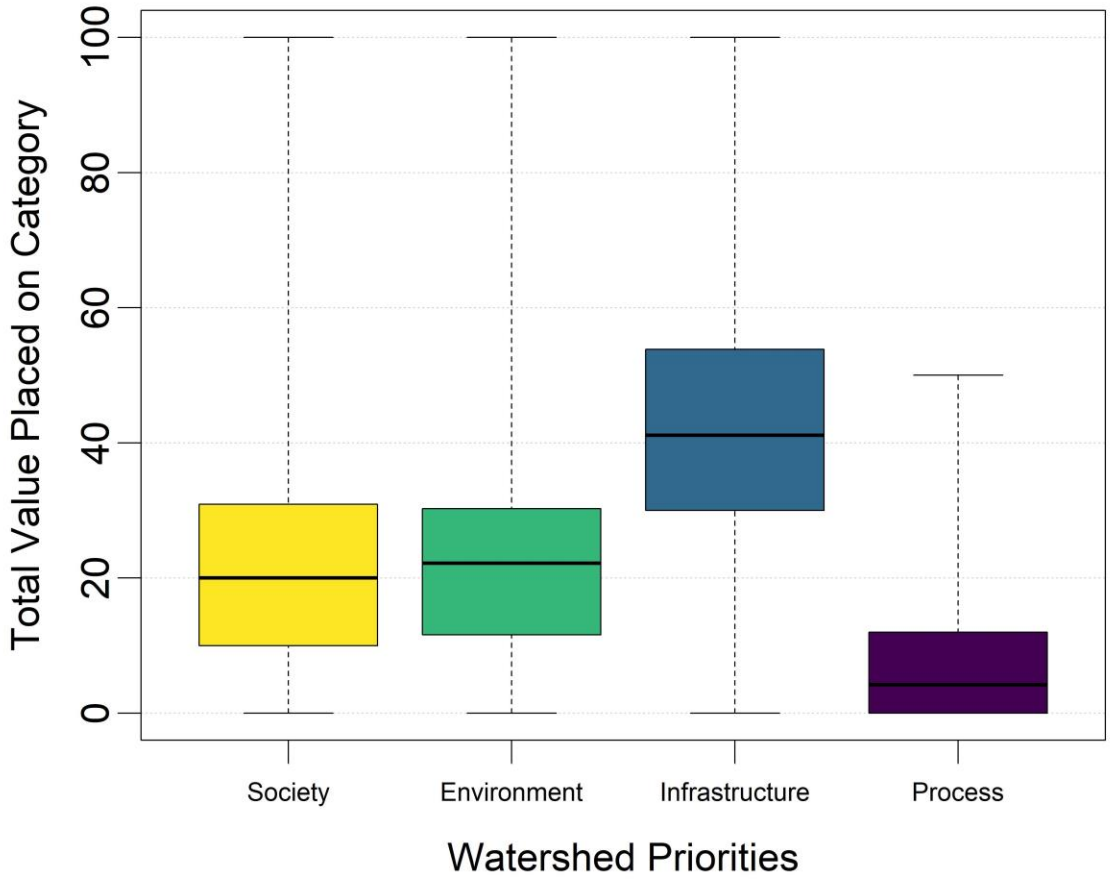
Middle Oconee River flow management



Data: McKay S.K. 2015. Quantifying trade-offs associated with hydrologic environmental flow methods. Journal of the American Water Resources Association, 51 (6), 1508-1518.

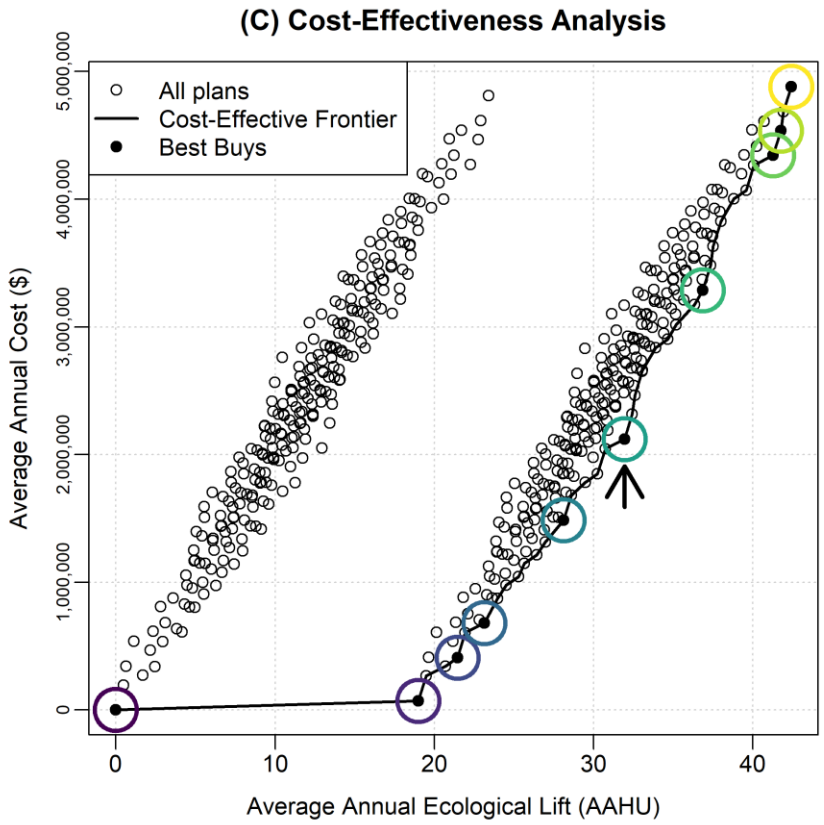
Working with Stakeholders:
Proctor Creek, Atlanta, Georgia

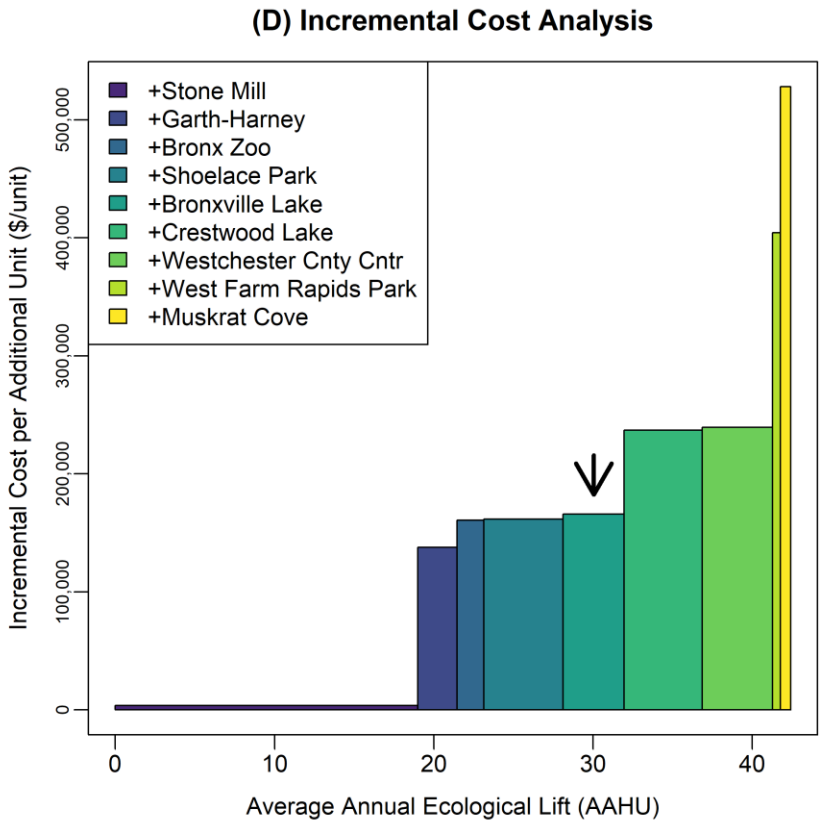




Data (revised from): USACE Mobile District. 2017. Integrated Water Resource Management Comprehensive Plan Recommendations. Chapter 7 in Proctor Creek Feasibility Study.
POC: Meredith Ladart (SAM)

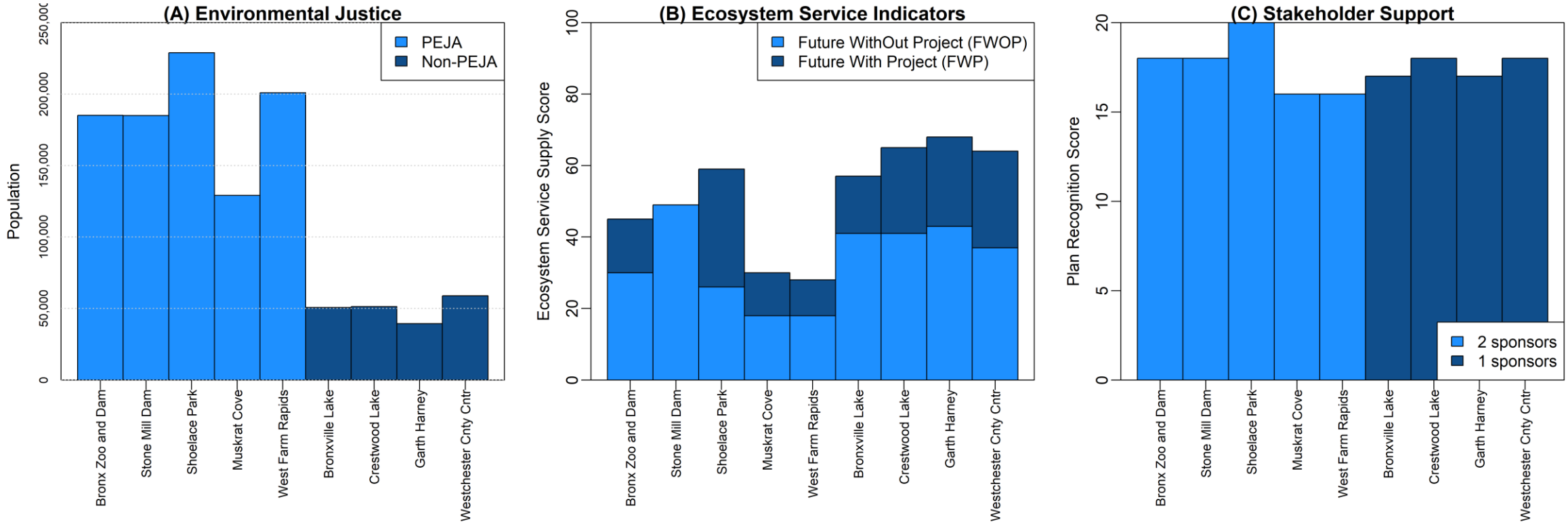
Informing Decision Making:
Hudson-Raritan Estuary Restoration





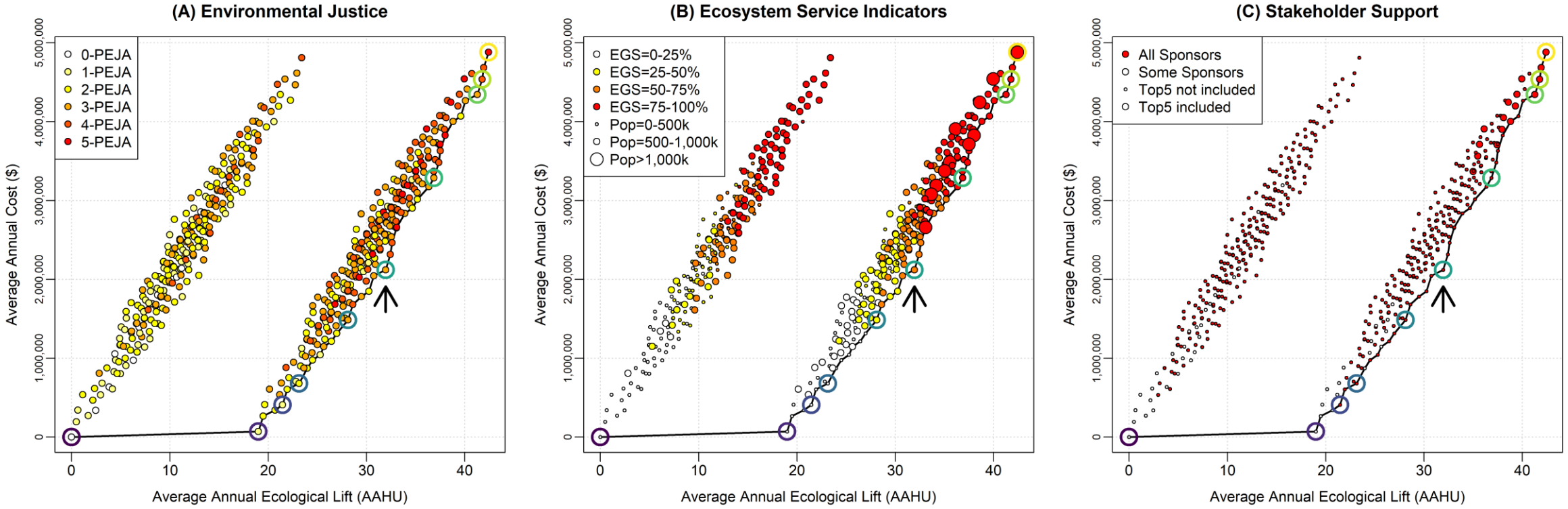
Methods: McKay et al. Incorporating multiple lines of evidence in urban stream restoration decision-making. In review at *JAWRA*.
POC: Lisa Baron (NAN)

Informing Decision Making:
Hudson-Raritan Estuary Restoration



Methods: McKay et al. Incorporating multiple lines of evidence in urban stream restoration decision-making. In review at *JAWRA*.
POC: Lisa Baron (NAN)

Informing Decision Making:
Hudson-Raritan Estuary Restoration



Methods: McKay et al. Incorporating multiple lines of evidence in urban stream restoration decision-making. In review at *JAWRA*.
POC: Lisa Baron (NAN)

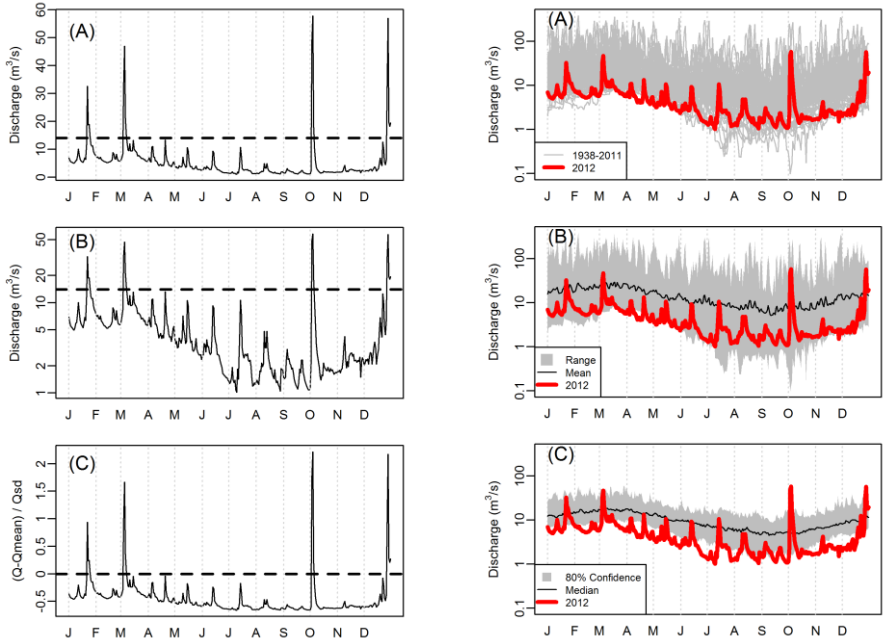
Selecting visualization methods

Adapted from other great resources:

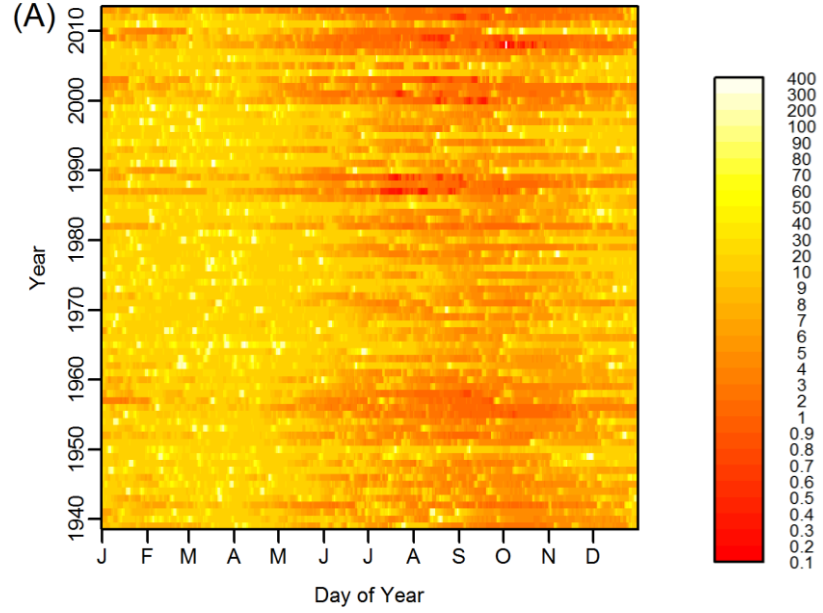
- Aigner W., Miksch S., Muller W., Schumann H., and Tominski C. 2007. Visualizing time-oriented data: A systematic view. Computers & Graphics, 31, 401-409.
- Aigner W., Miksch S., Schumann H., and Tominski C. 2011. Visualization of time-oriented data. Springer-Verlag, London.
- Kelleher C. and Wagener T. 2011. Ten guidelines for effective data visualization in scientific publications. Environmental Modelling & Software, 26, 822-827.
- Tufte E.R. 2001. The visual display of quantitative information. Graphics Press, Cheshire, Connecticut.
- Weissgerber et al. 2015. Beyond bar and line graphs: Time for a new data presentation paradigm. PLoS Biology, 13 (4), doi:10.1371/journal.pbio.1002128.

1. What is the best method for communicating a message?

- Purpose and objectives should always drive visualization method
  - Short-term data vs. long-term record?
  - Guide analysis or communicate outcomes?
  - Extremes or central tendencies?
  - Details or gist?



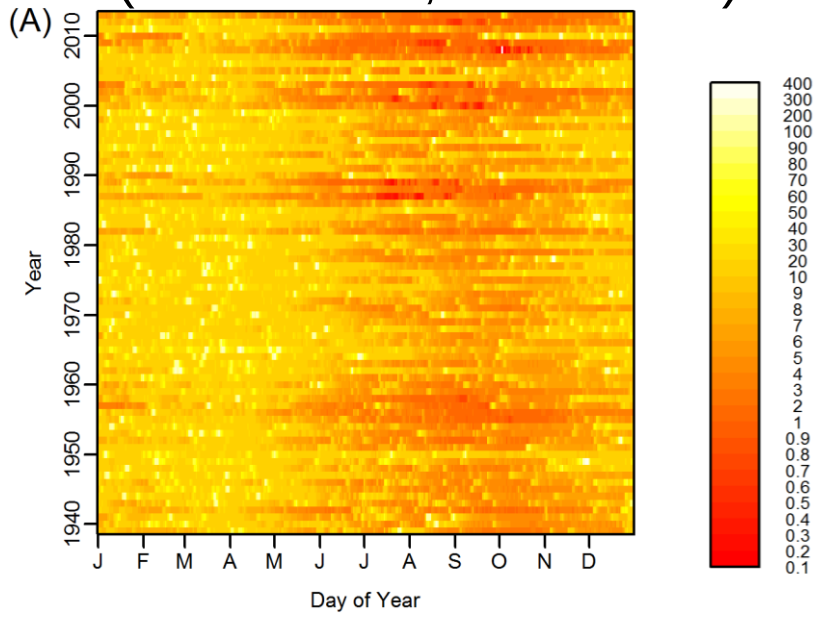
- Multiple methods can (and should) be applied simultaneously:
  - Explore a data set
  - Emphasize specific aspects of data
  - Highlight different elements of variability
  - Guide quantitative analyses



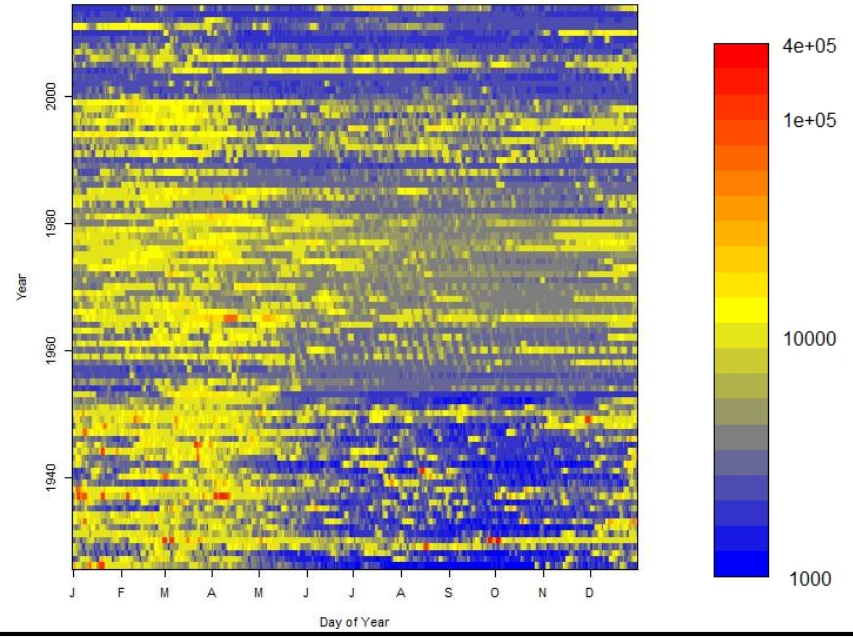
2. What are the relevant scales?

- Spatial and temporal scale vary widely in ecosystems
  - Length and intervals
  - Selection of units should be “natural” to the user
  - Directionality in color use (context-specific)
- Is historical (or spatial) context needed?
- Resolution of data and questions may drive toward a specific method

Academic, Drought Study
(Units = m^3/s, Red = DRY)

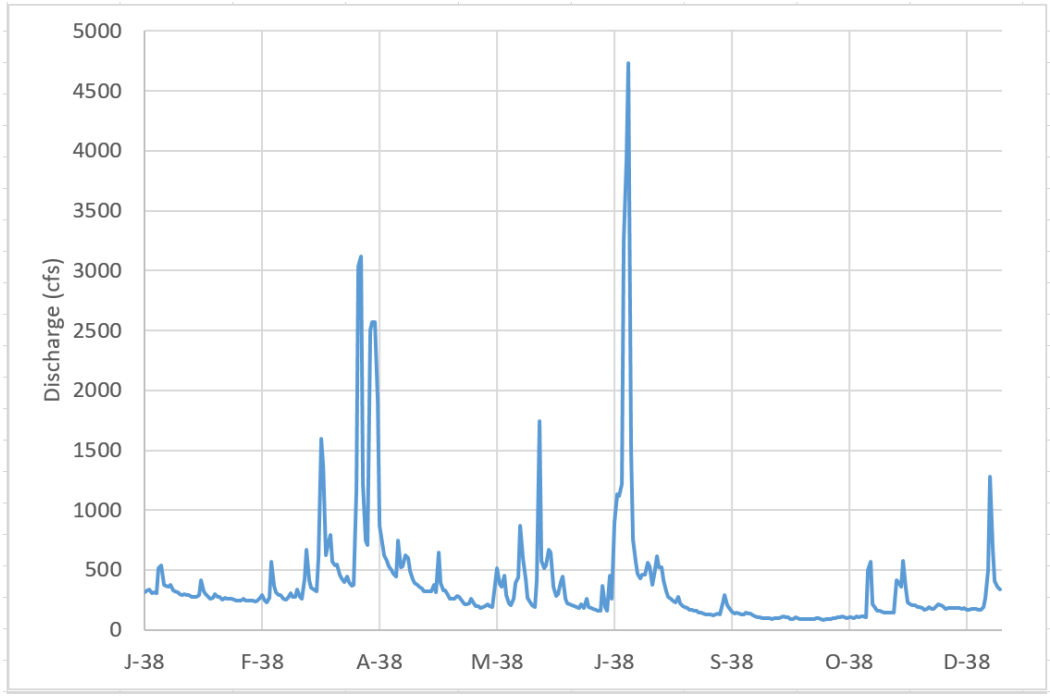


Applied, Flood Study
(Units = cfs, Red = WET)

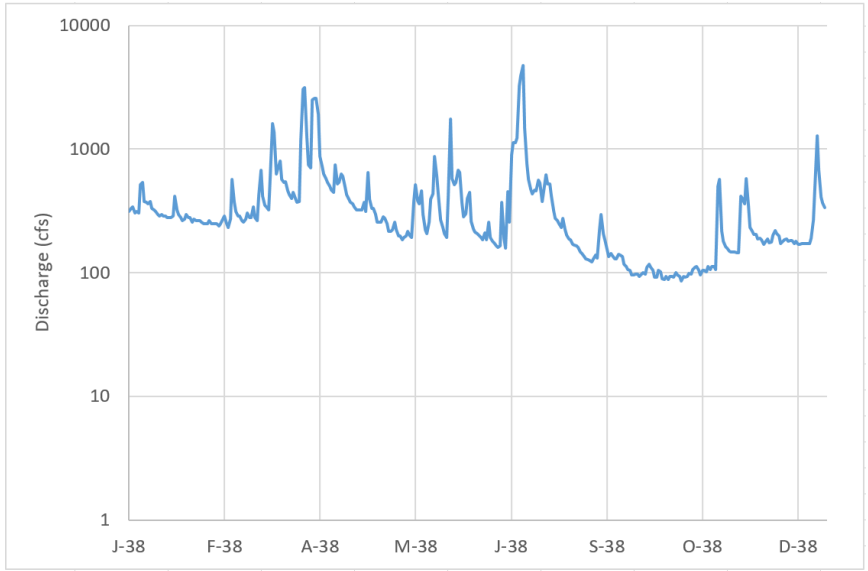


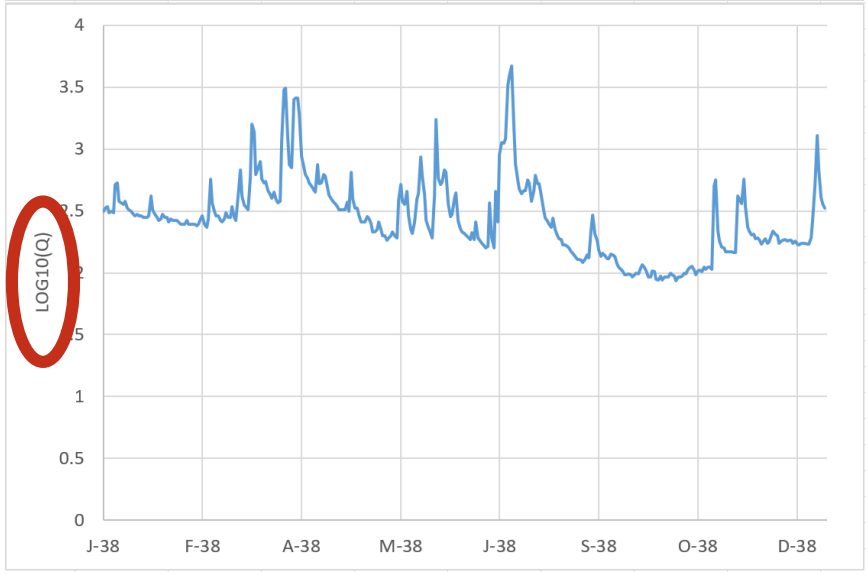
3. How are data distributed?

- Ecological data often occur over many orders of magnitude
  - e.g., changes in streamflow, boom-and-bust of a population
- Rescaling figures can be effective, but be careful not to obscure process (or units)



What is Log10(Q)?





4. What are the constraints of the visualization environment?

- Where are the figures being used?
  - Report vs. Presentation vs. Infographic
  - Screen / page size limits?

- How will users experience the figures?
  - Static, static with zoom, dynamic, or interactive

- Gotchas on medium selection
  - Color, bandwidth, user expectations,…

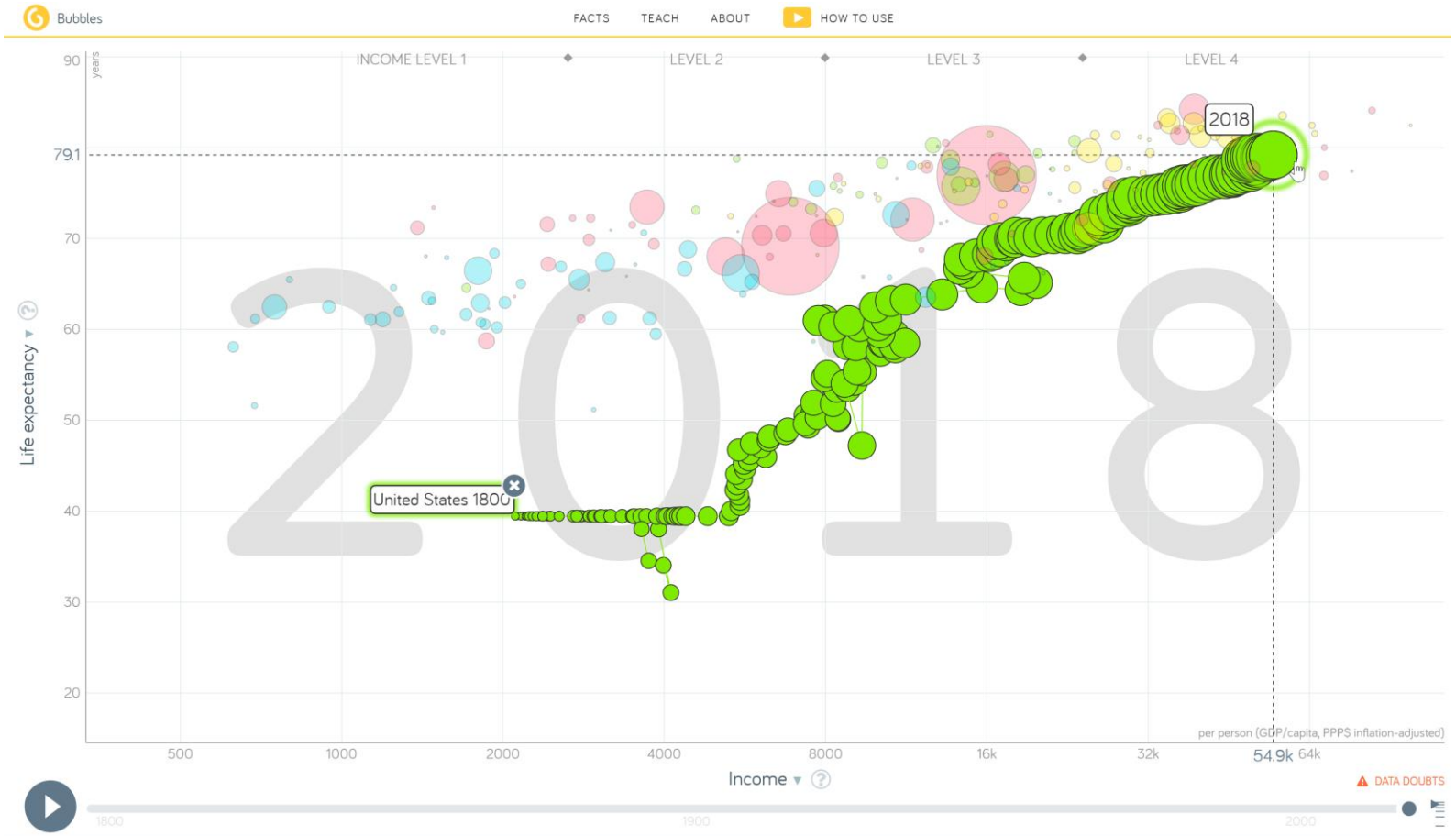
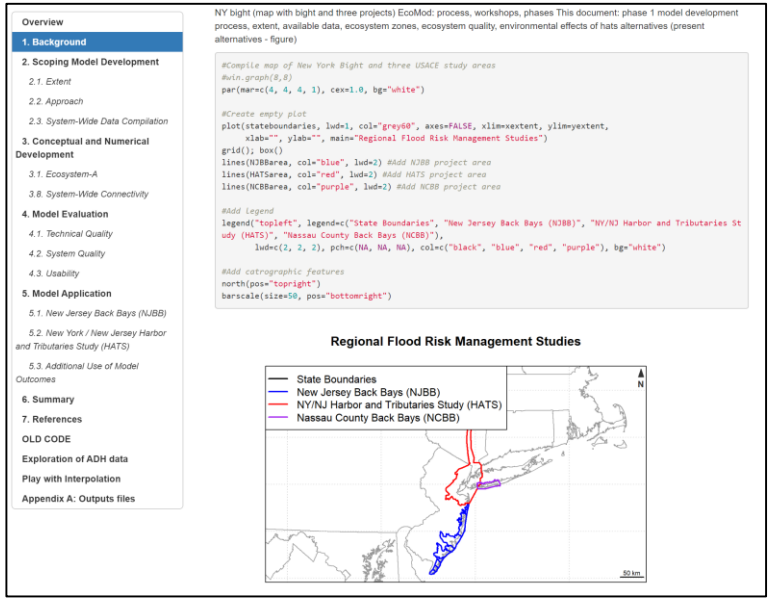
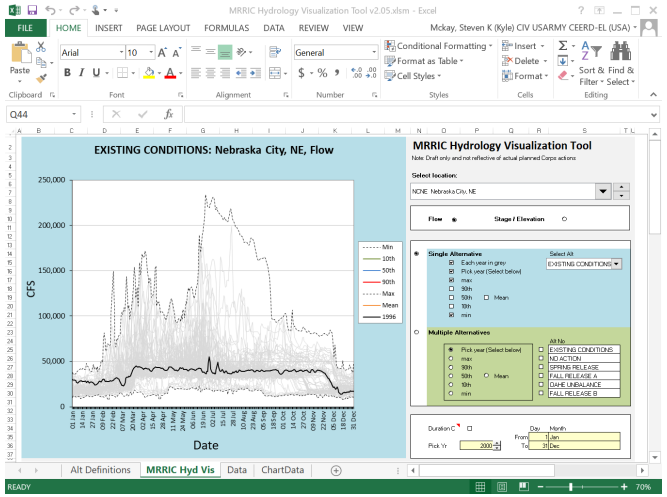


Figure: gapminder.org

5. What tools and expertise are available?

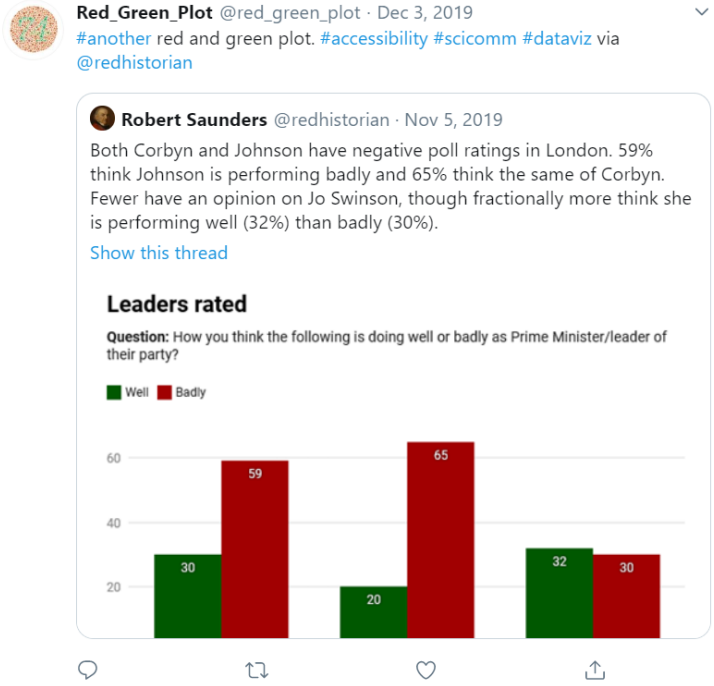
- A large variety of software is available to visualize data
  - e.g., Microsoft Excel, MATLAB, R Statistical Software, GIS
- Not all visualizations are easily conducted in all programs
- Personnel limitations (or assets)
- How many times will figures be generated?
  - Manual tools can require huge amounts of time
  - Move toward automated report generation via “reproducible research” methods (e.g., Rmarkdown)





Side notes and pet peeves

- Why should you care about color-blind friendly schemes?
  - Nationwide ~ 8% of men and 0.5% of women
  - USACE supervisors and Army officers >70% male
    - ▸ 6-7% chance the decision-maker can't distinguish red-green results
    - ▸ In a large vertical team meeting with 30+ attendees, 2 people may have some level of color-blindness!
- Figures require thinking, not just throwing data into Excel haphazardly (at the end of a complex analysis)
  - There are more techniques than line plots
  - Choose the right plot for the job
  - Never use the "smooth" line function
- Data sources: Can someone recreate your figure?
- Lying with data. Don't do it.
  - Don't bin the data in nefarious ways (e.g., 1-50%).
  - Don't truncate the scale to hide data points.
  - Label transformed axes clearly.
  - …



Fun web resources

- Time series visualization methods

http://survey.timeviz.net/

- A periodic table of visualization

www.visual-literacy.org/periodic_table/periodic_table.html

- Information is beautiful

http://www.informationisbeautiful.net/

- 35 incredible dataviz tools

http://www.creativebloq.com/design-tools/data-visualization-712402

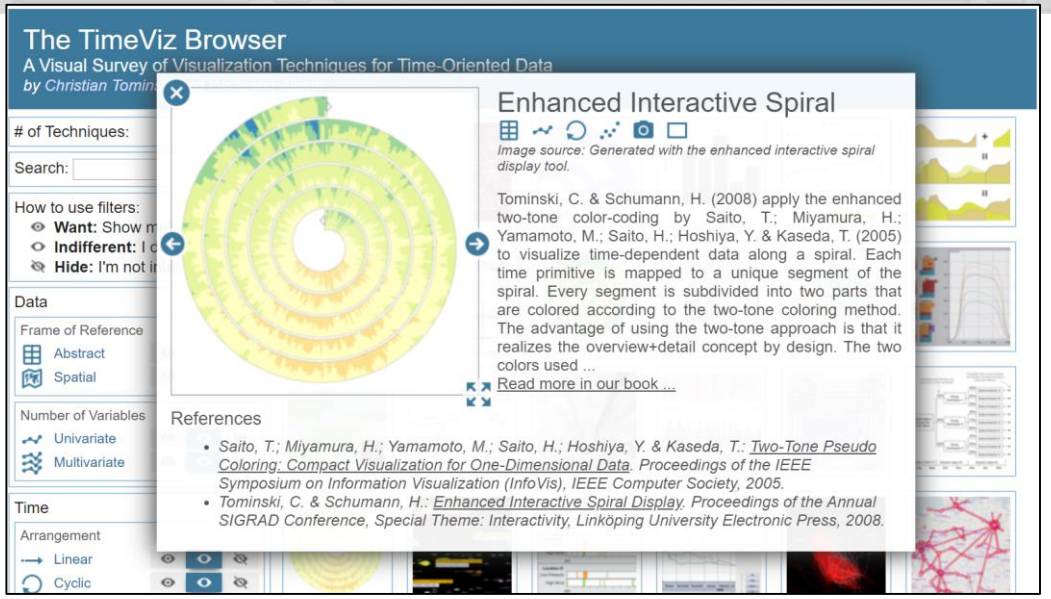
- Colorado State's USGS streamflow tools

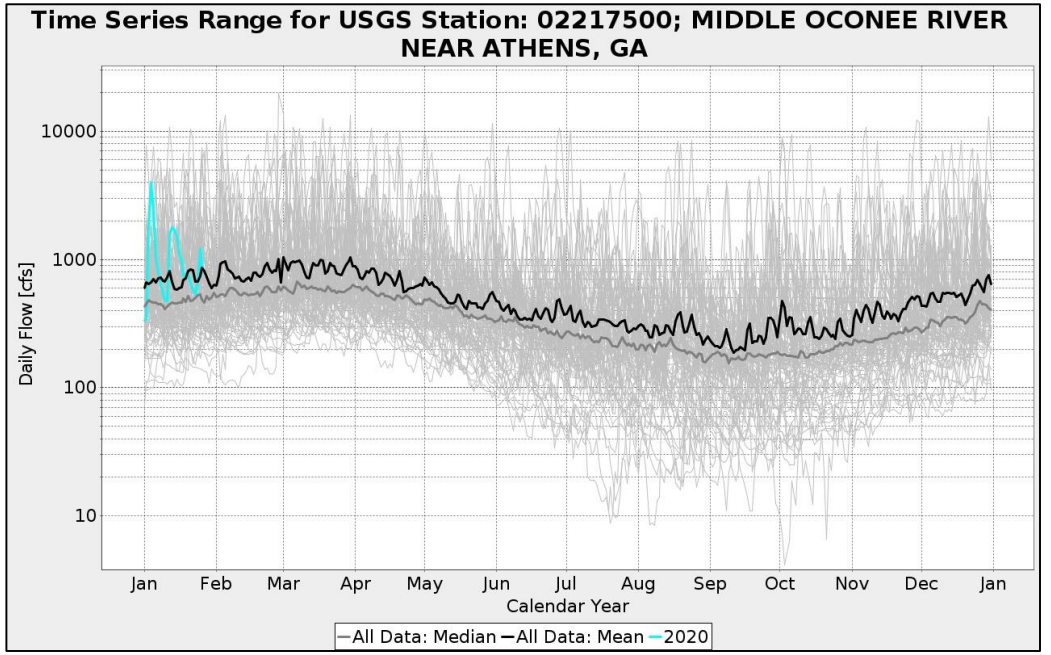
http://www.erams.com/flowanalysis/

- Gapminder

https://www.gapminder.org/

(and any Hans Rosling TED talk)





Thank you for your time!

Take-home messages

- The examples shown are flawed
- Don't let a 21st century analysis be poorly represented by a 18th century visualization
- Visualization can provide a powerful mechanism for communication and learning
- Many methods exist in readily-available software
- Visualization cannot substitute for rigorous quantitative analyses, but it can inform the analyst and guide the analyses

- McKay S.K. 2019. Visualization as a tool for ecological analysis. *Encyclopedia of Ecology*, 2nd edition, Vol. 2, pp. 213-220. Eds. S. Jorgensen and B. Fath. doi: 10.1016/B978-0-12-409548-9.10566-4.

- Many thanks to Todd Swannack and EMRRP!

- Find me here:
  - S. Kyle McKay, Ph.D., P.E.
  - Kyle.McKay@usace.army.mil
  - 917-790-8717