



Coeur d'Alene, Idaho
31 March – 4 April, 2003

Case Study: Classification of Western Streams

Presented by
Jeroen Gerritsen, Tetra Tech, Inc.

Reference and Degraded Site Designations (McIntyre)

Environmental data

Taxonomic Data

A priori and *a posteriori*
site classification
(Gerritsen)

Metric Data
(Frydenborg)

Metric Exploration
(Frydenborg)

Select Responsive Metrics
(Blocksom)

Develop Final Multimetric
(Blocksom)

Multimetric

A posteriori
site classification

Community
Cluster Groups

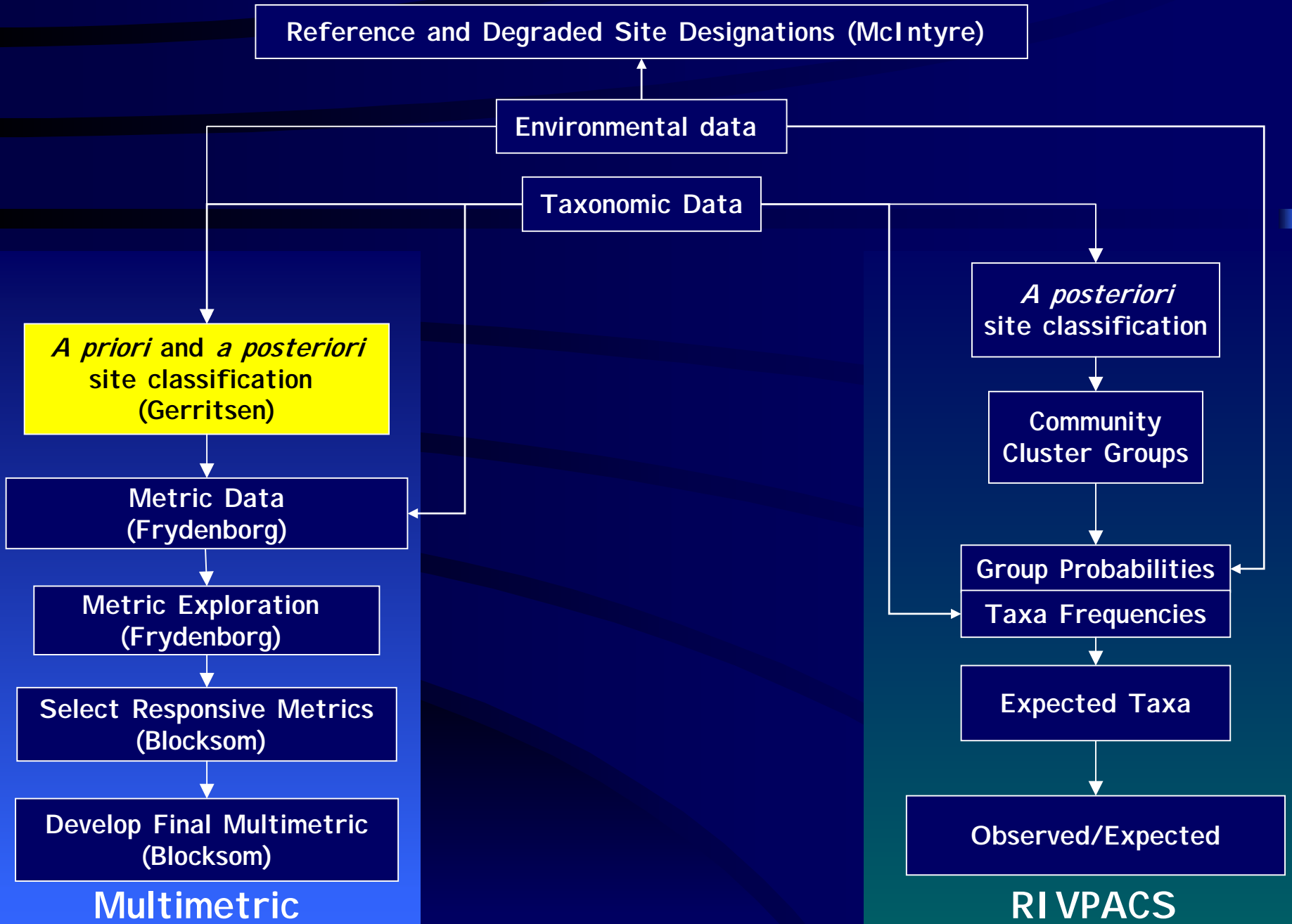
Group Probabilities

Taxa Frequencies

Expected Taxa

Observed/Expected

RIVPACS



Multimetric Index Development

- Database consisting of reference and stressed populations (sites)
- **Classify resource (reference sites)**
- Identify and test candidate metrics
- Select metrics for dimensionless index
- Select thresholds for assessment

Classification of Sites (Streams)

- **The intent of classification is to identify groups of sites that under ideal conditions would have comparable biological communities**
- **Classification should rely on those characteristics of sites that are intrinsic, or natural, and not the result of human activities**

Classification approaches

- *A priori* rule-based models
 - geographic regions
 - salinity zones
- *A posteriori* development of rules
 - Cluster analysis followed by discriminant models
- Gradient and mixed models
 - Elevation, catchment size, salinity, depth, etc.

Classification of Wyoming streams

Testing an *a priori* model

Middle Rockies

Western ranges

Tetons, Absaroka, Wind River

Wyoming, Salt River

Central – Bighorns

Bighorns

East – Black Hills

Black Hills

Southern Rockies

Medicine Bow, Laramie ranges

Wyoming Basin

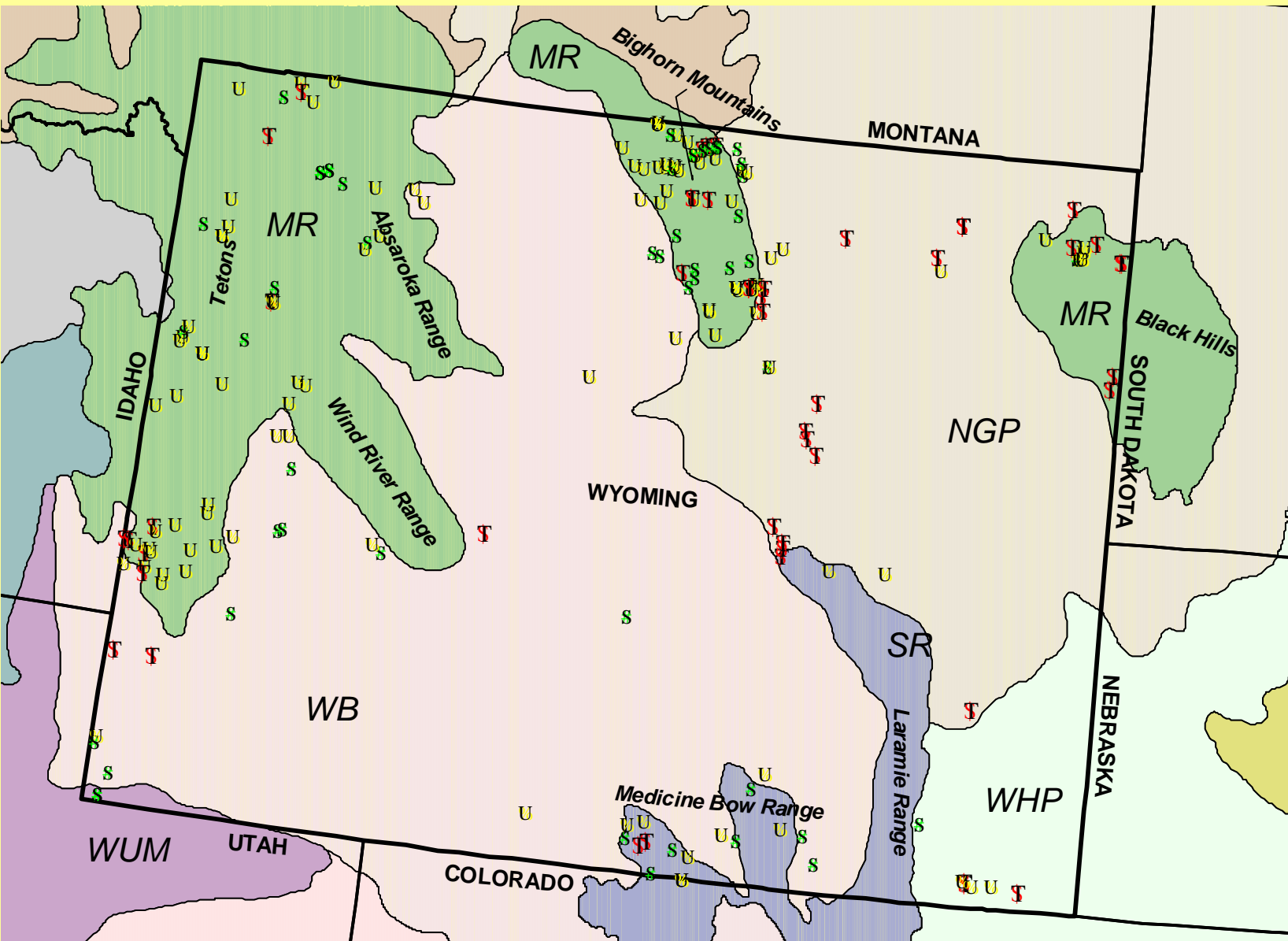
sagebrush high desert

Northwestern Great Plains

mostly tall grass prairie

Western High Plains

mostly short grass prairie



Wy sampling sites

- S Reference
- U Other
- F Impaired

Ecoregions

- SRB
- NBR
- WVFP
- MR
- WB
- WUM
- CP
- SR
- WHP
- ST
- CGP
- NGP
- NGP
- NSH

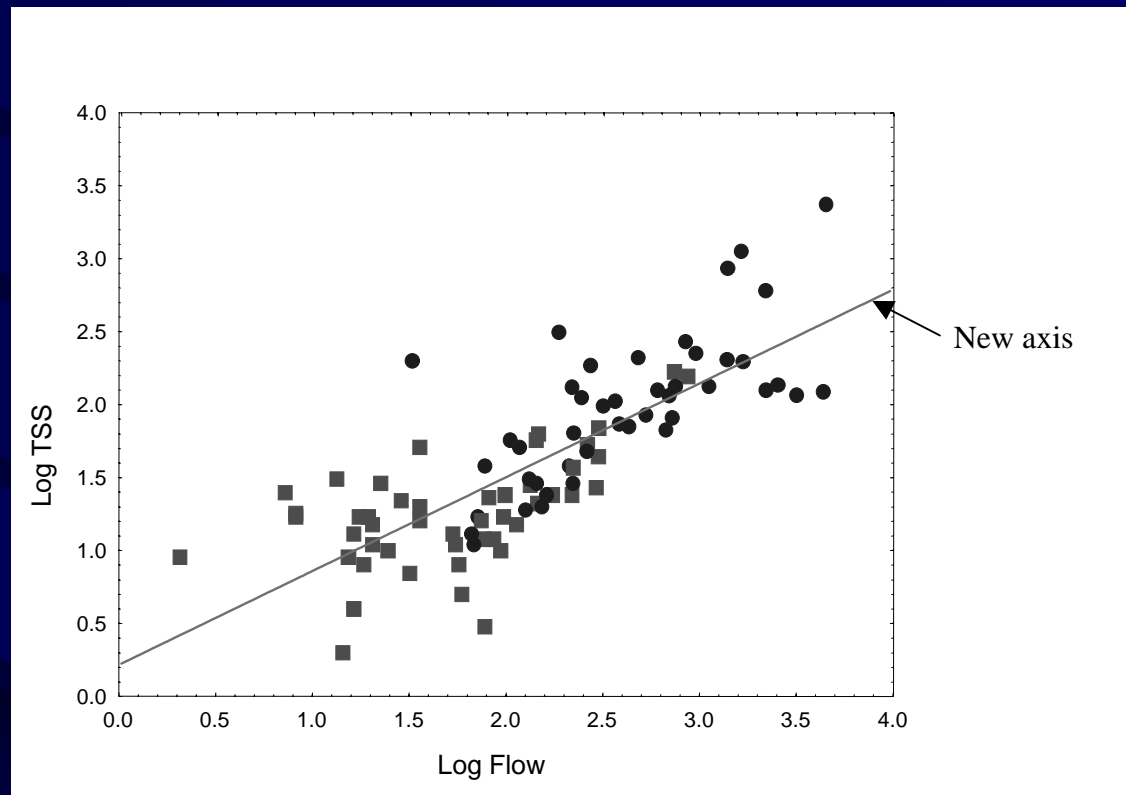


Other classifications

- Elevation
- Gradient
- Watershed area
- Climate
- Geology
- Latitude, longitude
- Natural water quality (alkalinity, color)
- Substrate

Ordination

- Putting things in order, according to their similarity
- Reduce dimensions: regression line is new axis
- What if we have 3 variables? 4 or more?



Ordination

- PCA (Principal Components Analysis)
 - Multivariate extension of regression
 - Assumption: normal distribution, linear
 - NOT suited for species data
- Correspondence analysis (CA)
 - Uses chi-square as similarity
- Non-metric multidimensional scaling (NMS)
 - Non-metric: converts distances to ranks, then does ordination on ranks
 - Recreates map using only distances between cities
 - Points close together are similar: use this to visually identify groups and structure

Similarity

- Ordination works on some measure of similarity (or dissimilarity)
- e.g., Jaccard similarity:

$$JI = \frac{\text{Taxa in common}}{\text{Total taxa}}$$

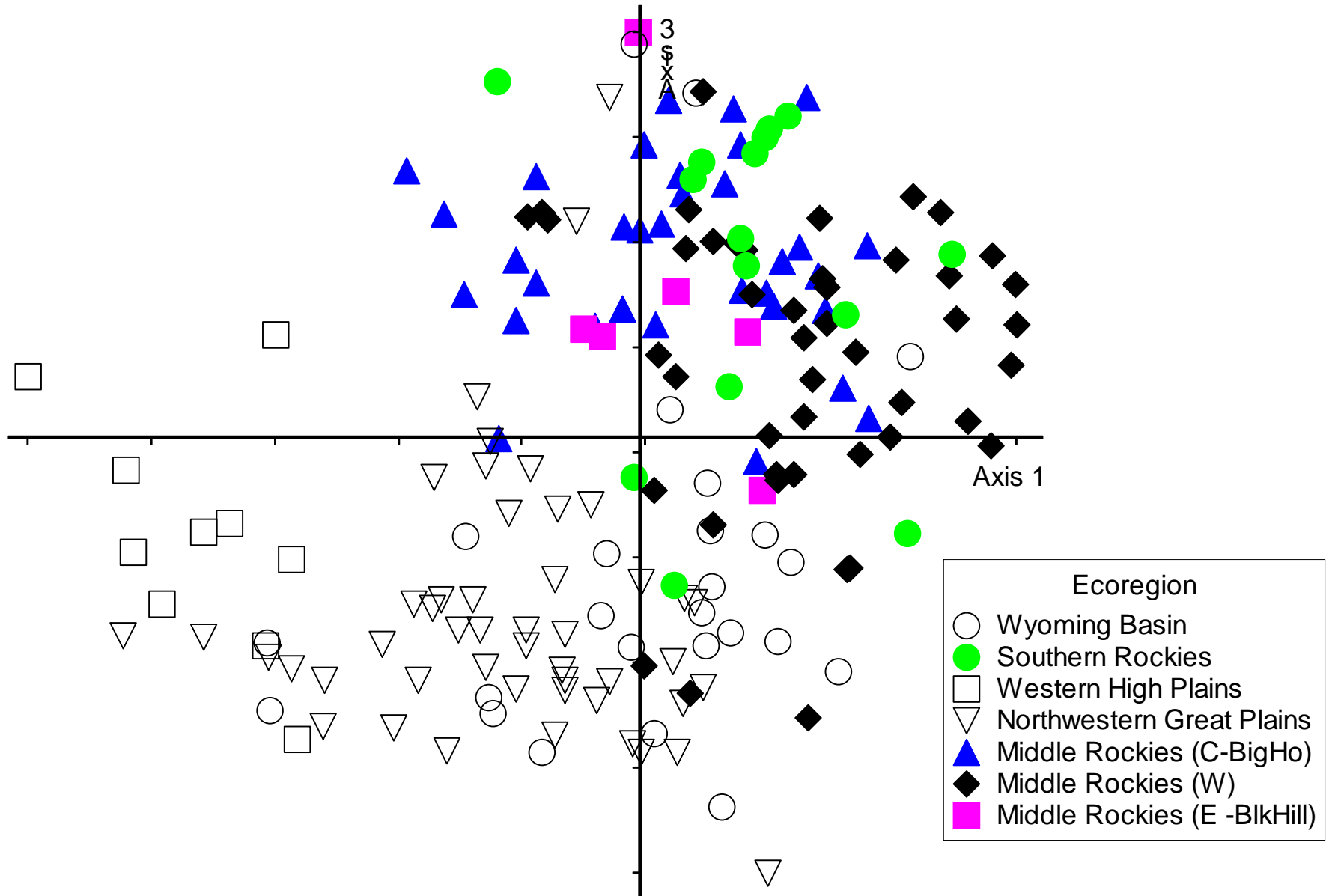
- There are many similarity indexes!

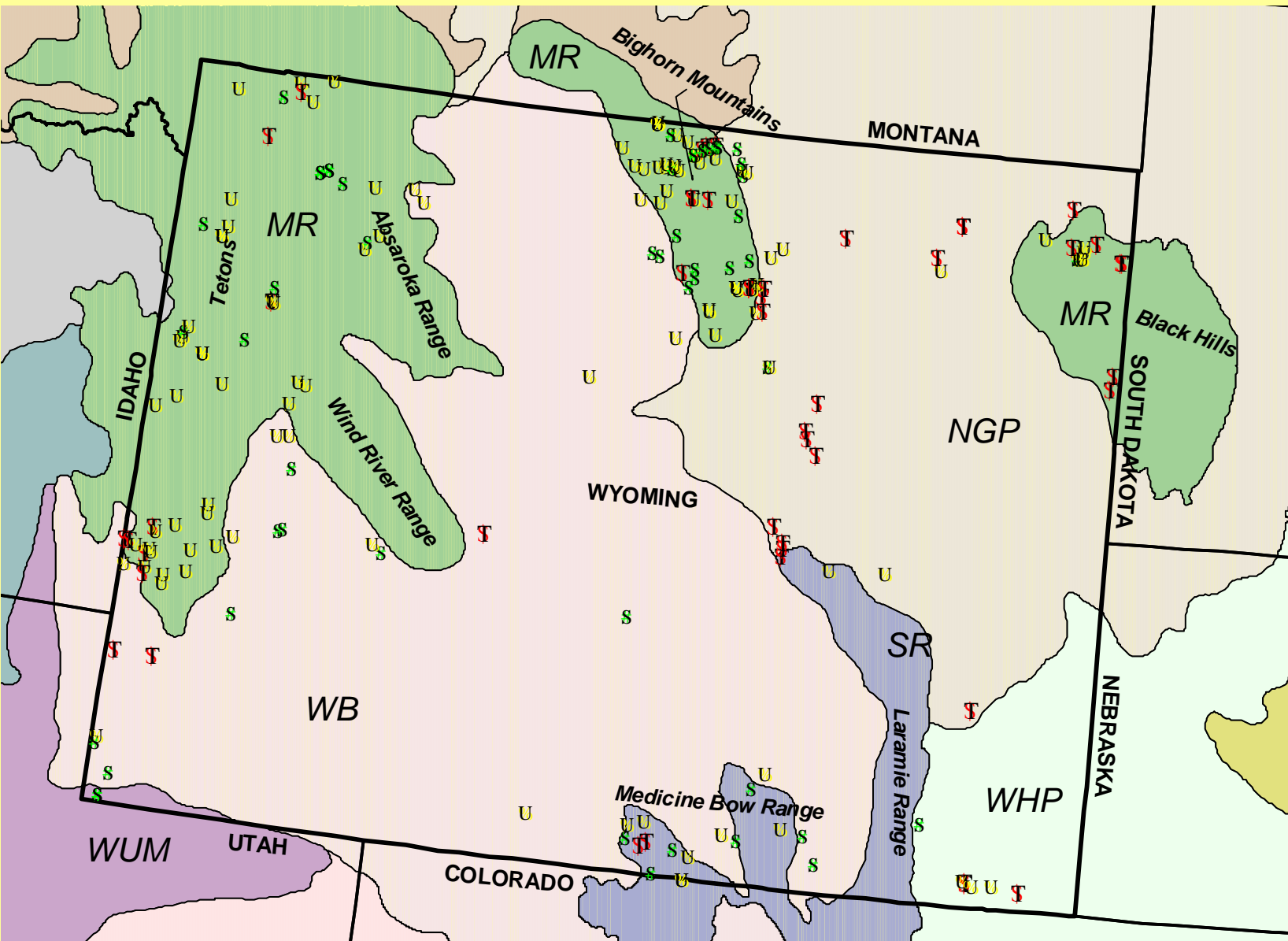
Ordination

Let's try it –

- NMS of Wyoming sites
- Similarity metric is proportional Bray-Curtis (a measure of % similarity)
- Plot sites in the reduced dimensions (called “ordination space”)
- Look for structure with respect to *a priori* classes

Wyoming Reference Sites





Wy sampling sites

- S Reference
- U Other
- F Impaired

Ecoregions

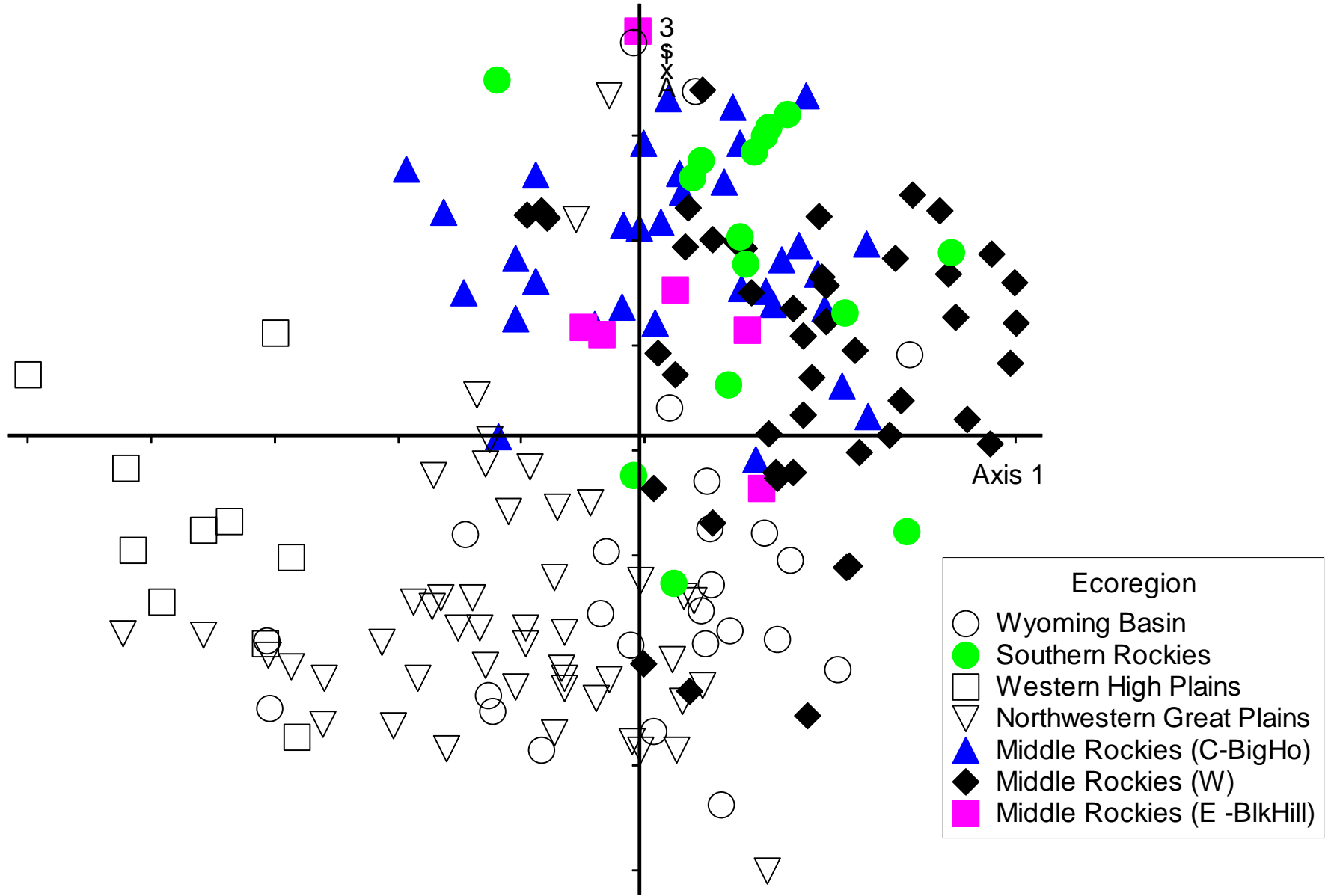
- SRB
- NBR
- WVFP
- MR
- WB
- WUM
- CP
- SR
- WHP
- ST
- CGP
- NGP
- NGP
- NSH



How do we read these?

- Points close together are similar; far apart are dissimilar
- Look for patterns in grouping of a priori classes
- Axes (in NMS) are **not** meaningful by themselves

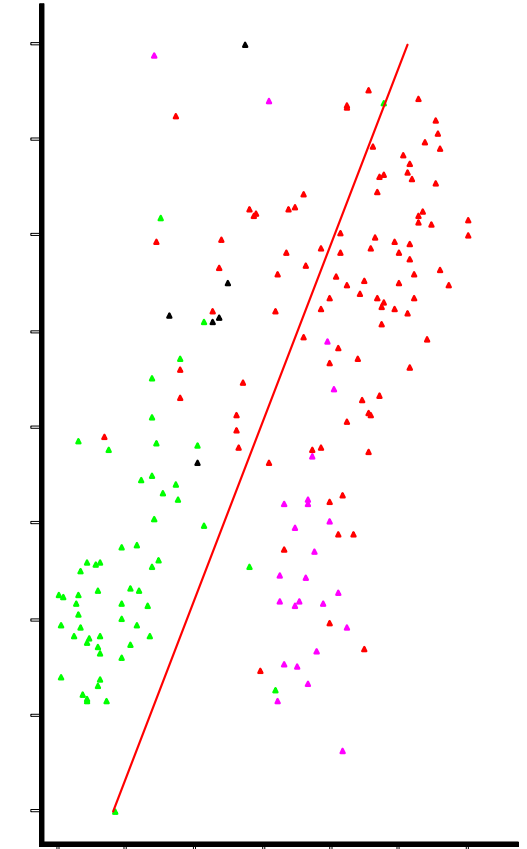
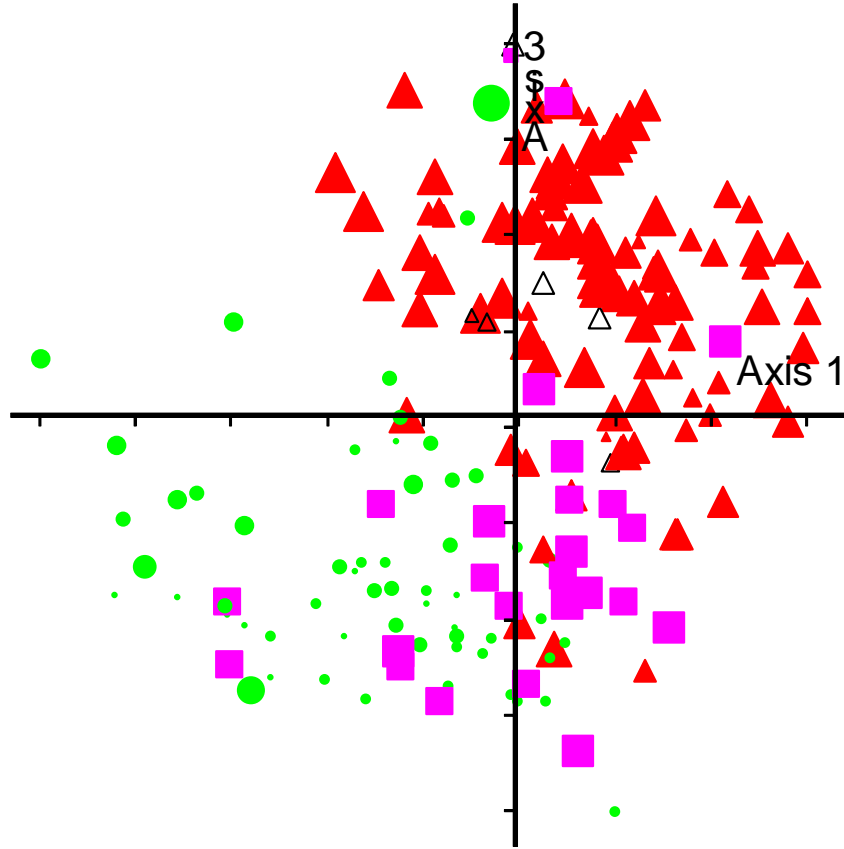
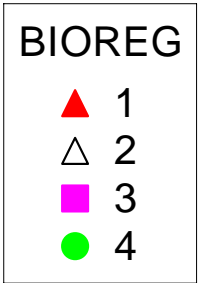
Wyoming Reference Sites



Now add other variables

- We consider a single continuous variable (e.g., elevation), and plot elevation against scores on the NMS axes to see if elevation is associated
- We can also scale the size of the symbols in the ordination plot to reflect the continuous variable (elevation)

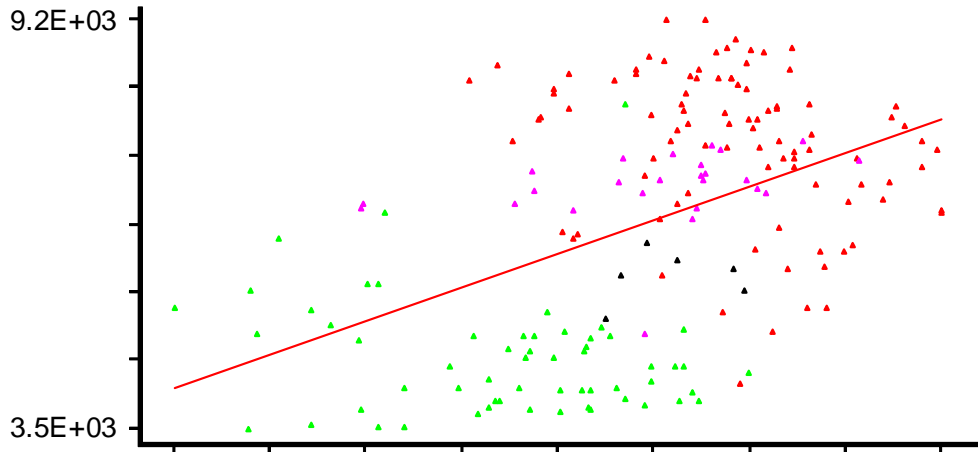
Wyoming reference sites



3.5E+03 9.2E+03

ELEVATIO

Axis 1
 $r = .458$ $\tau = .287$
Axis 3
 $r = .601$ $\tau = .445$



9.2E+03

3.5E+03

Classification exercise

1. Wyoming

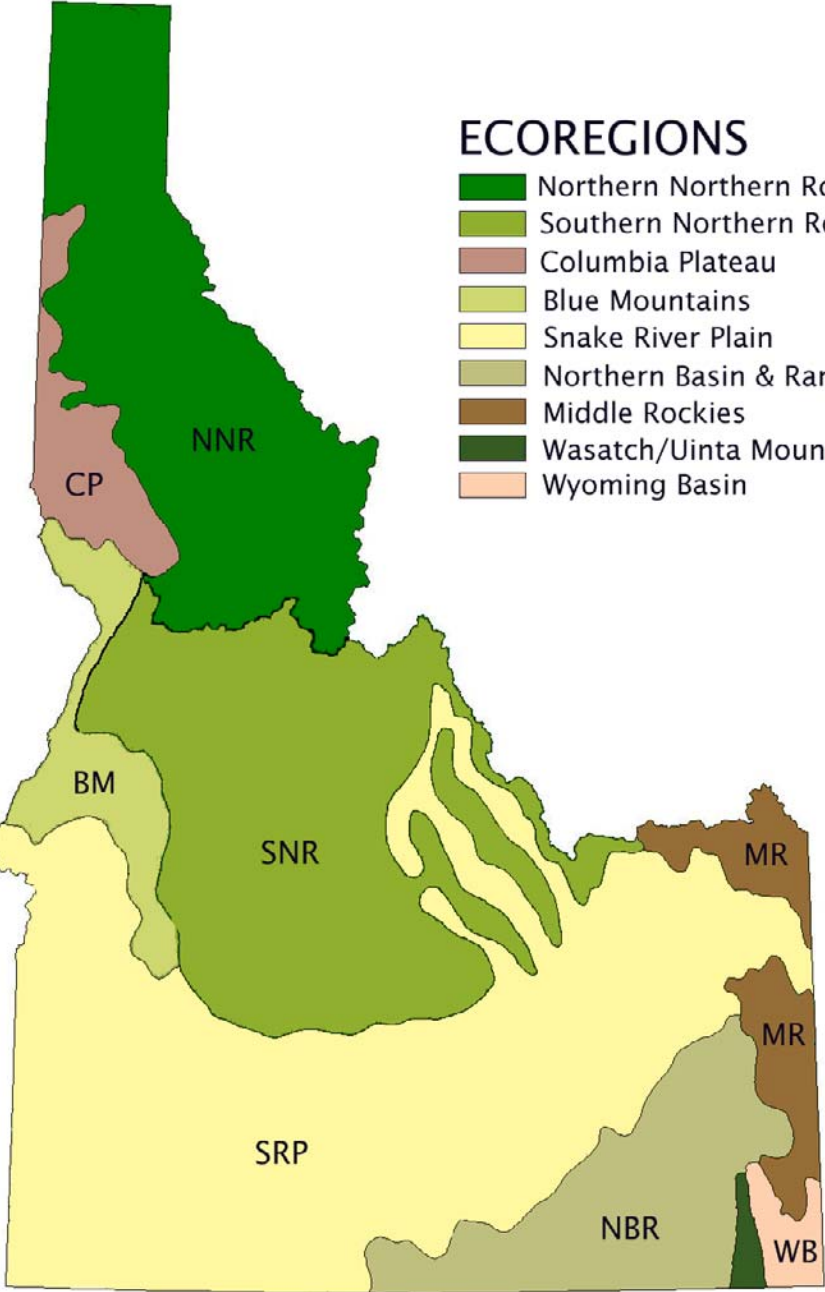
- **Objective: interpretation of classification information**
 - Pages 1-12 of handout (you have seen 2-5)
 - Look at pages 2-12: associations with other variables
 - Look at these associations and develop a conceptual model (in your head) of the factors that structure stream communities
 - Can you develop alternative classifications to the geographic one we have presented (bioregions)?
 - Which do you think is better?

Part 2: Idaho

- Now look at pp 13-21
- We will run through 14-15, and a map, as a group
- Try to develop a classification for Idaho

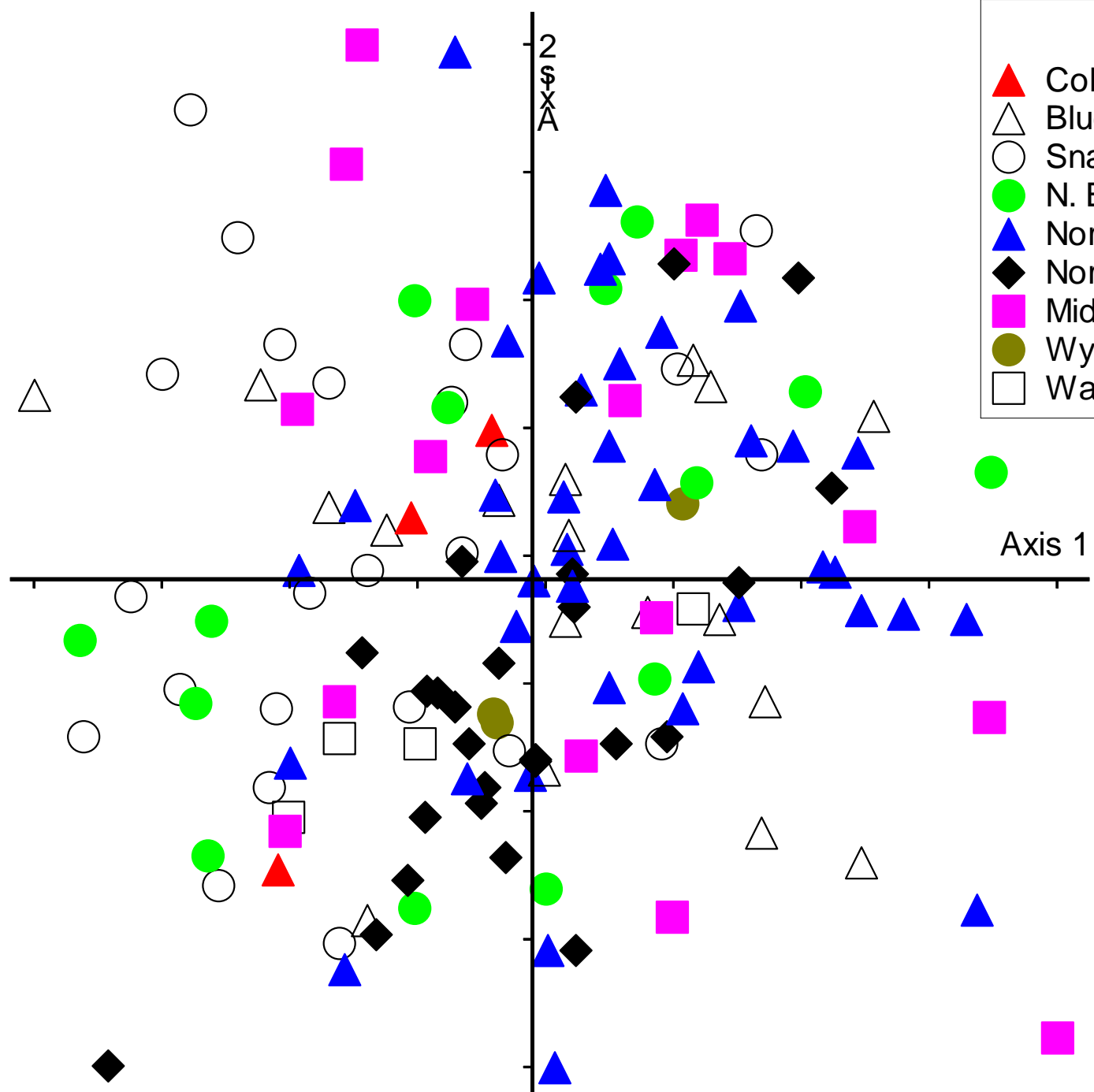
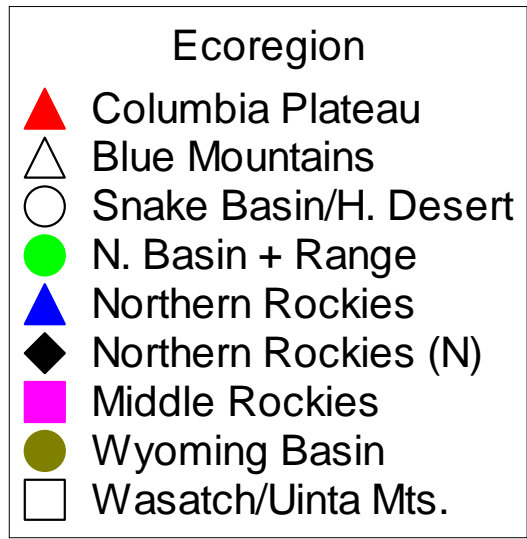
ECOREGIONS

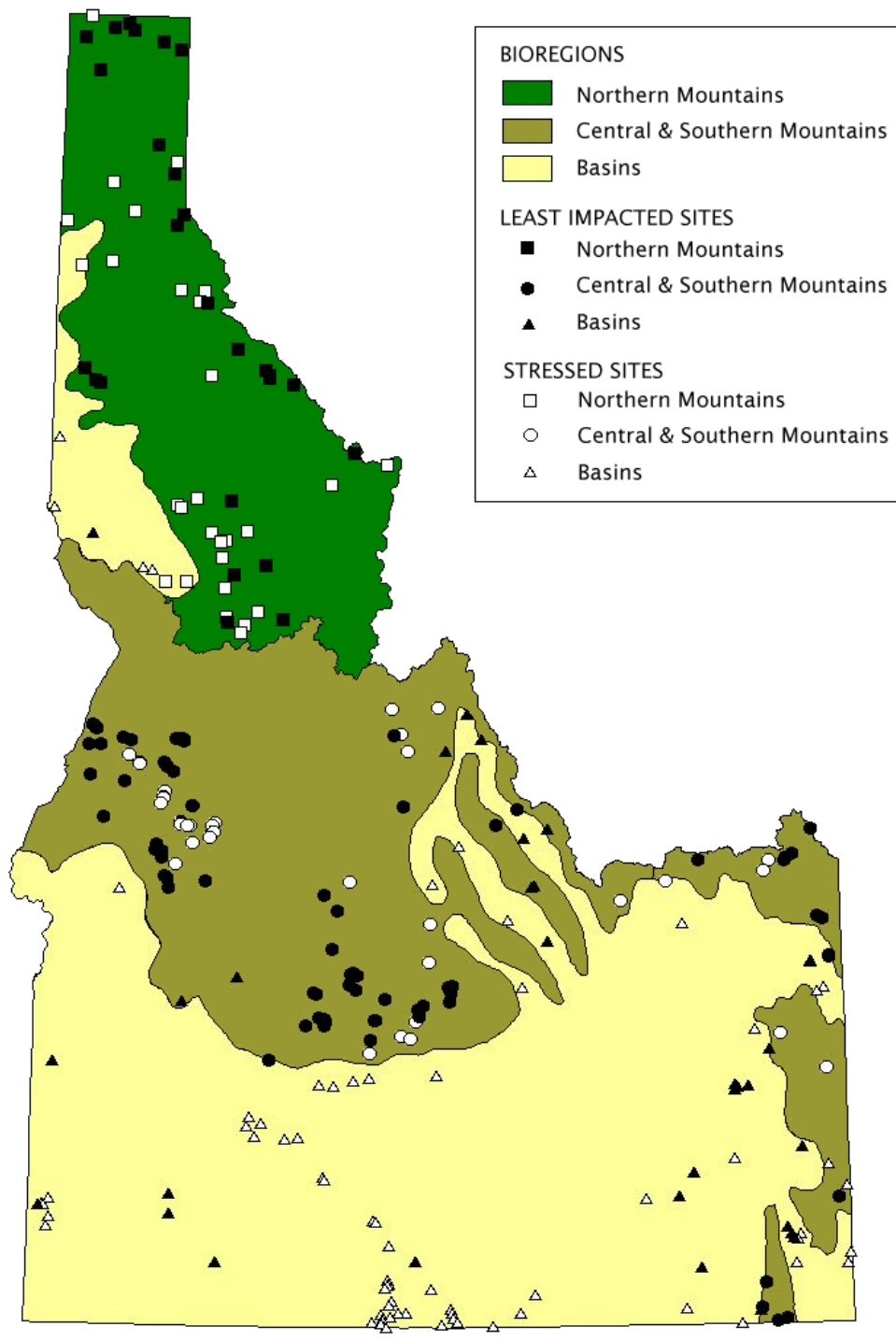
- Northern Northern Rockies
- Southern Northern Rockies
- Columbia Plateau
- Blue Mountains
- Snake River Plain
- Northern Basin & Range
- Middle Rockies
- Wasatch/Uinta Mountains
- Wyoming Basin



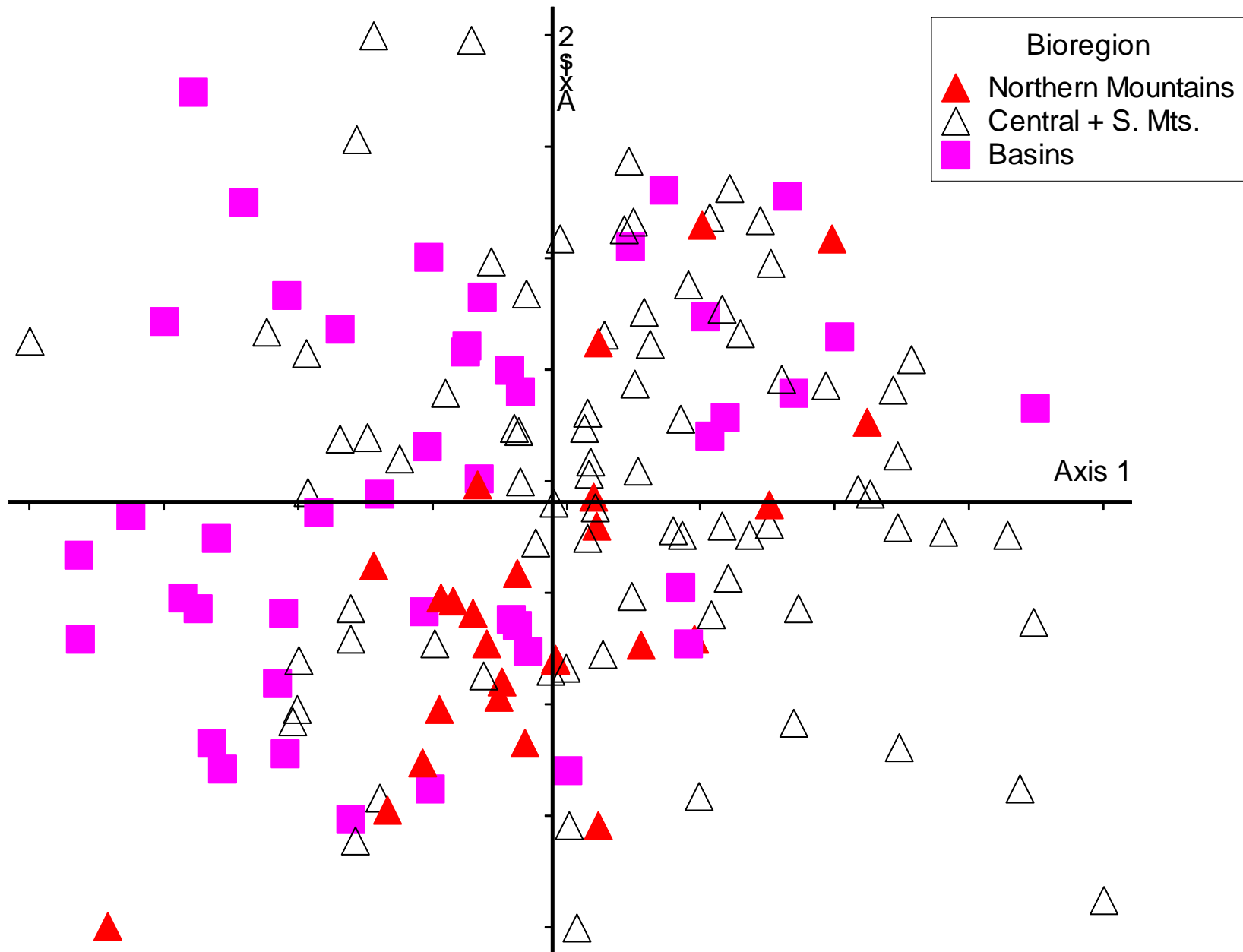
WUM

Idaho Reference Sites





Idaho reference sites



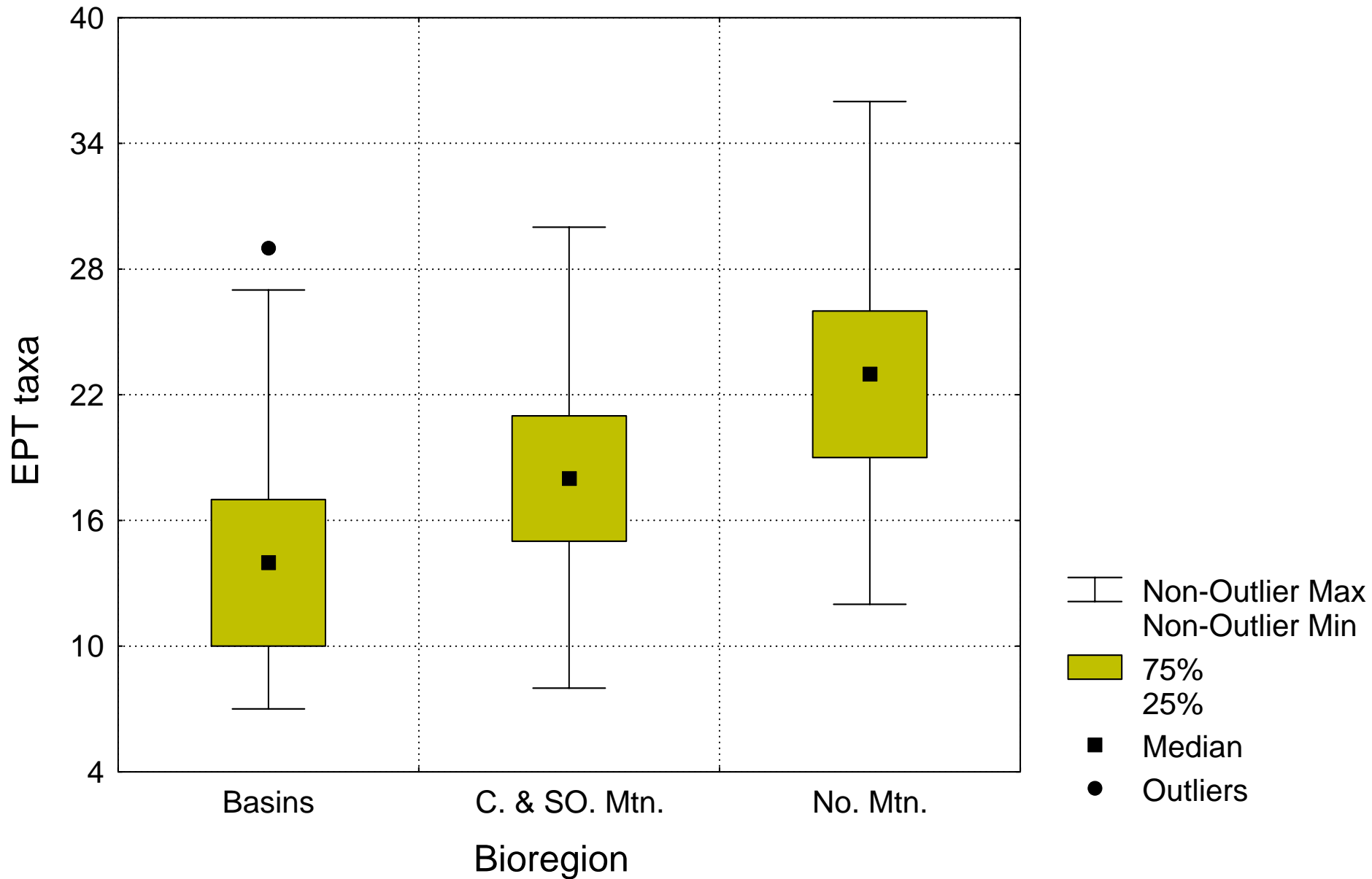
Idaho classification

- Stream classification in Idaho was not as easy!
- There was no clear distinction among regions
- Environmental variables (elevation, etc.) were equally nebulous
- Why?

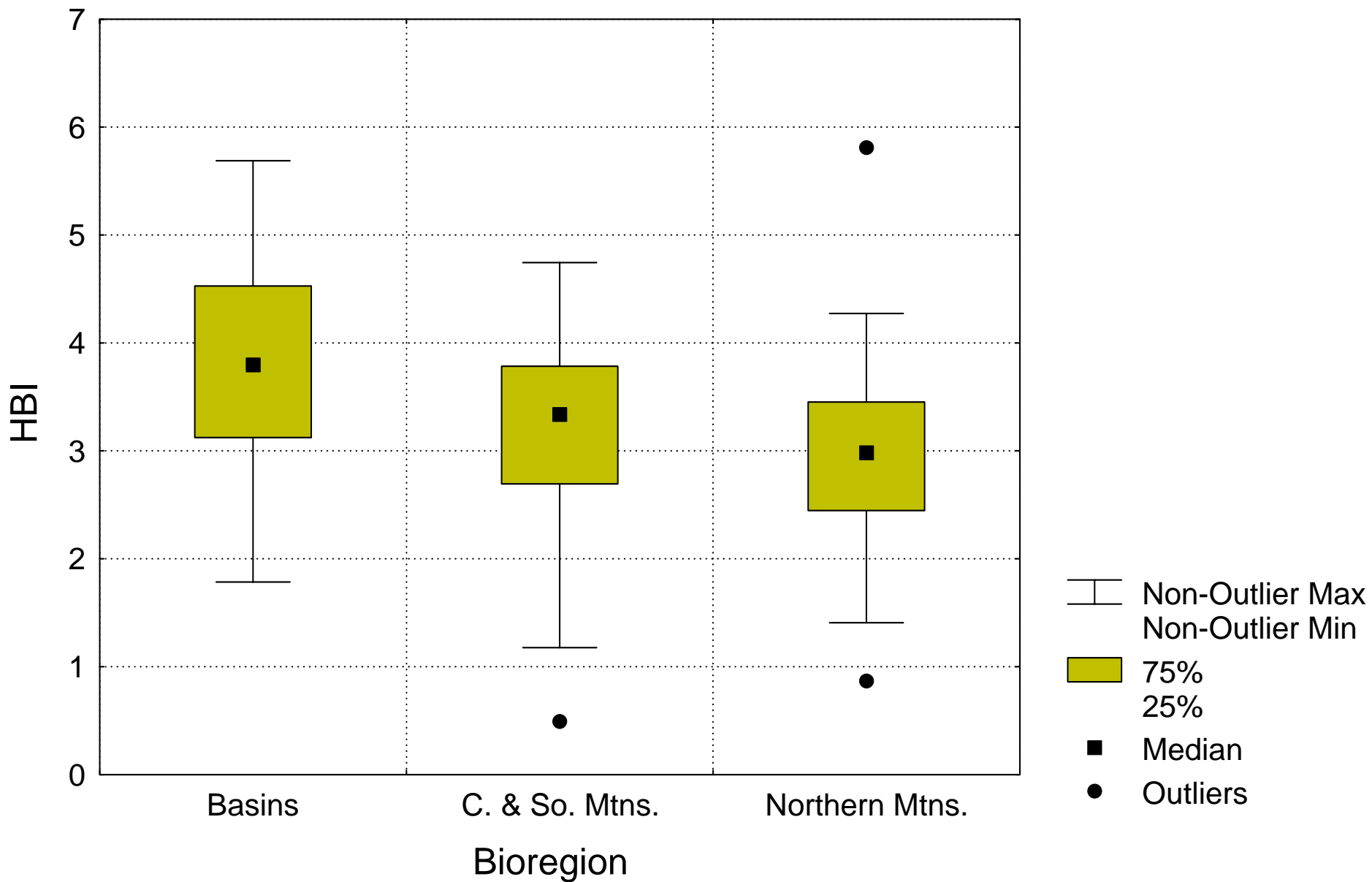
Classification

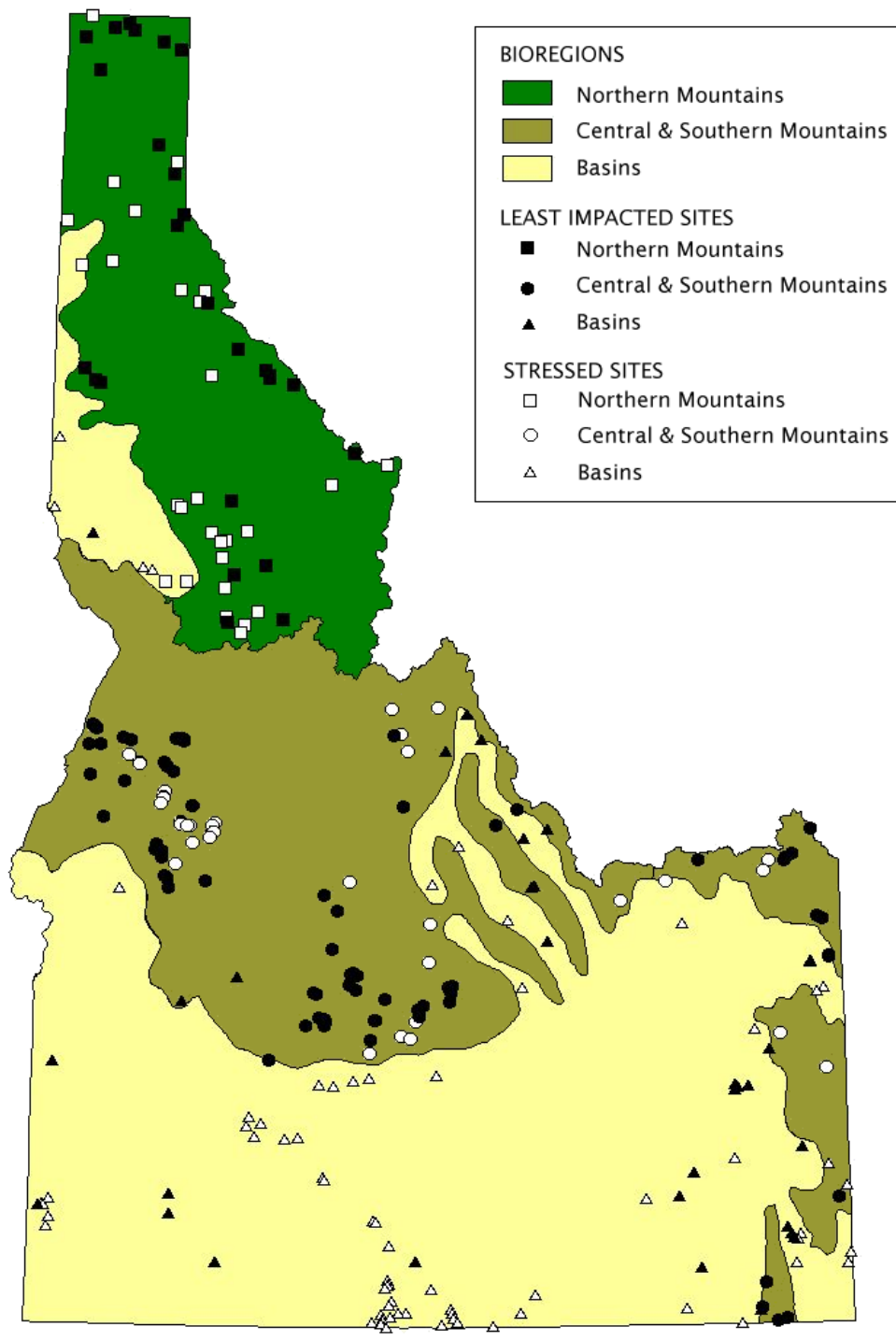
- We have built site classes so far on species composition. However, we will be building the index with metrics. Do the site classes make sense with metrics?

Idaho Reference Sites



Idaho Reference Sites





Idaho

- Classification is usually done with community composition. In Idaho, taxa were not predictable by geography or other variables
- Metric values **did** segregate on the geographic classes.
- For a multimetric index, classification must make sense for metrics: always check your classification with metric values!

