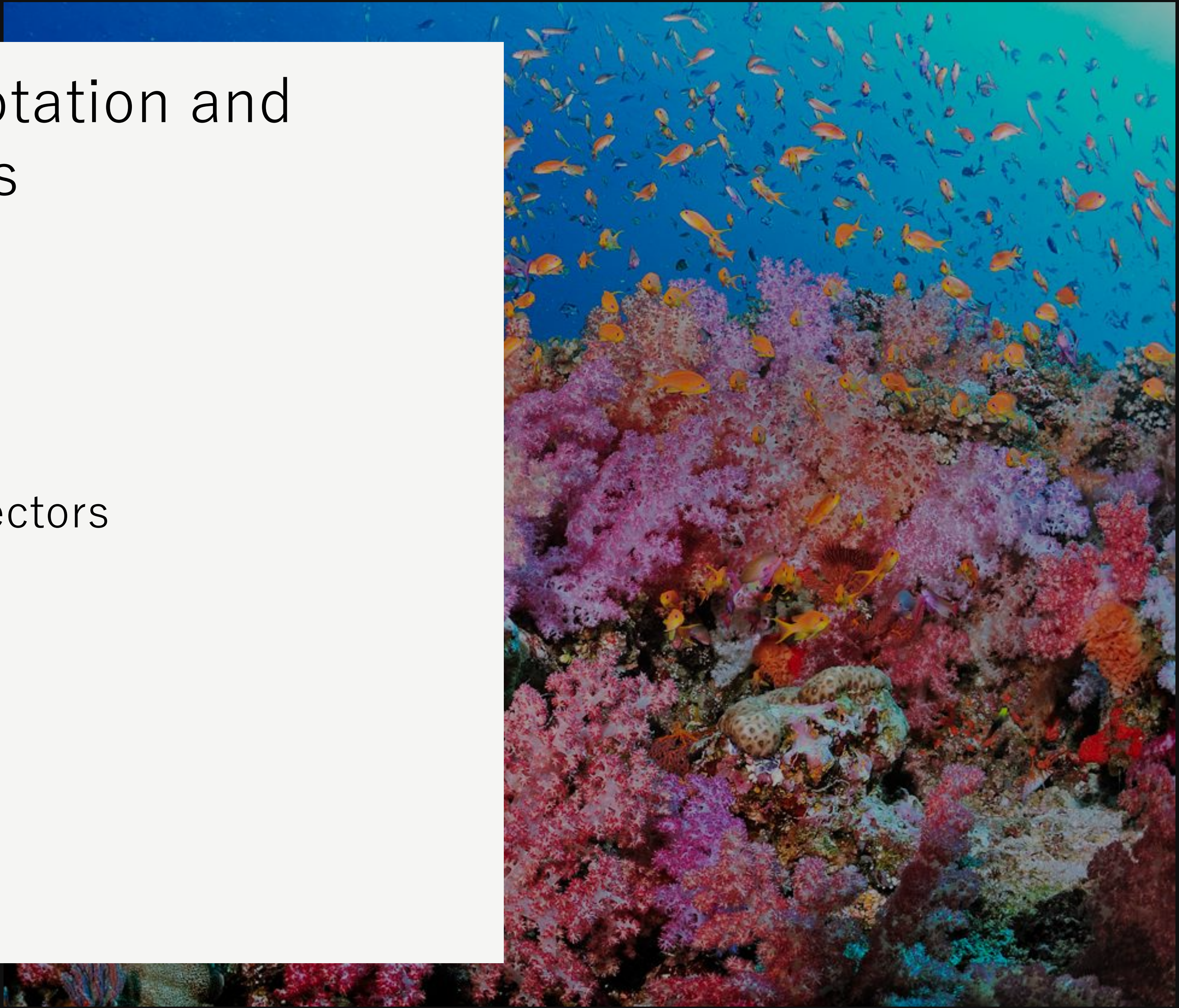
FW 599 Special Topics: Multivariate Analysis of Ecological Data in R

Lecture 3: Matrix Notation and Association Matrices

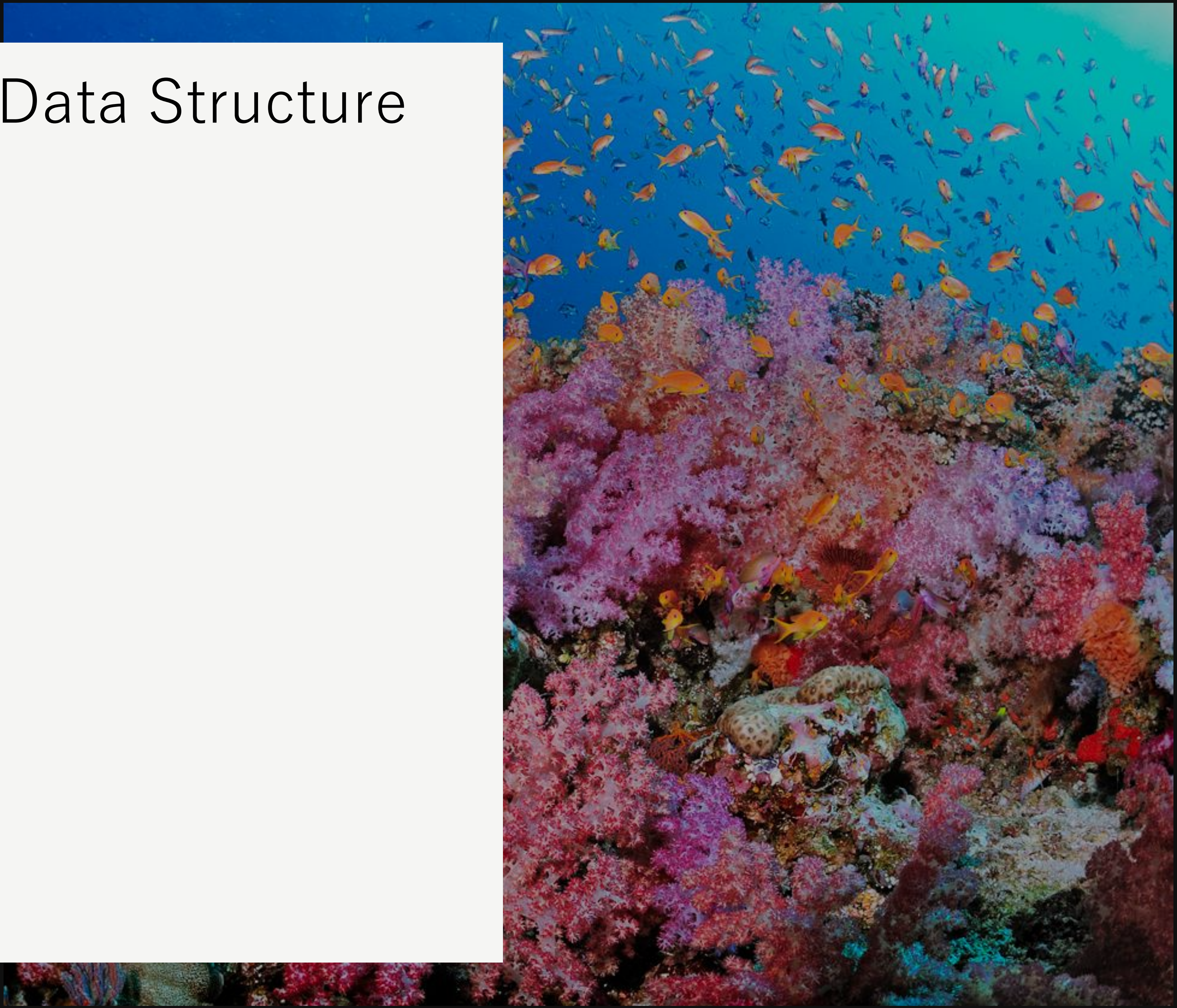
Tuesday, October 8, 2024

Lecture 3: Matrix Notation and Association Matrices

- Types of Matrices
- Matrix (Linear) Algebra
- Eigenvalues and Eigenvectors
- Association Matrices
- Q vs. R Analysis



Recap: Multivariate Data Structure

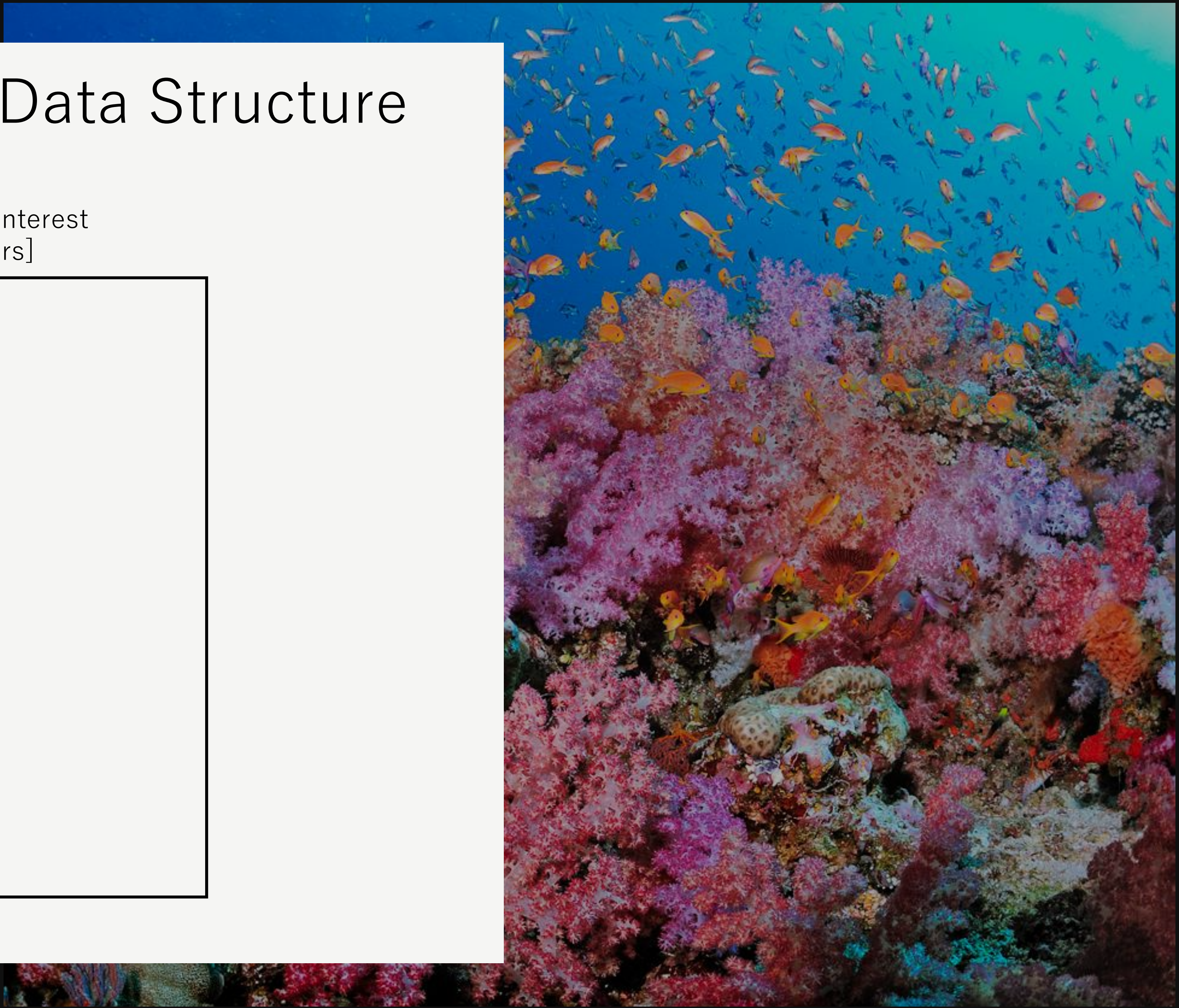


Recap: Multivariate Data Structure

Variable(s) of Interest
[Descriptors]

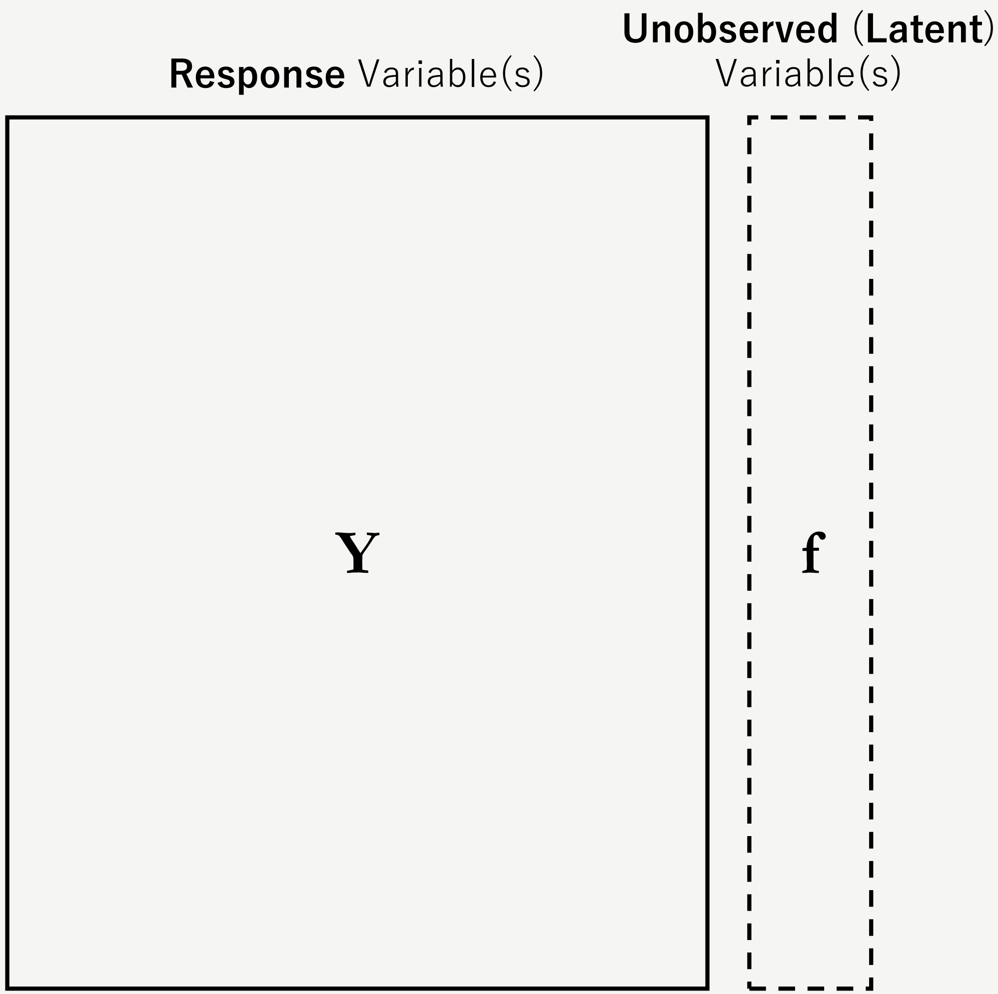
Sampling Units
[Objects]

Y



Recap: Multivariate Data Structure

Structural Methods: look for structure underlying the data matrix **Y**.



Recap: Multivariate Data Structure

Functional Methods: relate the response variable(s) **Y** as a function of the predictor variable(s) **X**.

Response Variable(s)	**Predictor** or **Explanatory** Variable(s)
$\mathbf{Y}$	$\mathbf{X}$

Matrix Notation

Matrix Notation: Ecological Matrices

Objects (row *i*): Defined *a priori.* E.g., sites, locations, individuals, observations

Variable(s) of Interest
[Descriptors]

Sampling Units
[Objects]

$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
.
.
.
.
$\mathbf{x}_i$
.
.
.
$\mathbf{x}_n$

Which variable can be increased to infinity?

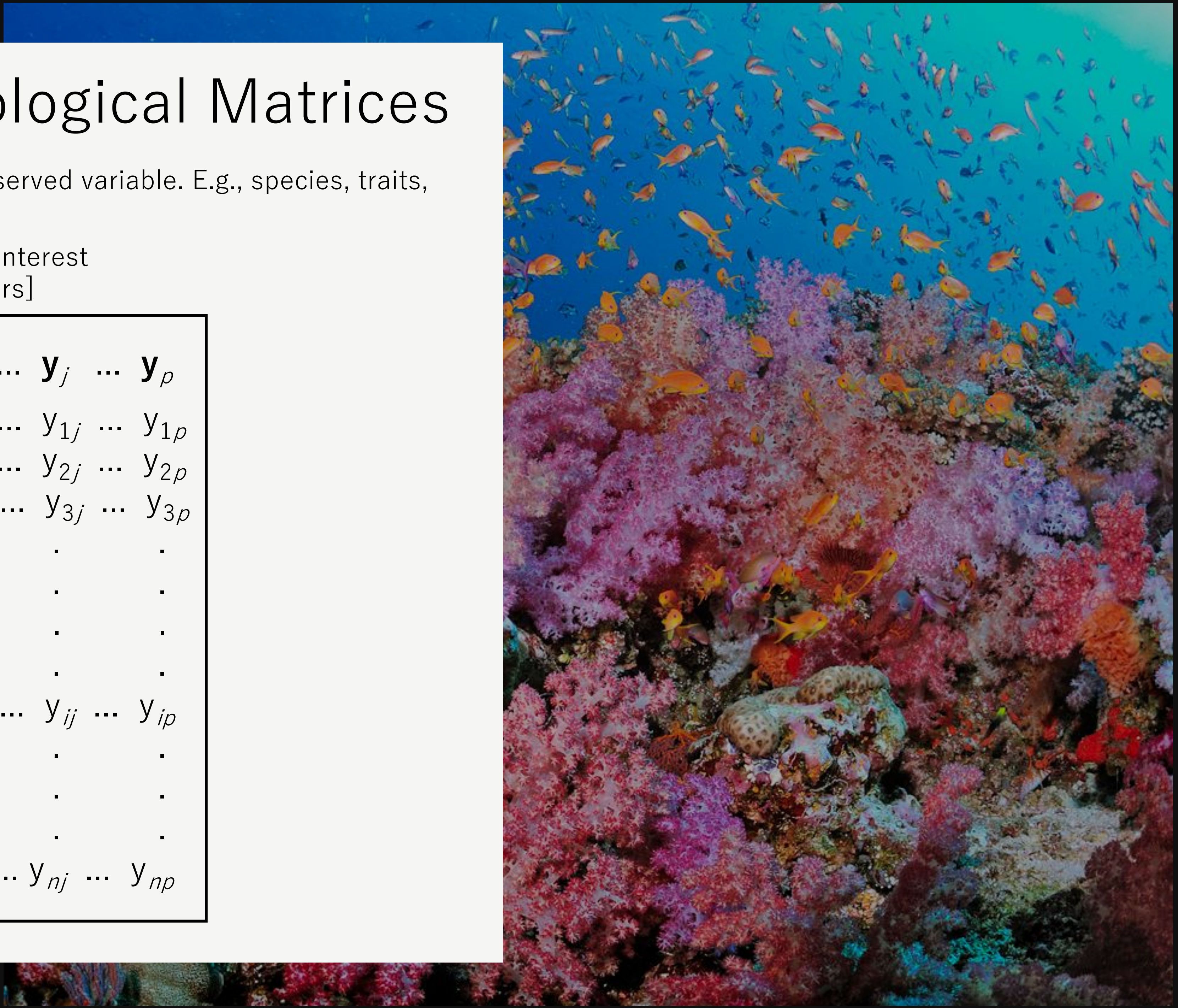
Matrix Notation: Ecological Matrices

Descriptors (column *j*): The measured or observed variable. E.g., species, traits, environmental characteristics

Variable(s) of Interest
[Descriptors]

Sampling Units
[Objects]

$$
\begin{array}{c|cccccccc|}
 & \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_j & \cdots & \mathbf{y}_p \\
\mathbf{x}_1 & y_{11} & y_{12} & y_{13} & \cdots & y_{1j} & \cdots & y_{1p} \\
\mathbf{x}_2 & y_{21} & y_{22} & y_{23} & \cdots & y_{2j} & \cdots & y_{2p} \\
\mathbf{x}_3 & y_{31} & y_{32} & y_{33} & \cdots & y_{3j} & \cdots & y_{3p} \\
. & . & . & . & & . & & . \\
. & . & . & . & & . & & . \\
. & . & . & . & & . & & . \\
. & . & . & . & & . & & . \\
\mathbf{x}_i & y_{i1} & y_{i2} & y_{i3} & \cdots & y_{ij} & \cdots & y_{ip} \\
. & . & . & . & & . & & . \\
. & . & . & . & & . & & . \\
. & . & . & . & & . & & . \\
\mathbf{x}_n & y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nj} & \cdots & y_{np}
\end{array}
$$



Matrix Notation: Ecological Matrices

A matrix of **order** (i.e., dimensions) n x p is denoted $\mathbf{Y}_{np}$ where any given **element** within the matrix is denoted y_{ij}.

Variable(s) of Interest
[Descriptors]

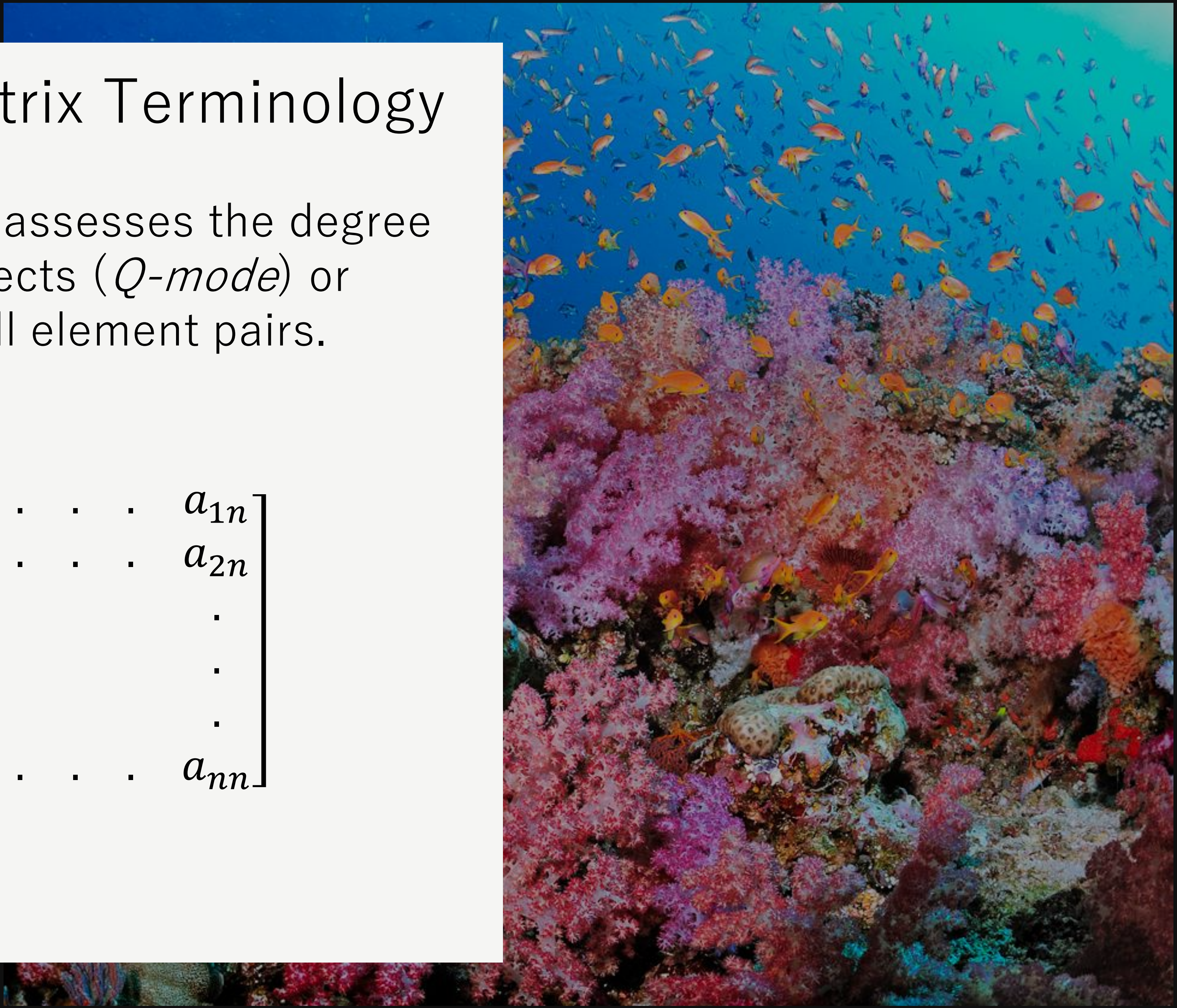
Sampling Units
[Objects]

$$
\begin{array}{c|cccccc}
 & \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 \cdots & \mathbf{y}_j & \cdots & \mathbf{y}_p \\
\hline
\mathbf{x}_1 & y_{11} & y_{12} & y_{13} \cdots & y_{1j} & \cdots & y_{1p} \\
\mathbf{x}_2 & y_{21} & y_{22} & y_{23} \cdots & y_{2j} & \cdots & y_{2p} \\
\mathbf{x}_3 & y_{31} & y_{32} & y_{33} \cdots & y_{3j} & \cdots & y_{3p} \\
. & . & . & . & . & & . \\
. & . & . & . & . & & . \\
. & . & . & . & . & & . \\
. & . & . & . & . & & . \\
\mathbf{x}_i & y_{i1} & y_{i2} & y_{i3} \cdots & y_{ij} & \cdots & y_{ip} \\
. & . & . & . & . & & . \\
. & . & . & . & . & & . \\
. & . & . & . & . & & . \\
\mathbf{x}_n & y_{n1} & y_{n2} & y_{n3} \cdots & y_{nj} & \cdots & y_{np}
\end{array}
$$

Matrix Notation: Matrix Terminology

An **association matrix (A)** assesses the degree of resemblance among objects (*Q-mode*) or descriptors (*R-mode*) for all element pairs.

$$\mathbf{A}_{nn} = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{12} & a_{22} & . & . & . & a_{2n} \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ a_{n1} & a_{n2} & . & . & . & a_{nn} \end{bmatrix}$$



Matrix Notation: Matrix Terminology

Association matrices are **square matrices** where the number of rows is equal to the number of columns (which is equal to the number of objects, *n* or descriptors, *p*).

$$\mathbf{A}_{nn} = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{12} & a_{22} & . & . & . & a_{2n} \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ a_{n1} & a_{n2} & . & . & . & a_{nn} \end{bmatrix}$$

Matrix Notation: Matrix Terminology

Association matrices are almost always **symmetric**, where elements in the upper right triangle are equal to elements in the lower left triangle ($a_{ij} = a_{ji}$).

$$\mathbf{A}_{nn} = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{12} & a_{22} & . & . & . & a_{2n} \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ a_{n1} & a_{n2} & . & . & . & a_{nn} \end{bmatrix}$$

Matrix Notation: Matrix Terminology

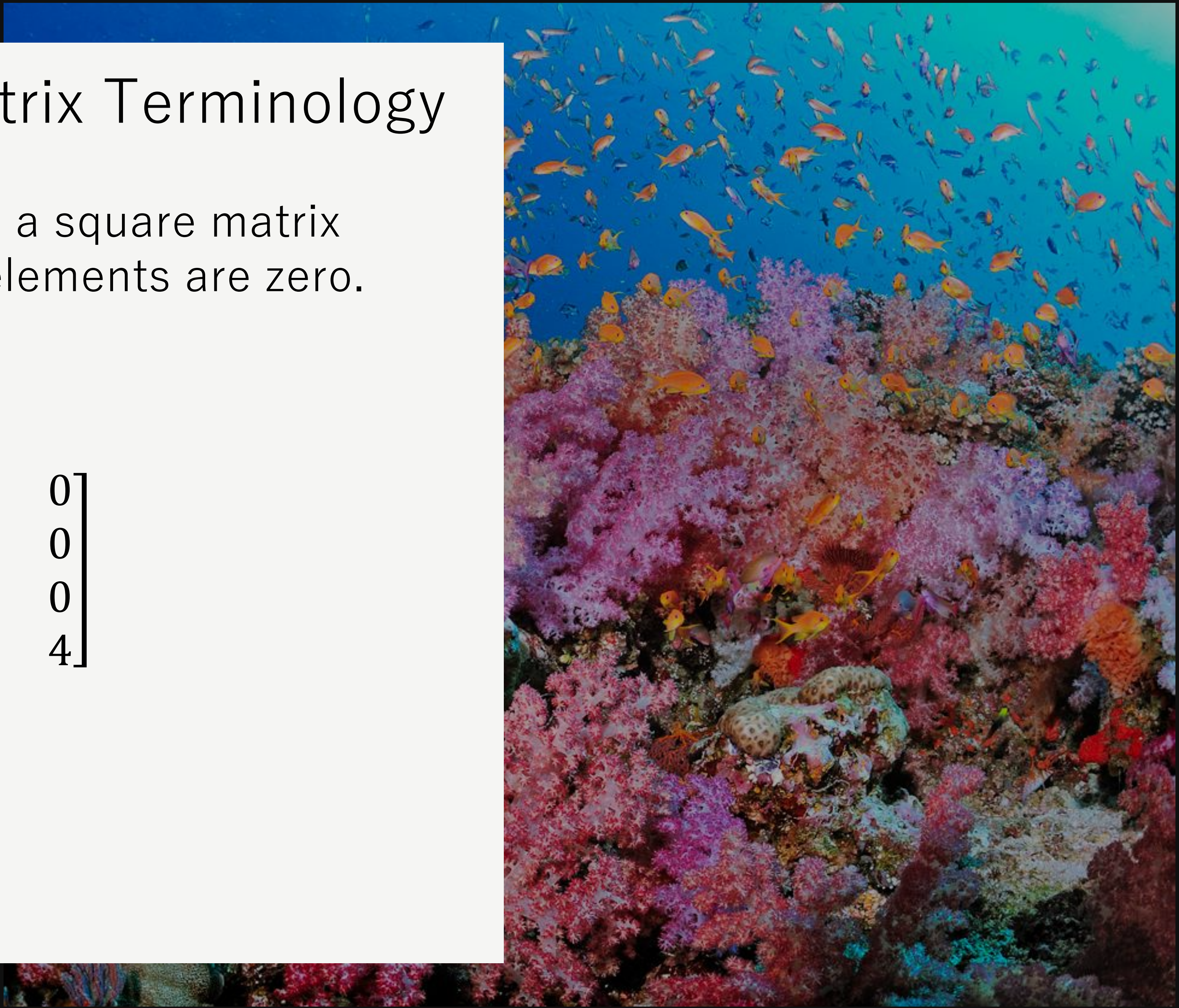
The measure of association between an object/descriptor and itself usually takes a value of **1** (similarity) or **0** (dissimilarity/distance).

$$\mathbf{A}_{nn} = \begin{bmatrix} 1 & a_{12} & . & . & . & a_{1n} \\ a_{12} & 1 & . & . & . & a_{2n} \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ a_{n1} & a_{n2} & . & . & . & 1 \end{bmatrix}$$

Matrix Notation: Matrix Terminology

- A **diagonal matrix** (**D**) is a square matrix where all non-diagonal elements are zero.

$$\begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$



Matrix Notation: Matrix Terminology

- A **diagonal matrix** (**D**) is a square matrix where all non-diagonal elements are zero.

$$\begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

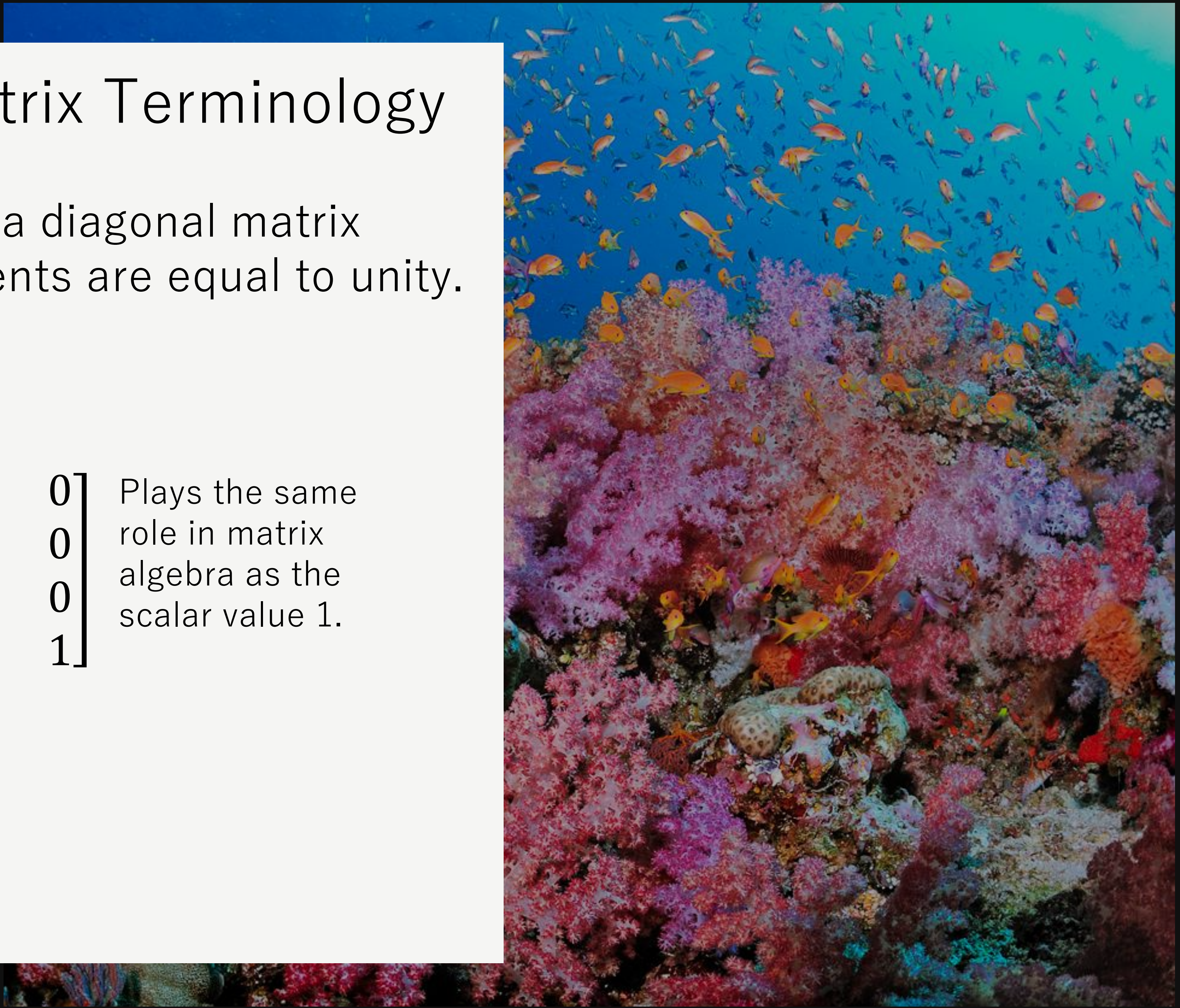
The **trace** of a matrix is the sum of its diagonal elements. In this case, trace = 20

Matrix Notation: Matrix Terminology

- An **identity matrix (I)** is a diagonal matrix where all diagonal elements are equal to unity.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Plays the same role in matrix algebra as the scalar value 1.

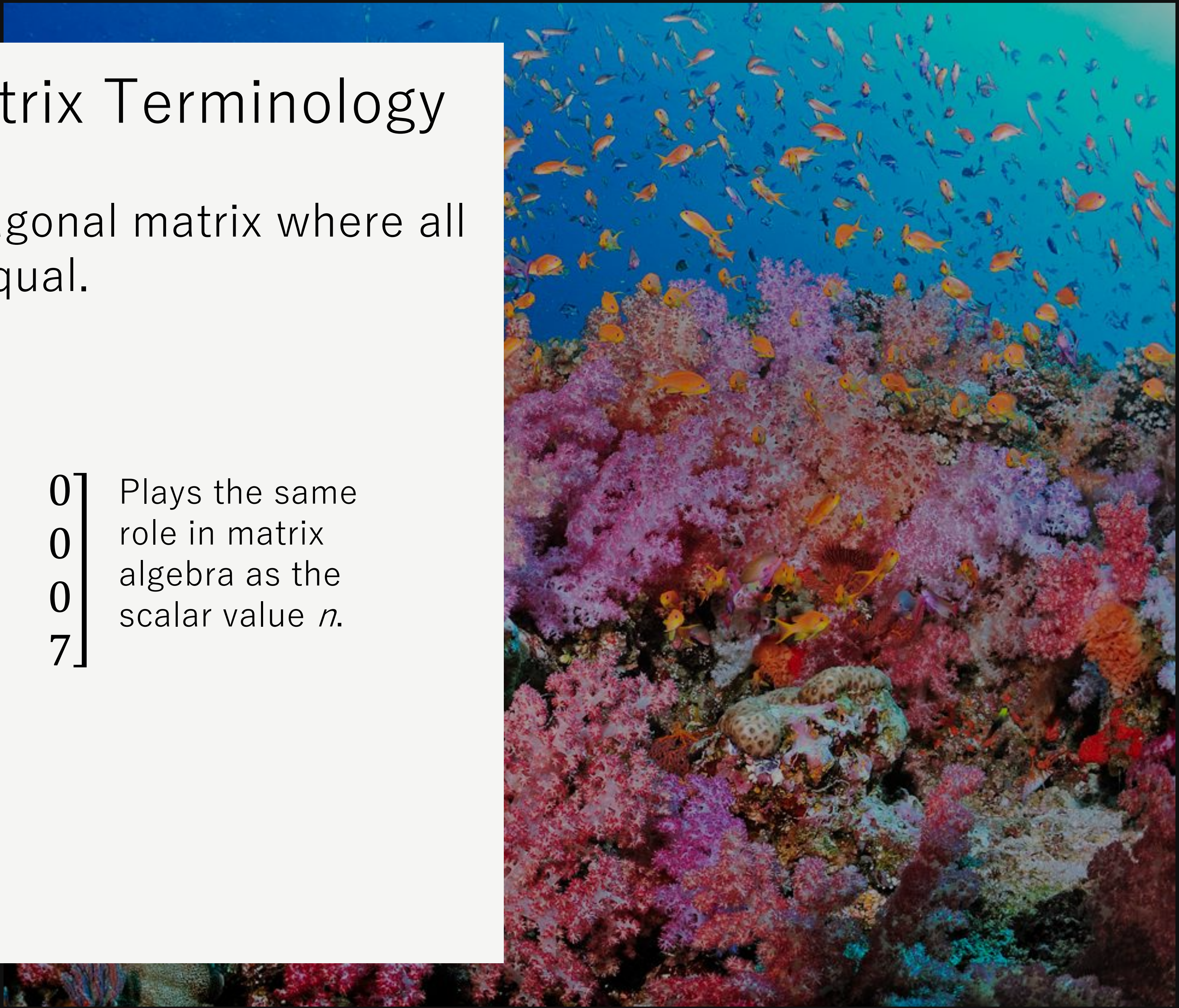


Matrix Notation: Matrix Terminology

- An **scalar matrix** is a diagonal matrix where all diagonal elements are equal.

$$\begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

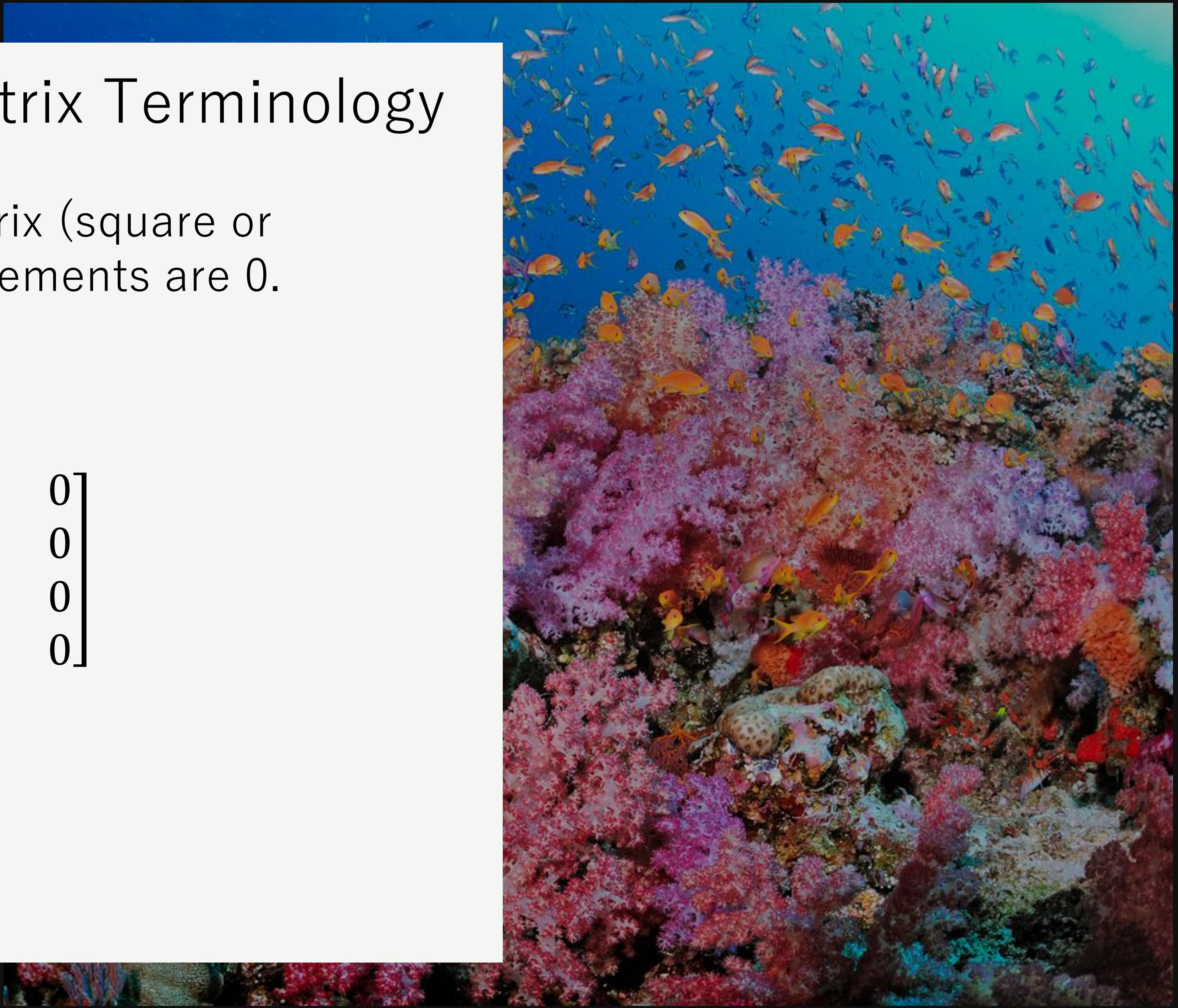
Plays the same role in matrix algebra as the scalar value *n*.



Matrix Notation: Matrix Terminology

- A **null matrix** is any matrix (square or rectangular) where all elements are 0.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



Matrix Notation: Matrix Terminology

- The **transpose** of a matrix **B** is denoted **B’** in which ‘$b_{ij} = b_{ji}$.

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$$

$$\mathbf{B}' = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

Matrix Notation: Matrix Terminology

- A square matrix that is equal to its transpose is **symmetric**.

$$\mathbf{B} = \begin{bmatrix} 1 & 4 & 7 \\ 4 & 1 & 2 \\ 7 & 2 & 1 \end{bmatrix}$$

Association matrices are almost always symmetric across the diagonal

Matrix Notation: Vectors and Scaling

- A **vector** is a **column matrix** with format (n x 1) or **row matrix** with format (1 x p).

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{bmatrix}$$

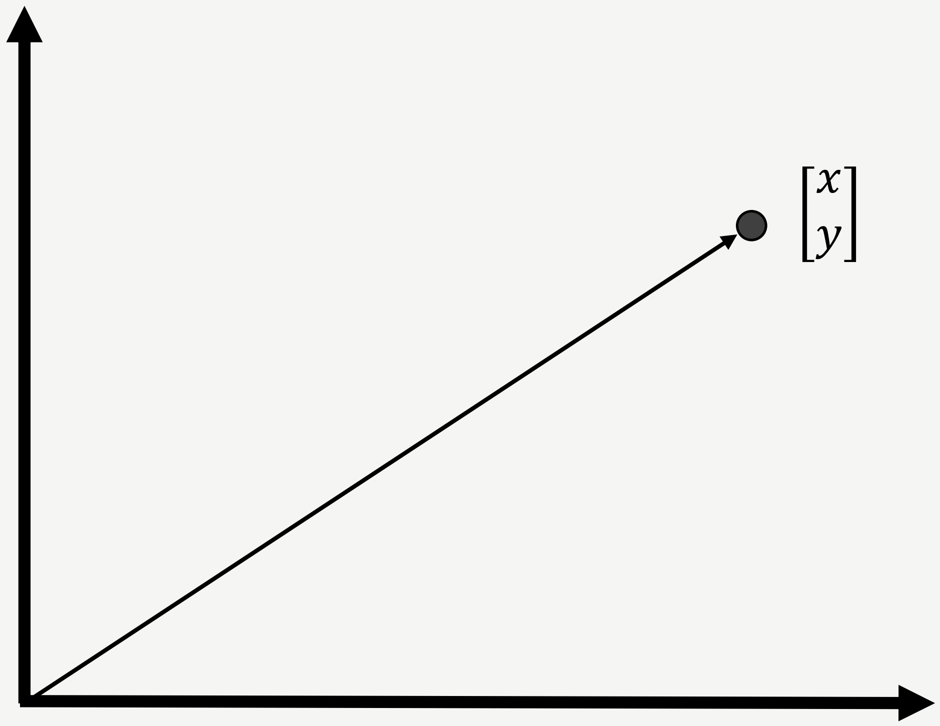
Matrix Notation: Vectors and Scaling

- A **vector** is a **column matrix** with format (*n* x 1) or **row matrix** with format (1 x *p*).

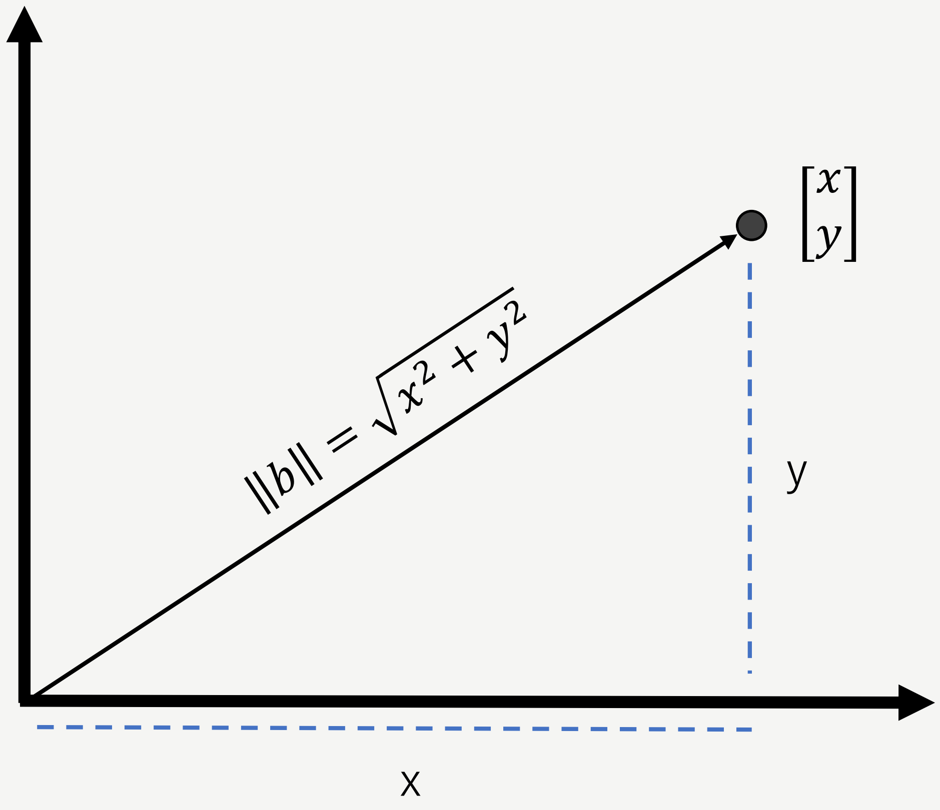
$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{bmatrix}$$

A vector graphically refers to the end-point of a line segment in *n*-dimensional Euclidean space.

Matrix Notation: Vectors and Scaling



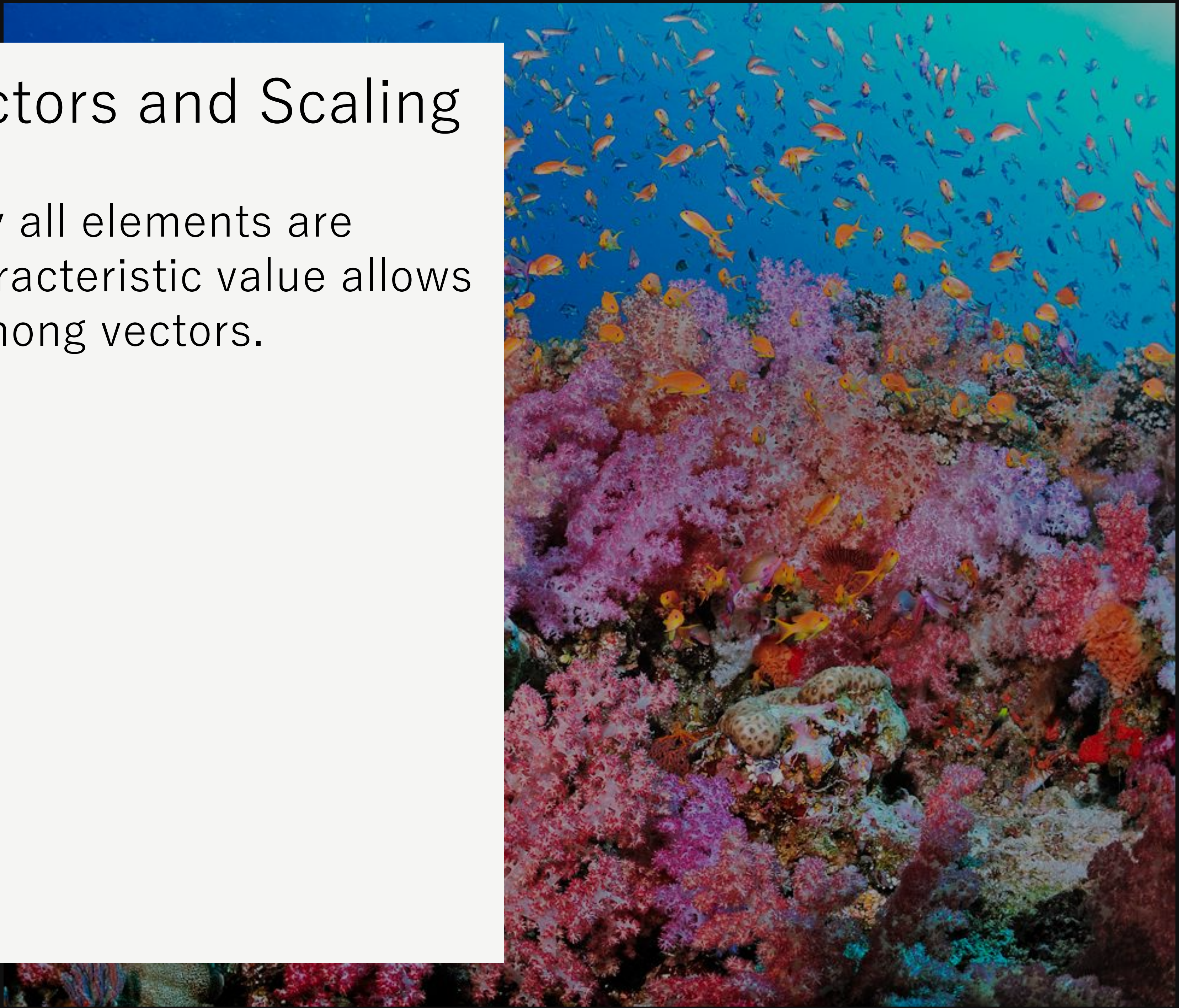
Matrix Notation: Vectors and Scaling



Thus, the length (or **norm**) of any vector (‖ **b** ‖) can be calculated using Pythagorean's theorem

Matrix Notation: Vectors and Scaling

- A **scaled** vector whereby all elements are divided by the same characteristic value allows for direct comparison among vectors.



Matrix Notation: Vectors and Scaling

- A **scaled** vector whereby all elements are divided by the same characteristic value allows for direct comparison among vectors.

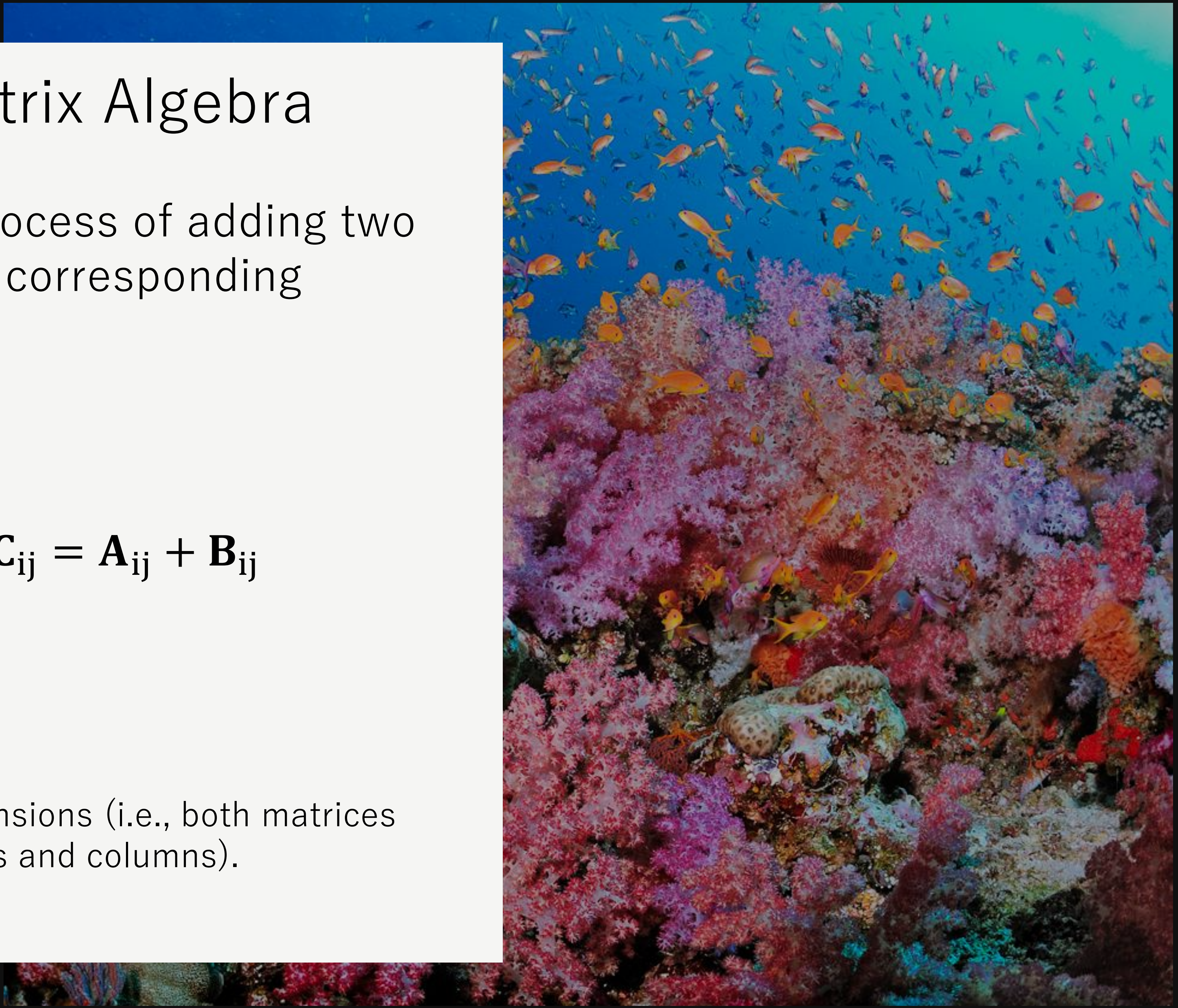
- A **normalized** vector is scaled by the length of the vector ($\| b \|$). *The length of any normalized vector in n-dimensional space is equal to 1.*

Matrix Notation: Matrix Algebra

- Matrix **addition** is the process of adding two matrices by adding their corresponding elements.

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \; \textit{where} \; \mathbf{C}_{\mathrm{ij}} = \mathbf{A}_{\mathrm{ij}} + \mathbf{B}_{\mathrm{ij}}$$

Matrices must be of the same dimensions (i.e., both matrices must have the same number of rows and columns).



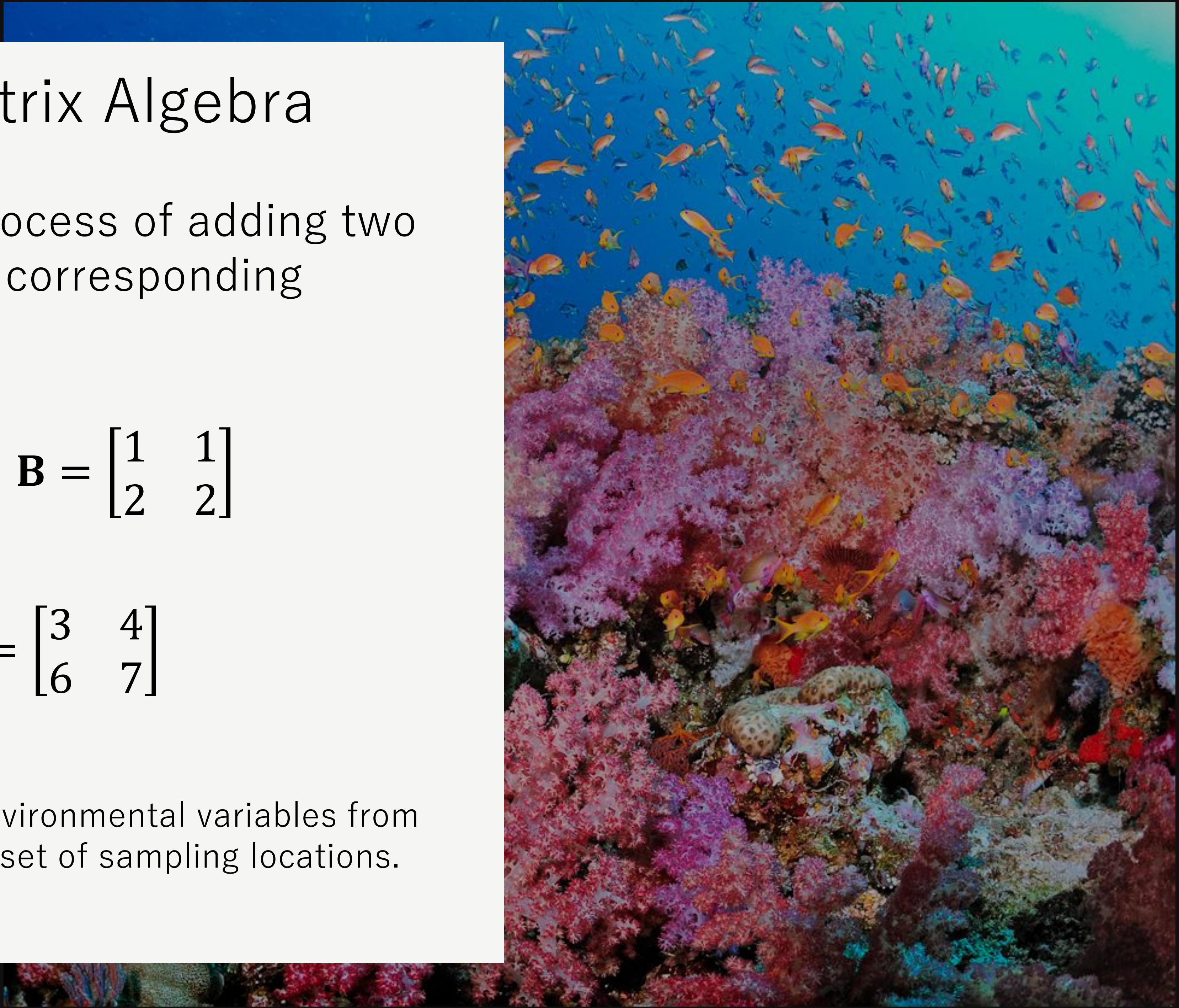
Matrix Notation: Matrix Algebra

- Matrix **addition** is the process of adding two matrices by adding their corresponding elements.

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 3 & 4 \\ 6 & 7 \end{bmatrix}$$

Example application: combining environmental variables from two different datasets for the same set of sampling locations.

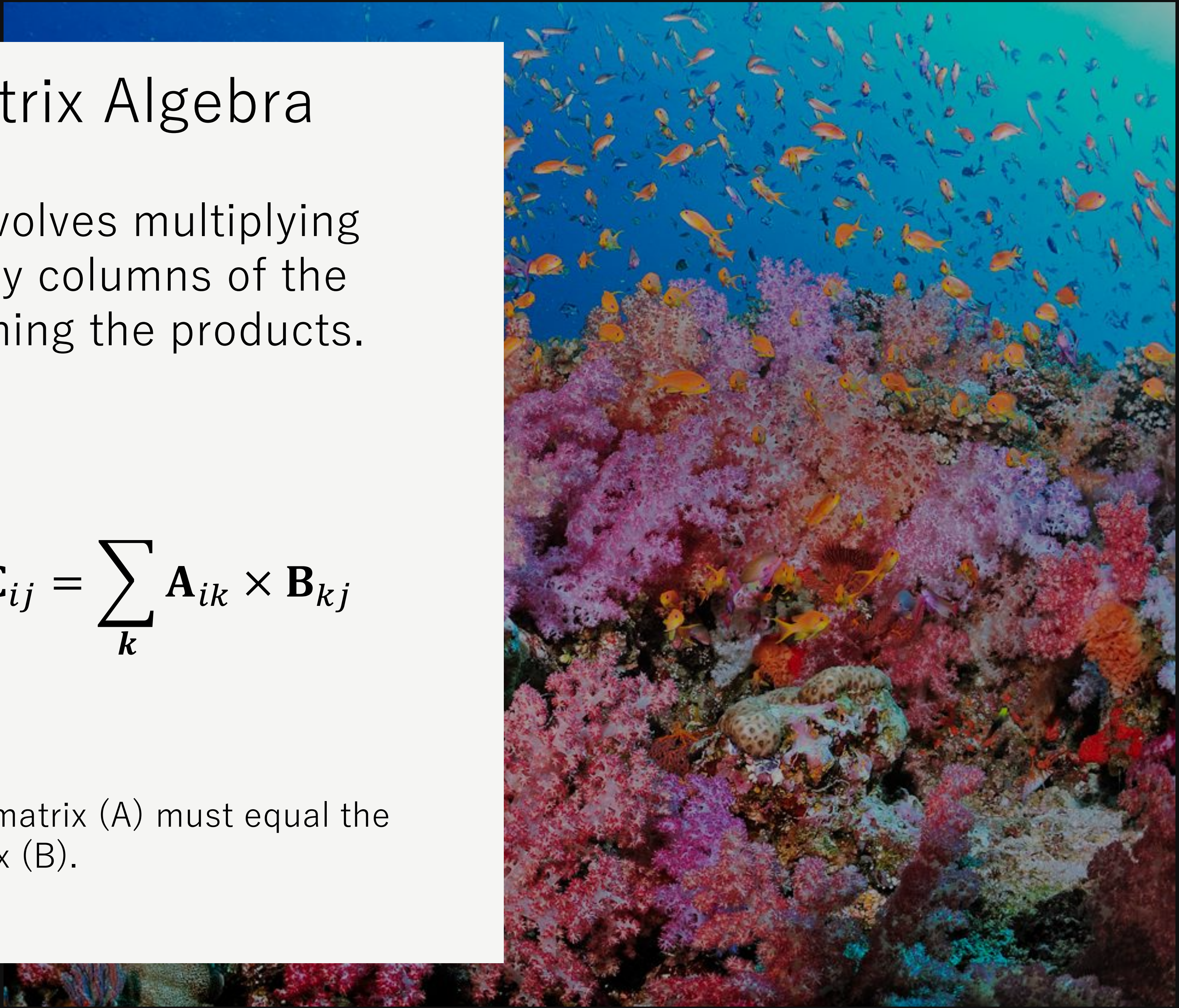


Matrix Notation: Matrix Algebra

- Matrix **multiplication** involves multiplying rows of the first matrix by columns of the second matrix and summing the products.

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} \; where \; \mathbf{C}_{ij} = \sum_{\boldsymbol{k}} \mathbf{A}_{ik} \times \mathbf{B}_{kj}$$

The number of columns in the first matrix (A) must equal the number of rows in the second matrix (B).

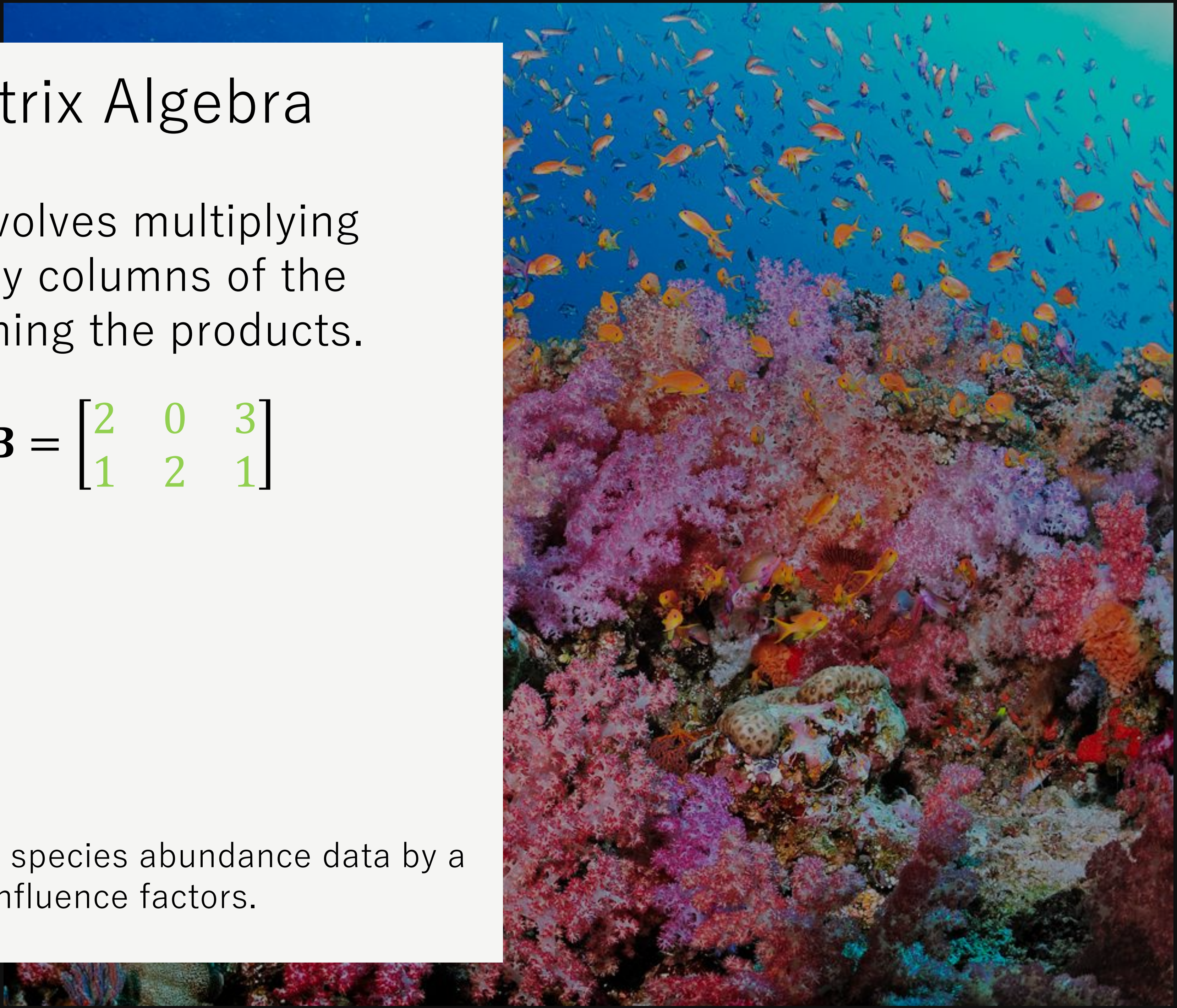


Matrix Notation: Matrix Algebra

- Matrix **multiplication** involves multiplying rows of the first matrix by columns of the second matrix and summing the products.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 2 & 0 & 3 \\ 1 & 2 & 1 \end{bmatrix}$$

Example application: Transforming species abundance data by a matrix representing environmental influence factors.



Matrix Notation: Matrix Algebra

- Matrix **multiplication** involves multiplying rows of the first matrix by columns of the second matrix and summing the products.

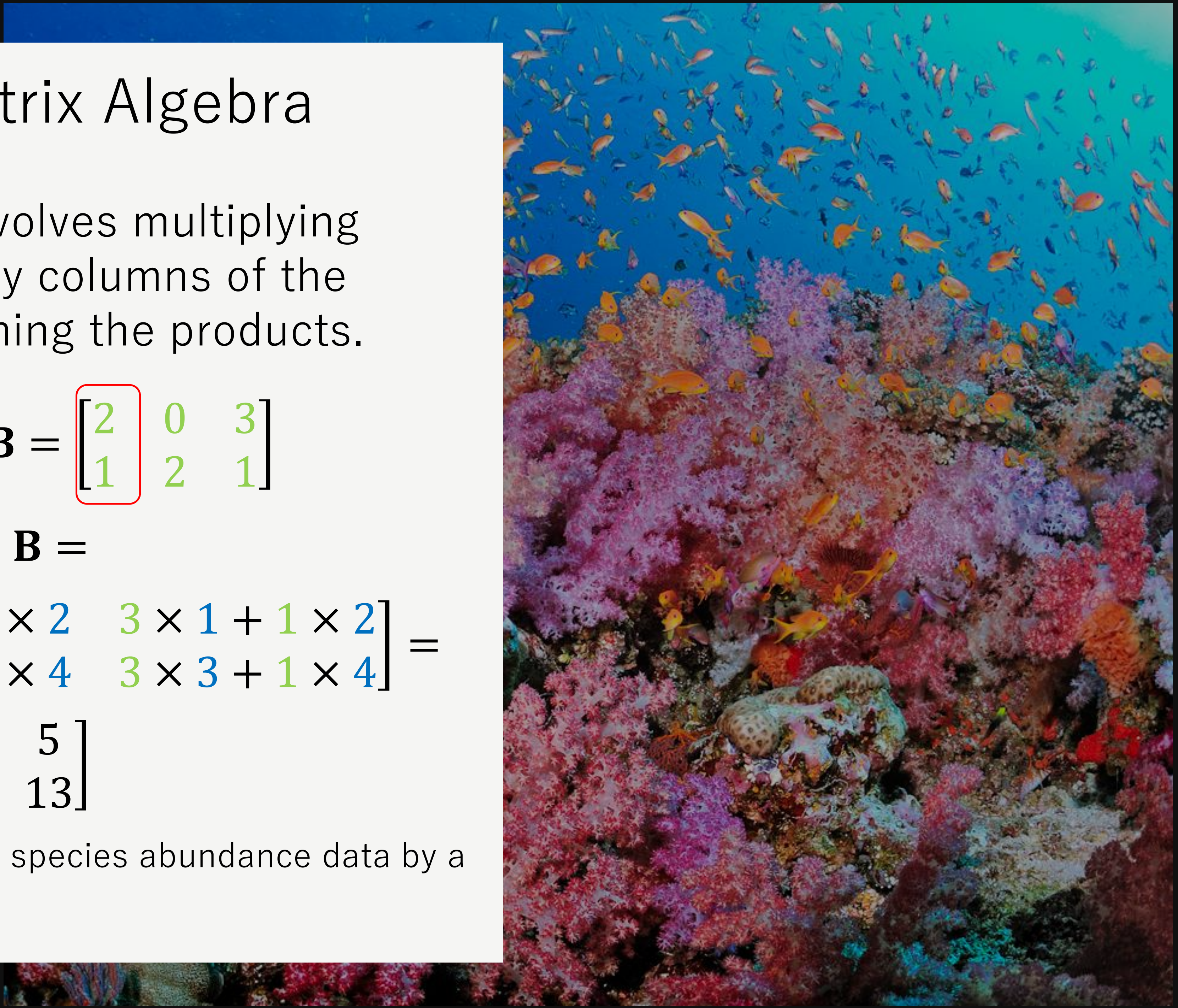
$$\mathbf{A} = \begin{bmatrix} \boxed{1 \quad 2} \\ 3 \quad 4 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \boxed{\begin{matrix} 2 \\ 1 \end{matrix}} & \begin{matrix} 0 \\ 2 \end{matrix} & \begin{matrix} 3 \\ 1 \end{matrix} \end{bmatrix}$$

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} =$$

$$\begin{bmatrix} 2 \times 1 + 1 \times 2 & 0 \times 1 + 2 \times 2 & 3 \times 1 + 1 \times 2 \\ 2 \times 3 + 1 \times 4 & 0 \times 3 + 2 \times 4 & 3 \times 3 + 1 \times 4 \end{bmatrix} =$$

$$\begin{bmatrix} 4 & 4 & 5 \\ 10 & 8 & 13 \end{bmatrix}$$

Example application: Transforming species abundance data by a matrix representing sampling effort.



Matrix Notation: Matrix Algebra

- The **dot product** (or **scalar product**) of two vectors is a single number obtained by multiplying corresponding entries and summing those products.

$$\mathbf{b} \bullet \mathbf{c} = \sum_{i=1}^{n} \mathrm{b}_i \times \mathrm{c}_i$$

Matrix Notation: Matrix Algebra

- The **dot product** (or **scalar product**) of two vectors is a single number obtained by multiplying corresponding entries and summing those products.

$$\mathbf{b} = \begin{bmatrix} 0.3 \\ 0.5 \\ 0.2 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} 10 \\ 15 \\ 5 \end{bmatrix}$$

$$\mathbf{b} \bullet \mathbf{c} = 0.3 \times 10 + 0.5 \times 15 + 0.2 \times 5 = \mathbf{11.5}$$

Example application: Transforming species abundance data by a matrix representing sampling effort.

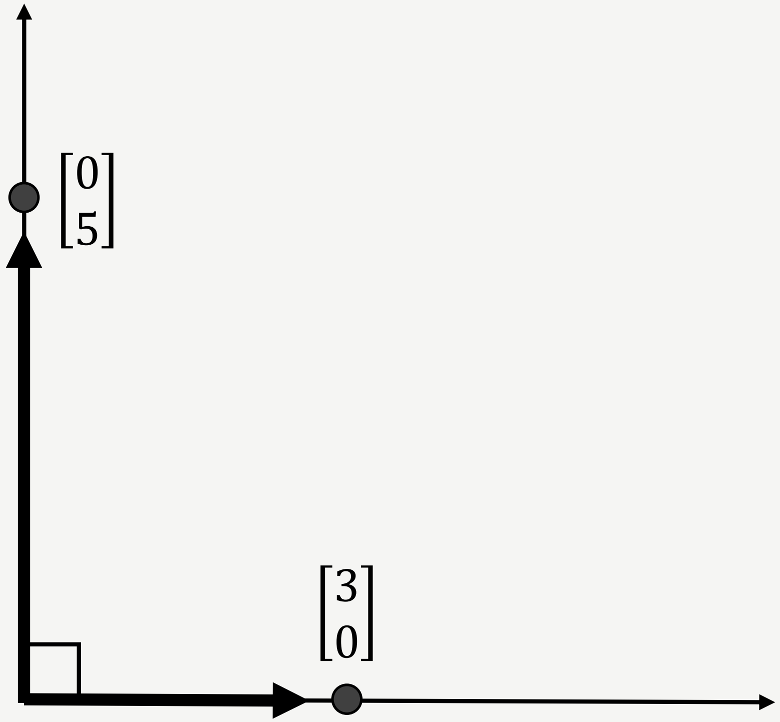
Matrix Notation: Matrix Algebra

- If the scalar product is zero, the two vectors are said to be **orthogonal** (at a 90° angle) from one-another.

$$\mathbf{b} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

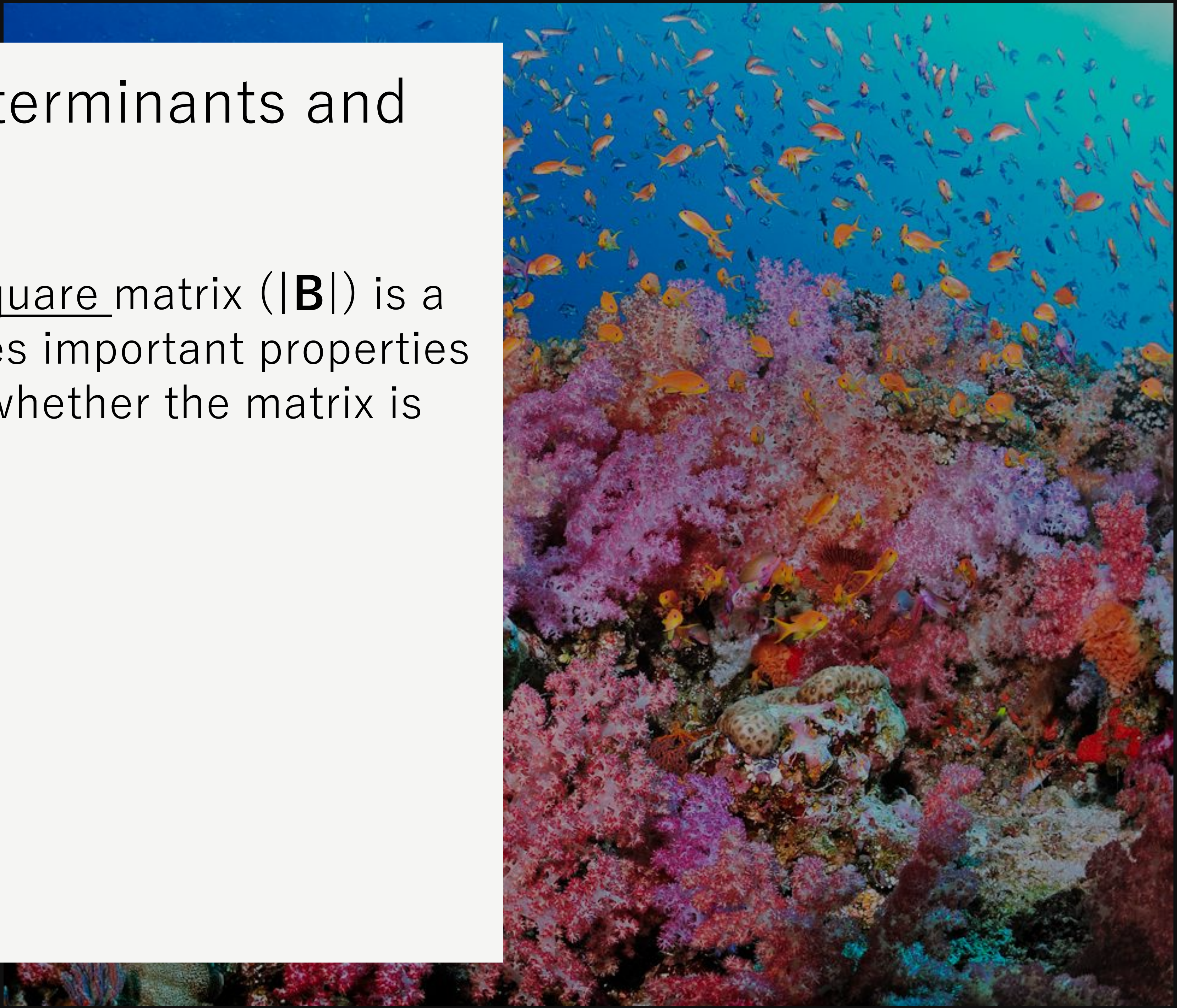
$$\mathbf{c} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

$$\mathbf{b} \bullet \mathbf{c} =$$
$$3 \times 0 + 0 \times 5 = \mathbf{0}$$



Matrix Notation: Determinants and Ranks

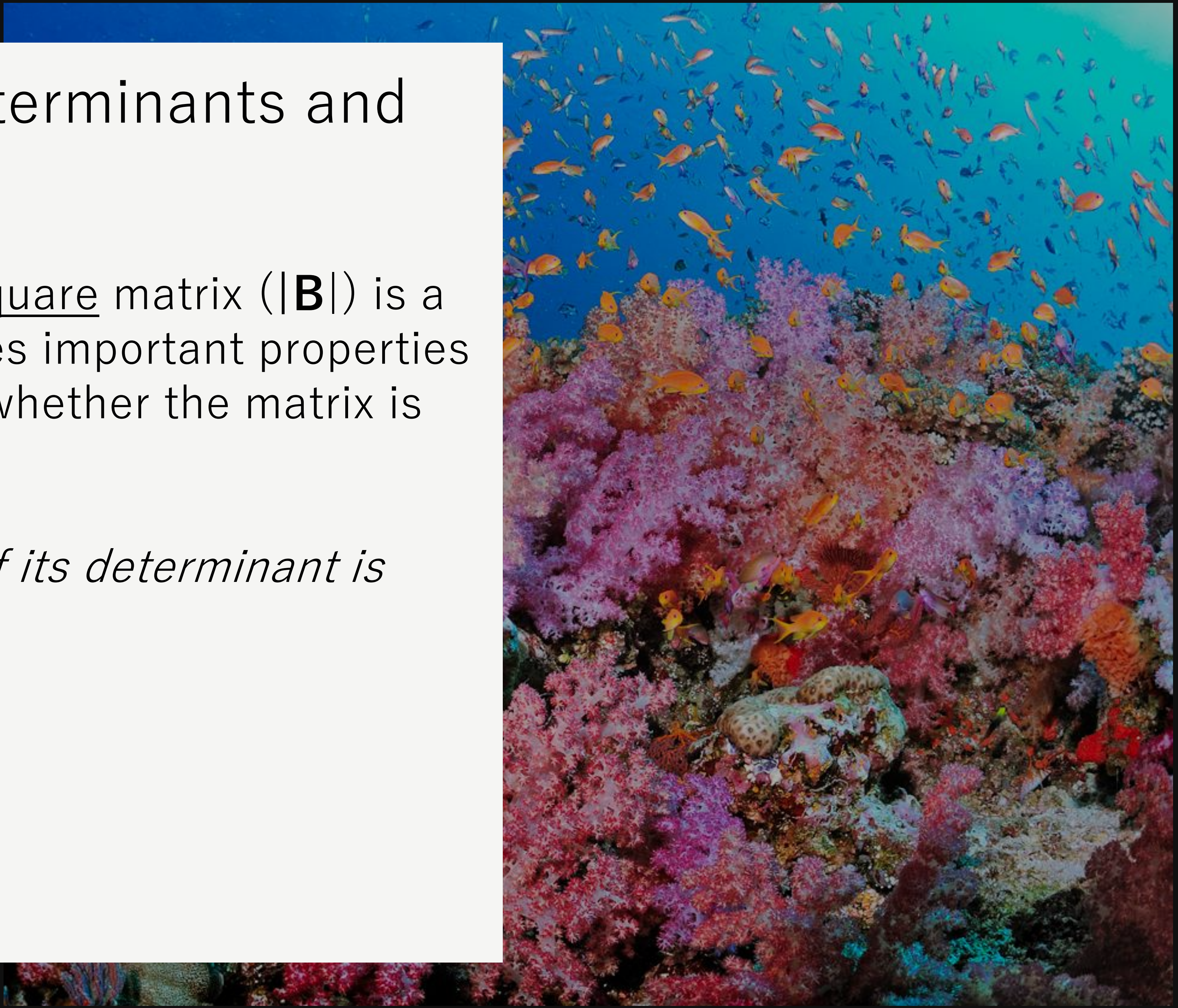
- The **determinant** of a square matrix ($|\mathbf{B}|$) is a scalar value that captures important properties of the matrix, including whether the matrix is invertible.



Matrix Notation: Determinants and Ranks

- The **determinant** of a square matrix (|**B**|) is a scalar value that captures important properties of the matrix, including whether the matrix is invertible.

A matrix is invertible only if its determinant is non-zero!



Matrix Notation: Determinants and Ranks

- For a 2 x 2 matrix: $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

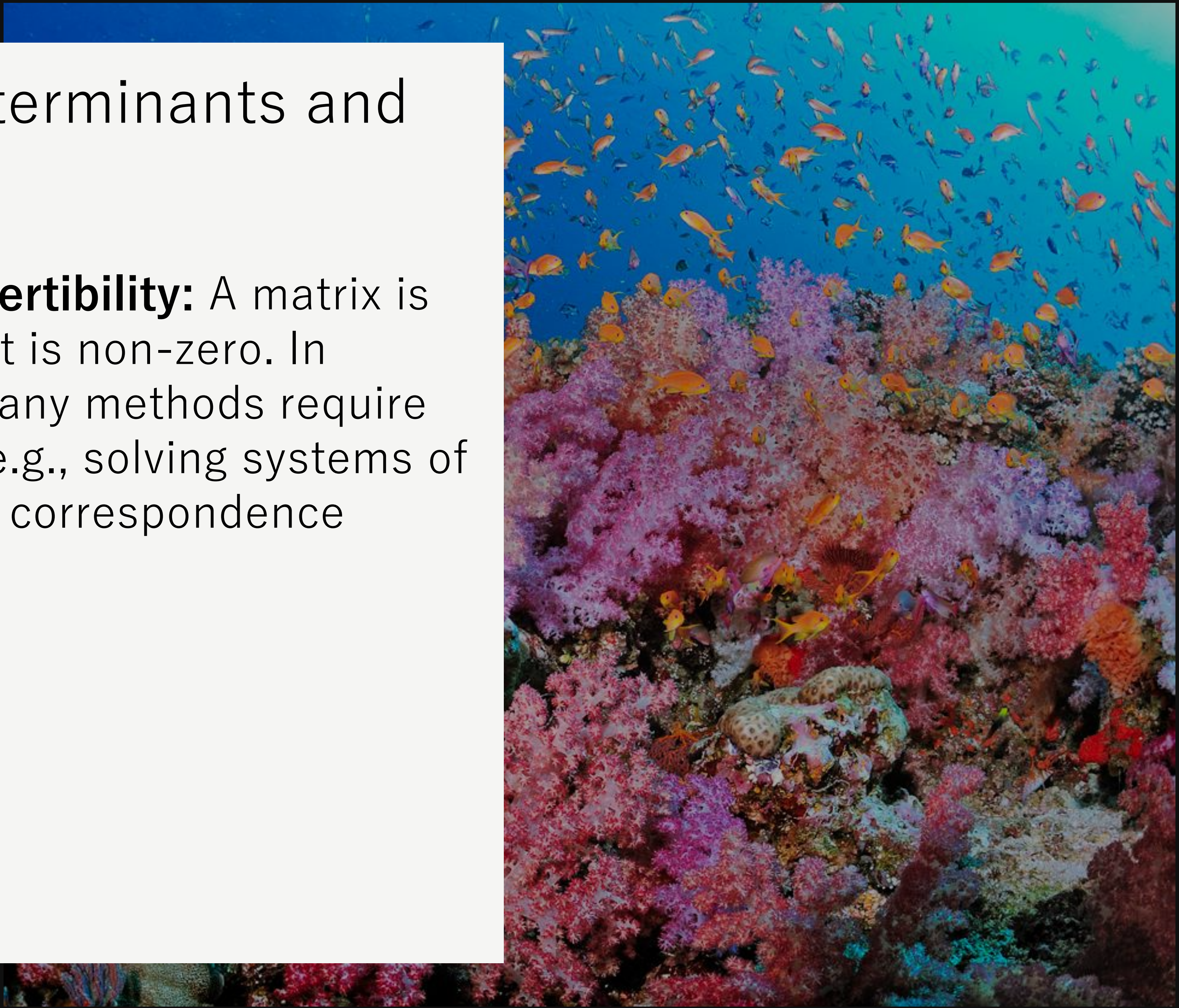
$$|\mathbf{A}| = ad - bc$$

- For a 3 x 3 matrix: $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

$$|\mathbf{A}| = a(ei - fh) - b(di - fg) + c(dh - eg)$$

Matrix Notation: Determinants and Ranks

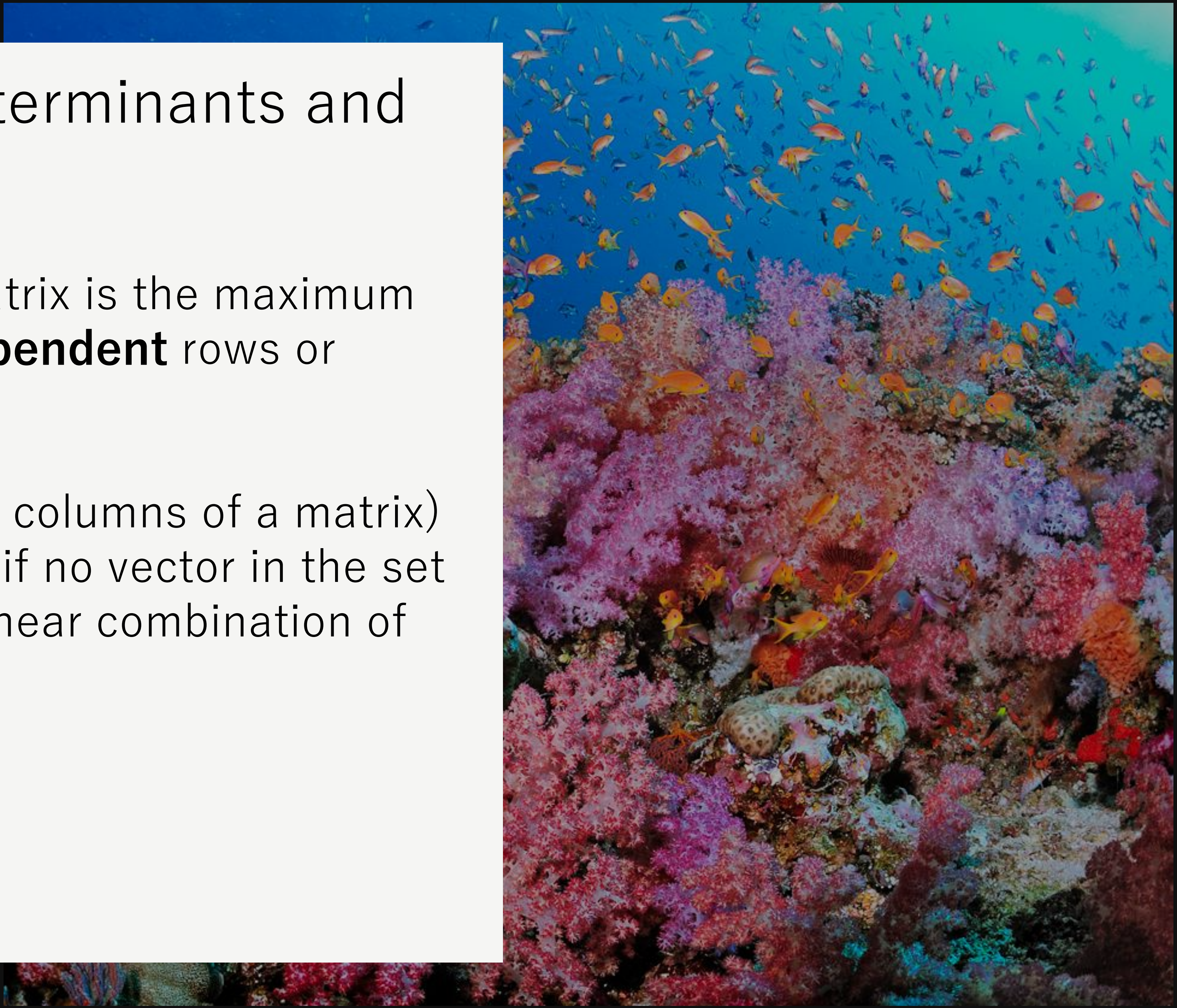
Understanding Matrix Invertibility: A matrix is **invertible** if its determinant is non-zero. In ecological data analysis, many methods require the inversion of matrices (e.g., solving systems of linear equations, canonical correspondence analysis).



Matrix Notation: Determinants and Ranks

- The **rank** of a <u>square</u> matrix is the maximum number of **linearly independent** rows or columns in the matrix.

- A set of vectors (rows or columns of a matrix) is **linearly independent** if no vector in the set can be expressed as a linear combination of the others.



Matrix Notation: Determinants and Ranks

- The **rank** of a square matrix is the maximum number of **linearly independent** rows or columns in the matrix.

- A set of vectors (rows or columns of a matrix) is **linearly independent** if no vector in the set can be expressed as a linear combination of the others.

A matrix with a rank lower than its order has a determinant of zero and is not invertible.

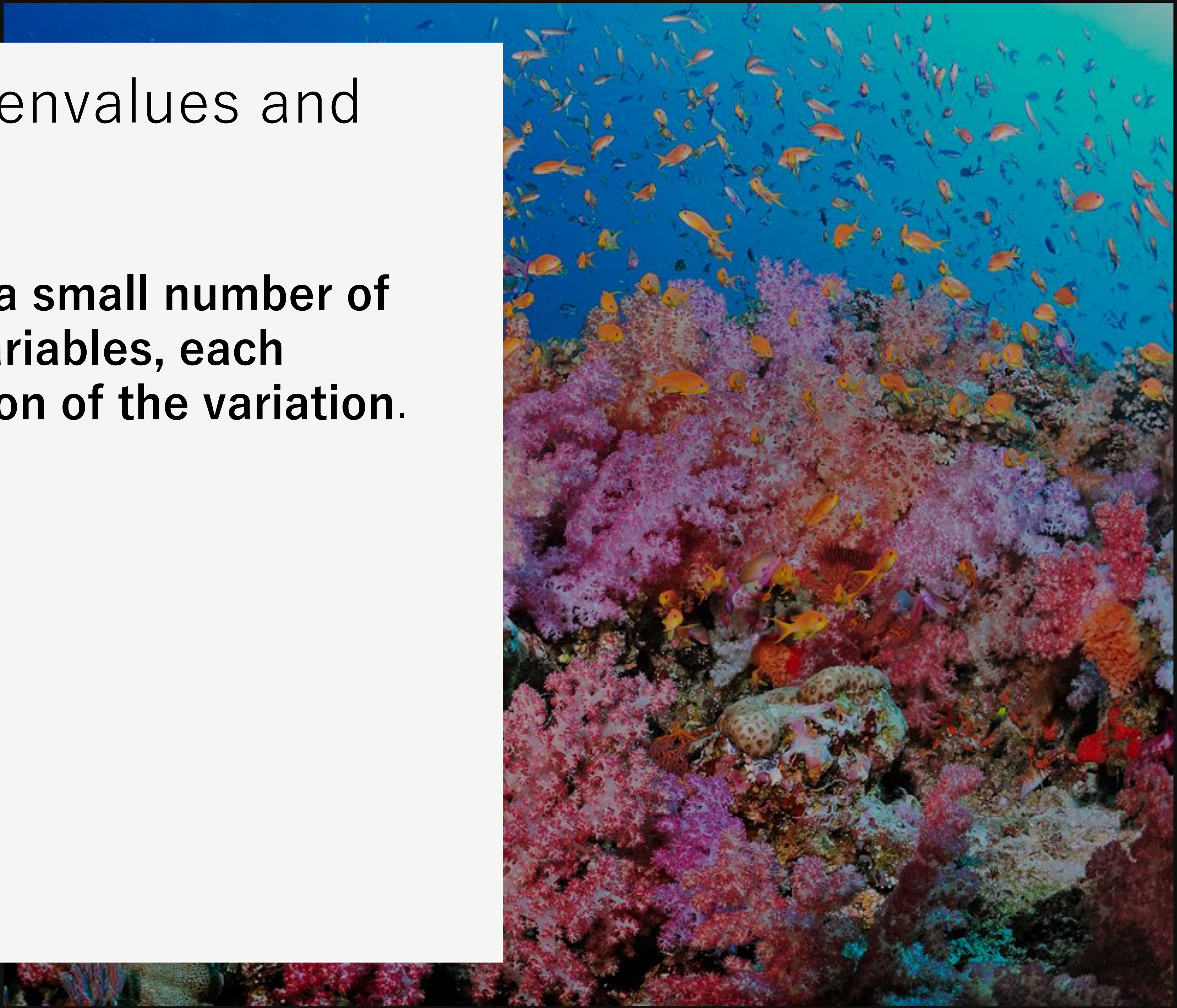
Matrix Notation: Eigenvalues and Eigenvectors

- **Eigenvalues** ($\boldsymbol{\lambda}$) are scalars that satisfy the equation $\mathbf{Av}=\boldsymbol{\lambda}\,\mathbf{u}$, where **A** is a square matrix (for example, an association matrix) and **u** is a non-zero vector.

- **Eigenvectors** (**u**) are non-zero vectors that, when multiplied by the matrix **A**, result in a vector that is a scalar multiple of itself.

Matrix Notation: Eigenvalues and Eigenvectors

- The goal is to **generate a small number of linearly independent variables, each explaining a large portion of the variation**.



Matrix Notation: Eigenvalues and Eigenvectors

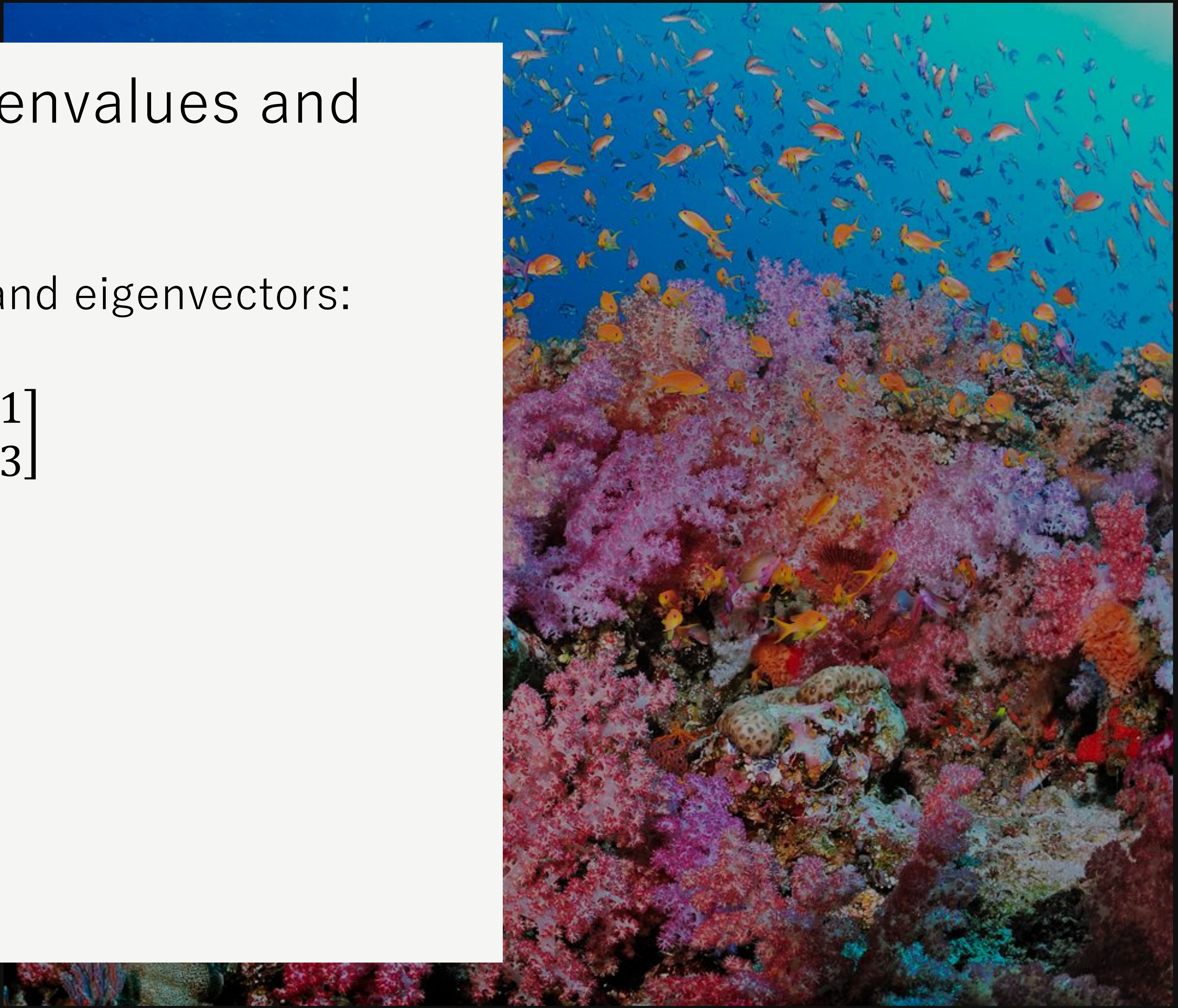
- The goal is to **generate a small number of linearly independent variables, each explaining a large portion of the variation**.

- i.e., generate a diagonal matrix equivalent to the square matrix **A**

Matrix Notation: Eigenvalues and Eigenvectors

- Solving for eigenvalues and eigenvectors:

$$\mathbf{A} = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$



Matrix Notation: Eigenvalues and Eigenvectors

- Solving for eigenvalues and eigenvectors:

1) Form the characteristic equation

$$|\mathbf{A} - \lambda \, \mathbf{I}| = \begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} = 0$$

Matrix Notation: Eigenvalues and Eigenvectors

- Solving for eigenvalues and eigenvectors:

2) Solve for eigenvalues ($\boldsymbol{\lambda}$)

$$|\mathbf{A} - \lambda\,\mathbf{I}| = \begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} = 0$$

$$(4-\boldsymbol{\lambda}) \times (3-\boldsymbol{\lambda}) - 2 \times 1 = 0$$

$$\boldsymbol{\lambda}^2 - 7\boldsymbol{\lambda} + 10 = 0$$

$$(\boldsymbol{\lambda} - 2) \times (\boldsymbol{\lambda} - 5) = 0$$

Matrix Notation: Eigenvalues and Eigenvectors

- Solving for eigenvalues and eigenvectors:

2) Solve for eigenvalues ($\boldsymbol{\lambda}$)

$$|\mathbf{A} - \lambda\,\mathbf{I}| = \begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} = 0$$

$$(4-\boldsymbol{\lambda}) \times (3-\boldsymbol{\lambda}) - 2 \times 1 = 0$$

$$\boldsymbol{\lambda}^2 - 7\boldsymbol{\lambda} + 10 = 0$$

$$(\boldsymbol{\lambda} - 2) \times (\boldsymbol{\lambda} - 5) = 0$$

$$\lambda_1 = 2, \lambda_2 = 5$$

Matrix Notation: Eigenvalues and Eigenvectors

- Solving for eigenvalues and eigenvectors:

3) Solve for eigenvectors ($\mathbf{u}$)

$$(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I})\mathbf{u} = 0$$

$$(\mathbf{A} - \boldsymbol{\lambda}_1\mathbf{I}) = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \quad (\mathbf{A} - \boldsymbol{\lambda}_2\mathbf{I}) = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

Matrix Notation: Eigenvalues and Eigenvectors

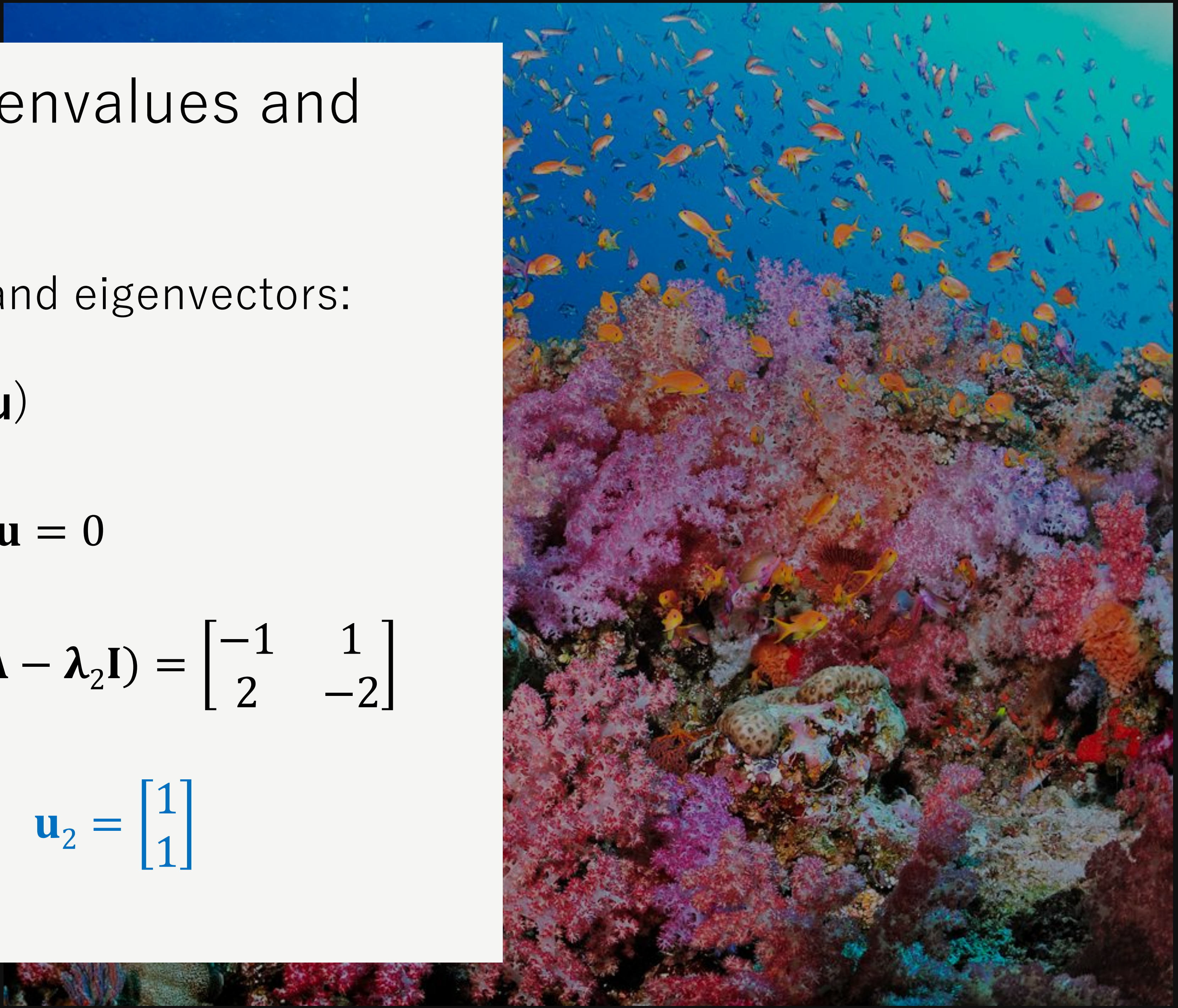
- Solving for eigenvalues and eigenvectors:

3) Solve for eigenvectors ($\mathbf{u}$)

$$(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I})\mathbf{u} = 0$$

$$(\mathbf{A} - \boldsymbol{\lambda}_1\mathbf{I}) = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \quad (\mathbf{A} - \boldsymbol{\lambda}_2\mathbf{I}) = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

$$\mathbf{u}_1 = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$



Matrix Notation: Eigenvalues and Eigenvectors

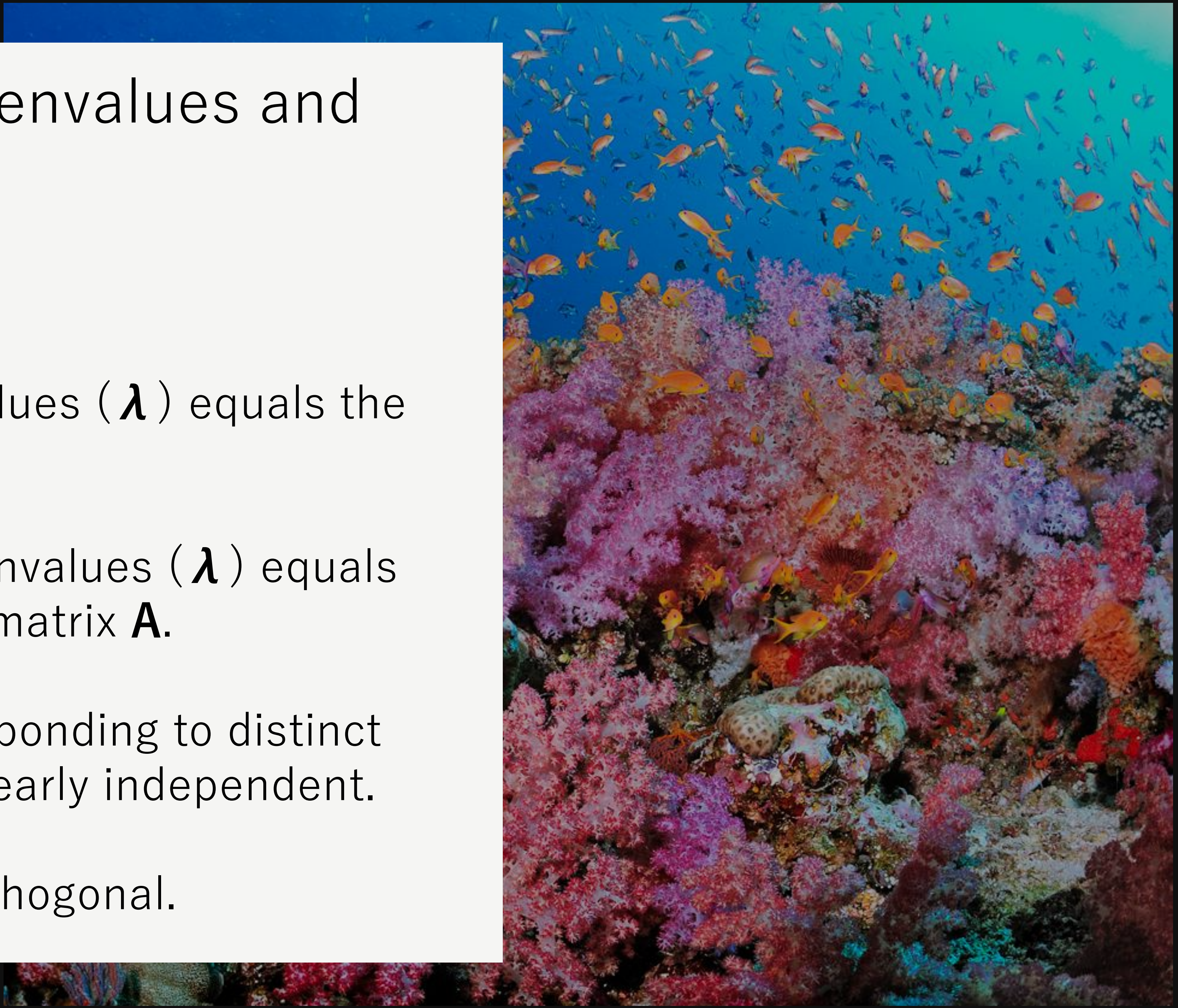
Properties:

1. The sum of the eigenvalues ($\boldsymbol{\lambda}$) equals the trace of the matrix **A**.

2. The product of the eigenvalues ($\boldsymbol{\lambda}$) equals the determinant of the matrix **A**.

3. Eigenvectors ($\mathbf{u}$) corresponding to distinct eigenvalues ($\boldsymbol{\lambda}$) are linearly independent.

4. Eigenvectors ($\mathbf{u}$) are orthogonal.



Association Matrices

Association Matrices: A Very Special Square (Usually) Matrix

An **association matrix (A)** assesses the degree of resemblance among objects (*Q-mode*) or descriptors (*R-mode*) for all element pairs.

Producing an association matrix is the first step in the numerical analysis of ecological data!

Association Matrices: A Very Special Square (Usually) Matrix

An **association matrix (A)** assesses the degree of resemblance among objects (*Q-mode*) or descriptors (*R-mode*) for all element pairs.

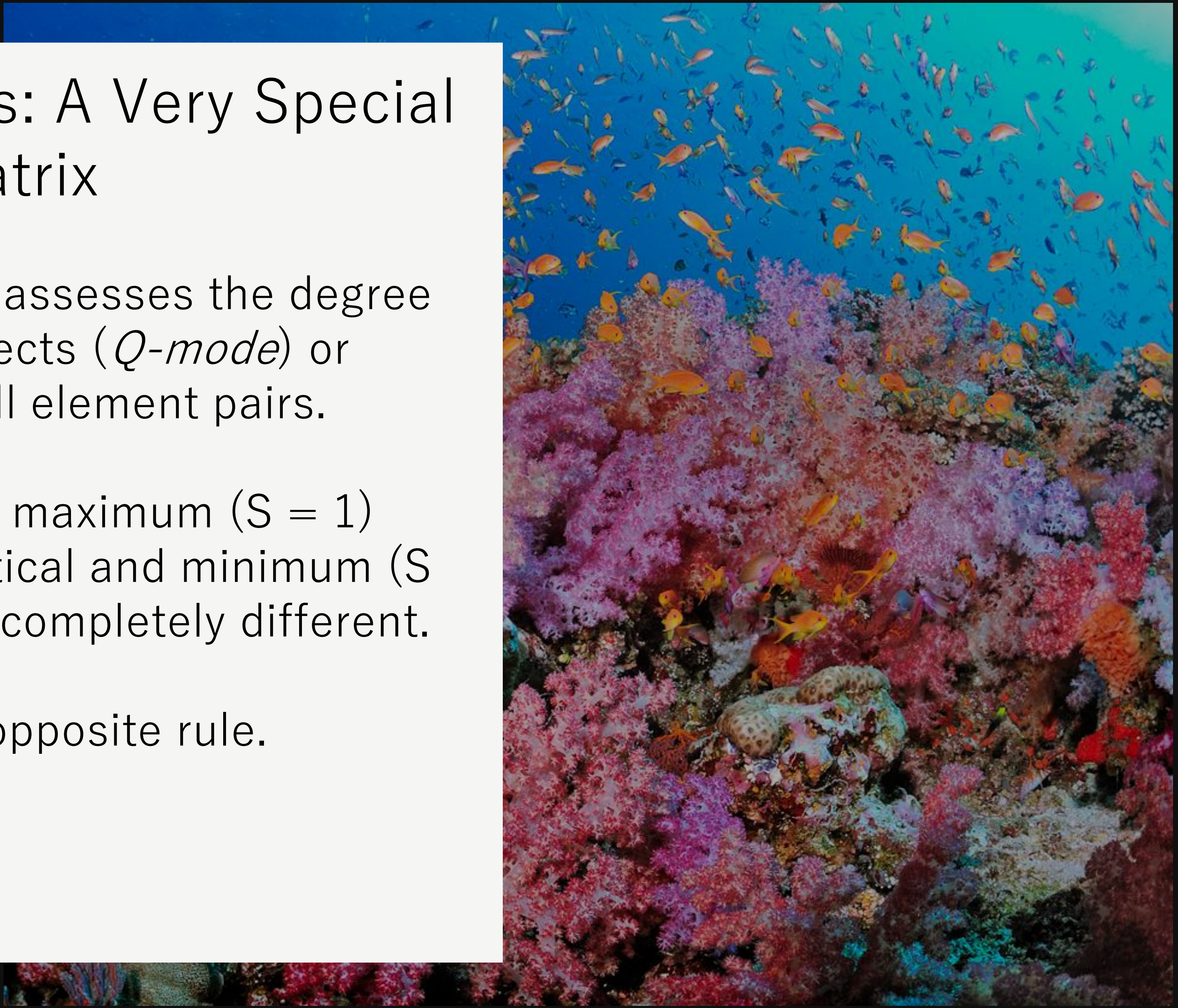
Similarity coefficients are maximum (S = 1) when two objects are identical and minimum (S =0) when two objects are completely different.

Association Matrices: A Very Special Square (Usually) Matrix

An **association matrix (A)** assesses the degree of resemblance among objects (*Q-mode*) or descriptors (*R-mode*) for all element pairs.

Similarity coefficients are maximum (S = 1) when two objects are identical and minimum (S =0) when two objects are completely different.

Dissimilarities follow the opposite rule.



Association Matrices: A Very Special Square (Usually) Matrix

An **association matrix (A)** assesses the degree of resemblance among objects (*Q-mode*) or descriptors (*R-mode*) for all element pairs.

Similarity coefficients are maximum (S = 1) when two objects are identical and minimum (S =0) when two objects are completely different.

Dissimilarities follow the opposite rule.

Distances may not be bound by a pre-determined upper limit, but can be normalized.

Association Matrices: The Double-Zero Problem

- If a species is present at two sites, it is generally an indicator of similarity (favorability, tolerability) between these two sites…

Association Matrices: The Double-Zero Problem

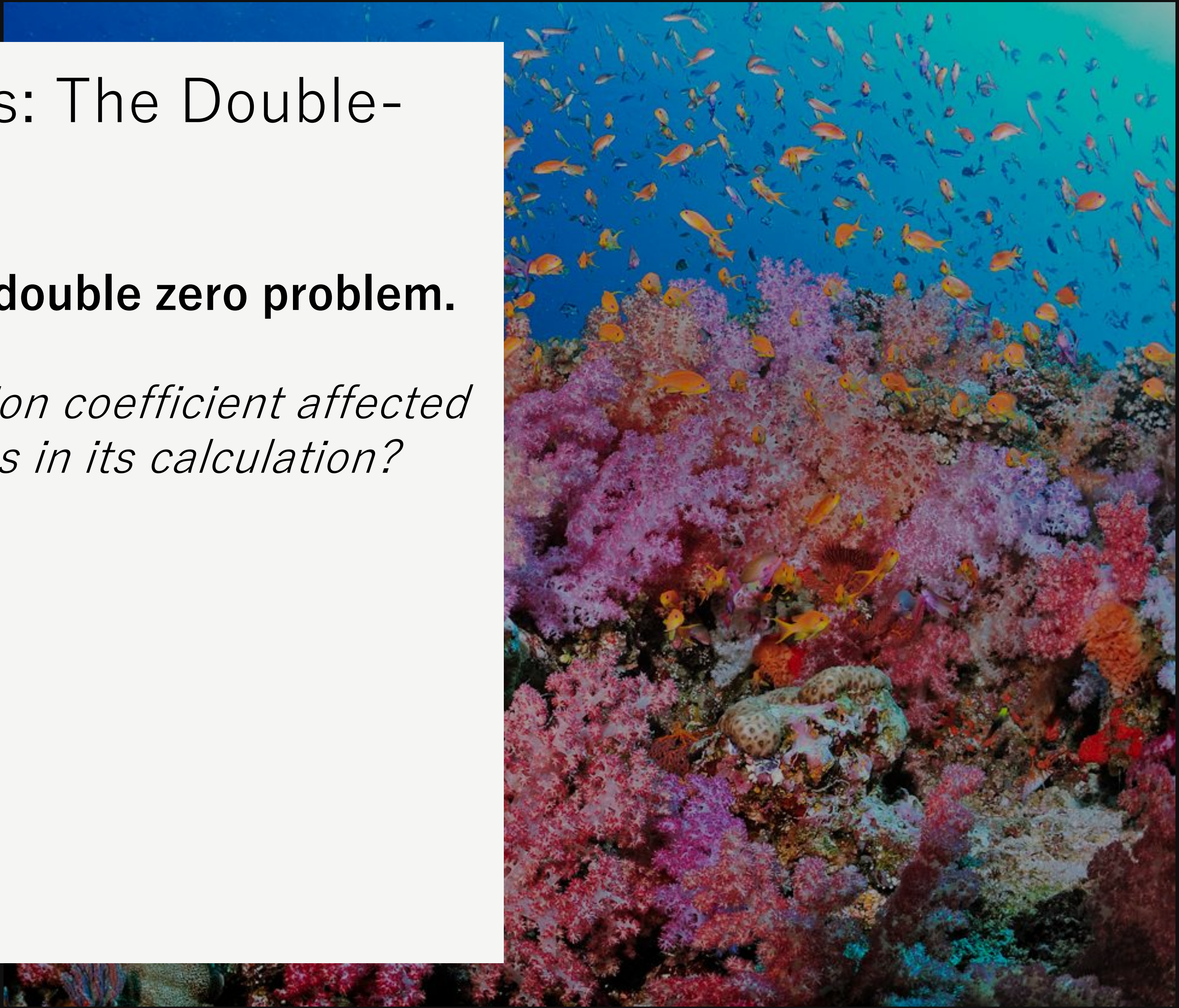
- If a species is present at two sites, it is generally an indicator of similarity (favorability, tolerability) between these two sites…
- ***However***, absences can occur for many reasons, indicate a variety of environmental conditions, and do not necessarily signify environmental similarity

Most scientists do not consider the absence of a species at two sites to provide useful information.

Association Matrices: The Double-Zero Problem

- Here, we encounter the **double zero problem.**

Is the value of an association coefficient affected by inclusion of double zeros in its calculation?



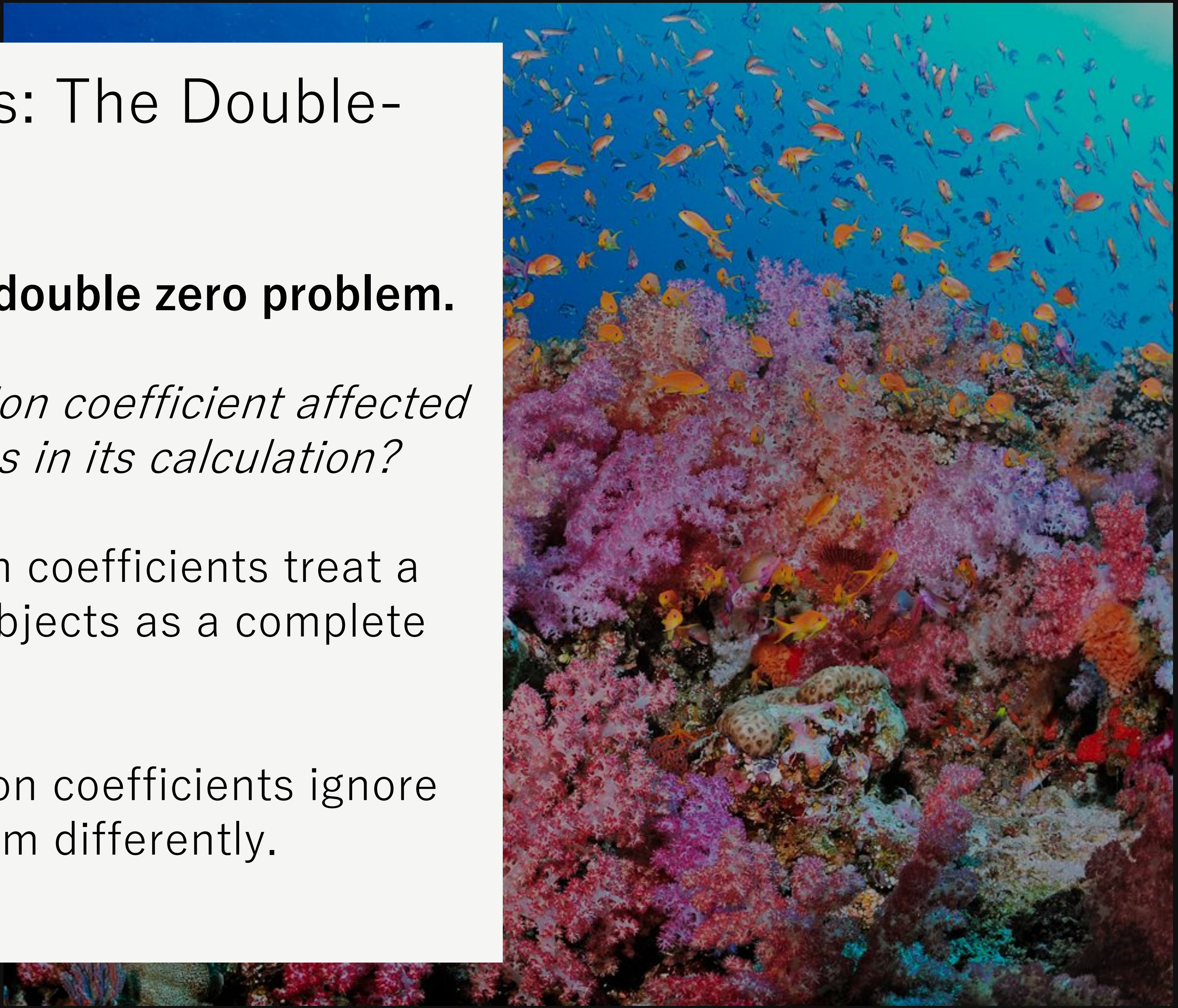
Association Matrices: The Double-Zero Problem

- Here, we encounter the **double zero problem.**

Is the value of an association coefficient affected by inclusion of double zeros in its calculation?

- **Symmetrical** association coefficients treat a zero value for a pair of objects as a complete similarity.

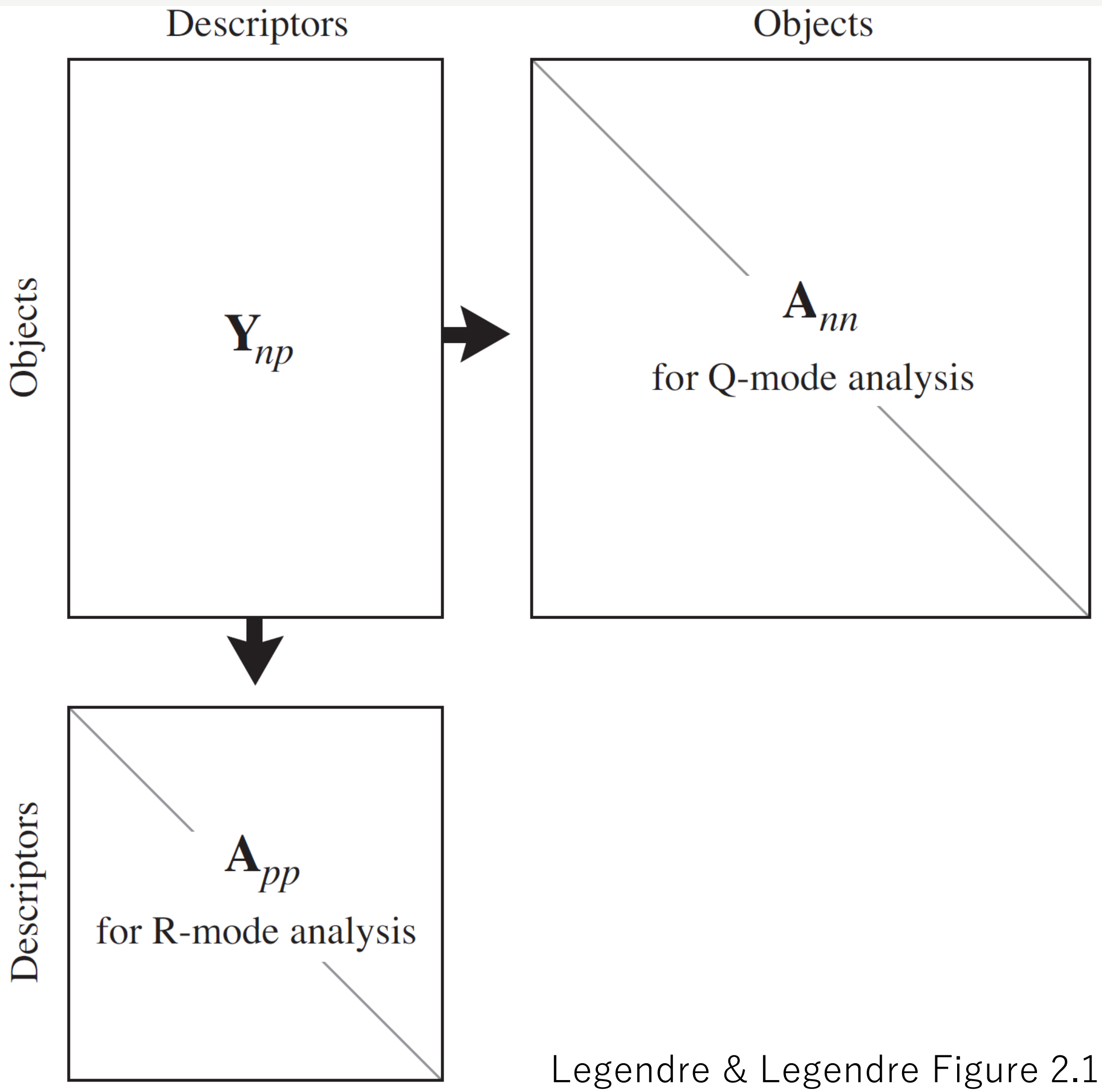
- **Asymmetrical** association coefficients ignore double zeros or treat them differently.



Association Matrices: Q vs. R Mode

An **association matrix (A)** assesses the degree of resemblance among objects (**Q-mode**) or descriptors (**R-mode**) for all element pairs.

Association Matrices: Q vs. R Mode



Legendre & Legendre Figure 2.1

Association Matrices: Relationships Among Objects (Q Mode)

- Binary, presence/absence
- Quantitative, metric
- Quantitative, semimetric

Association Matrices: Simple Matching Coefficient

		Object $\mathbf{x}_2$	
		1	0
Object $\mathbf{x}_1$	1	a	b
	0	c	d

$$S(\mathbf{x}_1, \mathbf{x}_2) = \frac{a + d}{a + b + c + d}$$

Use: Binary or presence/absence data
Association Type: Similarity
Range: 0–1
Symmetrical: Yes
Metric: Yes

Association Matrices: Jaccard's Coefficient

		Object $\mathbf{x}_2$	
		1	0
Object $\mathbf{x}_1$	1	a	b
	0	c	d

$$S(\mathbf{x}_1, \mathbf{x}_2) = \frac{a}{a + b + c}$$

Use: Binary or presence/absence data
Association Type: Similarity
Range: 0–1
Symmetrical: No
Metric: Yes

Association Matrices: Sørensen Coefficient

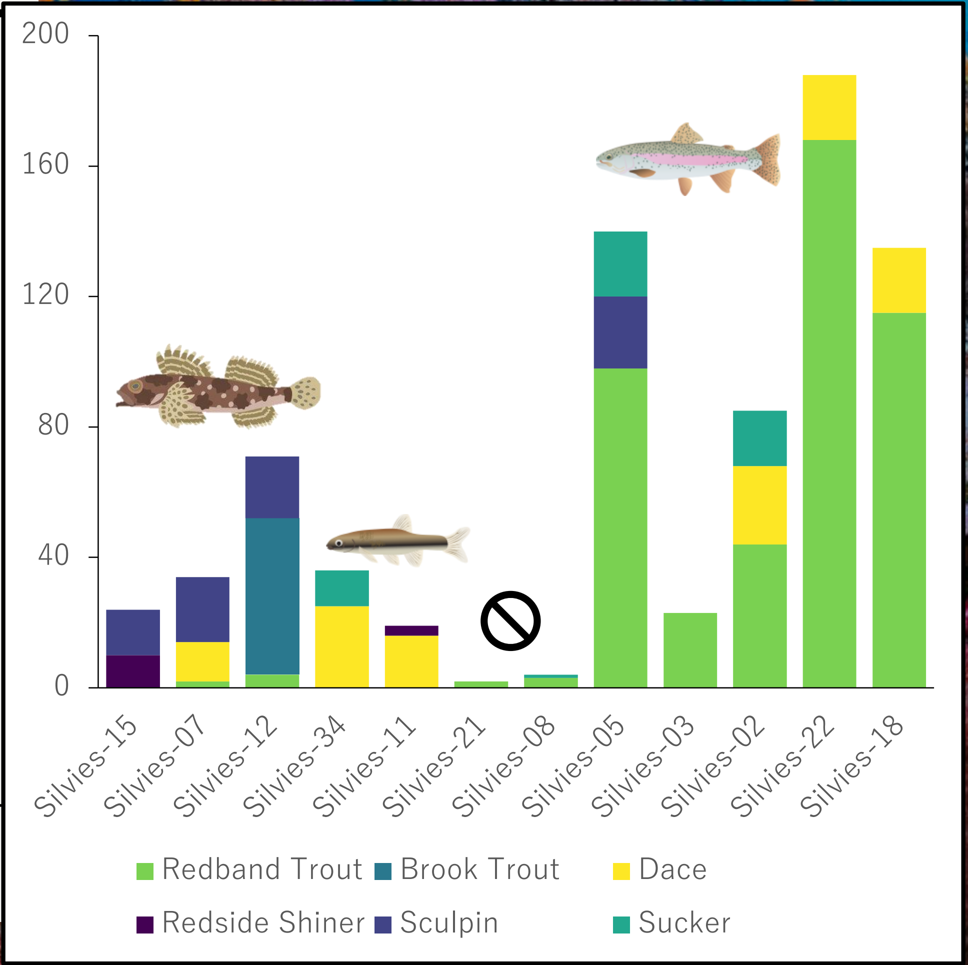
		Object $\mathbf{x}_2$	
		1	0
Object $\mathbf{x}_1$	1	a	b
	0	c	d

$$S\left(\mathbf{x}_1, \mathbf{x}_2\right) = \frac{2a}{2a + b + c}$$

Use: Binary or presence/absence data
Association Type: Similarity
Range: 0–1
Symmetrical: No
Metric: Semimetric

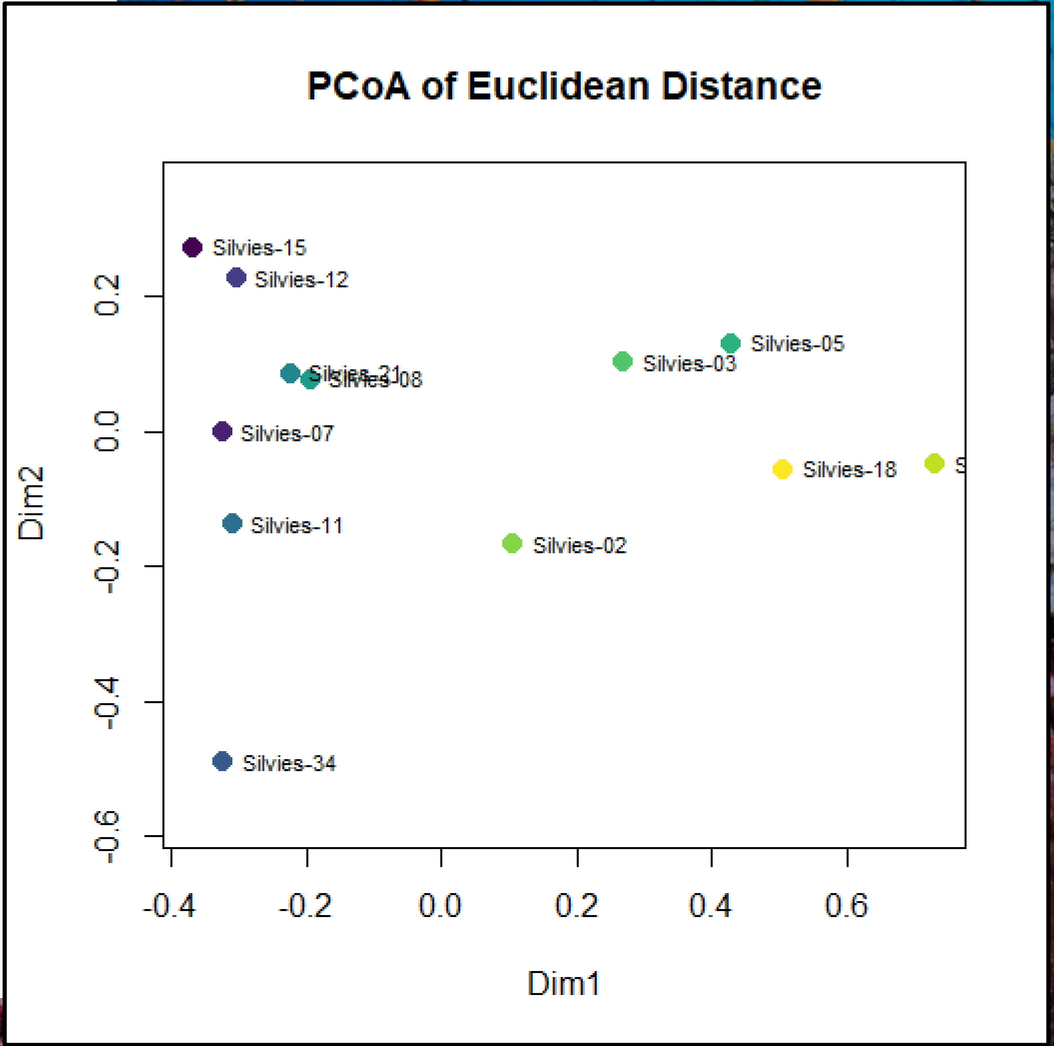
Association Matrices: Example Dataset

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



Association Matrices: Euclidean Distance

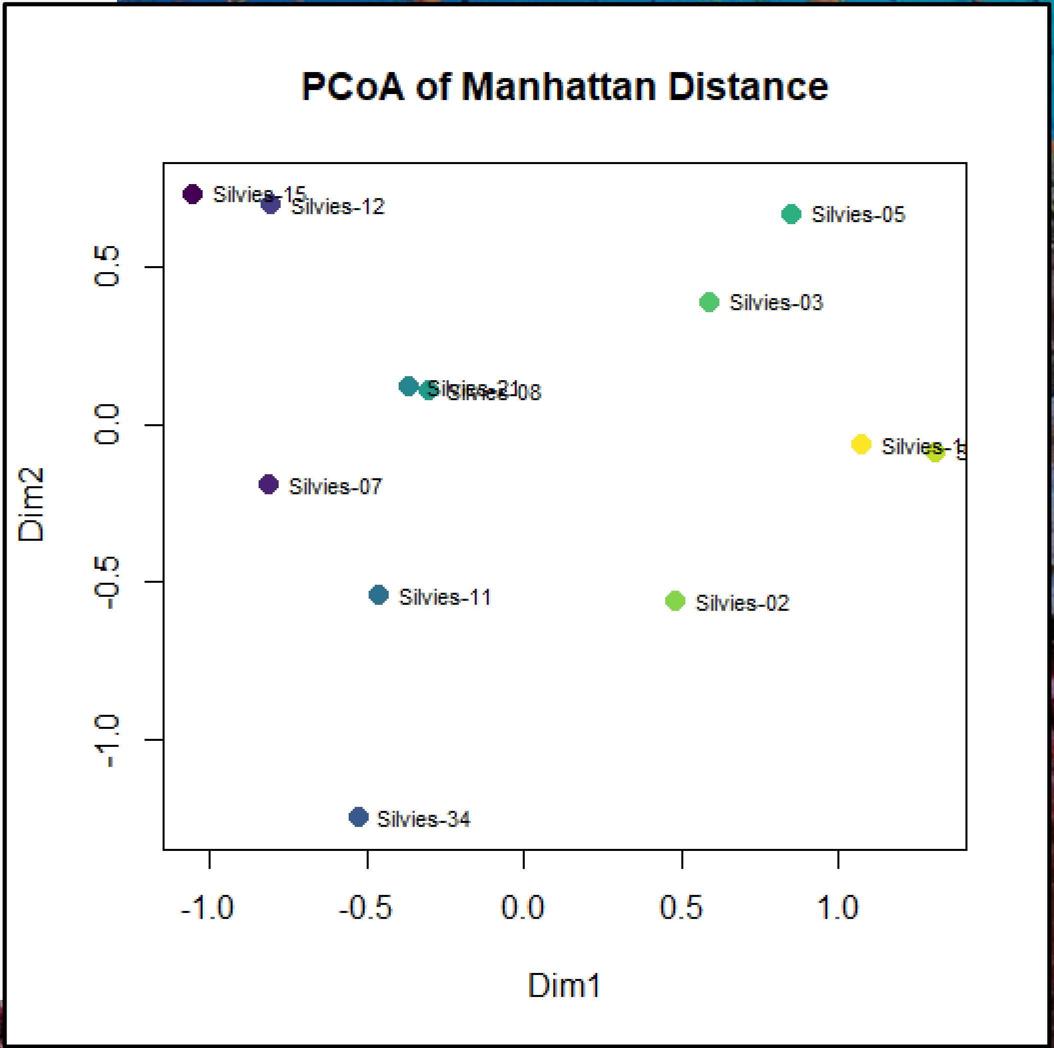
$$D(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^{p} (y_{1j} - y_{2j})^2}$$



Use: Quantitative environmental data (*do not use for species abundance*)
Association Type: Distance
Range: 0–∞ (value depends on descriptor scale)
Symmetrical: Yes
Metric: Yes

Association Matrices: Manhattan Distance

$$D\,(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^{p} |y_{1j} - y_{2j}|$$

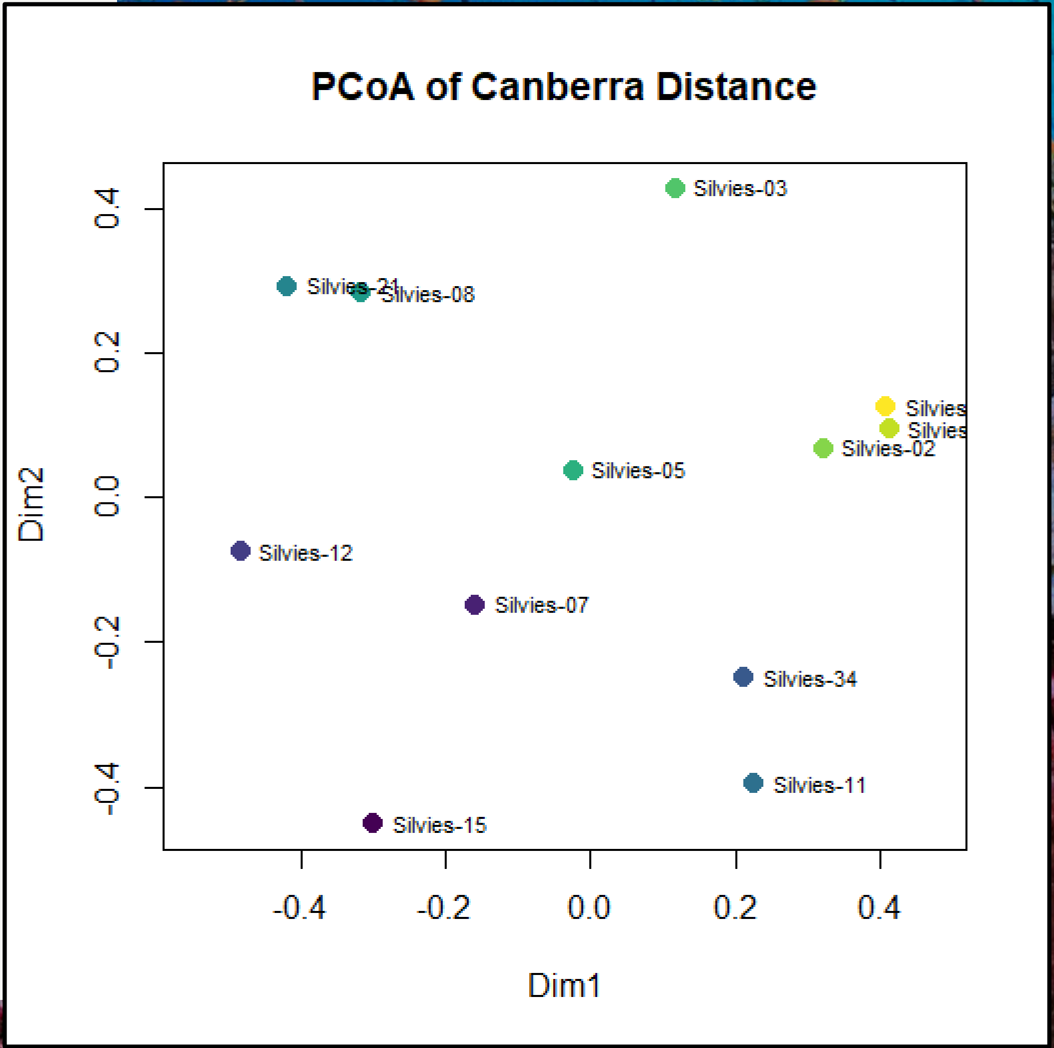


Use: Quantitative environmental data (*do not use for species abundance*)
Association Type: Distance
Range: 0–∞ (value depends on descriptor scale)
Symmetrical: Yes
Metric: Yes

Association Matrices: Canberra Distance

$$D\left(\mathbf{x}_1, \mathbf{x}_2\right) = \sum_{j=1}^{p} \left[\frac{|y_{1j} - y_{2j}|}{(y_{1j} - y_{2j})} \right]$$

Use: Quantitative data; *rarer species contribute more to differences than abundant species*
Association Type: Distance
Range: 0–∞ (value depends on descriptor scale)
Symmetrical: No
Metric: Yes

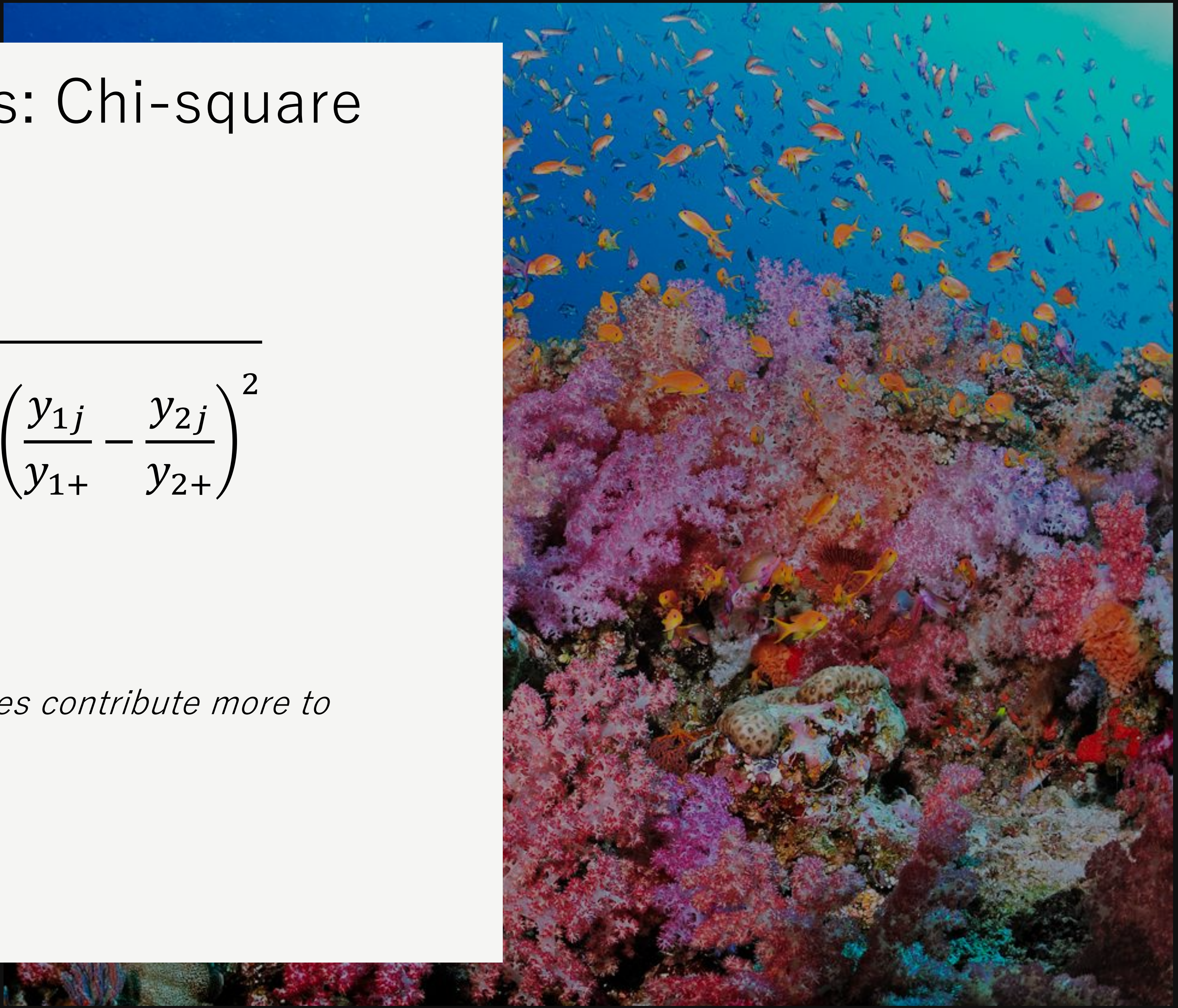


Association Matrices: Chi-square Coefficients

$$D\,(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^{p} \frac{1}{y_{+j}} \left(\frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right)^2}$$

X^2 Metric

Use: Quantitative data; *rarer species contribute more to differences than abundant species*
Association Type: Distance
Range: $0 - \sqrt{2}$
Symmetrical: No
Metric: Yes

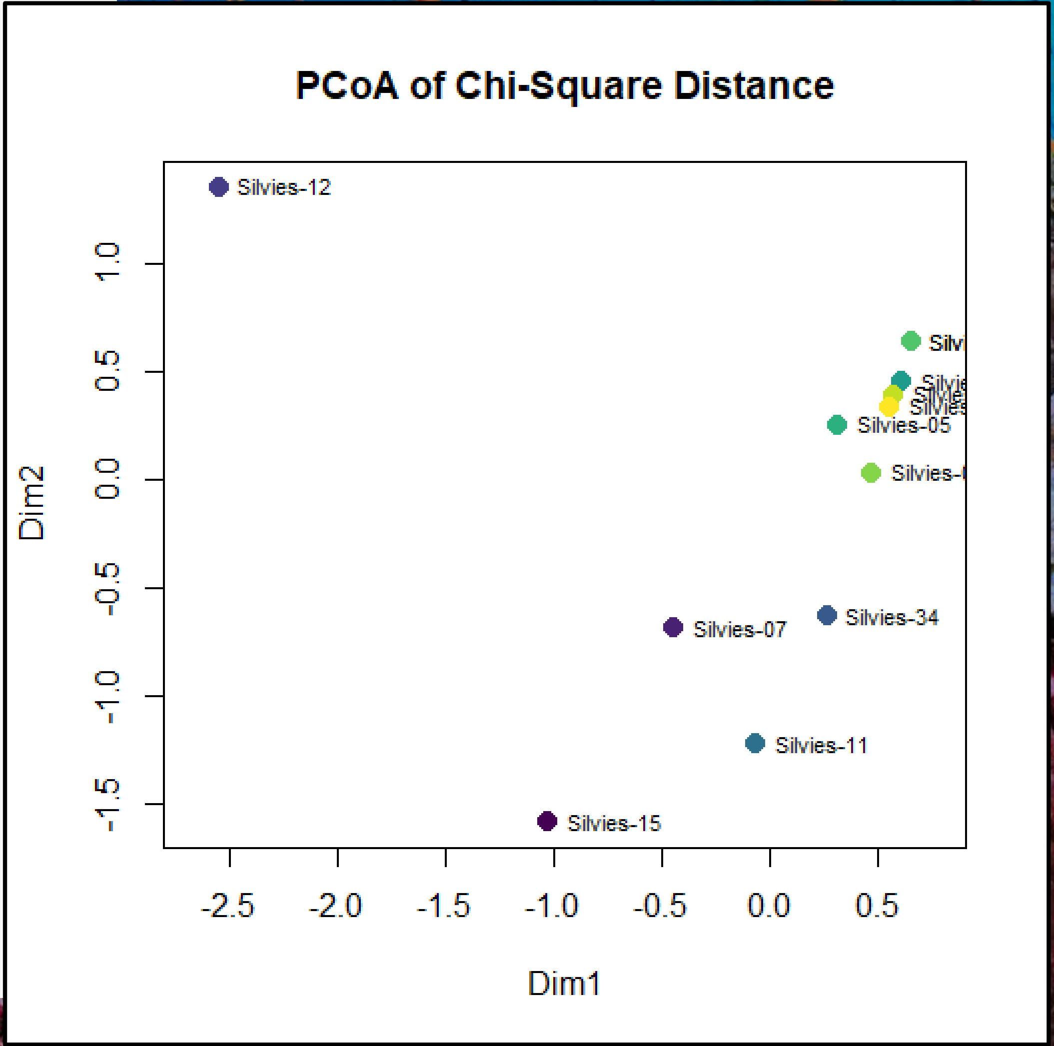


Association Matrices: Chi-square Coefficients

$$D\left(\mathbf{x}_1, \mathbf{x}_2\right) = \sqrt{y_{++}} \sqrt{\sum_{j=1}^{p} \frac{1}{y_{+j}} \left(\frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}}\right)^2}$$

X^2 Distance

Use: Quantitative data; *rarer species contribute more to differences than abundant species*
Association Type: Distance
Range: $0 - \sqrt{2y_{++}}$
Symmetrical: No
Metric: Yes



Association Matrices: Percentage Difference/Bray-Curtis Dissimilarity

$$D\,(\mathbf{x}_1, \mathbf{x}_2) = \frac{\sum_{j=1}^{p} |y_{1j} - y_{2j}|}{\sum_{j=1}^{p} (y_1 + y_{2j})}$$

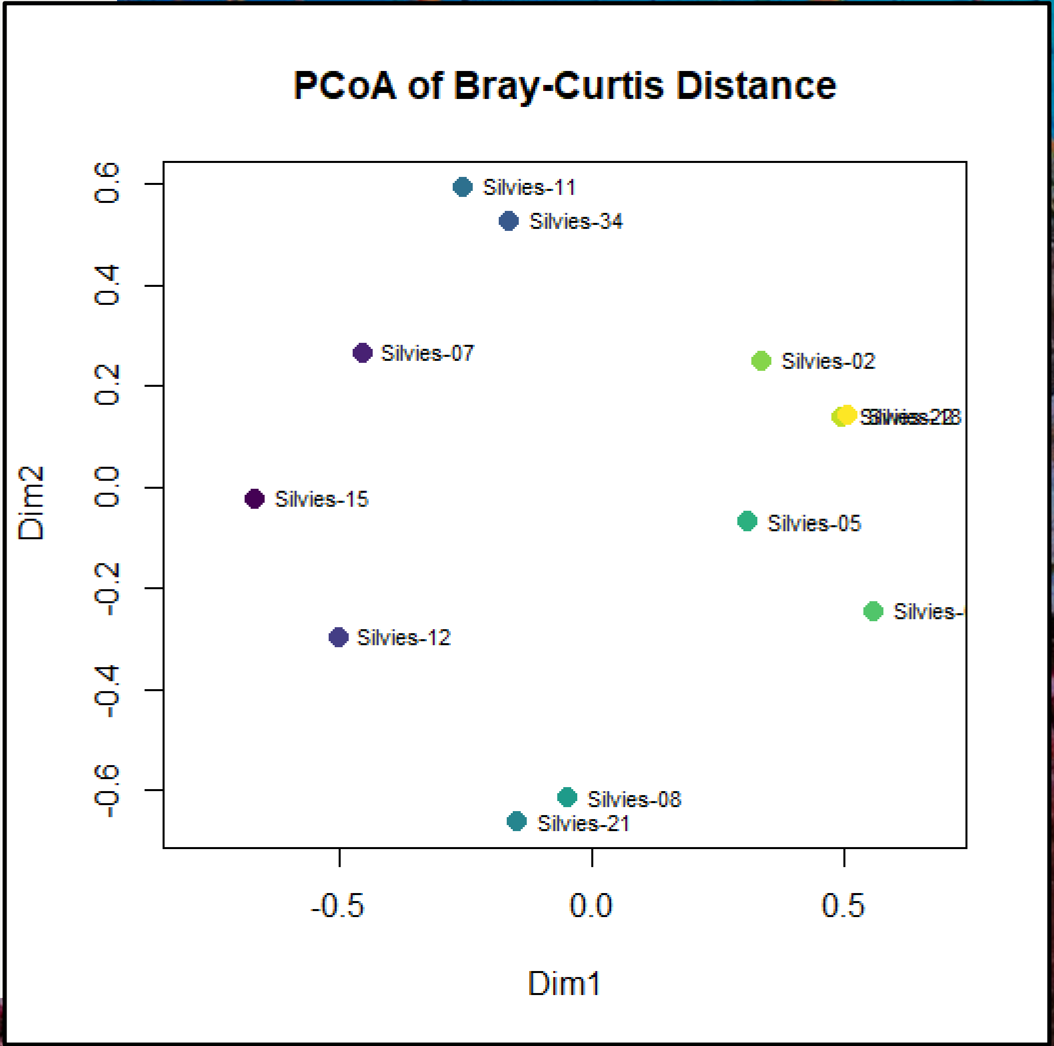
Use: Quantitative data; *particularly suited to species abundance*
Association Type: Distance (Similarity = 1-D)
Range: 0–1
Symmetrical: No
Metric: Semimetric

Association Matrices: Percentage Difference/Bray-Curtis Dissimilarity

$$\mathrm{D}\,(\mathbf{x}_1, \mathbf{x}_2) = \frac{\sum_{j=1}^{p} |y_{1j} - y_{2j}|}{\sum_{j=1}^{p} (y_1 + y_{2j})} = 1 - \frac{2W}{A + B}$$

Where *A* and *B* are the sums of abundances of all species at each of the two sites and *W* is the sum of the minimum abundances of the species.

Use: Quantitative data; *particularly suited to species abundance*
Association Type: Distance (Similarity = 1-D)
Range: 0–1
Symmetrical: No
Metric: Semimetric

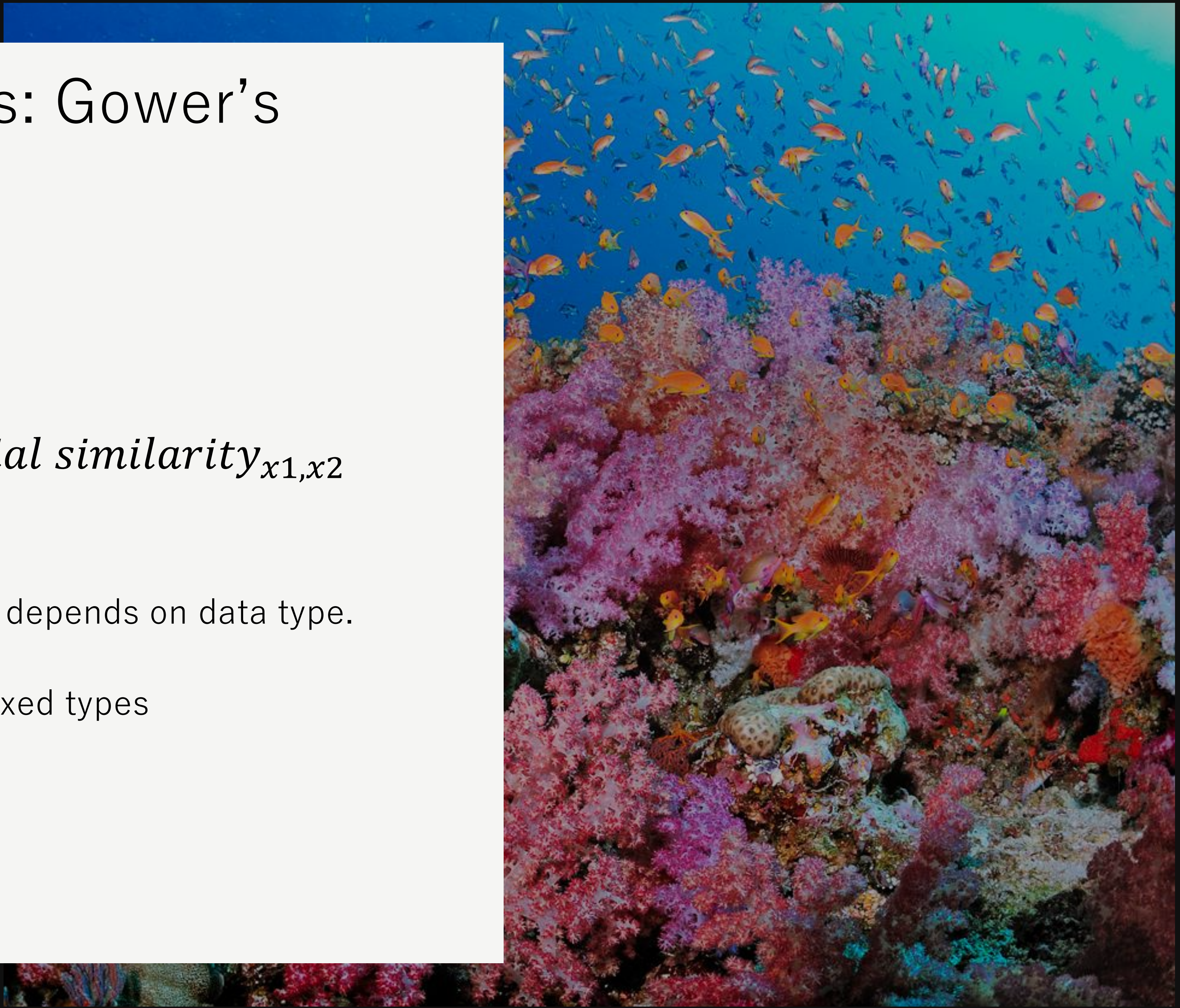


Association Matrices: Gower's Coefficient

$$\mathrm{D}\,(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{p}\sum_{j=1}^{p} partial\ similarity_{x1,x2}$$

Where the "partial similarity" value depends on data type.

Use: Quantitative descriptors of mixed types
Association Type: Similarity
Range: 0–1
Symmetrical: Yes
Metric: Yes



Association Matrices: Relationships Among Descriptors (R Mode)

- Pearson’s correlation coefficient
- Spearman’s *r*
- Kendall’s tau

Association Matrices: Relationships Among Descriptors (R Mode)

- **Pearson's correlation coefficient**
- Spearman's *r*
- Kendall's tau

Pearson's correlation coefficient is a **parametric** measure of dependence.

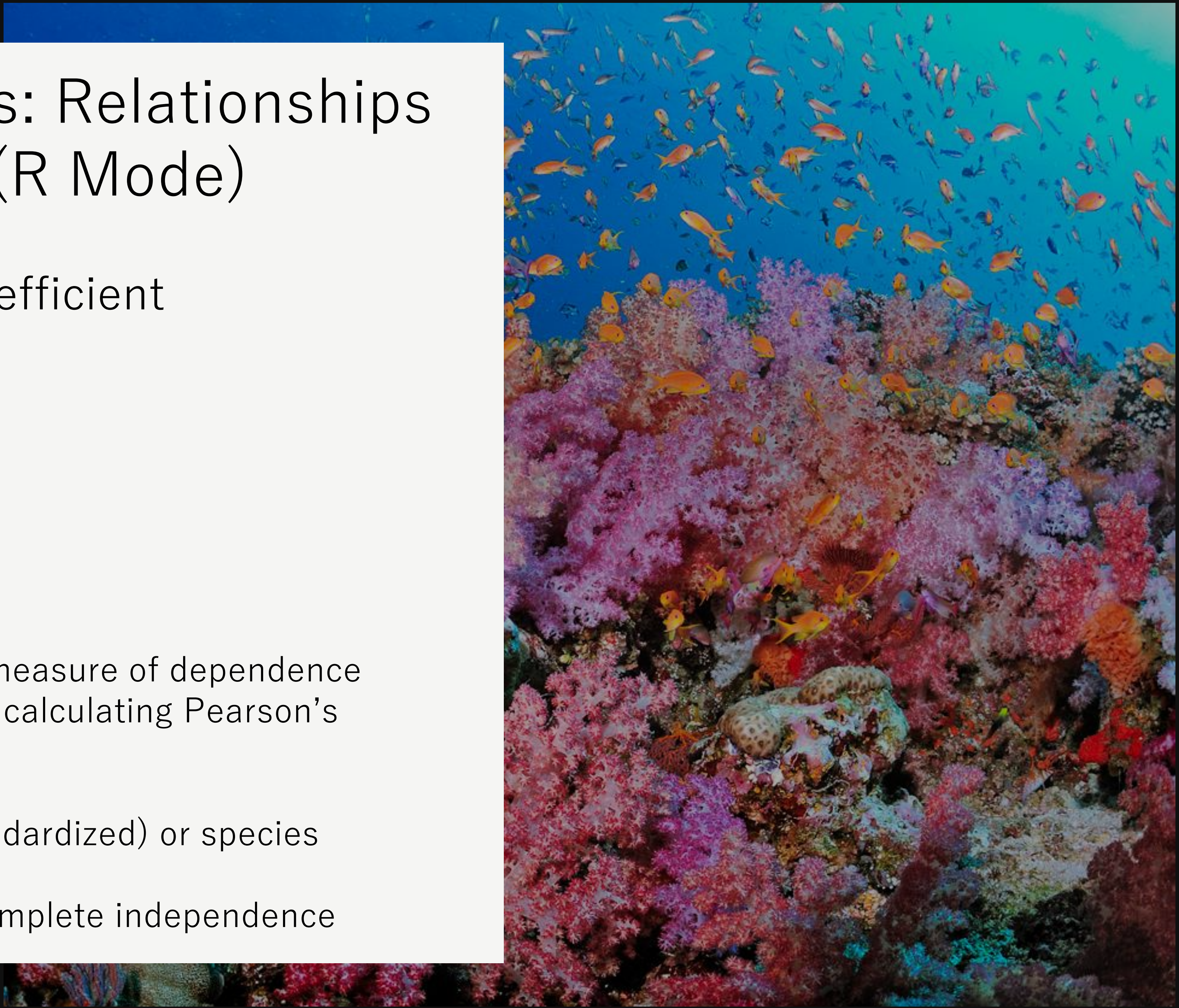
Use: Quantitative descriptors (standardized) or species abundances
Range: -1 – 1 where 0 indicates complete independence

Association Matrices: Relationships Among Descriptors (R Mode)

- Pearson's correlation coefficient
- **Spearman's *r***
- Kendall's tau

Spearman's *r* is a **nonparametric** measure of dependence whereby values are ranked prior to calculating Pearson's correlation coefficient.

Use: Quantitative descriptors (standardized) or species abundances
Range: -1 – 1 where 0 indicates complete independence



Association Matrices: Relationships Among Descriptors (R Mode)

- Pearson's correlation coefficient
- Spearman's *r*
- **Kendall's tau**

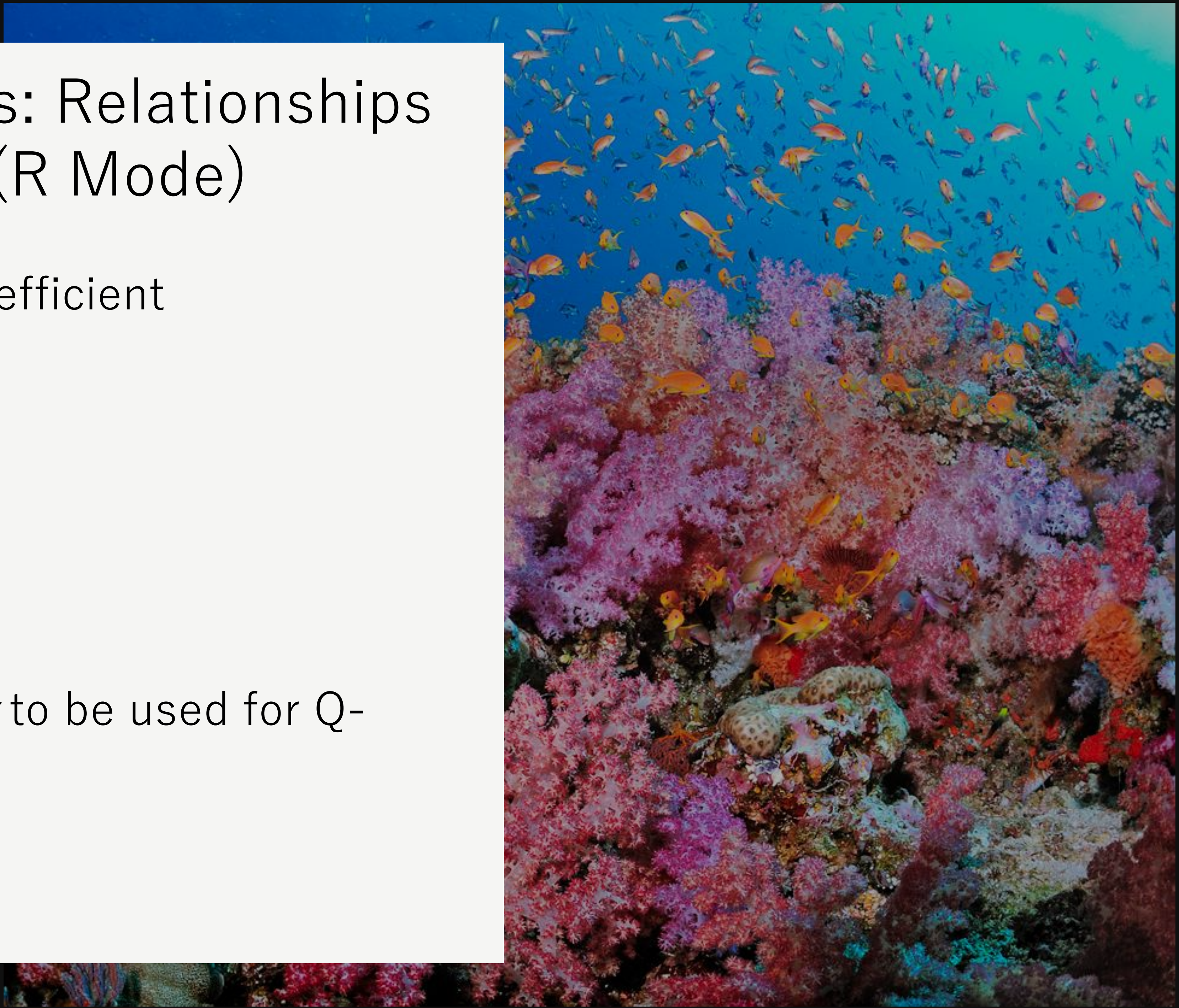
Kendall's tau is a **nonparametric** measure of dependence and is also a "rank" correlation coefficient.

Use: Quantitative descriptors (standardized)
Range: -1 – 1 where 0 indicates complete independence

Association Matrices: Relationships Among Descriptors (R Mode)

- Pearson's correlation coefficient
- Spearman's *r*
- Kendall's tau

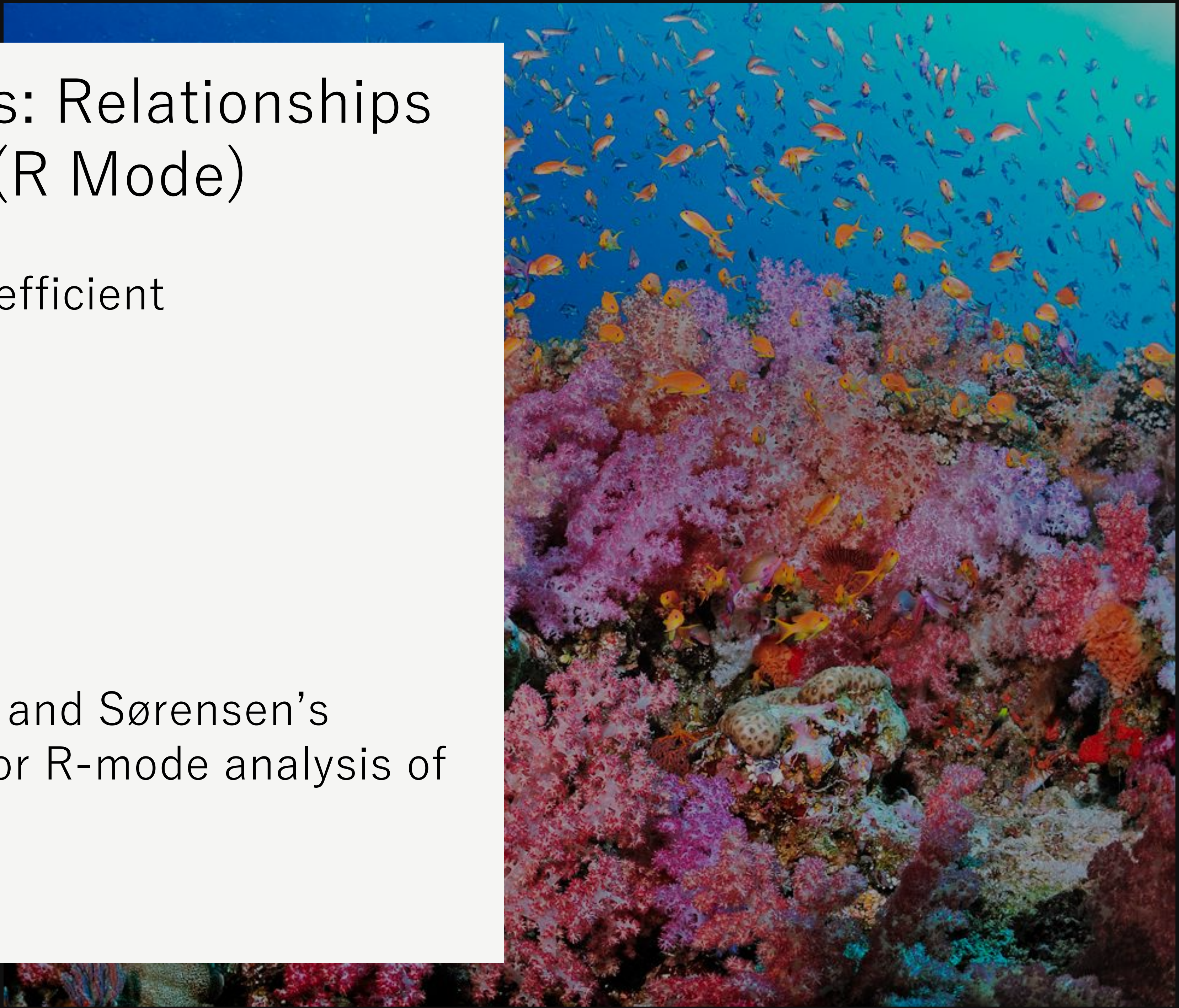
These coefficients are *not* to be used for Q-mode analysis.



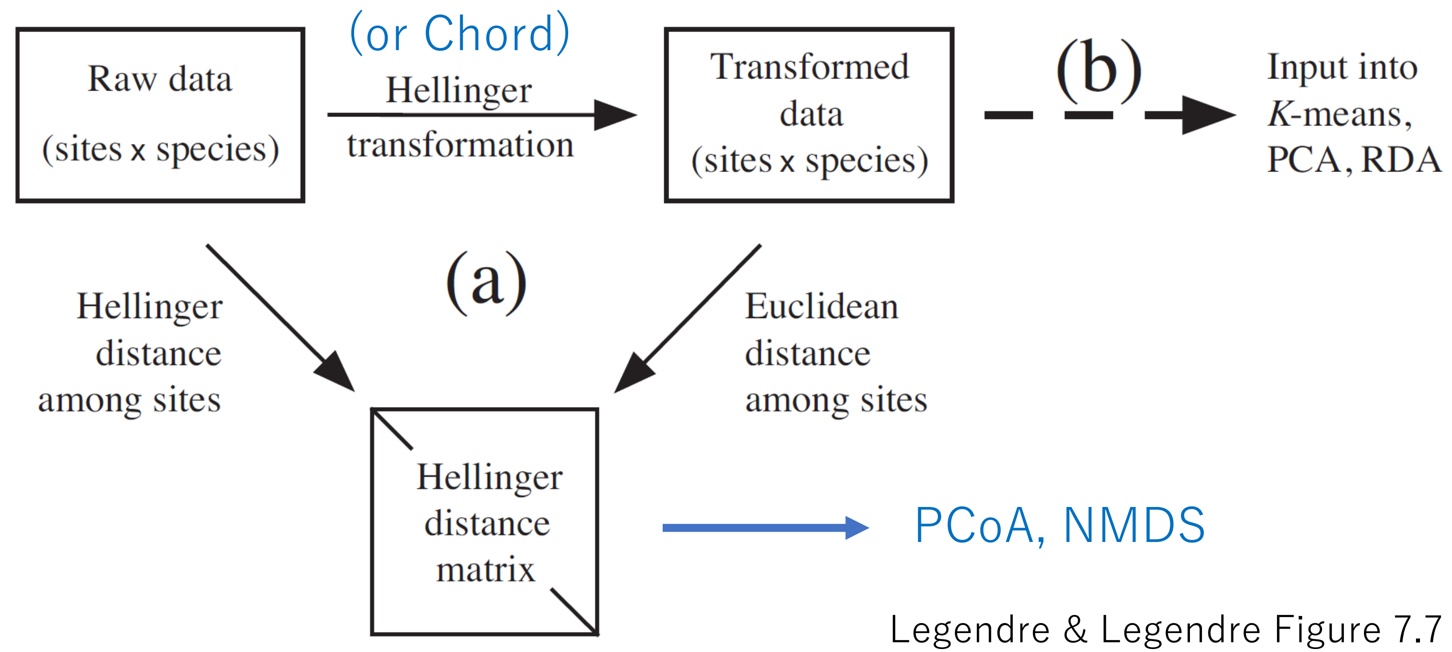
Association Matrices: Relationships Among Descriptors (R Mode)

- Pearson's correlation coefficient
- Spearman's *r*
- Kendall's tau

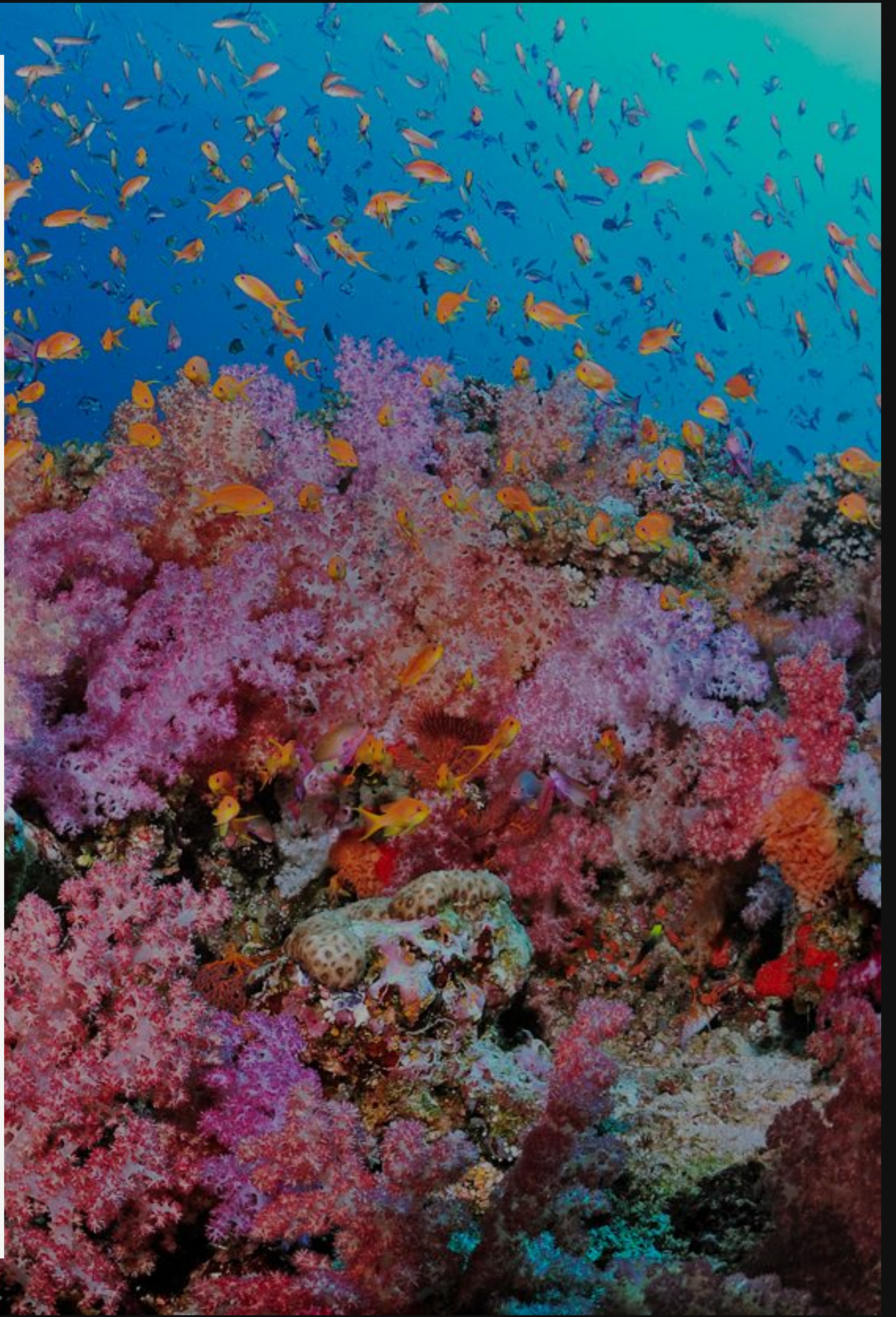
Simple matching, Jaccard, and Sørensen's coefficients can be used for R-mode analysis of presence/absence data.



Association Matrices: Transformations Revisited



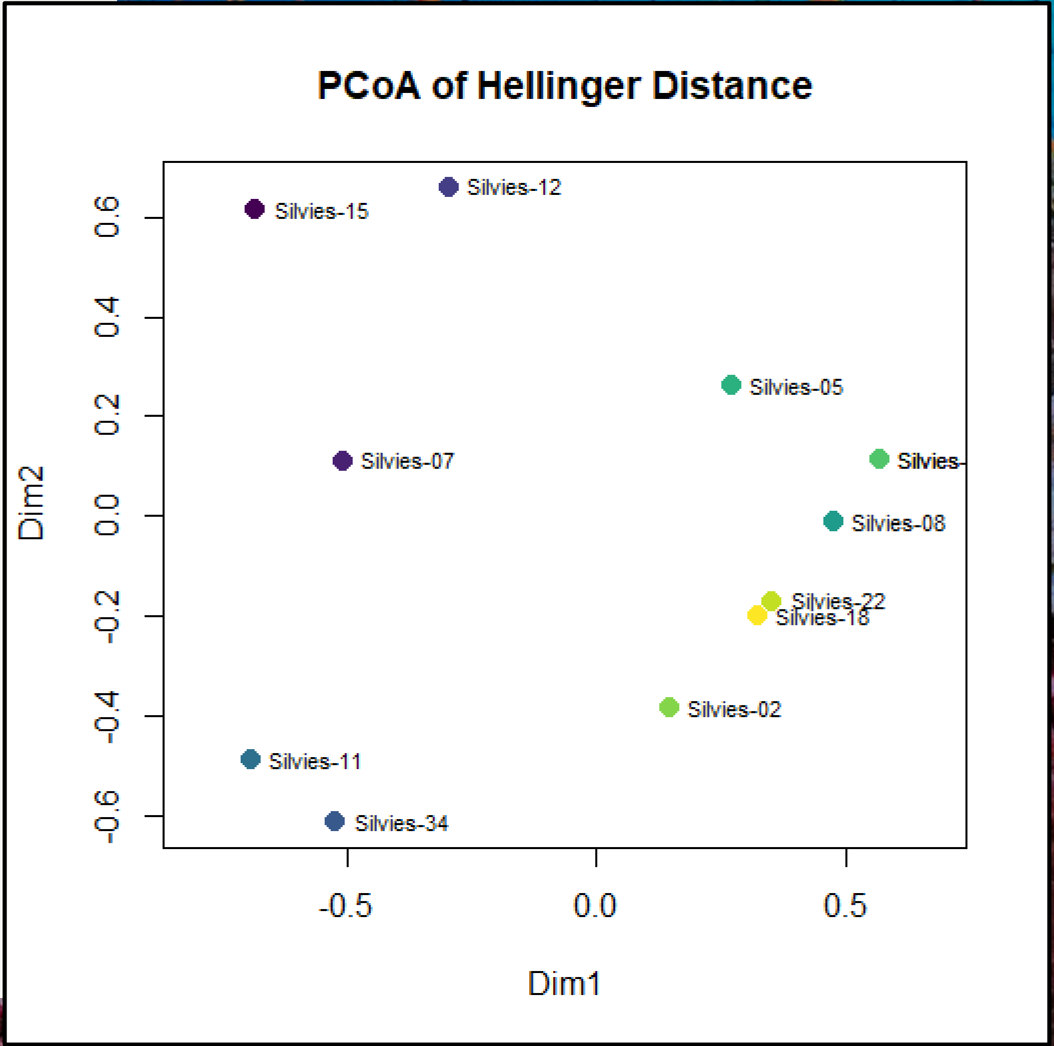
Legendre & Legendre Figure 7.7



Association Matrices: Hellinger Distance

$$D\left(\mathbf{x}_1, \mathbf{x}_2\right) = \sqrt{\sum_{j=1}^{p}\left[\sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}}\right]^2}$$

Use: Quantitative data; *well suited for species abundances*
Association Type: Distance
Range: $0 - \sqrt{2}$
Symmetrical: No
Metric: Yes



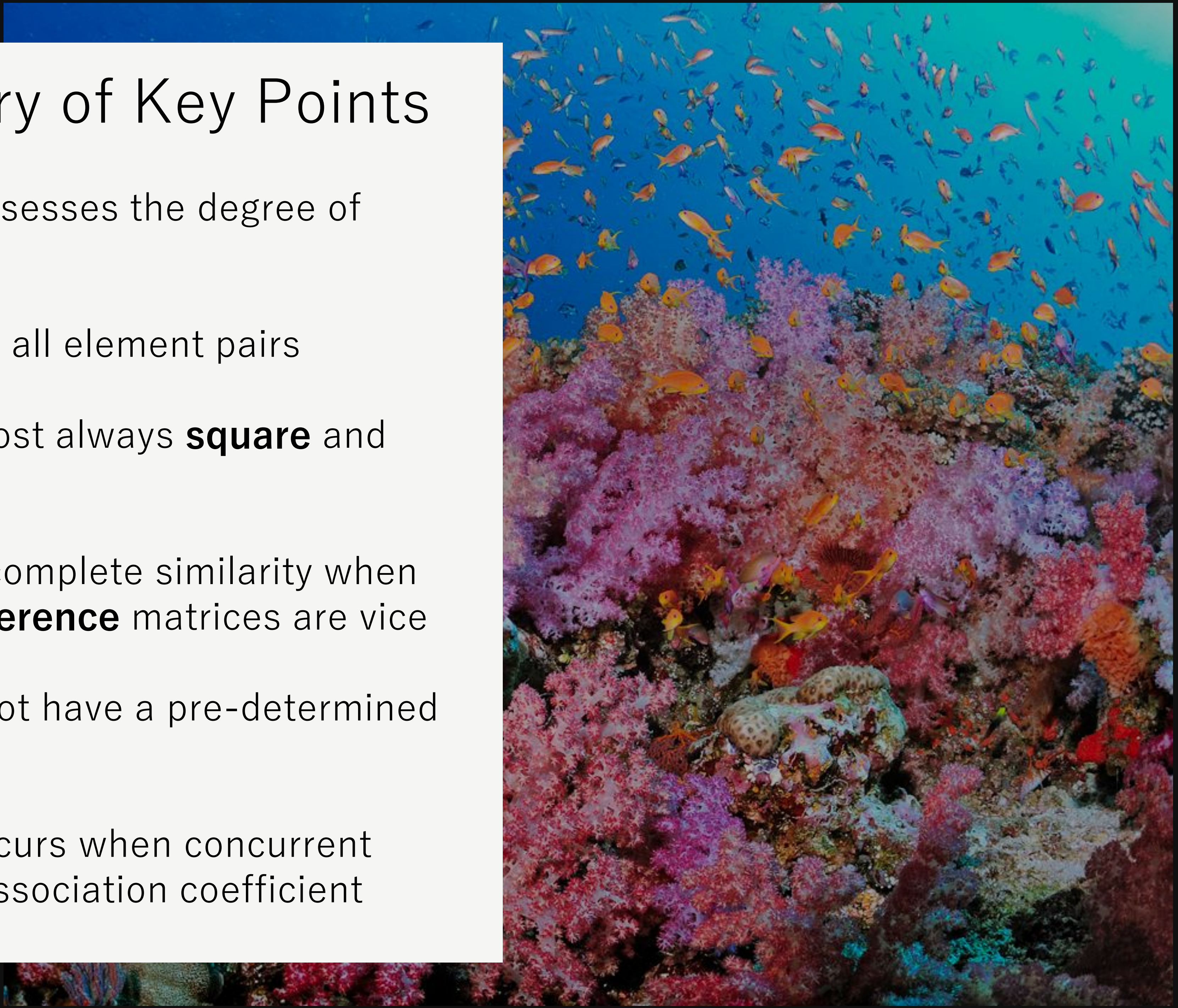
Conclusion: Summary of Key Points

- An **association matrix (A)** assesses the degree of resemblance among:
  - Objects (**Q-mode**) or
  - Descriptors (**R-mode**) for all element pairs

- Association matrices are almost always **square** and **symmetrical**

- **Similarity** matrices indicate complete similarity when the coefficient = 1, while **difference** matrices are vice versa
  - **Distance** matrices may not have a pre-determined upper bound

- The **double zero problem** occurs when concurrent absences count toward the association coefficient



Questions?