Stats in the Wild

Activity 3

Flora Explorer

Data moves

Grouping (using, creating or highlighting subsets)	**Filtering** (showing/ hiding subsets)	**Ordering** (sorting into an order)	**Summarising** (computing or calculating to describe a characteristic of a dataset)	
Linking (identifying corresponding case(s) in one representation and another representation)	**Inspecting** (hovering, clicking on or locating an object to gain information)	**Expanding datasets** (adding data, merging or joining datasets)	**Creating new variables** (e.g. rates/ proportions from existing data)	

Data forms

Choosing or creating a representation for a purpose

Finding and using relationships or patterns

Adapted from Hudson, R. A., Mojica, G. F., Lee, H. S., & Casey, S. (2024) *Data Moves as a Focusing Lens for Learning to Teach with CODAP*. Computers in the Schools, 1–26. https://doi.org/10.1080/07380569.2024.2411705

Overview

approx. 2-4 hours in total

1	2	3	4	5
Explore Scotland's native flora (plants)	**Collect** data in the wild	**Tidy and summarise** data	**Create** models of data	**Evaluate and critique** a model

You will need

Video notes: bit.ly/StatsWild3

- MWS Activity 3 Small Grids.pdf **(printed)**
- MWS Activity 3 Large Grids.pdf **(printed)**
- MWS Activity 3 Modelling Spreadsheet.xlsx
- Quadrat (metal grid)
- Microsoft Excel **(or other spreadsheet software)**
- Camera **(optional)**
- *Seek* app **(pre-downloaded)**

Why do this activity?

This activity focuses on **data modelling** – using data to create a statistical model that represents the real world. Data modelling activities encourage students to engage with meaningful phenomena by organising, structuring and visualising data. Through data modelling tasks, students are asked to answer questions *with* data rather than *about* data, supporting the development of **statistical literacy**.

This activity could be set as a **homework** task.

Possible approach

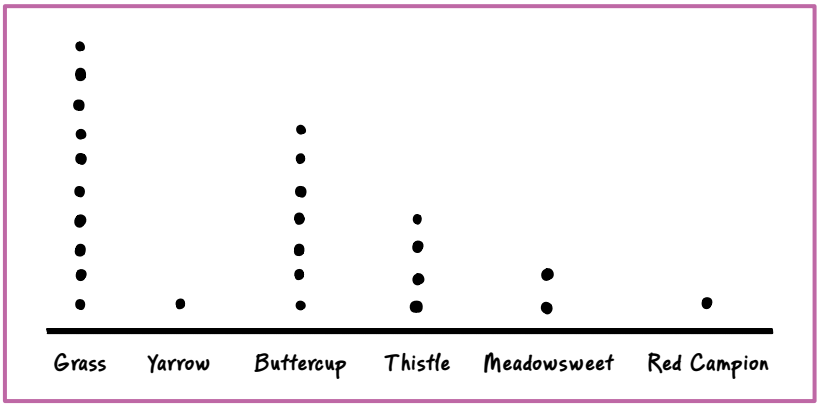
1 **Explore** Scotland's native flora (plants)

Ask students to **identify 5-10 plants** that are common in the local area and make some notes about features that might help identify them.

The aim of this activity is to get students to start thinking about how they might go about identifying different plants so that, in the later sections when they are collecting data, they are familiar with some common plants, and also have a way of identifying any plants they don't know when they encounter them. The **video notes** demonstrate several approaches to identifying plants that may be useful.

Encourage students to **compare** the plants they found. This could be a teacher-led, whole-class discussion. Students should be guided to consider:

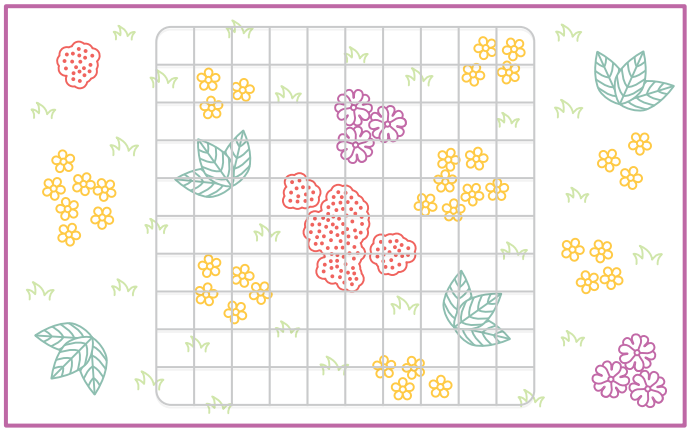
- Which plants are more common and which are less common?
- Which features help most in identifying plants?
- Which names to use for plants? (there is an opportunity to discuss a range of different names that have meaning to students)
- When and when not to group plants (e.g. different types of grass)



One approach could be to pool data and **co-create a dot plot** on a whiteboard. Draw a horizontal line and get students to shout out the name of a plant which can be added below the line, and then ask for a show of hands and add a dot for each student who identified that plant in their set.

2 **Collect** data in the wild

In this section, students go outside and map the plants in an open space **using a quadrat** (metal grid). This may require a little planning in terms of finding outdoor areas that are safe, bounded, not too far away and still offer opportunities to see a variety of plants.

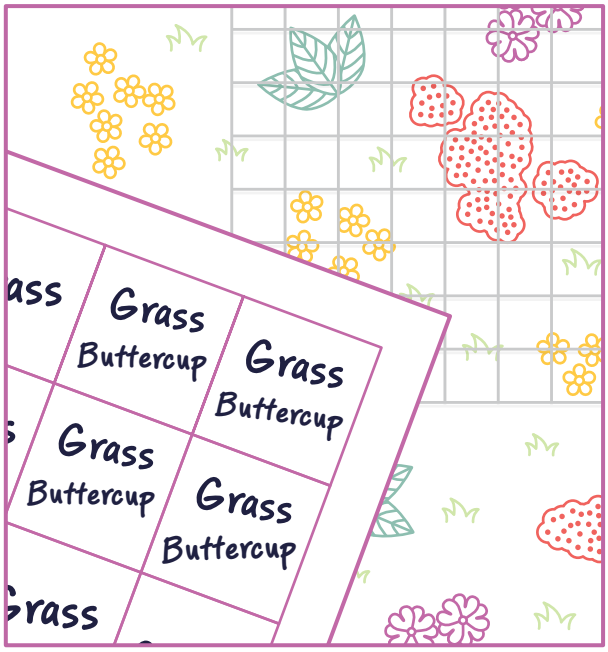


In small groups, students are aiming to collect **5–10 samples of data** by placing a quadrat on the ground and cataloguing the plants in each grid square, then moving it to a new spot and doing the same again. In a later section they will be asked to create models based on "one plant type per grid square", but for this first part it is okay if they record more than one plant per grid square – **encourage them to notice this** as they record.

Students should be instructed to take care **not to disturb** wildlife any more than is necessary, and to **not remove** any creatures or plants from their habitat.

Encourage students to experiment with approaches to recording their data using the large 10x10 grids provided. They could try:

- Shading regions as a one-to-one map of the grid
- Using shorthand with a key to note down plant names in the grid squares
- Listing the different plant names in each grid square
- Recording counts of each type of plant in each grid square
- Recording proportions of each type of plant in each square
- Recording the most common plant type in each grid square



If students have access to cameras, it might be a good idea to **take a picture of each quadrat in place** so that they can check any queries or details again later.

Students can use the *Seek* app (either outside if pre-downloaded, or back in the classroom) to **identify** each plant type. *This app does not require registration or user details and is for users aged 12+.*

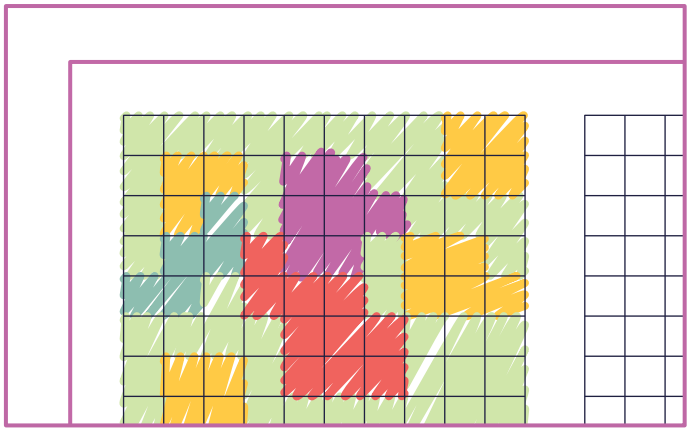
When students return to the classroom, it is important that they double-check the names of the plants they have photographed by researching them online, or using other sources of information such as books, since apps can misclassify and it is good practice to double-check classifications.

3 **Tidy and summarise** data

Lead a discussion focused on the process of data collection. Some questions you may choose to ask include:

- How did you decide where to place the quadrat?
- How might that have affected whether the sample was representative of the whole green space?
- How did you choose to record the data? Why that approach?
- What did you find challenging? (e.g. how to record a plant that was contained across two or more grid squares)
- Did anything limit the accuracy of your data? (e.g. uneven ground, tall plants)

Give the students a new set of blank 10x10 grids and ask them to fill one in for each of their samples, this time using a "1 plant type per grid square" approach. It may be that students have already used this approach – if so, this step can be skipped.



They will need to consider:

- How to decide which plant type to record if there were multiple in a grid square (e.g. the plant which covered more than half the square, or the plant that covered the largest amount of the square)
- How to fill out the grids (e.g. write the plant type in each square, create a key and colour-code the squares)

One key issue here may be where a plant is common but is never the main plant in a grid square - for example daisies amongst grass.

Students should be encouraged to discuss how well the 10x10 grids represent the real-world. If they have taken photos of the quadrats laid across the plants, this is a good time to compare them.

The students are now ready to **create their dataset** using the spreadsheet provided (`Data` tab). In the first column they should list all the plant types recorded across all the 10x10 grids. Each column represents one sample, and they should count the number of grid squares containing each plant type and record it. The table is designed to calculate the mean of each plant type and a proportion for each plant based on the means.

You may choose to **lead a discussion** to help the students to consider:

- What the 'average' of each plant represents
- Whether the averages are a good representation of the data
- Whether there are any particular averages that over- or under-represent an individual plant type, and why this might be the case

4 **Create** models of data

In this section the students will use their data to **create a model** of the quadrats which they can use to simulate the green space. The `Model` tab in the spreadsheet is already set up with formulas so that students can focus on considering the proportions that emerged from their own data.

The model works by randomly generating 100 integers between 1 and 10 in a grid and then using conditional formatting to colour code the numbers such that the proportion of each colour reflects the proportions of each type of plant in the model.

The model building process is as follows:

1. Write a list of all the plant types recorded.
2. Look at the `Proportion of models` column in the `Data` tab.
3. For each plant type, round the proportion to the nearest 10.
4. In the `Model` tab, fill out the ***Model table*** to reflect the rounded proportion. For example, if Rye Grass was assigned a rounded proportion of 30%, add the word Rye Grass to 3 of the 10 rows.
5. Set up conditional formatting in the simulated quadrat. (There is more information in the **video notes** for how to do this).

Pressing the `F9` key will refresh the random numbers and generate a new simulated quadrat.

Once students' models are working, they can record the data for 5-10 simulated quadrats in the table by counting the total number of grid squares containing each plant type. They may also like to take screenshots of their simulated quadrats and save them in a new document.

5 **Evaluate and critique** a model

In this final task, students **critique the model** in terms of how well it reflects:

- The data used to generate it
- The real-world

The model has numerous flaws and students should be **encouraged to discuss these** and **propose ways they may be improved**. As an extension activity for students who are confident in the use of Excel, you may choose to give them some time to improve the models by, for example, extending the model table to twenty rows and rounding proportions to the nearest five.

Some examples of flaws in the model include:

- Any biases identified in the original sample are replicated in the model
- Rounding proportions to 10 percent may over- or under-represent certain plant types
- Plant types with less than 5% coverage may not be included in the model
- Using one-plant-per-square reduces the resolution of the model
- Plant types may tend to cluster in areas on the real quadrats but are randomly spread in the simulated quadrats

Example classroom dialogue

I'm going to pick a colour for each plant type and colour in the grid so that it matches where all the plants are...

What should I do if a plant covers more than one grid square?

I'm not sure how to decide which plant to write down when the two plants in the grid are exactly half...

If I round to the nearest ten I get loads of plants that are zero. Then they don't show up on the model...

References and further reading

Bargagliotti, A., Franklin, C. A., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II).* American Statistical Association and National Council of Teachers of Mathematics. https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx

Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM, 50*(7), 1113–1123. https://doi.org/10.1007/s11858-018-0989-2

Macey, D., & Hornby, W. (2018). *Teaching Statistics.* Cambridge University Press. https://www.cambridgebookshop.co.uk/products/teaching-statistics?srsltid=AfmBOooZpGVsg-riGqi3aTJu2A09X_wGSR9o9KezHWFauEAQdHGcSvYn