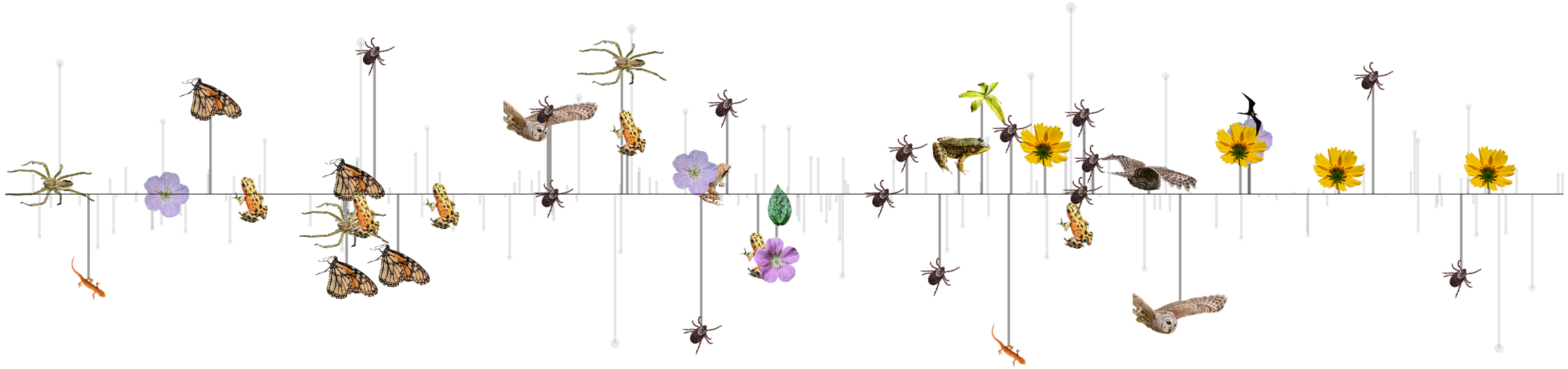
Intro to Quantitative Ecology
UMass Amherst – Michael France Nelson

Deck 2: Quantitative Ecology – The Scientific Process



Announcements

- Time to finish Rmd/Introductions group assignment today

What is Quantitative Ecology?

What is Quantitative Ecology?

Quantitative	Ecology
• "Quantitative methods emphasize objective measurements and the statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques. Quantitative research focuses on gathering numerical data and generalizing it across groups of people or to explain a particular phenomenon." • USC Libraries	From the Ecological Society of America: • "Ecology is the study of the relationships between living organisms, including humans, and their physical environment; it seeks to understand the vital connections between plants and animals and the world around them. Ecology also provides information about the benefits of ecosystems and how we can use Earth's resources in ways that leave the environment healthy for future generations"

What is Quantitative Ecology?

"In the broad sense, the term quantitative ecology applies to any mathematical or numerical treatment of the topic, …"

> Schneider, David C.. Quantitative Ecology. Elsevier Science.

In this course, we'll focus on learning a set of quantitative techniques that we can use to turn ecological data into knowledge.

The Scientific Process

The Scientific Process: Gardener's Stages

Gardener arranges the scientific into four stages:

1. **Planning**
2. Data Recording
3. Data Exploration (and analyses)
4. Reporting Results

The Scientific Process: Gardener's Stages

Key Points

- This is an idealized outline: We really love it when science works this way!
- The process is iterative
- We frequently need to revisit one or more phases

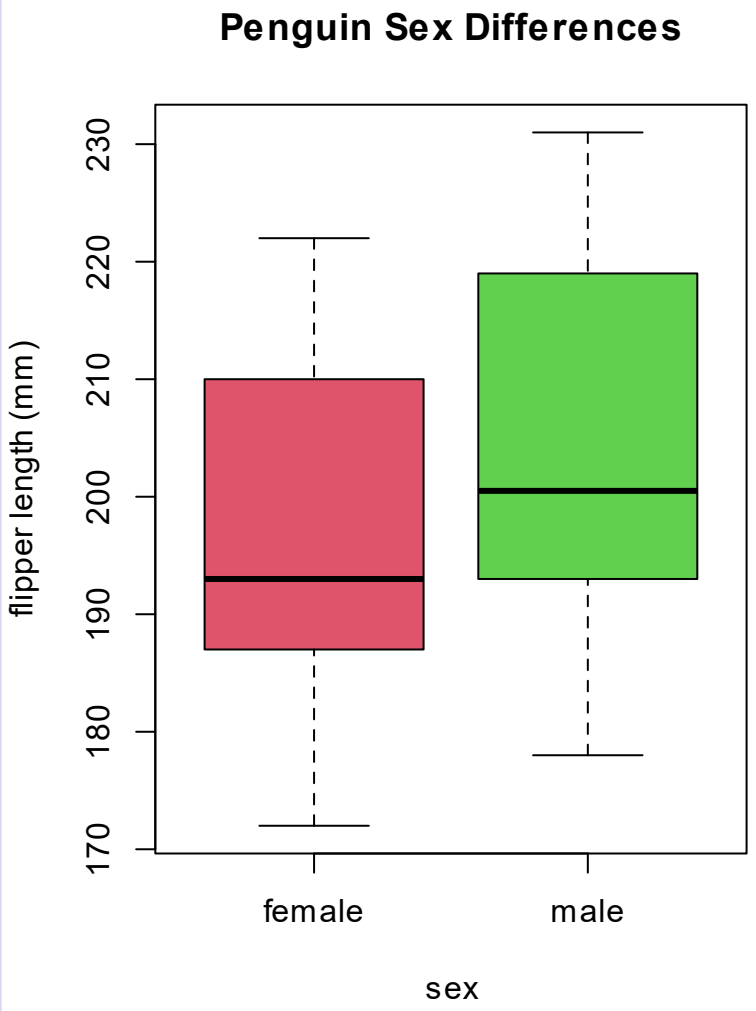
Planning: Defining our Question

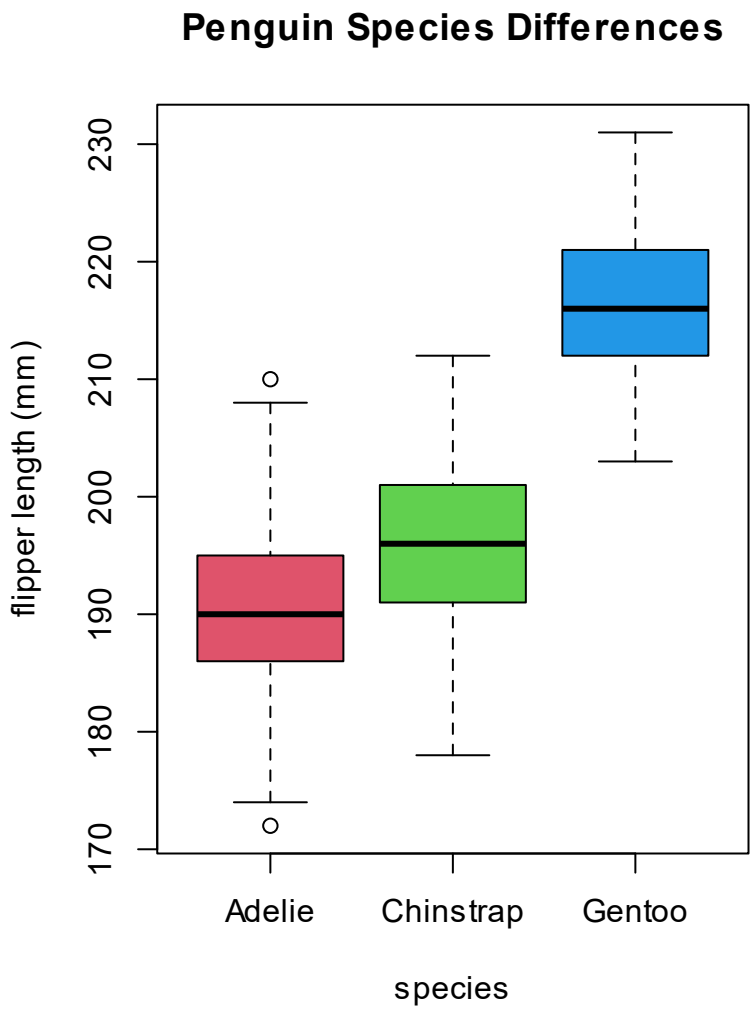
What are we interested in?

- Differences?
  - Are different groups *significantly different* from one-another?
- Associations?
  - Do different combinations categories tend to occur together?
- Correlations?
  - Do different variables tend to vary in a *coordinated way*?
  - NOTE: We may be interested in more than one of these!

Differences: Is A Different From B?

- Are male penguins heavier than females?
- Do Adelie and Chinstrap penguins have different flipper lengths?





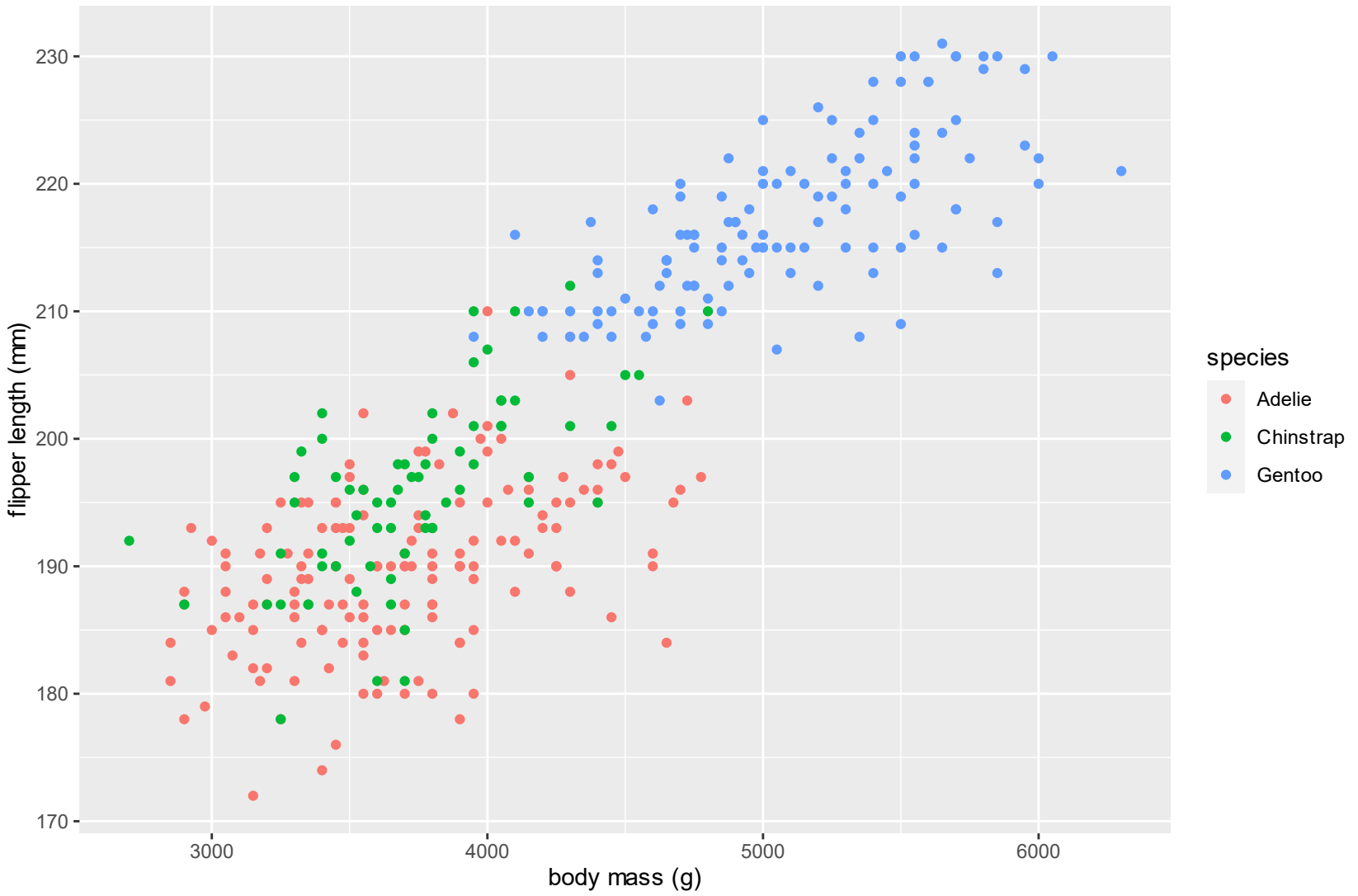
Associations: Do categories co-ocurr?

- Do different penguin species occur on different islands?

	Island		
	Biscoe	Dream	Torgersen
species			
Adelie	44	56	52
Chinstrap	0	68	0
Gentoo	124	0	0

Correlations: Do quantities vary together?

- Do heavier penguins have longer flippers?



Hypotheses and Hypothesis Testing

What is a hypothesis?

Null Hypothesis	Alternative Hypothesis
Null hypotheses: nothing *interesting* is happening, for example: • Male and female penguins do not differ in body mass. • The three penguin species are equally likely to occur on each island. • Penguin body mass is not related to flipper length.	The alternative hypotheses are usually what we think is *actually* happening, for example: • Male and female penguins have different body masses. This is a nondirectional hypothesis. • Male penguins are *heavier* than female penguins. A directional hypothesis. • Heavier penguins tend to have longer flippers. A directional hypothesis.

Hypotheses and Falsifiability

We usually don't think the null hypothesis is true!

- But… we need rigorous criteria to quantify our evidence against a null hypothesis.
- For a proposition to be a *scientific hypothesis* we must be able to define what evidence would support or refute the hypothesis.

Null and alternative hypotheses come in pairs.

- The alternative hypothesis is usually our research motivation.
- An example:
  - **Null:** All penguins species have the same size bills.
  - **Alternative:** Different penguin species have different bill sizes.
- Another example:
  - Penguins are not sexually dimorphic for body mass.
  - Male penguins are heaver than females.

R’s Basic Data Structures

Vectors, matrices, and data frames

R's Basic Data Types

Numeric Types	Non-Numeric Types

• Integers
  • Whole numbers like 1, 2, 457
• Decimals
  • 1.1, 3.14, 2.00001
• Missing values
  • NA, NULL, NaN
	• Text
	  • Letters or symbols like "this", "gr8"
	  • Numbers represented as text: "eight", "one"
	• Boolean (logical)
	  • True/False values
	• Missing values
	  • NA, NULL, NaN

Vectors and Matrices

- Vectors are 1-dimensional collections of elements
  - Indexed using square brackets with one number
  - Each element in a vector must be of the same type.
  - Use the c() function to create a vector.
- Matrices are 2-dimensional collections of elements
  - Rows and columns, indexed using square brackets with 2 numbers:
    - [row, col]
  - All elements in the matrix must be **same type.**
  - Use matrix(), cbind(), and rbind() to build a matrix

Creating and Subsetting Vectors and Matrices

Vectors

```
# Create a numeric vector
vec1 = c(1, 2, 37)

# Logical test: which element equals 2
vec1 == 2

# Create a logical vector
vec2 = c(TRUE, FALSE, TRUE, TRUE)

# Retrieve element 4 from vec2
vec2[4]
```

Matrices

```
# Create a matrix of zeroes
mat1 = matrix(0, nrow = 3, ncol = 4)

# Create a matrix from vec1
mat2 = matrix(vec1, nrow = 3, ncol = 5)

# Retrieve element in row 2, column 3
mat2[2, 3]

# Retrieve all elements in row 3
mat2[3, ]
```

Data Frame Properties

- 2-dimensional data structure
  - Rows and columns
  - Indexed using square brackets with two numbers
- Follows the row-data paradigm (more on this in week 3)
  - Rows are observations
  - Columns are attributes (or variables)
- Unlike matrices, each column can be a different data type:
  - Numeric, integer, factor, character, logical (a.k.a Boolean)
- Each column has a name – you can extract a column by its name.

Data Frame Operations: Creation

Syntax	Example

• Use the data.frame() function to build from scratch.
• Use the c() function to create the columns.
• You can name your columns.
	``# Create a data.frame
	df1 = data.frame(
	  col_1 = c(1, 2, 5),
	  col_2 = c(TRUE, TRUE, FALSE),
	  col_3 = c("Mike", "Meg", "Chris")
	)```

Data Frame Operations: Subset by Name

Syntax	Example

• Use the dollar sign and the name of the column.	`> df1$col_2`
	`[1]  TRUE  TRUE FALSE`
	`> df1$col_1`
	`[1] 1 2 5`
	`> df1$col_3`
	`[1] "Mike"  "Meg"  "Chris"`
	`>`

Data Frame Operations: Subset by Position

Syntax	Example

• Use square brackets with the row and column indices.
• Leave either the row or column index blank to get entire rows or columns
	`> # get element in row 1, col 2`
	`> df1[1, 2]`
	`[1] TRUE`
	`> # get all elements in row 2`
	`> df1[2, ]`
	`  col_1 col_2 col_3`
	`2     2  TRUE   Meg`
	`>`

Data Frame Operations: Logical Subset

Syntax	Example

• Use the subset() function with a logical expression
• Double equals sign == checks for equality
• != checks for 'not equal'
	`> subset(df1, col_3 == "Mike")`
	`  col_1 col_2 col_3`
	`1     1  TRUE  Mike`
	`> subset(df1, col_3 != "Mike")`
	`  col_1 col_2 col_3`
	`2     2  TRUE   Meg`
	`3     5 FALSE Chris`
	`>`

Before Thursday's Class

- Read the Desert Shrubs assignment
- Come to class with your R questions!

Notes on RMD Syntax

- When you want a line break in the rendered document, you must include a blank line in the RMD:

```
13 ## Demo lines
14
15 Line 1
16 Line 2
17 Line 3
18
```

Demo lines

Line 1 Line 2 Line 3

```
13 ## Demo lines
14
15 Line 1
16
17 Line 2
18
19 Line 3
20
```

Demo lines

Line 1

Line 2

Line 3

Notes on RMD Syntax

- You must include a space between the number sign and the header text:

```
13 ##Demo lines
14
15 Line 1
16
17 Line 2
18
19 Line 3
20
```

##Demo lines

Line 1

Line 2

Line 3

```
13 ## Demo lines
14
15 Line 1
16
17 Line 2
18
19 Line 3
20
```

Demo lines

Line 1

Line 2

Line 3

In-class Assignment

Vectors and Data Frames Practice

In-class Assignment: Instructions

Form groups of 4-5 and complete the group self-selection.

Choose a scribe who will submit the group's document.

Finish last Thursday's assignment (if needed)

Locate the assignment (link is on GitHub)

Work through the questions. Be sure to discuss your answers as a group.

Announcements

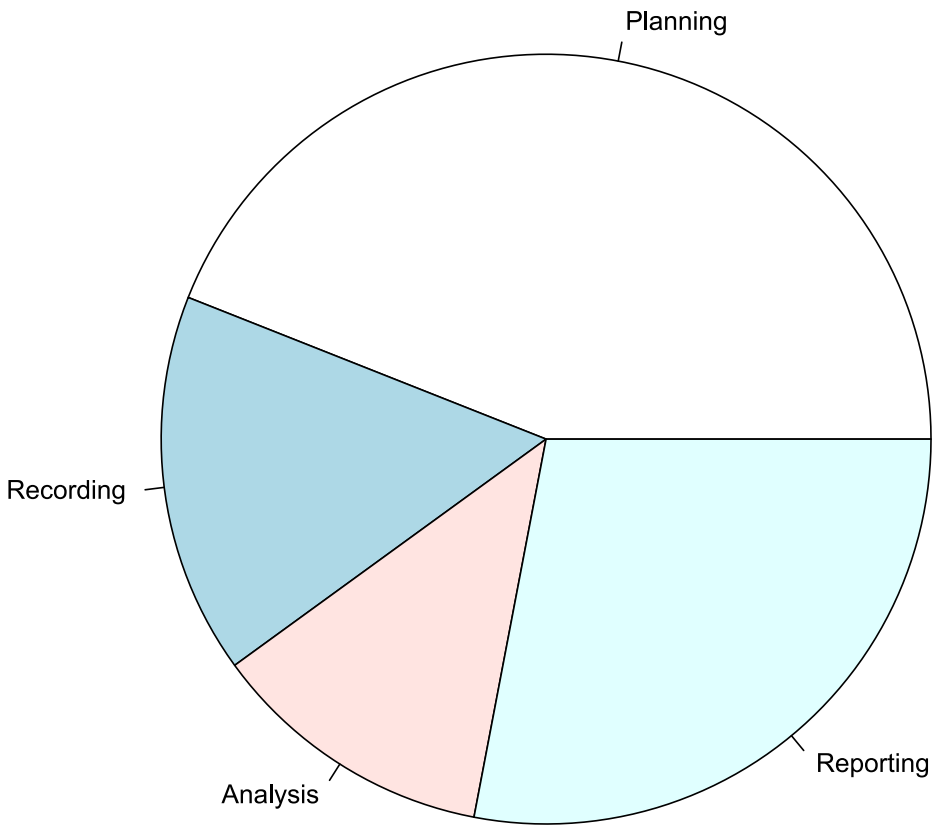
- Office Hours: Attend at least 1 session this semester.
  - Worth 5% of your grade!
  - If the scheduled office hours don't work for you, we'll make an appointment that does!
- DataCamp assignment
  - Don't submit the certificate
  - Do your work in an R script file (this is what you'll submit)

Skills: Barraquand et al. reading

- R programming
- Building and interpreting models
- Visualizing data
- General math/statistical knowledge
- General computer skills
- Data organization/management
- Hypothesis testing
- Sampling

Important Stages

- A plurality of you chose the planning stage!
- Reporting was the second most popular.



Bias and Skew

These definitions are different than in everyday language

- Bias: unrepresentative sampling.
- Skew: a nonsymmetrical distribution (we'll cover what this means)

Wong-Baker


Wong-Baker FACES® Pain Rating Scale
0
No Hurt
2
Hurts Little Bit
4
Hurts Little More
6
Hurts Even More
8
Hurts Whole Lot
10
Hurts Worst

Data Types and Scales

This topic seems simple at first, but there is a lot of hidden subtlety and complexity.

Data Type Examples	Data Scale Examples

• Numeric: integer, decimal
• Logical (True/False)
• Categorical
  • Categorical data may, or may not, have an intrinsic ordering.
	• Ratio and interval (numeric type)
	• Ordinal and nominal (categorical type)
	  • Ordinal: categories have an order
	  • Nominal: no sensible way to order
	• Periodic (numeric type)
	  • “Wraps around”
	• Wong-Baker (categorical type)
	  • Is it ordinal or nominal?

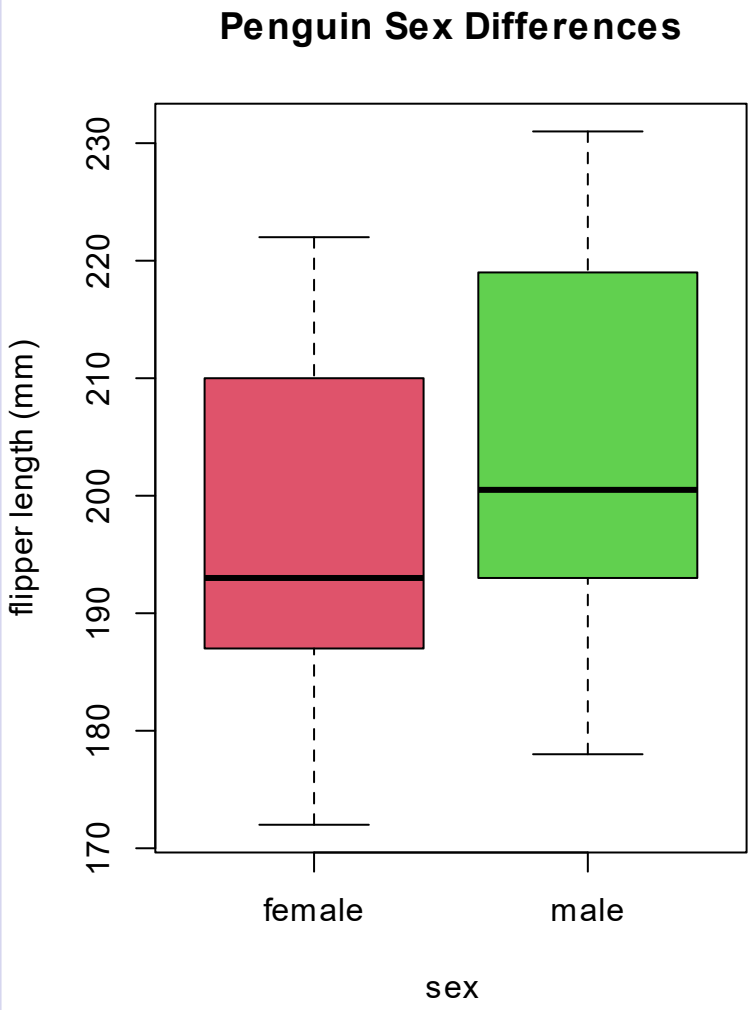
Data Types and Scales

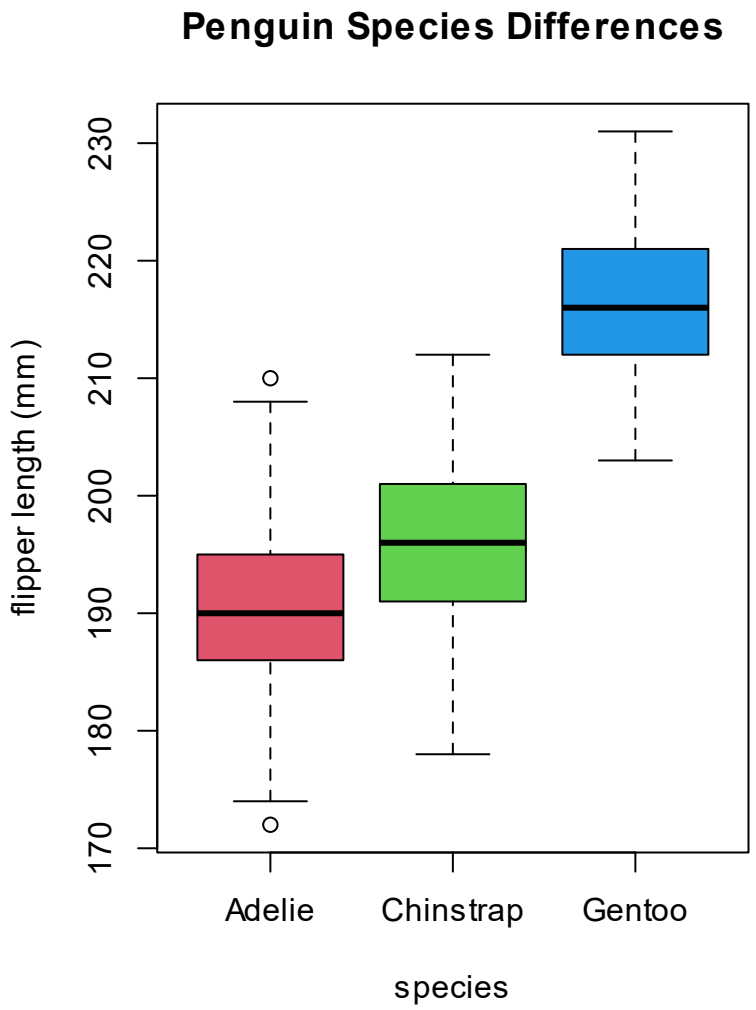
What data types can you use for each of the three question types?

- Differences?
- Associations?
- Correlations?

Differences: Numeric Data

- Real numbers and integers, for example
  - Length in mm
  - Mass in g
  - Age in days
  - Temperature in C
- Interval or ratio scale
  - Temperature in Kelvin vs. temperature in C
- T-tests, ANOVA





Associations: Integer Data

- Counts (integers)
- Chi-square tests

	Island		
	Biscoe	Dream	Torgersen
species			
Adelie	44	56	52
Chinstrap	0	68	0
Gentoo	124	0	0

Correlations: Numeric Data

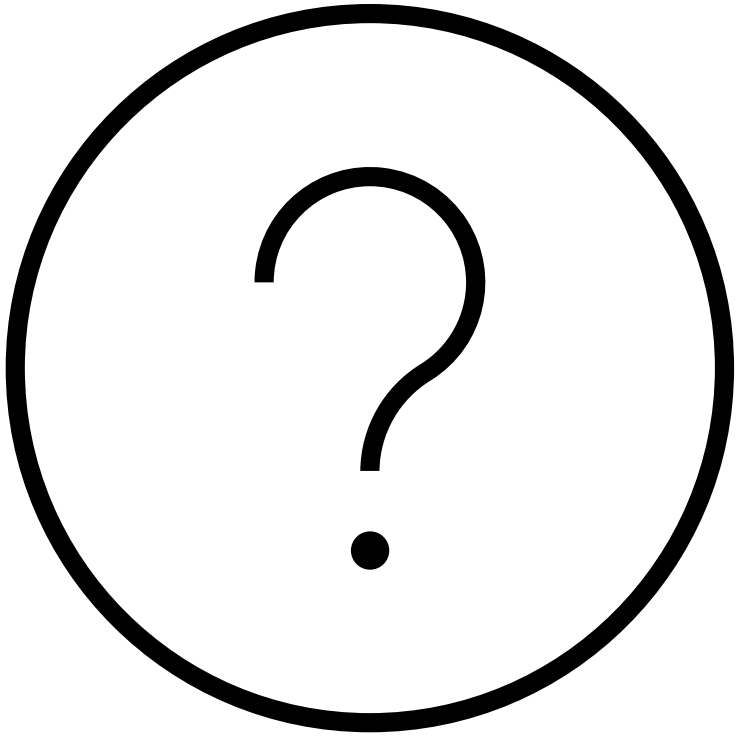
- Two dimensional: x- and y-axes
- Integer or real numbers
- Regression, Correlation test, ANCOVA

What's Wrong With Bias?

How are the general and statistical definitions of bias different?

Bias can lead to *non-representative* sampling.

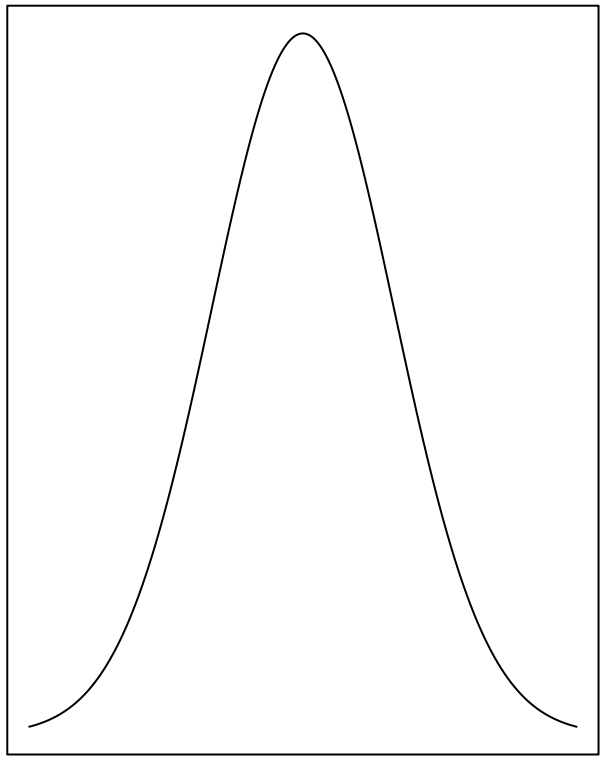
- A good sampling strategy can help.
- But… we can never totally eliminate the potential to collect an unrepresentative sample.
- We'll learn all about *sampling error* and the factors that contribute to it!

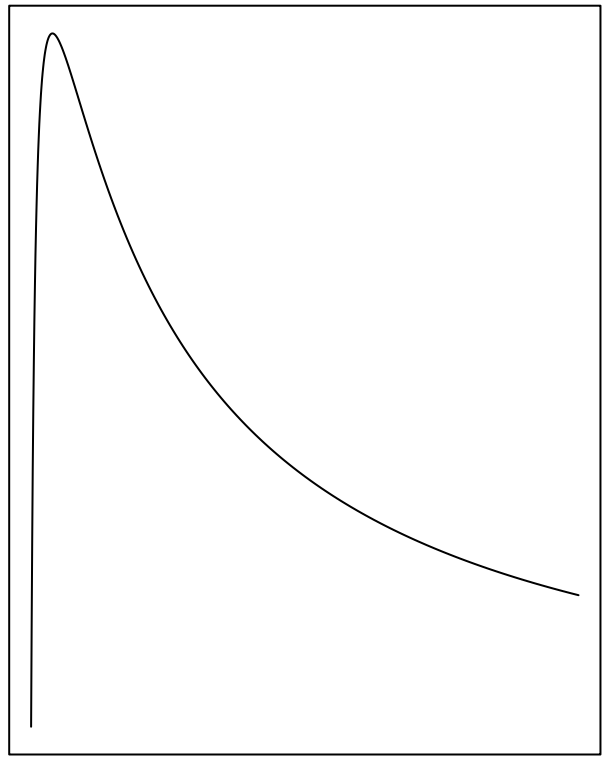


What About Skewed Data?

Skewed Data

- ‘Skew’ has a specific statistical definition:
- A skewed distribution is not symmetrical.
- Right- and left-skew





Statistical Definitions vs. Common Uses

The everyday uses of many of these are heavily *value laden*.

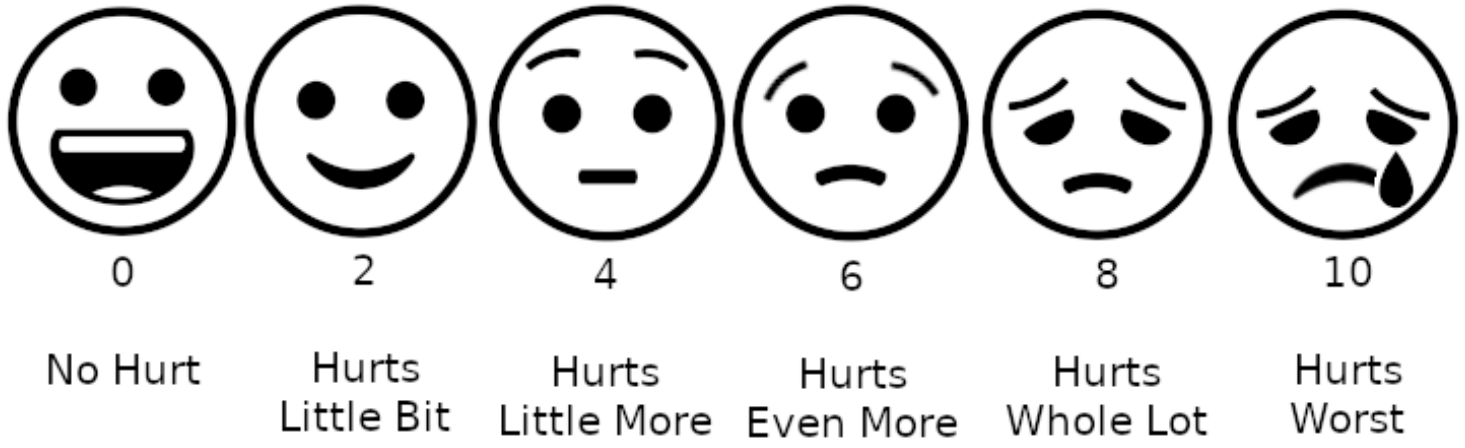
Statistical usage aims to be *value neutral*.

- Theory and hypotheses.
- Bias, accuracy, representative sampling.
- Random vs. haphazard
- Skewed data

Is it possible to be *value neutral*?

The Wong-Baker Scale

- What is the data type?
- What is the data scale?



DAFOR Scale: Vegetation Cover

D for Dominant

A for Abundant

F for Frequent

O for Occasional

R for Rare

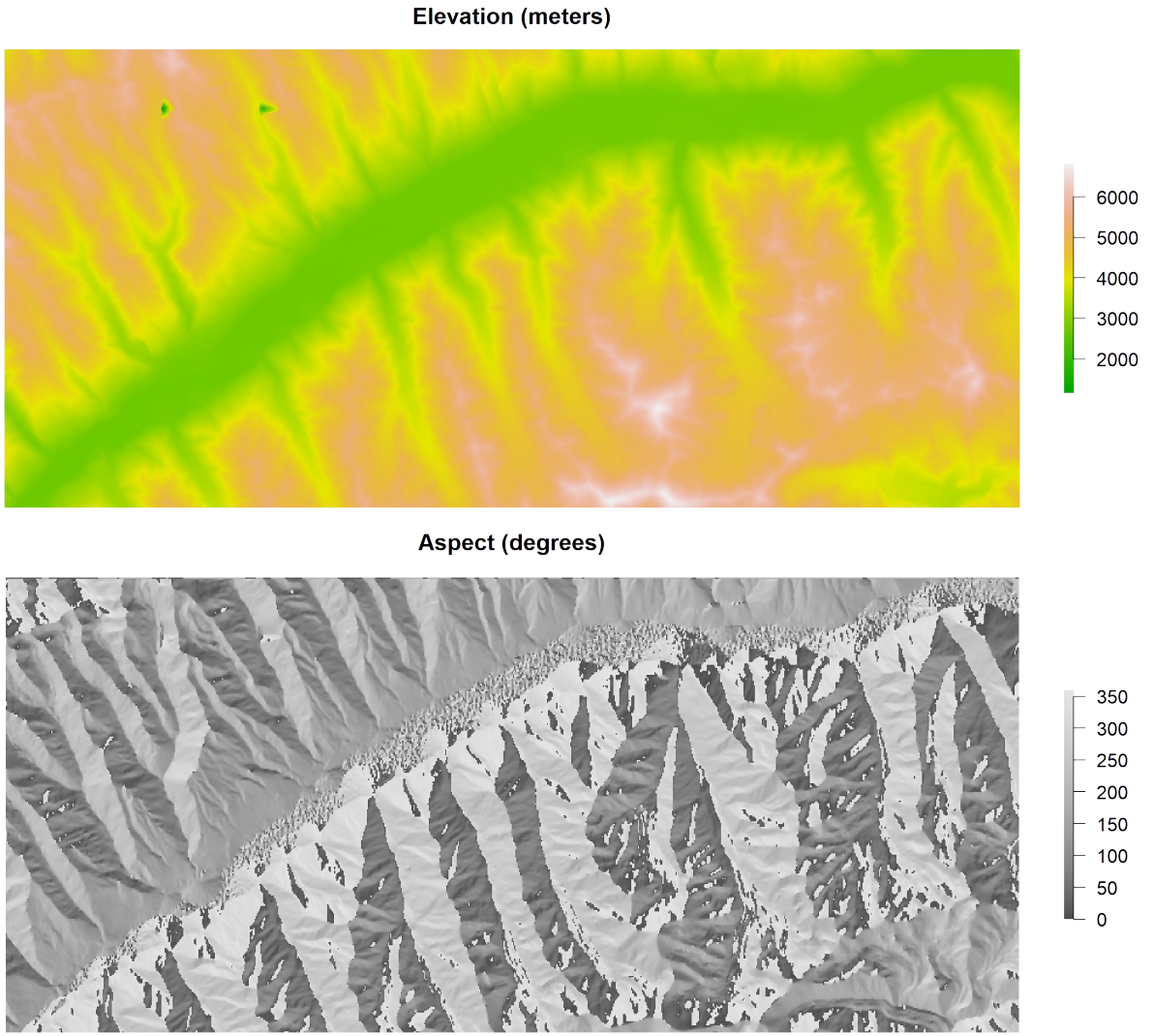
from botanicalkeys.co.uk

- What is the data type?
- What is the data scale?

Where Can I Find Periodic Data?

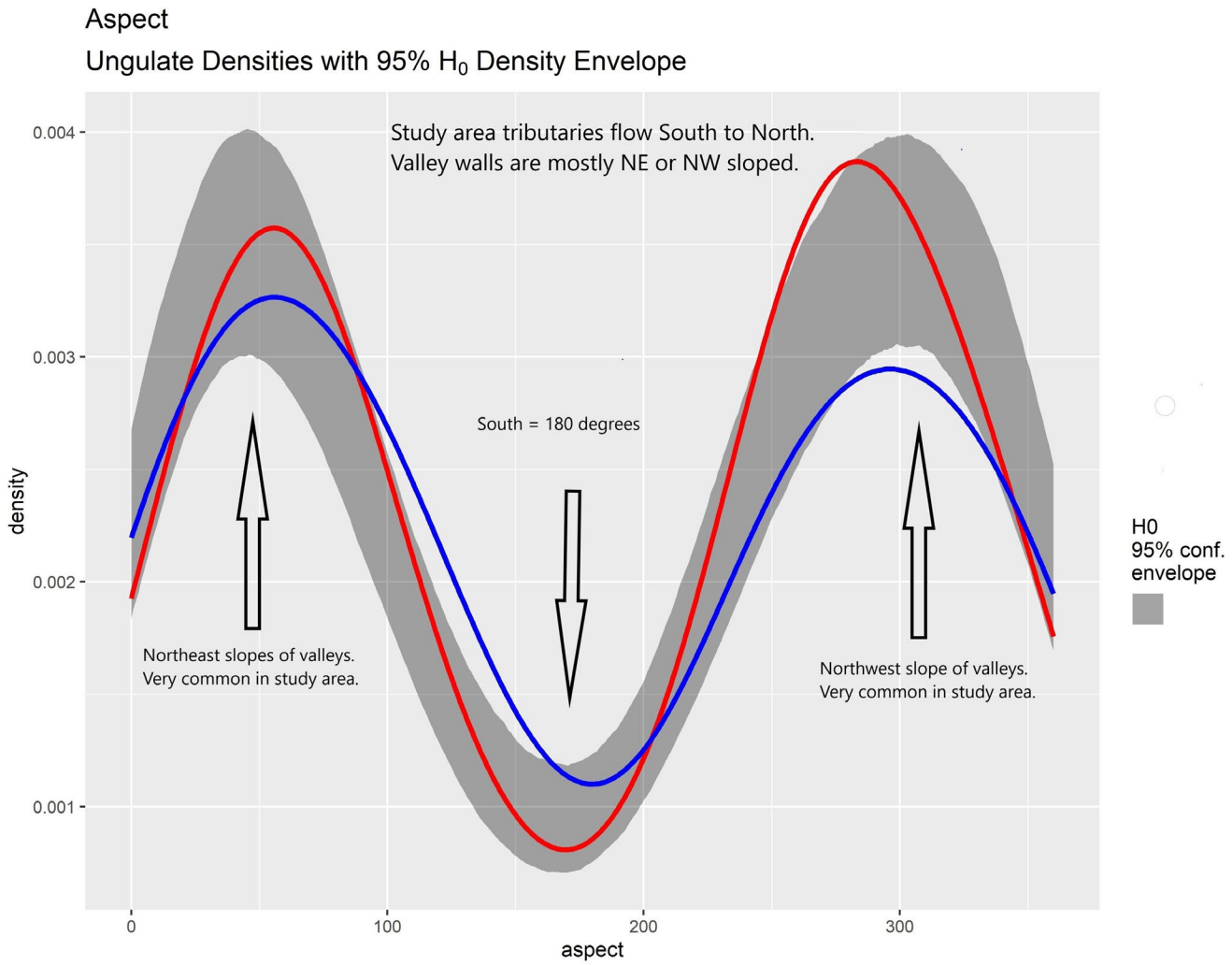
It can be difficult to find examples of periodic data.

- Aspect is a good example.
  - What is aspect?
- Resampling example on next slide:
  - x-axis is aspect
  - y-axis is likelihood to find an animal at a particular axis.
  - Gray ribbon represents likely values of aspect, selected *at random* by repeated simulation.
  - Lines are smoothed occurrence data for 2 ungulate species.
  - 0/360 degrees is North.
  - 180 degrees is South.



Periodic Data Example

- Herbivore preference for aspect (direction of slope):



Sampling Data

Why do we need to sample?

Some key questions:

- Why is sampling necessary?
- How can we minimize sampling bias?
- What are my sampling options?
- What is the best sampling scheme?
- What the heck is a quadrat?

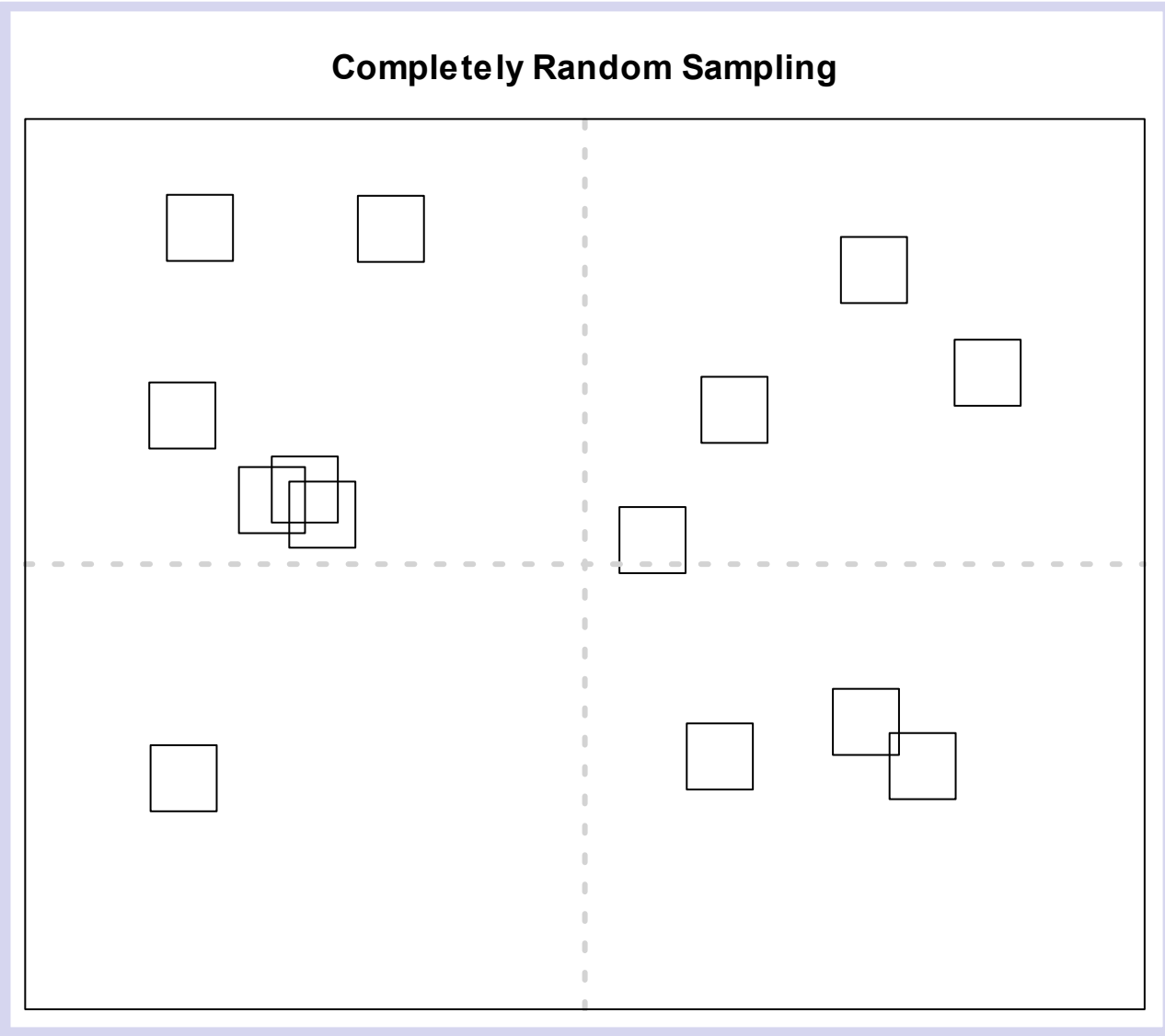
Sampling Methods

Gardener lists four (among many other possibilities):

- Random
- Systematic
- Mixed.
- Haphazard

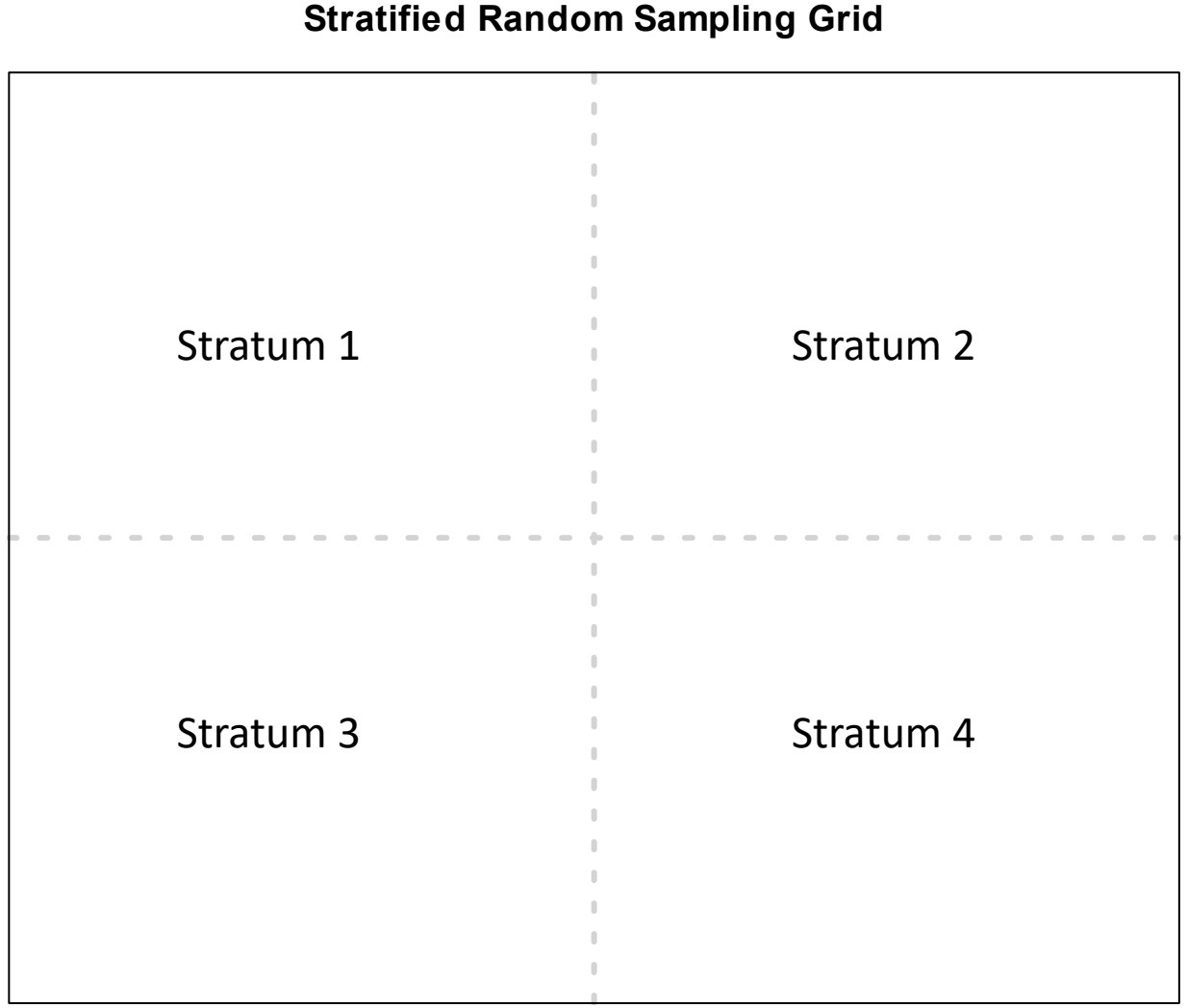
Random Sampling

- To replace or not to replace?
- In reference to random sampling with replacement, Gardener states:
  - “Obviously you do not need to place the quadrat a second time and count the buttercups again, you simply copy the data.”
  - Do you agree?
- Randomness can look less random than you expect…
  - Notice the clumps and wide spaces.

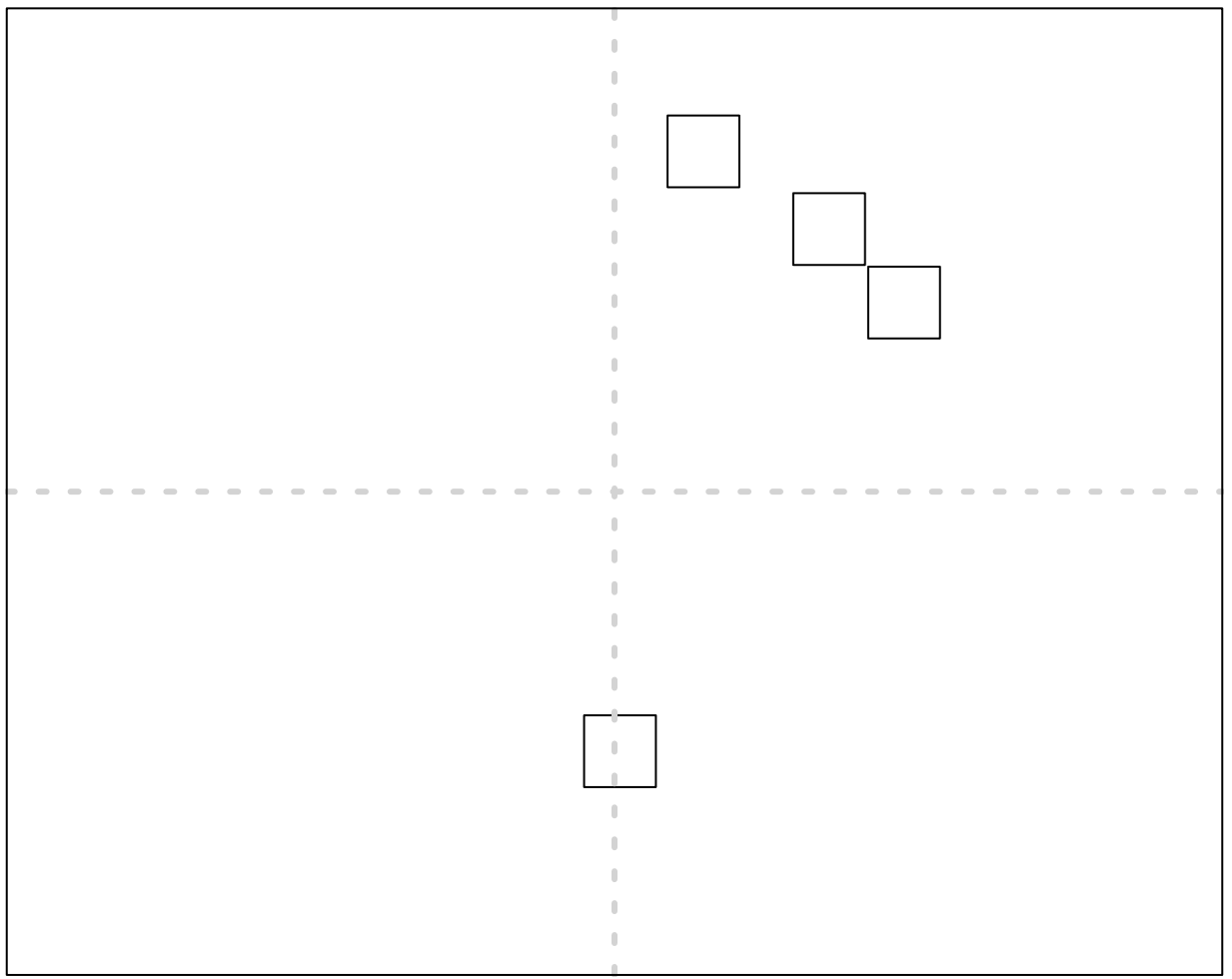


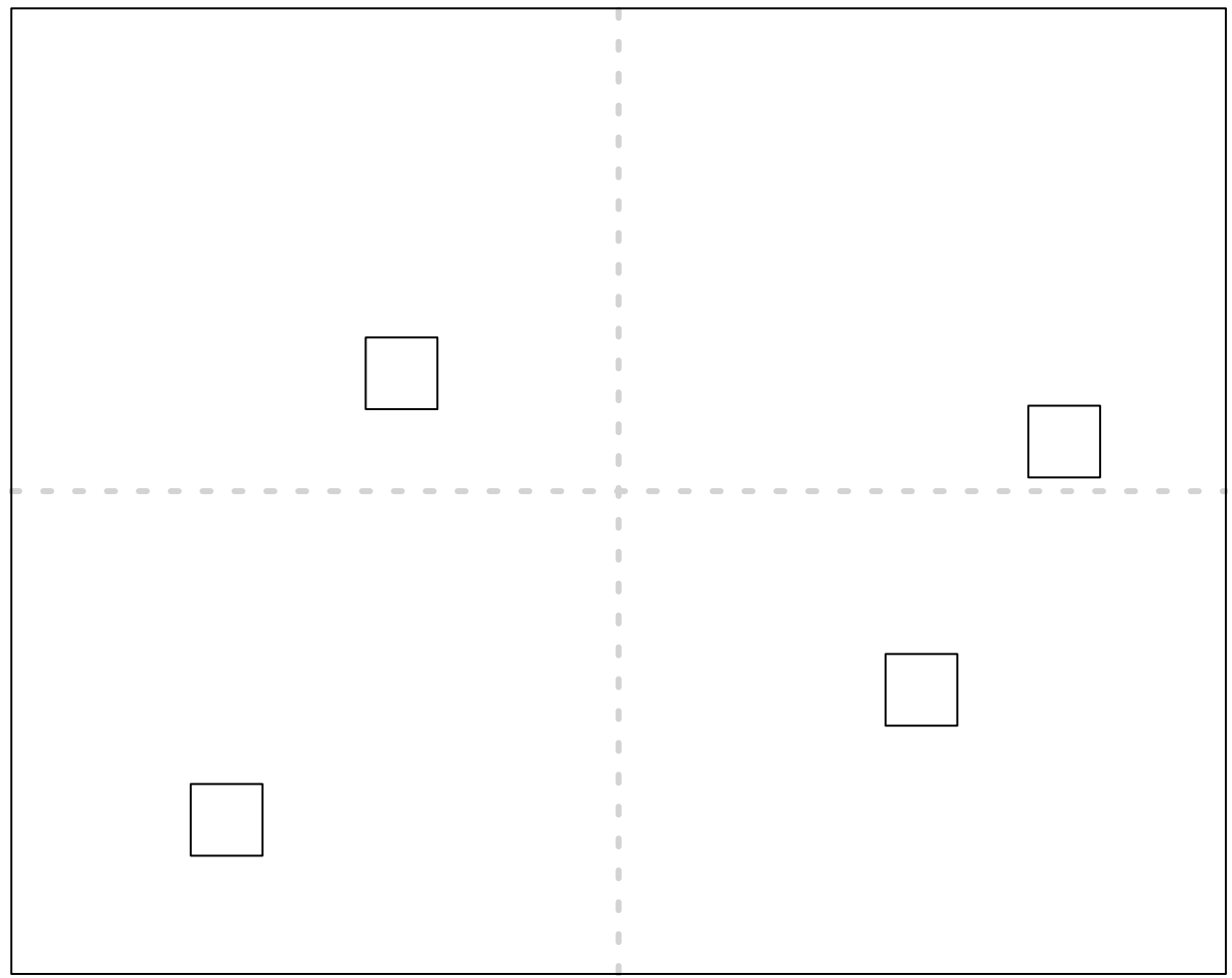
Stratified Random Sampling

- Completely random sampling can lead to clumping or large gaps, especially when the number of quadrats is small.
- The procedure is like random sampling, except you first divide your area up into **strata**, within which you select random sampling units.



Completely Random and Stratified Random Sampling



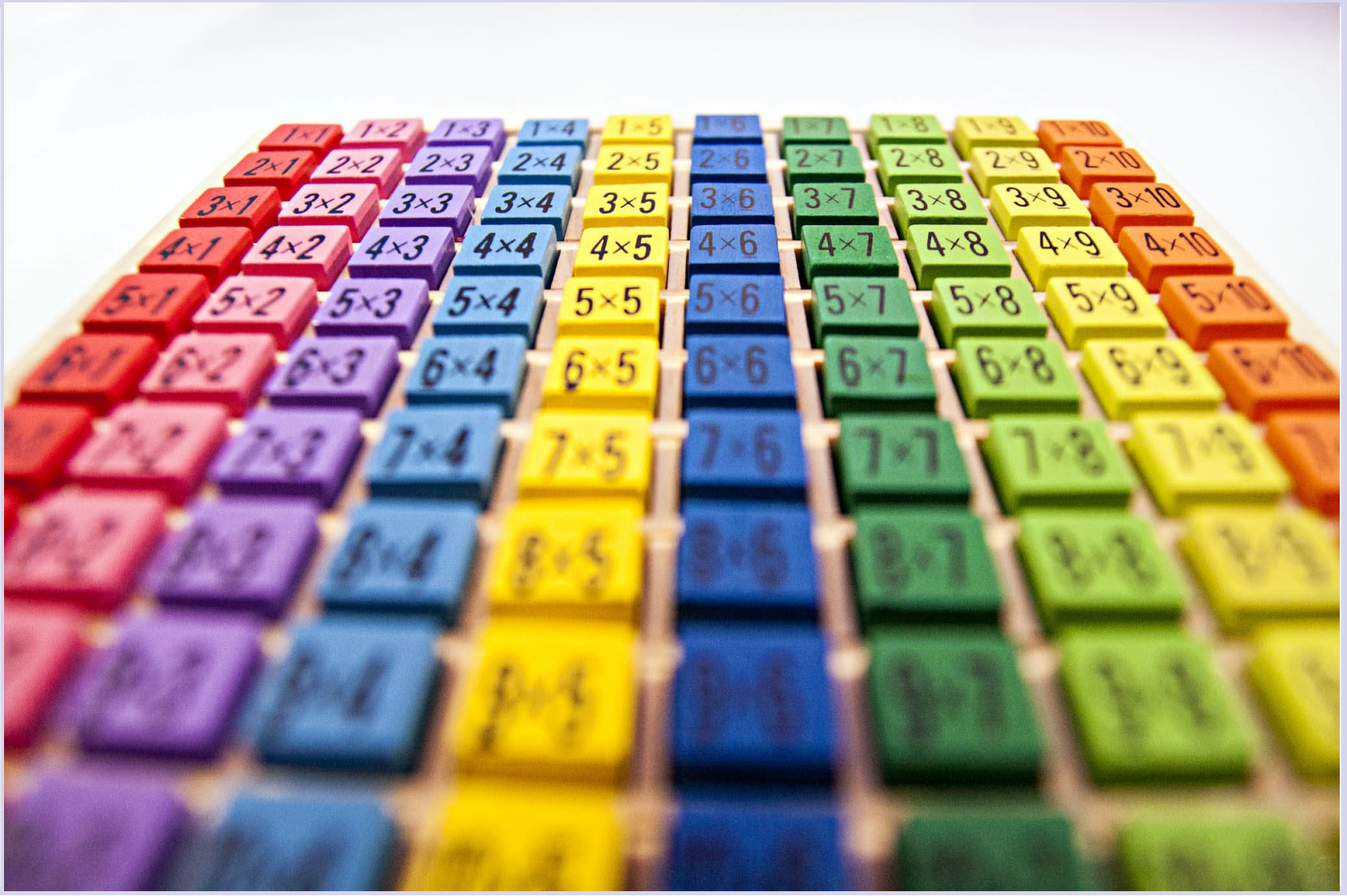


Systematic, Mixed, and Haphazard Sampling

- When is systematic sampling useful?
- Transects
  - Line
  - Belt
  - Interrupted belt
- Can you think of a scenario in which you would use haphazard sampling?

Key Concepts From This Slide Deck

- What is quantitative ecology?
- Scientific process: Planning stage
- Data types
- Data scales
- Vectors, data frames, and matrices in R
- Sampling



For Next Tuesday

Desert Shrubs Assignment

Desert Shrubs

1. Find instructions on GitHub
2. Form groups, self-select in Moodle
3. Choose a sampling scheme
4. Check out the plot randomizer script on GitHub