Tools for effective science communication

Interpreting your data

Dylan Taillie & Caroline Donovan

CMC Introduction to Data Interpretation Workshop
January 30, 2018
Interpretation

- Evaluating and analyzing your data in order to communicate it in a meaningful way with your selected audience
Kinds of data

- Nominal
  - Non-numerical
  - Qualitative

Types of aquatic plants
- Emergent
- Submergent
- Floating leaf
- Free floating
Kinds of data

• Ordinal
  -- Numerical
  -- Quantitative
Kinds of data

• Interval
  – Basic WQ data
  – Distance between numbers

• Ratio
  – Similar to interval
  – Absolute zero
Precision vs Accuracy

- Accuracy is how close a measurement is to a real value
- Precision is when repeated measurements closely match each other
Precision vs Accuracy

• Accuracy is how close a measurement is to a real value
• Precision is when repeated measurements closely match each other

<table>
<thead>
<tr>
<th>WQ Parameter</th>
<th>Equipment</th>
<th>Precision</th>
<th>Accuracy</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissolved oxygen</td>
<td>LaMotte 5860</td>
<td>0.2 mg/L</td>
<td></td>
<td>0 – 10+ mg/L</td>
</tr>
<tr>
<td>Dissolved oxygen</td>
<td>Ex. LaMotte 1761</td>
<td>0.01 mg/L</td>
<td>± 2% FS</td>
<td>0 – 20 mg/L</td>
</tr>
<tr>
<td>Nitrate-nitrogen</td>
<td>Hach NI-14 1416100</td>
<td>0.01 mg/L (0 – 1 mg/L); 0.1 mg/L (1 – 10 mg/L)</td>
<td></td>
<td>0 – 1 mg/L; 1 – 10 mg/L</td>
</tr>
<tr>
<td>Nitrate-nitrogen</td>
<td>LaMotte 3110</td>
<td>0.25, 0.5, 1, 2, 4, 6, 8, 10 mg/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrate-nitrogen</td>
<td>LaMotte 3354</td>
<td>0, 1, 2, 4, 6, 8, 10, 15 mg/L</td>
<td></td>
<td>0 – 10 mg/L</td>
</tr>
</tbody>
</table>
Activity - Cleaning your data in excel

• Cleaning is required before you interpret your data
• What does cleaning my data mean?
  – Formatting your spreadsheet so it’s consistent
  – Flagging unusual or duplicate data
Excel activity – step 1

• Are there headers with units associated with the data values?
• Do the rows and columns have the right widths and heights to view all the data or is that even needed?
• Are there any missing data that are in another spreadsheet or on another tab that need to be incorporated here?
Why are there two blank lines?

Should this be Conductivity?

Should you add more tabs so that the data is organized by indicator? Or by surface vs bottom measurements? Or something else?

How are you going to address data values that are not numbers?
Excel activity – step 2

• How are the data organized? By date or sampling station? Which way is the best way to look at and interpret the data?
• Are there any duplicate entries? Why? Do you delete them altogether or save them “just in case”? How do you organize and structure your files to do this?
• Are there any unusual data? You can sort the data from high to low and determine if any values are outside the expected range. This could be due to typing errors, instrument error, or they could be genuine outliers
The column headers are bolded and each label makes sense and has appropriate units.

Extra decimal point:

Is this the correct sampling date? Check original fieldsheets or verify with field staff/volunteer monitor.

This row is duplicated. Do you delete it or keep it and flag it?

<table>
<thead>
<tr>
<th>Site</th>
<th>Date</th>
<th>Temperature (°C)</th>
<th>Depth (ft)</th>
<th>Salinity</th>
<th>O₂ (mg/L)</th>
<th>O₂ Saturation</th>
<th>CondBustivity</th>
<th>Light Trans</th>
<th>Sigma-T</th>
<th>Fluorometer</th>
<th>pH400-700nm light</th>
<th>PAR</th>
<th>Oxidation Reduction Poten</th>
<th>Lat/Lng</th>
</tr>
</thead>
<tbody>
<tr>
<td>E6</td>
<td>3/1/2011</td>
<td>0.0</td>
<td>11</td>
<td>32.0</td>
<td>74.7</td>
<td>62.4</td>
<td>3.7</td>
<td>1.2</td>
<td>2.1</td>
<td>1.7</td>
<td>6.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td>3/1/2011</td>
<td>0.0</td>
<td>32.0</td>
<td>11.4</td>
<td>4.65</td>
<td>3.7</td>
<td>1.2</td>
<td>1.2</td>
<td>2.1</td>
<td>1.7</td>
<td>6.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

University of Maryland
CENTER FOR ENVIRONMENTAL SCIENCE INTEGRATION AND APPLICATION NETWORK
Excel activity – step 3

• Are there any cells that need to be changed from numbers to text or vice versa so that Excel can read them correctly?

• How are your latitude and longitude written? Is it in a format that works for you or for someone who will be doing GIS mapping?
This spreadsheet has been cleaned and simplified. Only columns pertaining to dissolved oxygen were kept. The tab is labelled appropriately. The columns have been sorted first by date, then by station.
Using statistics to describe your data

• What are descriptive statistics?
  – Tools to provide basic summarized information about your data
  – Mean, median and mode

<table>
<thead>
<tr>
<th>Sample site</th>
<th>Caddisfly (count)</th>
<th>Mayfly (count)</th>
<th>Stonefly (count)</th>
<th>Clams (count)</th>
<th>Aquatic worms (count)</th>
<th>Crayfish (count)</th>
<th>% Sensitive to Pollution</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>2</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>GP1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>16</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>GP2</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>GP3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td>D1</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>49</td>
</tr>
<tr>
<td>B11</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>mean</td>
<td>2.4</td>
<td>3.3</td>
<td>2.7</td>
<td>10.3</td>
<td>9.0</td>
<td>7.4</td>
<td>22.4%</td>
</tr>
<tr>
<td>median</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>22%</td>
</tr>
<tr>
<td>mode</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Using statistics to describe your data (cont.)

- **Range**
  - The total spread of all values in a dataset

![Median and Quartiles diagram]

- First quartile (Q1)
- Lower quartile
- Median (Q2)
- Upper quartile
- Third quartile (Q3)

**Interquartile range (IQR):**

\[ Q1 - Q3 \]
Using statistics to describe your data (cont.)

- Outliers
  - Data values that fall outside the general distribution of the data
- Standard deviation
  - Distance from mean
  - Variability
- Standard error
  - Type of SD
  - Depends on sample size
Using statistics to describe your data (cont.)

- Bell curves
  - Normal distribution
  - 95% of data within 2 SD
Using statistics to describe your data (cont.)

- Non-normal distribution
Using statistics to describe your data (cont.)

• Correlation
  – Two variables related
  – Temperature & DO
Displaying data

- Data in tables – why use a table?
- Parts of a table
## Formatting a table for your audience

### Example 3: Basic Table with Science Communication Principles Embedded

<table>
<thead>
<tr>
<th>Sample site</th>
<th>Caddisfly</th>
<th>Mayfly</th>
<th>Stonefly</th>
<th>Clams</th>
<th>Aquatic worms</th>
<th>Crayfish</th>
<th>% Sensitive to Pollution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anacostia 1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>2</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Anacostia 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Gunpowder 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>16</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Gunpowder 2</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>Gunpowder 3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td>Davis Branch 1</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>49</td>
</tr>
<tr>
<td>Big Elk 11</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>
Data in graphs

• Graphing data is the easiest way to visualize your data
• Help you too:
  – See relationships between different measurements in the data
  – Identify outliers
  – Visualize and identify trends
Types of graphs

- Bar graph
- Line graph
- Pie graph
- Comparison bar graph
Choosing a graph and formatting – bar graph

EXAMPLE 2: AFTER FORMATTING, BAR GRAPH

Gracilaria is nitrogen-limited

Response to treatment (% change in productivity)

Chincoteague Bay
Sinepuxent Bay
Newport Bay
Assawoman Bay
Isle of Wight Bay

Control
+ nitrogen
+ phosphorus
+ nitrogen + phosphorus
Choosing a graph and formatting – line graph

EXAMPLE 4: AFTER FORMATTING, LINE GRAPH

Zostera seagrass grows fastest in summer months

Graph showing growth rate (seagrass growth units) from January to December.
Choosing a graph and formatting – pie graph

Land use by area

- Shrubland: 25%
- Woody wetlands: 18%
- Herbaceous wetlands: 15%
- Agriculture: 11%
- Grassland: 10%
- Deciduous forest: 10%
- Evergreen forest: 5%
- Developed: 3%
- Water: 3%
### Submerged aquatic vegetation (SAV) area (hectares) in the York River

<table>
<thead>
<tr>
<th>Sample site and year</th>
<th>0-10%</th>
<th>10-40%</th>
<th>40-70%</th>
<th>70-100%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>York tidal fresh 2005</td>
<td>7</td>
<td>16</td>
<td>21</td>
<td>81</td>
<td>125</td>
</tr>
<tr>
<td>York oligohaline 2005</td>
<td>1</td>
<td>10</td>
<td>8</td>
<td>34</td>
<td>53</td>
</tr>
<tr>
<td>York total 2005</td>
<td>8</td>
<td>50</td>
<td>29</td>
<td>115</td>
<td>178</td>
</tr>
<tr>
<td>York tidal fresh 2010</td>
<td>2</td>
<td>22</td>
<td>13</td>
<td>73</td>
<td>110</td>
</tr>
<tr>
<td>York oligohaline 2010</td>
<td>0</td>
<td>28</td>
<td>32</td>
<td>143</td>
<td>203</td>
</tr>
<tr>
<td>York total 2010</td>
<td>2</td>
<td>50</td>
<td>45</td>
<td>216</td>
<td>313</td>
</tr>
<tr>
<td>York tidal fresh 2015</td>
<td>37</td>
<td>14</td>
<td>16</td>
<td>167</td>
<td>234</td>
</tr>
<tr>
<td>York oligohaline 2015</td>
<td>15</td>
<td>135</td>
<td>44</td>
<td>188</td>
<td>367</td>
</tr>
<tr>
<td>York total 2015</td>
<td>72</td>
<td>149</td>
<td>60</td>
<td>355</td>
<td>601</td>
</tr>
</tbody>
</table>

**SAV area in the York River**

**SAV area by sample site**

**Total SAV area**

**2015 SAV area**
Data in figures

- Help readers visually connect information
- Connect numbers from graphs to general patterns and trends or show information on a geographic scale
Data in figures

• Parts of a figure
  – Maps
  – Graphs
  – Photos
  – Text
  – Caption
  – Title
Overall eutrophic condition of coastal lagoons in the United States, early 2000s

- Low: 1
- Moderate low: 4
- Moderate: 9
- Moderate high: 2
- High: 4

Insufficient data for analysis (8)

- Great South Bay
- Barnegat Bay–Little Egg Harbor Estuary
- New Jersey Inland Bays
- Delaware Inland Bays
- Northern Maryland Coastal Bays
- Southern Maryland Coastal Bays
- Albemarle Sound
- Pamlico Sound
- Bogue Sound
- Indian River
- Biscayne Bay
Figure example

Flow rate & nitrogen concentrations in Bassett Creek

- Orange dots: Stormflow nitrogen sample
- Blue dots: Baseflow nitrogen sample
- Blue line: Daily mean streamflow
- Blue dashed line: Daily mean baseflow

Y-axis: Daily mean streamflow and baseflow (cubic feet per second)
X-axis: Month & Year (2002-2004)
Right Y-axis: Nitrite+nitrate (mg N L\(^{-1}\))
Summary

• Start small, let the data lead you
  – What kind of data are you collecting?
  – What is the best way to organize your data?
  – What is the best graph for your data?
  – Will a figure help explain your data?

• Thanks for joining us and let Caroline and I know if you have any questions!
Now go back and choose the best display type for your parameter!