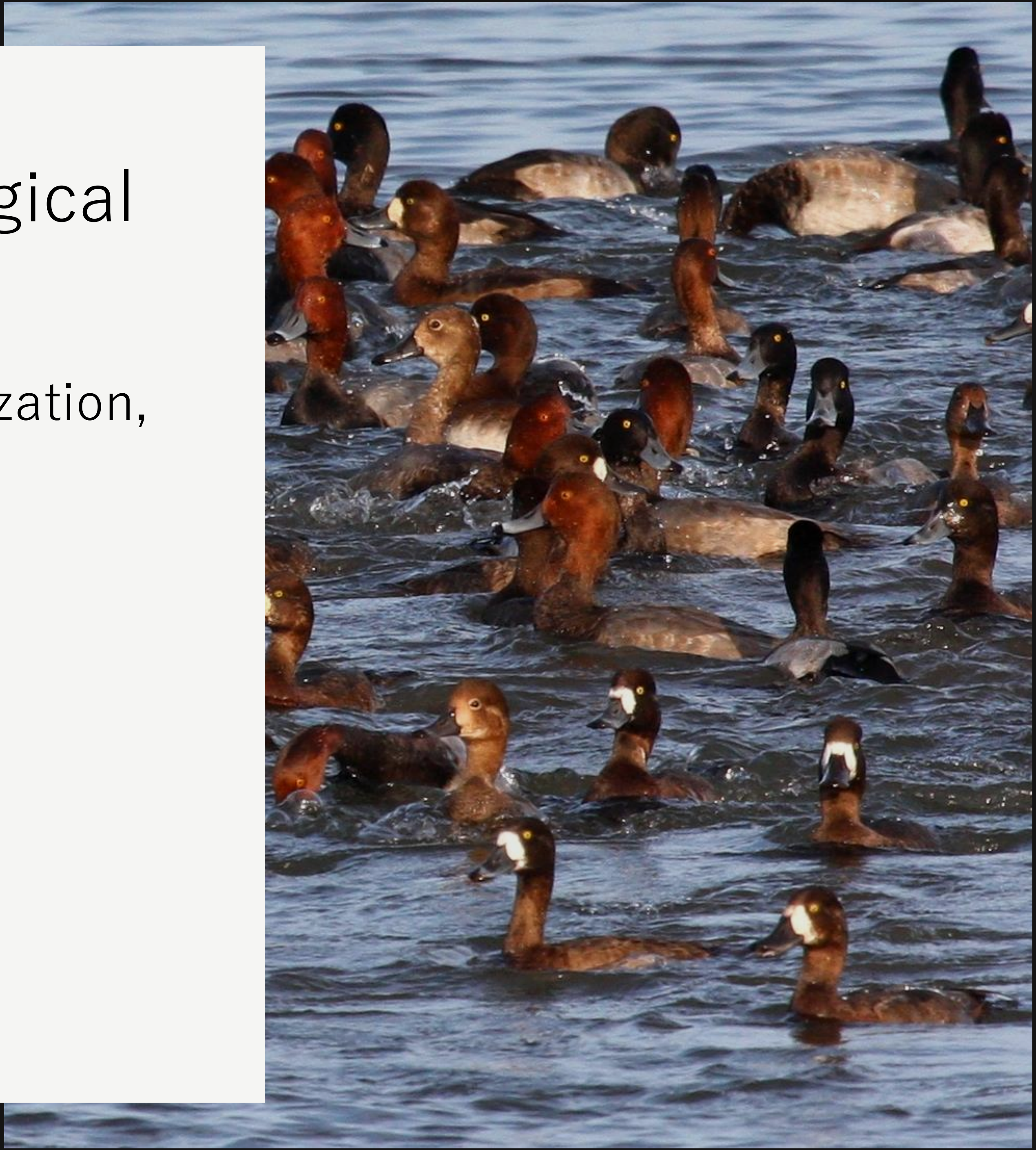
# FW 599 Special Topics: Multivariate Analysis of Ecological Data in R

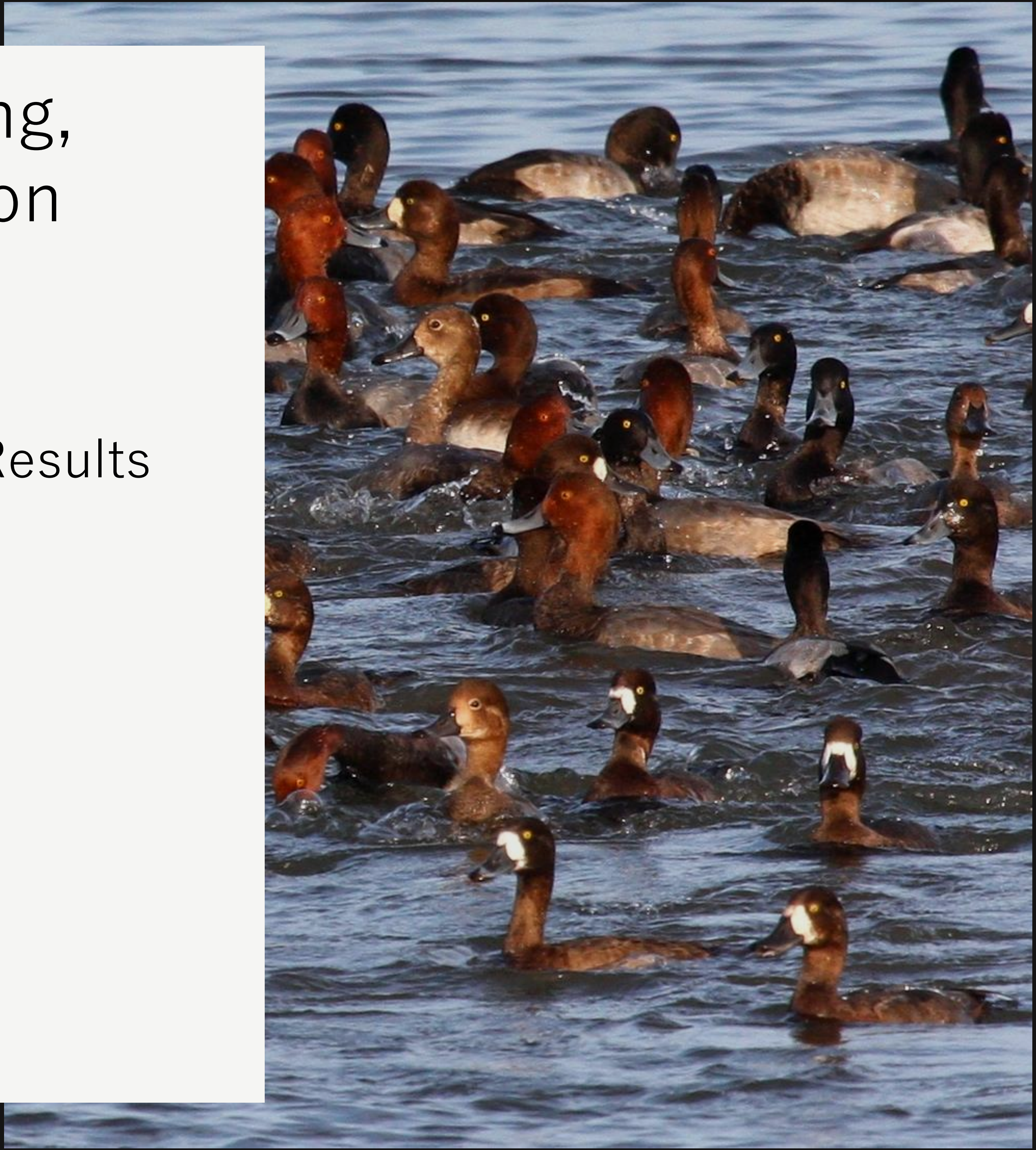
## Lecture 8: Variance Partitioning, Visualization, and Interpretation

Thursday, October 24, 2024

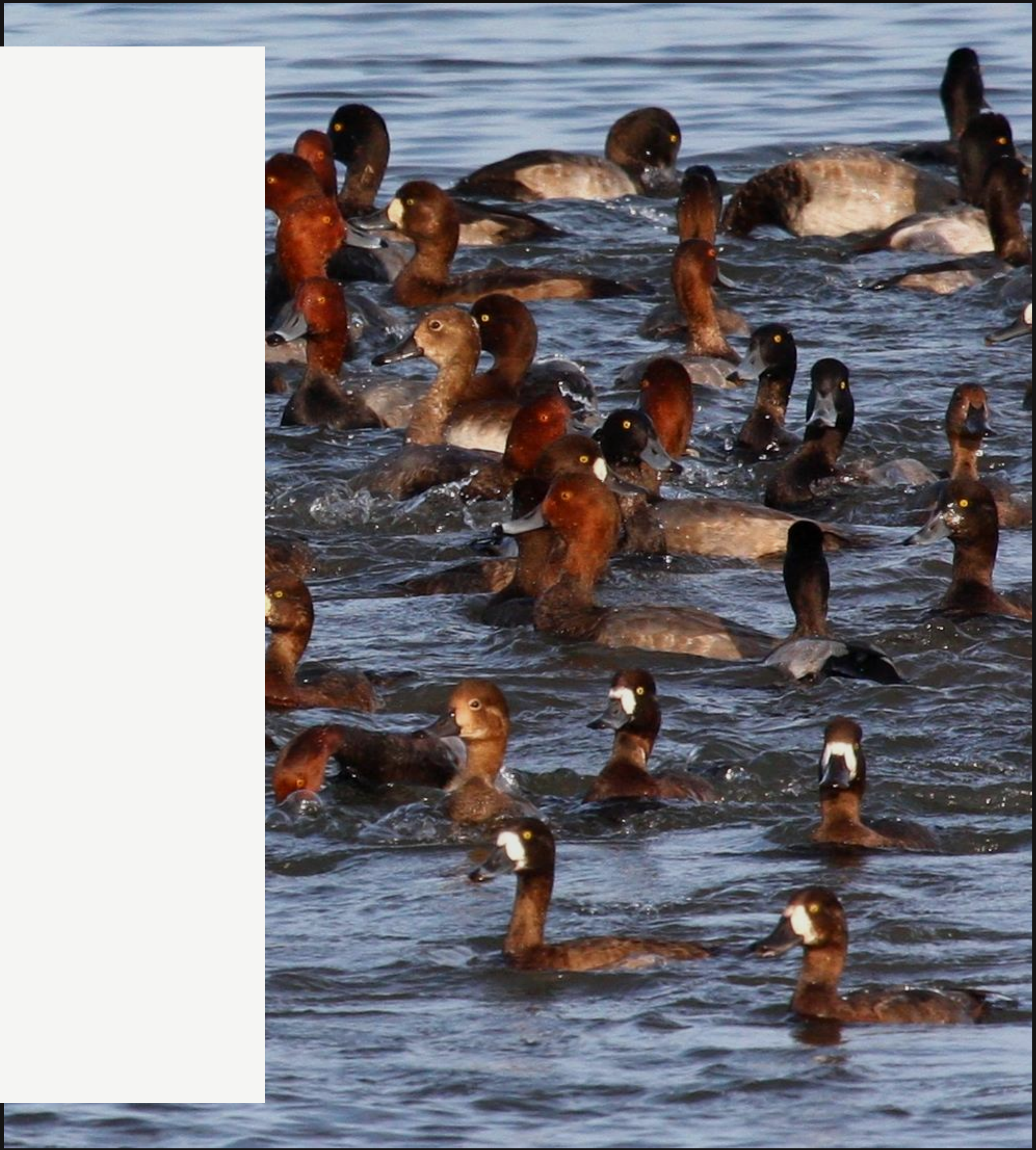


# Lecture 8: Variance Partitioning, Visualization, and Interpretation

- Variance Partitioning in Ordination
- Effective Visualization of Ordination Results
- A Very Light Introduction to Inference
- Inferring Causation



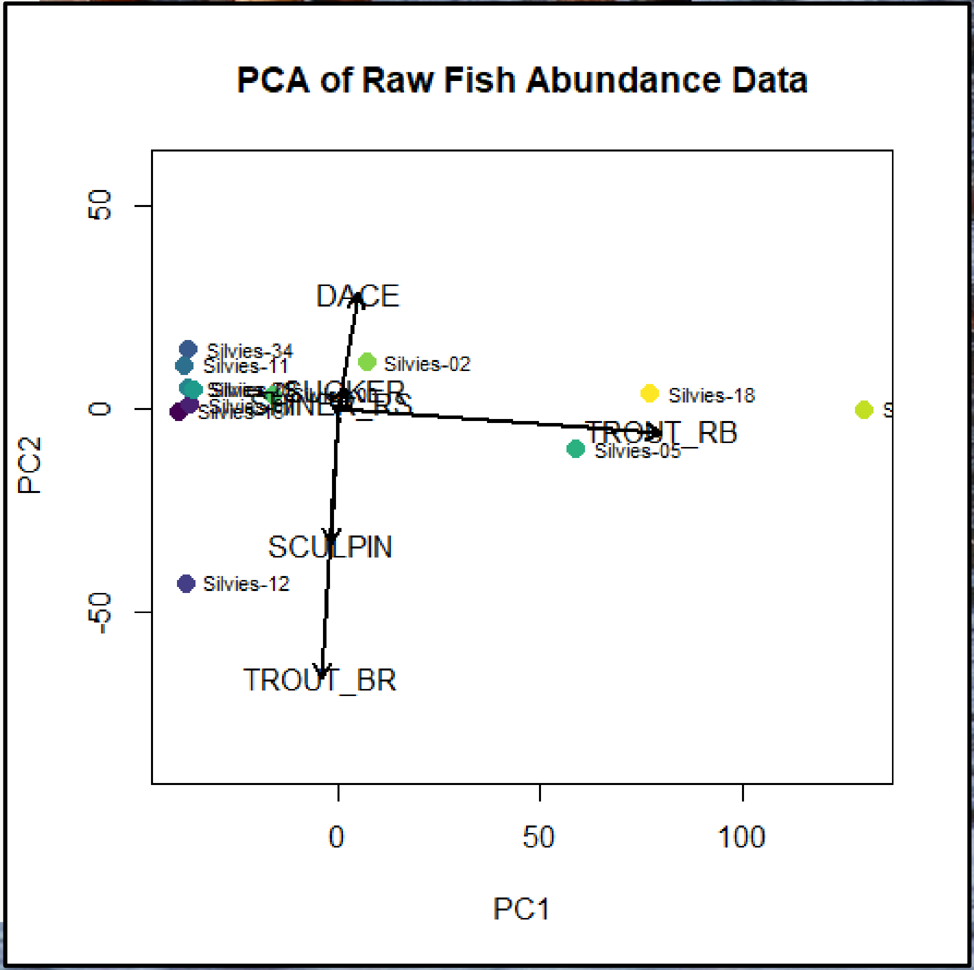
# Recap: Ordination



# Recap: Ordination

**Principal Component Analysis** uses eigenanalysis to reduce the dimensionality of large, ecological datasets while retaining as much information as possible.

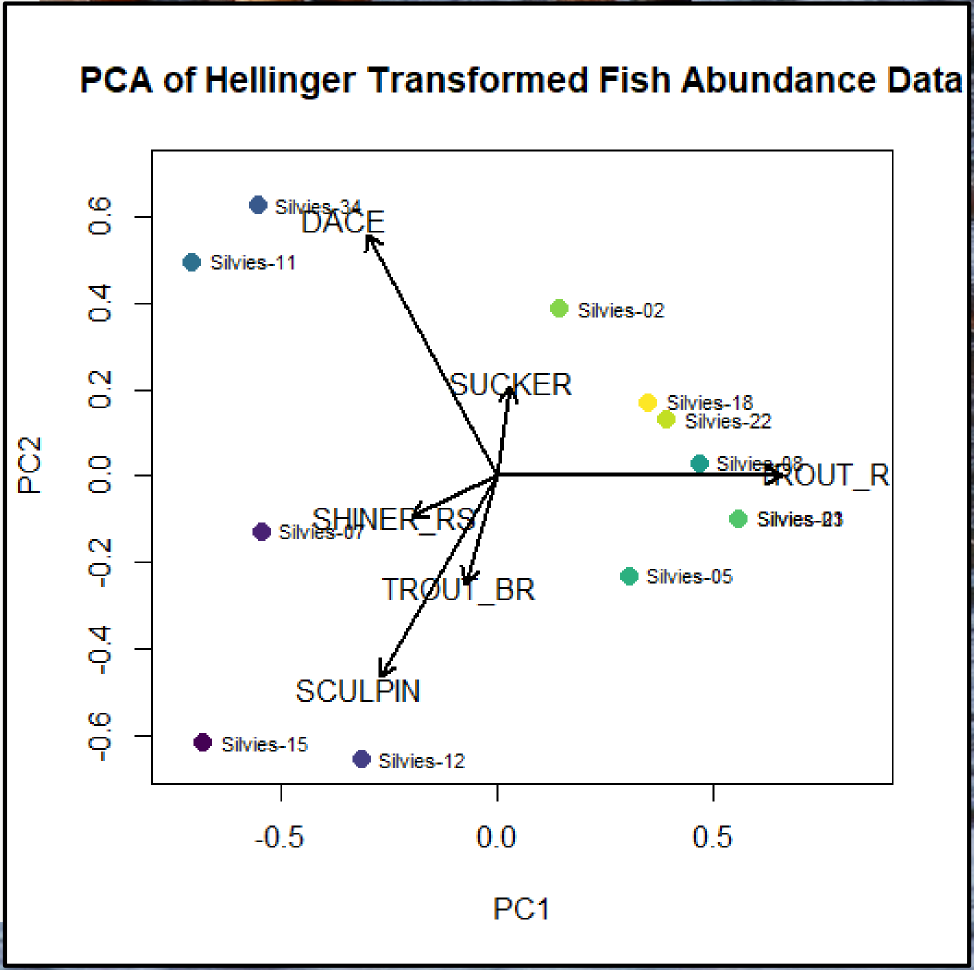
- Re-projects data in multidimensional space
- Maximizes the variance explained by the first principal axes (eigenvectors)



# Recap: Ordination

**Principal Component Analysis** uses eigenanalysis to reduce the dimensionality of large, ecological datasets while retaining as much information as possible.

- Re-projects data in multidimensional space
- Maximizes the variance explained by the first principal axes (eigenvectors)



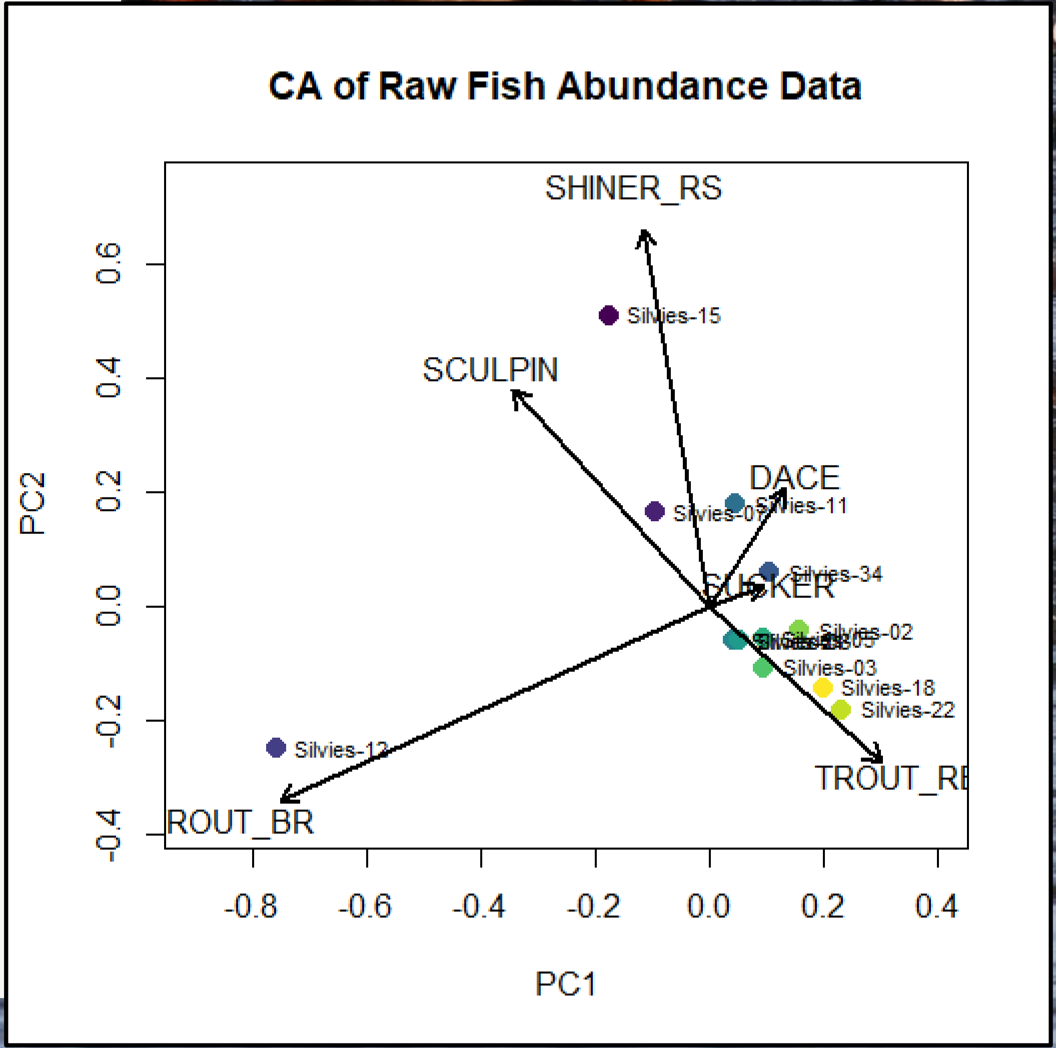
# Recap: Ordination

**Correspondence Analysis** is an alternative ordination method that preserves the chi-square distances ($\chi^2$) among objects in the principal axes.

Designed for dimensionally homogeneous, non-negative data (such as counts).

Computed on a two-way contingency table instead of a dispersion matrix.

Excludes double zeros!

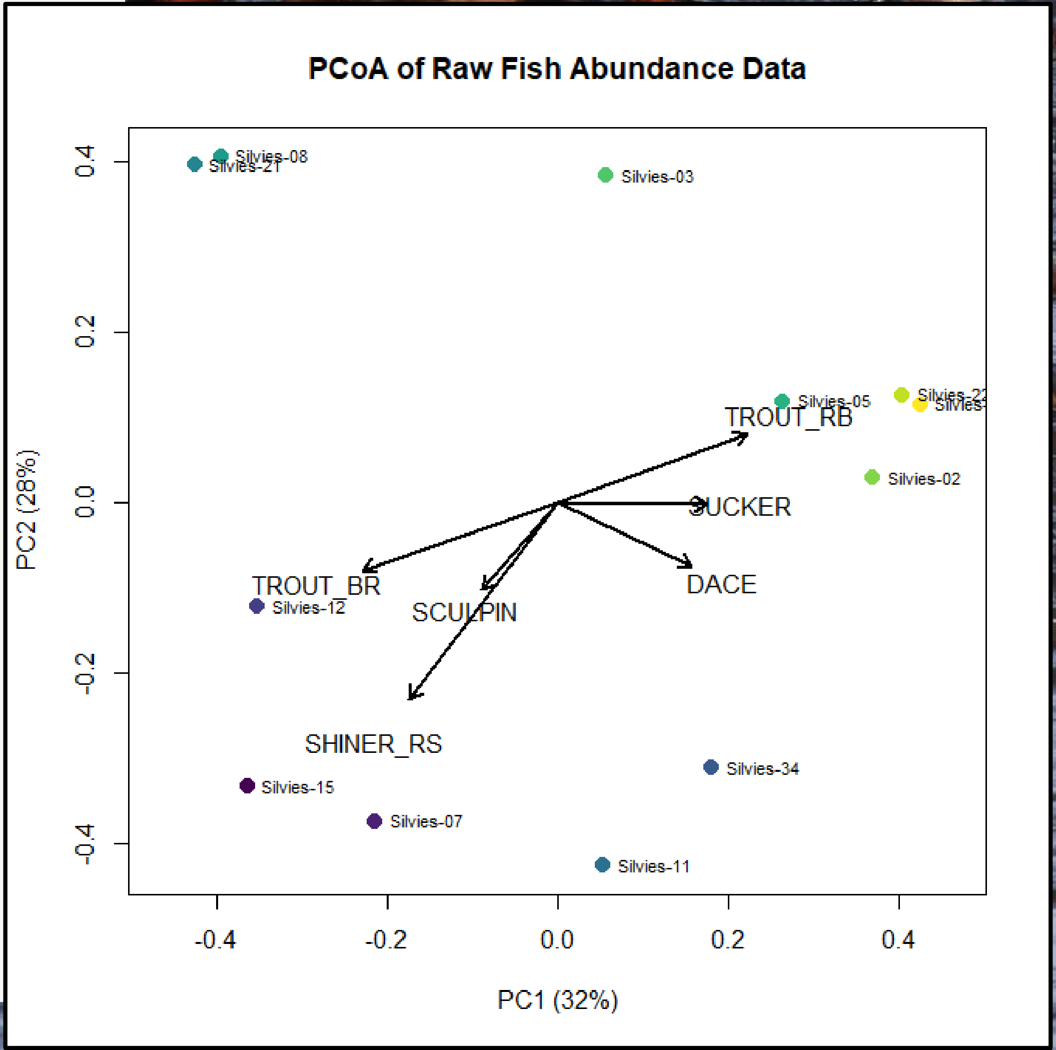


# Recap: Ordination

**Principal Coordinate Analysis** or **Metric Multidimensional Scaling**.

**Principal coordinates** are mediated through a distance function that has been computed among objects.

Improved performance if distances are **metric** (i.e., do not violate the “triangle inequality”)

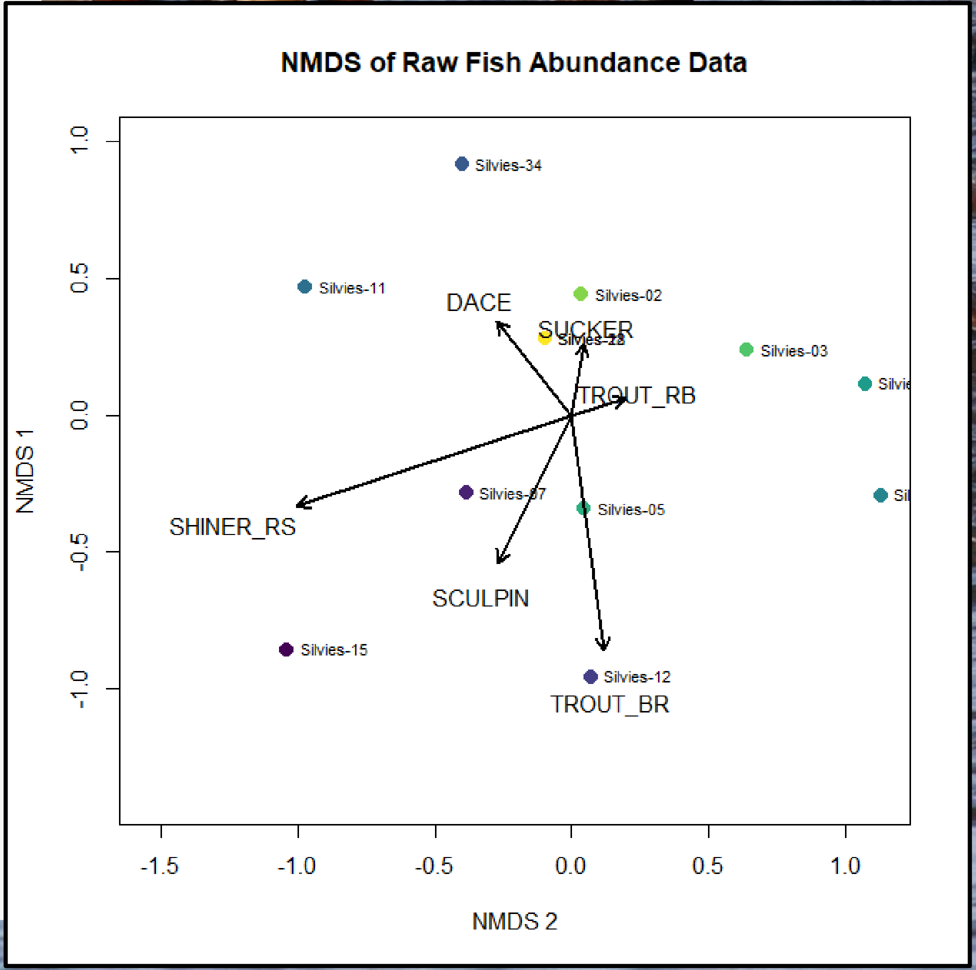


# Recap: Ordination

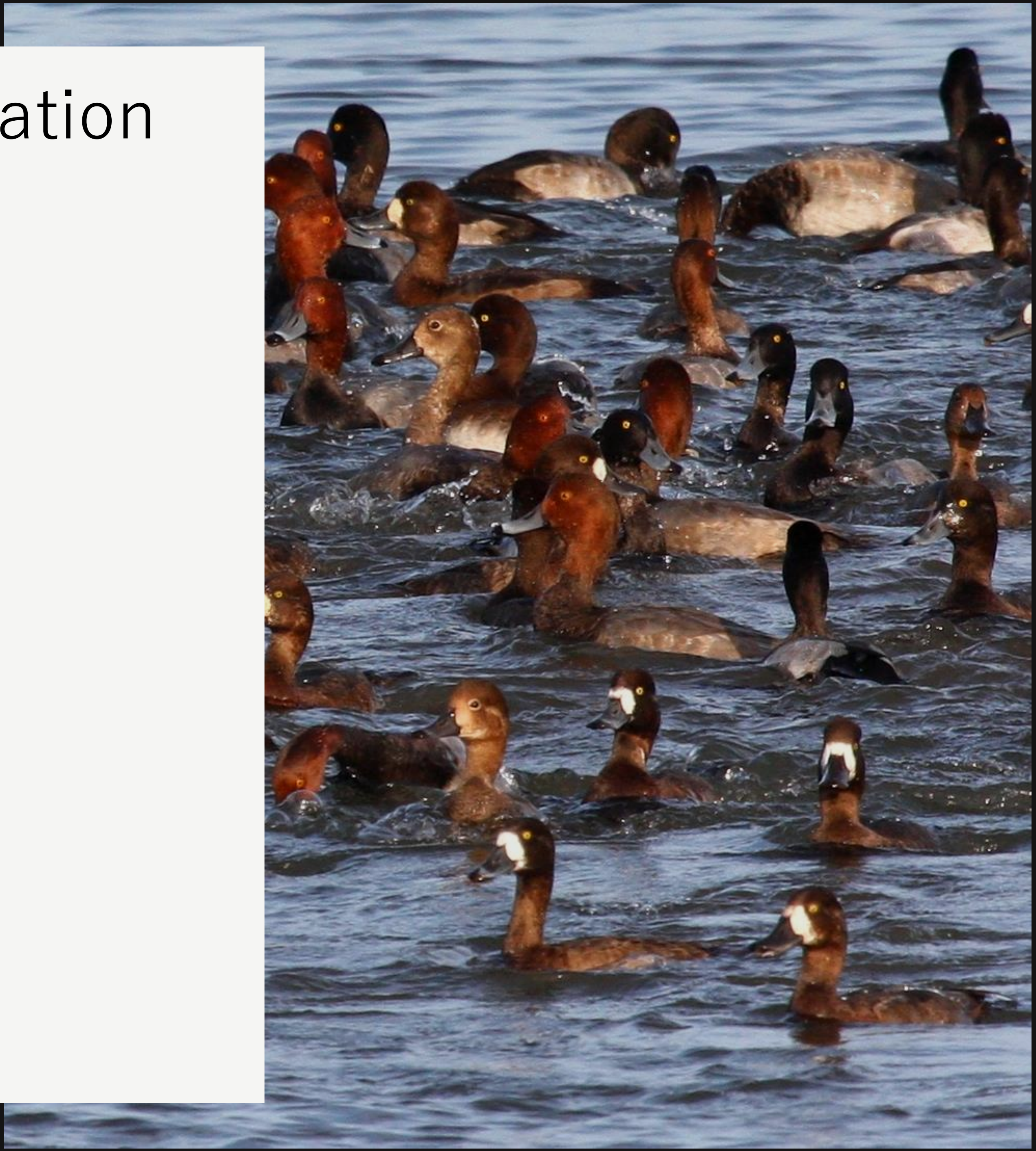
**Nonmetric Multidimensional Scaling** preserves rank order dissimilarities between objects in a low-dimensional space.

Useful for examining *relative*, rather than *absolute*, distances among data points.

NMDS axes *do not* maximize variability in space and are *arbitrary*!

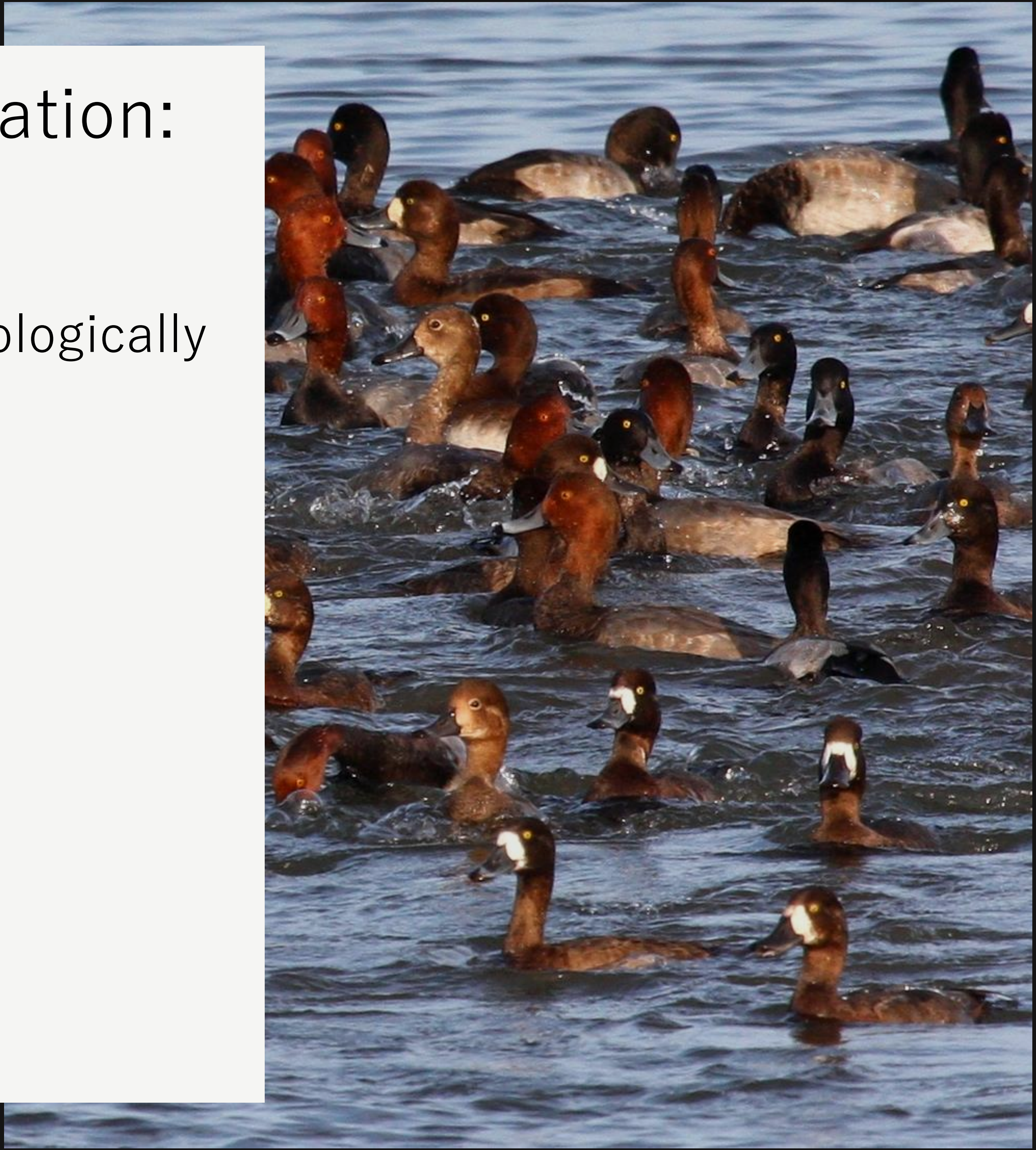


# Making Inferences from Ordination



# Making Inferences from Ordination: Objectives

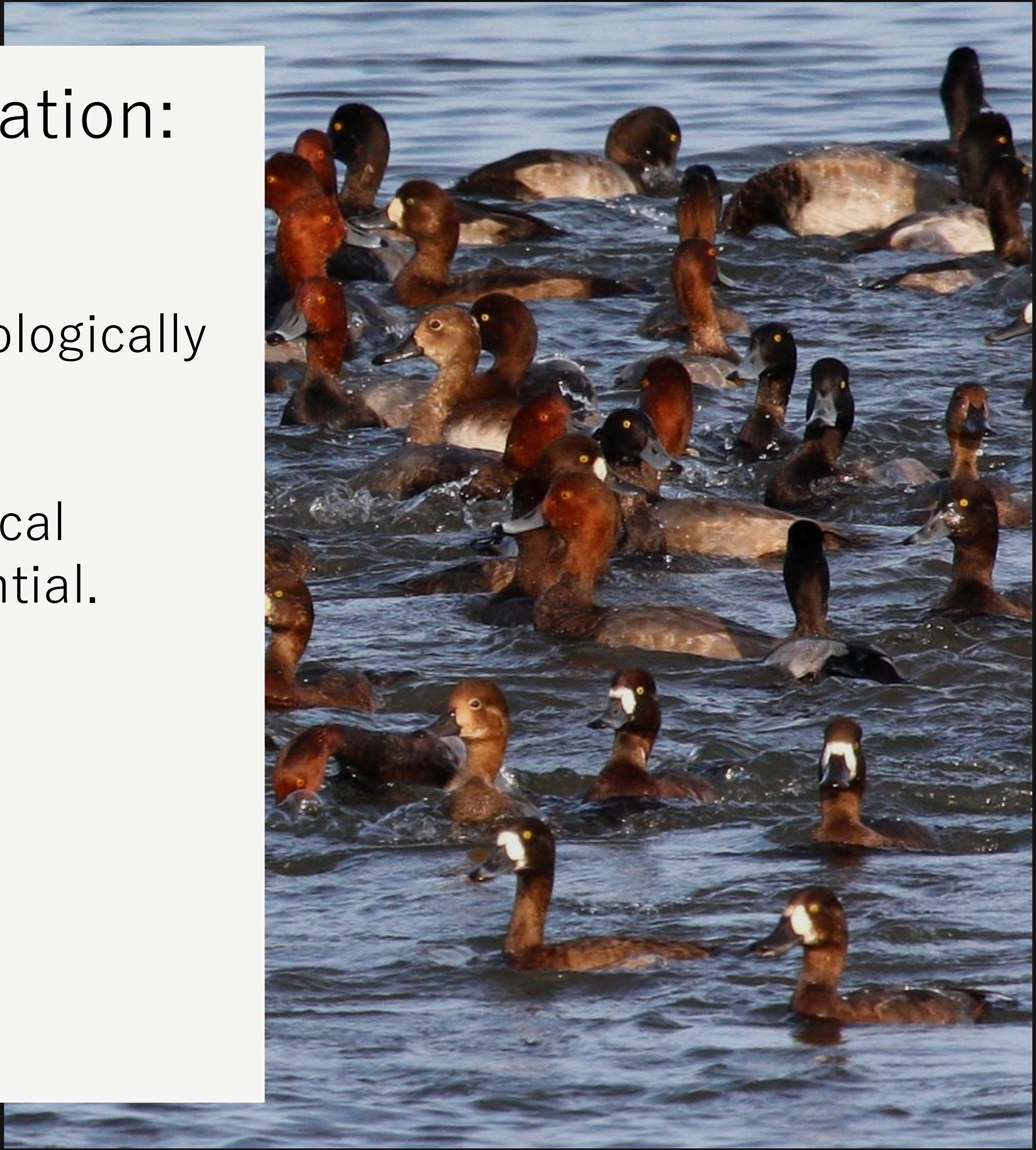
How do we translate our results into ecologically meaningful insights?



# Making Inferences from Ordination: Objectives

How do we translate our results into ecologically meaningful insights?

**Interpretation:** links patterns to ecological processes. Can be exploratory *or* inferential.

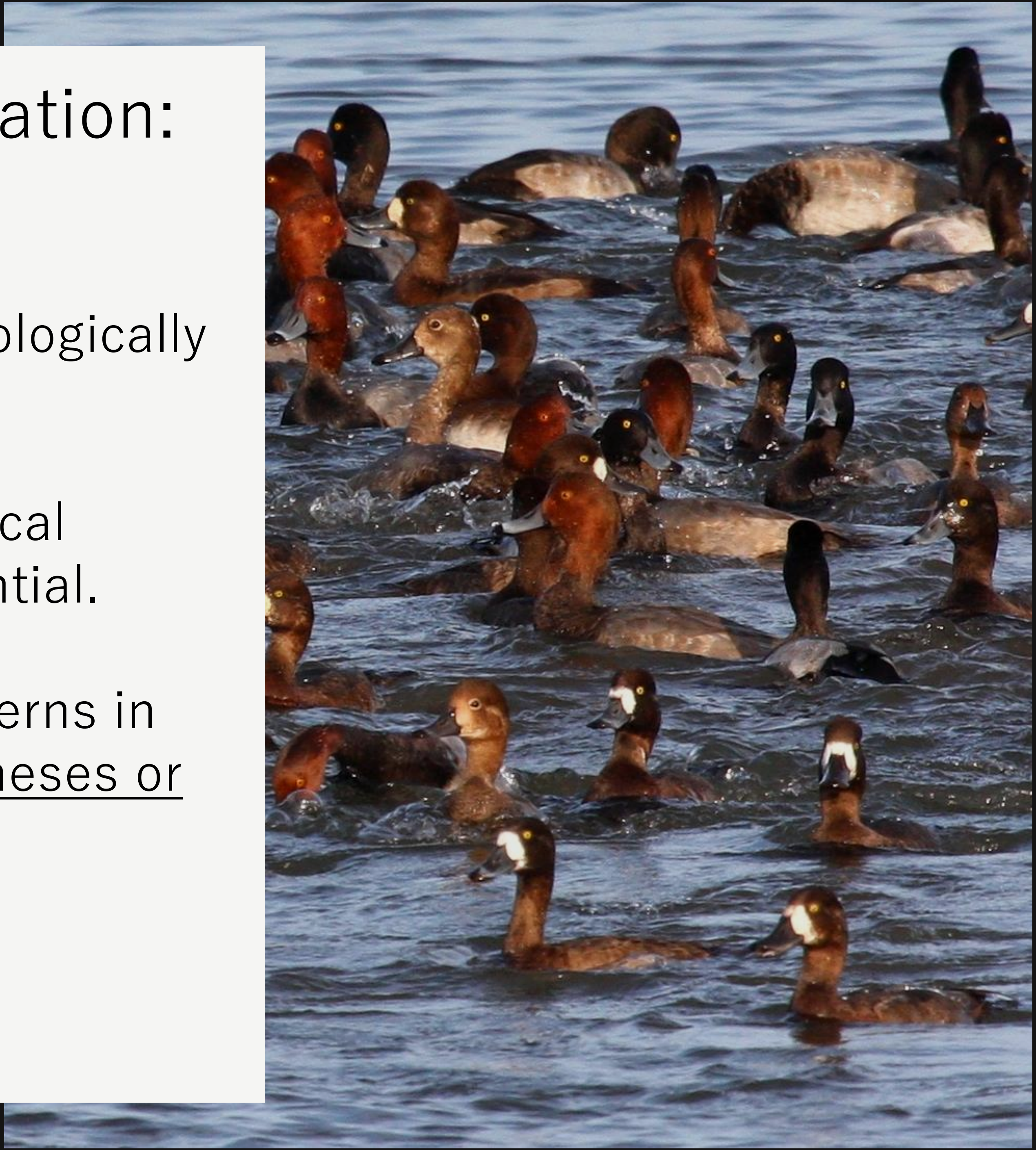


# Making Inferences from Ordination: Objectives

How do we translate our results into ecologically meaningful insights?

**Interpretation:** links patterns to ecological processes. Can be exploratory *or* inferential.

**Inference:** draws conclusions from patterns in complex datasets, usually <u>to test hypotheses or identify key explanatory variables</u>.

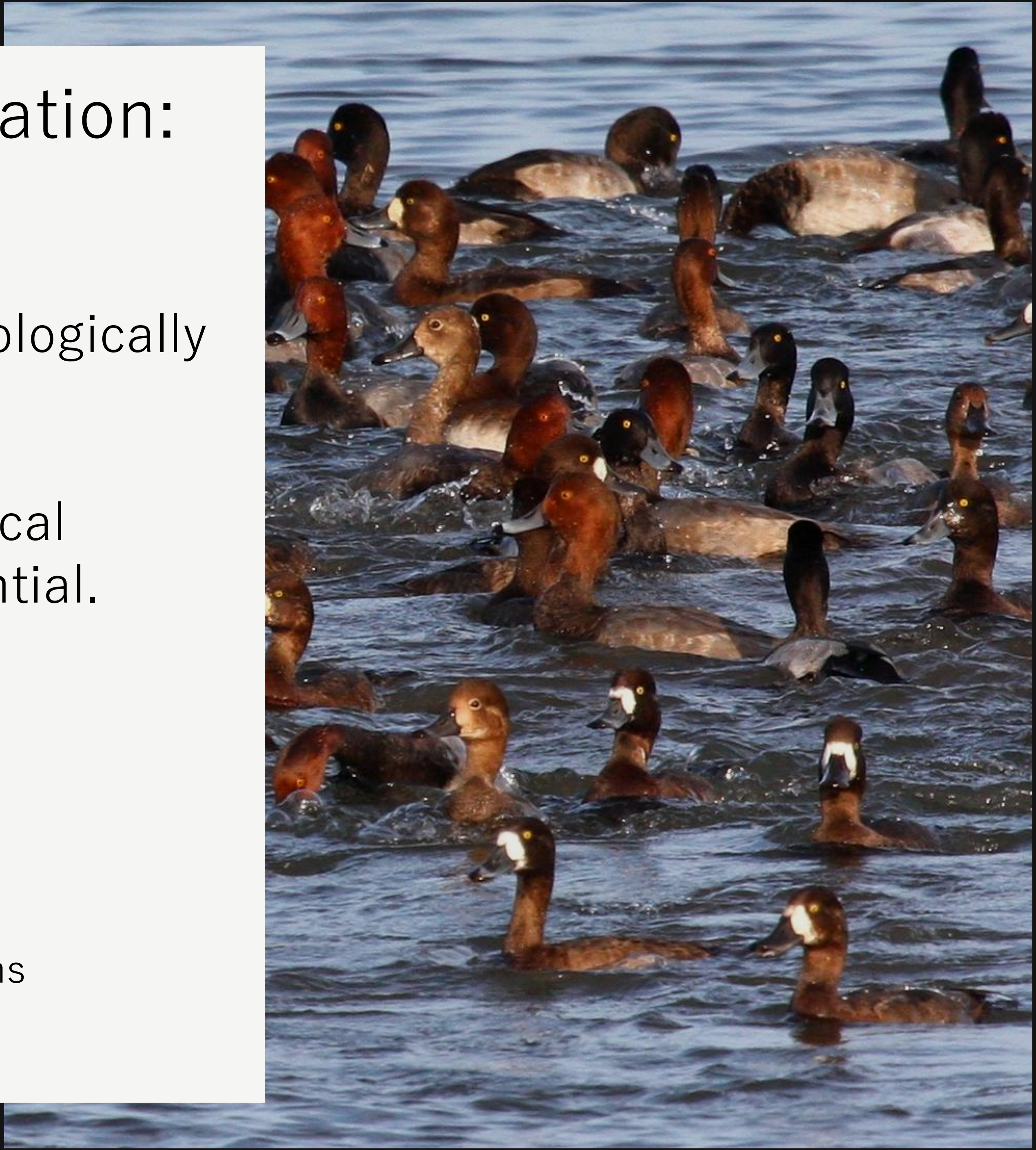


# Making Inferences from Ordination: Objectives

How do we translate our results into ecologically meaningful insights?

**Interpretation:** links patterns to ecological processes. Can be exploratory *or* inferential.

- Interpret principal axes and components
- Examine loadings and weights
- Identify clusters and groupings
- Visualize using biplots and ordination diagrams

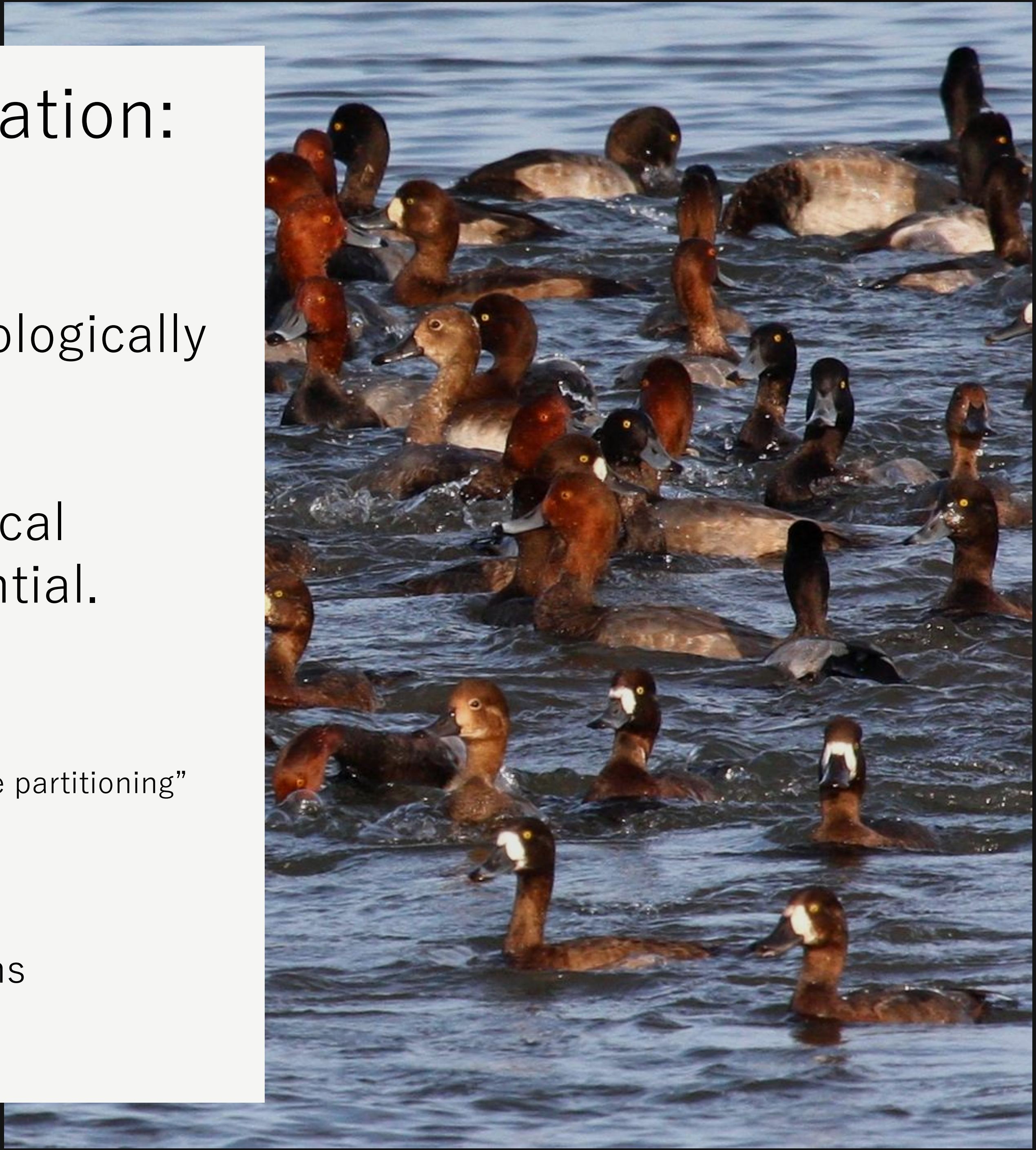


# Making Inferences from Ordination: Objectives

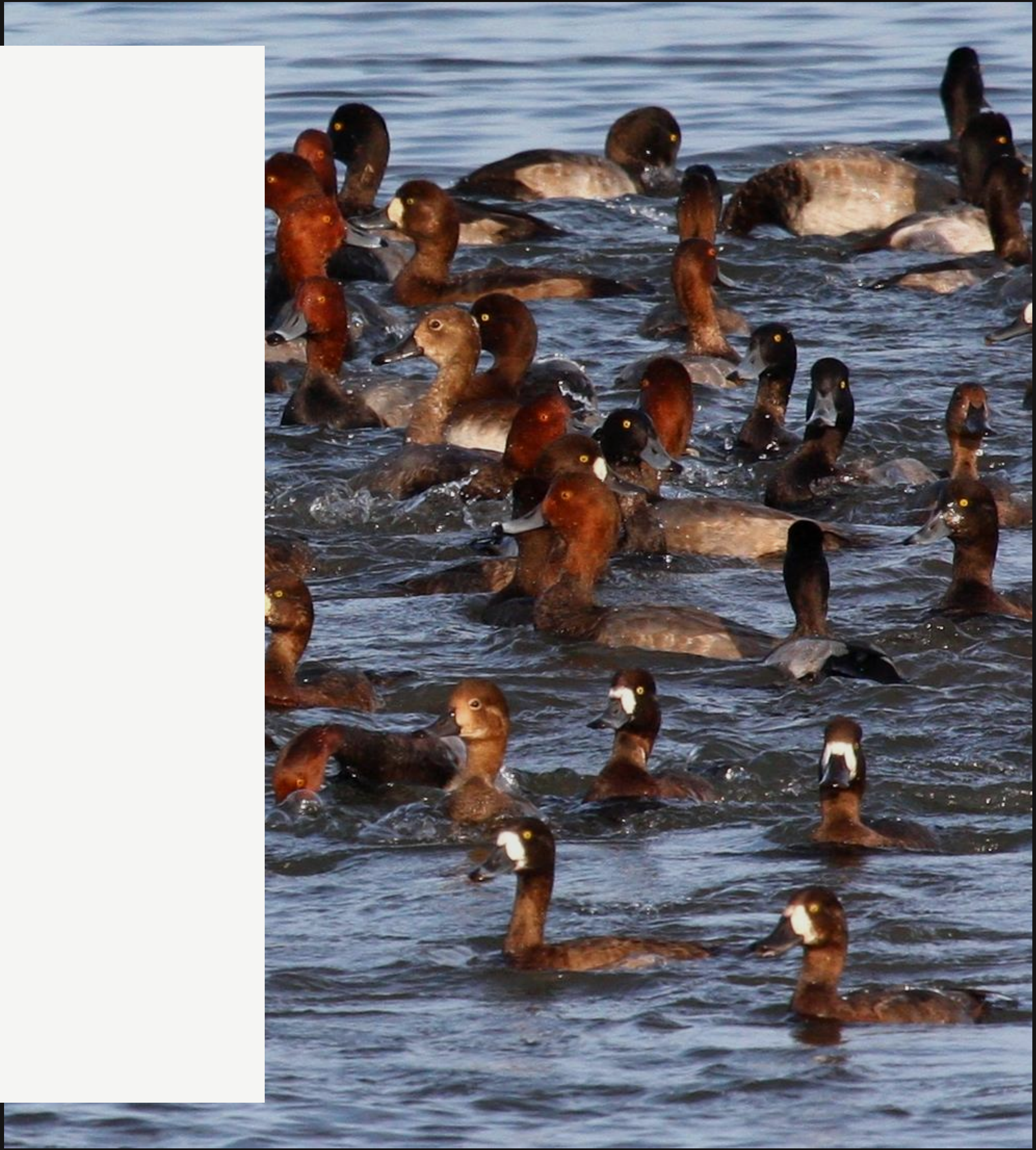
How do we translate our results into ecologically meaningful insights?

**Interpretation:** links patterns to ecological processes. Can be exploratory *or* inferential.

- Interpret principal axes and components
- Examine loadings and weights ← "Variance partitioning"
- Identify clusters and groupings
- Visualize using biplots and ordination diagrams

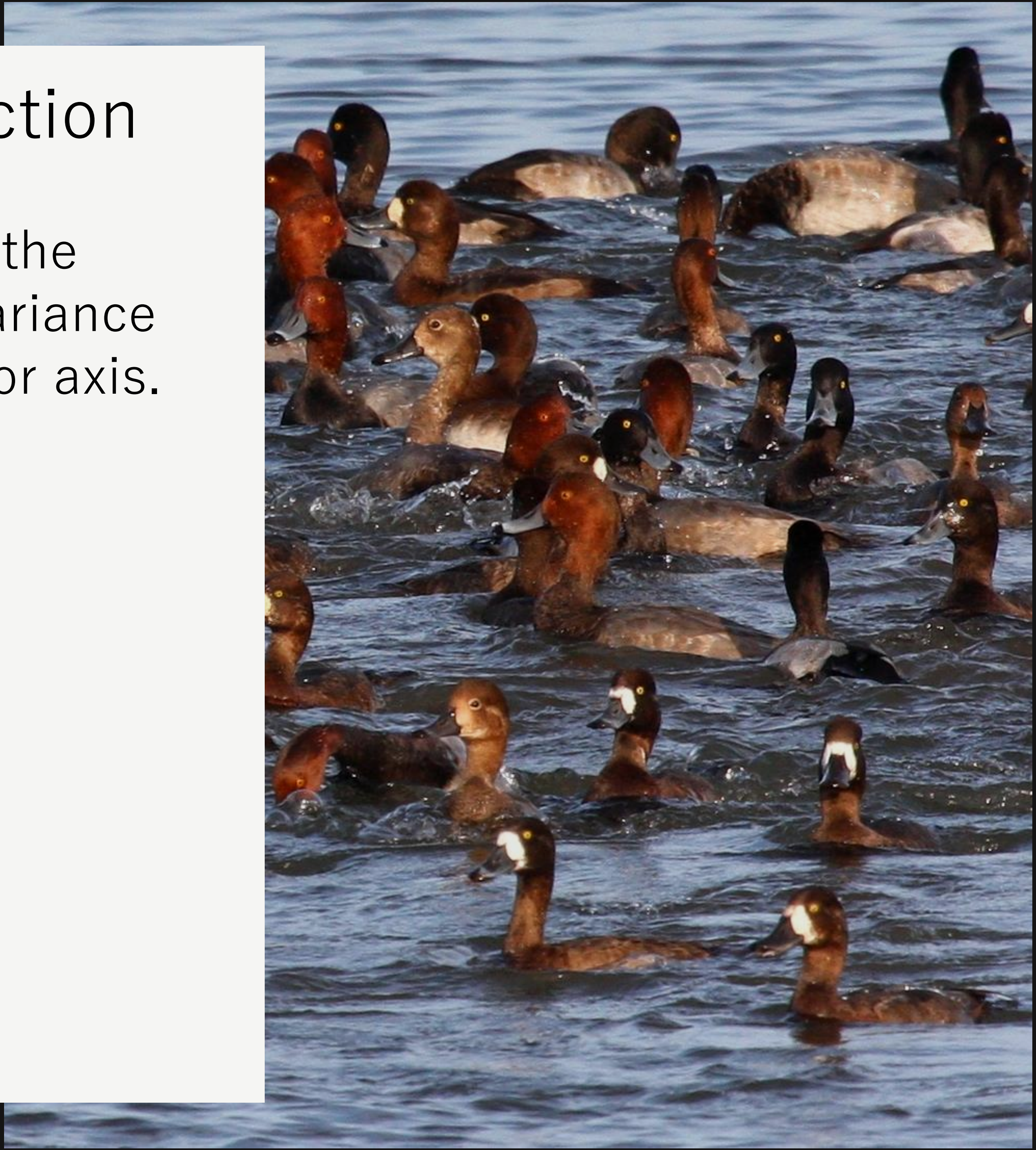


# Variance Partitioning



# Variance Partitioning: Introduction

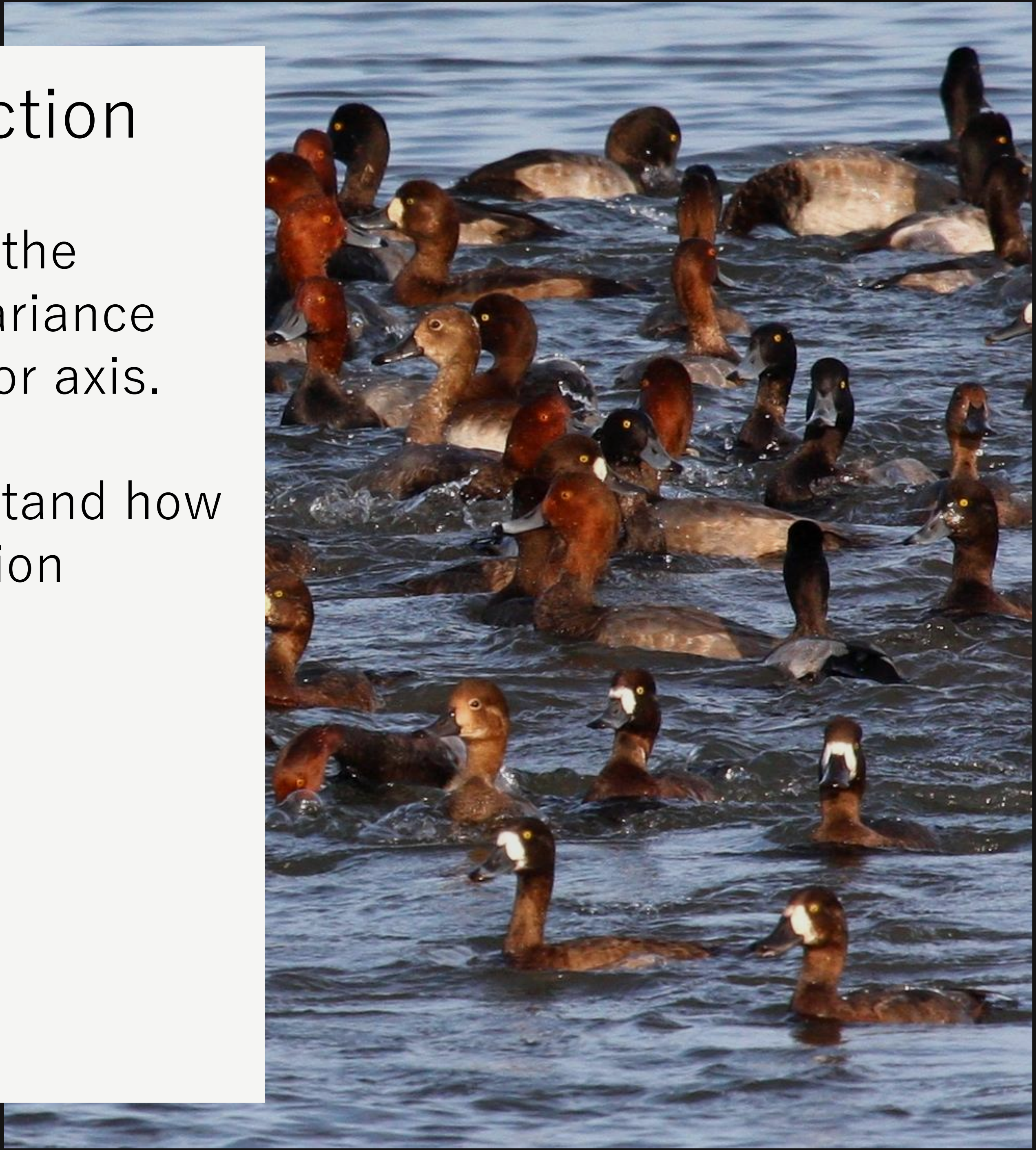
In ordination methods like PCA and CA, the **eigenvalues** represent the amount of variance explained by each principal component or axis.



# Variance Partitioning: Introduction

In ordination methods like PCA and CA, the **eigenvalues** represent the amount of variance explained by each principal component or axis.

**Variance partitioning** is used to understand how different factors contribute to the variation observed in a dataset.

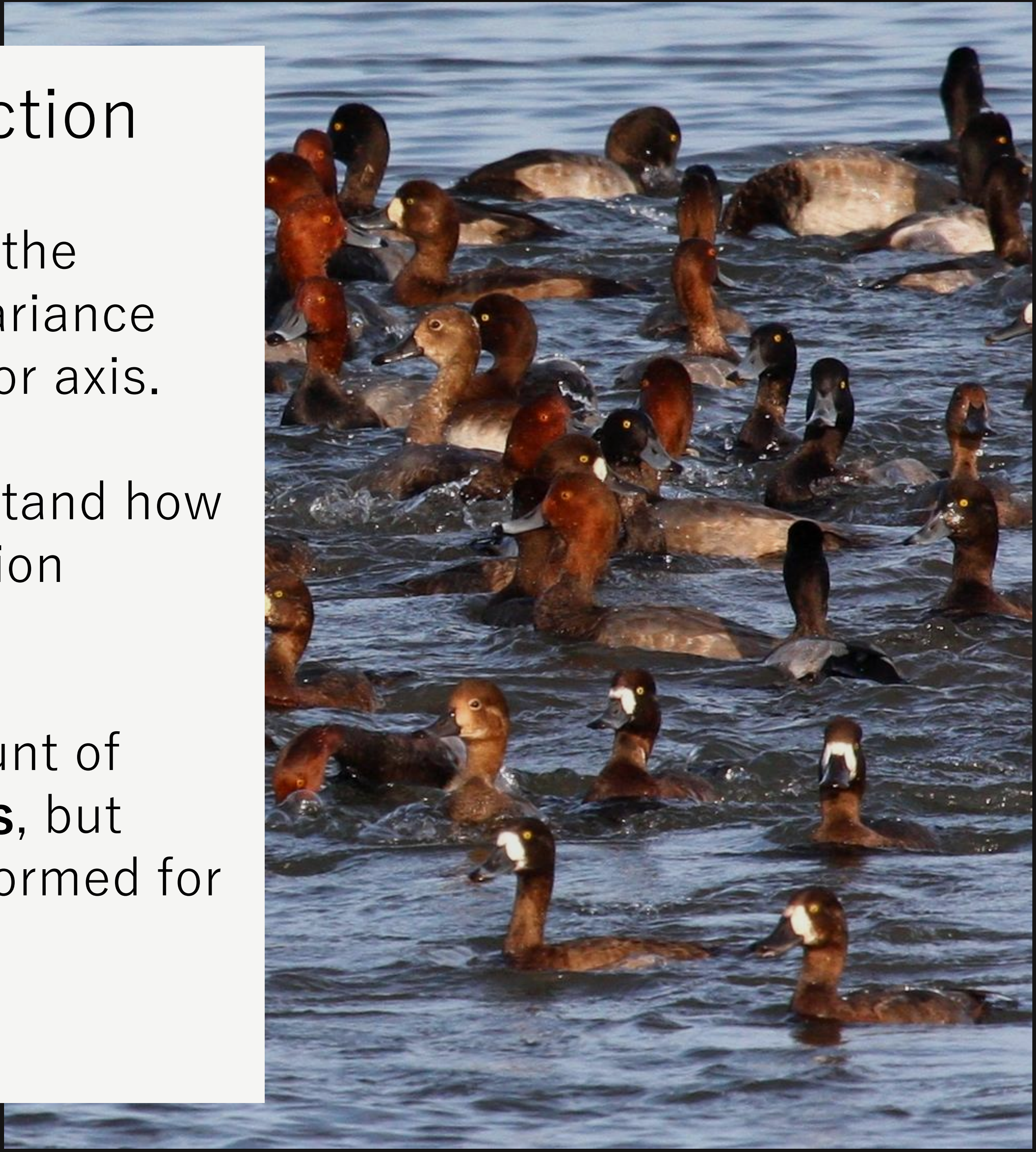


# Variance Partitioning: Introduction

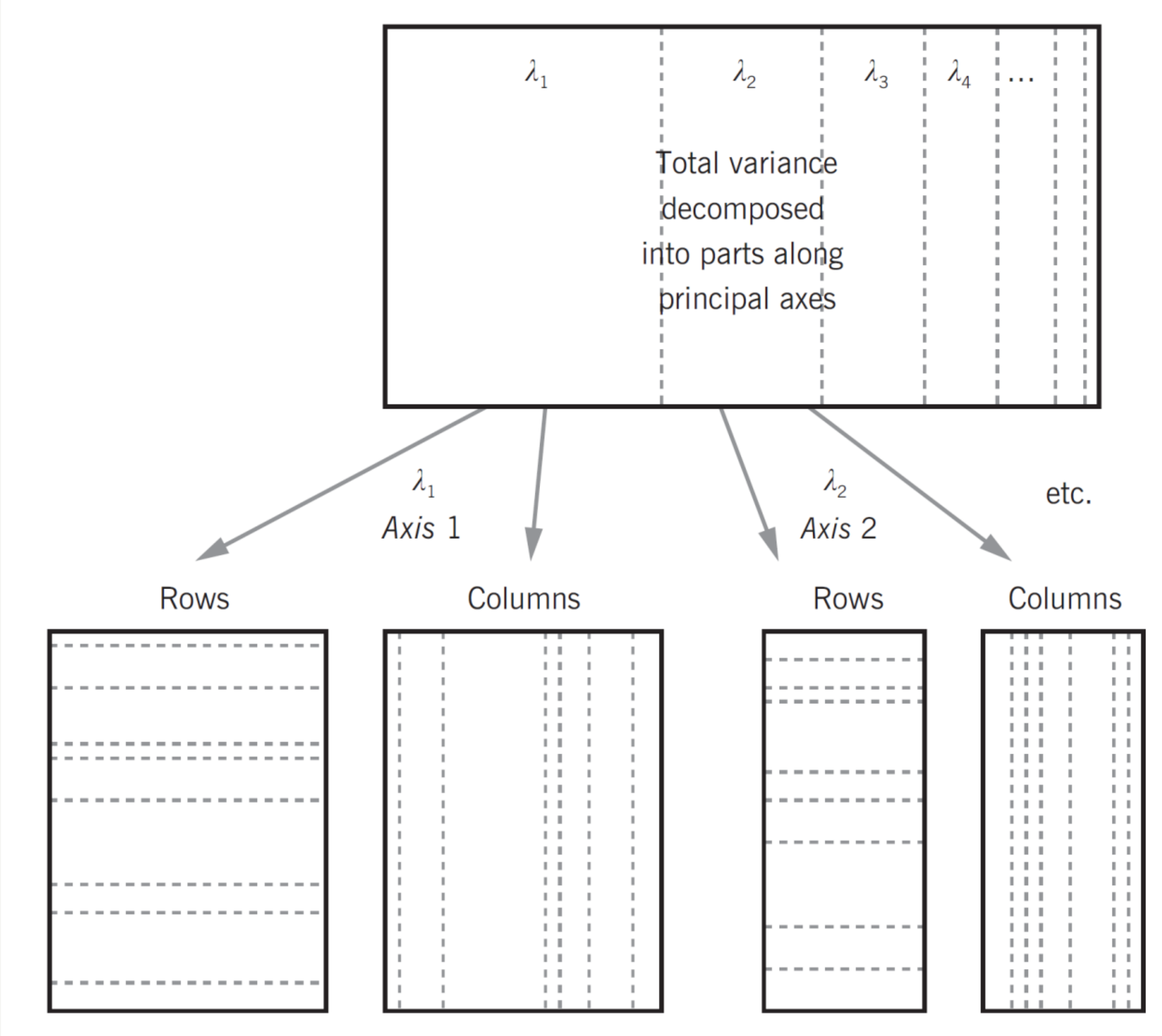
In ordination methods like PCA and CA, the **eigenvalues** represent the amount of variance explained by each principal component or axis.

**Variance partitioning** is used to understand how different factors contribute to the variation observed in a dataset.

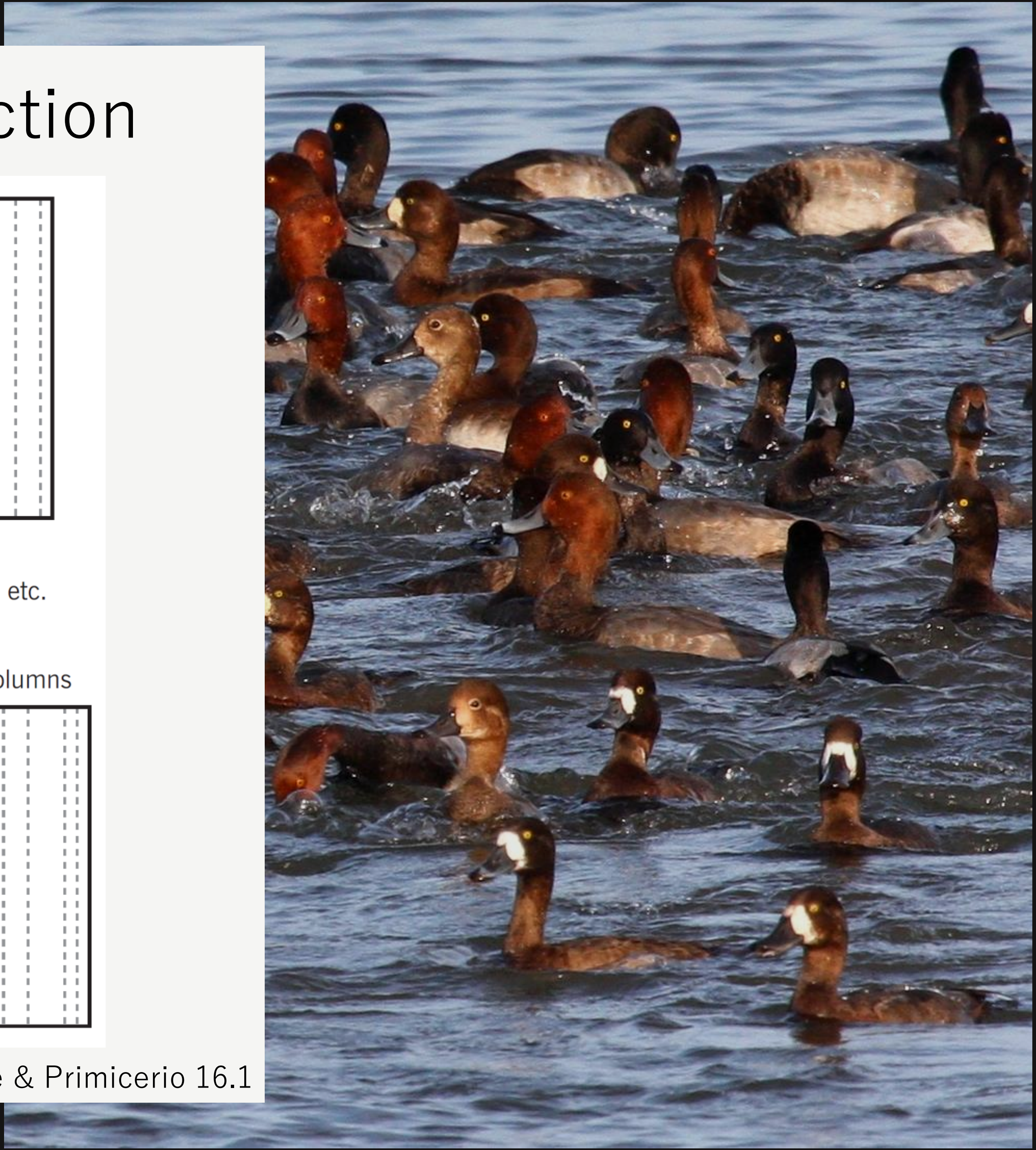
- We are usually interested in the amount of variation explained by the **descriptors**, but variance partitioning can also be performed for objects and individual observations.



# Variance Partitioning: Introduction

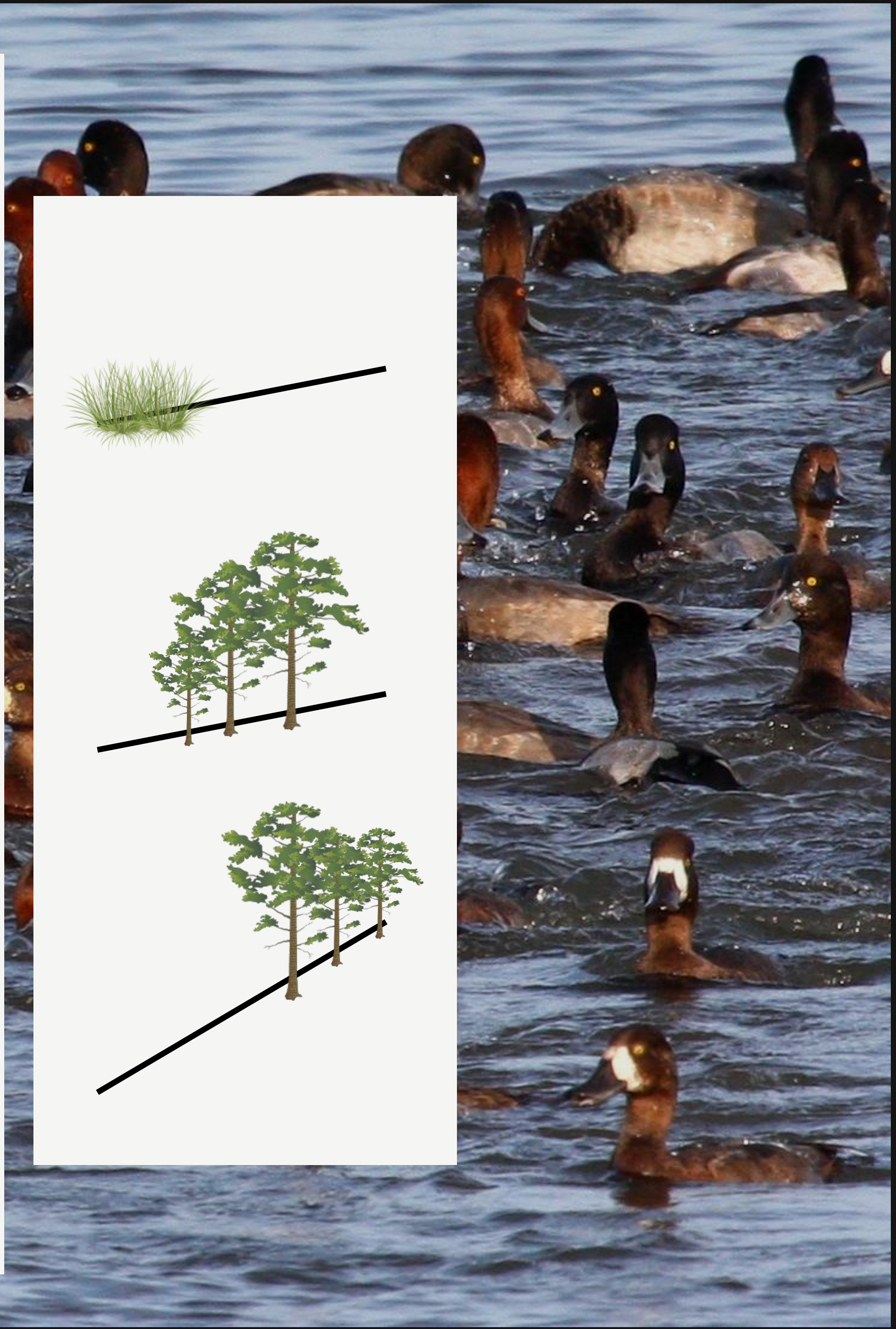



Greenacre & Primicerio 16.1



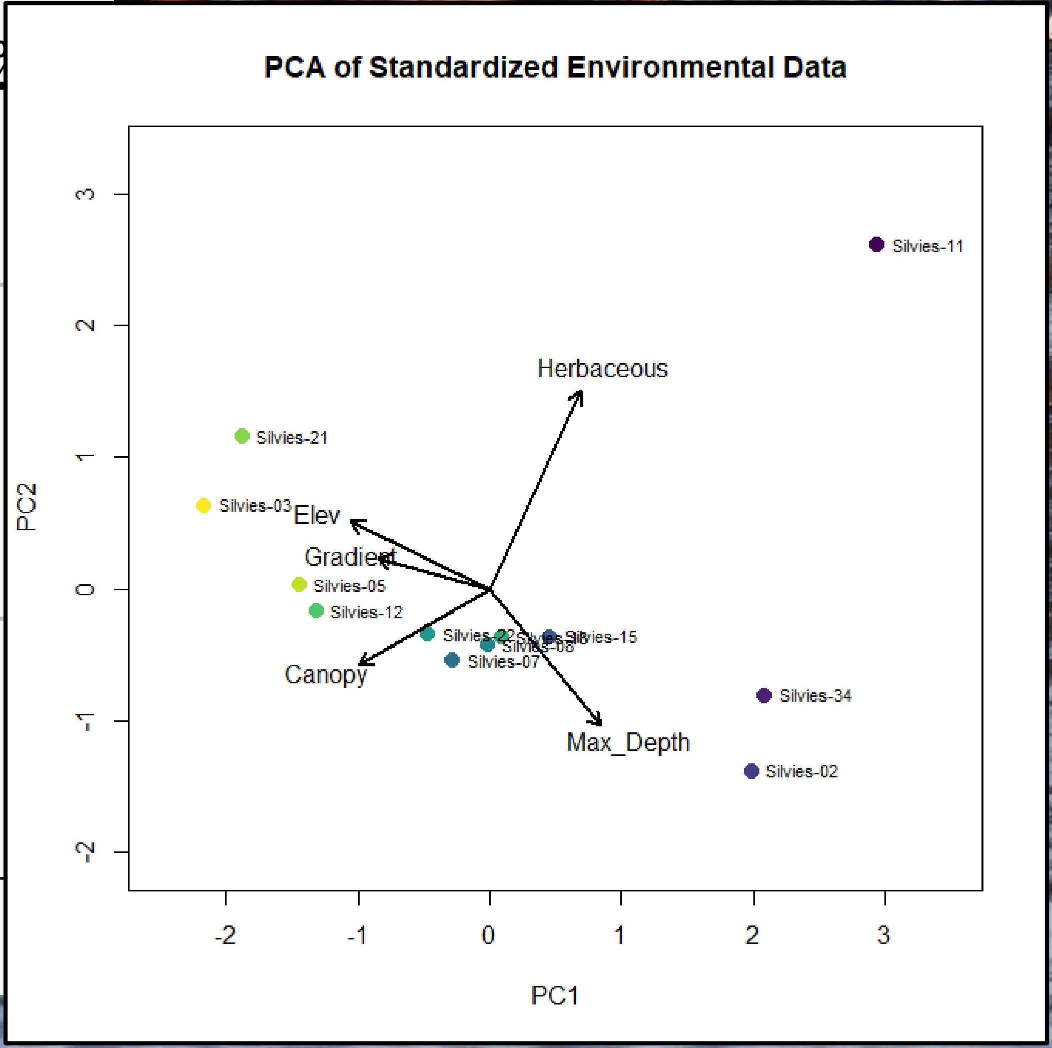
# Variance Partitioning: Diagnostics

| Site ID | Max Depth (m) | Gradient (%) | Elevation (m) | Canopy (%) | Herb (%) |
|---|---|---|---|---|---|
| Silvies-11 | 0.45 | 0.3 | 1439 | 0.0 | 55.1 |
| Silvies-34 | 0.78 | 1.1 | 1487 | 0.0 | 0.0 |
| Silvies-02 | 0.71 | 0.4 | 1372 | 29.6 | 0.0 |
| Silvies-15 | 0.40 | 0.2 | 1471 | 41.1 | 0.0 |
| Silvies-07 | 0.50 | 1.3 | 1547 | 52.3 | 0.0 |
| Silvies-08 | 0.40 | 0.6 | 1492 | 51.4 | 0.0 |
| Silvies-22 | 0.42 | 0.9 | 1555 | 54.7 | 0.0 |
| Silvies-18 | 0.42 | 0.5 | 1510 | 46.2 | 0.0 |
| Silvies-12 | 0.52 | 3.2 | 1658 | 51.9 | 0.0 |
| Silvies-21 | 0.18 | 2.4 | 1713 | 37.5 | 0.0 |
| Silvies-05 | 0.45 | 5.5 | 1565 | 46.7 | 0.0 |
| Silvies-03 | 0.20 | 3.3 | 1634 | 59.0 | 0.0 |



# Variance Partitioning: Diagnostics

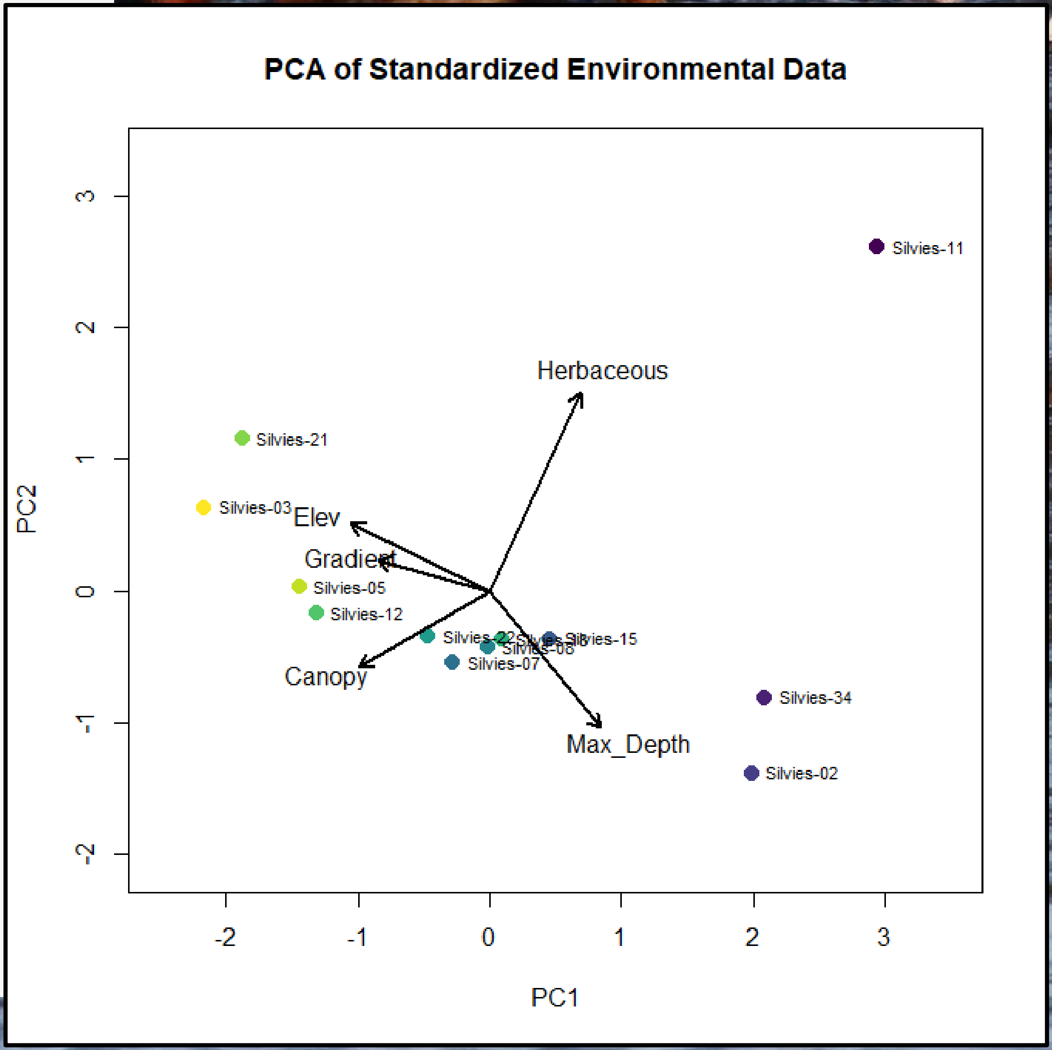
| Site ID | Max Depth (m) | Gradient (%) | Elevation (m) | Canopy (%) | Herb ( |
|---|---|---|---|---|---|
| Silvies-11 | 0.45 | 0.3 | 1439 | 0.0 | 55.1 |
| Silvies-34 | 0.78 | 1.1 | 1487 | 0.0 | 0.0 |
| Silvies-02 | 0.71 | 0.4 | 1372 | 29.6 | 0.0 |
| Silvies-15 | 0.40 | 0.2 | 1471 | 41.1 | 0.0 |
| Silvies-07 | 0.50 | 1.3 | 1547 | 52.3 | 0.0 |
| Silvies-08 | 0.40 | 0.6 | 1492 | 51.4 | 0.0 |
| Silvies-22 | 0.42 | 0.9 | 1555 | 54.7 | 0.0 |
| Silvies-18 | 0.42 | 0.5 | 1510 | 46.2 | 0.0 |
| Silvies-12 | 0.52 | 3.2 | 1658 | 51.9 | 0.0 |
| Silvies-21 | 0.18 | 2.4 | 1713 | 37.5 | 0.0 |
| Silvies-05 | 0.45 | 5.5 | 1565 | 46.7 | 0.0 |
| Silvies-03 | 0.20 | 3.3 | 1634 | 59.0 | 0.0 |



# Variance Partitioning: Diagnostics

$\lambda_1 = 2.69$ $\lambda_2 = 1.09$ $\lambda_3 = 0.78$

$\lambda_4 = 0.31$ $\lambda_5 = 0.12$

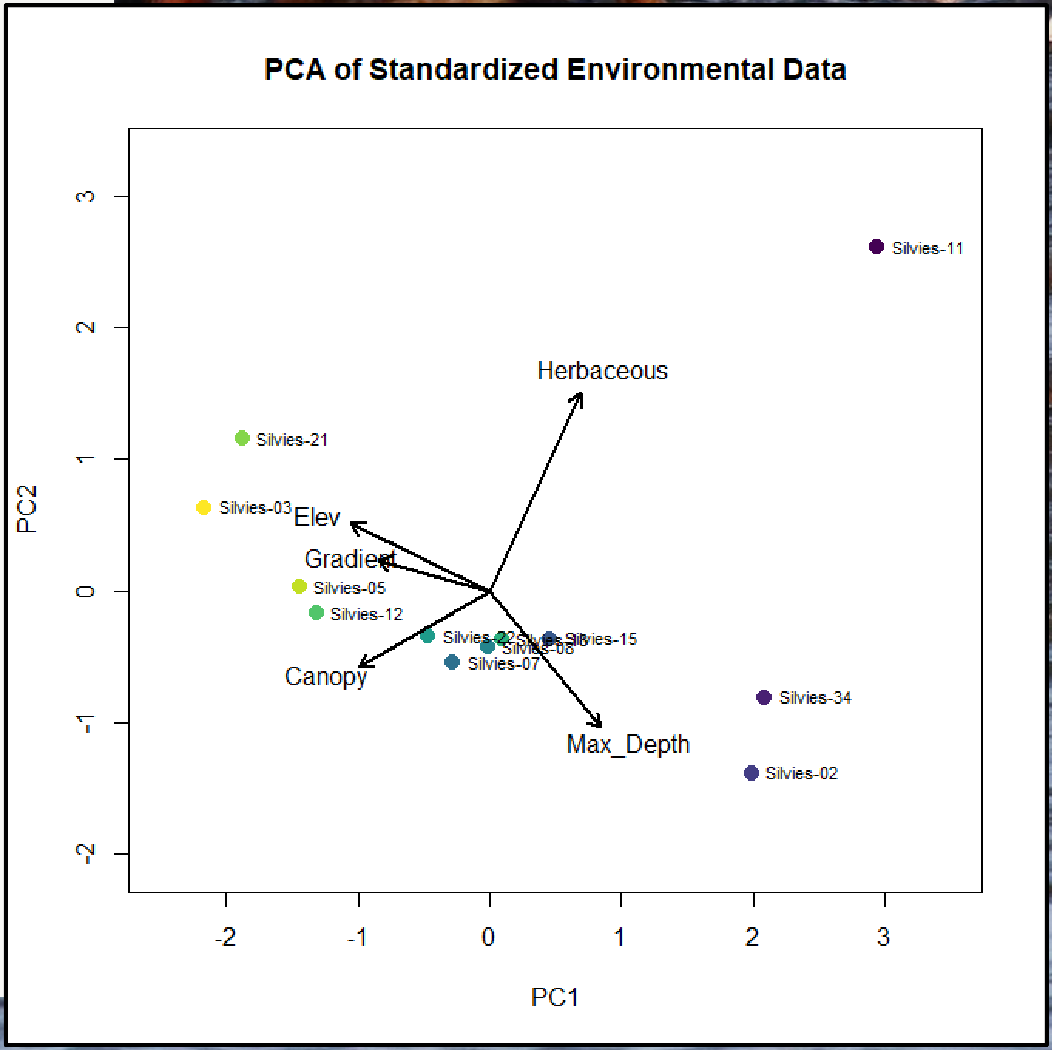


# Variance Partitioning: Diagnostics

$\lambda_1 = 2.69$ $\lambda_2 = 1.09$ $\lambda_3 = 0.78$

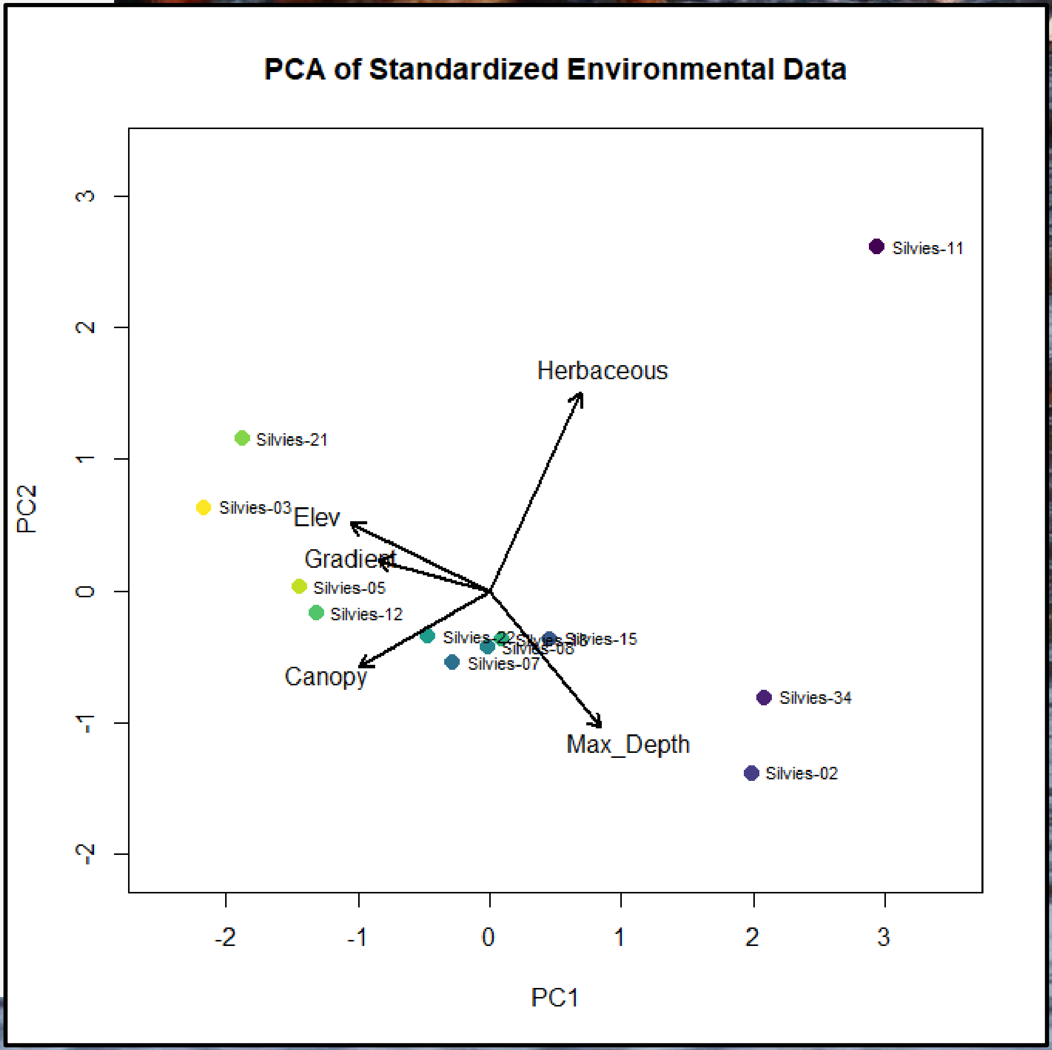
$\lambda_4 = 0.31$ $\lambda_5 = 0.12$

What is the contribution of each environmental variable to the principal axes?



# Variance Partitioning: Diagnostics

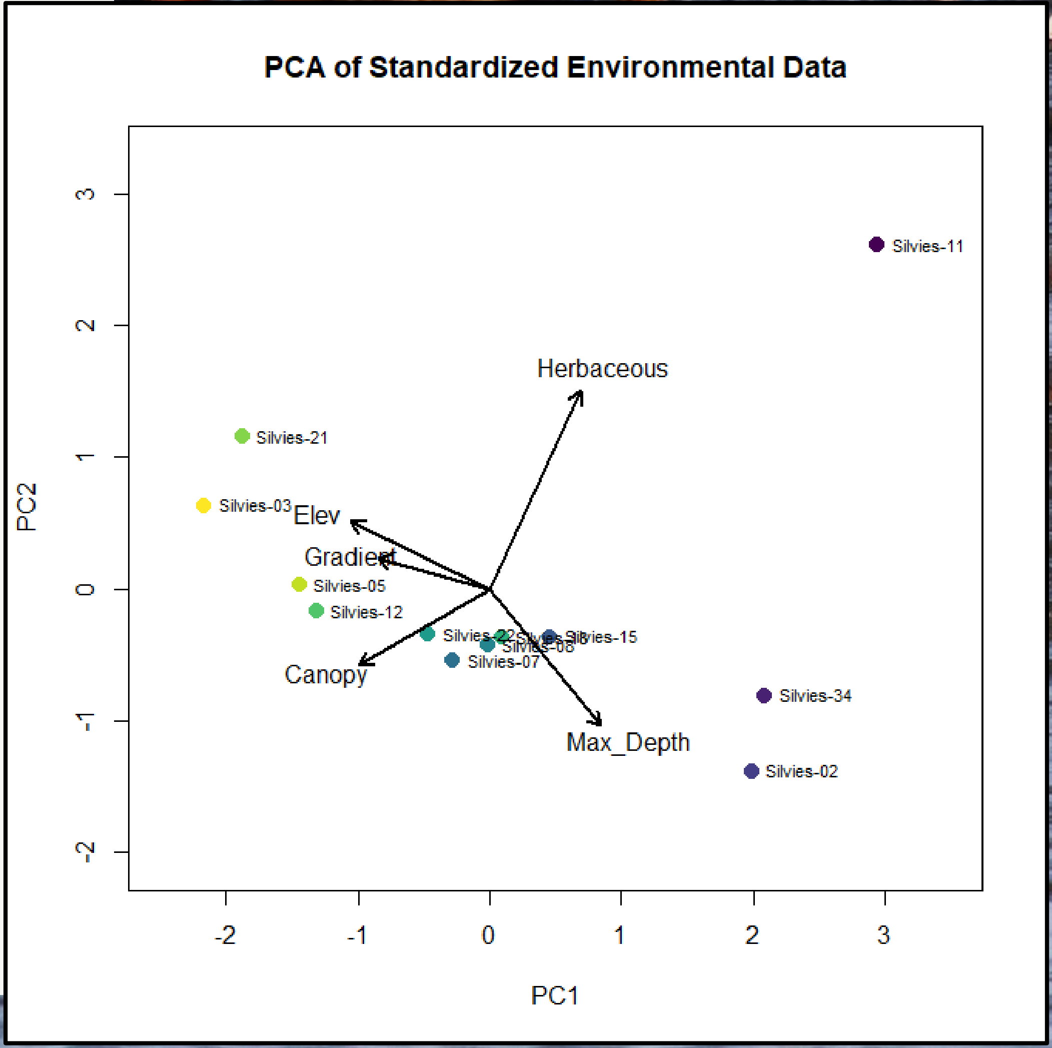
| Site ID | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Max Depth | 0.48 | 0.29 | 0.19 | 0.00 | 0.04 |
| Gradient | 0.48 | 0.01 | 0.43 | 0.06 | 0.01 |
| Elevation | 0.75 | 0.07 | 0.03 | 0.12 | 0.03 |
| Canopy | 0.66 | 0.09 | 0.13 | 0.09 | 0.03 |
| Herbaceous | 0.32 | 0.62 | 0.00 | 0.04 | 0.02 |
| Sum | 2.69 | 1.09 | 0.78 | 0.31 | 0.12 |



# Variance Partitioning: Diagnostics

| Site ID | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Max Depth | 0.18 | 0.27 | 0.25 | 0.00 | 0.31 |
| Gradient | 0.18 | 0.01 | 0.55 | 0.20 | 0.05 |
| Elevation | 0.28 | 0.07 | 0.03 | 0.40 | 0.22 |
| Canopy | 0.25 | 0.08 | 0.17 | 0.17 | 0.23 |
| Herbaceous | 0.12 | 0.57 | 0.00 | 0.12 | 0.20 |
| Sum | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

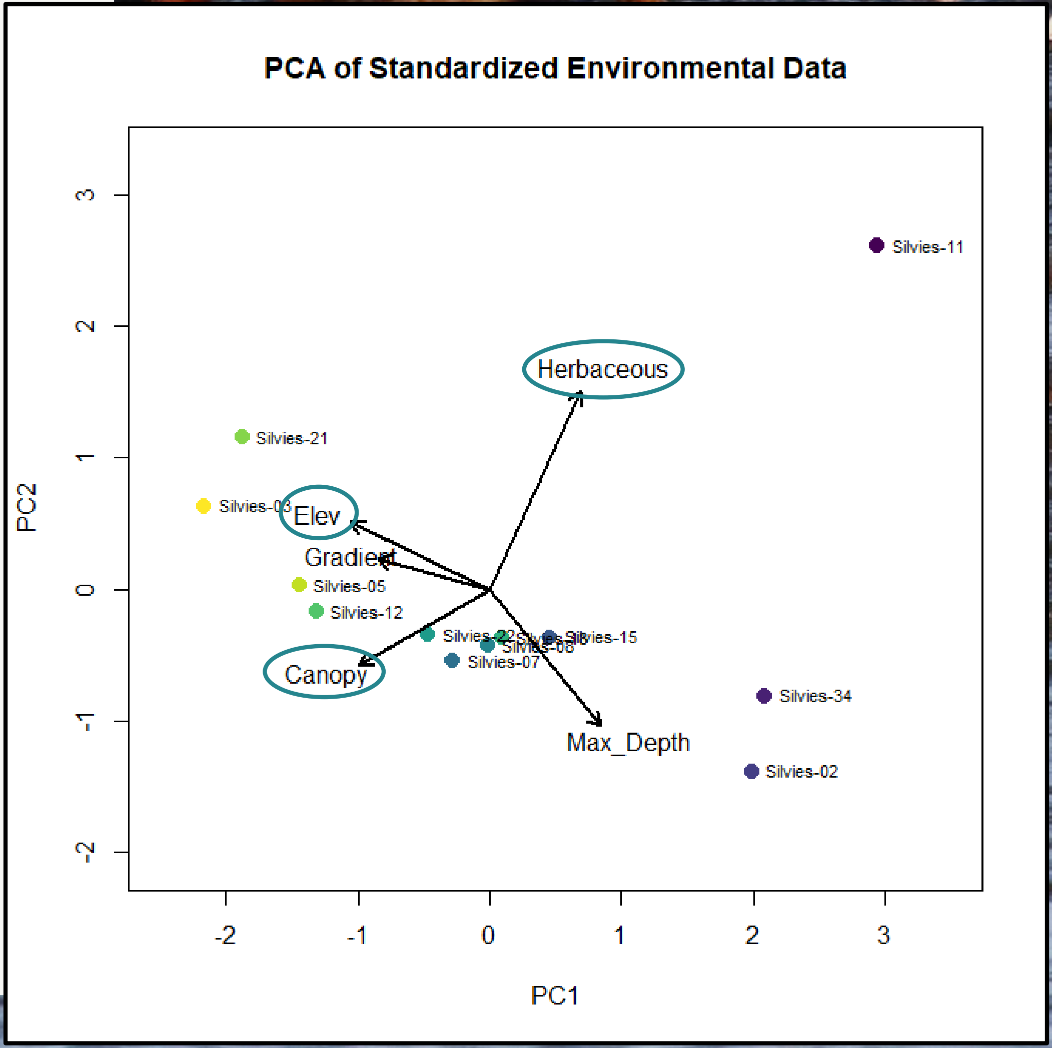
This is equivalent to the squared **loadings** (the eigenvectors)! Thus, the squared loadings explain each descriptor's relative contribution to the variance.



# Variance Partitioning: Diagnostics

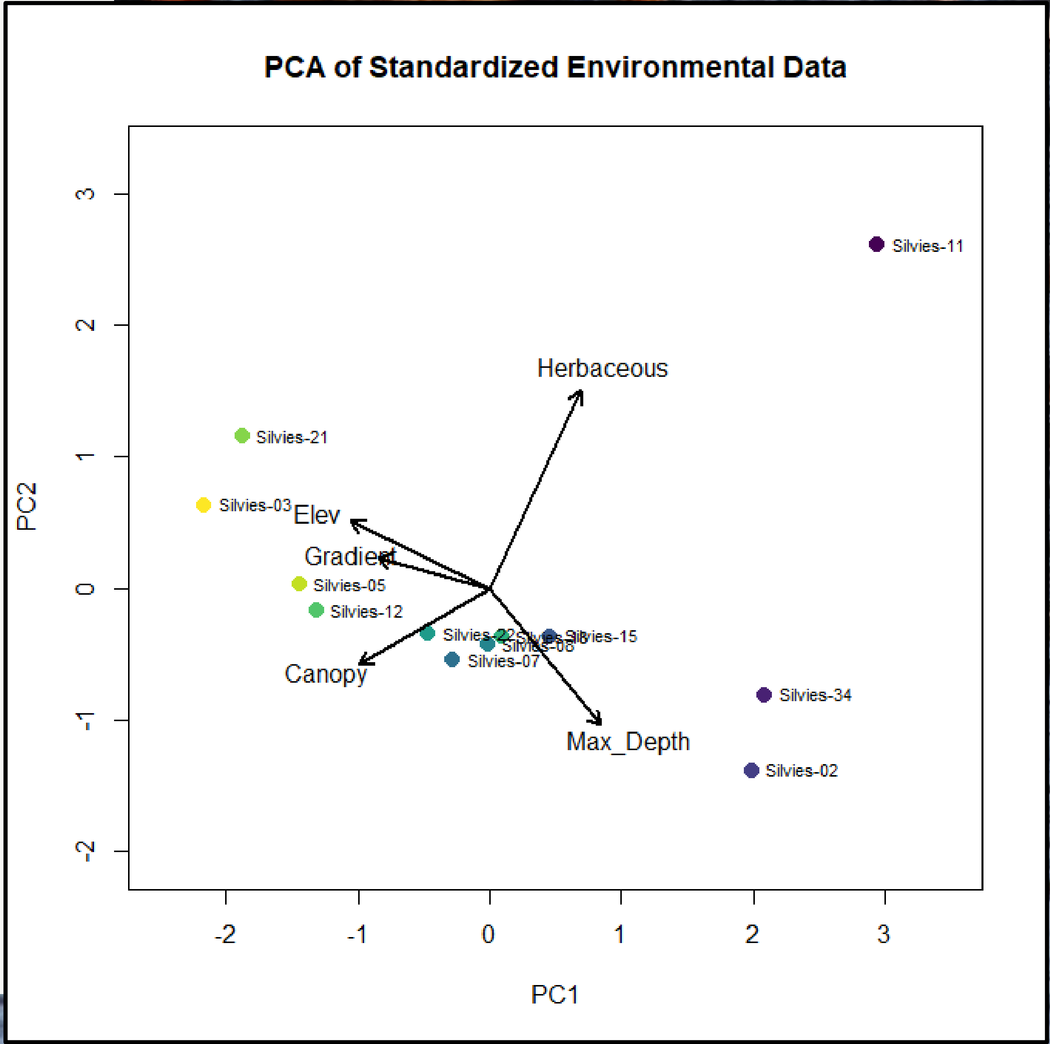
| Site ID | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Max Depth | 0.18 | 0.27 | 0.25 | 0.00 | 0.31 |
| Gradient | 0.18 | 0.01 | 0.55 | 0.20 | 0.05 |
| Elevation | 0.28 | 0.07 | 0.03 | 0.40 | 0.22 |
| Canopy | 0.25 | 0.08 | 0.17 | 0.17 | 0.23 |
| Herbaceous | 0.12 | 0.57 | 0.00 | 0.12 | 0.20 |
| Sum | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

This is equivalent to the squared **loadings** (the eigenvectors)! Thus, the squared loadings explain each descriptor's relative contribution to the variance.



# Variance Partitioning: Diagnostics

| Site ID | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Max Depth | 0.48 | 0.29 | 0.19 | 0.00 | 0.04 |
| Gradient | 0.48 | 0.01 | 0.43 | 0.06 | 0.01 |
| Elevation | 0.75 | 0.07 | 0.03 | 0.12 | 0.03 |
| Canopy | 0.66 | 0.09 | 0.13 | 0.09 | 0.03 |
| Herbaceous | 0.32 | 0.62 | 0.00 | 0.04 | 0.02 |
| Sum | 2.69 | 1.09 | 0.78 | 0.31 | 0.12 |



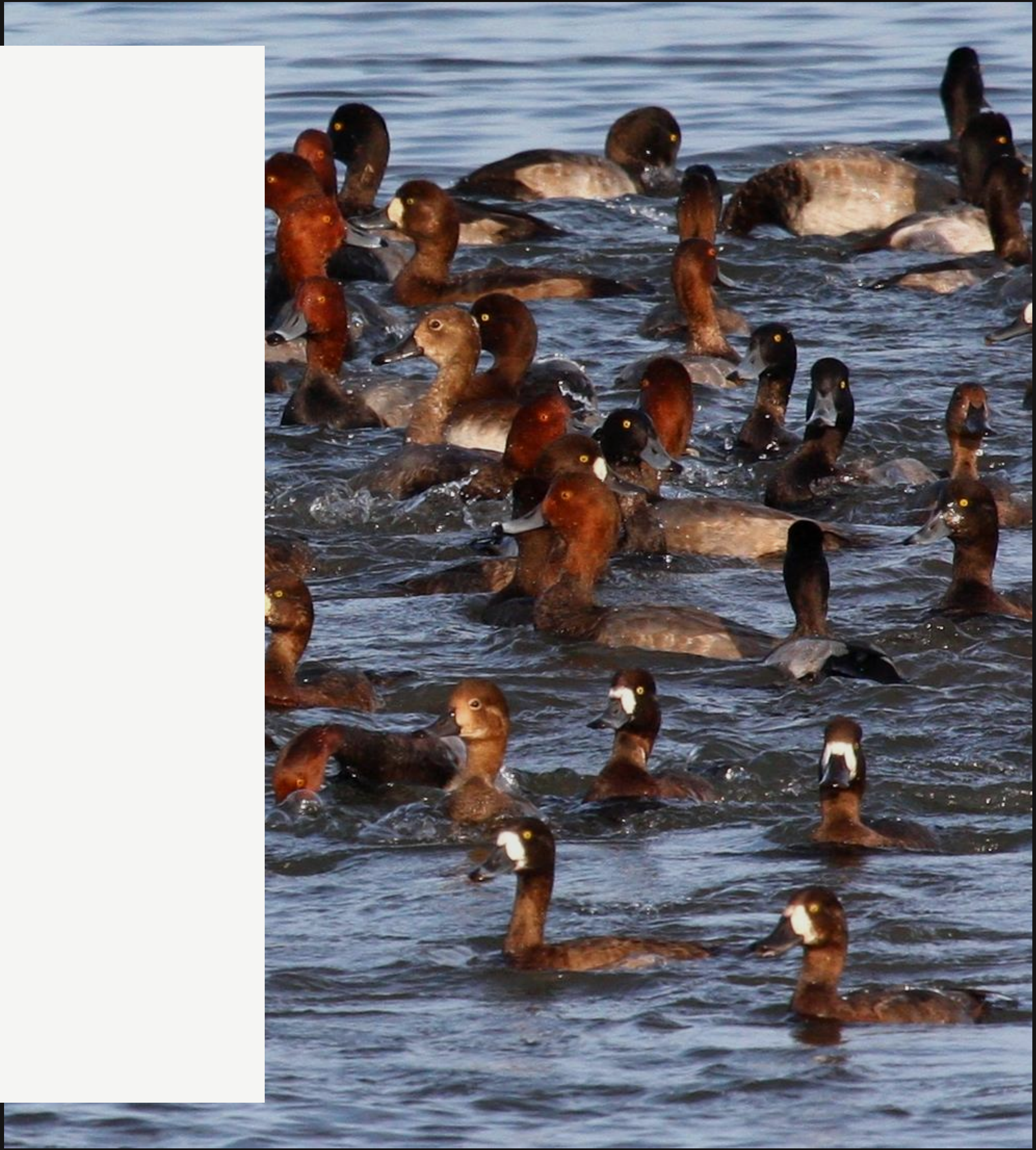
# Making Inferences from Ordination: Objectives

How do we translate our results into ecologically meaningful insights?

**Interpretation:** links patterns to ecological processes. Can be exploratory *or* inferential.

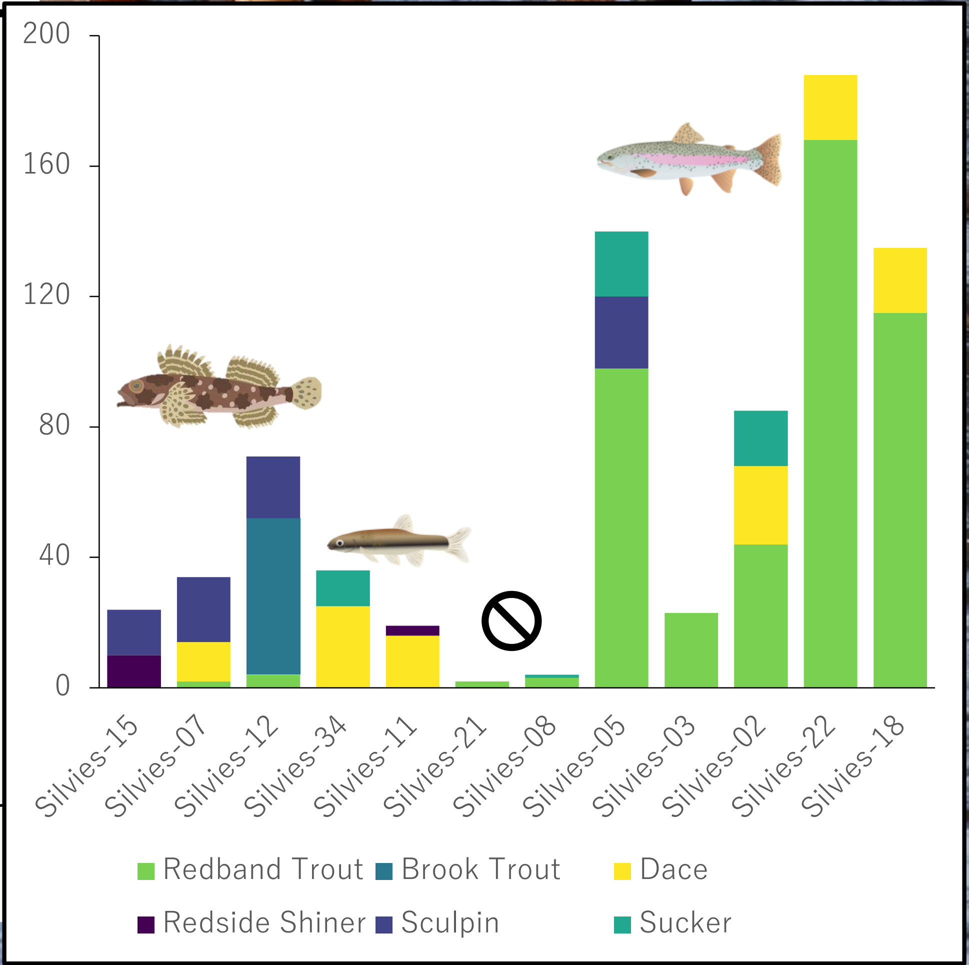
- Interpret principal axes and components
- Examine loadings and weights
- Identify clusters and groupings
- Visualize using biplots and ordination diagrams

# Effective Biplots



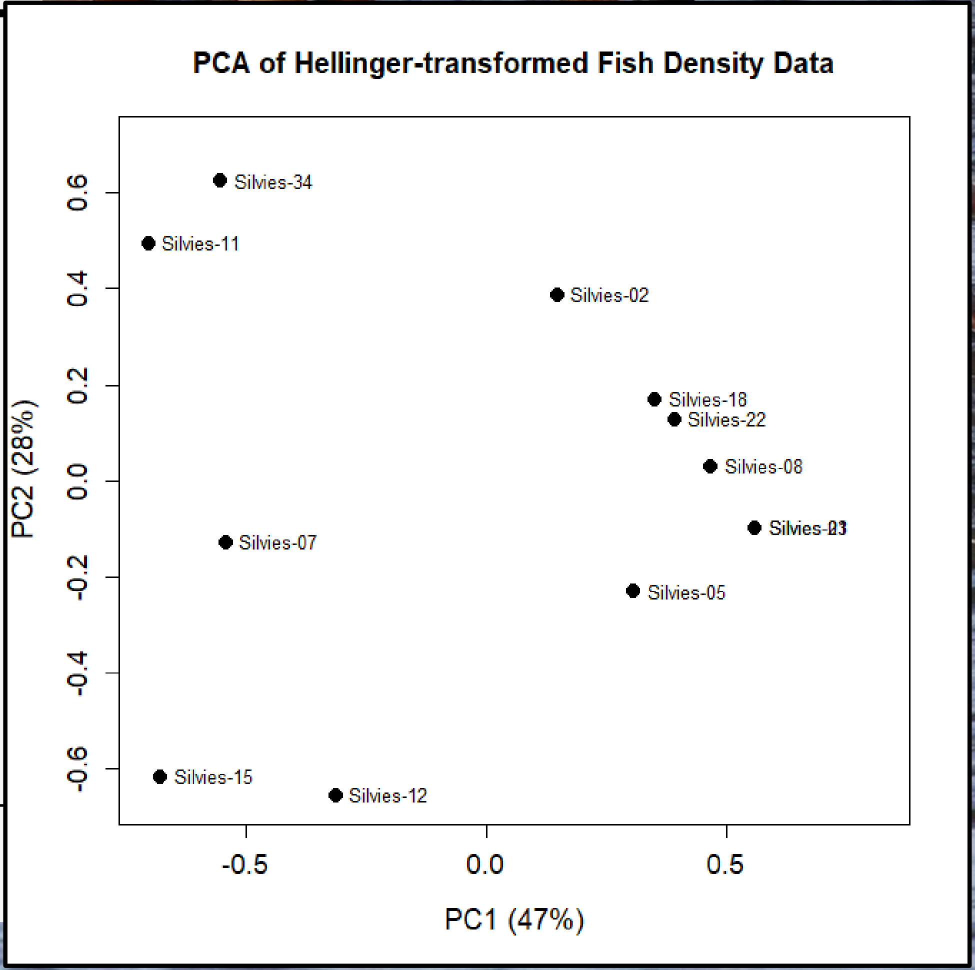
# Effective Biplots: Anatomy of a Biplot

| Site ID | Redband Trout | Brook Trout | Dace | Redside Shiner | Sculpin | Sucker |
|---|---|---|---|---|---|---|
| Silvies-15 | 0 | 0 | 0 | 10 | 14 | 0 |
| Silvies-07 | 2 | 0 | 12 | 0 | 20 | 0 |
| Silvies-12 | 4 | 48 | 0 | 0 | 19 | 0 |
| Silvies-34 | 0 | 0 | 25 | 0 | 0 | 11 |
| Silvies-11 | 0 | 0 | 16 | 3 | 0 | 0 |
| Silvies-21 | 2 | 0 | 0 | 0 | 0 | 0 |
| Silvies-08 | 3 | 0 | 0 | 0 | 0 | 1 |
| Silvies-05 | 98 | 0 | 0 | 0 | 22 | 20 |
| Silvies-03 | 23 | 0 | 0 | 0 | 0 | 0 |
| Silvies-02 | 44 | 0 | 24 | 0 | 0 | 17 |
| Silvies-22 | 168 | 0 | 20 | 0 | 0 | 0 |
| Silvies-18 | 115 | 0 | 20 | 0 | 0 | 0 |



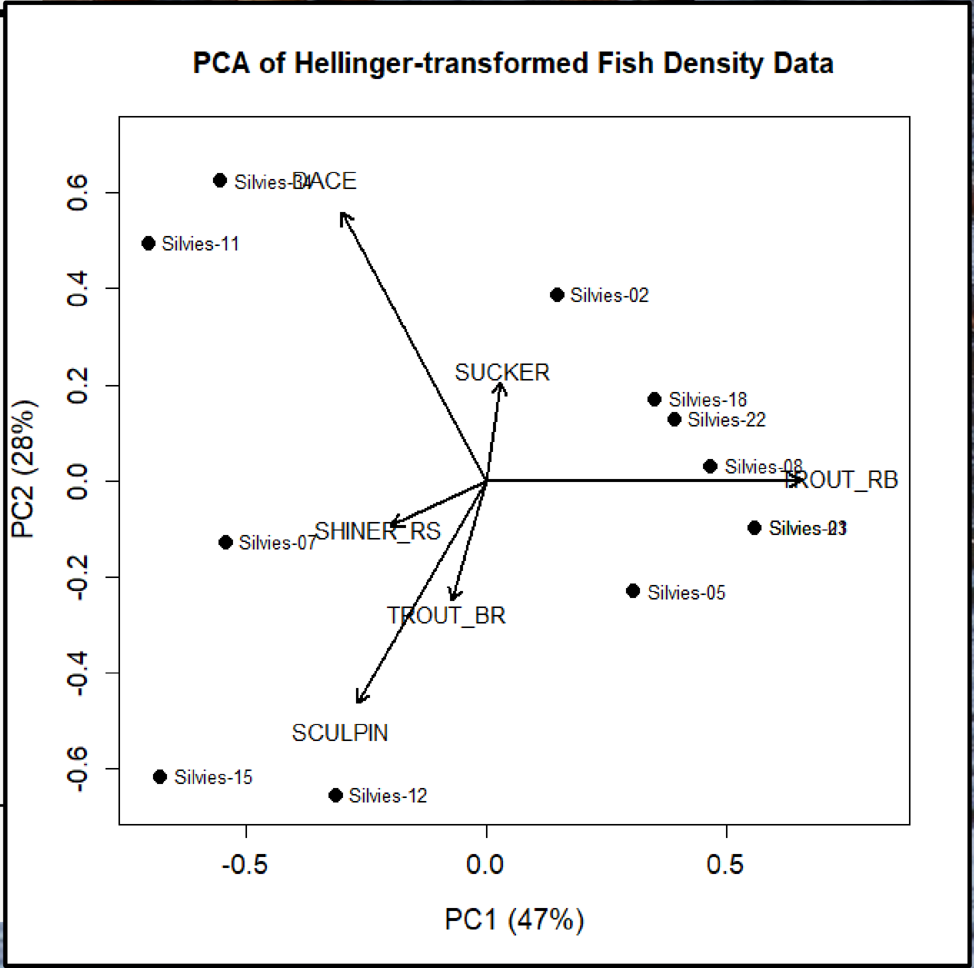
# Effective Biplots: Anatomy of a Biplot

| Site ID | Redband Trout | Brook Trout | Dace | Redside Shiner | Sculpin | Sucker |
|---|---|---|---|---|---|---|
| Silvies-15 | 0 | 0 | 0 | 10 | 14 | 0 |
| Silvies-07 | 2 | 0 | 12 | 0 | 20 | 0 |
| Silvies-12 | 4 | 48 | 0 | 0 | 19 | 0 |
| Silvies-34 | 0 | 0 | 25 | 0 | 0 | 11 |
| Silvies-11 | 0 | 0 | 16 | 3 | 0 | 0 |
| Silvies-21 | 2 | 0 | 0 | 0 | 0 | 0 |
| Silvies-08 | 3 | 0 | 0 | 0 | 0 | 1 |
| Silvies-05 | 98 | 0 | 0 | 0 | 22 | 20 |
| Silvies-03 | 23 | 0 | 0 | 0 | 0 | 0 |
| Silvies-02 | 44 | 0 | 24 | 0 | 0 | 17 |
| Silvies-22 | 168 | 0 | 20 | 0 | 0 | 0 |
| Silvies-18 | 115 | 0 | 20 | 0 | 0 | 0 |



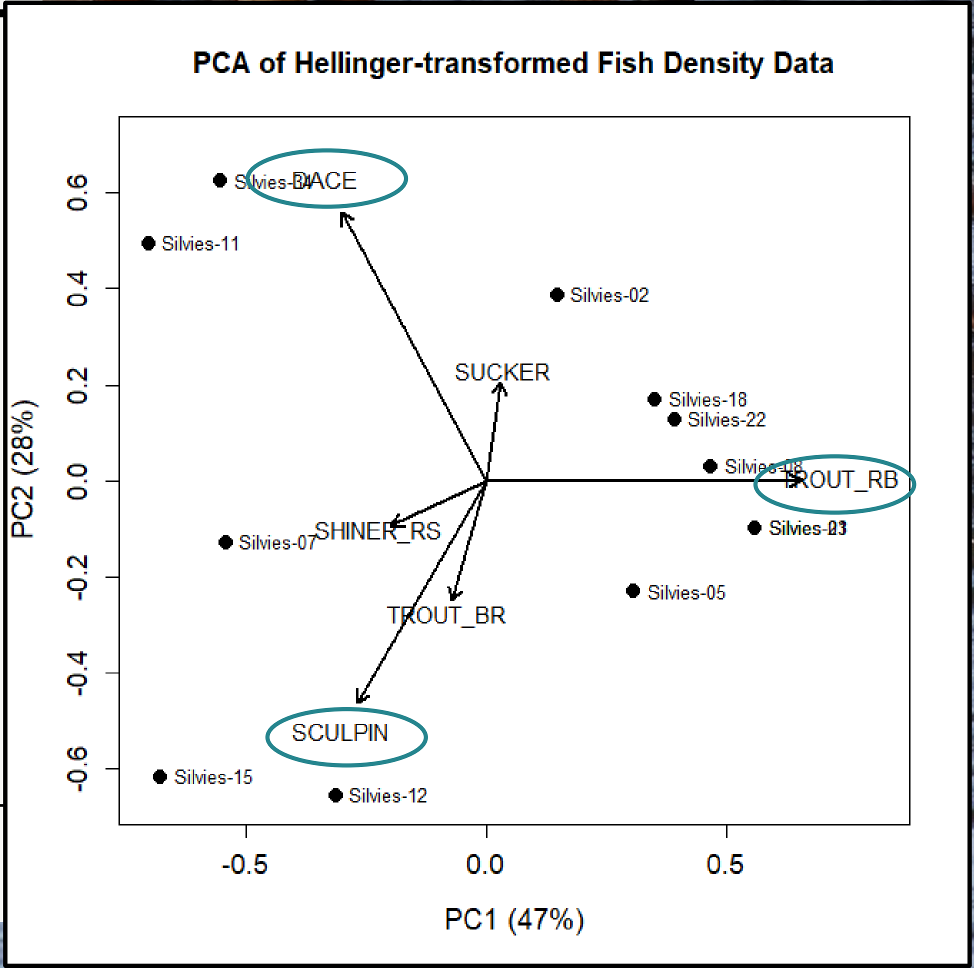
# Effective Biplots: Anatomy of a Biplot

| Site ID | Redband Trout | Brook Trout | Dace | Redside Shiner | Sculpin | Sucker |
|---|---|---|---|---|---|---|
| Silvies-15 | 0 | 0 | 0 | 10 | 14 | 0 |
| Silvies-07 | 2 | 0 | 12 | 0 | 20 | 0 |
| Silvies-12 | 4 | 48 | 0 | 0 | 19 | 0 |
| Silvies-34 | 0 | 0 | 25 | 0 | 0 | 11 |
| Silvies-11 | 0 | 0 | 16 | 3 | 0 | 0 |
| Silvies-21 | 2 | 0 | 0 | 0 | 0 | 0 |
| Silvies-08 | 3 | 0 | 0 | 0 | 0 | 1 |
| Silvies-05 | 98 | 0 | 0 | 0 | 22 | 20 |
| Silvies-03 | 23 | 0 | 0 | 0 | 0 | 0 |
| Silvies-02 | 44 | 0 | 24 | 0 | 0 | 17 |
| Silvies-22 | 168 | 0 | 20 | 0 | 0 | 0 |
| Silvies-18 | 115 | 0 | 20 | 0 | 0 | 0 |



# Effective Biplots: Anatomy of a Biplot

| Site ID | 68% on PC1<br>Redband Trout | Brook Trout | 49% on PC2<br>Dace | Redside Shiner | 33% on PC2<br>Sculpin | Sucker |
|---|---|---|---|---|---|---|
| Silvies-15 | 0 | 0 | 0 | 10 | 14 | 0 |
| Silvies-07 | 2 | 0 | 12 | 0 | 20 | 0 |
| Silvies-12 | 4 | 48 | 0 | 0 | 19 | 0 |
| Silvies-34 | 0 | 0 | 25 | 0 | 0 | 11 |
| Silvies-11 | 0 | 0 | 16 | 3 | 0 | 0 |
| Silvies-21 | 2 | 0 | 0 | 0 | 0 | 0 |
| Silvies-08 | 3 | 0 | 0 | 0 | 0 | 1 |
| Silvies-05 | 98 | 0 | 0 | 0 | 22 | 20 |
| Silvies-03 | 23 | 0 | 0 | 0 | 0 | 0 |
| Silvies-02 | 44 | 0 | 24 | 0 | 0 | 17 |
| Silvies-22 | 168 | 0 | 20 | 0 | 0 | 0 |
| Silvies-18 | 115 | 0 | 20 | 0 | 0 | 0 |

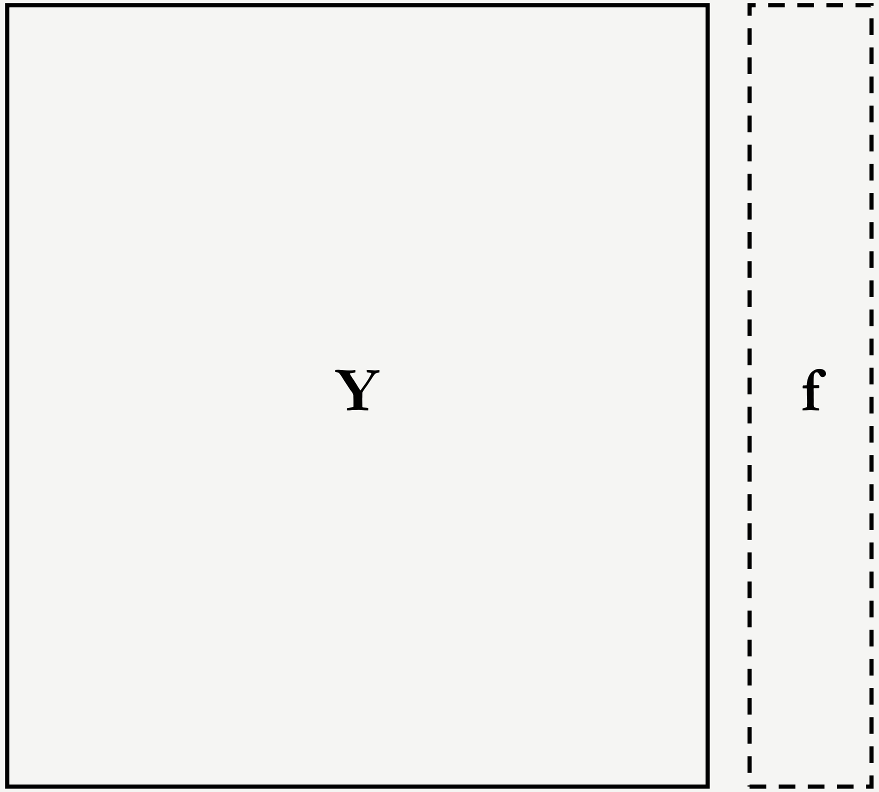


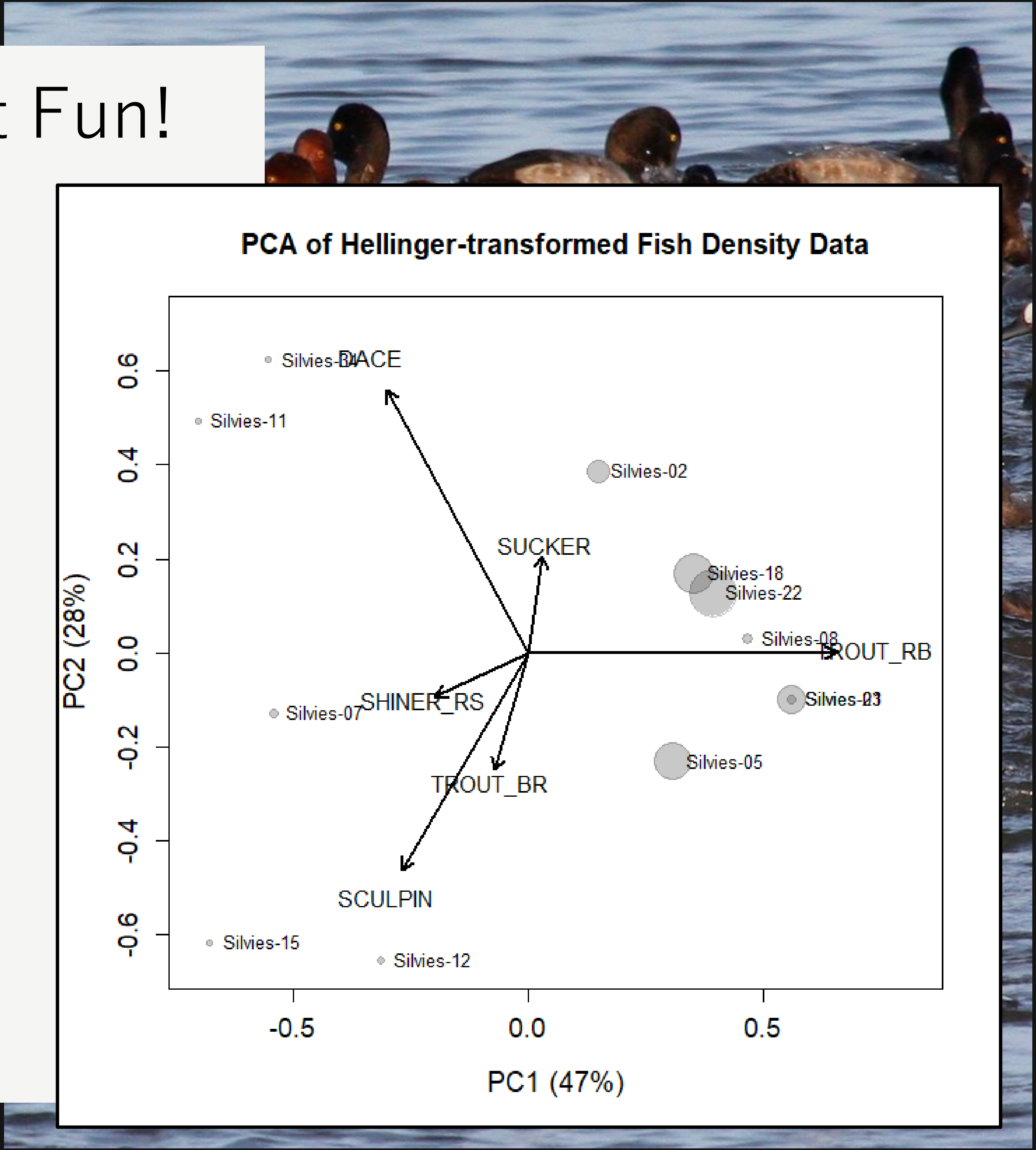
# Effective Biplots: Now Make it Fun!

Point size by species abundance

**Response** Variable(s)

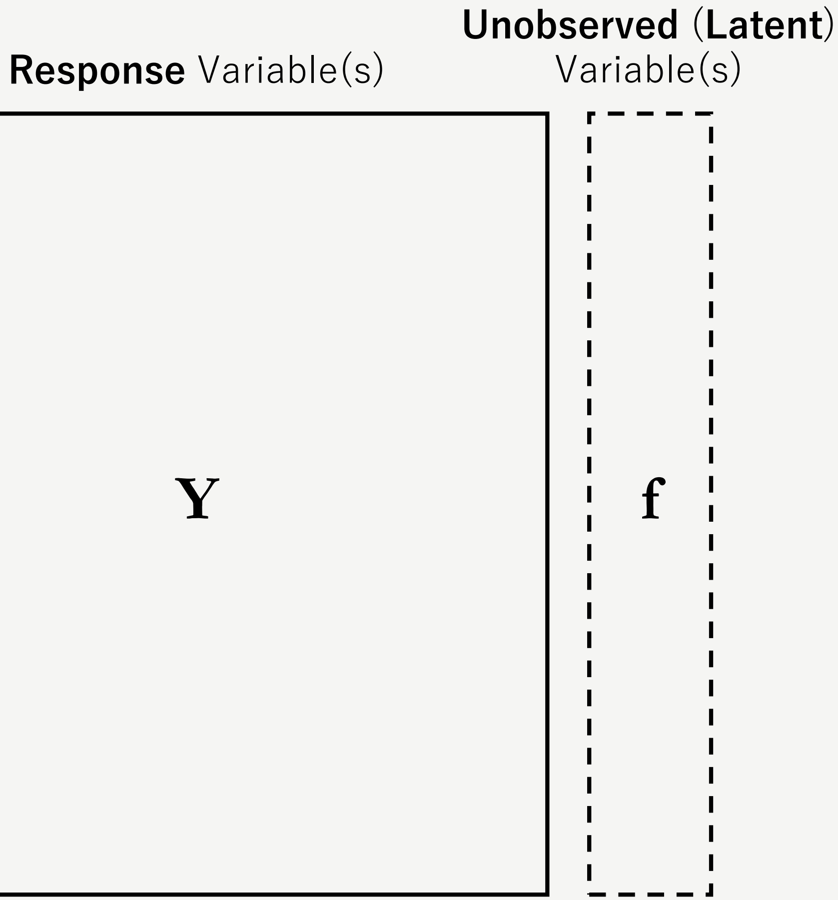
**Unobserved** (**Latent**) Variable(s)

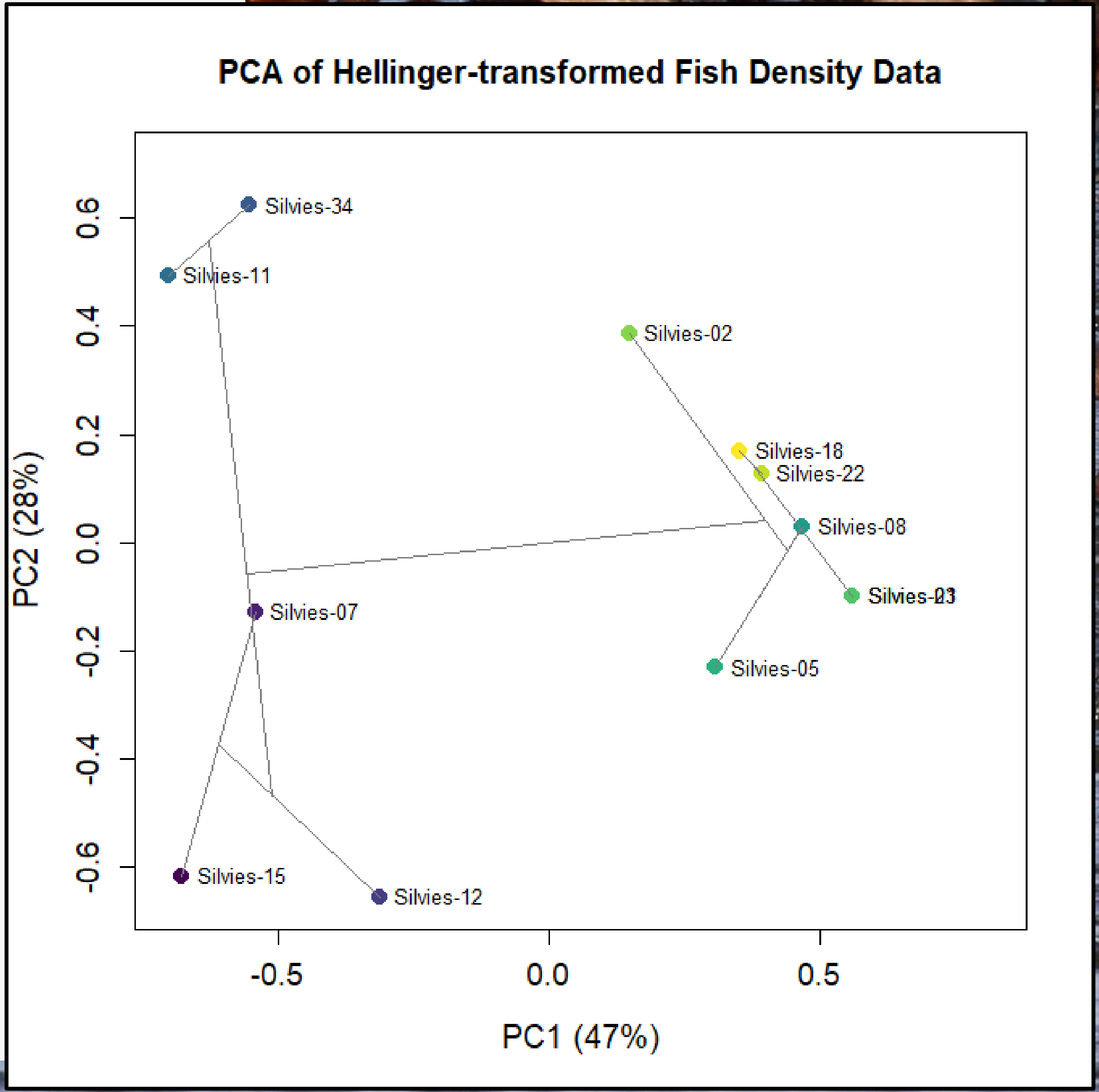


# Effective Biplots: Now Make it Fun!

Combine ordination and clustering output!



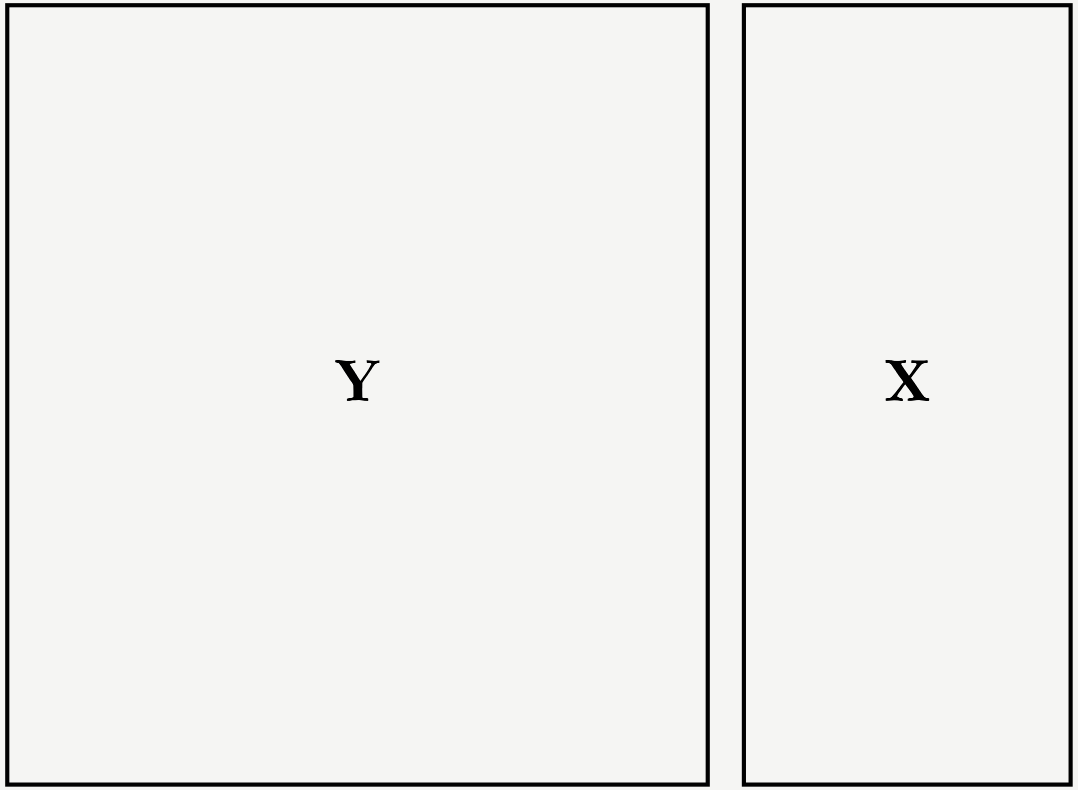



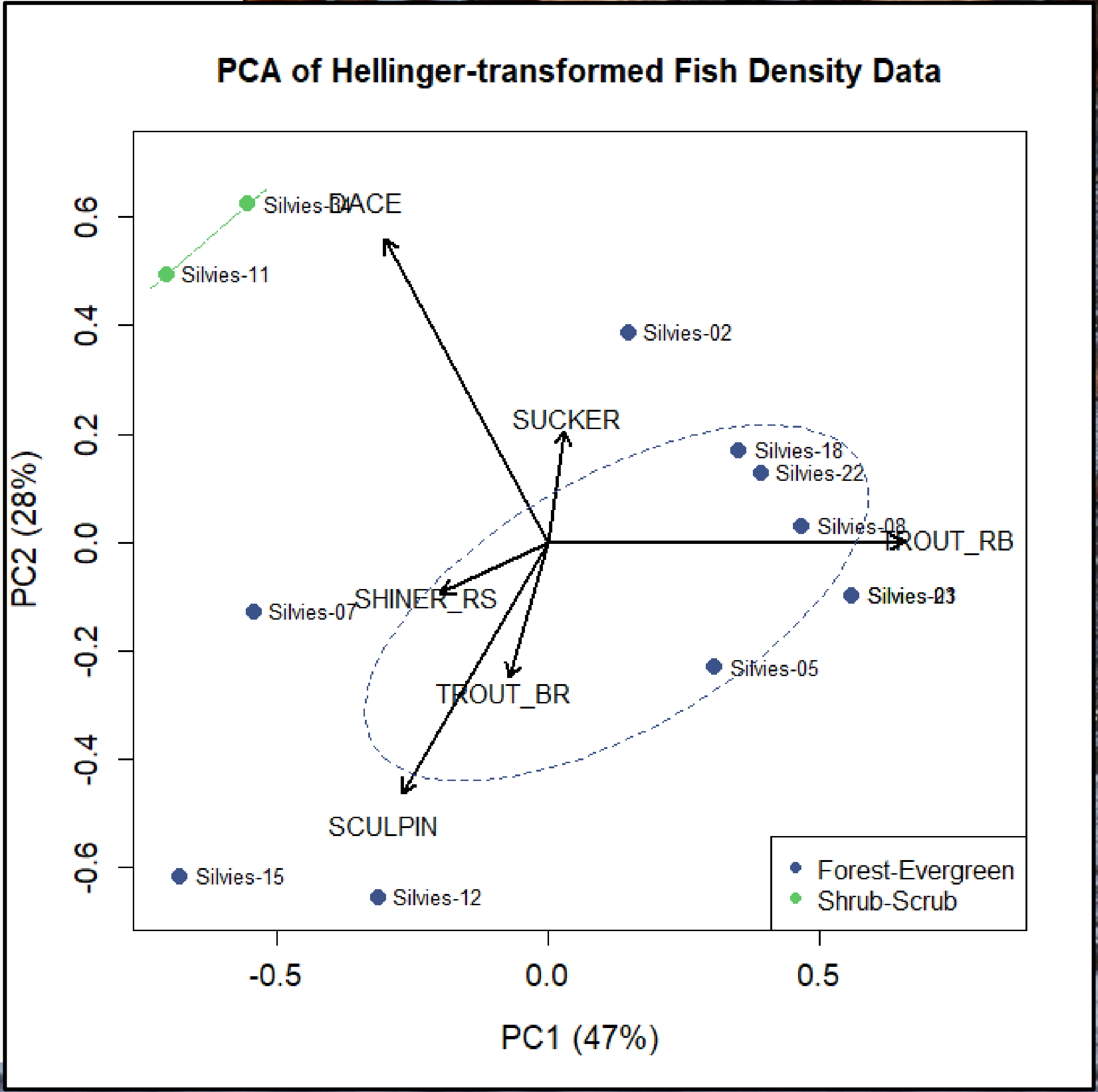
# Effective Biplots: Now Make it Fun!

Color/ellipse by habitat type

**Response** Variable(s)

**Predictor** or
**Explanatory** Variable(s)

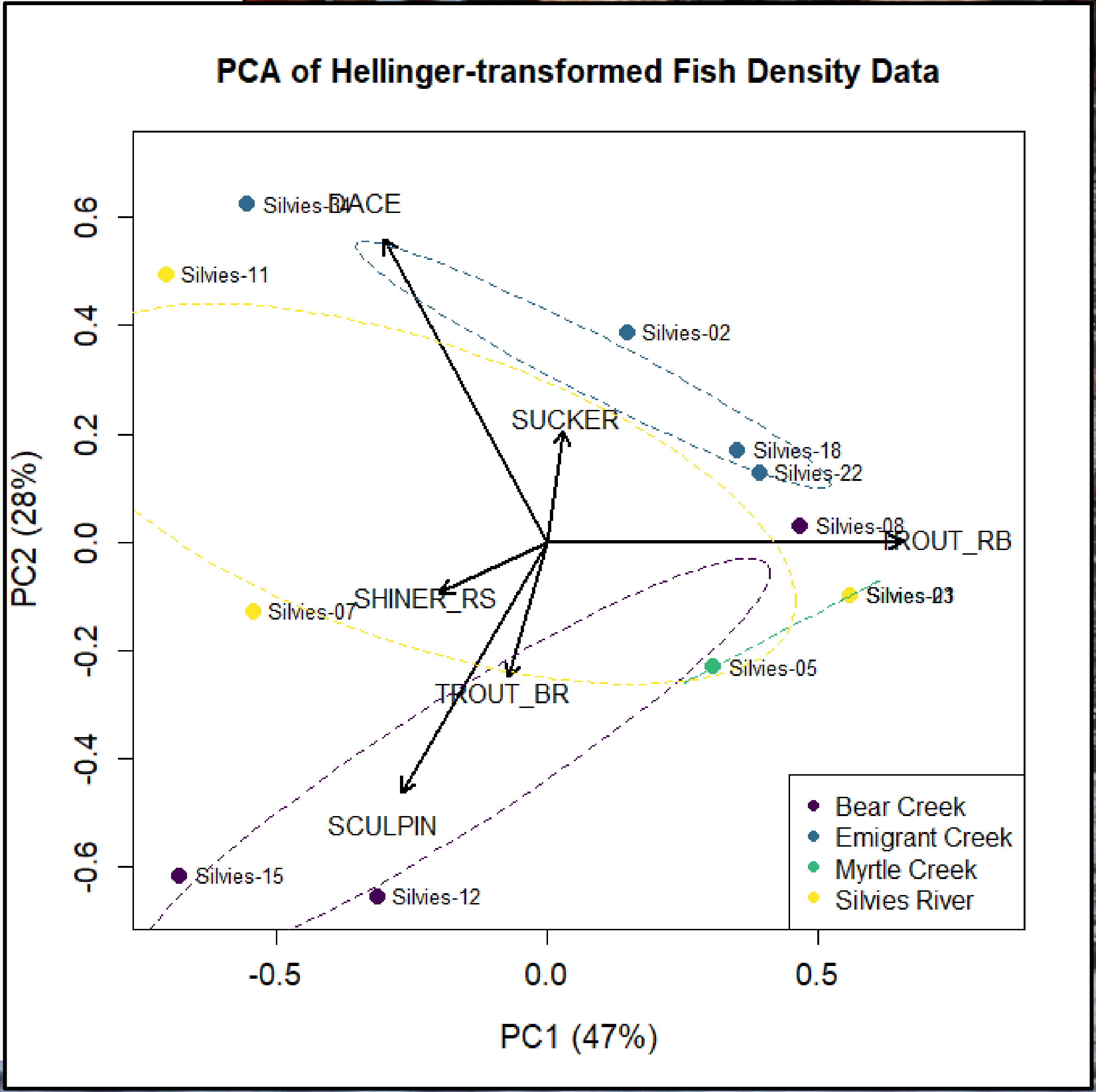


# Effective Biplots: Now Make it Fun!

Color/ellipse by tributary

| **Response** Variable(s) | **Predictor** or **Explanatory** Variable(s) |
| --- | --- |
| **Y** | **X** |



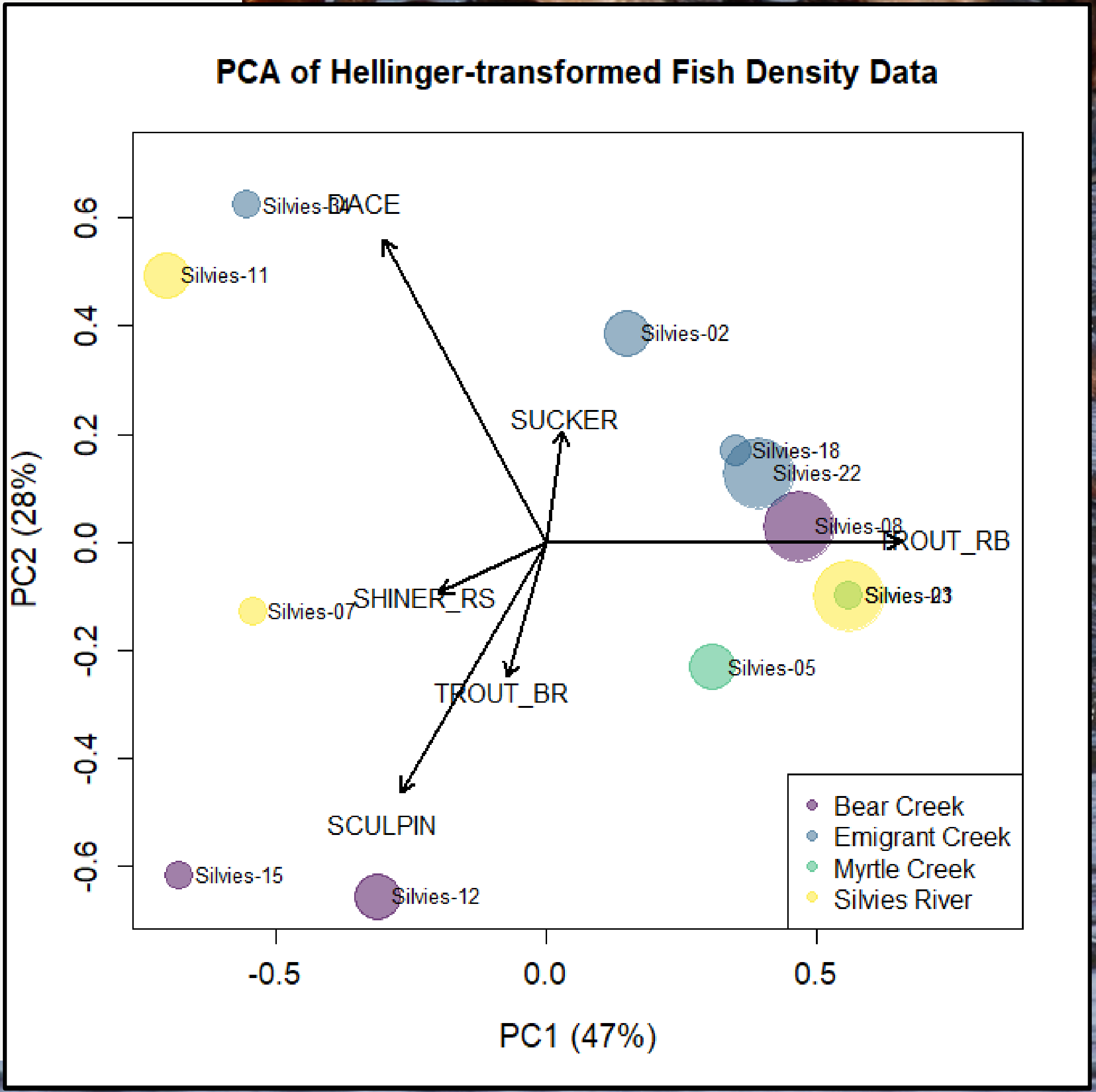
# Effective Biplots: Now Make it Fun!

Color/ellipse by tributary

**Response** Variable(s)

**Predictor** or **Explanatory** Variable(s)

| Y | X |
|---|---|

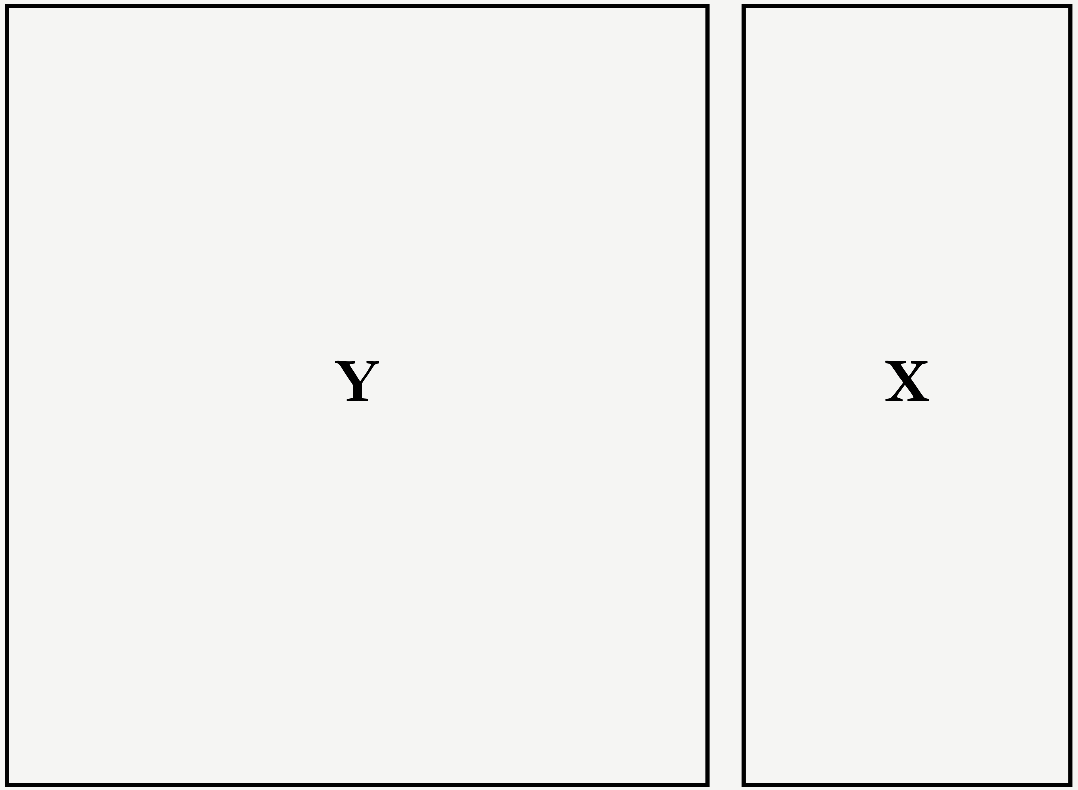


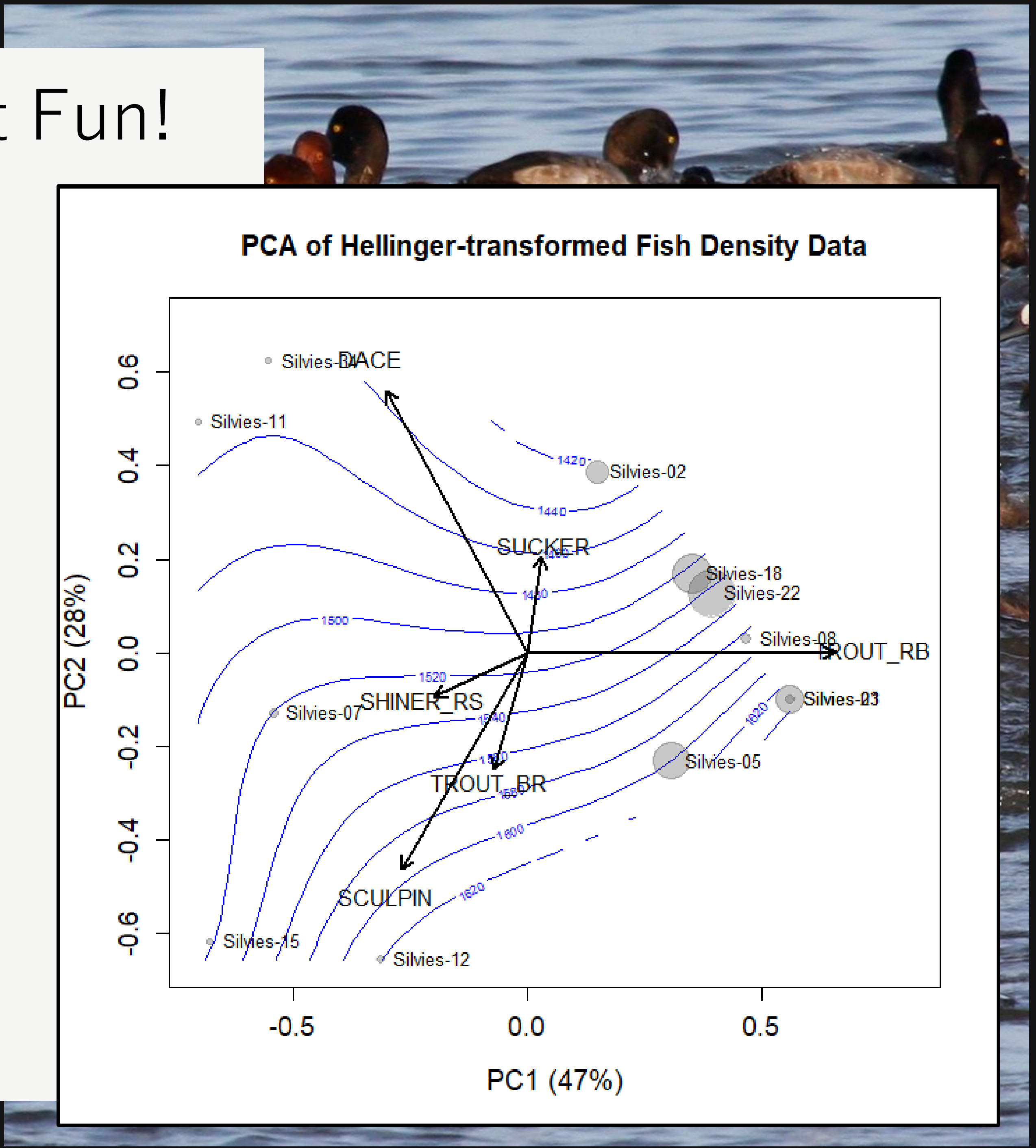
# Effective Biplots: Now Make it Fun!

## Gradient by elevation

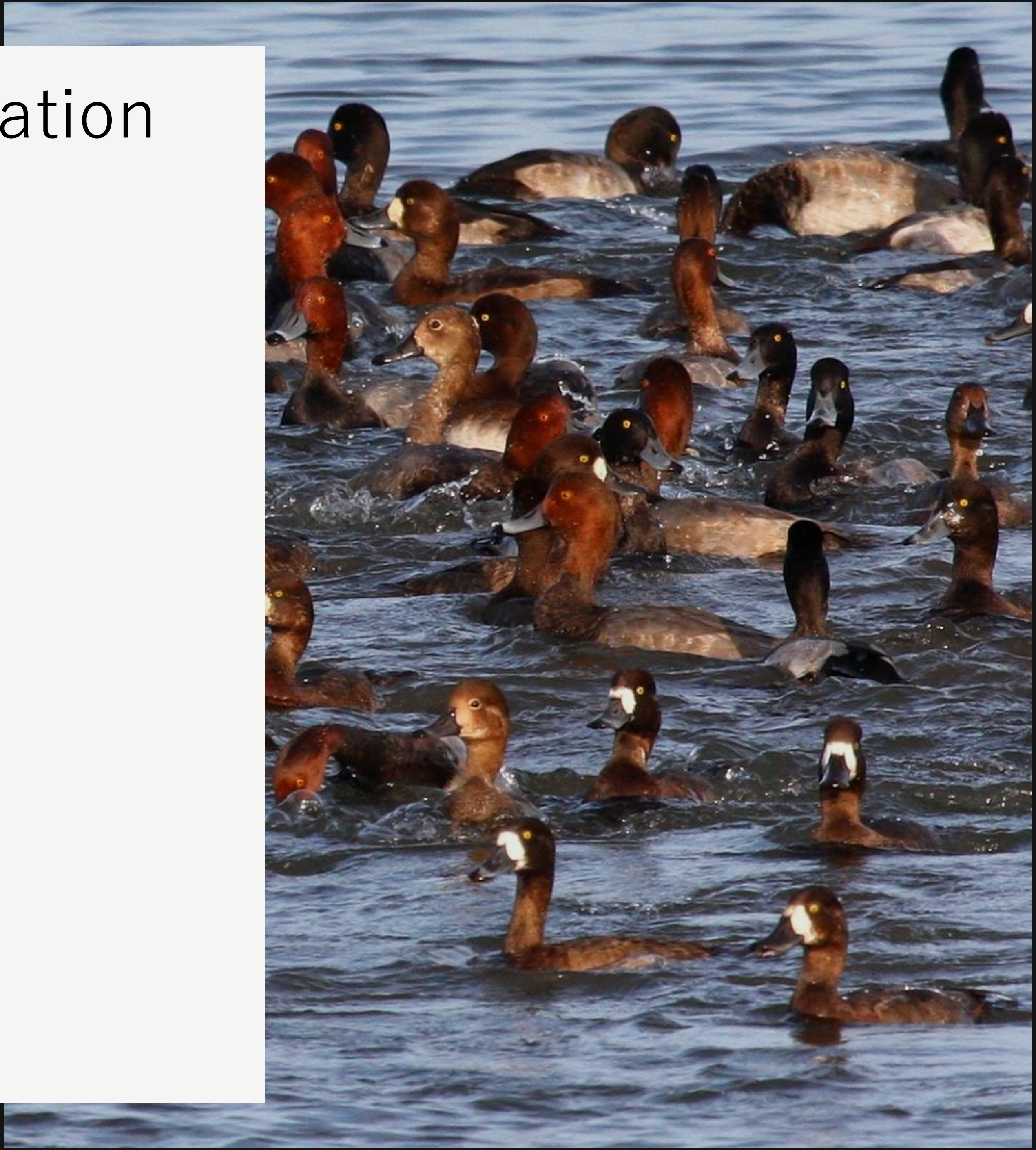
**Response** Variable(s)

**Predictor** or
**Explanatory** Variable(s)



# Making Inferences from Ordination

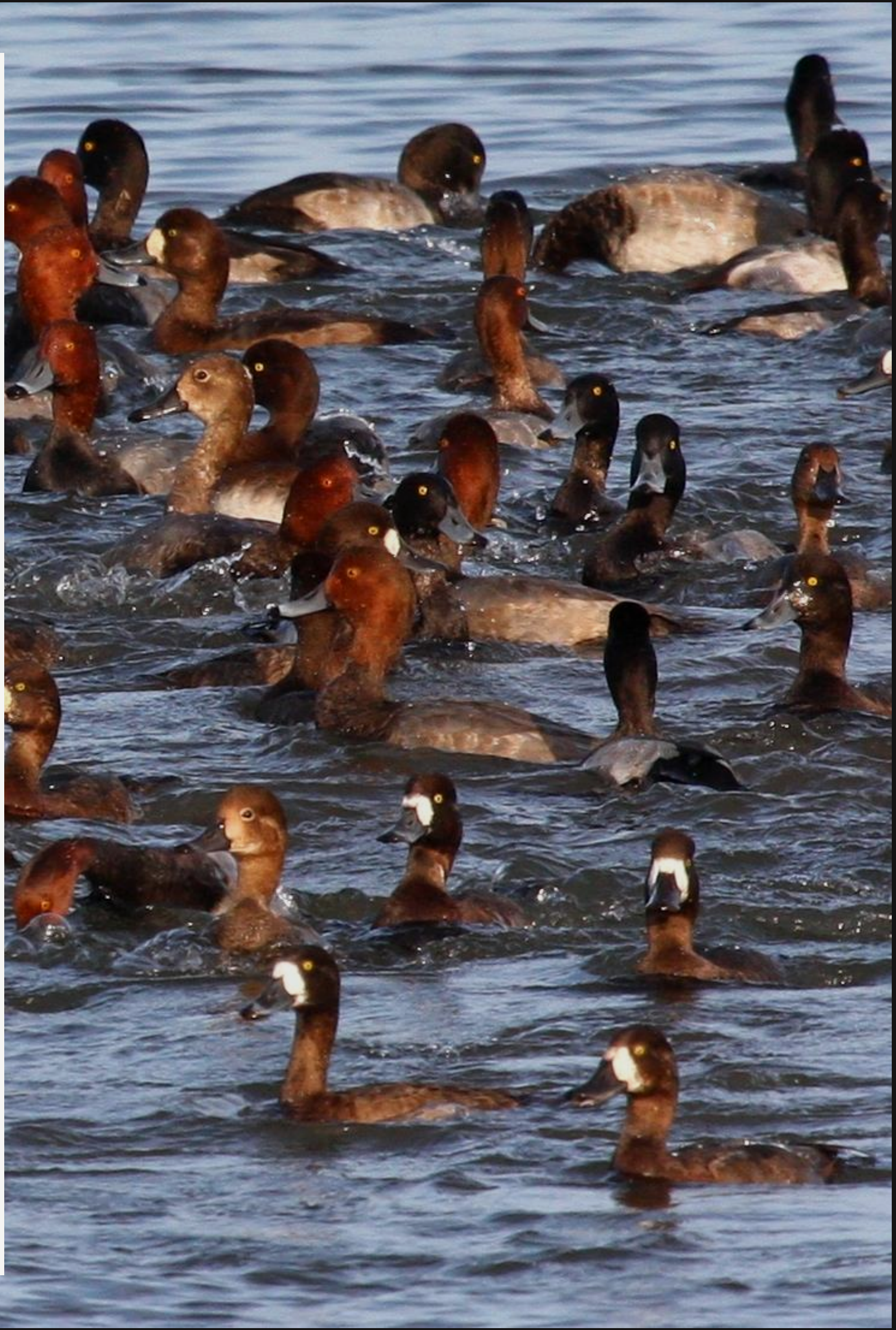


# Making Inferences from Ordination: Objectives

How do we translate our results into ecologically meaningful insights?

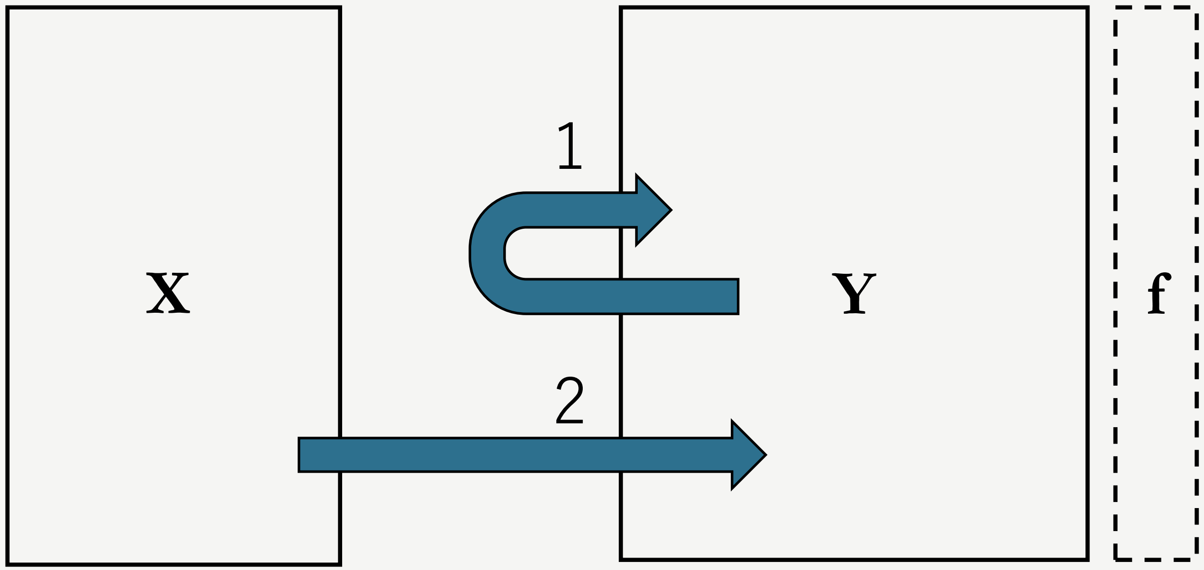
**Interpretation:** links patterns to ecological processes. Can be exploratory *or* inferential.

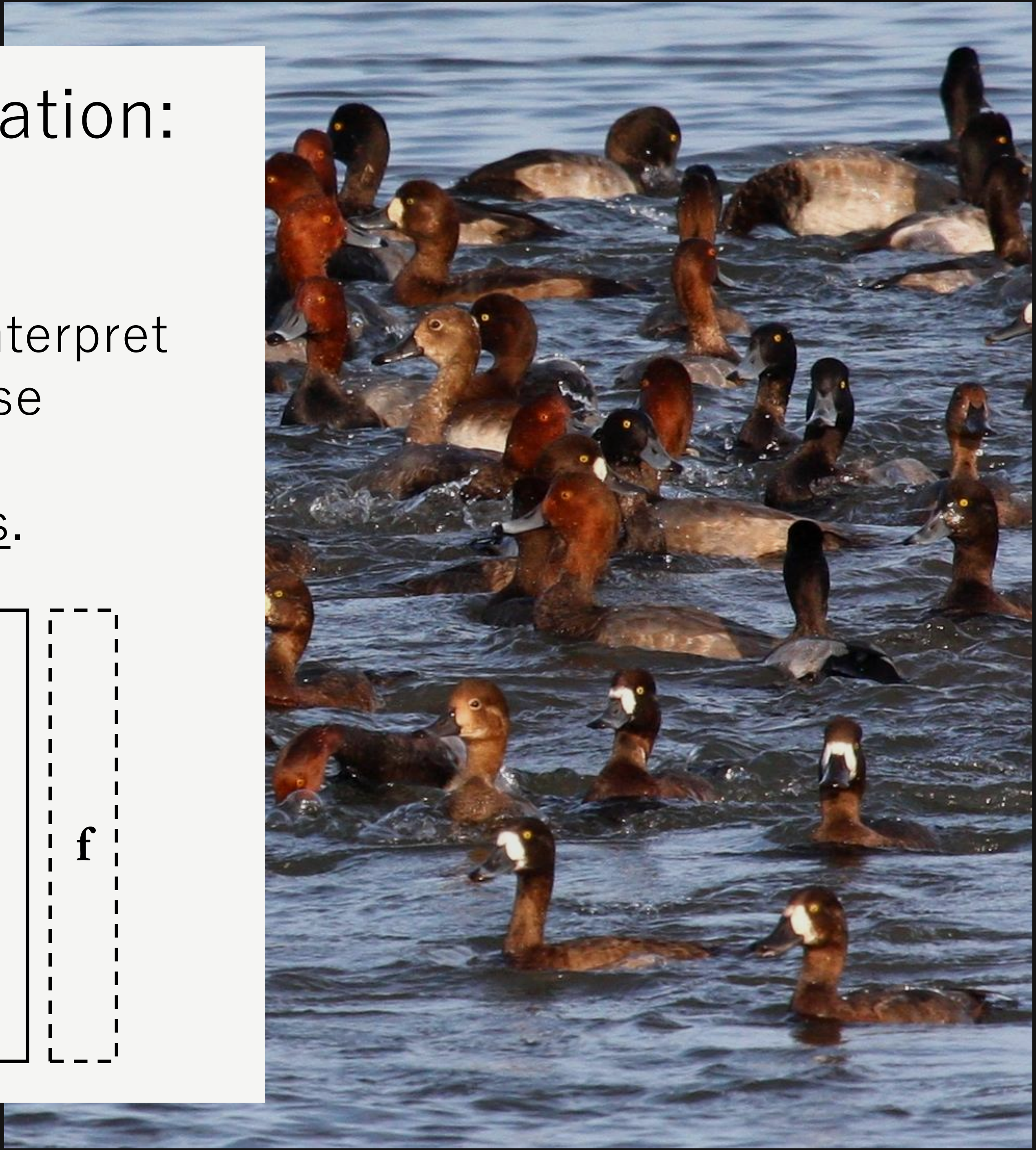
**Inference:** draws conclusions from patterns in complex datasets, usually to test hypotheses or identify key explanatory variables.



# Making Inferences from Ordination: Indirect Gradient Analysis

The goal of **indirect comparison** is to interpret the structure of the descriptors (response variables) using either the descriptors themselves or another set of descriptors.

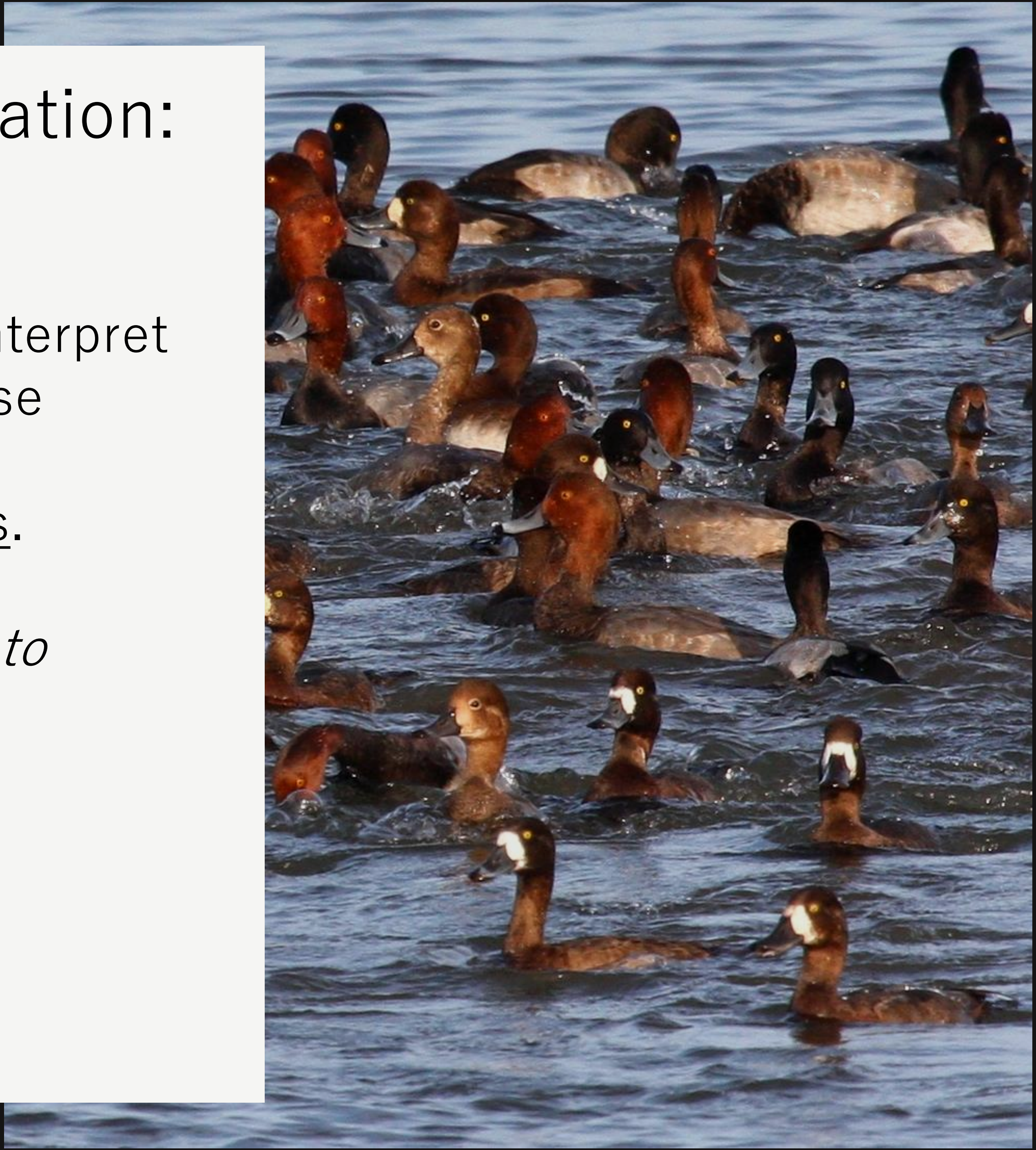


# Making Inferences from Ordination: Indirect Gradient Analysis

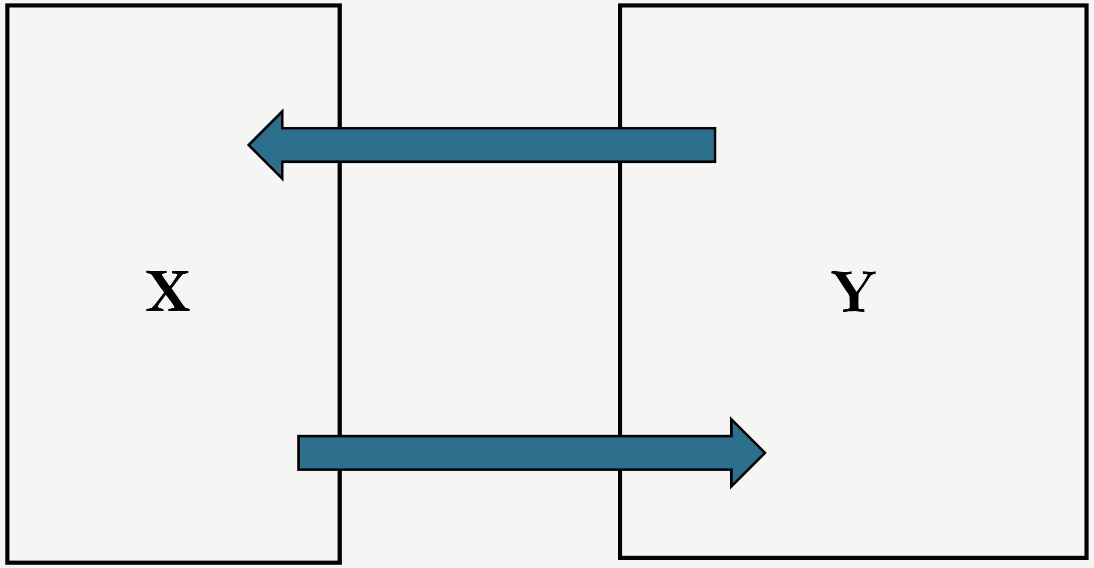
The goal of **indirect comparison** is to interpret the structure of the descriptors (response variables) using either the descriptors themselves or another set of descriptors.

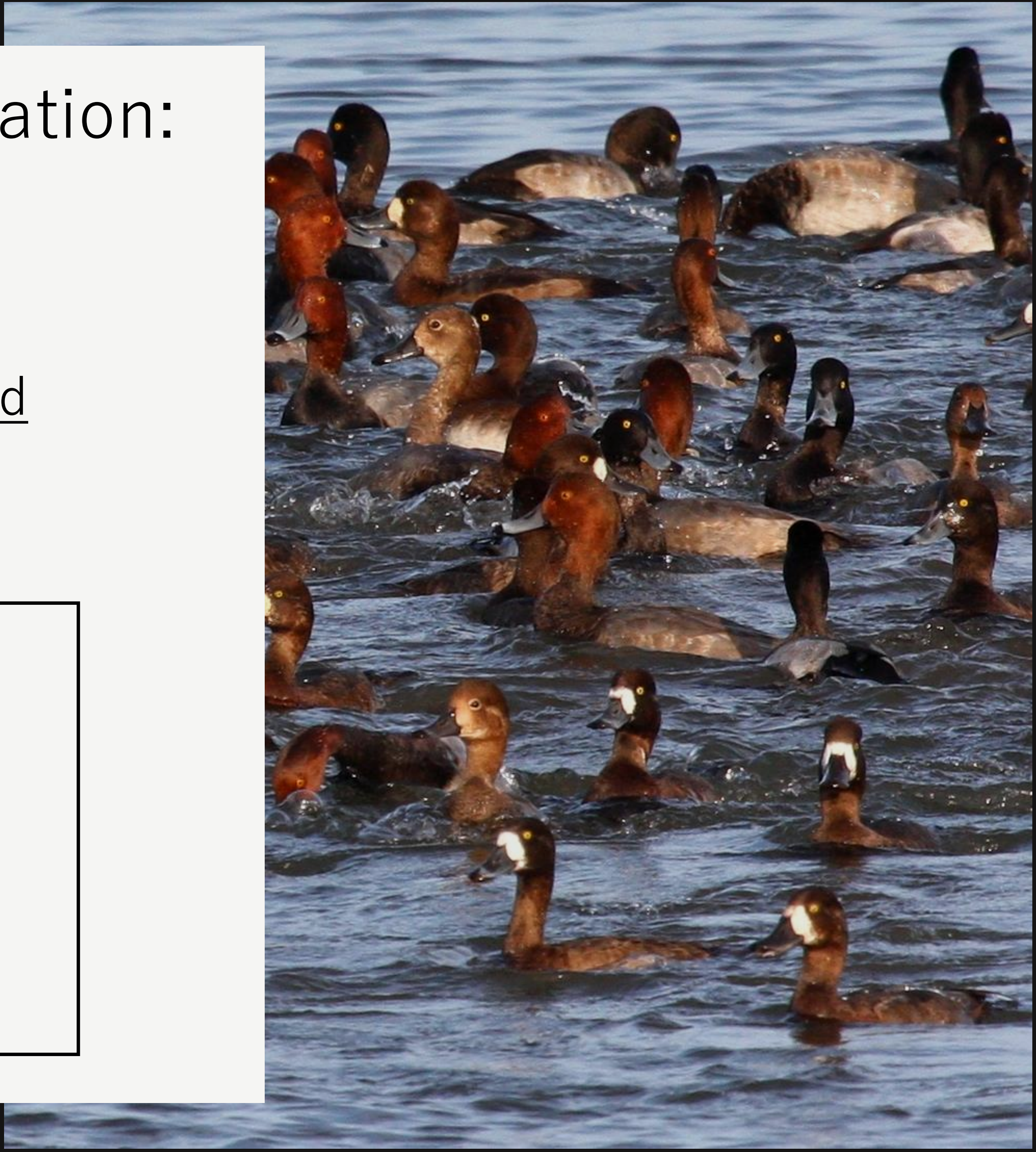
*Indirect comparison does not allow one to conduct formal tests of significance!!!*



# Making Inferences from Ordination: Direct Gradient Analysis

The goal of **direct comparison** is to simultaneously analyze the response and explanatory data matrices.

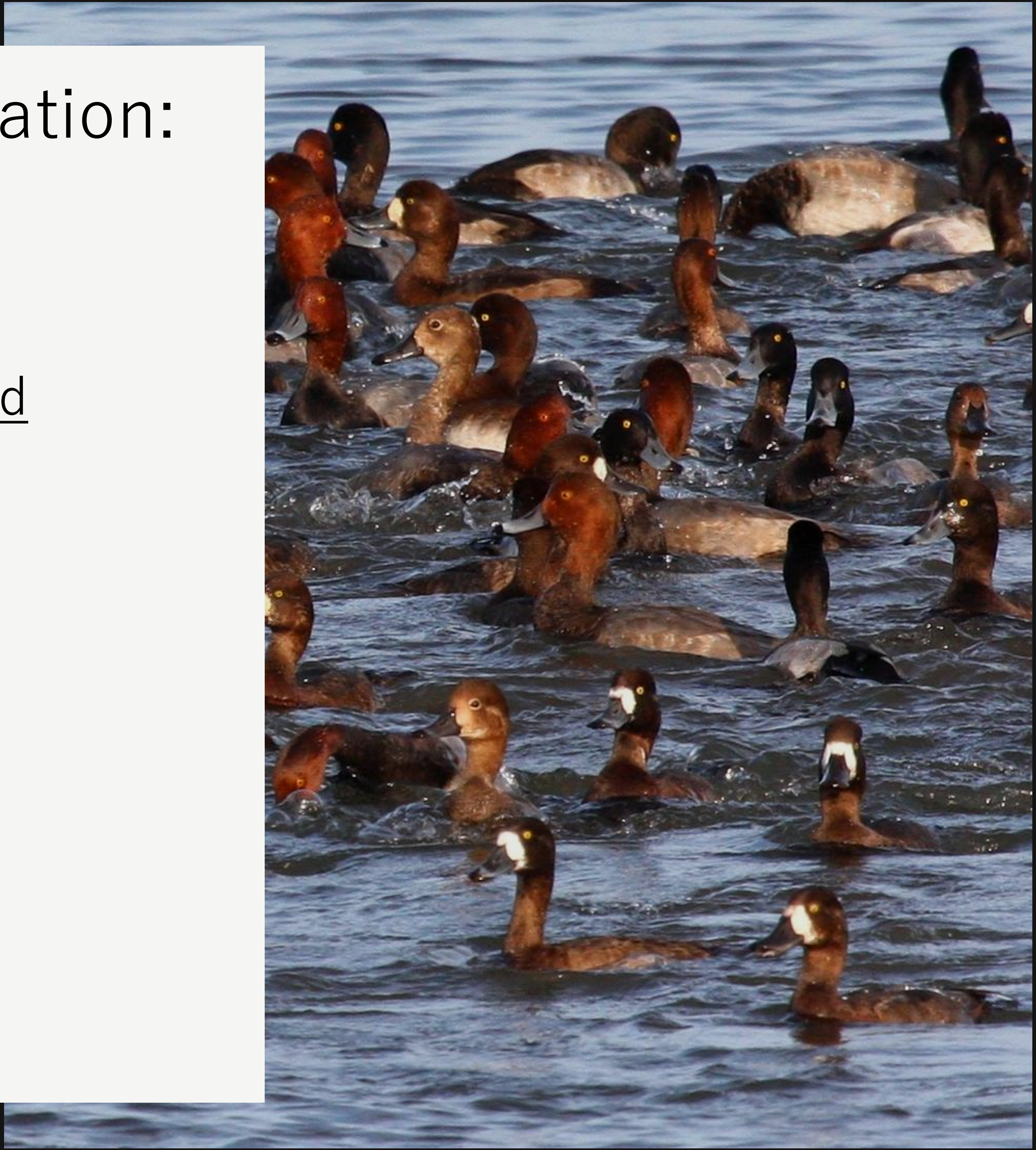


# Making Inferences from Ordination: Direct Gradient Analysis

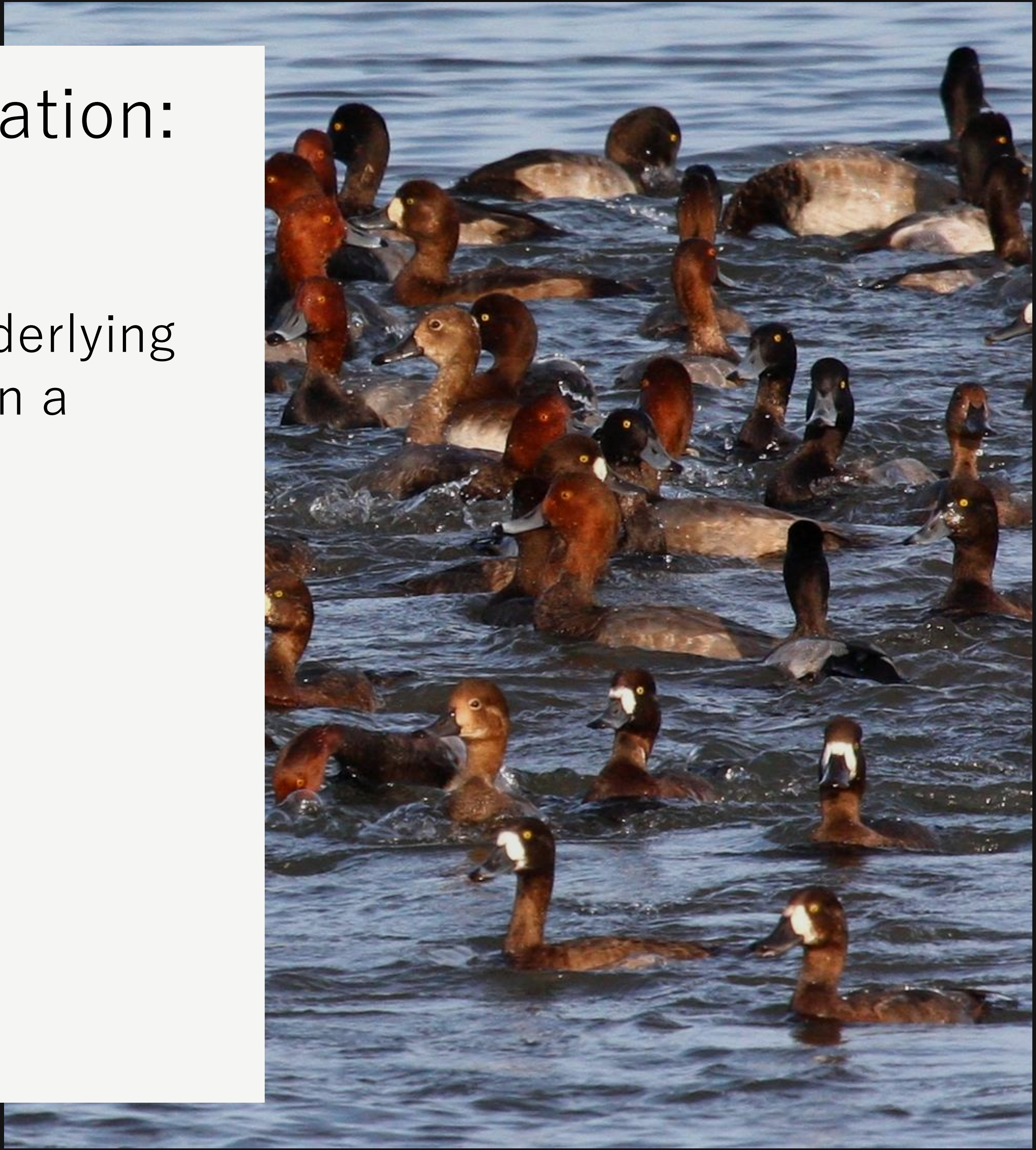
The goal of **direct comparison** is to simultaneously analyze the response and explanatory data matrices.

- Canonical Analysis (e.g., RDA, CCA)
- Mantel Test



# Making Inferences from Ordination: Explanatory

**Explanatory data analysis** looks for underlying relationships, patterns, and trends within a dataset.

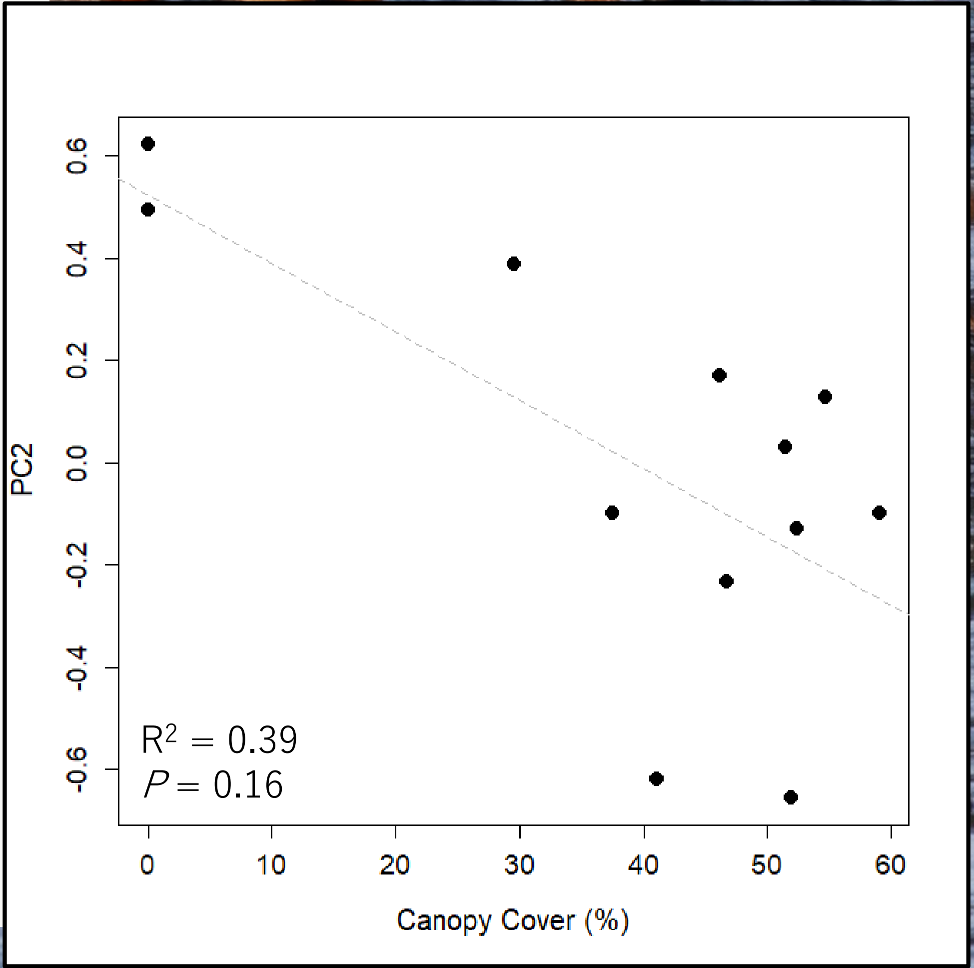


# Making Inferences from Ordination: Explanatory

**Explanatory data analysis** looks for underlying relationships, patterns, and trends within a dataset.

1) **Indirect Comparison:** Treat principal axes/coordinates or clustering partitions as response variables in a regression analysis.

2) **Direct Comparison:** Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA).

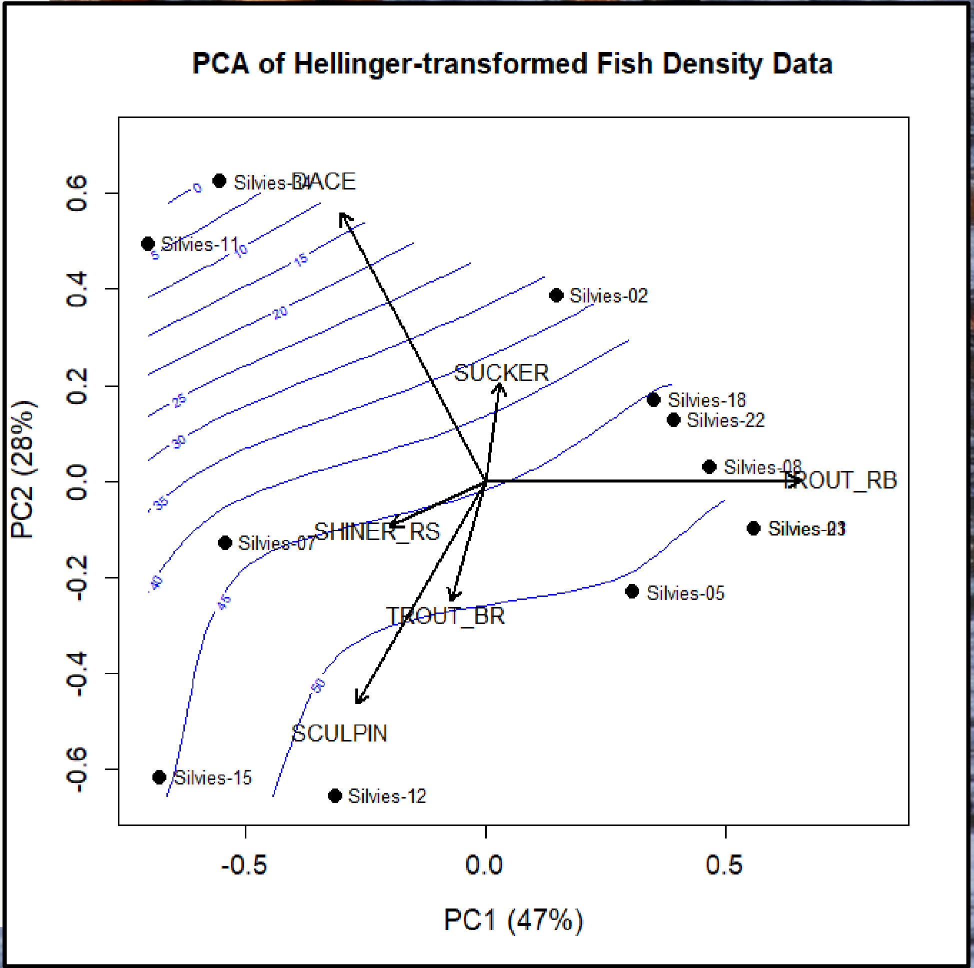


# Making Inferences from Ordination: Explanatory

**Explanatory data analysis** looks for underlying relationships, patterns, and trends within a dataset.

1) **Indirect Comparison:** Treat principal axes/coordinates or clustering partitions as response variables in a regression analysis.

2) **Direct Comparison:** Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA).

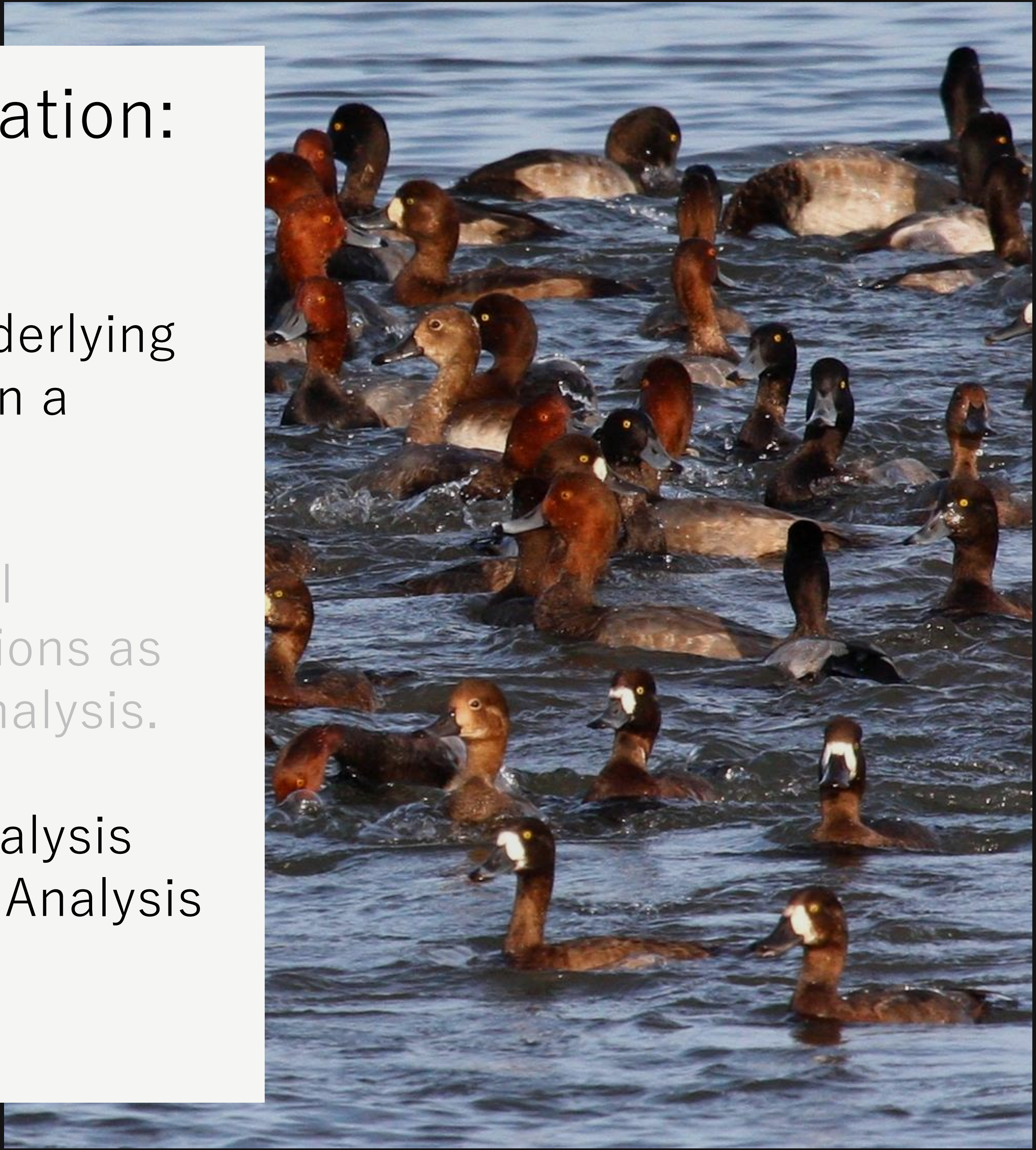


# Making Inferences from Ordination: Explanatory

**Explanatory data analysis** looks for underlying relationships, patterns, and trends within a dataset.

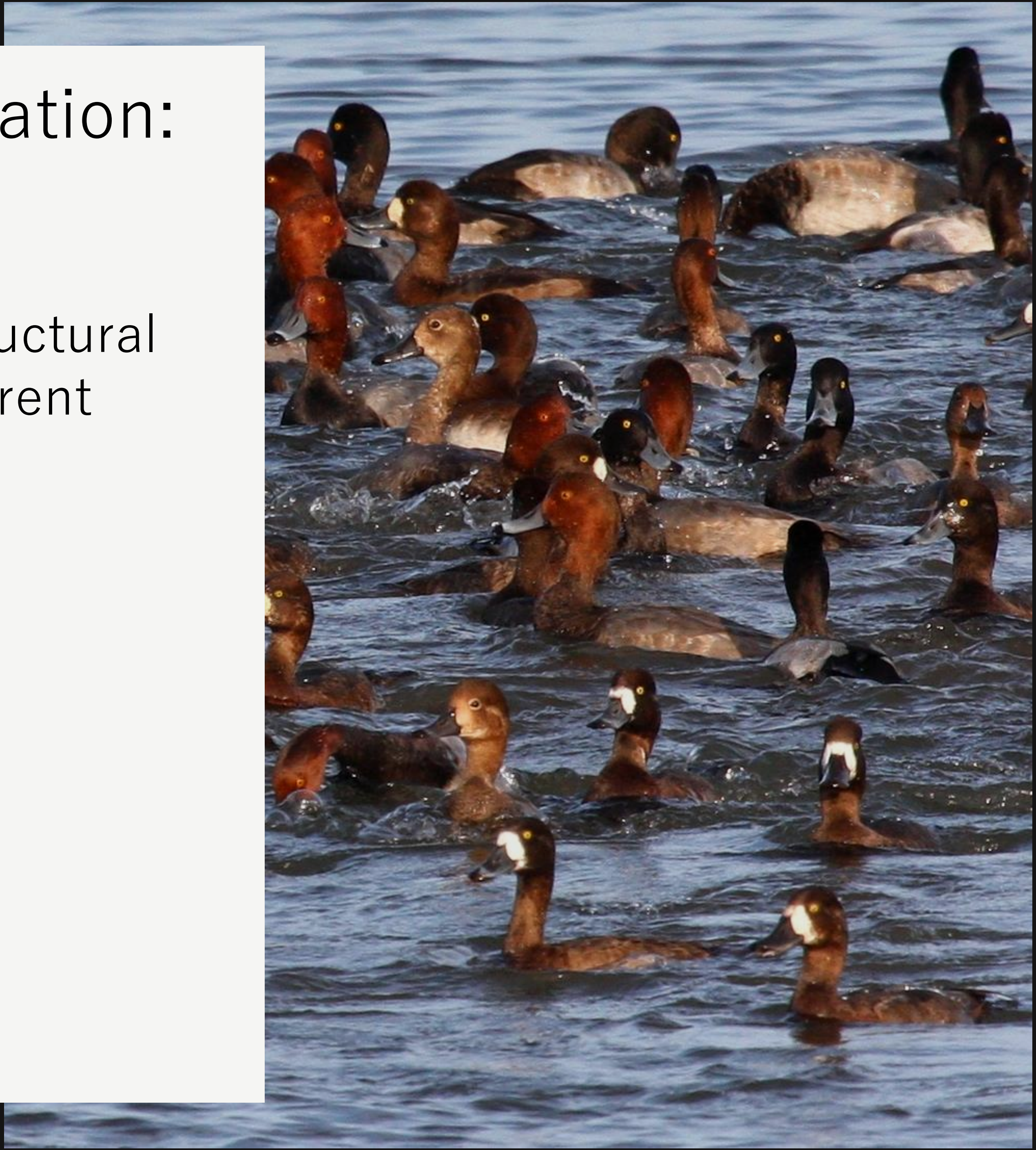
1) **Indirect Comparison:** Treat principal axes/coordinates or clustering partitions as response variables in a regression analysis.

2) **Direct Comparison:** Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA). **Next lecture!!!**



# Making Inferences from Ordination: Forecasting

**Ecological forecasting** extrapolates structural relationships among descriptors to different sites, time periods, etc.



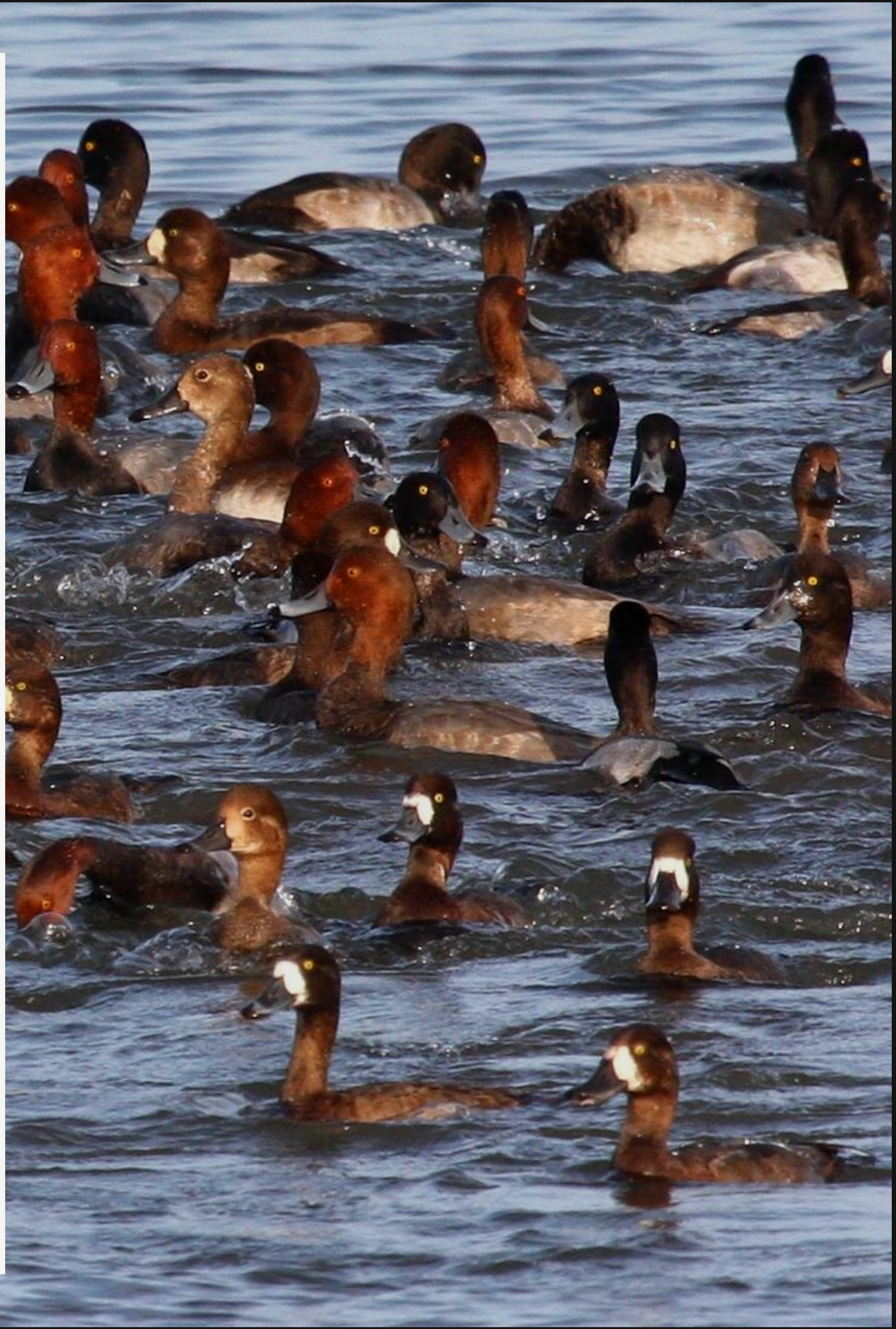
# Making Inferences from Ordination: Forecasting

**Ecological forecasting** extrapolates structural relationships among descriptors to different sites, time periods, etc.

1) Regression (indirect comparison)

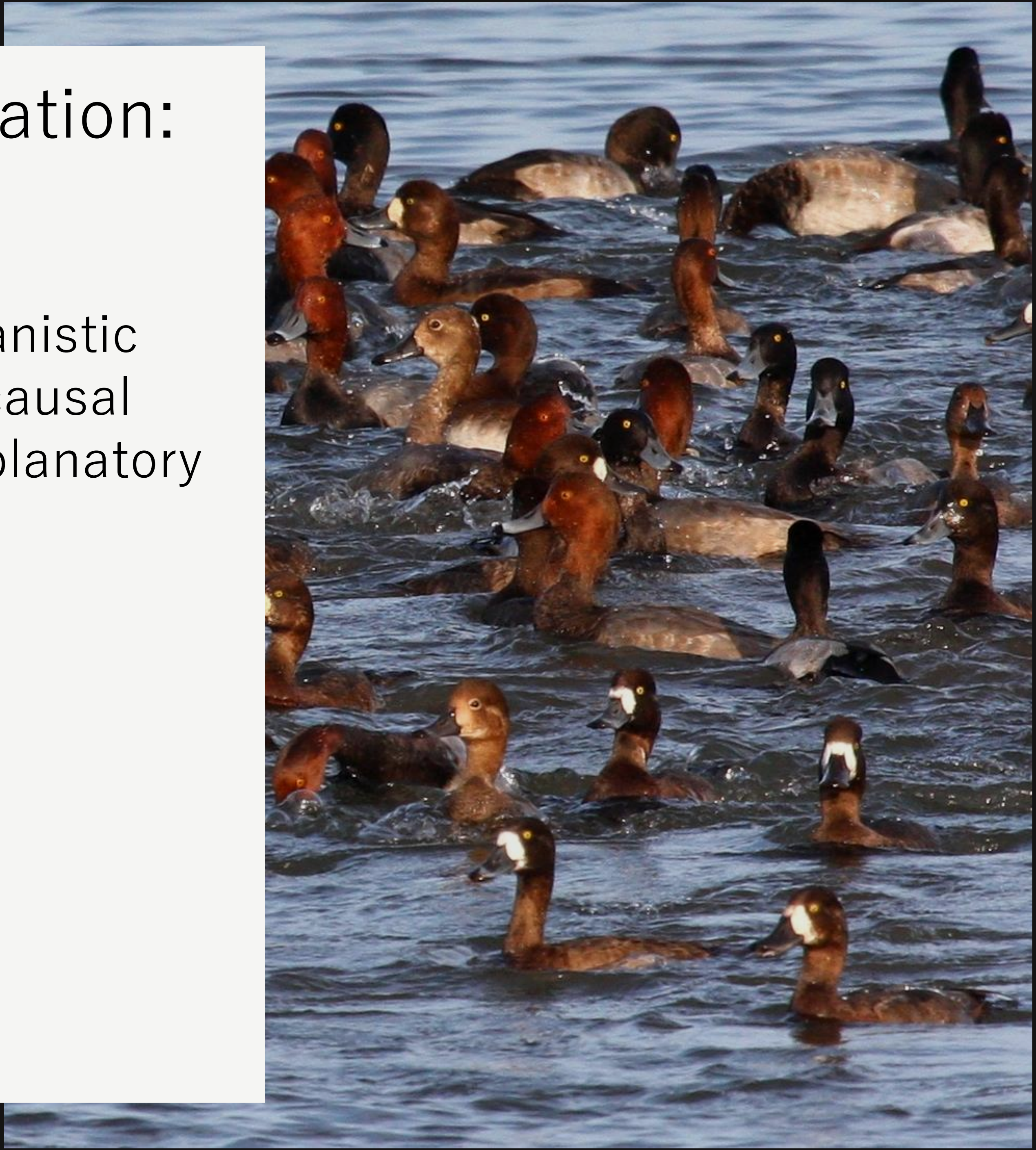
2) Canonical analysis (direct comparison; RDA, CCA)

3) Decision analysis: classification and regression trees



# Making Inferences from Ordination: Prediction

**Ecological prediction** relies on a mechanistic (functional) understanding to describe causal relationships between response and explanatory variables.

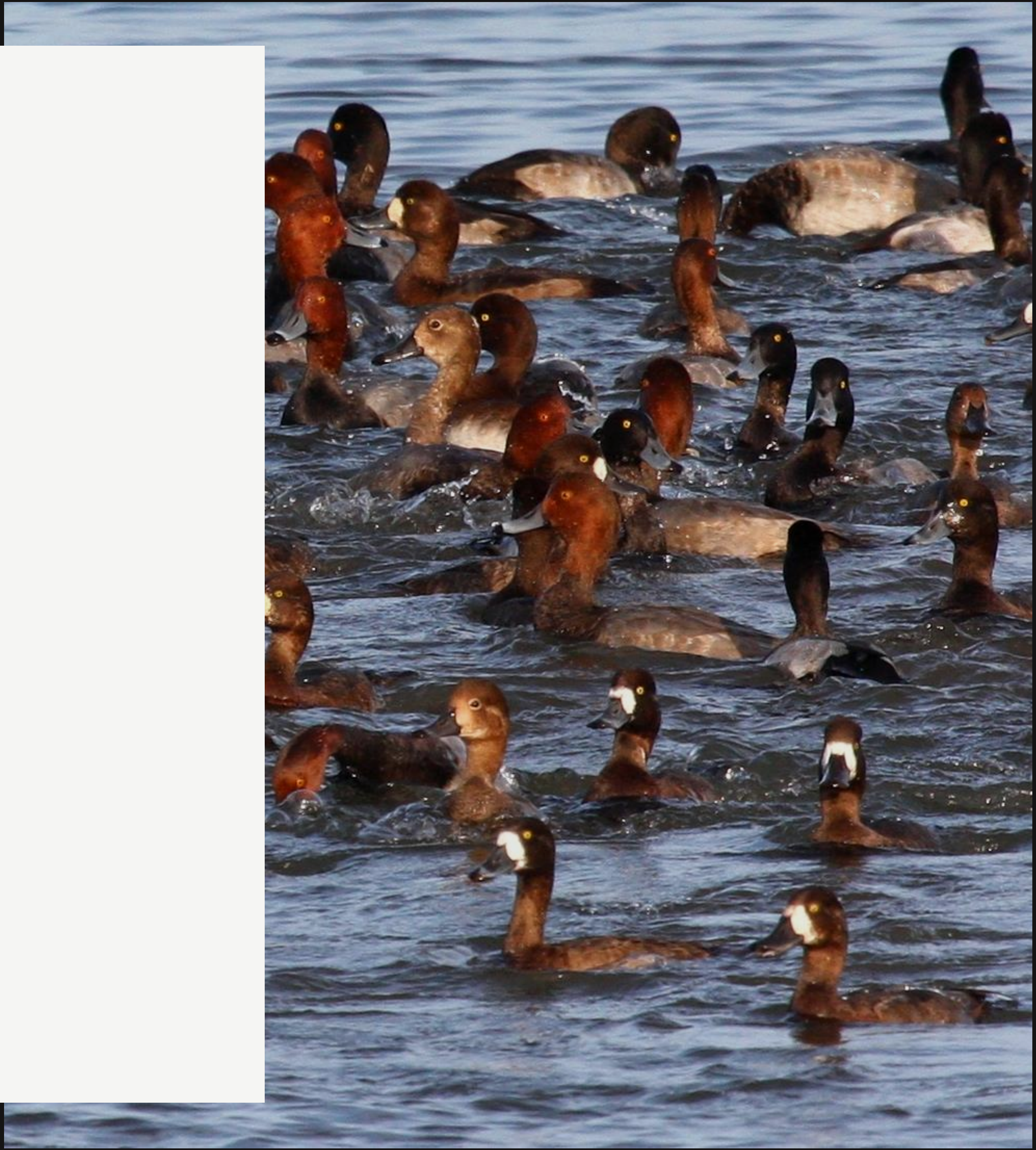


# Making Inferences from Ordination: Prediction

**Ecological prediction** relies on a mechanistic (functional) understanding to describe causal relationships between response and explanatory variables.

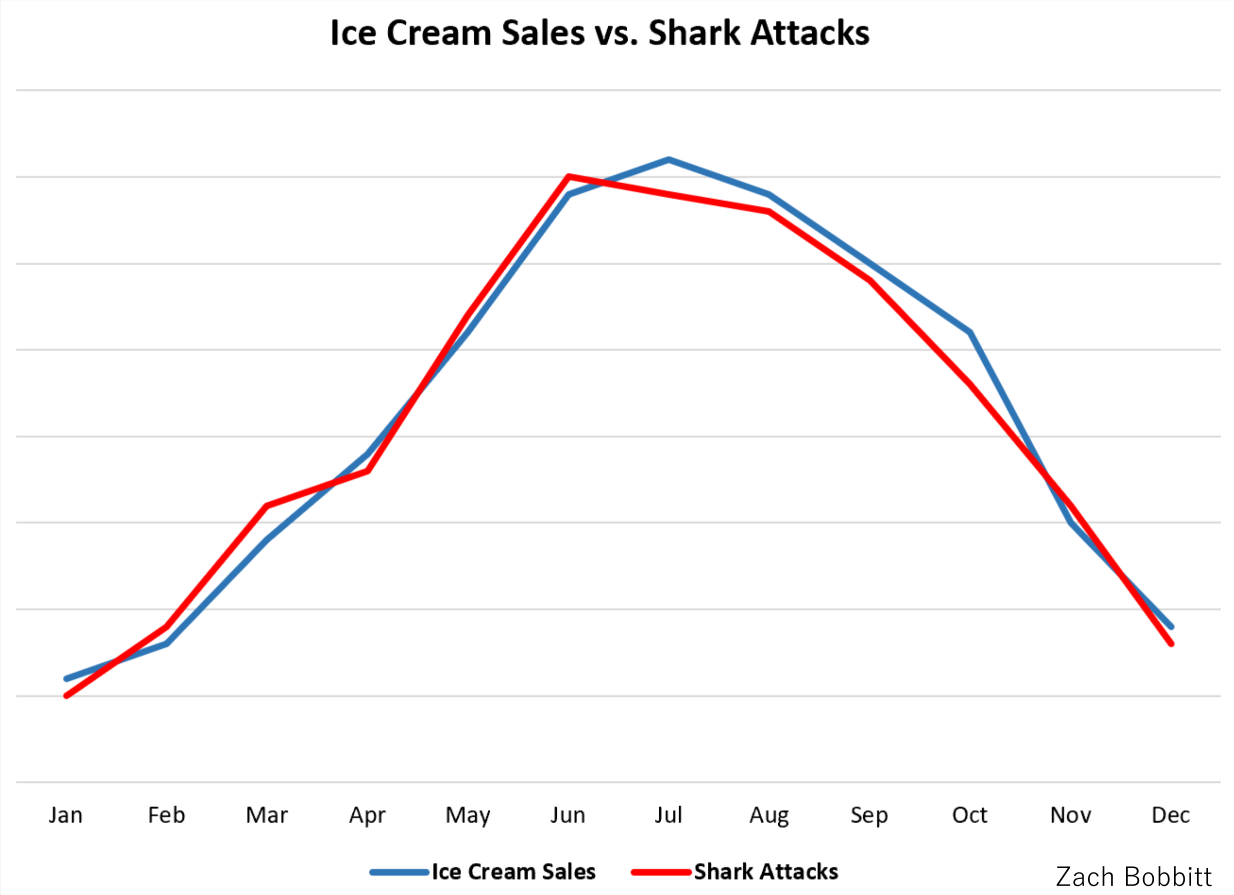
**Forecasting techniques may only be used for prediction when there is reason to believe the relationships between explanatory and response variables are causal in nature.**

# Inferring Causation

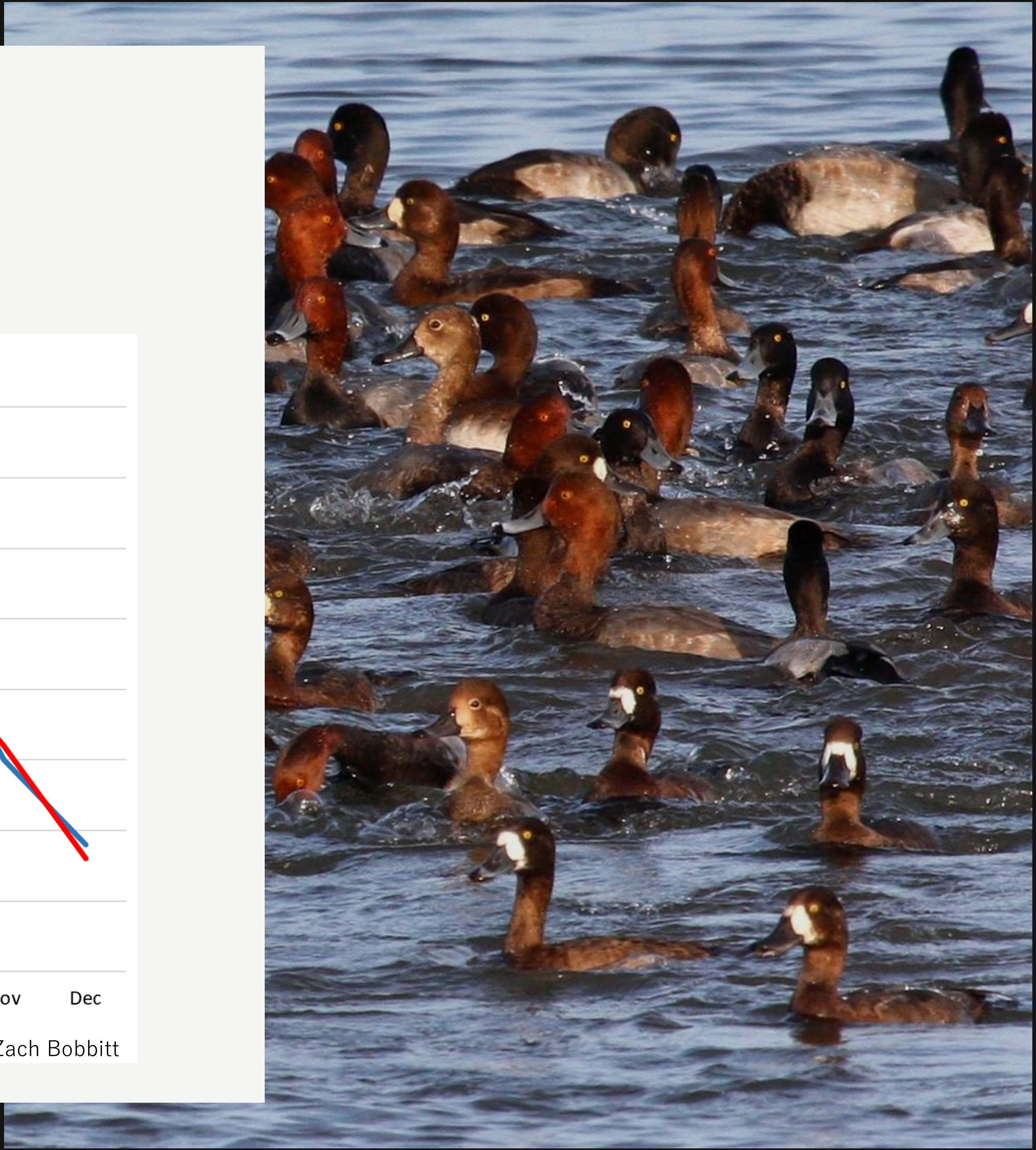


# Inferring Causation

"Correlation does not imply causation."

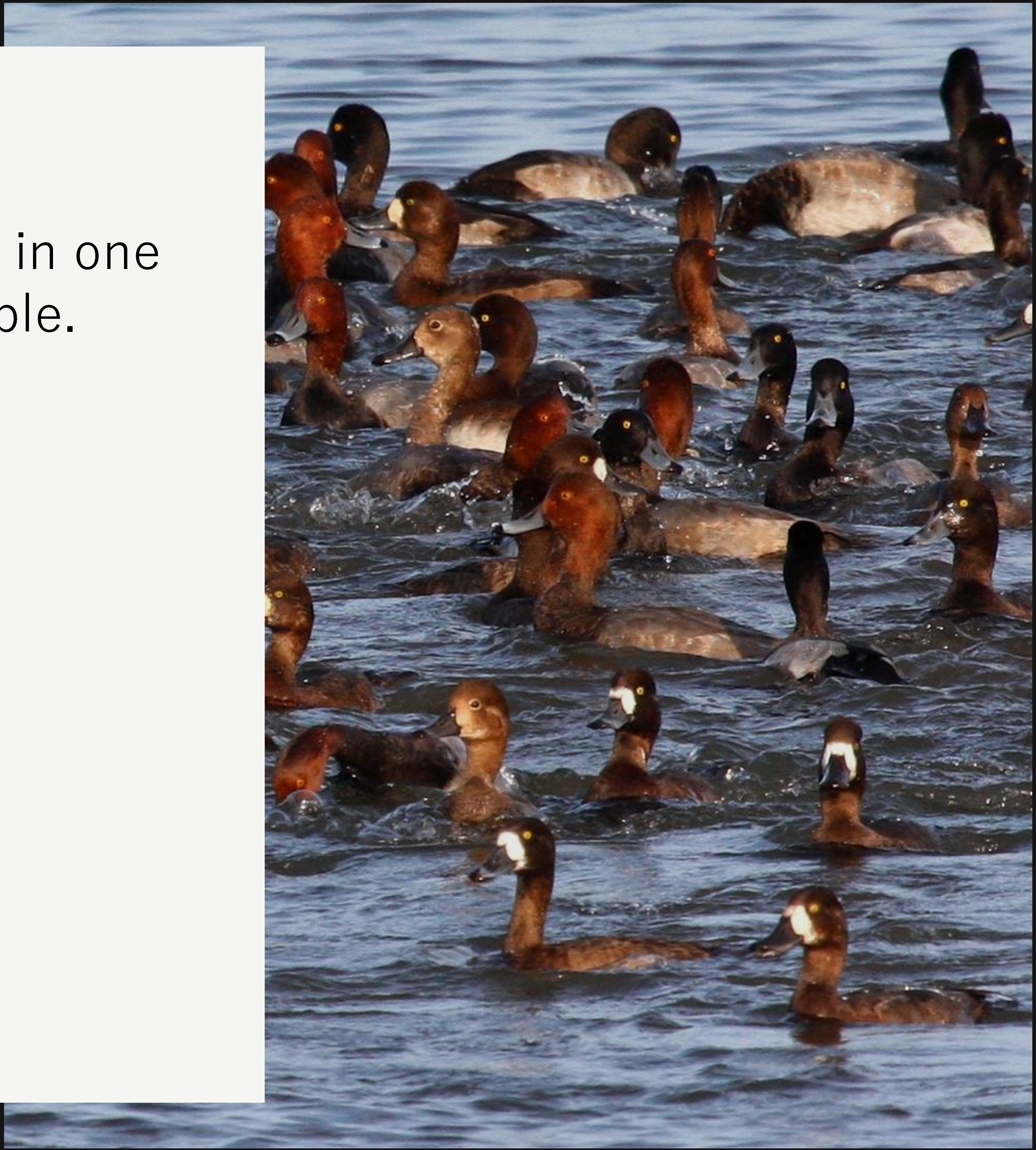



Zach Bobbitt



# Inferring Causation

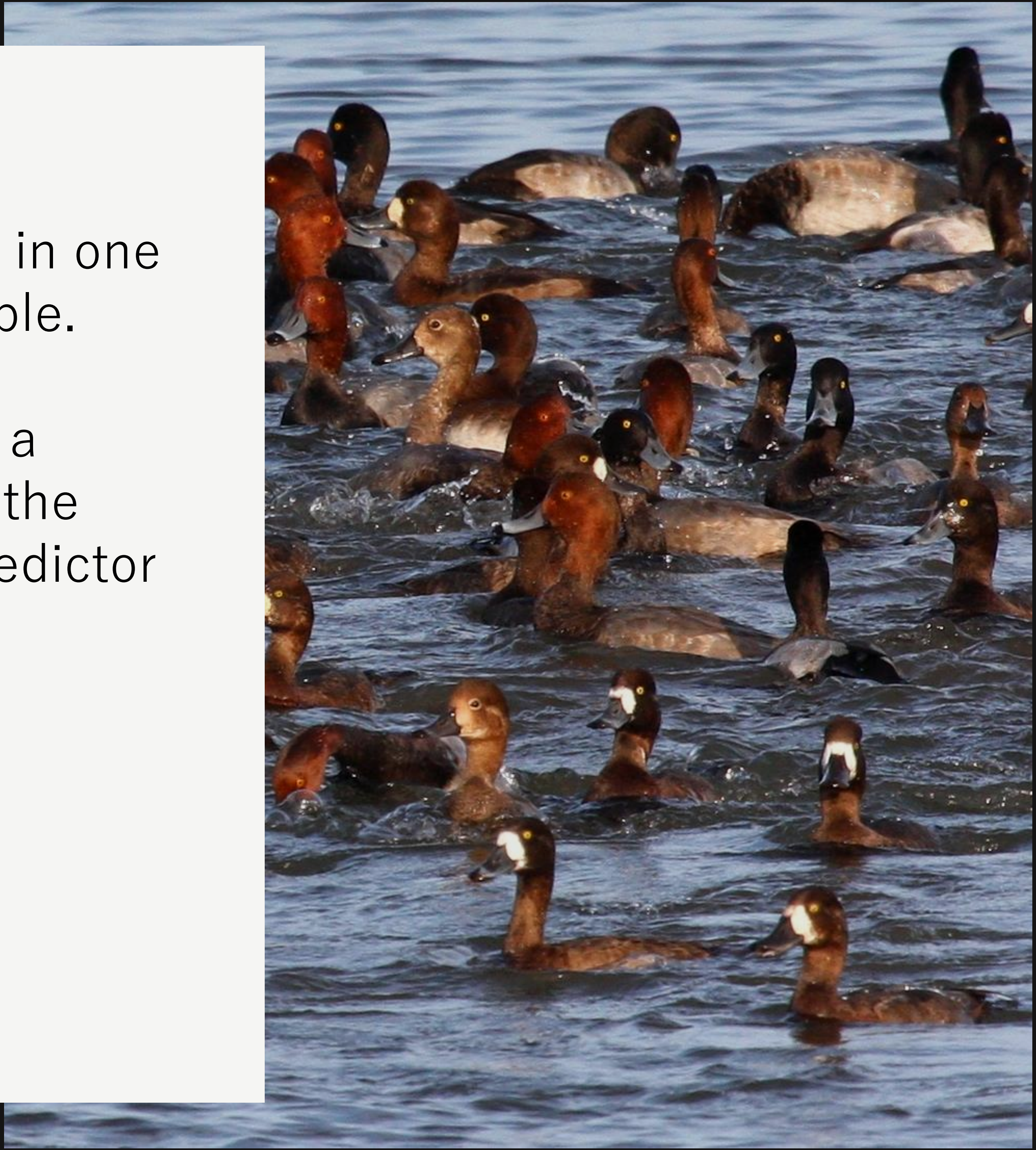
**Causality:** The hypothesis that changes in one variable cause changes in another variable.



# Inferring Causation

**Causality:** The hypothesis that changes in one variable cause changes in another variable.

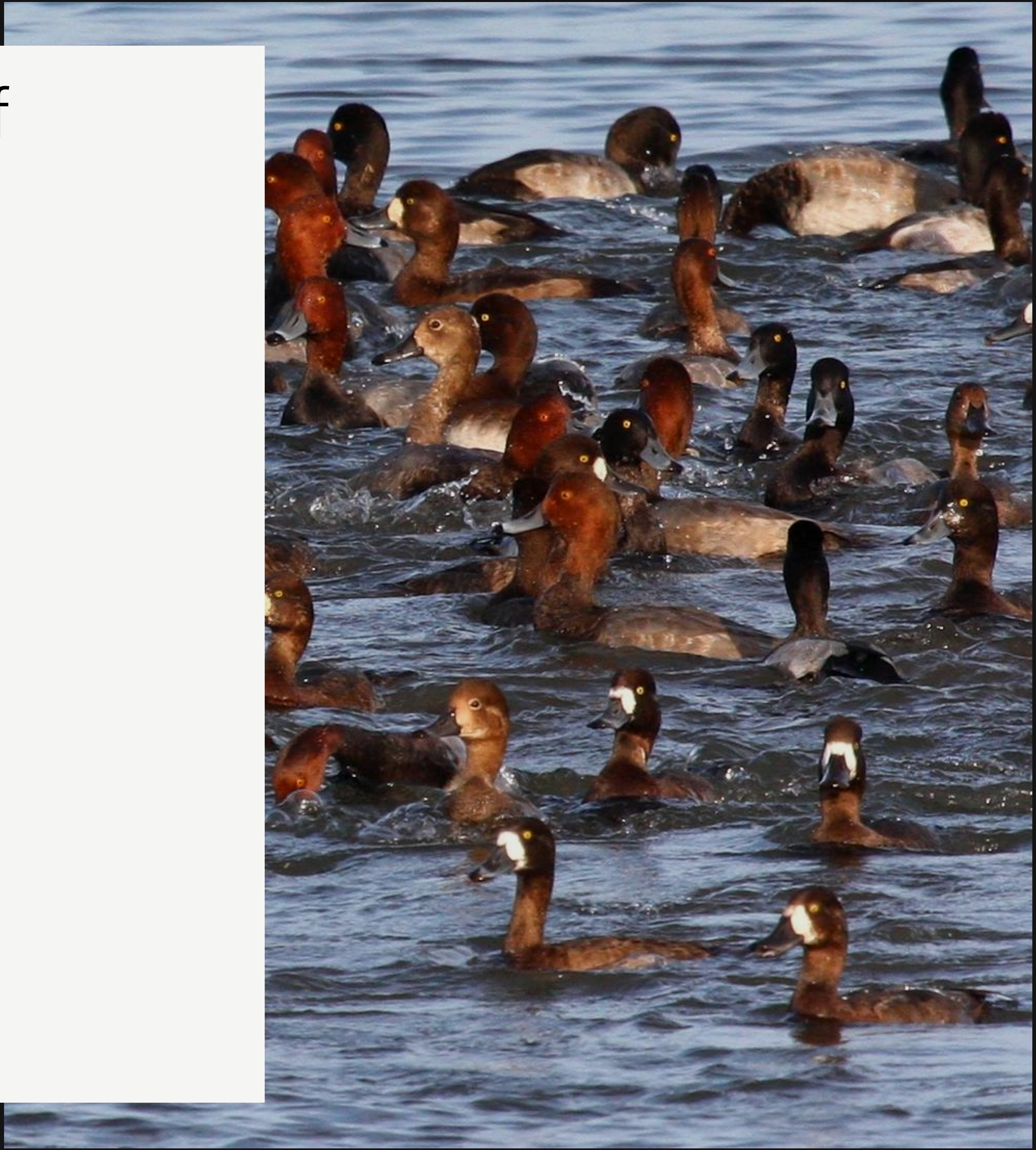
The causality hypothesis is supported if a significant proportion of the variation in the response variable is explained by the predictor variable (correlation coefficient; $R^2$).



# Inferring Causation: Models of Causality

Causal Chain

**A**
**↓**
**B**
**↓**
**C**



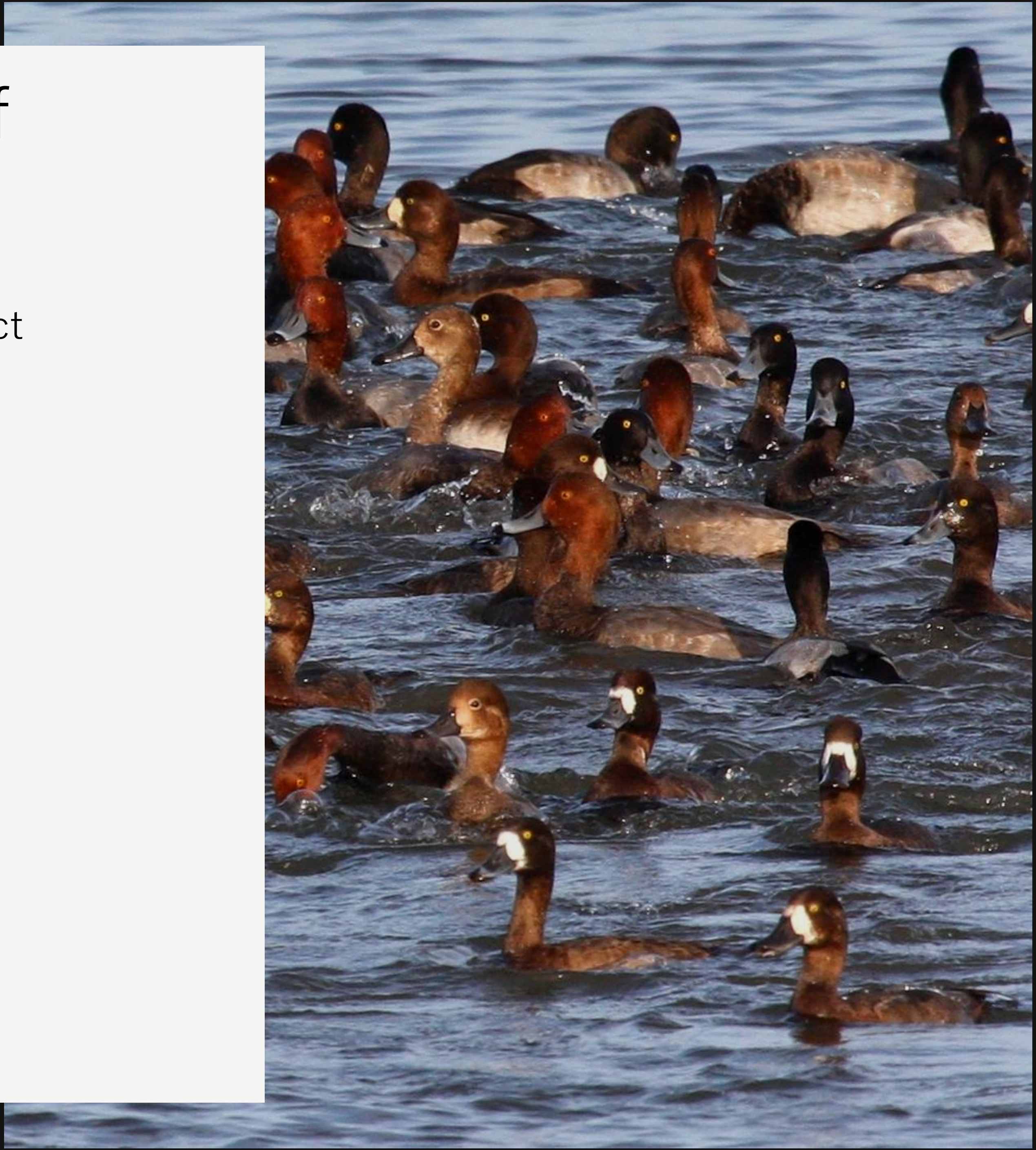
# Inferring Causation: Models of Causality

Causal Chain

$$A \rightarrow B \rightarrow C$$

Double Effect

$$A \leftarrow B \rightarrow C$$



# Inferring Causation: Models of Causality

Causal Chain

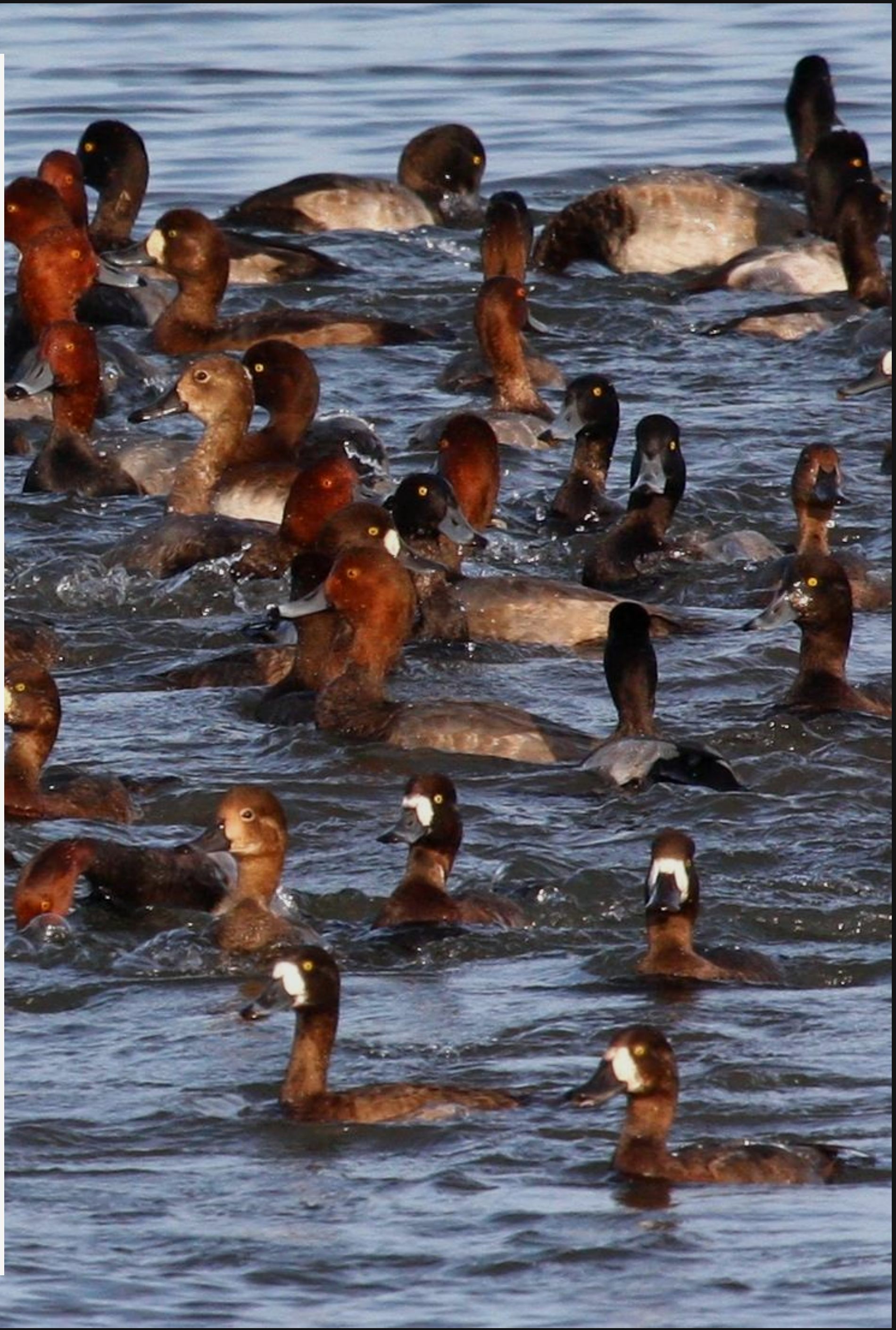
$$A \rightarrow B \rightarrow C$$

Double Effect

$$A \leftarrow B \rightarrow C$$

Double Cause

$$A \rightarrow B \leftarrow C$$



# Inferring Causation: Models of Causality

### Causal Chain

$$\begin{tikzcd} A \arrow[d] \\ B \arrow[d] \\ C \end{tikzcd}$$

### Double Effect

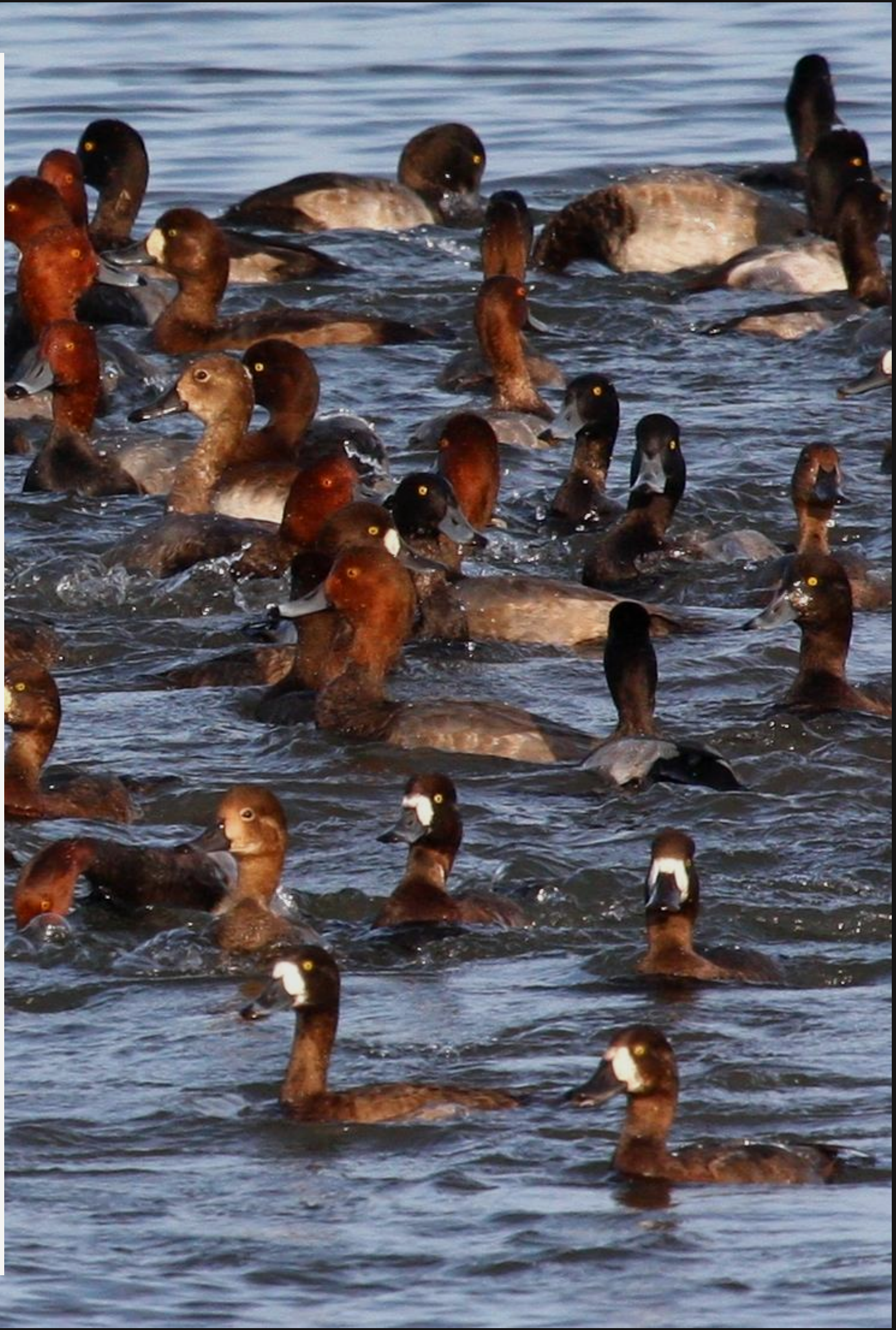
$$\begin{tikzcd} & B \arrow[dl] \arrow[dr] & \\ A & & C \end{tikzcd}$$

### Double Cause

$$\begin{tikzcd} A \arrow[dr] & & C \arrow[dl] \\ & B & \end{tikzcd}$$

### Triangular

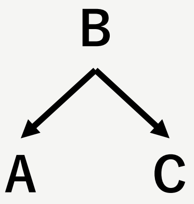
$$\begin{tikzcd} A \arrow[rr] \arrow[dr] & & C \arrow[dl] \\ & B & \end{tikzcd}$$



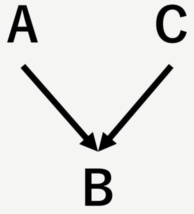
# Inferring Causation: Models of Causality

Causal Chain

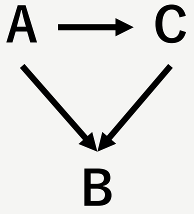

Double Effect

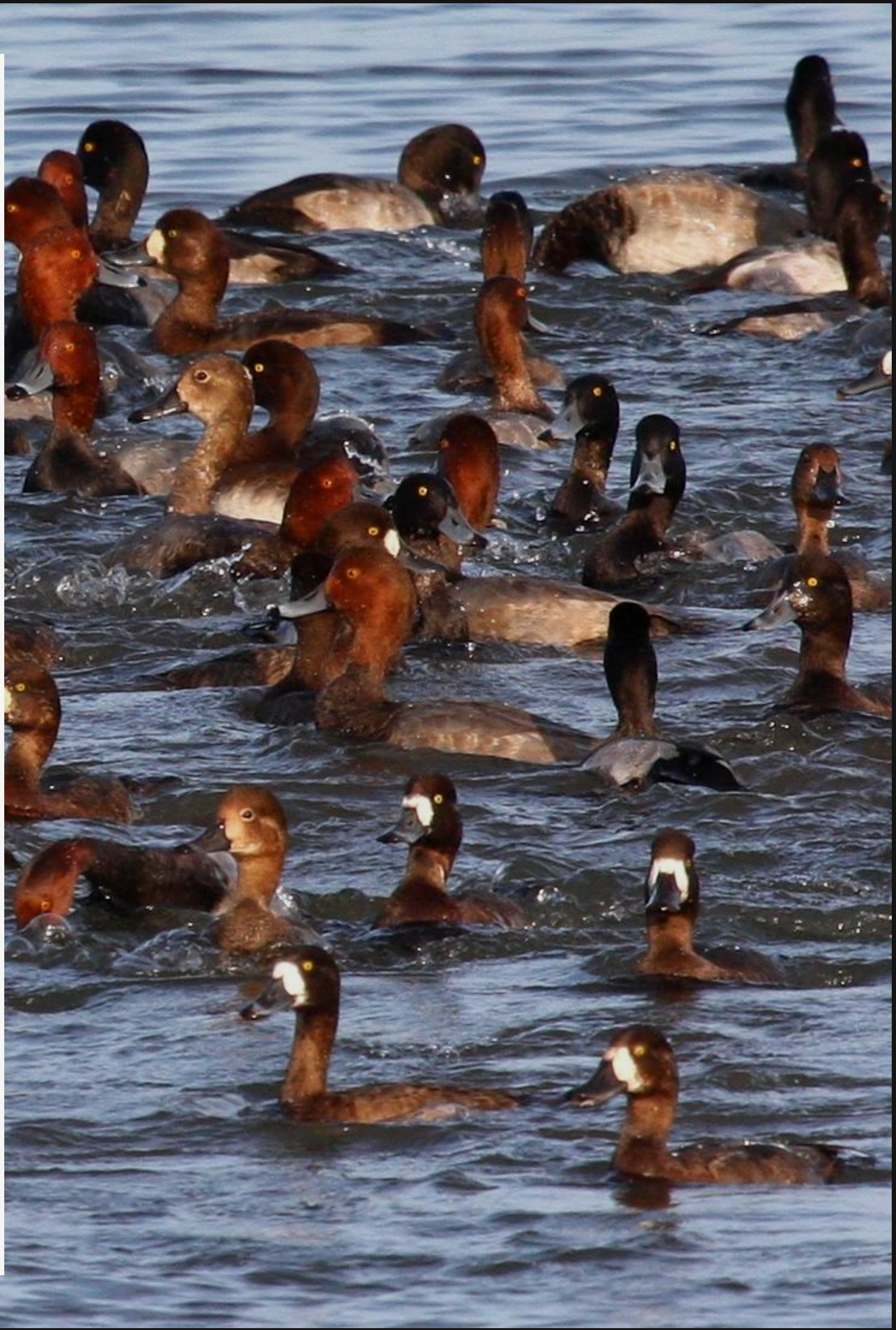



Double Cause




Triangular



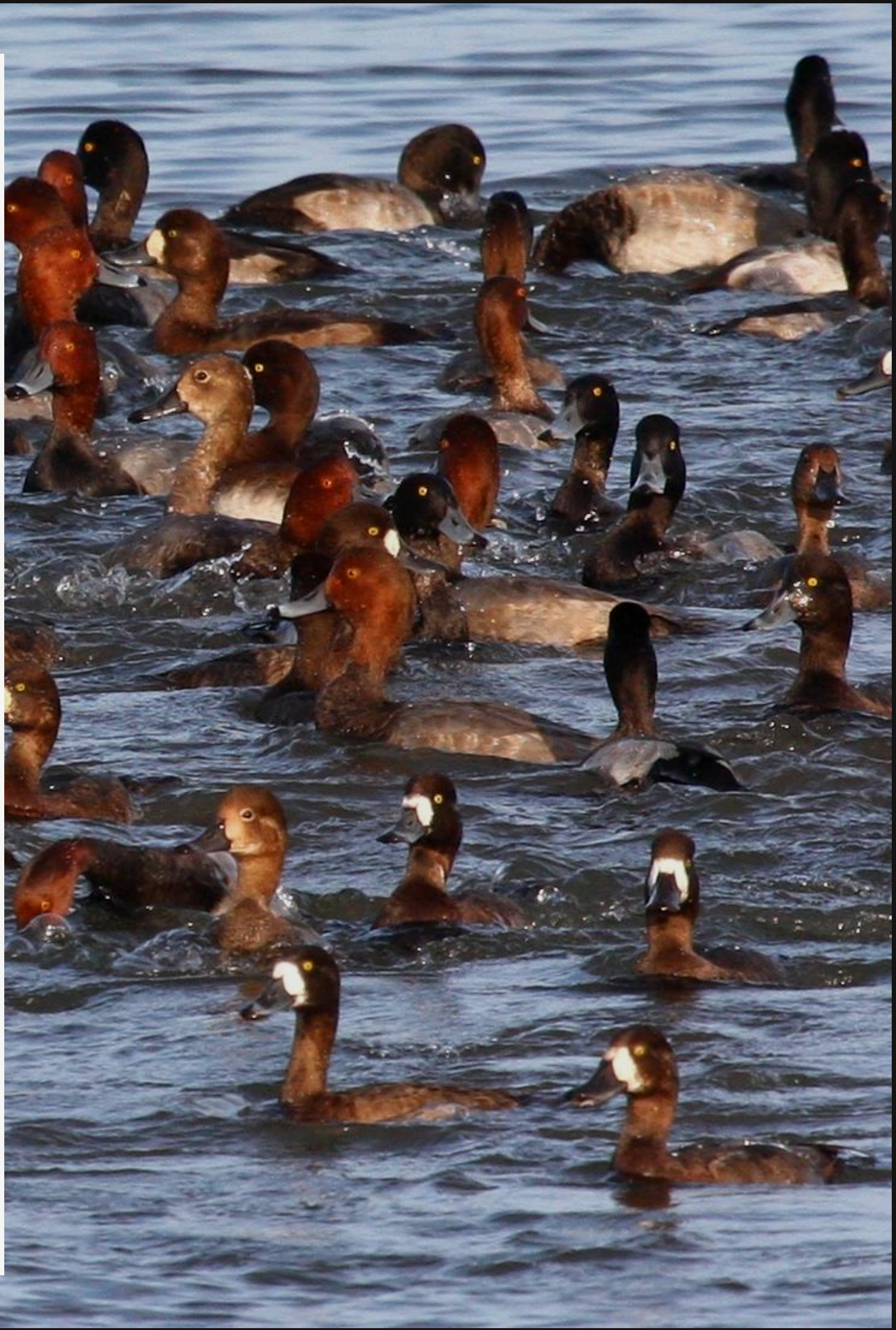



# Conclusion: Summary of Key Points

- **Interpretation** links patterns to ecological processes
  - Interpret principal components
  - Examine loadings
  - Identify clusters and groupings
  - Visualize using biplots

- **Inference** is used to test hypotheses or identify key explanatory variables
  - **Explanatory** vs. **forecasting** vs. **predictive** analyses
  - **Indirect** vs. **Direct Comparison** (gradient analysis)

- Use **simple** and **partial correlation coefficients** to infer causality
  - Even so, structures can be unclear



# Questions?