

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

Term: Fall Quarter, 2024

Time: Tuesday/Thursday 2:00 pm – 3:50 pm

Location: Nash 032

Instructor: Melanie Davis (melanie.davis@oregonstate.edu)

Course Website:

<https://canvas.oregonstate.edu/courses/1977576>



Thursday, September 26, 2024

- Class introductions
- Go through syllabus, schedule, assignments, expectations for the quarter
- Demo course website, class data sets
- R Markdown/Quarto language and resources
- Other course resources
- Lecture: Intro to Multivariate Analysis



Who are you?

- Name (preferred)
- Department
- Major advisor
- Research project
- One mystery you're hoping to unravel during class
- One fun fact about yourself



Who are you?

- Name (preferred)
- Department
- Major advisor
- Research project
- One mystery you're hoping to unravel during class
- One fun fact about yourself

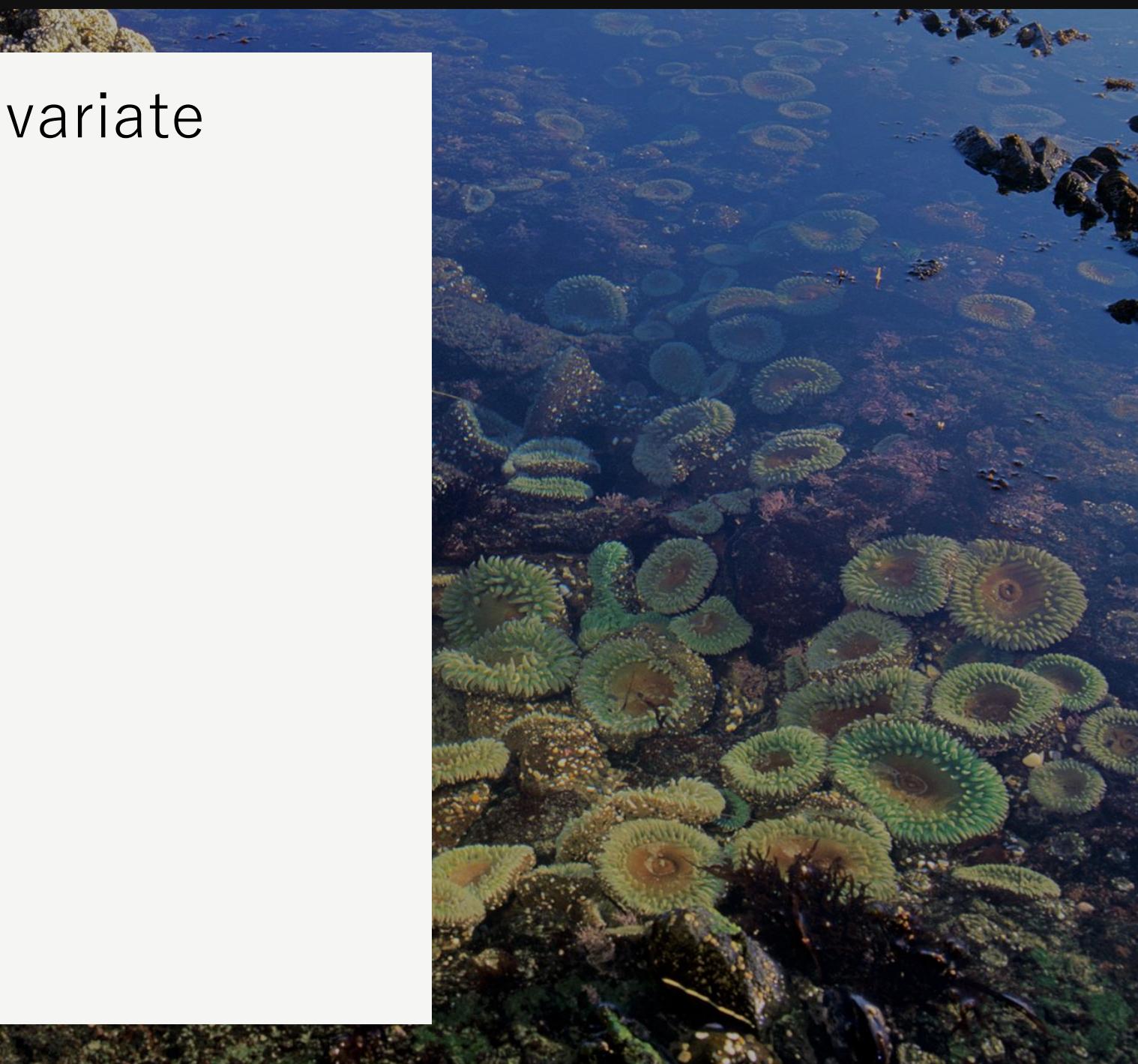


Course Resources

- Canvas: <https://canvas.oregonstate.edu/courses/1977576>
- R Markdown/Quarto: <https://quarto.org/docs/guide/>
- Textbooks: On Canvas!



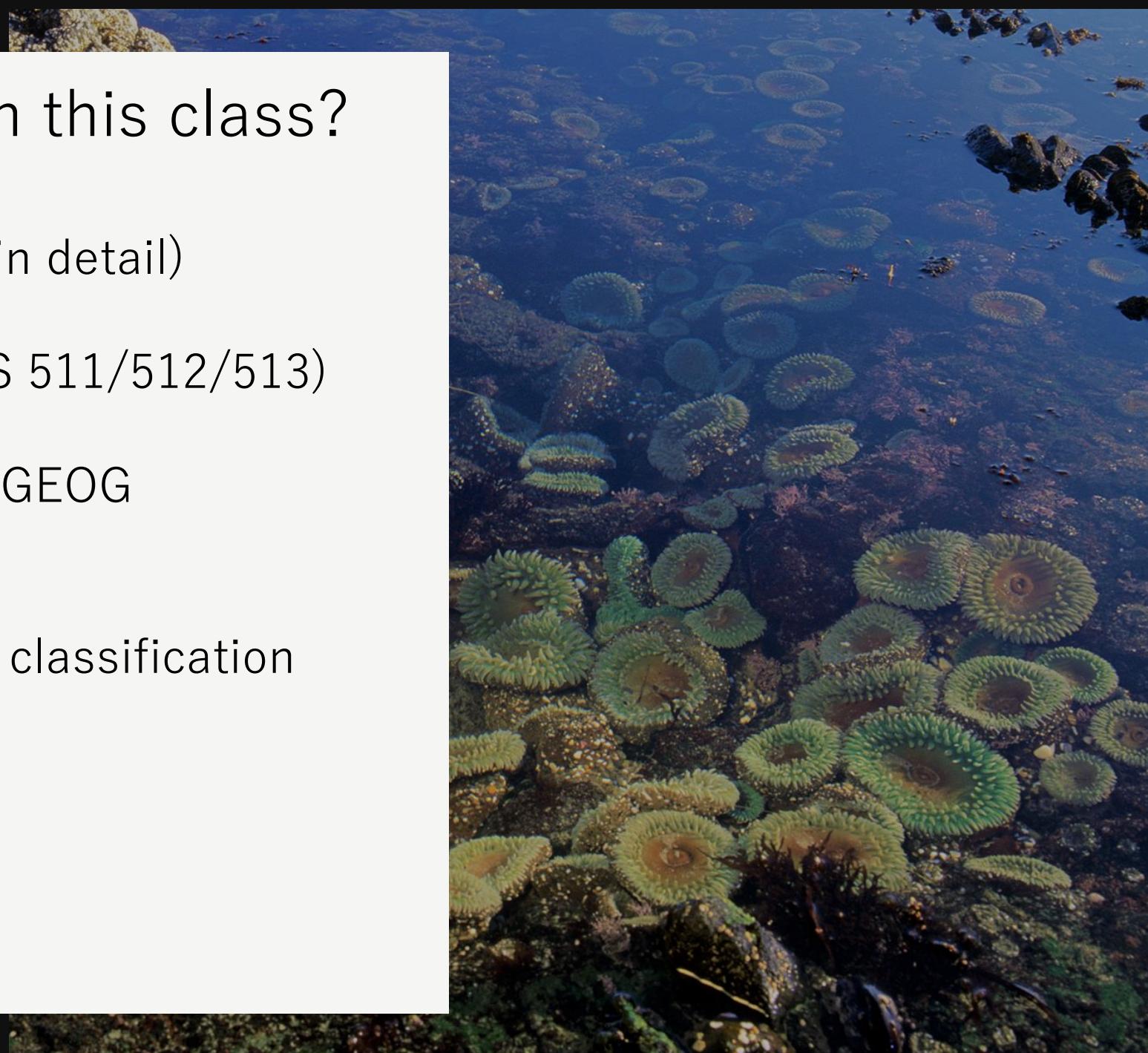
Lecture 1: Intro to Multivariate Analysis



Lecture 1: Intro to Multivariate Analysis

- What is it?
- Why is it important?
- How do ecologists use it?



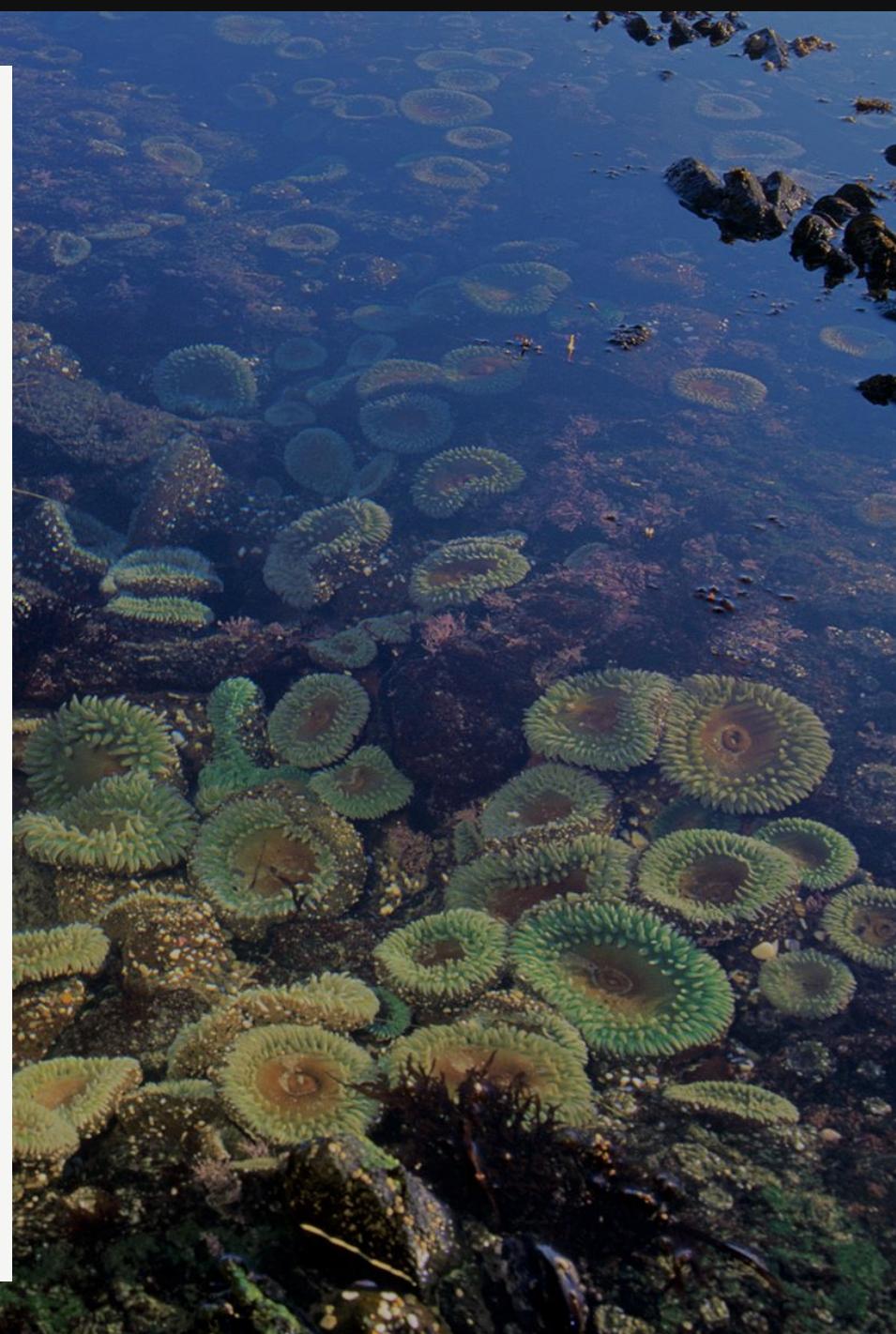


What is NOT covered in this class?

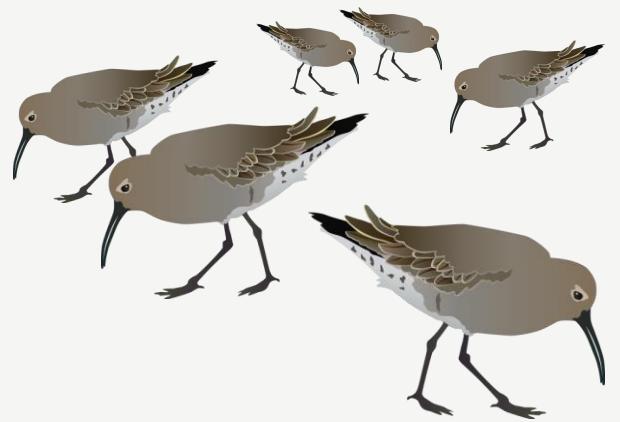
- Study design (at least, not in detail)
- Univariate methods (STATS 511/512/513)
- Advanced spatial analysis (GEOG 560/561/566)
- Remote sensing and image classification (GEOG 580/581)

Multivariate Analysis:

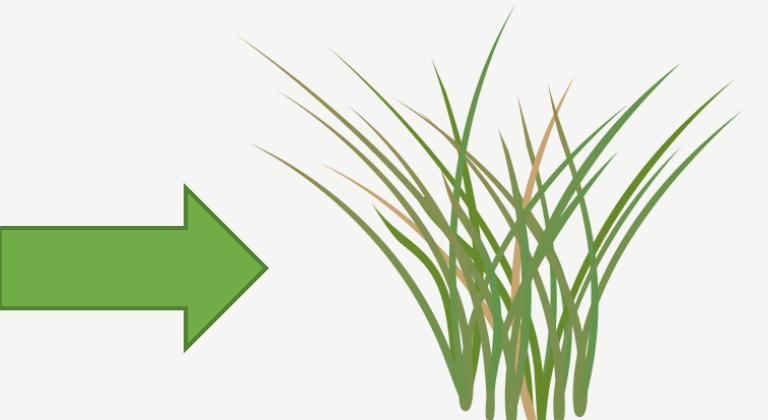
A set of techniques that is used to examine multiple variables simultaneously



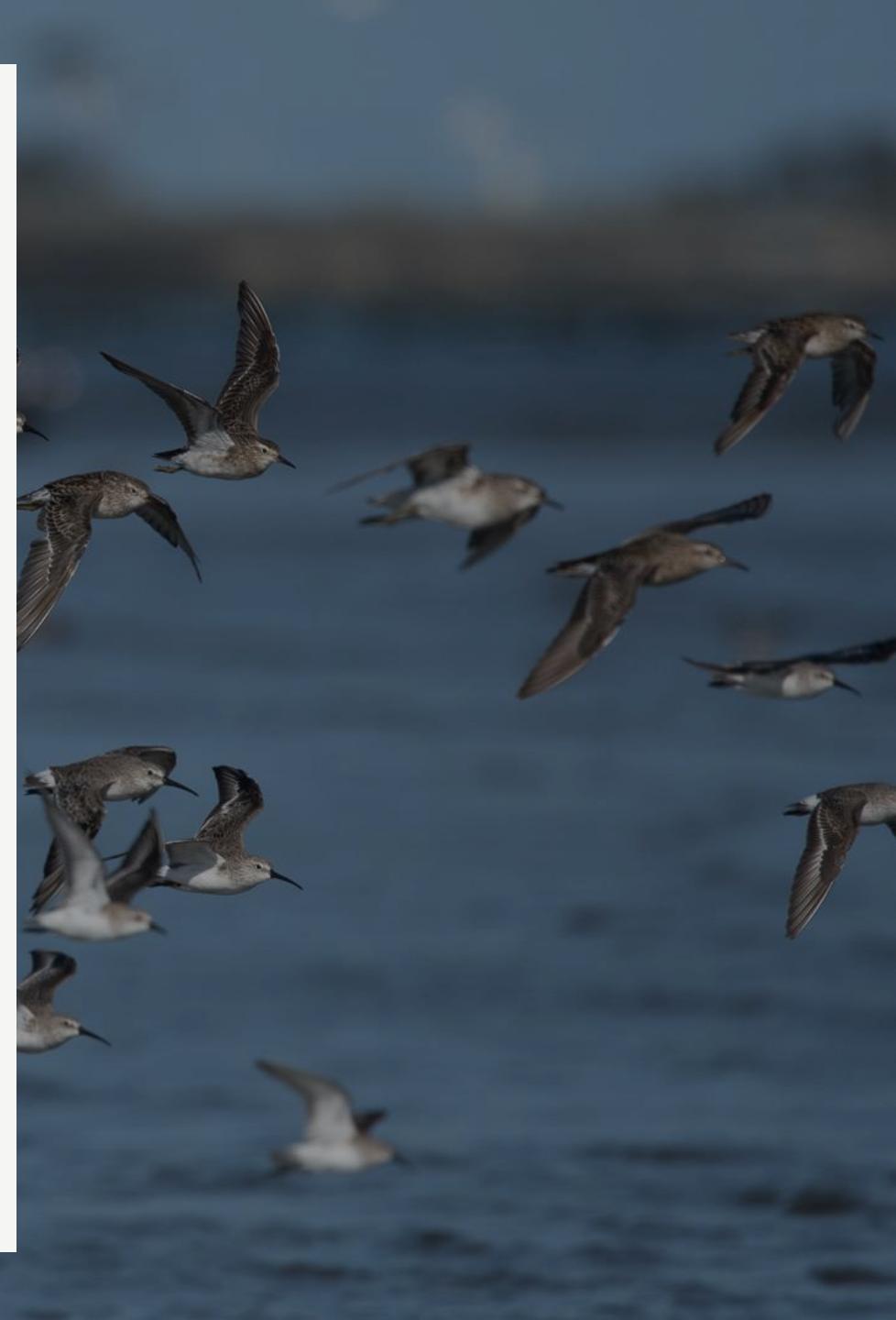
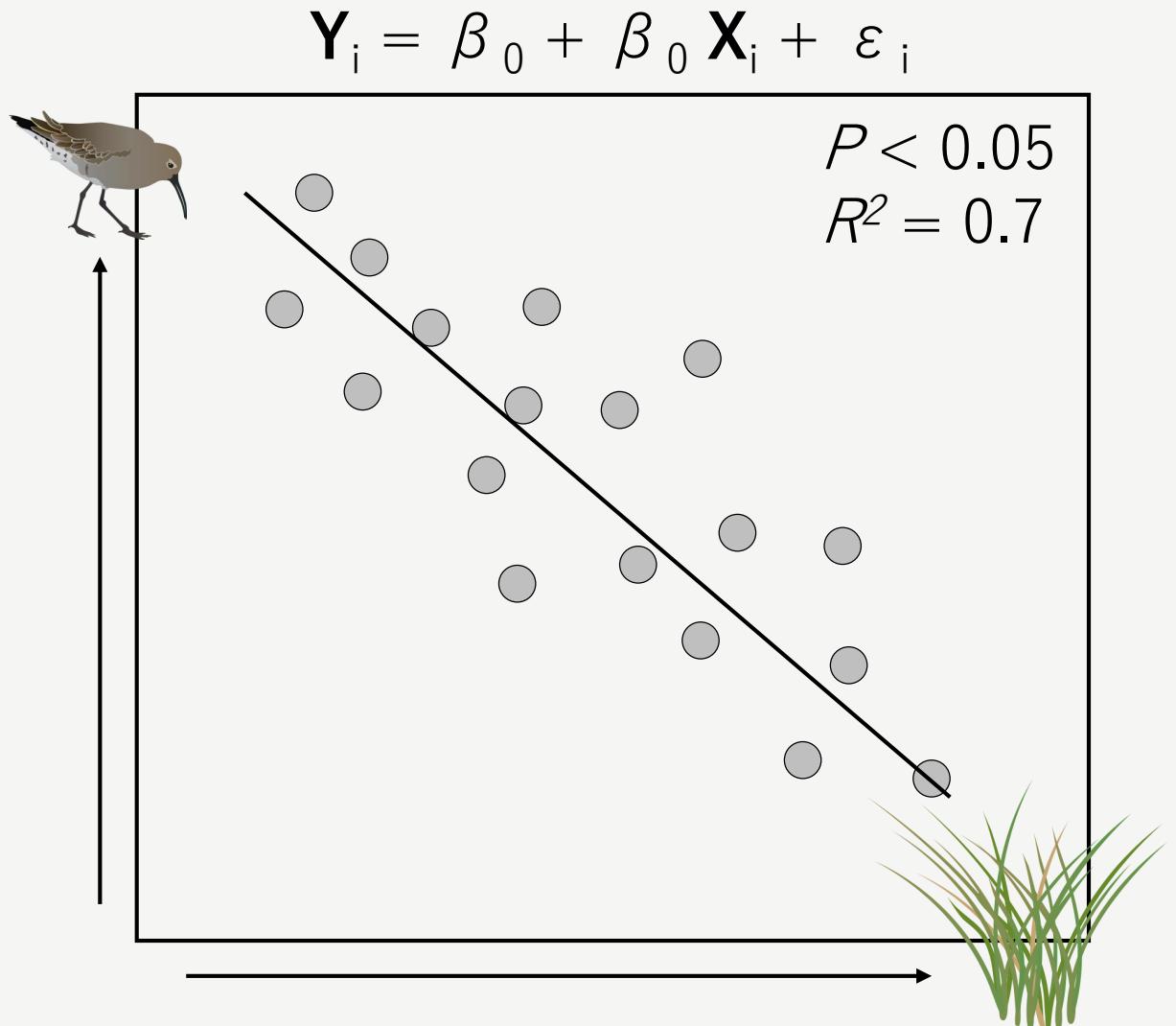
Response (**Y**)



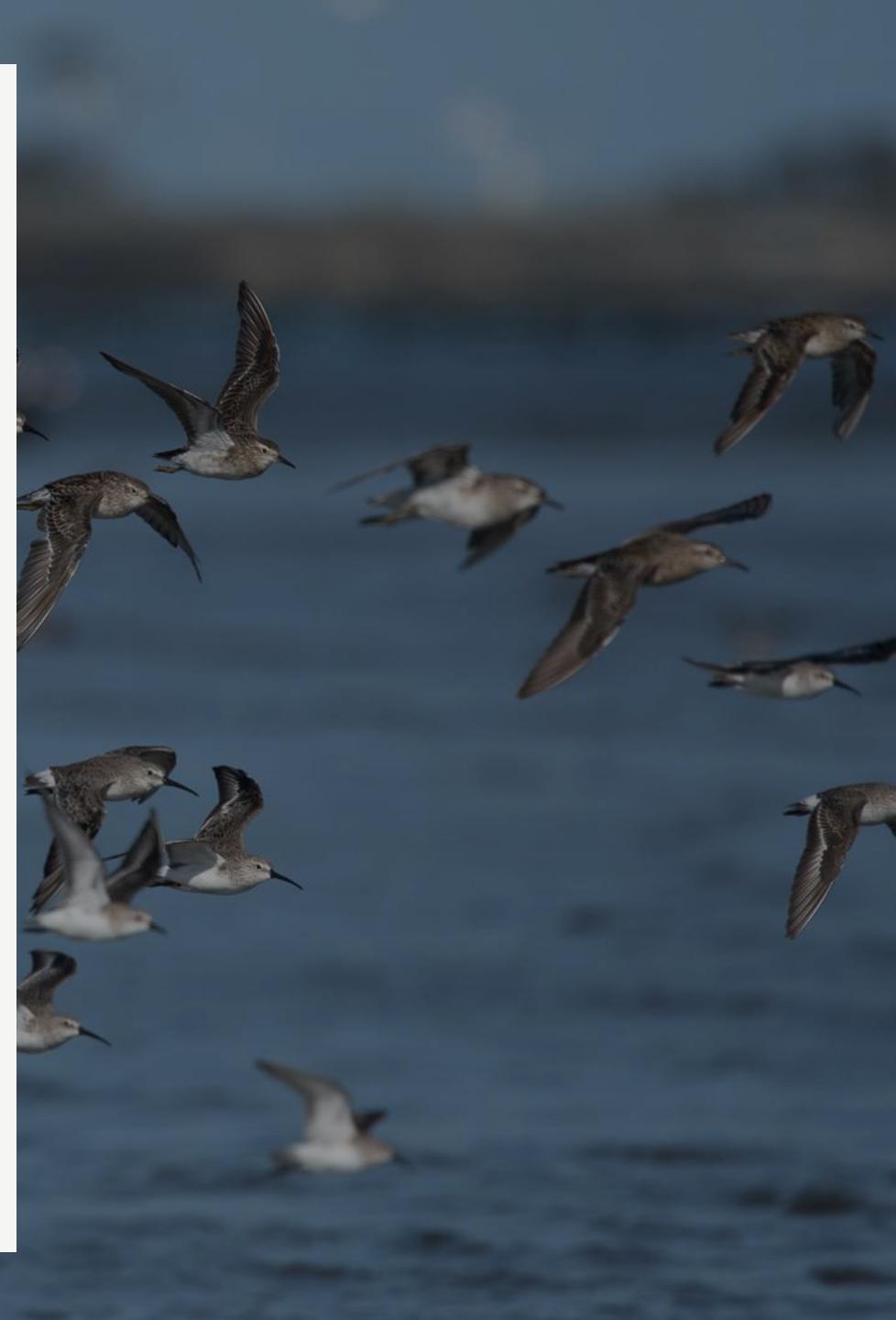
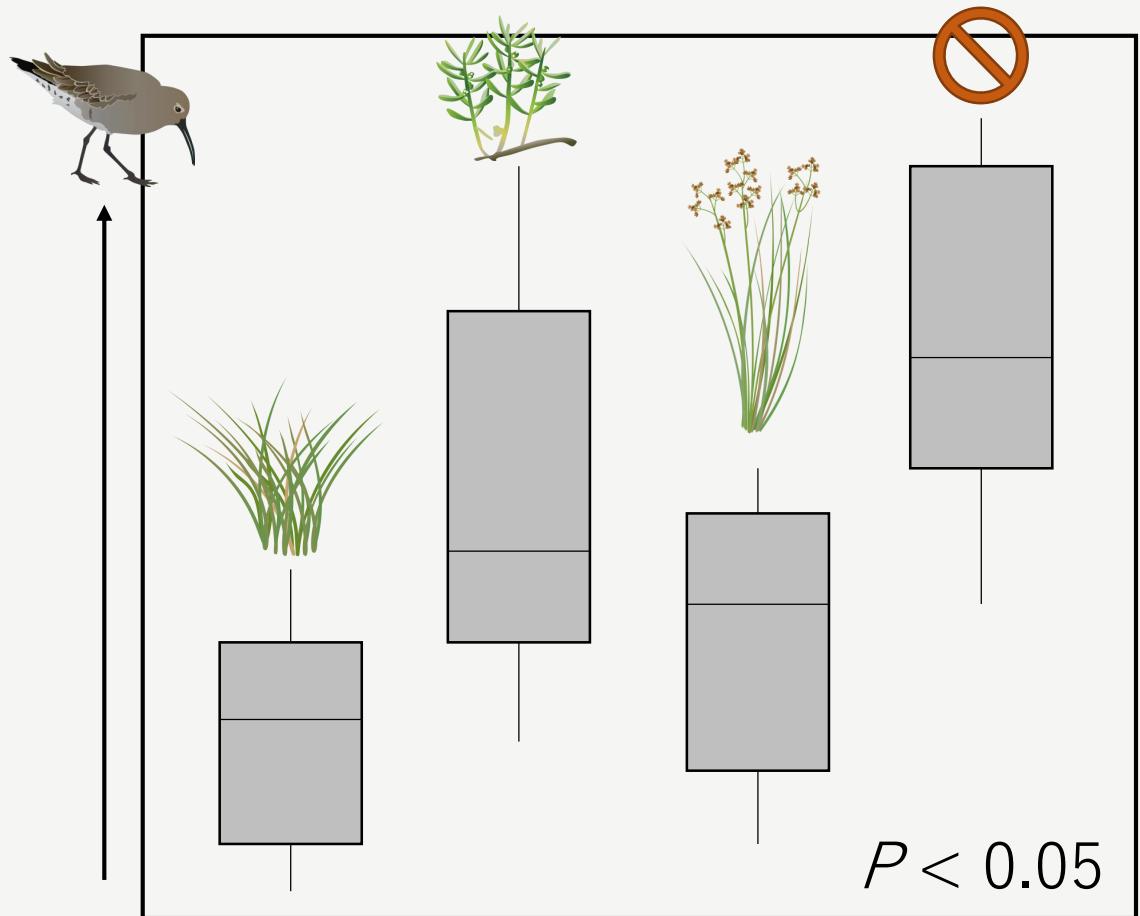
Predictor (**X**)



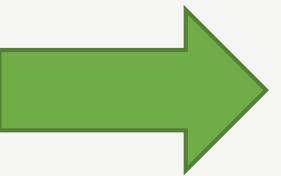
Univariate: Linear Model



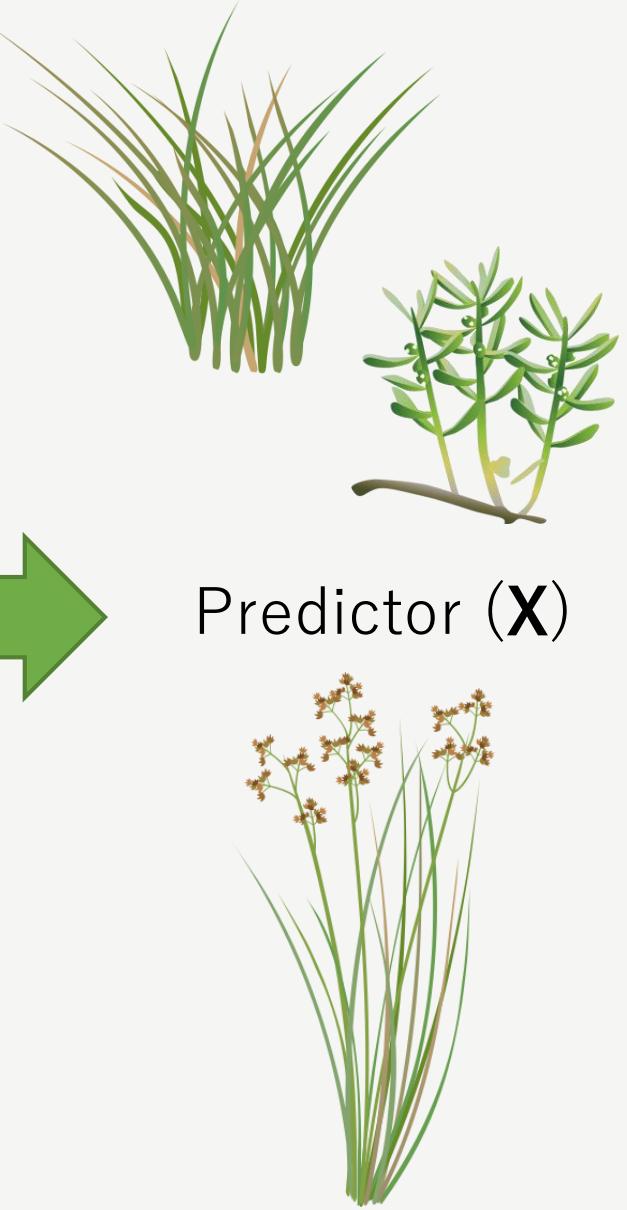
Univariate: Analysis of Variance

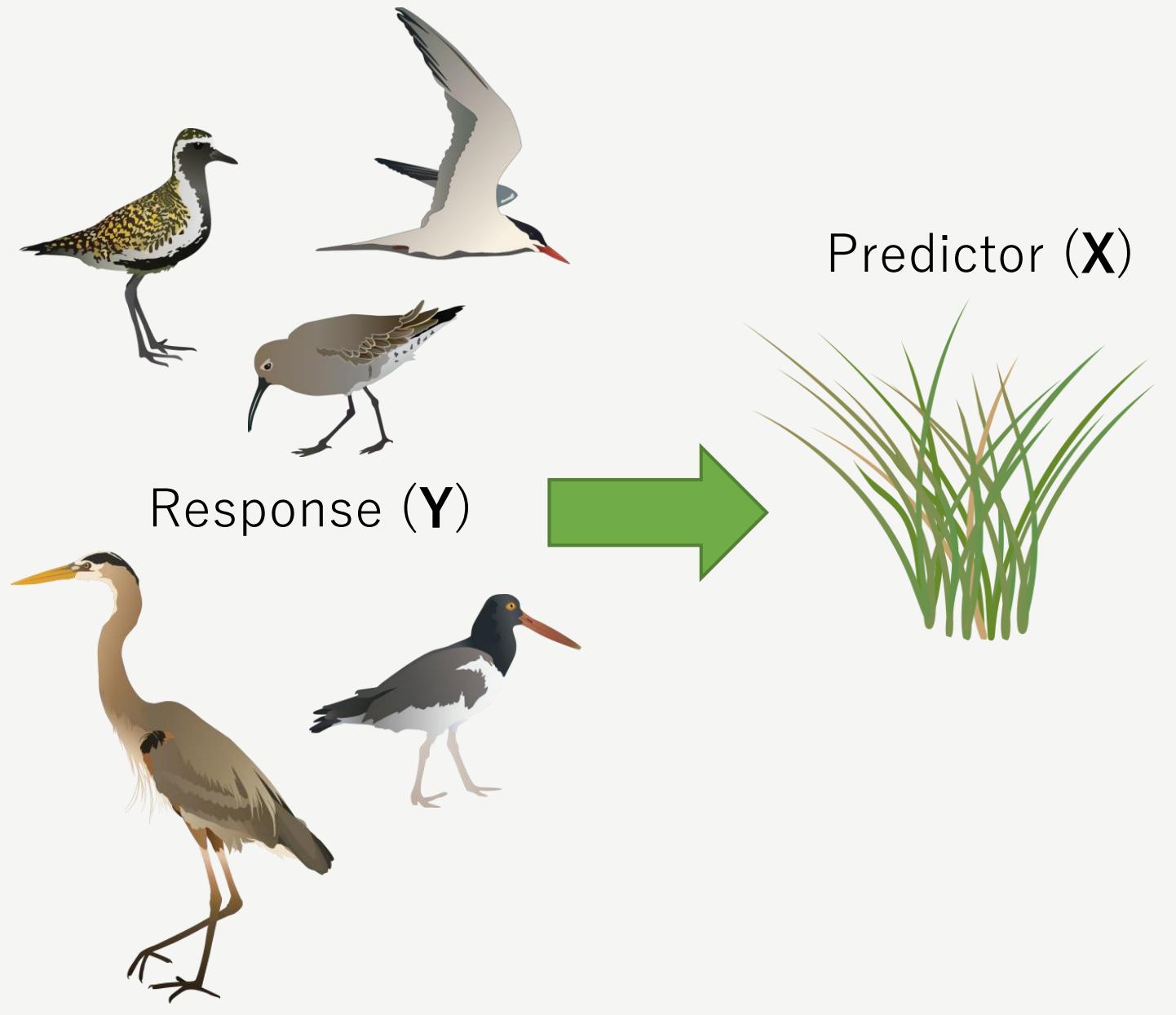


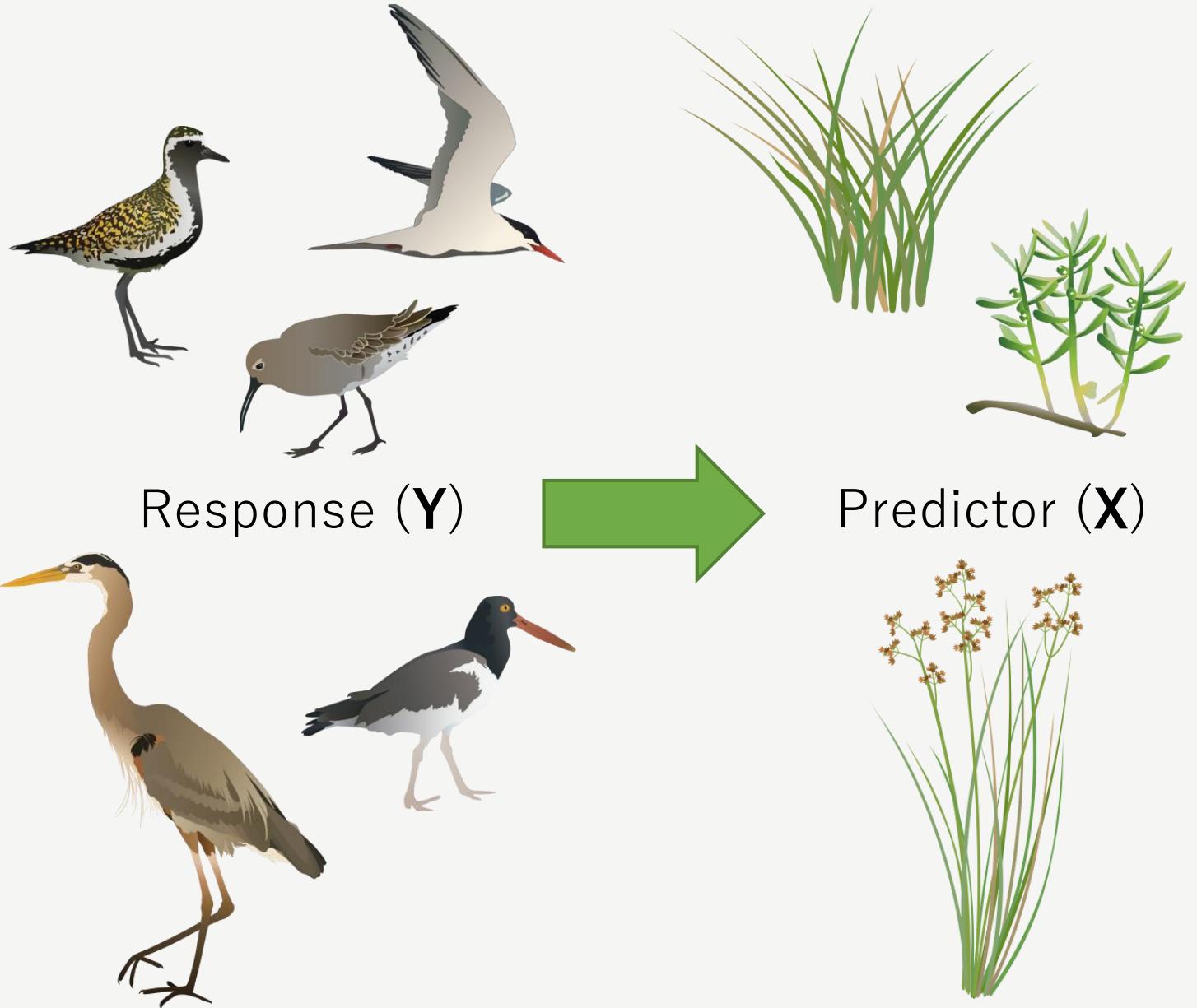
Response (**Y**)



Predictor (**X**)

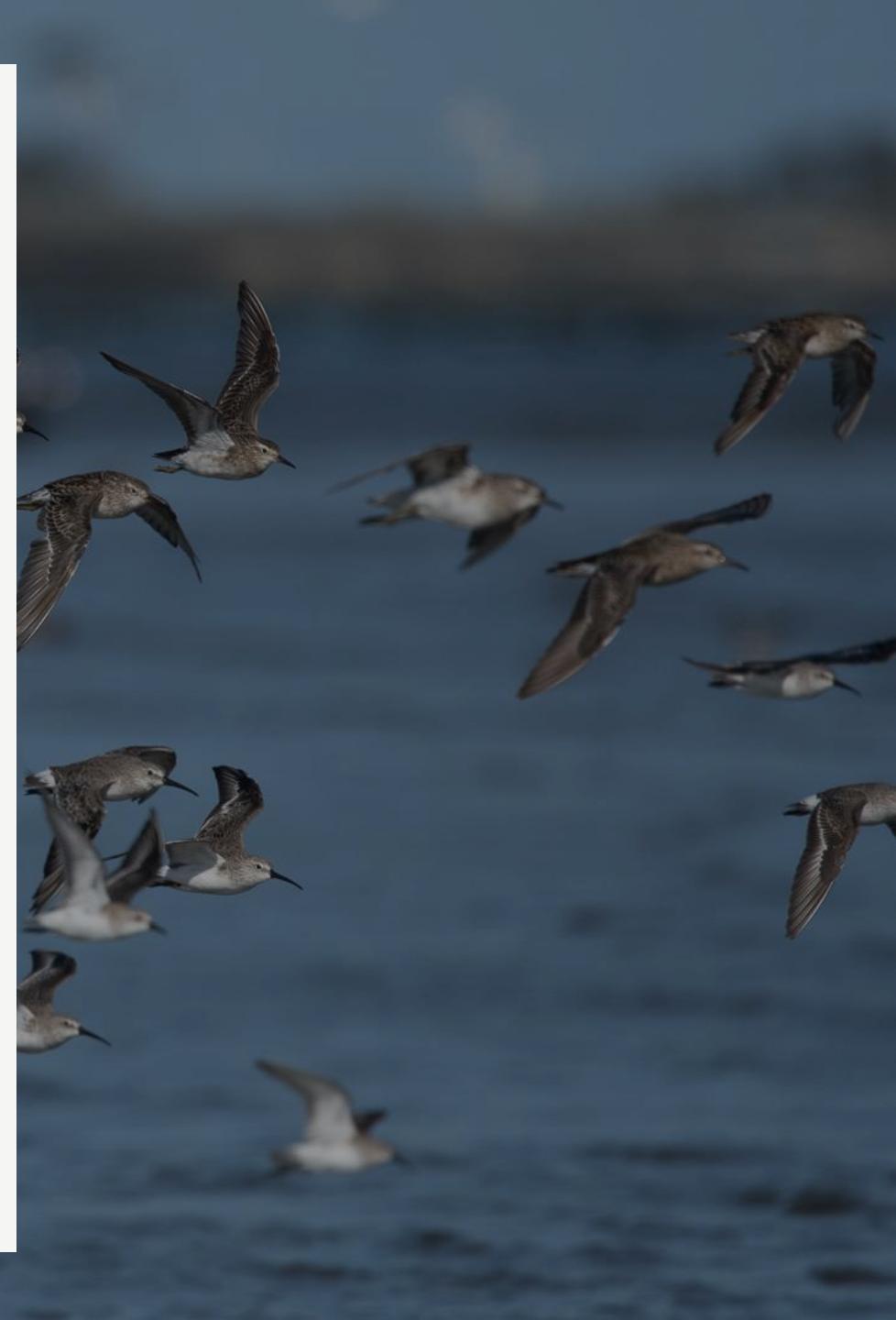
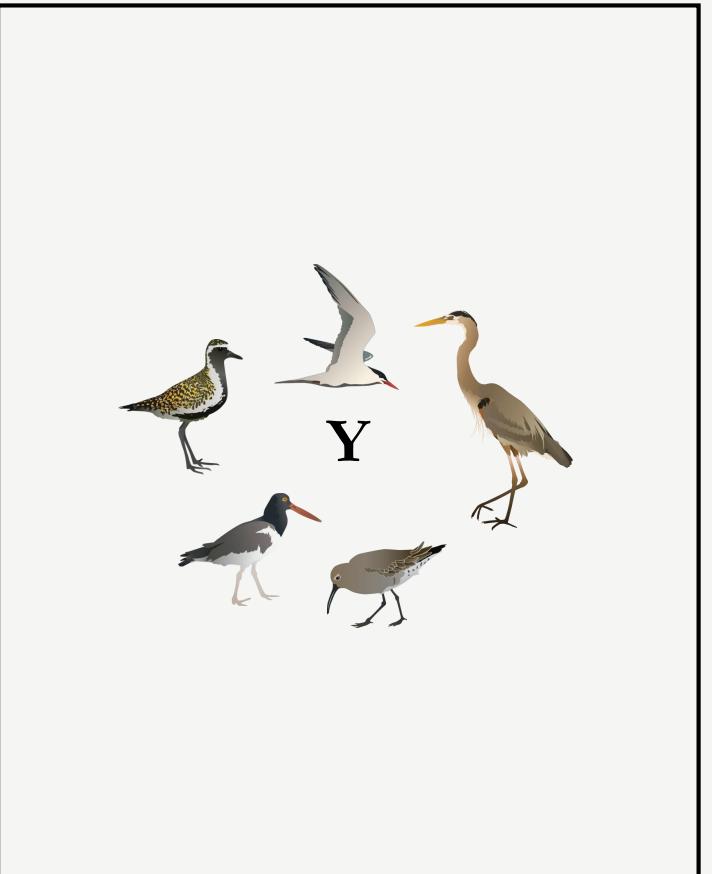






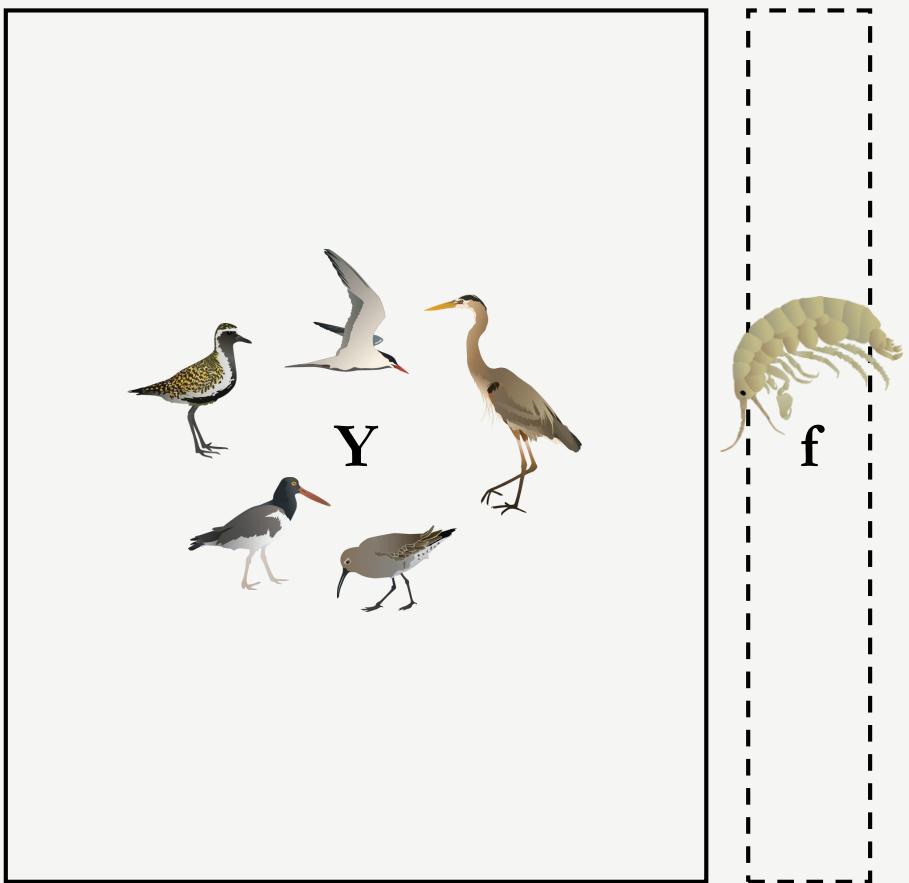
Multivariate Analysis

It all starts with a matrix...



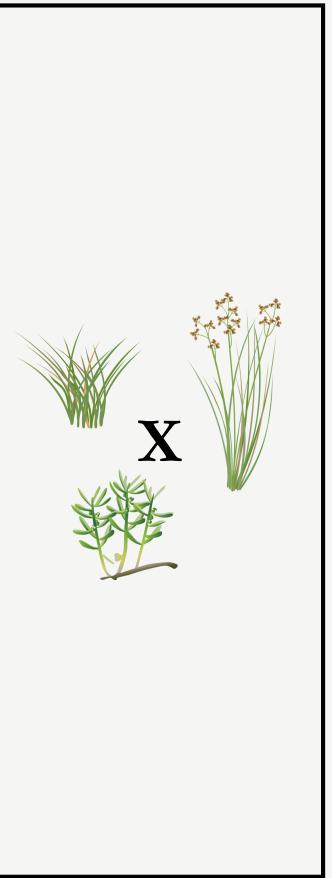
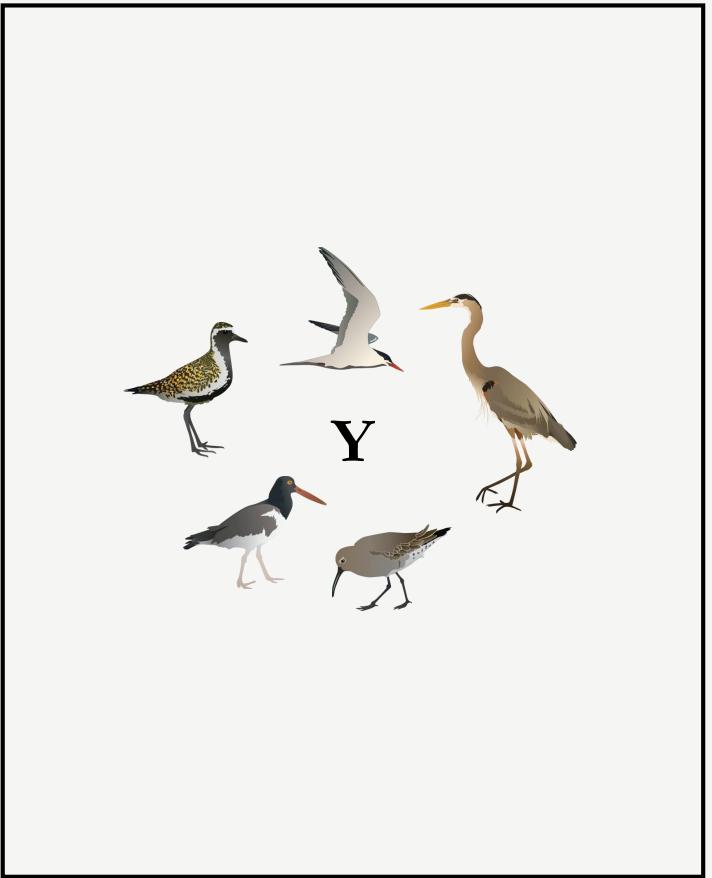
Multivariate Analysis

Then we add a latent process (or processes)...



Multivariate Analysis

Or even another matrix...

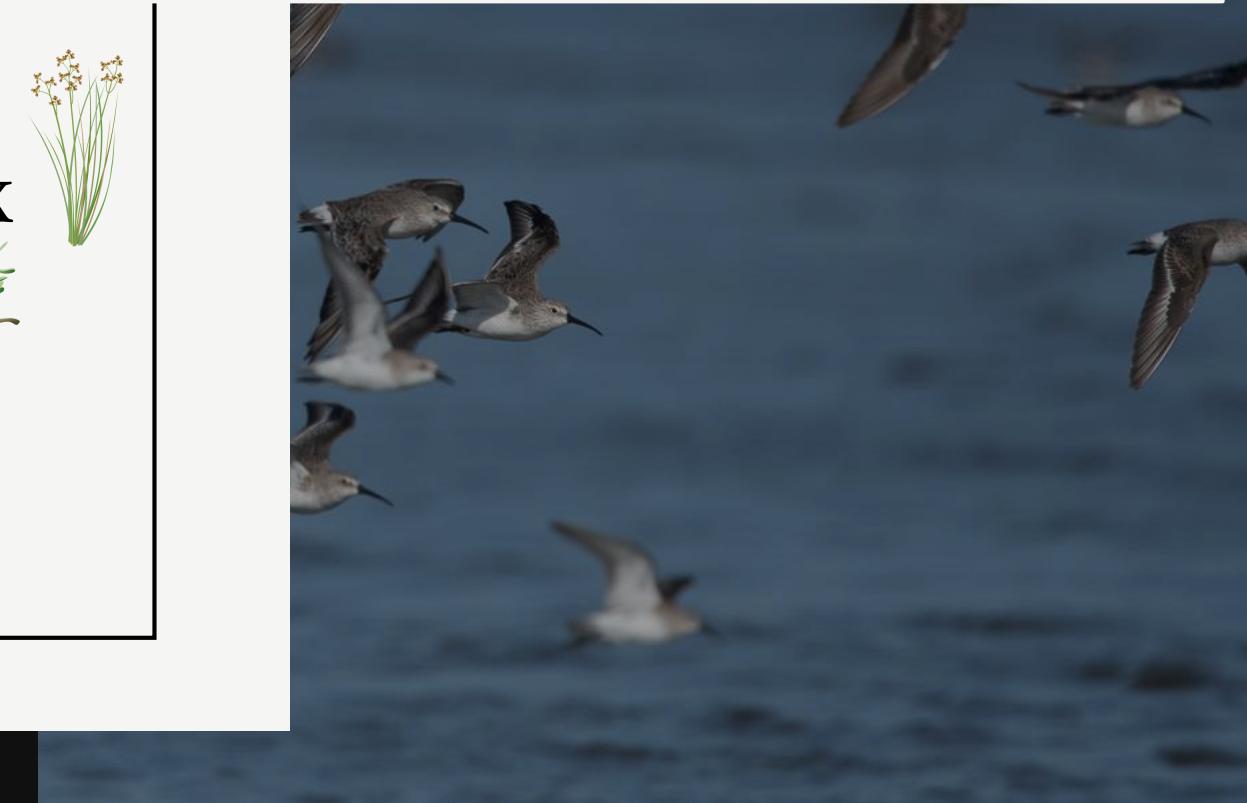
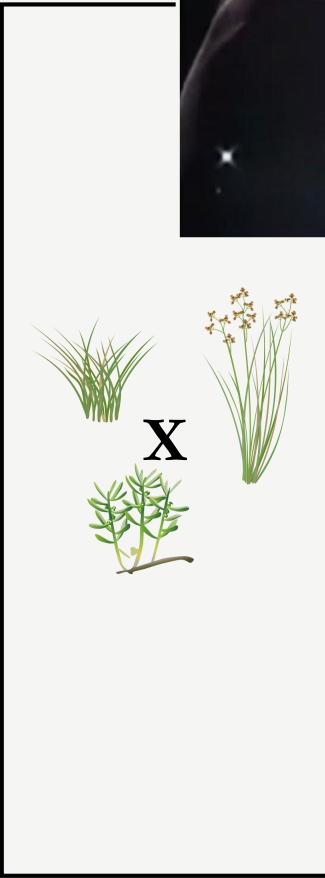
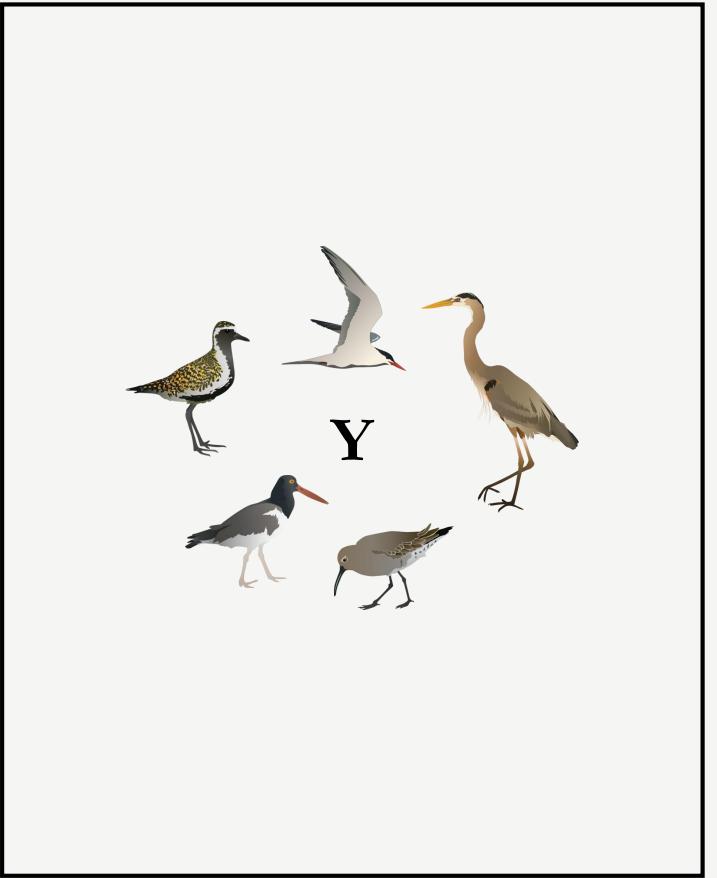
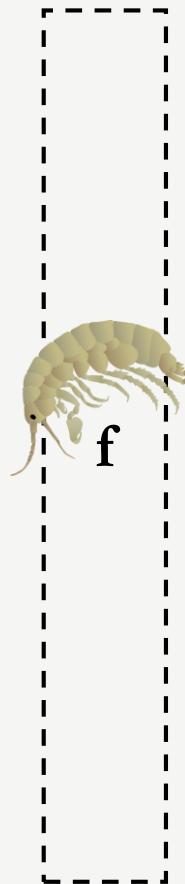


Multivariate Analysis

Lots of matrices!?!?



What do we do with all these matrices???



What do we do with all these matrices???

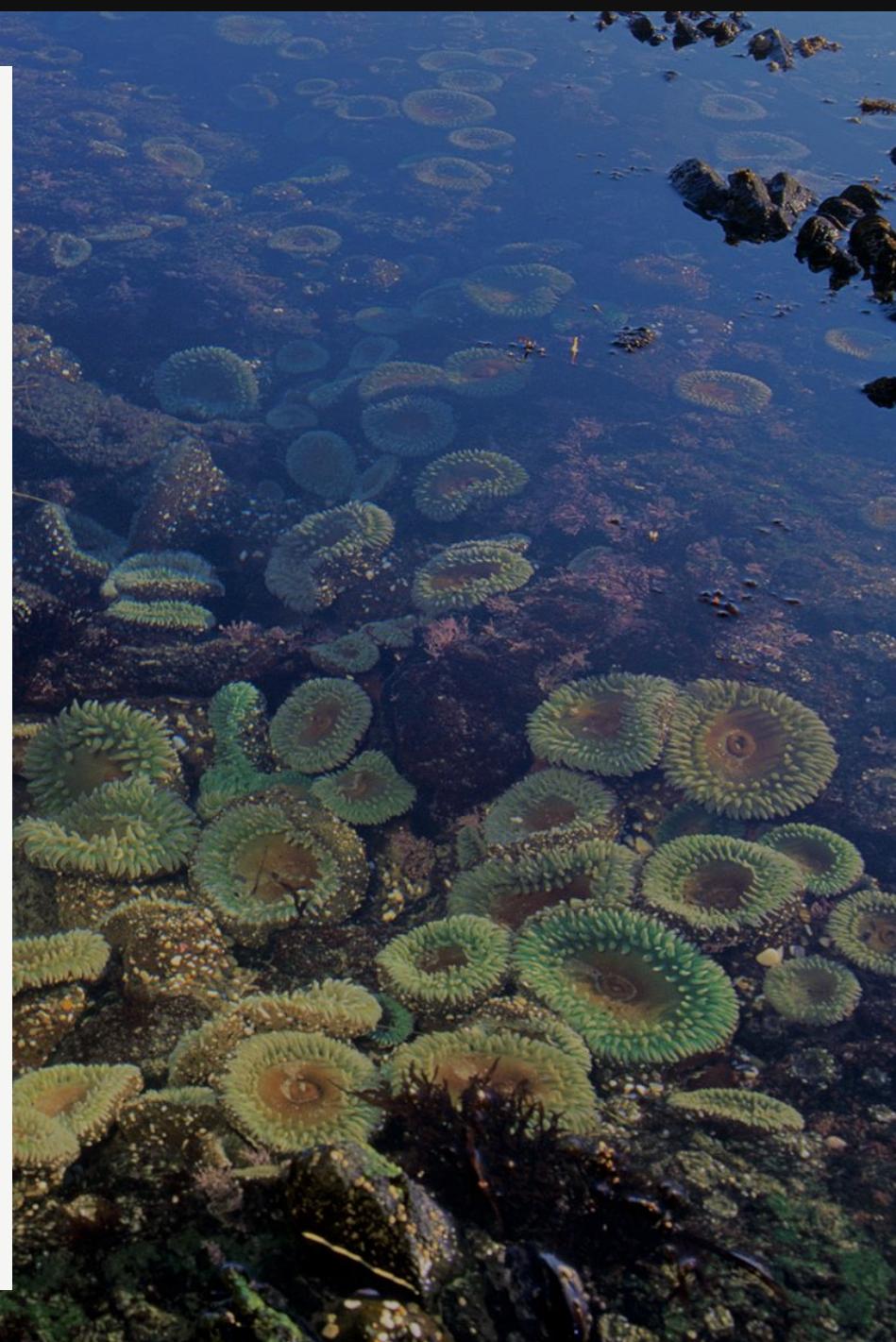


**MULTIVARIATE ANALYSIS IS
WHAT WE DO WITH ALL THESE
MATRICES!**



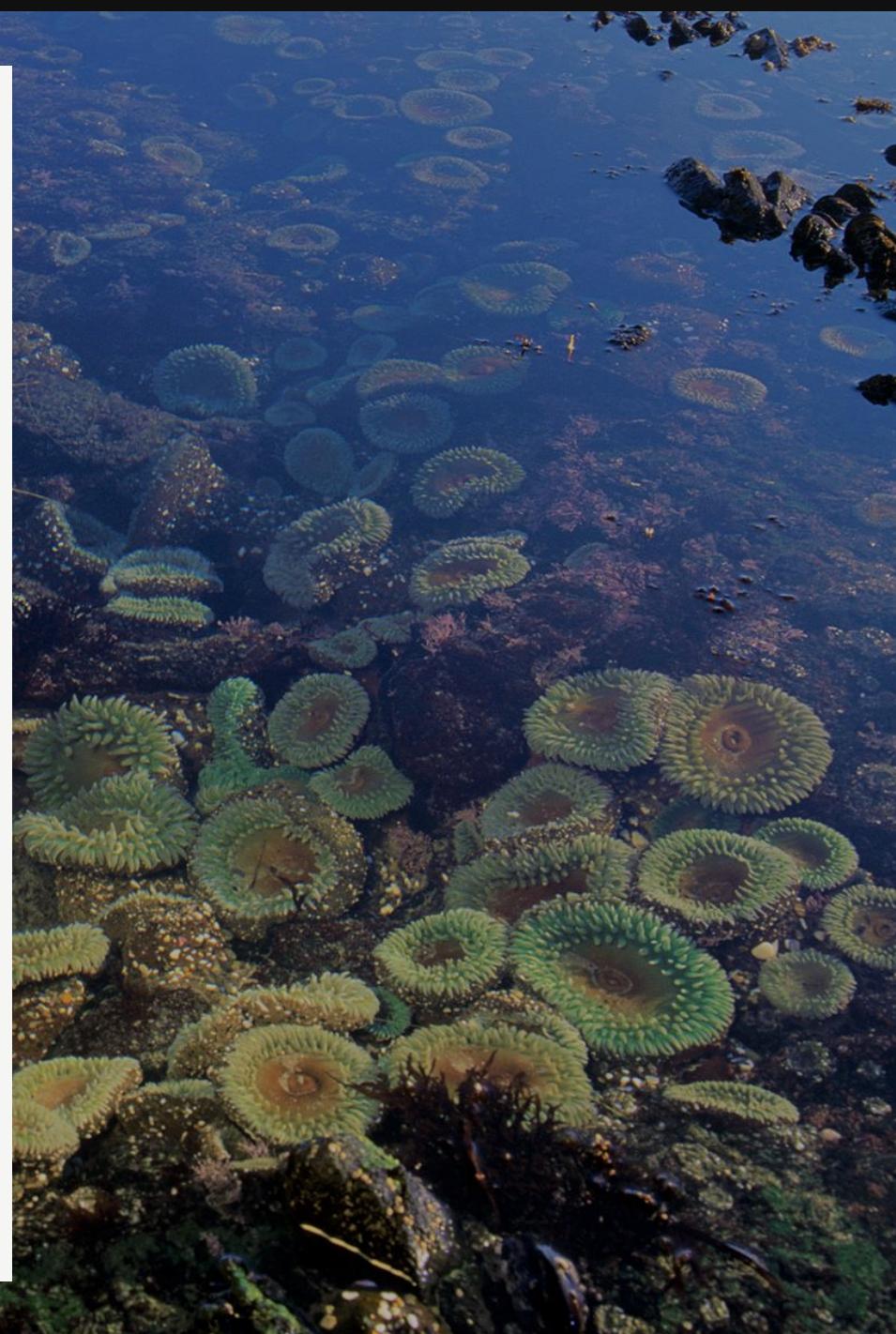
Importance of considering interactions and relationships among multiple variables:

- Complex interactions: *ecological systems are complex and interdependent*



Importance of considering interactions and relationships among multiple variables:

- Complex interactions
- Multidimensional patterns: *understanding the holistic picture*



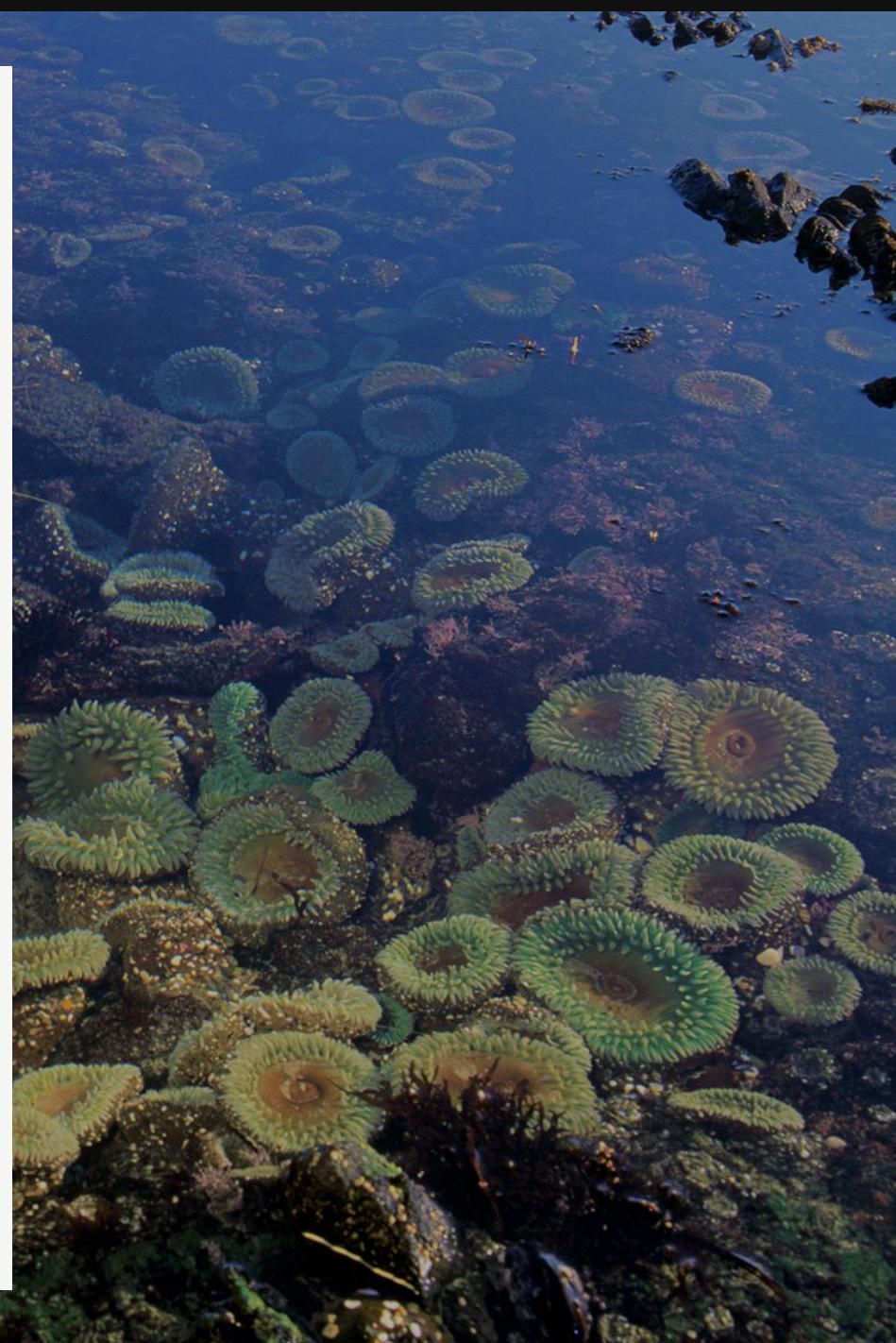
Importance of considering interactions and relationships among multiple variables:

- Complex interactions
- Multidimensional patterns
- Reduction of bias: *reducing the risk of biased estimates and Type I errors*



Importance of considering interactions and relationships among multiple variables:

- Complex interactions
- Multidimensional patterns
- Reduction of bias
- Informed decision-making: *making more informed decisions about conservation*





Purpose and Applications in Ecology



Community Composition



Diet and Trophic Ecology



Species Distributions and Habitat Suitability



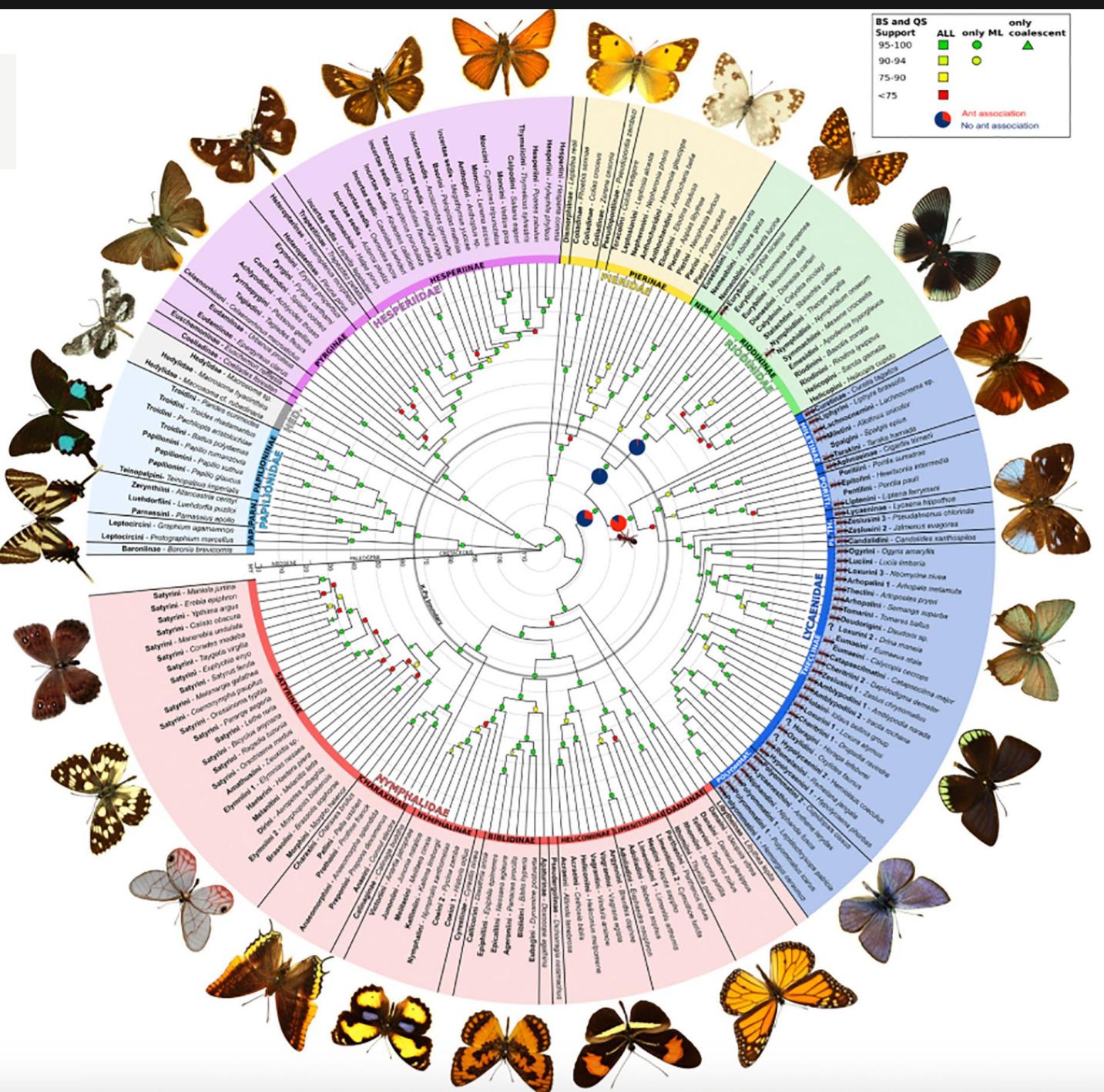
Habitat Classification and Landscape-Level Patterns



Identifying Spatiotemporal Patterns

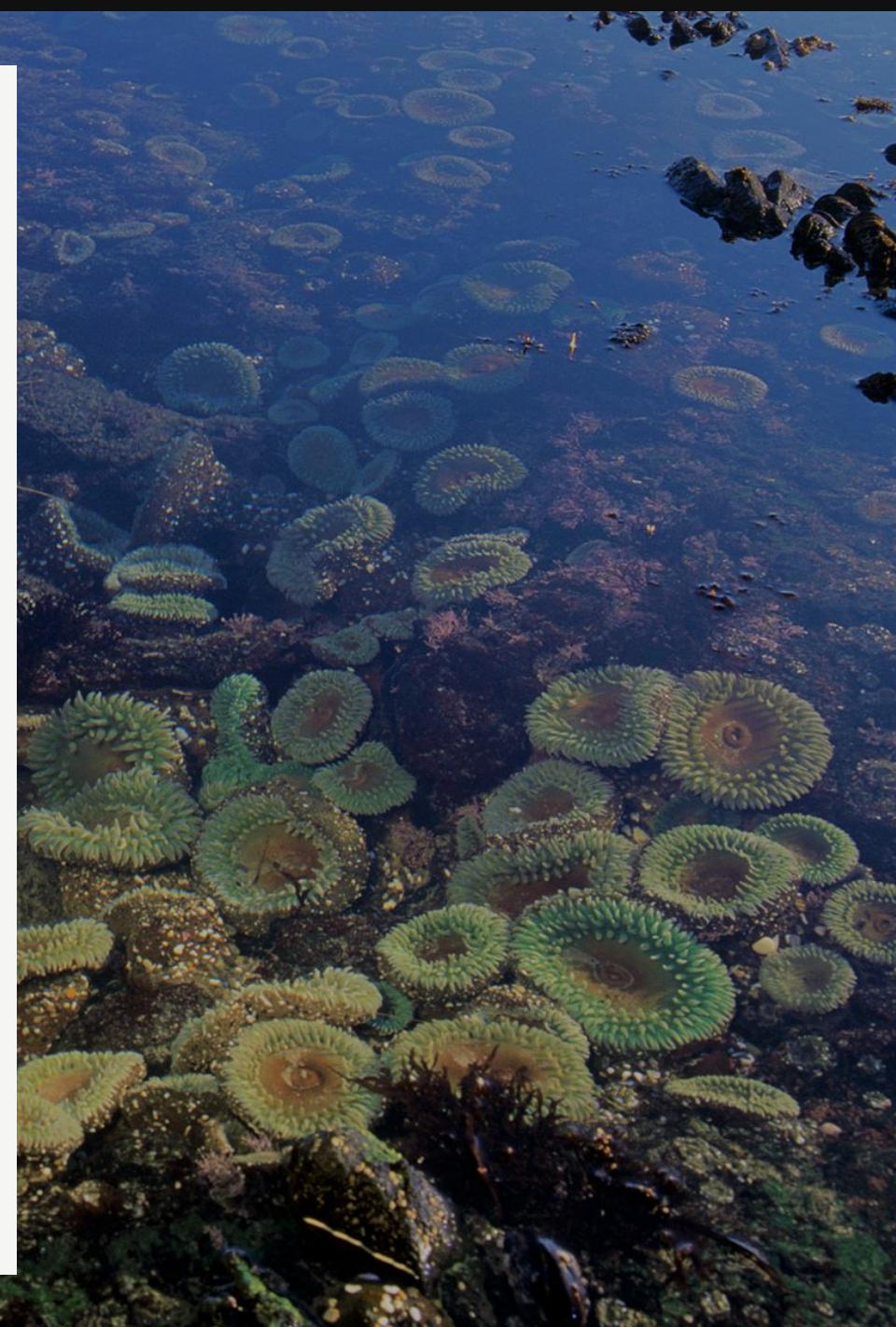
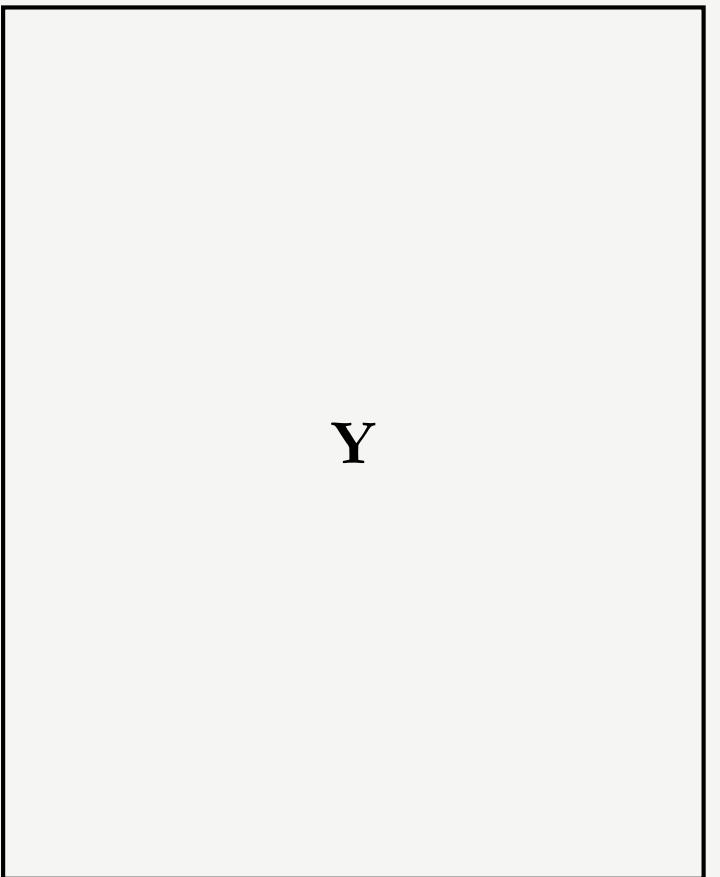


Phylogeny and Evolution

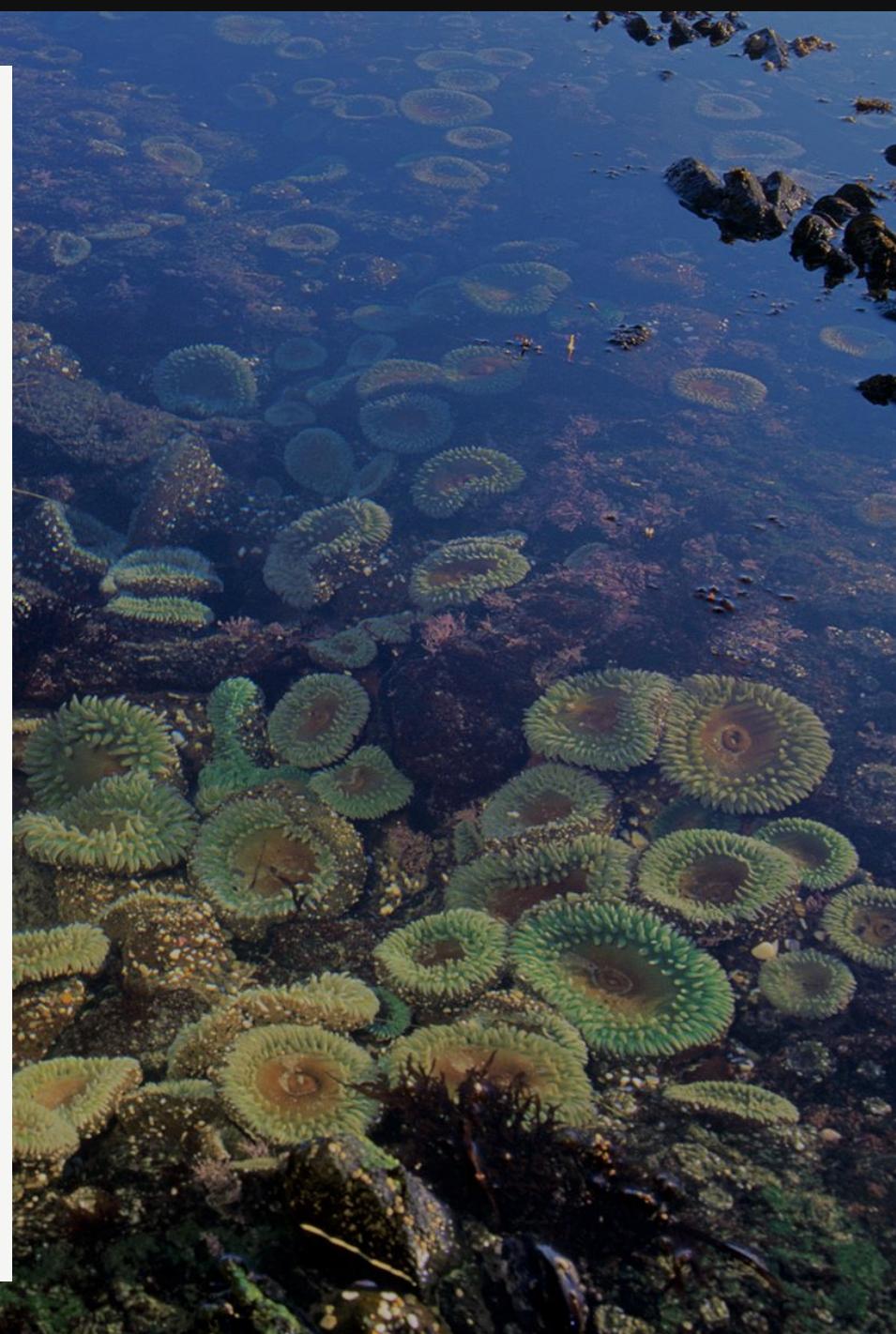
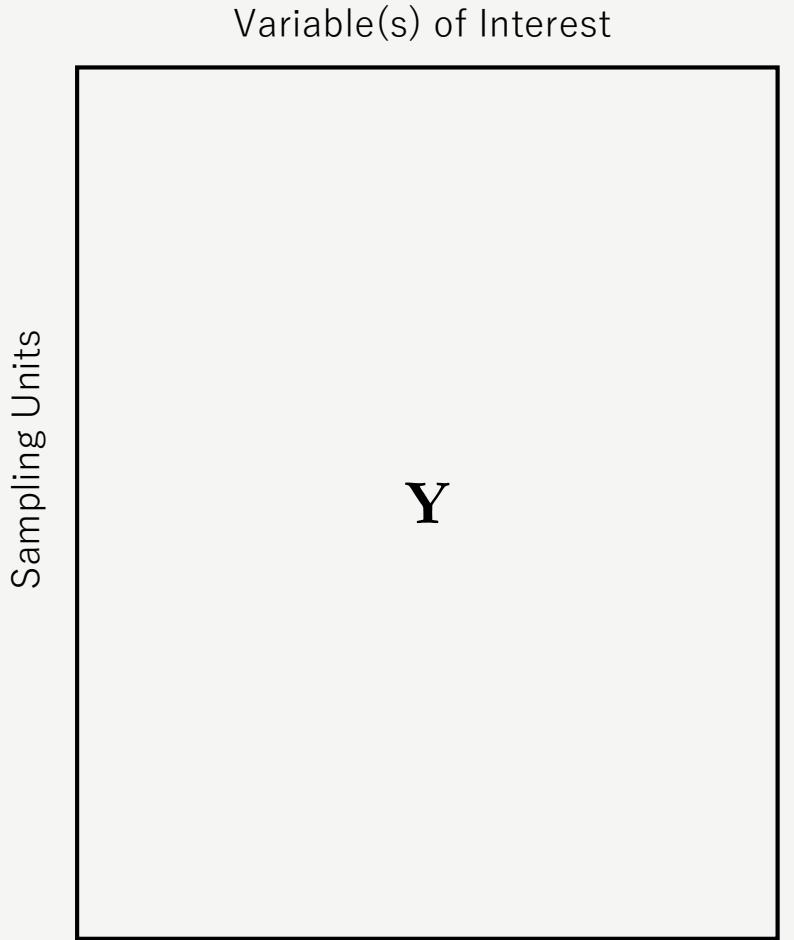


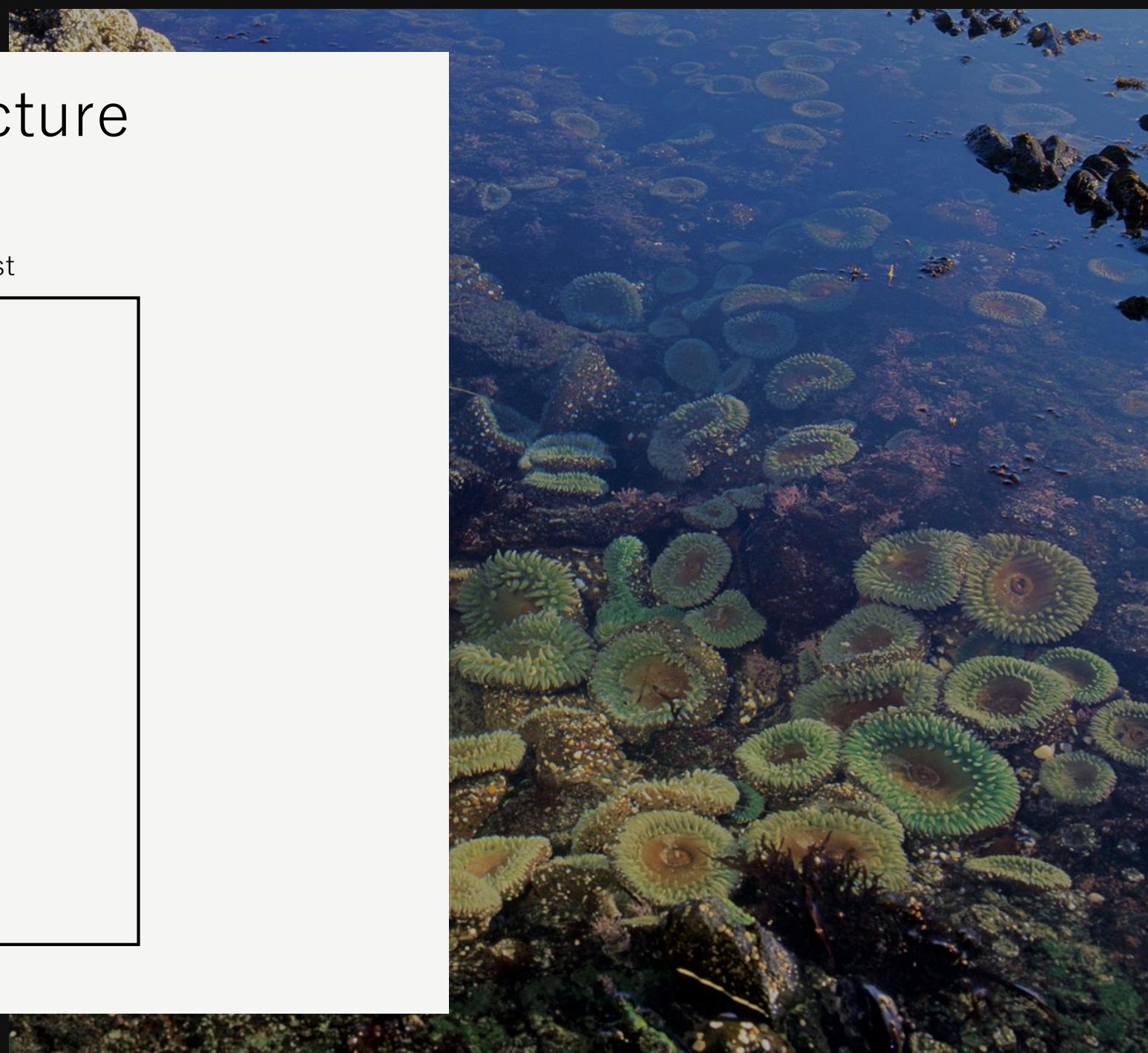


Multivariate Data Structure

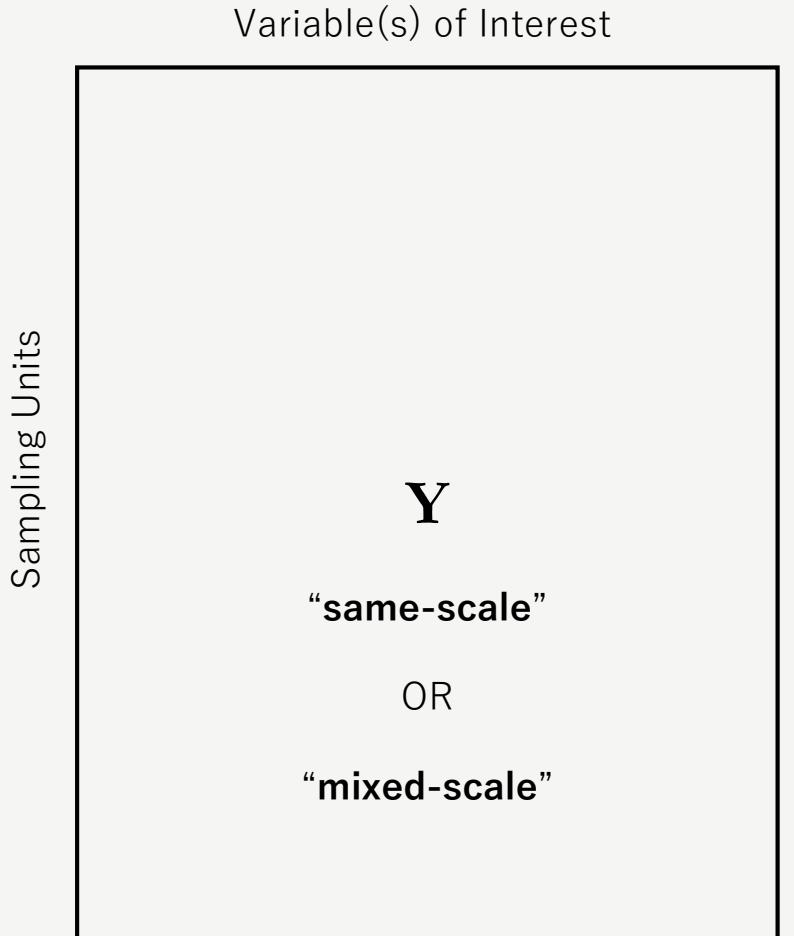


Multivariate Data Structure





Multivariate Data Structure



Multivariate Data Structure

Sampling Units

Variable(s) of Interest

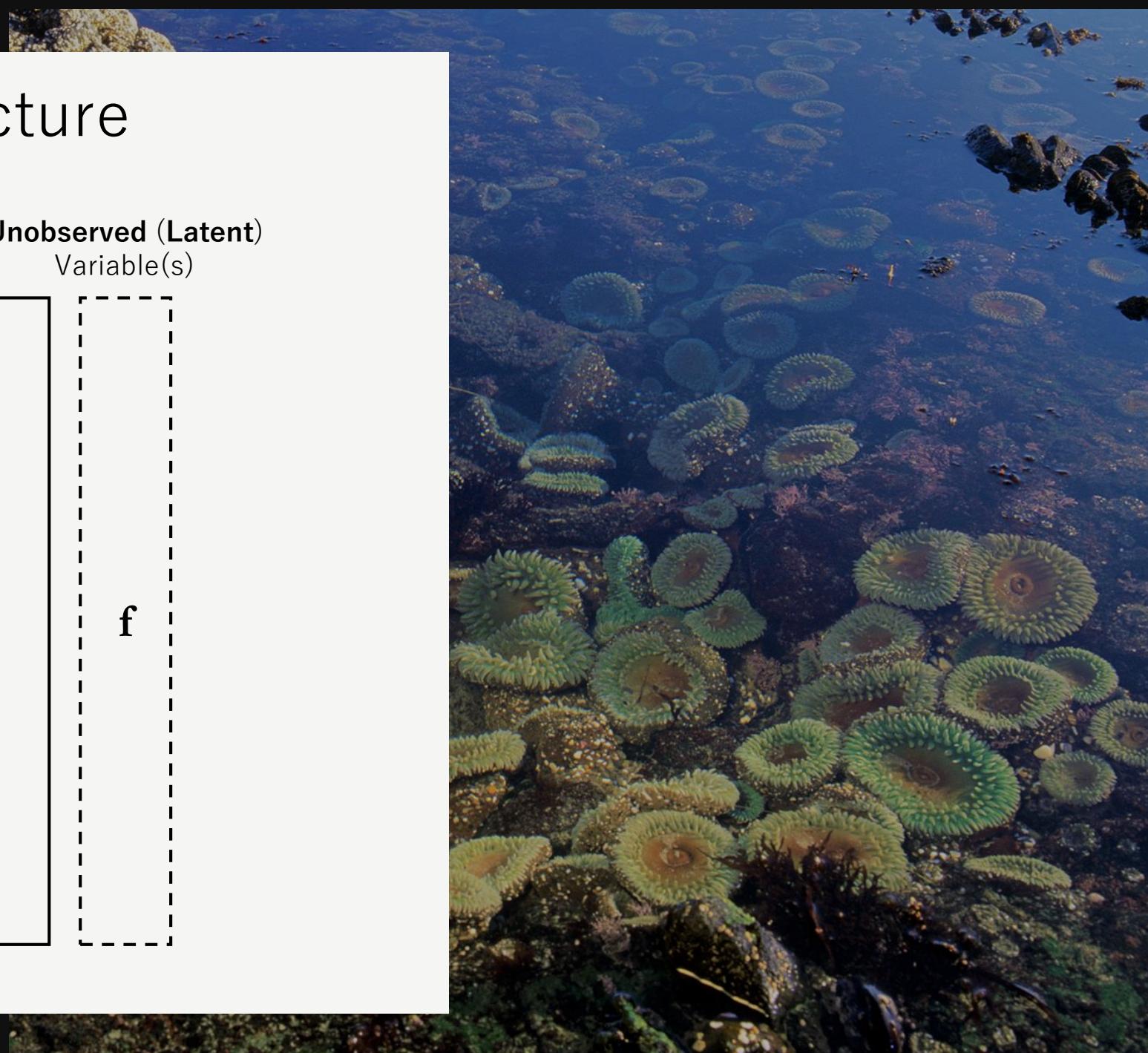
Y

“continuous”

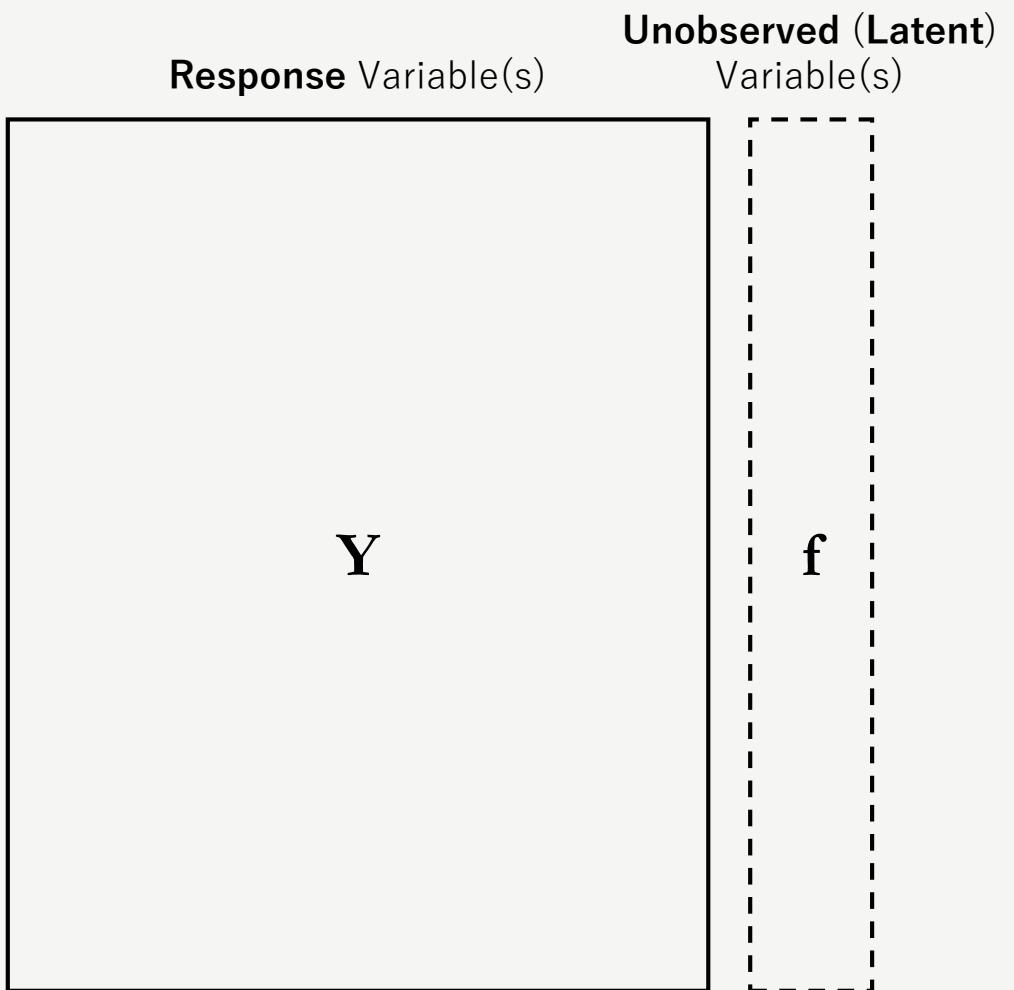
AND/OR

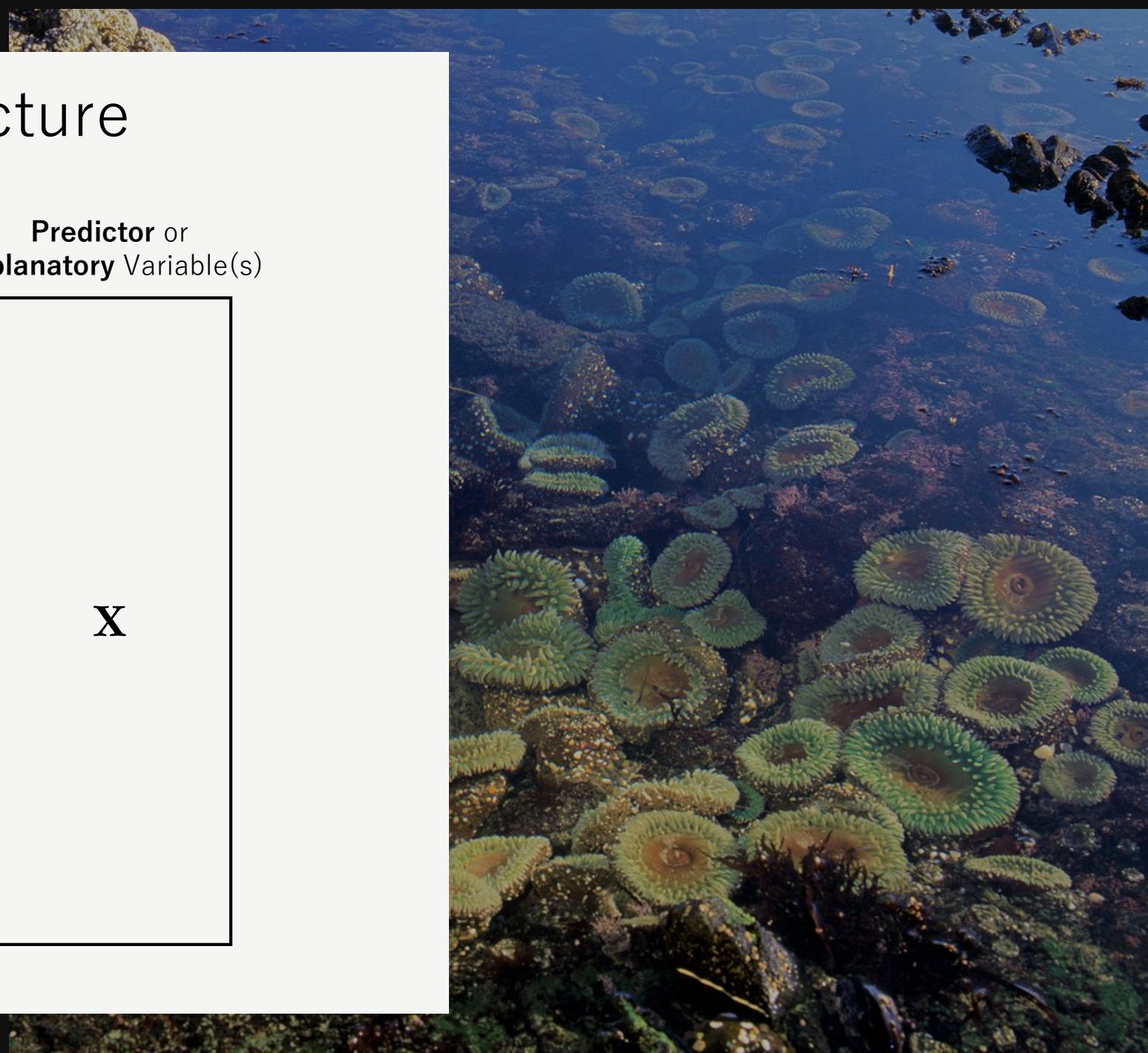
“categorical”



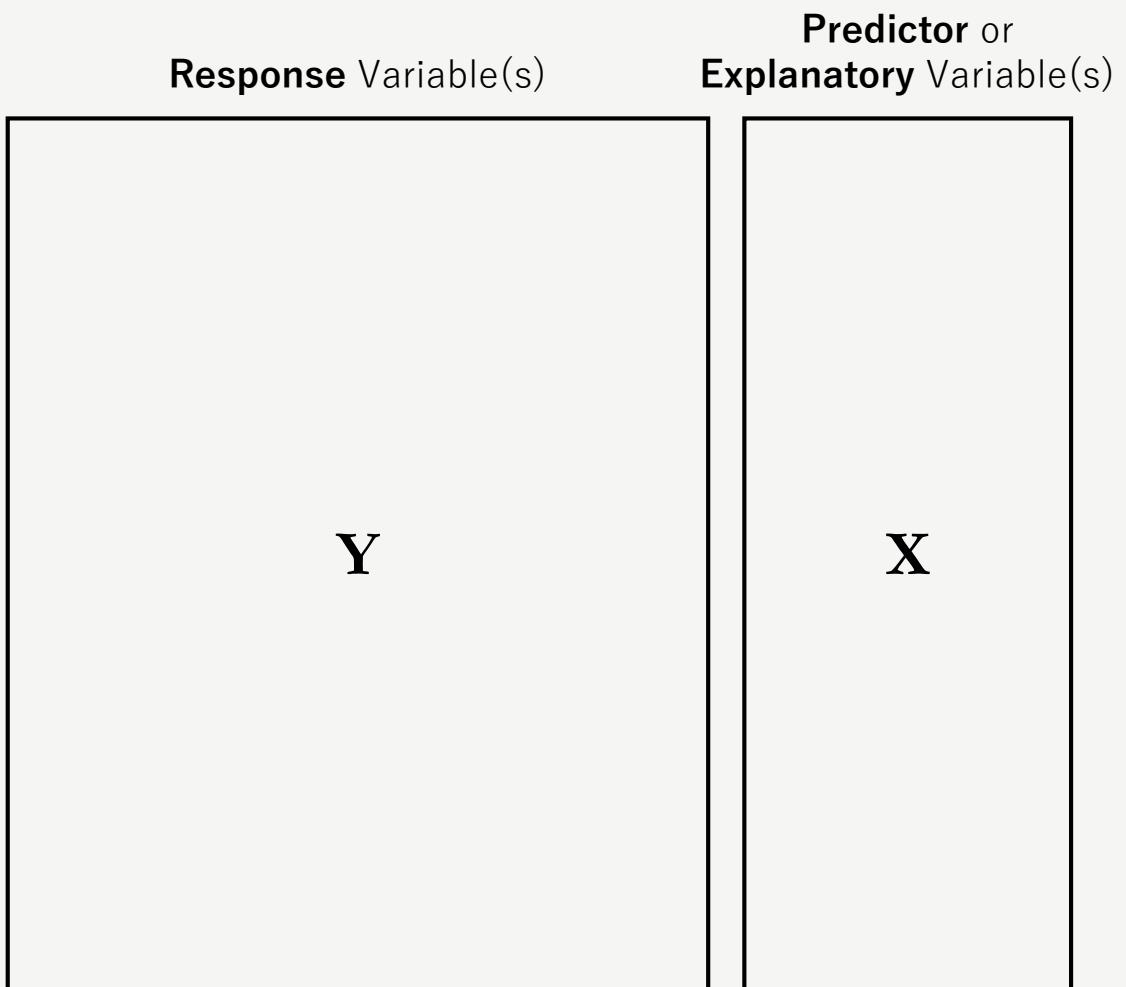


Multivariate Data Structure



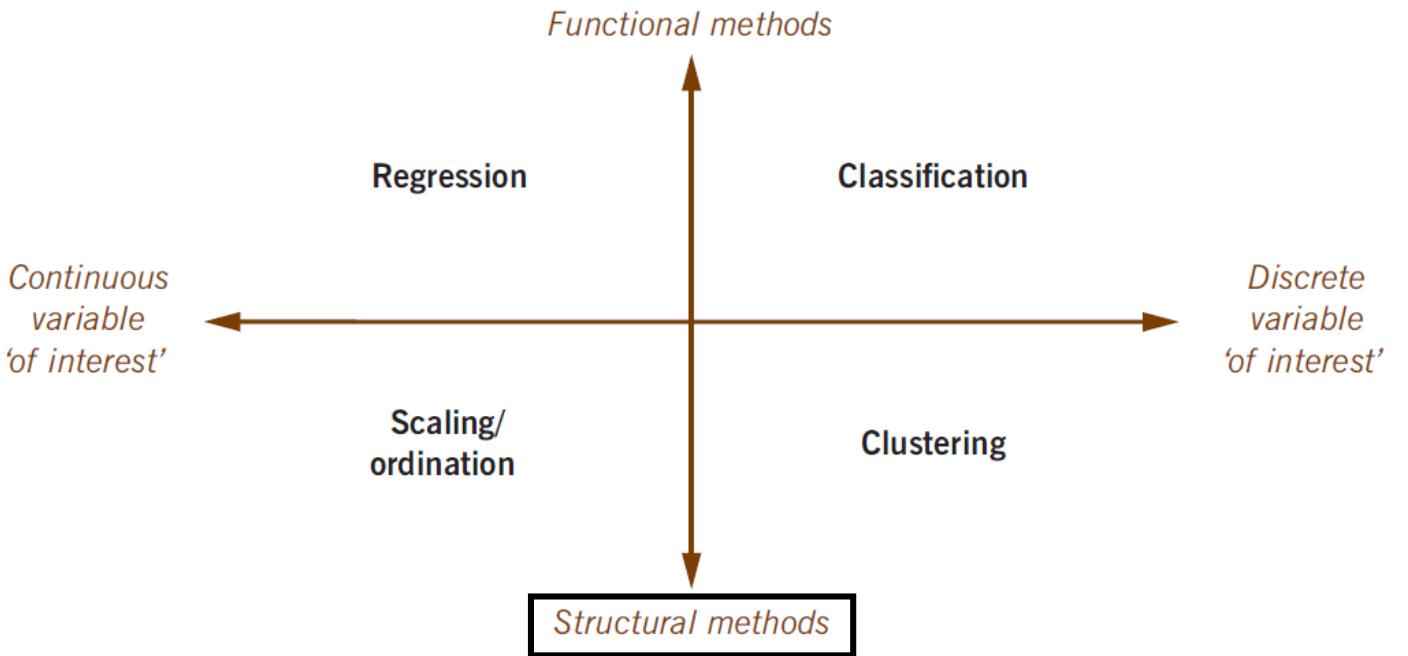


Multivariate Data Structure

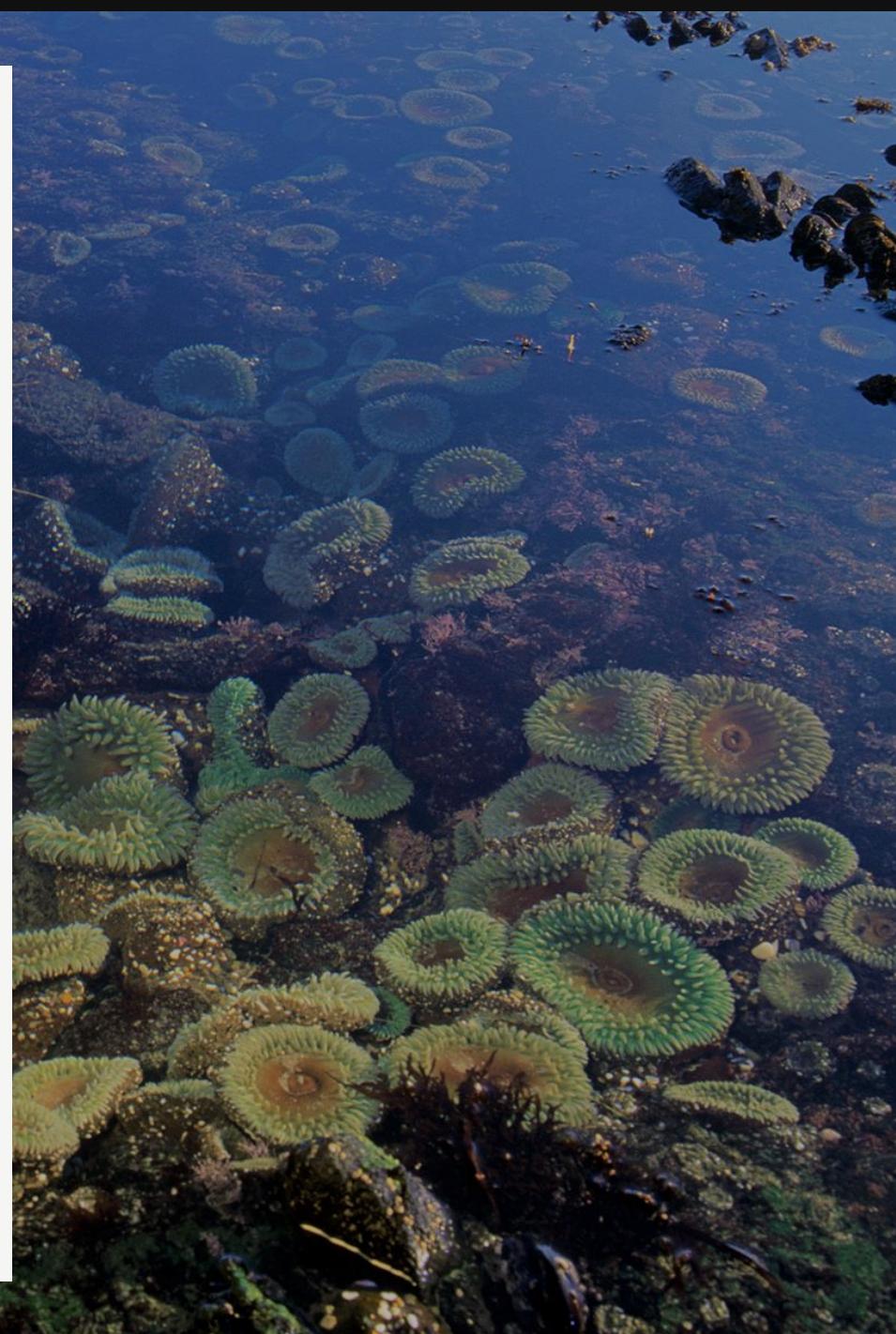


The “Four Corners”

Structural Methods: look for structure underlying the data matrix \mathbf{Y} .

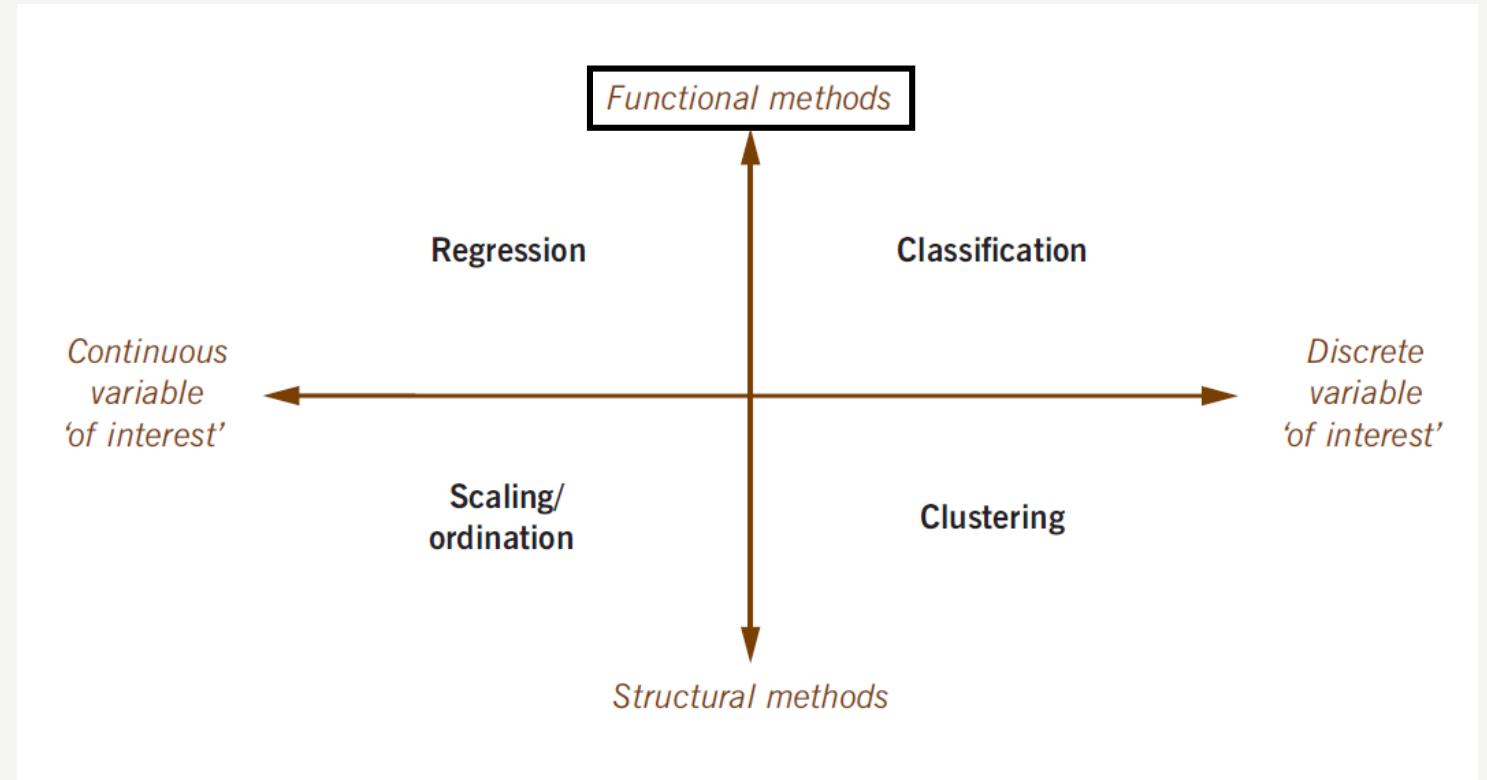


Greenacre and Primicerio

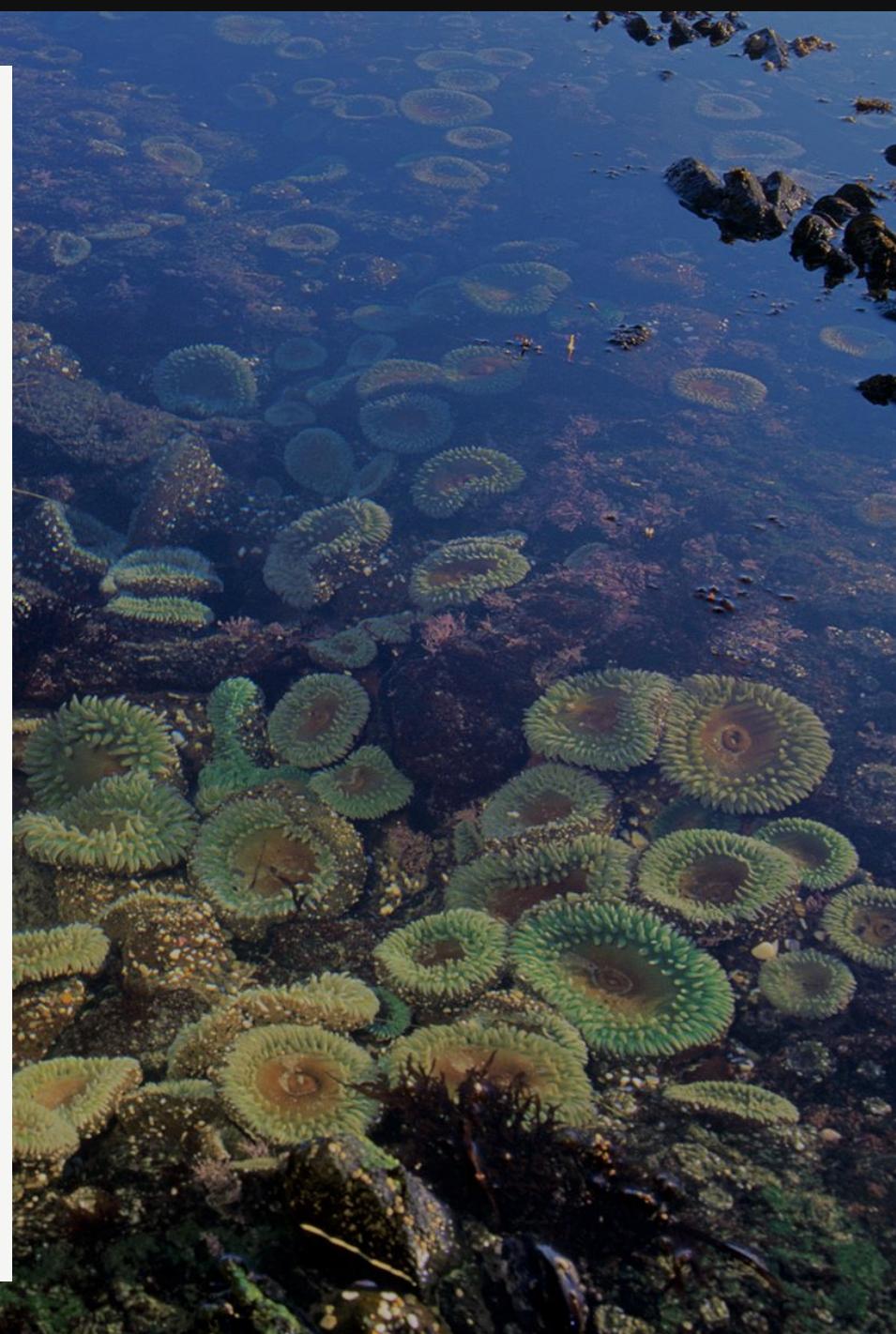


The “Four Corners”

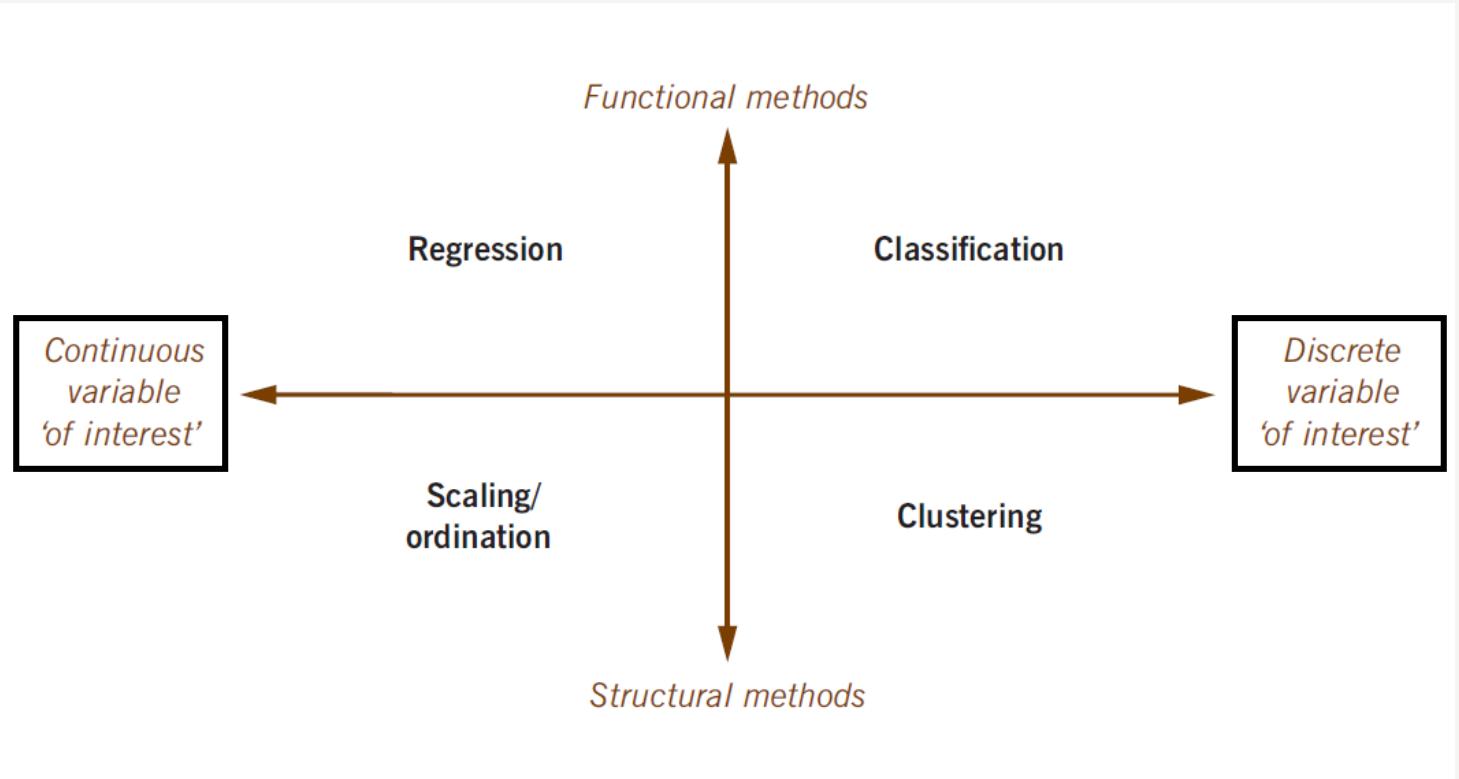
Functional Methods: relate the response variable(s) \mathbf{Y} as a function of the predictor variable(s) \mathbf{X} .



Greenacre and Primicerio



The “Four Corners”

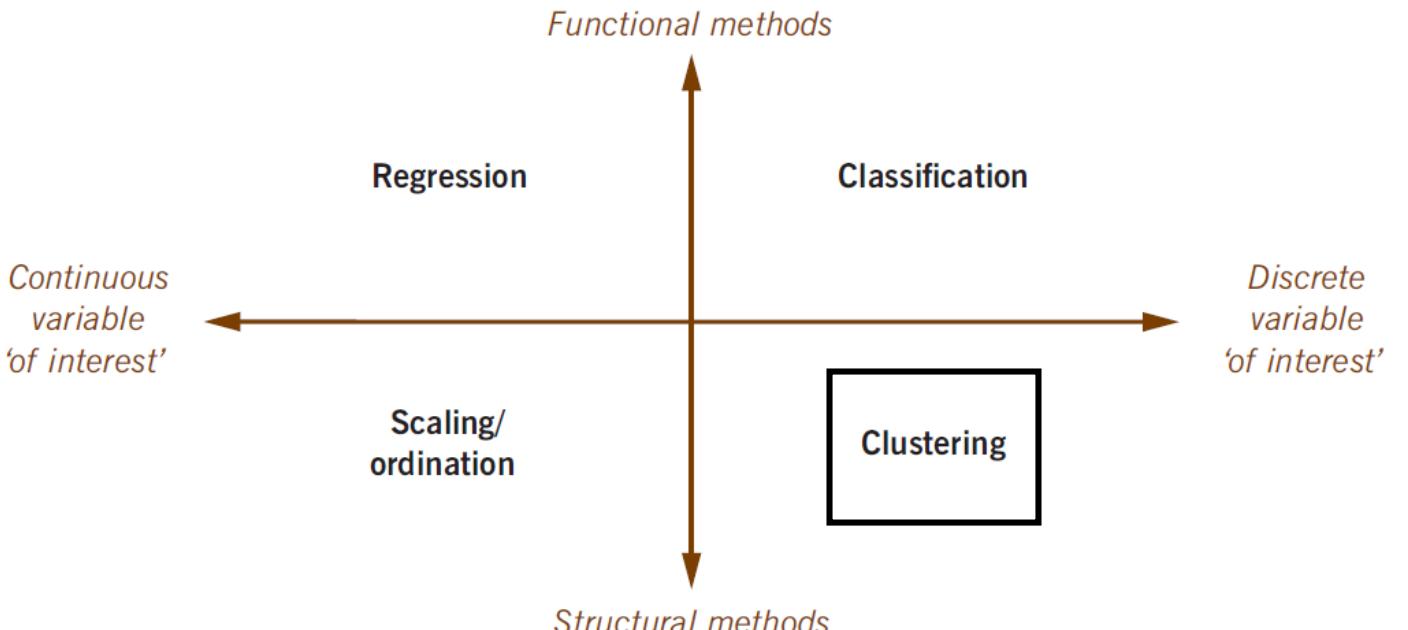


Greenacre and Primicerio



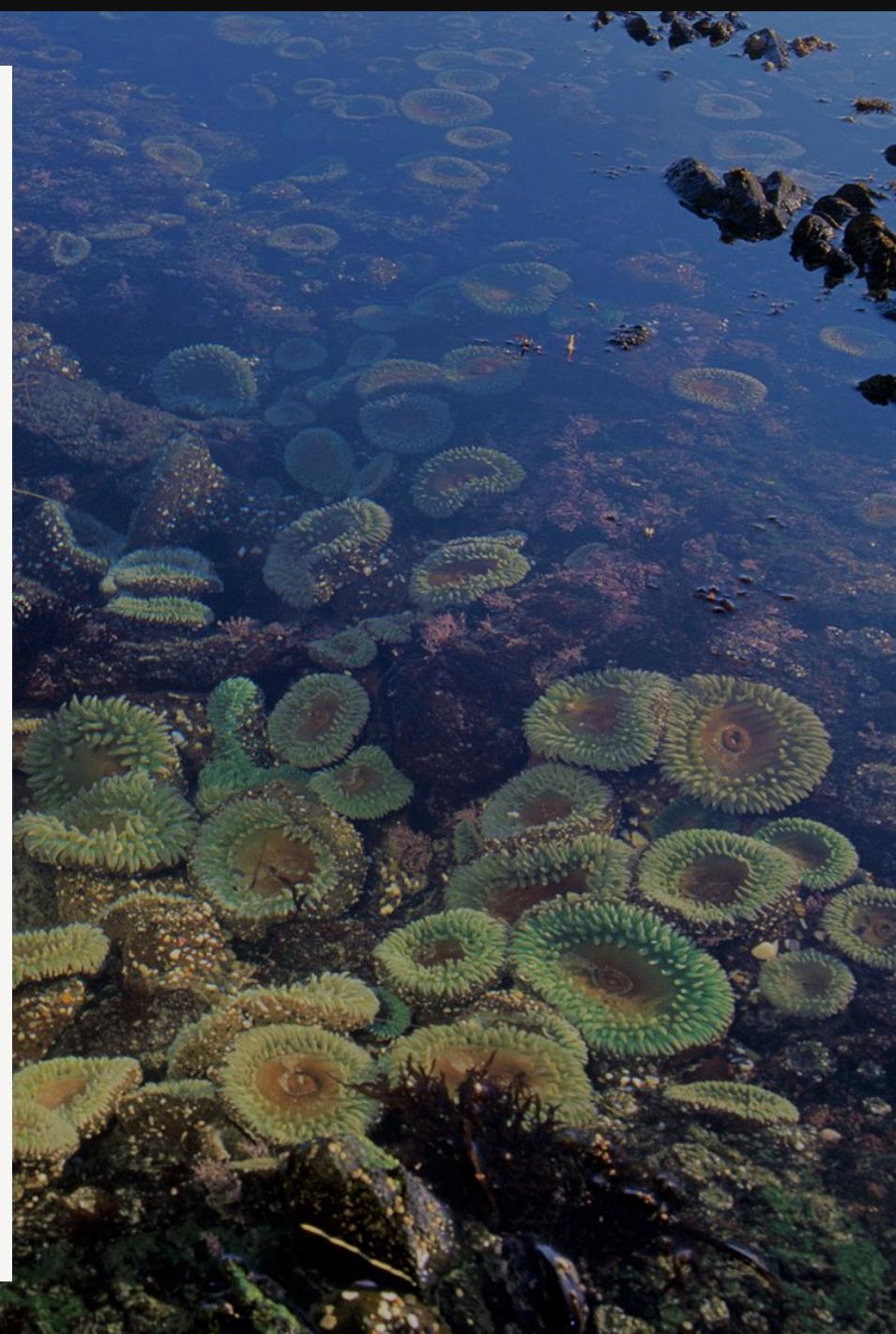
The “Four Corners”

Clustering: Structural methods uncovering a latent categorical variable. Also referred to as **unsupervised** learning.



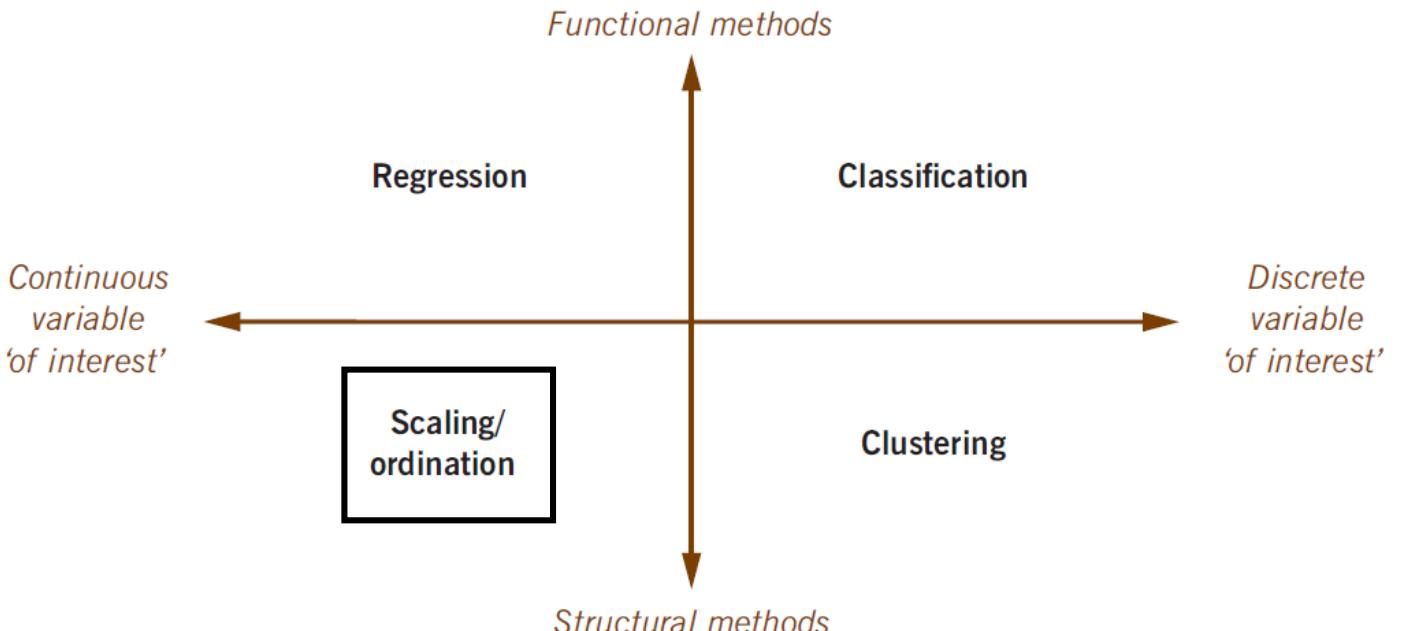
Can we look for similarities across all measured variables and come up with a grouping?

Greenacre and Primicerio



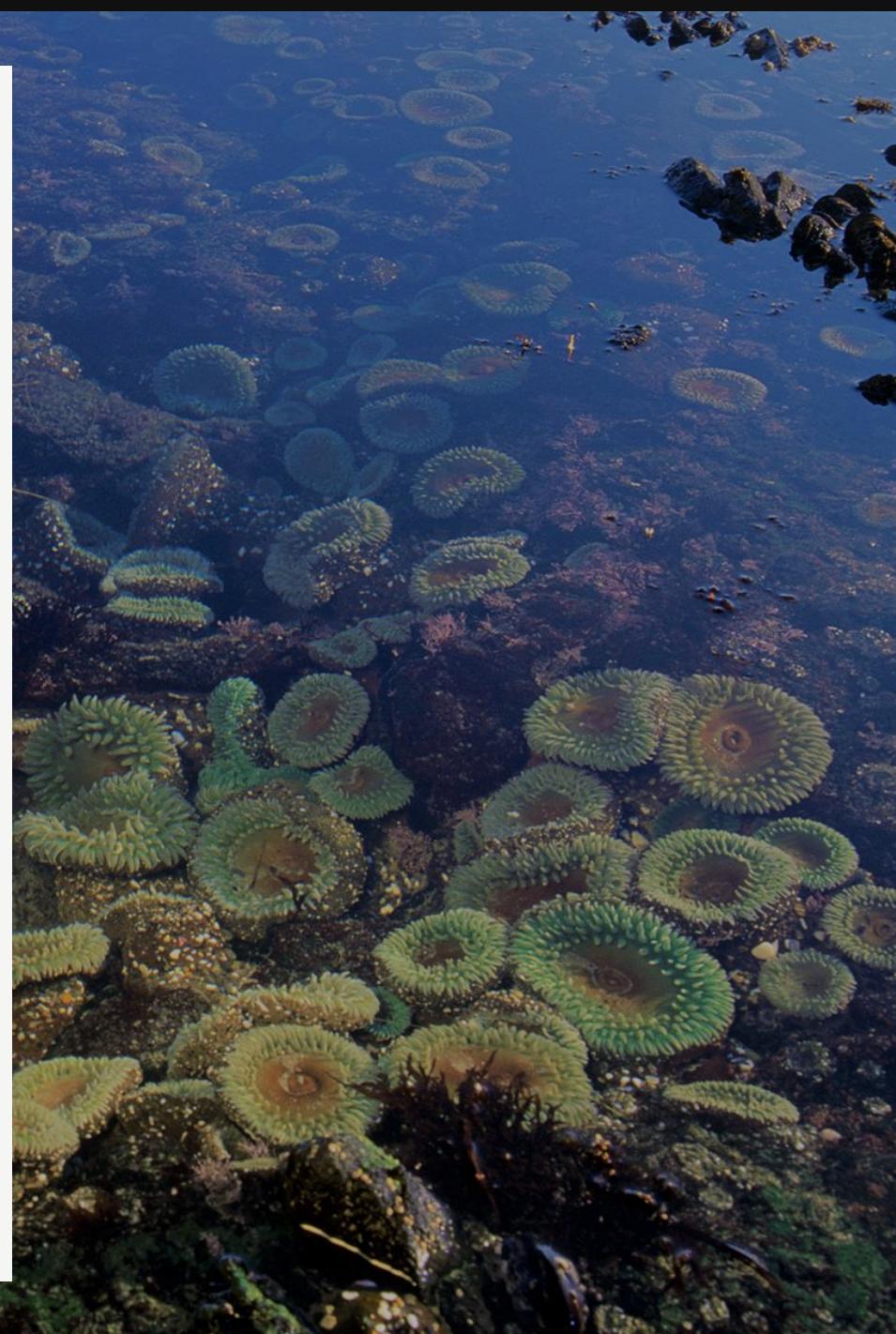
The “Four Corners”

Ordination: Structural methods uncovering a latent continuous variable. Also referred to as **supervised** learning.



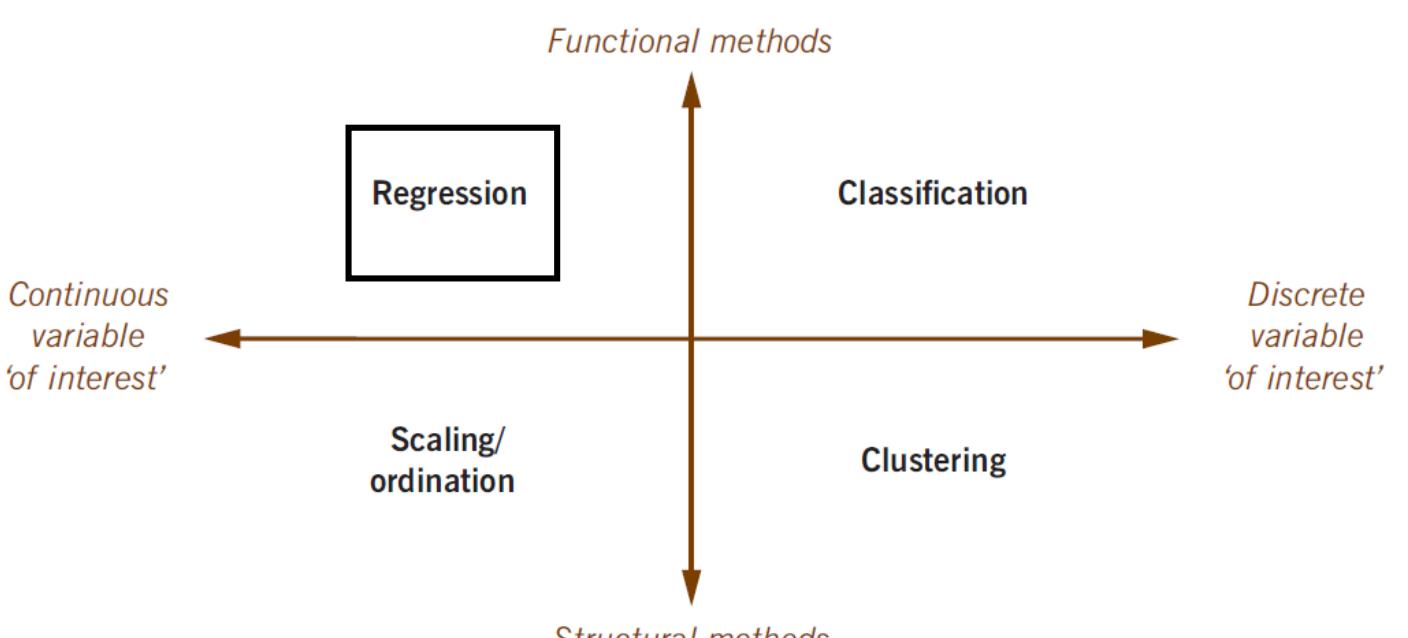
Can we identify structures in the data of a continuous nature?

Greenacre and Primicerio



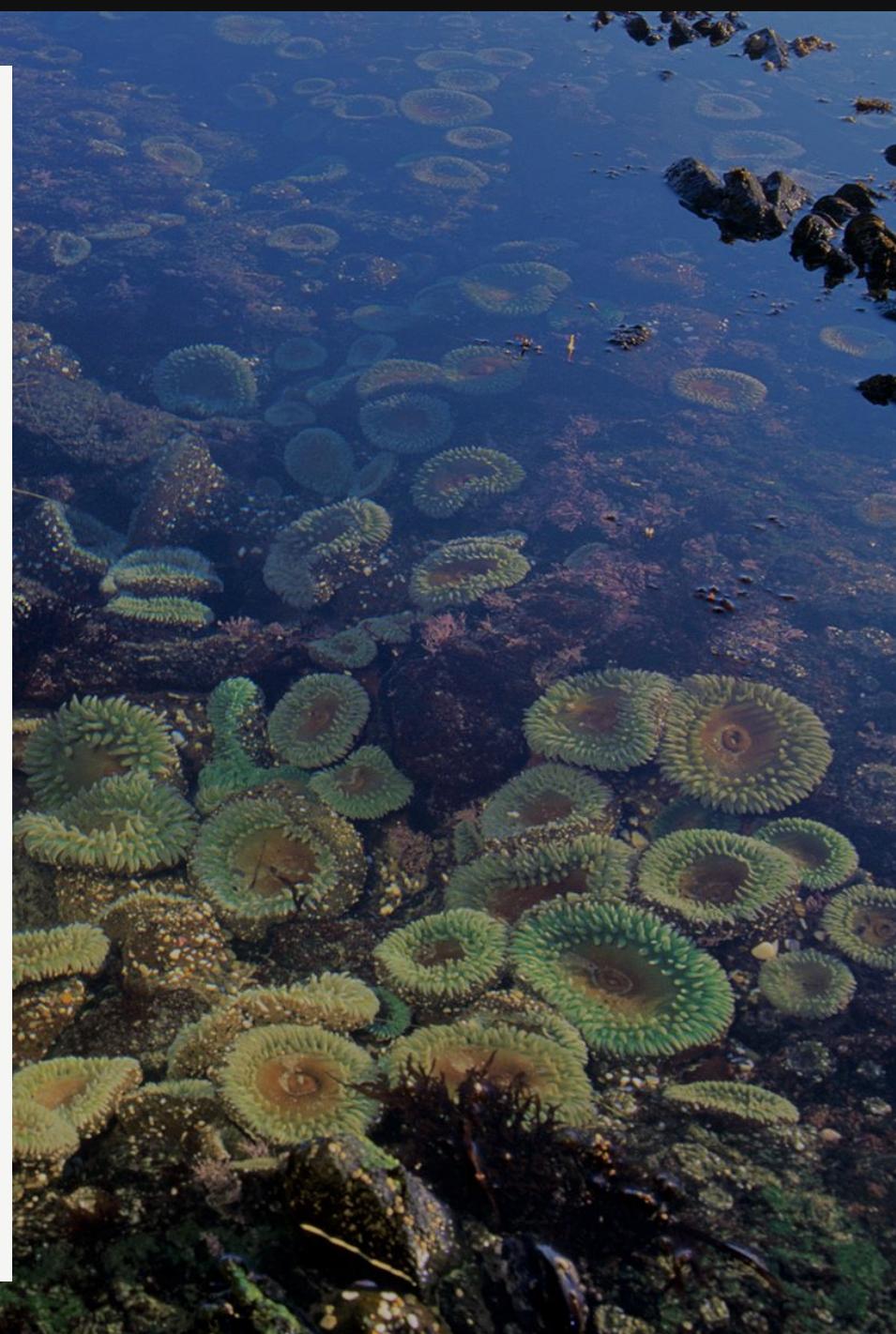
The “Four Corners”

Regression: Functional methods explaining a given continuous variable.



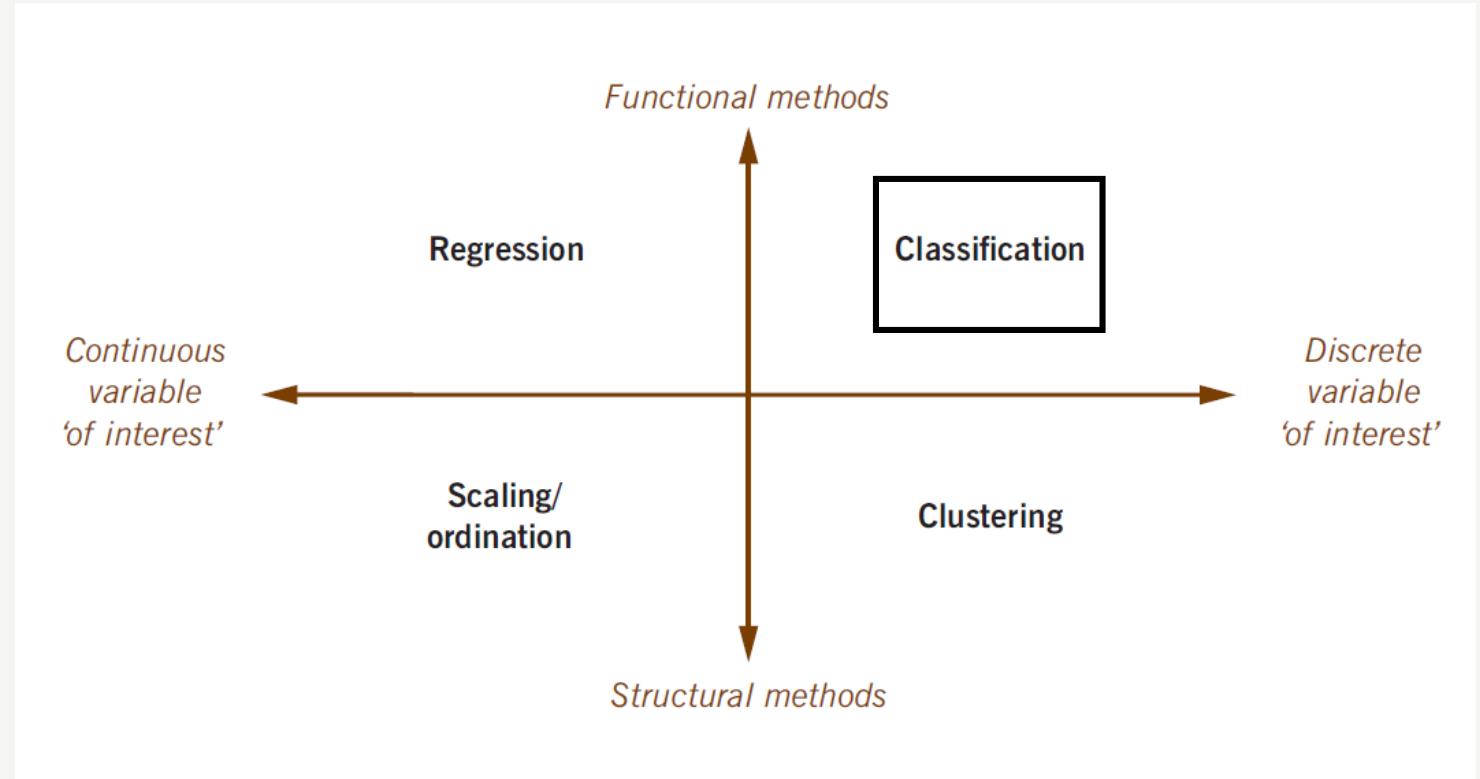
Can we use the multivariate data to explain one or more continuous response variables?

Greenacre and Primicerio



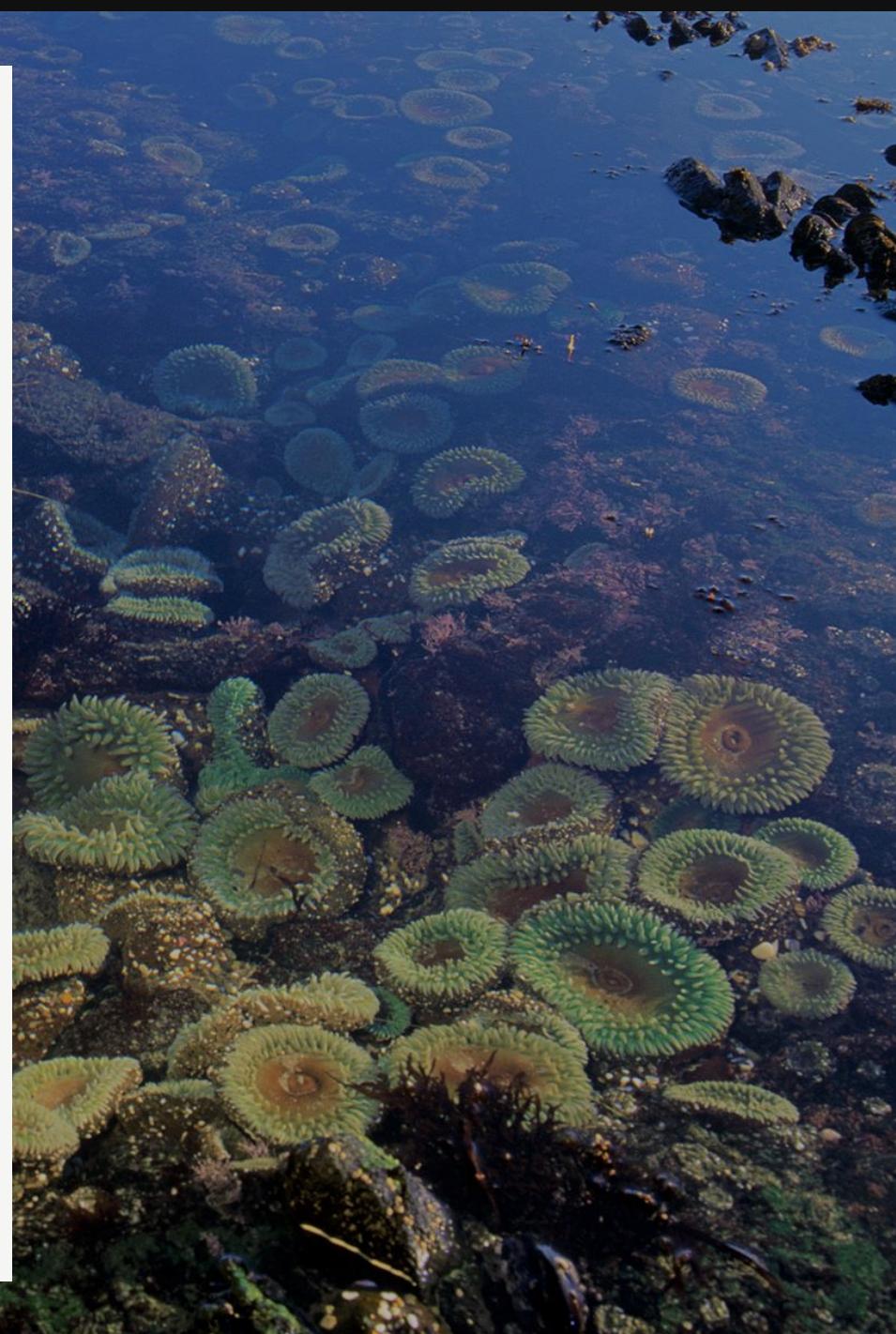
The “Four Corners”

Classification: Functional methods explaining a given categorical variable.



Can we model or predict a variable that takes on a small number of discrete values?

Greenacre and Primicerio



Flavors of Multivariate Analysis

Exploratory

Descriptive

Confirmatory

Inferential

Predictive

Flavors of Multivariate Analysis

Purpose: To uncover patterns, trends, and relationships in the data without prior hypotheses.

- Often the first step in data analysis
- Used to understand the underlying structure of the data, identify anomalies, and generate hypotheses for further study
- Open-ended and data-driven

Techniques: Structural methods including Principal Component Analysis (PCA), Nonmetric Multidimensional Scaling (NMDS), and Cluster Analysis

Exploratory

Descriptive

Confirmatory

Inferential

Predictive

Flavors of Multivariate Analysis

Purpose: To describe structure and relationships within the data.

- Often used as part of preliminary analysis
- Helpful for presenting data in a way that is easy to understand and interpret

Techniques: Structural methods including PCA, NMDS, and Cluster Analysis

Exploratory

Descriptive

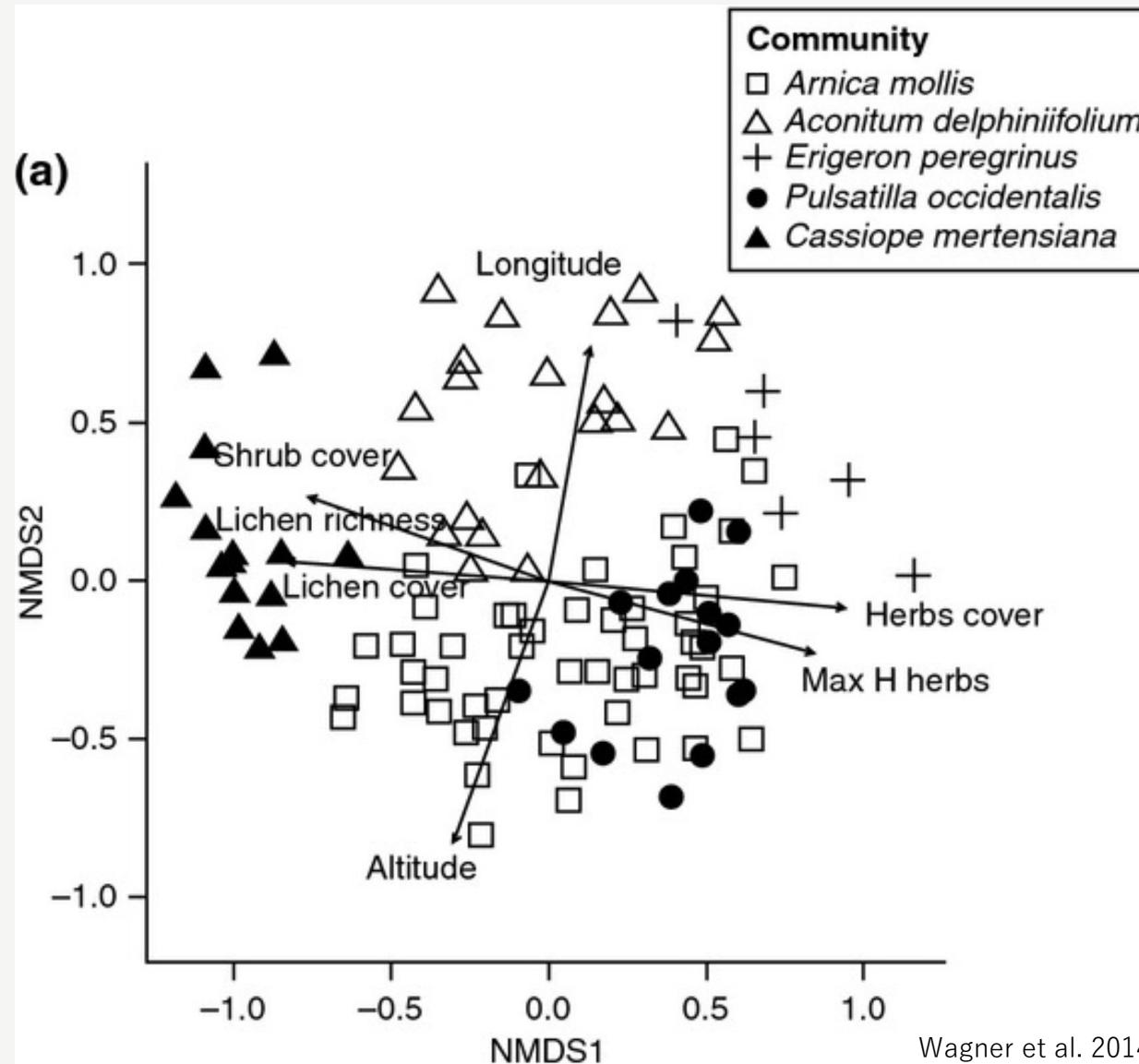
Confirmatory

Inferential

Predictive

Flavors of Multivariate Analysis

Exploratory



Confirmatory

Descriptive

Inferential

Predictive

Flavors of Multivariate Analysis

Purpose: To test specific hypotheses about the relationships among variables.

- Begins with a clear, testable hypothesis
- Often relies on statistical tests
- Used to validate findings from exploratory analysis

Techniques: Functional methods including Multivariate Analysis of Variance (MANOVA/PERMANOVA) and Canonical Correspondence Analysis (CCA)

Exploratory

Descriptive

Confirmatory

Inferential

Predictive

Flavors of Multivariate Analysis

Purpose: To infer population parameters and test hypotheses about the data.

- Aims to draw conclusions about the larger population and assess the reliability of these conclusions
- Involves the calculation of probabilities to estimate the likelihood that observed patterns are due to chance
- Accounts for sampling error and uncertainty

Techniques: Functional methods including MANOVA, Discriminant Function Analysis (DFA), and CCA

Exploratory

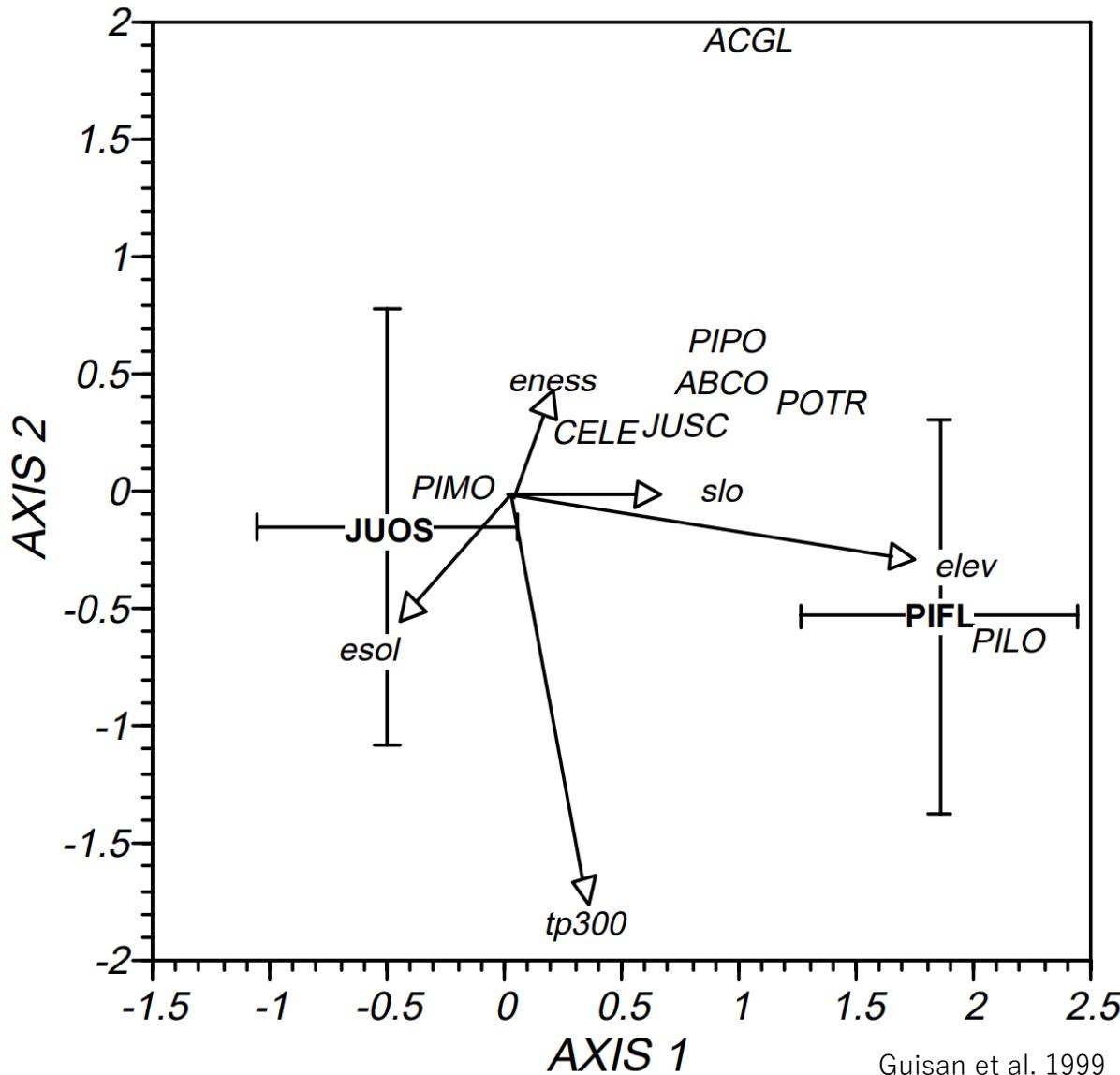
Descriptive

Confirmatory

Inferential

Predictive

Flavors of Multivariate Analysis



Exploratory

Descriptive

Confirmatory

Inferential

Predictive

Flavors of Multivariate Analysis

Purpose: To predict the value of a response variable based on multiple explanatory variables

Techniques: Functional methods including CCA, Classification and Regression Trees, Random Forests, and Neural Networks

Exploratory

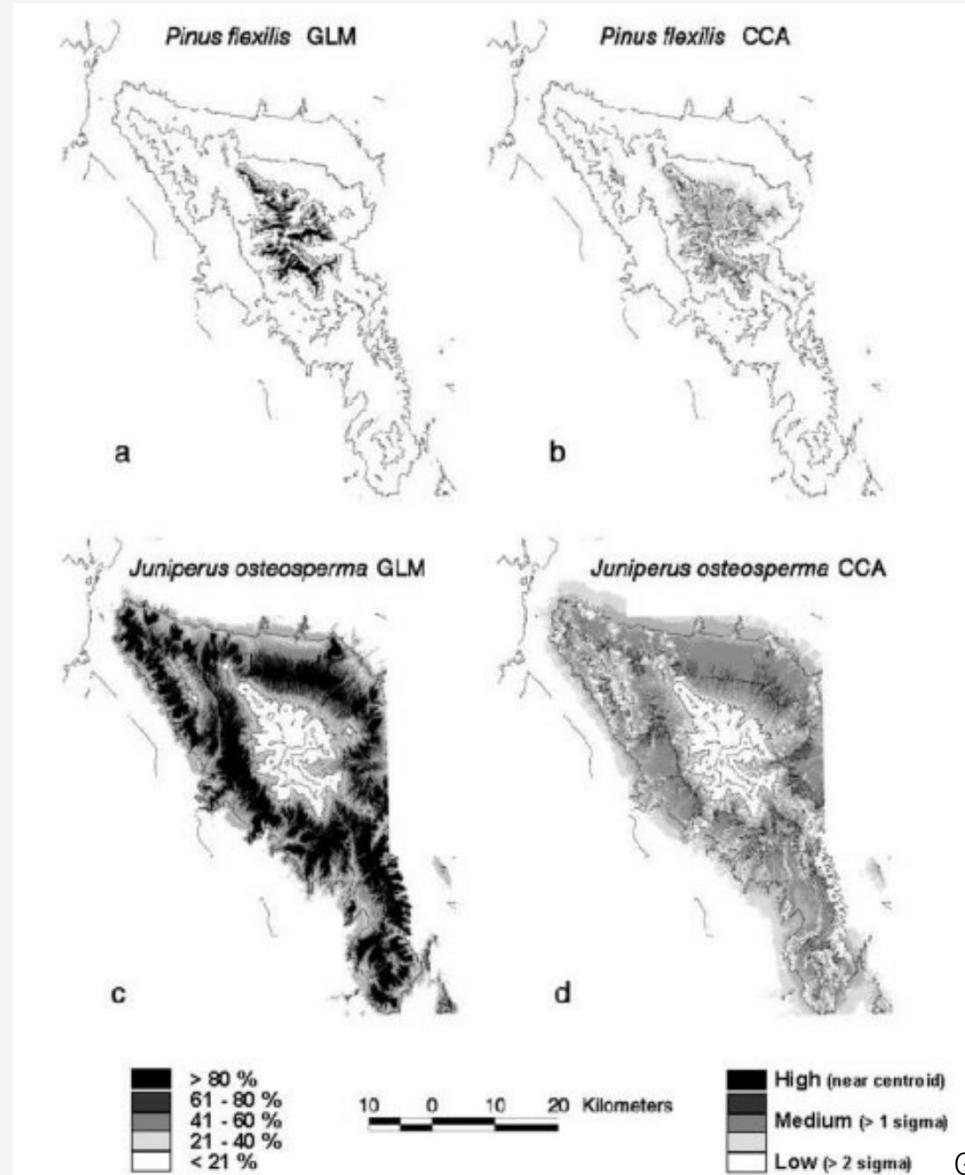
Descriptive

Confirmatory

Inferential

Predictive

Flavors of Multivariate Analysis



Exploratory

Descriptive

Confirmatory

Inferential

Predictive

Who figured this stuff out, anyway?

1901: Principal Component Analysis (Pearson)

1932: Cluster Analysis (Driver, Kroeber)

1936: Canonical Correlation Analysis (Hotelling)

1936: Discriminant Analysis (Fisher)

1951: K-nearest Neighbor Algorithm for Classification and Regression (Fix, Hodges)

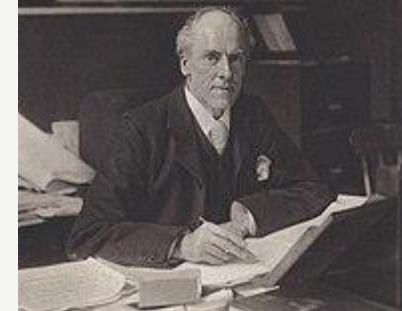
1952: Multidimensional Scaling (Torgerson)

2001: Random Forests (Breiman, Cutler)

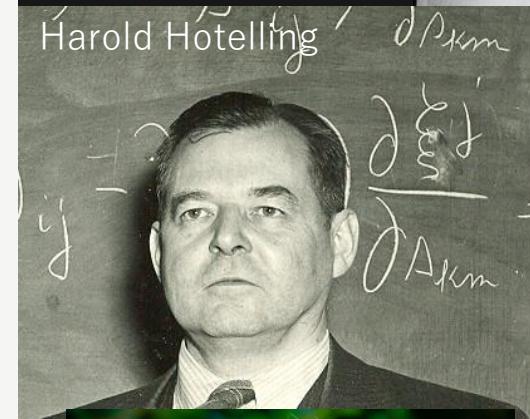
2010s: Neural Networks and Machine Learning (Hinton, LeCun, Bengio)



Karl Pearson



Ronald Fisher

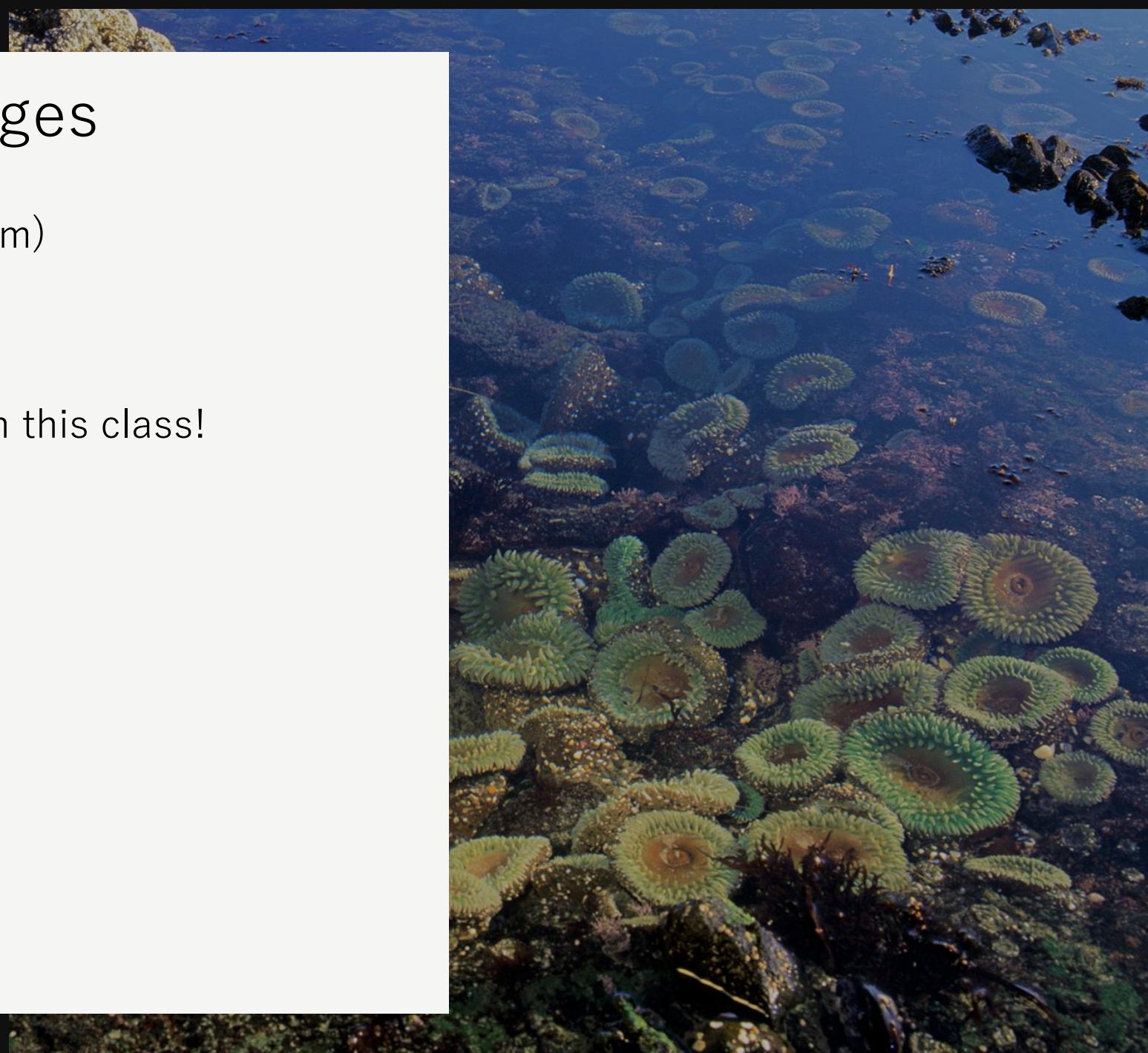


Adele Cutler



Evelyn Fix





Programs and R Packages

- PC-ORD (Bruce McCune's program)
- SPSS Statistics
- SAS (old school)
- R/R Studio: What we'll be using in this class!



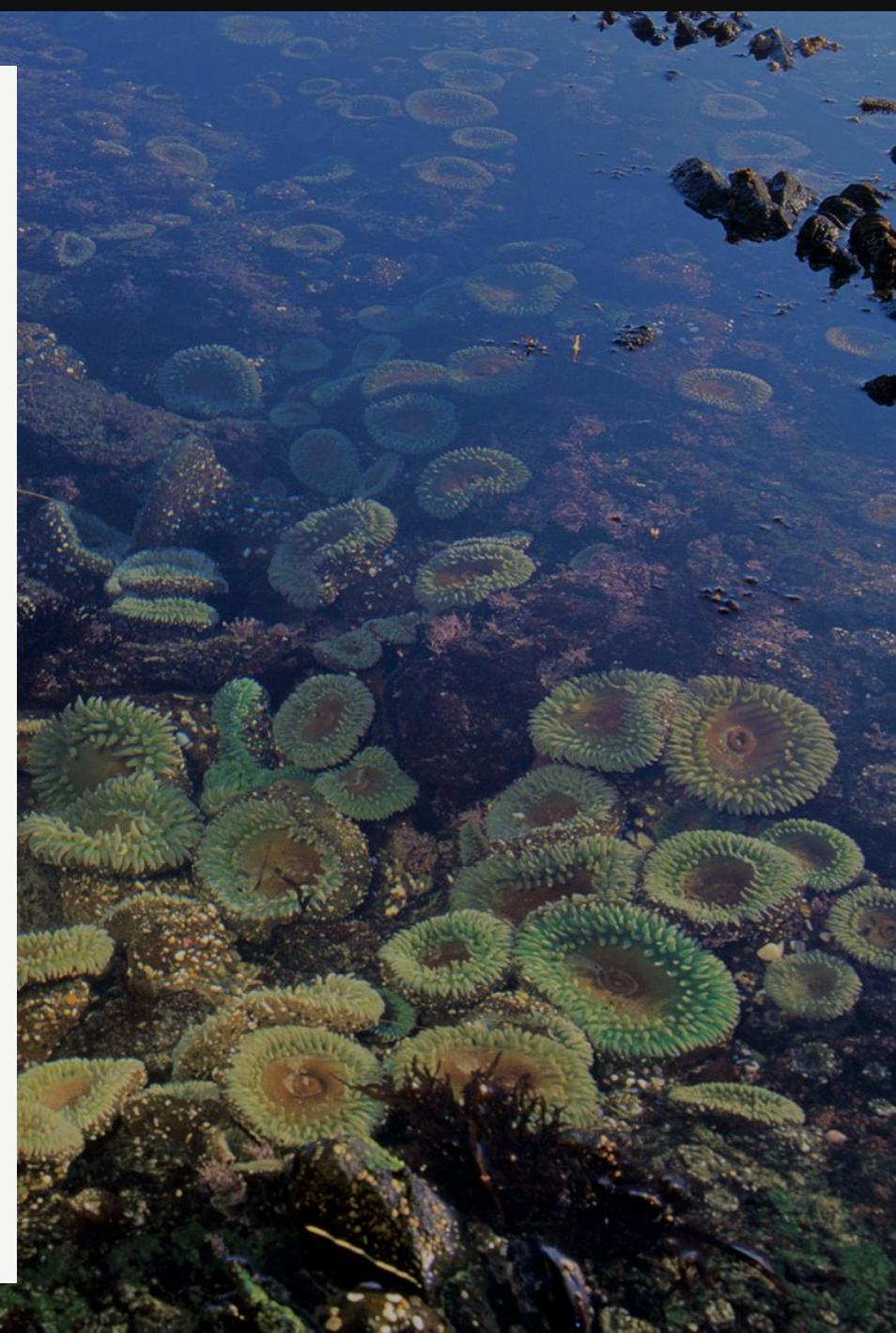
Programs and R Packages

- PC-ORD (Bruce McCune's program)
- SPSS Statistics
- SAS (old school)
- R/R Studio: What we'll be using in this class!
 - **"vegan"**: the quintessential multivariate stats package for ecological data sets
 - **"ade4"**: mostly used for ordination
 - **"cluster"**: methods for cluster analysis
 - **"mixOmics"**: specifically designed for 'omics data
 - **"dplyr"** and **"tidyverse"**: some helpful data manipulation packages
 - **"ggplot2"**: for making your data visualizations look pretty



Conclusion: Summary of Key Points

- Data structure
 - **Same-scale or mixed-scale?**
 - **Continuous or categorical?**
- **Structural methods:** response variable(s) related to latent variable(s)
- **Functional methods:** response variable(s) related to predictor variable(s)
- Types of Analysis:
 - **Exploratory**
 - **Descriptive**
 - **Confirmatory**
 - **Inferential**
 - **Predictive**



Conclusion: What to Expect in Future Lectures

- Next Week: Data Exploration and Pre-Processing
- **Structural methods:**
 - Cluster Analysis (Week 3)
 - Ordination: PCA (Week 4)
 - Ordination: PCoA, NMDS (Week 5)
- **Functional methods:**
 - RDA/CCA (Week 6)
 - Regression: (PER)MANOVA, ANOSIM (Week 7)
 - Classification and Regression Trees (Week 8)
 - Random Forests (Week 8)
- **Advanced methods & errata** (Week 9)



Questions?

