Processing Benthic Macroinvertebrate Data in R

Emma Jones, Lucy Baker, and Jason Hill
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

Monitoring statewide 2001 - present
~ 60 ProbMon Sites / Year

735 paired benthic and water chemistry/habitat data points (2001 – 2016)
Stressor Analysis in Virginia

- 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)

- 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:

<table>
<thead>
<tr>
<th>Probability of Stress to Aquatic Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
</tr>
<tr>
<td>Medium</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>None</td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Dissolved Oxygen</th>
<th>Probability of Stress to Aquatic Life</th>
<th>Concentration (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>&lt; 7</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>&gt; 7, &lt; 8</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>&gt; 8, &lt; 10</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>&gt; 10</td>
</tr>
</tbody>
</table>
Context is Everything
### 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.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>pH</th>
<th>DO</th>
<th>TN</th>
<th>TP</th>
<th>TotalHabitat</th>
<th>LRBS</th>
<th>MetalsCCU</th>
<th>SpCond</th>
<th>TDS</th>
<th>DSulfate</th>
<th>DChloride</th>
<th>DPotassium</th>
<th>DSodium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>8.261</td>
<td>9.709</td>
<td>0.677</td>
<td>0.01786</td>
<td>115.9</td>
<td>0.1875</td>
<td>0.1872</td>
<td>492.5</td>
<td>315</td>
<td>31.7</td>
<td>19.4</td>
<td>1.305</td>
<td>9.63</td>
</tr>
<tr>
<td>Median</td>
<td>8.05</td>
<td>9.6</td>
<td>0.665</td>
<td>0.02</td>
<td>120</td>
<td>0.1875</td>
<td>0.1872</td>
<td>520</td>
<td>315</td>
<td>31.7</td>
<td>19.4</td>
<td>1.305</td>
<td>9.63</td>
</tr>
</tbody>
</table>

### 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 ($n = 735$)

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_BenthicStressorAnalysis

Report Available here:
www.deq.virginia.gov/Programs/Water/WaterQualityInformationTMDLs/WaterQualityMonitoring/ProbabilisticMonitoring.aspx
Part 2: Introduction to Rmarkdown

Hands on component if you want to follow along

What is Rmarkdown?

```
## Site Overview

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.

For instance, `prettyStreamName` was sampled `nrow(bugData)` times for benthic macroinvertebrates and `nrow(envData)` times for ambient parameters between `format(min(envData$date), format= "%B %Y")` and `format(max(envData$date), format="%B %Y")`.

Alternatively, you could turn that information into a table to save space. Here are a few quick stats from the datasets.

```
```r
# Organize bug data
bugTable <- select(bugData, StationID, StreamName, Location, FamSCI) %>% # keep only parameters of interest
  mutate(`Median VSCI` = format(median(FamSCI), digits = 3), # calculate median
    `Number of Samples` = n() %>% # and the number of samples
    select(-c(FamSCI)) %>% # remove individual scores for overview
    distinct(`Number of Samples`, .keep_all = TRUE) # keep only 1 row
kable(t(siteTable), format = "html")
```
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).

Use case: Landowner Reports for Benthic and Field Data
Resources

RStudio RMarkdown Introduction
https://rmarkdown.rstudio.com/lesson-1.html

RStudio Cheat Sheets

Data Carpentry
Part 3: Discussion
Contact Information

Emma Jones  emma.jones@deq.virginia.gov

All code and datasets utilized in this workshop available here: