Processing Benthic Macroinvertebrate Data in R

Emma Jones, Lucy Baker, and

Jason Hill

VIRGINIA DEPARTMENT OF
ENVIRONMENTAL QUALITY

Expectation Setting

What this session is:

- Inspiration for using R for reproducible workflows
- Discussion on automating reports
  - General how to, when, why
- **Rmarkdown primer**
- Showcase some Virginia DEQ tools capitalizing on reproducible reports

What this session is not:

- Introduction to R
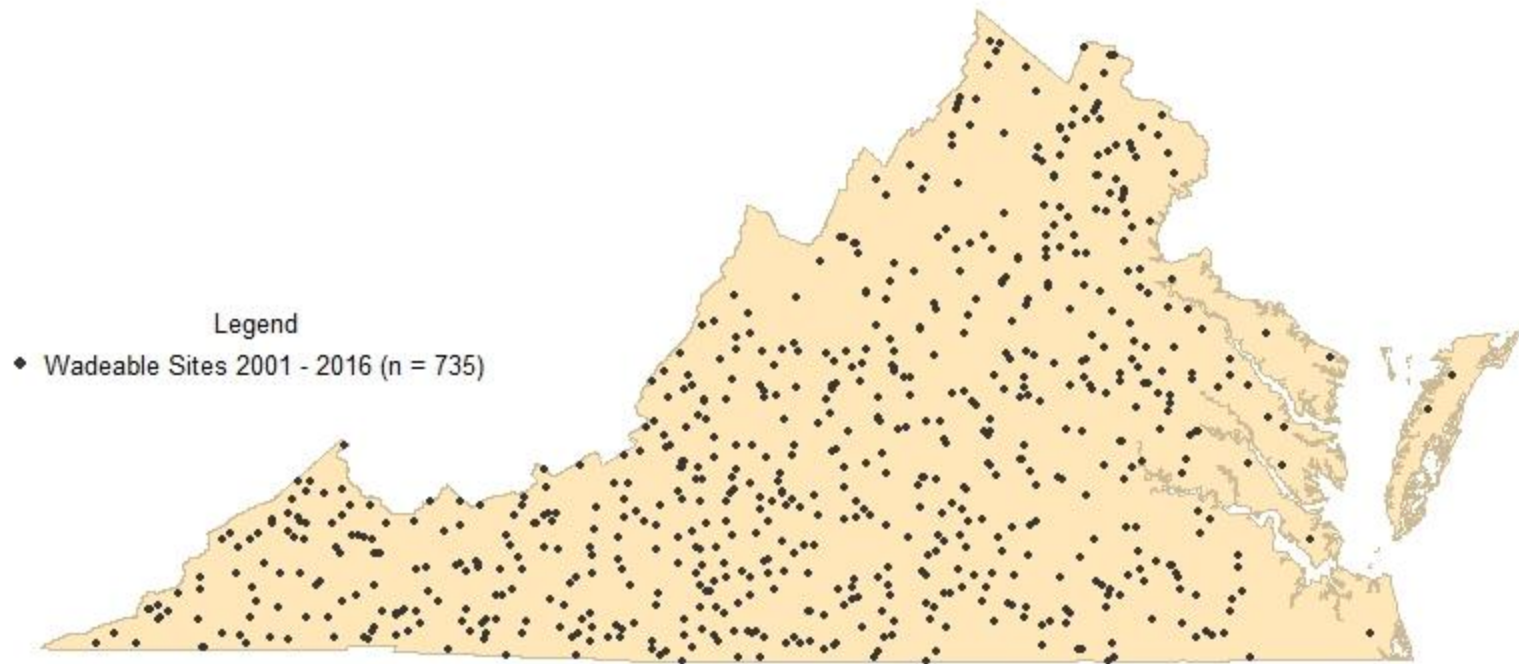- Programming 101
- Shiny tutorial
- Troubleshooting sesh

Part 1: Benthic Stressor Analysis Tool

Freshwater Probabilistic Monitoring in Virginia

Legend
♦ Wadeable Sites 2001 - 2016 (n = 735)

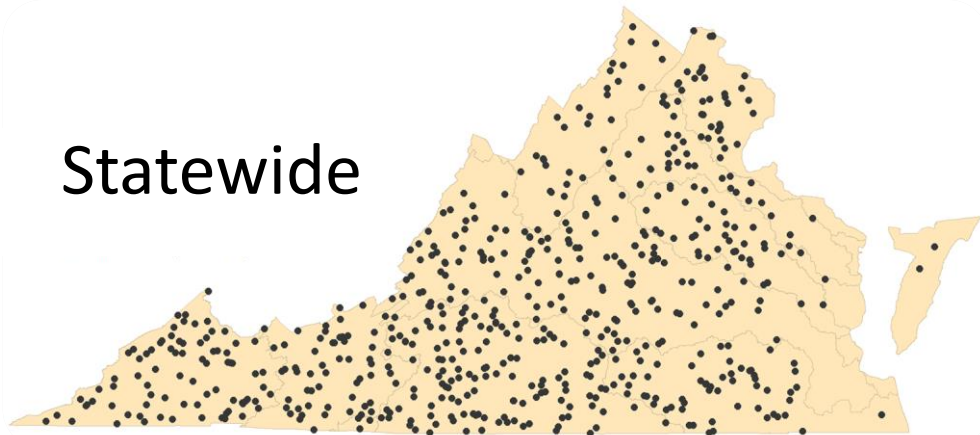Monitoring statewide 2001 - present
~ 60 ProbMon Sites / Year

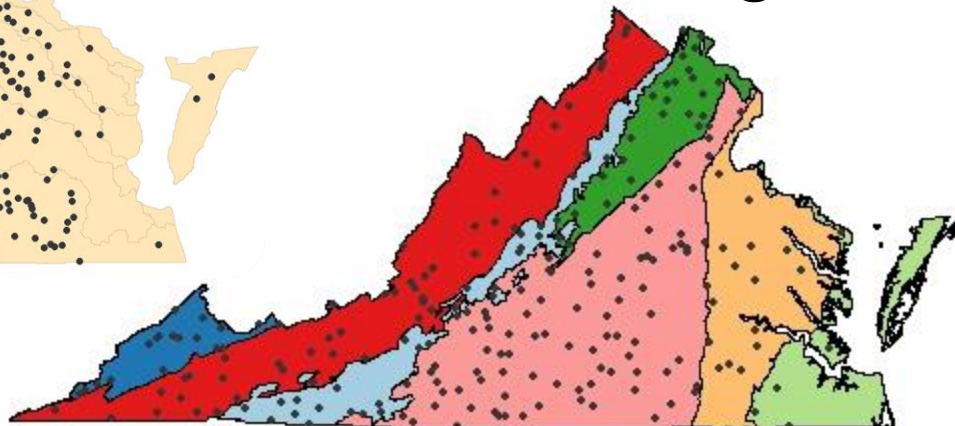735 paired benthic and water chemistry/habitat data points
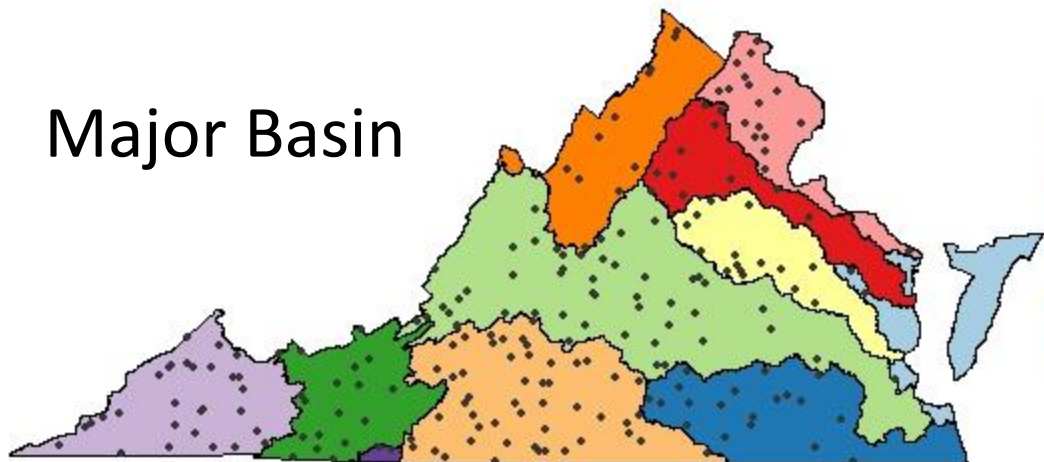(2001 – 2016)

Analysis Scale

Statewide

US EPA Level III Ecoregion

Major Basin

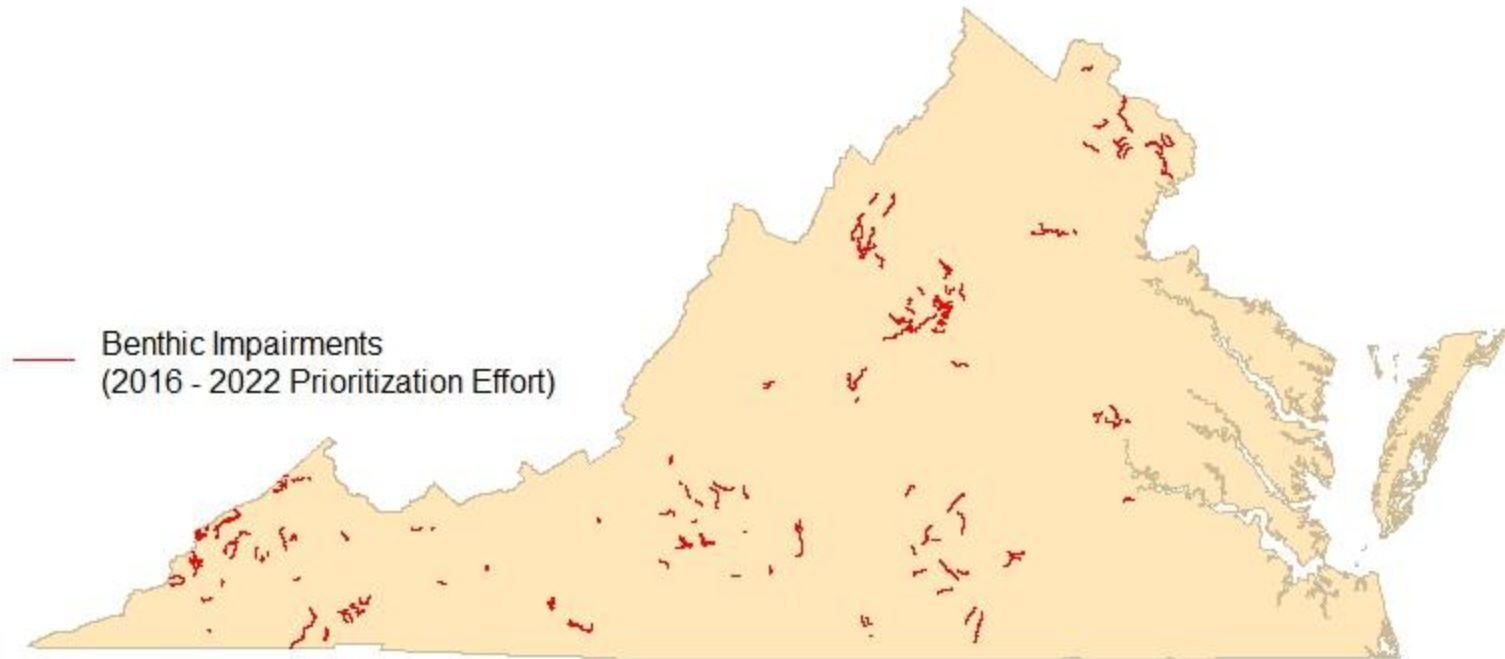Stream Order

Stressor Analysis in Virginia

Benthic Impairments
(2016 - 2022 Prioritization Effort)
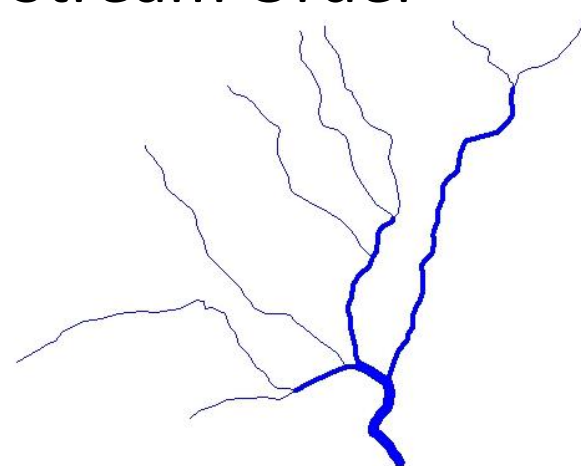
- Virginia's Prioritization effort includes 204 benthic macroinvertebrate community impaired segments (assessment units with benthic cause)
  - Identified as either "TMDL" or "TMDL alternative"
  - **Commitment to EPA for completion: 2016-2022**
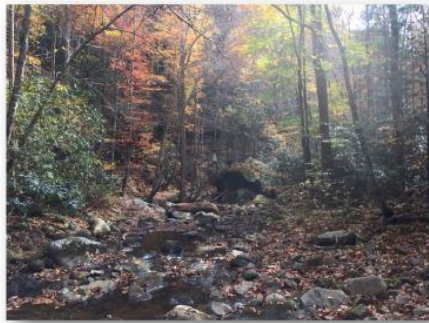- Stressor analyses need to be developed internally or by a contractor

Stressor Analysis in Virginia
(continued)

Stressor Analysis in Virginia:

Data Collection and Stressor Thresholds

Water Quality Monitoring, Biological Monitoring and
Water Quality Assessment Programs

Department of Environmental Quality

Richmond, Virginia

March 2017

VDEQ Technical Bulletin WQA/2017-001
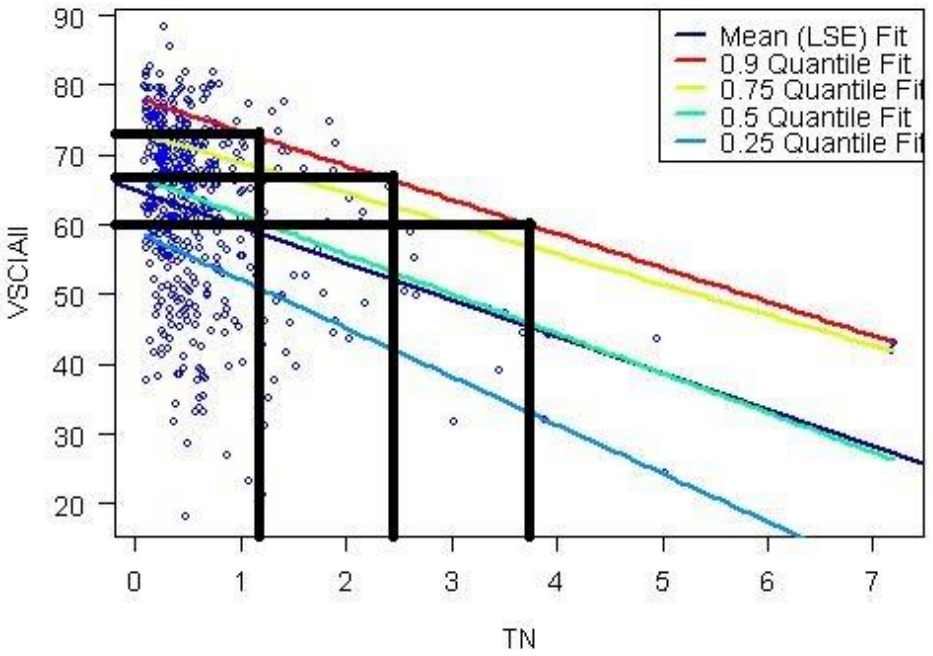
- Identifies the cause of the benthic macroinvertebrate community shift
- Weight-of-evidence approach
- Relies on all available data
- Parameters classified as…
  – Non-stressor
  – Possible stressor
  – Most probable stressor
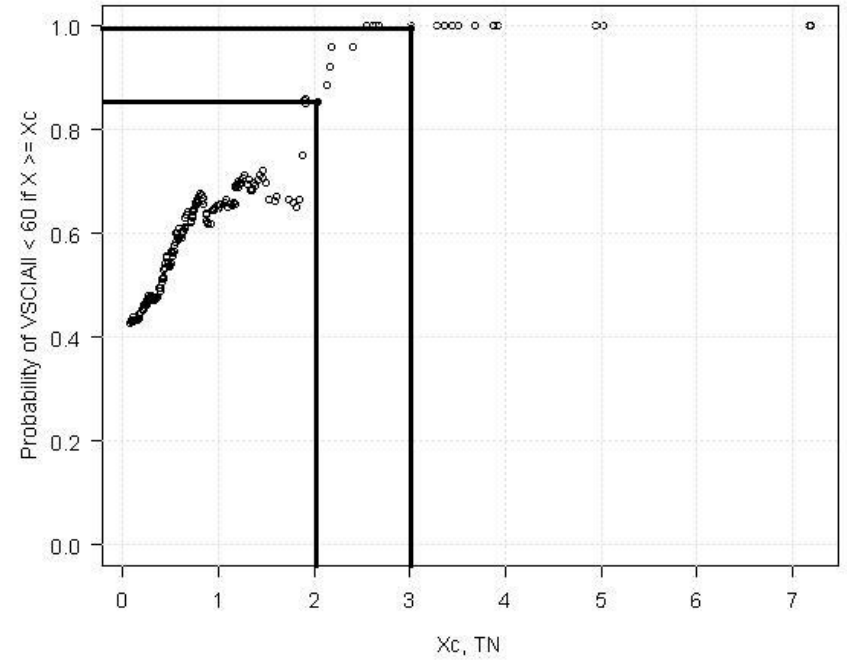- Multiple stressors may be identified

Developing Stressor Thresholds: Statistical Approach

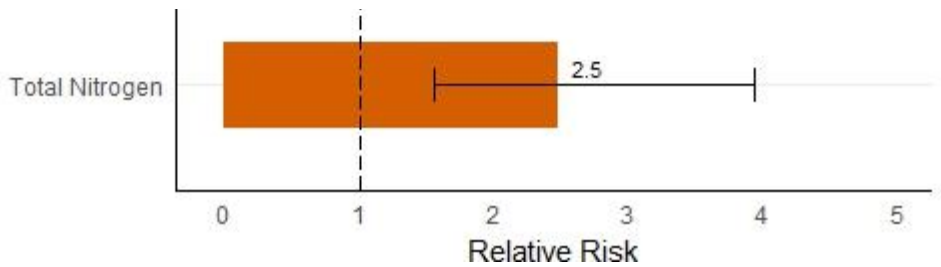**Quantile Regression**

Conditional Probability

Relative Risk

Developing Stressor Thresholds

- Probabilistic Data used to define parameter thresholds:

Probability of Stress to Aquatic Life
High
Medium
Low
None

Stressor Parameters

- Dissolved Oxygen
- pH
- Total Phosphorus
- Total Nitrogen
- Total Habitat
- Ionic Strength
  - Dissolved Sulfates
  - Dissolved Chloride
  - Dissolved Potassium
  - Dissolved Sodium
  - Specific Conductance / Total Dissolved Solids
- Relative Bed Stability (Quantitative Habitat analysis)
- Dissolved Metals (Cumulative Criterion Unit)

Dissolved Oxygen	
Probability of Stress to Aquatic Life	Concentration (mg/L)
High	< 7
Medium	> 7, < 8
Low	> 8, < 10
None	> 10

Context is Everything
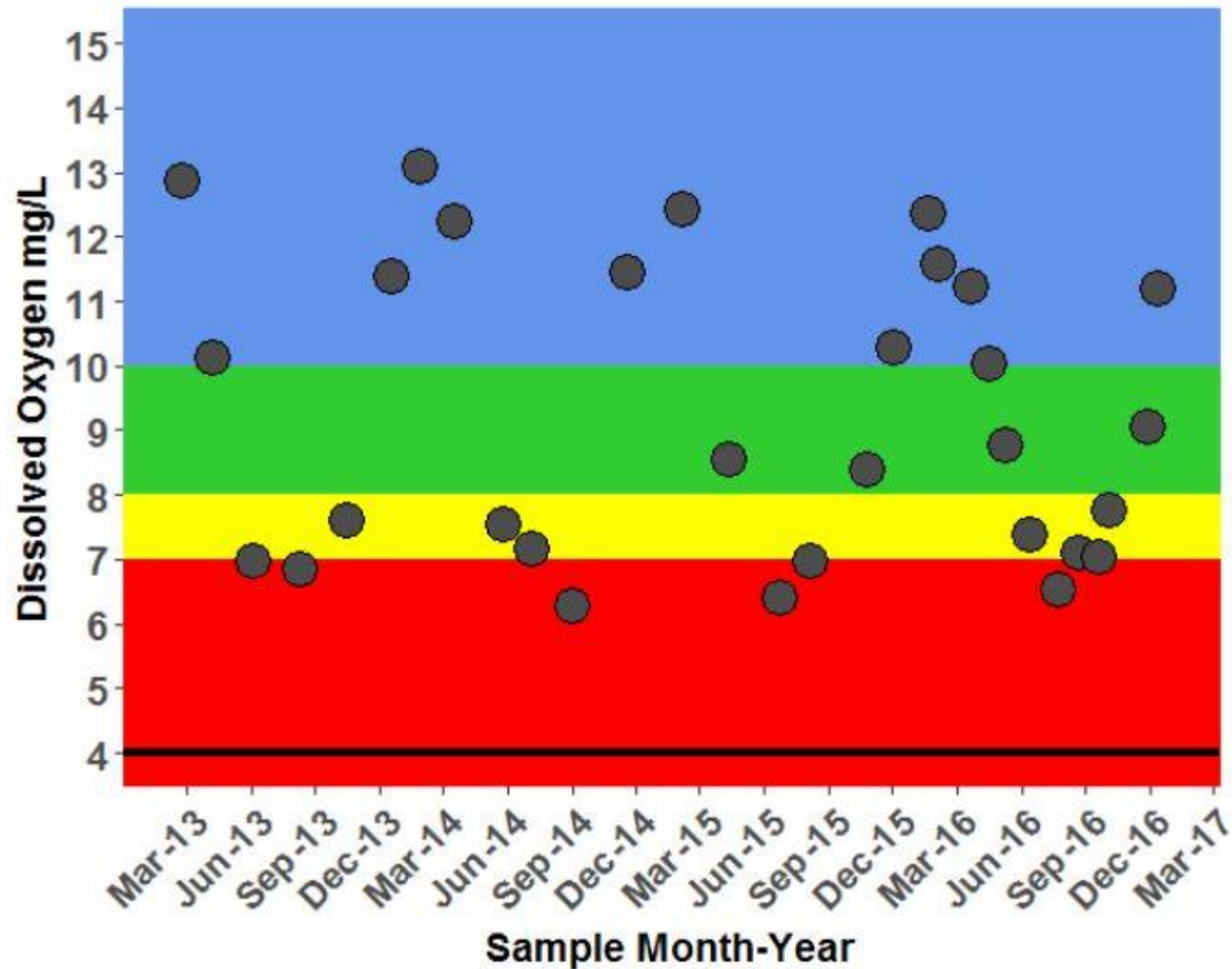
Composite Table | pH Summary | DO Summary | TN Summary | More ▾

Composite Table

You can export the table below as a .csv, .xlsx, or .pdf by clicking the corresponding button below. The Copy button copies all table data for you to put into any spreadsheet program. If you want the color background formatting, you need to manually select the table with your cursor to copy all associated formatting to a spreadsheet program.

[Copy] [CSV] [Excel] [PDF]

Statistic	pH	DO	TN	TP	TotalHabitat	LRBS	MetalsCCU	SpCond	TDS	DSulfate	DChloride	DPotassium	DSodium
Average	8.261	9.709	0.8771	0.01786	116.9	0.1875	0.1872	492.5	315	31.7	19.4	1.305	9.63
Median	8.05	9.6	0.865	0.02	120	0.1875	0.1872	520	315	31.7	19.4	1.305	9.63

Risk Category

Risk Category
High Probability of Stress to Aquatic Life
Medium Probability of Stress to Aquatic Life
Low Probability of Stress to Aquatic Life
No Probability of Stress to Aquatic Life

Report Output:

Click below to save a .HTML version of all the tables and graphics associated with the input station. You can save this to a .pdf after initial HTML conversion (File -> Print -> Save as PDF).

[⬇ Generate CDF Report]
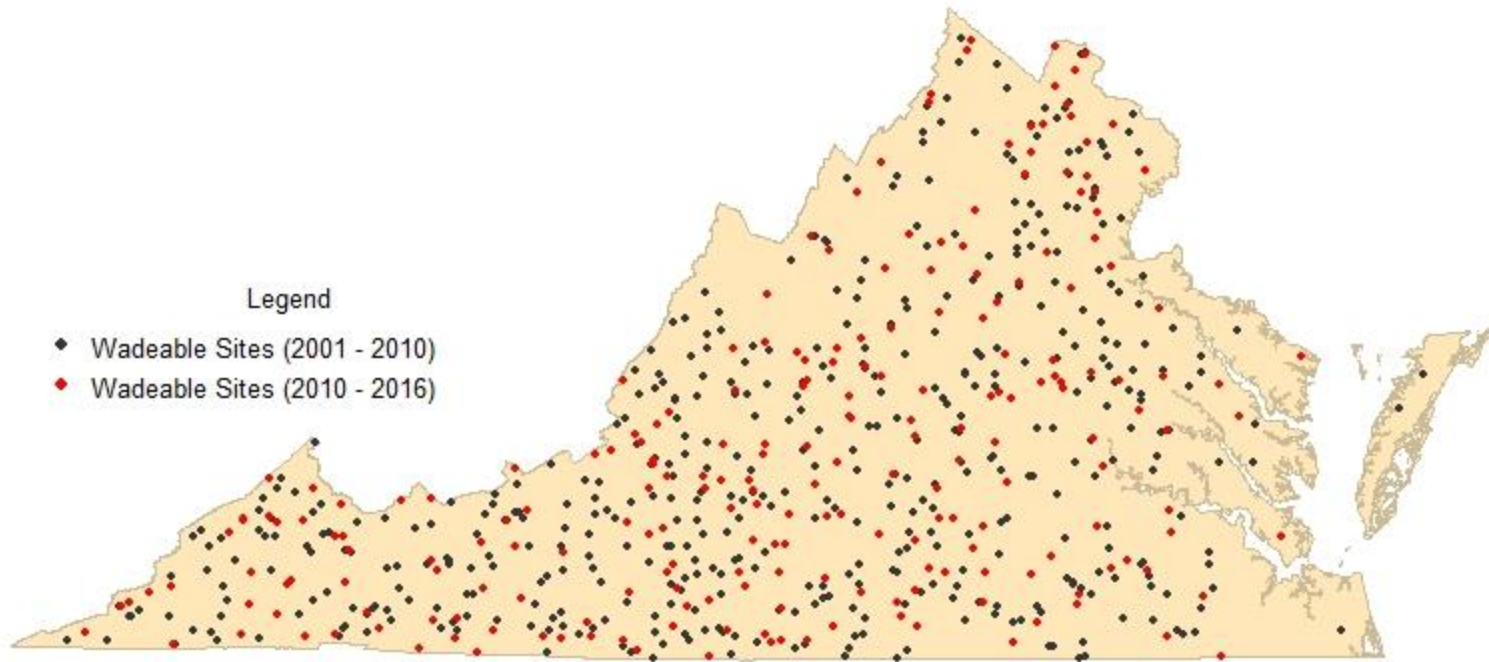
Tool Benefits

Anticipated:

- **Standardize thresholds statewide**
- Increase biological knowledge of TMDL coordinators
- **Simplify analytical updates with increasing n**
- Standardize data collection for follow up monitoring
- Standardize data manipulation/analyses
- Expedite data manipulation/analyses
- **Standardize reporting process and products**
- Expedite reporting process
- **Cost savings**

Unanticipated:

- Inadvertently developed nesting rationale and landowner report tools
- Build culture of reproducible reports/research
- Introduce automation to regular business practices
- Initiate open source culture within VDEQ
- **Gateway app** for the development of additional analytical applications for cross media business needs

Probabilistic Monitoring Sites: 2001-2016 (*n* = 735)

Legend
* Wadeable Sites (2001 - 2010)
* Wadeable Sites (2010 - 2016)

Paired benthic and water chemistry/habitat data points
Published report (2001 – 2010): ***n* = 474**
Interactive Application (2001 – 2016): ***n* = 735**

Tool Benefits

Anticipated:

- **Standardize thresholds statewide**
- Increase biological knowledge of TMDL coordinators
- **Simplify analytical updates with increasing n**
- Standardize data collection for follow up monitoring
- Standardize data manipulation/analyses
- Expedite data manipulation/analyses
- **Standardize reporting process and products**
- Expedite reporting process
- **Cost savings**

Unanticipated:

- **Inadvertently developed nesting rationale and landowner report tools**
- Build culture of reproducible reports/research
- Introduce automation to regular business practices
- Initiate open source culture within VDEQ
- **Gateway app** for the development of additional analytical applications for cross media business needs

Tool Uses (so far)

Benthic Stressor Reports for:
- Catawba Creek (preliminary EPA approval)
- Unnamed Tributary to Roanoke River
- Mountain Run
- Crane Creek
- Briery Creek
- Smith River
- Naked Creek
- Lynch Creek
- Reed Creek
- Allens Branch
- Devil Fork
- Bark Camp Branch

- Monitoring/Landowner Reports
- Assessment/Monitoring planning
- Dissolved metals assessment
- BCG/Tolerance document
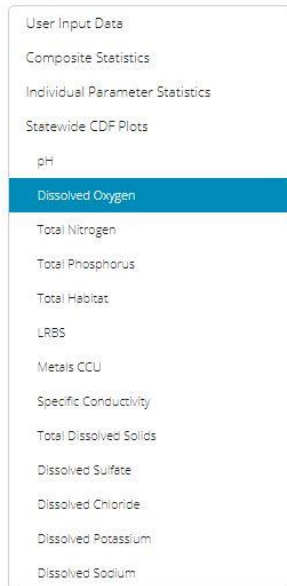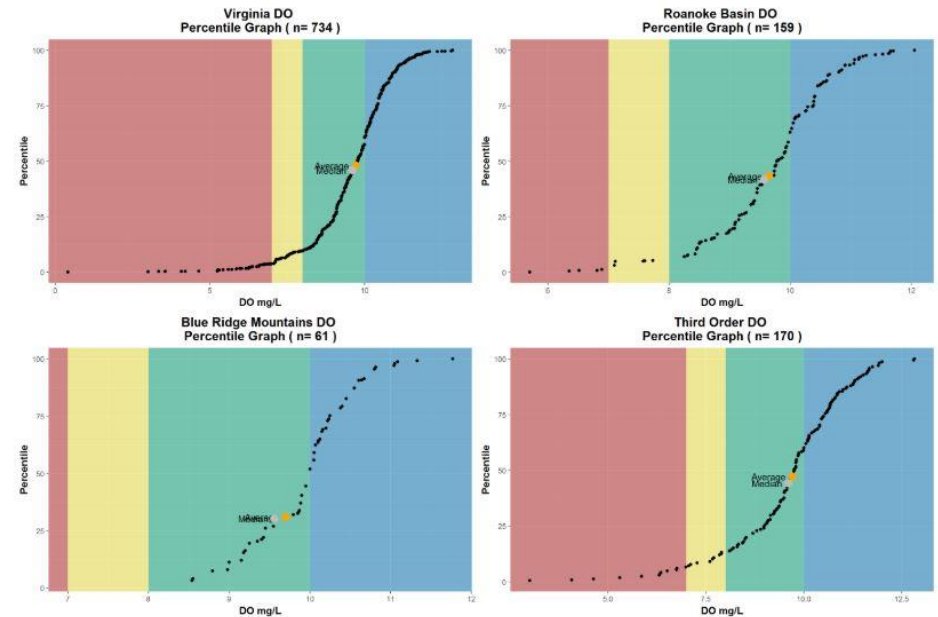
Stressor Tool Demo

All code and datasets necessary to run tool:

www.github.com/VDEQ/VDEQ_Benthic StressorAnalysis

Report Available here: www.deq.virginia.gov/Programs /Water/WaterQuality InformationTMDLs/ WaterQualityMonitoring/ ProbabilisticMonitoring.aspx

Part 2: Introduction to Rmarkdown

Hands on component if you want to follow along

http://github.com/EmmaVJones/ NCTC_Processing-Benthic- Macroinvertebrate-Data-in-R

What is Rmarkdown?

```
60 ▾  ## Site Overview
61
62   This is the section where you could put some basic description information about
     the site. You can also start to use the inline text feature and have a generic
     sentence that is automatically updated each time you enter a new dataset.
63
64   For instance, `r prettyStreamName` was sampled `r nrow(bugData)` times for benthic
     macroinvertebrates and `r nrow(envData)` times for ambient parameters between `r
     format(min(envData$date), format="%B %Y")` and `r format(max(envData$date),
     format="%B %Y")`.
65
66   Alternatively, you could turn that information into a table to save space. Here
     are a few quick stats from the datasets.
67
68 ▾  ```{r Quick Site overview, echo=FALSE}                              ⚙ ≡ ▸
69
70   # Organize bug data
71   bugTable <- select(bugData, StationID, StreamName, Location, FamSCI) %>% # keep
     only parameters of interest
72     mutate(`Median VSCI` = format(median(FamSCI), digits = 3), # calculate median
73             `Number of Samples` = n()) %>% # and the number of samples
74     select(-c(FamSCI)) %>% # remove individual scores for overview
75     distinct(`Number of Samples`, .keep_all = TRUE) # keep only 1 row
76   |
77   kable(t(siteTable), format = "html")
78
79   ```
80
```

Chunk options

There are a variety of options to affect how the code chunks are treated.

- Use `echo = FALSE` to avoid having the code itself shown.
- Use `results = "hide"` to avoid having any results printed.
- Use `eval = FALSE` to have the code shown but not evaluated.
- Use `warning = FALSE` and `message = FALSE` to hide any warnings or messages produced.
- Use `fig.height` and `fig.width` to control the size of the figures produced (in inches).

https://datacarpentry.org/genomics-r-intro/XX-knitr-markdown/index.html

syntax

Plain text
End a line with two spaces to start a new paragraph.
italics and _italics_
bold and __bold__
superscript^2^
~~strikethrough~~
[link](www.rstudio.com)

Header 1

Header 2

Header 3

Header 4

Header 5

Header 6

ellipsis: ...
inline equation: $A = \pi*r^{2}$
image:

horizontal rule (or slide break):

***

> block quote

* unordered list
* item 2
    + sub-item 1
    + sub-item 2

1. ordered list
2. item 2
    + sub-item 1
    + sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

becomes

Plain text
End a line with two spaces to start a new pa
italics and *italics*
bold and **bold**
superscript^2
~~strikethrough~~
link

Header 1

Header 2

Header 3

Header 4

Header 5

ellipsis: ...
inline equation: $A = \pi * r^2$

image:

horizontal rule (or slide break):

> block quote

* unordered list
* item 2
    o sub-item 1
    o sub-item 2

1. ordered list
2. item 2
    o sub-item 1
    o sub-item 2

Table Header	Second Header
Table Cell	Cell 2
Cell 3	Cell 4

https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf

Use case: Landowner Reports for Benthic and Field Data

Resources

RStudio RMarkdown Introduction

https://rmarkdown.rstudio.com/lesson-1.html

RStudio Cheat Sheets

https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf

Data Carpentry

https://swcarpentry.github.io/r-novice-gapminder/15-knitr-markdown/

Part 3: Discussion

Contact Information

Emma Jones   emma.jones@deq.virginia.gov

All code and datasets utilized in this workshop available here:

https://github.com/EmmaVJones/ NCTC_Processing-Benthic-Macroinvertebrate-Data-in-R