Assignment

• Use the spaeth.csv data from earlier:
• Eliminate species with 2 or fewer occurrences, then log transform the data
• Calculate Bray-Curtis similarity with the log transformed data
  • Perform PCoA with k=3
    • Report total variance explained, and the percent variance explained by each axis
  • Repeat PCoA with k=4
    • Report total variance explained, and the percent variance explained by each axis
• Plot the first two axes of PCoA with different symbols or colors for creek.

• What species have the highest correlation with axis 1 and 2? How does that help interpret your plot?
Ordination object

- \texttt{cmdscale} function returns
  - points – k axes scores for each sample
  - eigenvalues
  - Goodness of Fit measure

- How do we assess how well the analysis did?
  - Goodness of Fit (GOF) ranges from 0-1, and can be interpreted similarly to $r^2$
  - We can also calculate $r^2$ directly from the scores.

---

### Partitioning variation

- Total of all eigenvalues = 9.14
- Relative importance of each axis given by dividing each eigenvalue by the total...
  
<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Proportion of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.43</td>
<td>0.48</td>
</tr>
<tr>
<td>2.76</td>
<td>0.30</td>
</tr>
<tr>
<td>1.28</td>
<td>0.14</td>
</tr>
<tr>
<td>0.654</td>
<td>0.07</td>
</tr>
<tr>
<td>0.19</td>
<td>0.02</td>
</tr>
<tr>
<td>0.076</td>
<td>0.008</td>
</tr>
<tr>
<td>0.052</td>
<td>0.006</td>
</tr>
<tr>
<td>0.029</td>
<td>0.003</td>
</tr>
</tbody>
</table>

- Variance explained by each axis is different.
- For each axis, assess $r^2$ for the euclidean distance among all samples on that axis to the original distance matrix
- This loop does this for all k axes

```r
pct_var<-0
for(k in 1:2){
pct_var[k]<-100*cor(vegdist(prin_coord$points[,k],method="euclidean"),bray_distance)^2
}
pct_var
```
What does ordination space represent?

Samples plotted in two dimensional ordination space.

The size of the symbol is proportional to species 5 abundance.
Raw data, Second sample file

GOF values

Highly Structure Dataset

Randomized Dataset

K=3
Eigenvalues: 90% in first 3
GOF 0.85-0.89
Correlations between axes and
bray Curtis matrix 55.5 19.5
and 12.7 (sum 87.7)

K=3
Eigenvalues: 54% in first 3
GOF 0.45-0.54
Correlations between axes and
bray Curtis matrix 22.1 11.9
and 3.9 (sum 12.5)
Reporting PCoA Results

- Dimensions of original data matrix
  - In our example - 10 species and 29 samples
- Transformations or handling of missing/rare data
  - None in our example
- Type of distance matrix
  - Bray Curtis
- Number of dimensions retained
  - 3
- Total variance explained
- Relative importance of axes

Other Examples of PCoA this semester

- Very often you want to show groups of samples. This can be done with point colors or symbols:

```r
sample_groups <- factor(sample_groups)
sample_colors <- c("red", "green", "blue")
plot(prin_coord$points[,1:2], pch=19, col=sample_colors[
```
```r
plot(prin_coord$points[,1:2], pch = sample_symbols[sample_groups], col = "black")
```