Mapping Innovation

Activity 4.2.1. Filtering a real ocean dataset down to a sample you can defend

You use a live geographic information system to find deep-sea sponges that meet five numeric criteria at once, then justify the sample you picked. Four of the five criteria can be filtered. The fifth cannot, and that is the part people fail.

WORTH 5 POINTS | DOES NOT HELP ON THE WEBXAM | AT A COMPUTER. NO LAB HAZARDS | ABOUT 60 MINUTES

What this is and why it is worth doing

Marine sponges make byproducts with antibacterial, antiviral, antifungal and anticancer activity. Different species make different substances, so a lab that wants a real shot at finding something needs variety in one dive rather than one species collected many times.

You are working with the Drug Discovery Lab at Innovation, Inc. Your job is to find deep-sea regions with high sponge **biodiversity** and pick a sample that satisfies every criterion the experimental protocol imposes. The dataset is real: the National Oceanic and Atmospheric Administration's National Database for Deep-Sea Corals and Sponges, served through ArcGIS Online as the PLTW Sponge Hunters map.

DOES THIS HELP ON THE WEBXAM? NO, AND IT IS STILL GOOD WORK

WebXam 072110 weighs two domains only, and both are laboratory practice: handling, preparation, storage and disposal, and biotechnology research and experiments. Filtering a dataset is neither. There is no bench and no experiment here.

What it is, is genuine data literacy. Applying several numeric conditions at once to a real scientific dataset, noticing which condition a filter cannot express, and writing a justification somebody could check, are things you will do for the rest of your life whether or not a state exam measures them. Do it because it is useful, not because it is on a test.

How long it takes and what you need first

- **About 60 minutes.** Roughly 15 to learn the map, 30 to filter and select, 15 to write.
- **A computer with a browser.** A mouse is much easier than a touchscreen for panning and clicking individual points.
- **A way to take screenshots.** You have to submit the pop-up boxes proving each criterion, so know how to screenshot before you start.
- The PLTW Sponge Hunters ArcGIS map, linked from the activity. Open it in its own tab.
- No safety equipment. This one is at a desk.

What a GIS actually is

A **geographic information system** organises layers of information into a map. Every feature on that map carries two kinds of information, and telling them apart is the whole skill.

KIND	WHAT IT IS	EXAMPLE ON A HOSPITAL MAP
Spatial information	Where the thing is.	Each circle sits at a hospital's location, so you can see how the hospitals sit relative to each other.
Attribute information	Everything else about the thing.	Clicking a circle reveals the hospital name, type of care, city, state, county, zip, licensed beds, staffed beds, ICU beds, bed utilisation rate and potential capacity increase.

GIS locates things by **longitude and latitude in decimal degrees**, the same way a maths coordinate system locates a point with x and y. More decimal places means a more precise location. The value of a GIS is that you can lay unrelated layers on top of each other and see relationships nobody was looking for.

> **THE TWELVE FIELDS IN THE SPONGE LAYER**
>
> **CatalogNumber**, a unique record identifier. **ScientificName**, the genus and species as a Latin binomial, or the lowest level anyone could identify. **VernacularNameCategory**, the common name category. **Ocean**, the ocean basin. **Latitude** and **Longitude** in decimal degrees, ideally to five places. **DepthInMeters**. **ObservationDate**. **Density**, individuals or colonies per square metre. **LocationAccuracy**, the estimated uncertainty of the coordinates in metres. **Salinity** in practical salinity units. **Oxygen**, dissolved oxygen.

SPONGE CLASS	WHAT IT IS
Demosponge	A large, diverse group of structure-forming sponges, many reaching large sizes.
Glass sponge	A deep-water structure former capable of building large reefs.
Calcareous sponge	Characterised by spicules made of calcium carbonate. A few species live deep, most are shallow.
Homoscleroform sponge	A small group that grows very large in deep water.
Sponge (unspecified)	Nobody could determine the class. This is the one you filter out.

The five criteria

YOUR SAMPLE MUST SATISFY ALL OF THESE AT ONCE

1. Every sampled sponge is from one of the **four specified classes**, and you must have at least **two different classes**.
2. Depth of **at least 100 metres**.
3. Dissolved oxygen **greater than 0.1 mL/L**.
4. Salinity **less than 35.5** practical salinity units.
5. Sampled classes must be **within 250 miles of each other**.

Four of those are filters. The fifth is not. Criteria one through four are conditions on fields in the data, so a filter expression can express them. Criterion five is a relationship between the points you chose, which no field holds, so no filter can test it. You verify that one with the Measurement tool, by hand, after filtering.

The steps

1. Research where deep-sea sponges are found before you open the map. What water temperatures or salinity do they prefer? Do they avoid turbulent water? Record three to five preferences.

   Done when: three to five written preferences exist, with where you found them.

2. Open the PLTW Sponge Hunters map in its own tab. Zoom, pan and click individual points to see the pop-ups. Zoom in if no icons appear. The Home icon returns you to the default view if you get lost.

   Done when: you have opened at least five different pop-ups and know what fields they hold.

3. Open the Filter tool in the right-hand map tools panel and click **Add expression**. A filter expression has three parts: a **field**, an **operator**, and a **value**.

   Done when: the Filter panel is open.

4. Build the first expression: **VernacularNameCategory excludes sponge (unspecified)**. Save it. The symbols for filtered-out values disappear from view.

   Done when: unspecified sponges are gone from the map.

5. Add expressions for criteria two, three and four. **Set the filters to Match all conditions, not match any.** This is the single most common mistake in this activity.

   Done when: four expressions exist and the panel says match all.

6. Watch the ArcGIS Measurement video, then use the Measurement tool to check criterion five: are the classes you are considering within 250 miles of each other?

   Done when: you have measured an actual distance between two candidate points and written the number down.

7. Select your sample. Record the species name and location information for each sponge class you are harvesting.

   Done when: at least two different classes are selected and both are written down with coordinates.

8. Screenshot the pop-up box for each sample so that every criterion can be verified from the image. Depth, oxygen, salinity and class must all be legible.

   Done when: each sample has a screenshot in which all four field values can be read.

9. Write one to two paragraphs justifying your sample selection.

   Done when: the justification names each criterion and says how your sample meets it.

What finished looks like

1. Your three to five researched sponge preferences with sources.
2. Species name and location for every sponge class you selected.
3. A readable screenshot per sample proving criteria one through four.
4. Your measured distance proving criterion five, with a screenshot of the measurement.
5. One to two paragraphs justifying the selection.
6. Two short answers. What other reasons might you filter map data, with examples? How can GIS advance biomedical science?
7. Your response to the ethical scenario below.

THE ETHICAL SCENARIO

GIS lets a public health agency track disease spread by tracking the movement of infected people. That is genuinely useful and it maps individuals who are sick, which infringes their privacy. How should an agency balance individual privacy against data accessibility? Use what you know about **HIPAA**, the United States rules governing identifiable health information, to inform your answer.

WORKED EXAMPLE OF THE JUSTIFICATION FORMAT. INVENTED COORDINATES, NOT A REAL SELECTION

> We selected two classes from a single region in an invented basin. Sample one is a glass sponge at 44.10000 N, 130.20000 W, depth 640 m, dissolved oxygen 0.42, salinity 34.1. Sample two is a demosponge at 43.60000 N, 129.80000 W, depth 410 m, dissolved oxygen 0.38, salinity 34.3. Screenshots of both pop-ups are attached.
>
> Both samples clear every numeric criterion with room to spare rather than sitting on the line: the shallower of the two is more than four times the 100 m minimum, both oxygen values are close to four times the 0.1 mL/L floor, and both salinity values sit more than a full unit below the 35.5 ceiling. We chose points with margin on purpose, because **LocationAccuracy** in this dataset is an uncertainty in metres and several records carry uncertainty large enough to matter. A point that just barely clears a threshold might not clear it.
>
> For criterion five we measured 39 miles between the two points using the Measurement tool, well inside 250. We rejected a third candidate, a calcareous sponge with better oxygen values, because it measured 310 miles from the other two. Adding a third class would have improved biodiversity and broken the distance criterion, and the criteria are not negotiable.

Where people get stuck

Filters set to match any instead of match all.
With match any, a point that satisfies one criterion and fails the other three stays on the map. Everything looks fine and your whole selection is invalid. Check the setting before you select.

Trying to filter criterion five.
No field in the dataset holds the distance between two points you have not chosen yet. It cannot be a filter. Measure it.

Reading the oxygen units carelessly.
The dataset describes the Oxygen field as being in mg/L or mL/L, and the criterion is stated in mL/L. Those are not the same unit. Note which one the records you picked are in, and if you cannot tell, say so in your justification rather than pretending.

Picking one class and calling it a sample.
The criterion says at least two different classes. One class in eight locations fails.

Selecting points that only just clear a threshold.
A record at 101 m depth or 0.11 oxygen has almost no margin, and LocationAccuracy tells you these records carry uncertainty. Prefer margin, and say that you did.

Screenshots that crop off the field you were proving.
The screenshot is the evidence. If depth is not legible in the image, that criterion is unproven no matter what you typed.

Truncating coordinates.

Decimal degrees carry their precision in the decimal places. Copy what the pop-up shows.

Skipping the research in step 1 and filtering blind.

Knowing what conditions sponges prefer is what tells you where to look first. Without it you are panning a whole ocean at random.

How to submit and how it is credited

THIS ITEM IS WORTH 5 POINTS

Seven deliverables, mostly screenshots and short writing. Remember the 20 point per semester cap across all extra credit.

- **Where it goes.** The same place as everything else. Your class form on the portal, or handed to Mr. Mendoza in person. Nothing about extra credit needs a different route.
- **What to hand in.** Photographs of your work where the handout asks for them, plus your written pages. Photographs of handwritten pages are fine. A photograph you cannot read is not.
- **Put your name and period on every page and every photograph filename.** Unlabelled work cannot be credited and it is the single most common way points get lost.
- **When.** Any day after the pack opens, right up to the last week of the semester. There is no deadline inside that window and no penalty for going slowly.
- **Feedback.** You get it back with a note. If a piece is short of full credit you may fix the named thing and resubmit once.

Words this activity uses

Geographic information system

GIS. A tool that organises layers of located information into a map and lets you analyse them together.

Spatial information

Where something is.

Attribute information

Everything about a mapped feature other than its location.

Decimal degrees

Longitude and latitude written as decimal numbers. More decimal places means a more precise location.

Filter expression

A condition applied to a map layer, built from a field, an operator, and a value. It hides features that do not match.

Match all conditions

Every expression has to be true for a feature to stay visible. The opposite, match any, keeps features that satisfy just one.

Biodiversity

The variety of life. More variety means more untested chemistry to look at.

Practical salinity unit

psu. A measure of how much salt is dissolved in ocean water.

Dissolved oxygen

Oxygen dissolved in water, reported in this dataset in mg/L or mL/L.

LocationAccuracy

The estimated uncertainty of a record's coordinates, in metres. A reminder that a point on a map is an estimate.

Latin binomial

A two-part scientific species name, genus then species, such as Primnoa pacifica.

Spicule

A small structural element in a sponge. Calcareous sponges have calcium carbonate ones.

HIPAA

The United States rules governing what may be done with identifiable health information.

Therapeutic

A treatment, therapy or drug.