Data Processing
Making the numbers play nice with VB

Module 1
Variable Types

Variable Combinations

Variable Transformations
Independent Variables

→ MUST be quantitative for Virtual Beach

→ Should influence or explain bacteria somehow

→ Can be categorical numbers (ranked 1-3)
Variable Types

Quantitative
- Ratio
- Interval

Qualitative
- Ordinal
- Nominal
Rainfall (inches or cm)

Quantitative – Ratio
Turbidity

<10, 200, 1500 NTU

Quantitative – Ratio
Temperature (°C, °F)

Quantitative – Interval
Wind Direction (degrees)

Source: NOAA
Sky Conditions - sunny, partly cloudy, cloudy

<table>
<thead>
<tr>
<th>Ranked Categories</th>
<th>sunny</th>
<th>mostly sunny</th>
<th>partly sunny</th>
<th>mostly cloudy</th>
<th>cloudy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Qualitative – Ordinal
Variable Combinations

Interaction Terms
Multiply terms that interact with each other:
- Gulls & Wave Height
- Tributary Input & Alongshore current

Combine Categorical Value
Sum categories converted to numbers

Change-in-flow Variables
Subtract variables that occur over time
Explanatory variables are assumed to be independent from each other.

Collinear (non-independent) variables may over influence the model.

Variables may interact: e.g. river discharge & current.

Solution: Combine collinear variables into "Interaction term".
### Combined Categories

<table>
<thead>
<tr>
<th>Turbidity</th>
<th>Clear</th>
<th>Slightly Turbidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Day 2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Day 3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Turbid</th>
<th>Opaque</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Turbid + Opaque = ...**

<table>
<thead>
<tr>
<th>Turbidity</th>
<th>Clear</th>
<th>Slightly Turbidity</th>
<th>Turbid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Day 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Day 3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Change flow variables

Subtract one continuous data point from another to create proxy variables for changes in flow

EX.: Minimum and maximum stream flow over a day

\[(\text{Trib24 hr max}) - (\text{Trib24 hr min}) = \text{stream “flashiness”}\]
Variable Transformations

Including directional data

Making non-linear variables “more linear”
Nonlinearity

Since changes in *E. coli* concentrations can occur over orders of magnitude, it is a useful modeling practice to transform this variable.
Directional Data

Beach Orientation

Optional, but useful!
Beach Orientation in Compass Degrees

Wind & Current Directions given in numbers

Wind = 45
Current = 225
Computing Alongshore & Onshore components

Variable Types  Variable Combinations  Variable Transformations

45
-A
+O

+O
-A

0
270
180
90

Beach  Water

+O
-A
-O

+O

-A
Computing Alongshore & Onshore components

Variable Types  Variable Combinations  Variable Transformations

Map Controls
- 43.74
- 27.71

Map Settings
- Type
- Brightland

Beach Orientation
- Remove 1st Beach Marker
- Remove 2nd Beach
- Remove Water Marker
- Beach Orientation: 21.93

Current Location
- 43.7407313808212
- 27.7071511745453

Loading

- Beach
- Water

Variable Types
- Variable Combinations
- Variable Transformations

21.93
Computing Alongshore & Onshore components