Humpback whales: what the data reveals

Please refer to the online lesson plan on the DT Hub to access all website links and additional resources.



**YEARS 7-8**

Dr Michele Thums and her team have studied the behaviour of whales in Pender Bay in the Kimberley region of Western Australia.

Humpback whales migrate northwards from Antarctica along the WA coast to the Kimberley, where they breed. The data collected as part of this project attempted to identify crucial habitats for the whales, including breeding and feeding grounds. For this project, we are considering just one of the datasets from the study, which consists of observations of humpback whales in Pender Bay from 2009–12.

Observers stand at an observation point on the shore of Pender Bay in groups of usually 2–4 people, and record the whales they can see during 5-minute sweeps of the bay every 20 minutes. The observations typically happen during whale season – roughly from June to November.

The observations haven't been completely consistent. Sometimes there are more observers. Observations can be more or less frequent – sometimes they start in June, sometimes in July. sometimes they finish in October, sometimes in November. Sometimes observers recorded environmental conditions such as temperature and wind direction. Sometimes they didn't.

The lesson follows an inquiry process where students use the dataset to answer relevant questions about the whale population. They consider what other data they would need in order to effectively examine the impact on humpback whales of sonar activity and noise from development.

This lesson was created and developed in partnership with **Pawsey Supercomputing Centre** and **Western Australian Marine Science Institution** (**WAMSI) Kimberley Marine Research Program**.

**Keywords:** Data science, Dataset, Fauna data, Humpback whales, STEM, Science, Scientific method, Higher-order thinking, Spreadsheets, Data representation, Data analysis, Data visualisation, Excel, Numbers, Cross-curriculum priorities, Sustainability

**Learning intentions**

* Use relevant techniques to clean up data (eg remove blank rows)
* Save, store and use CSV files
* Use spreadsheeting software to sort and filter data to help answer inquiry questions
* Analyse data and look for patterns
* Create graphs such as a bar graph to present data
* Select the most suitable graphing format to visualise and present data
* Use higher-order thinking skills (interpreting, analysis, inferring, summarising, evaluating)
* Understand the scientific method used to collect data about the whales
* Design further experiments to collect extra data needed to understand the whale population

Learning hook

1. Use a relevant hook to motivate students to discover more about humpback whales.
* Use a quiz to stimulate interest in Humpback whales and discussion of some key concepts.
* A dilemma could be introduced to encourage students to consider a range of issues and concerns to determine the appropriate use of a World Heritage area. In the student task we have used the dilemma: ‘Should sonar activity and other noisy development be allowed in Pender Bay?’



Girls in focus: Research has shown that girls are interested in careers that have a positive social impact. This topic shines a spotlight on a contemporary issue that requires a balance of social, ethical, economic, and conservation perspectives.

Learning input

1. Briefly discuss students’ knowledge or any experiences of whales and what they know in particular about humpback whales. Discuss the kinds of noisy development that might occur in Pender Bay, in particular, naval sonar activities and oil and gas exploration. What impact might this have on the whale population? Discussions may also cover an ecological focus such as their distribution, habitat, feeding habits, and behaviours, as well as tourism/fishing/boating and human interactions. Provide an opportunity to locate and refer to relevant resources on the difference between baleen and toothed whales (humpback whales are baleen whales).
2. Introduce the scientific study conducted to gather data to understand whale behaviour and distribution along the West Australian coast. Provide the dataset ‘Whale data’. Organise students into collaborative teams to investigate the dilemma and develop their team response.

Girls in focus: Research has shown that girls value interaction and collaboration. Promote collaboration and recognition of the varied skills within the team, ensuring that all students are given the opportunity to manipulate the data set.

1. Use relevant video and text resources to provide a context for the way in which scientists (male and female) collect data to be used in conservation-related decisions. Feature the work of Michele Thums in this project. Use the video: Michele Thums ABC interview: humpback whales in Pender Bay.

Girls in focus: Research has shown that when girls are exposed to positive STEM role models, their interest increases, along with an improved self-concept related to STEM fields. When showing scientists in action and use of technology, provide a balanced representation of males and females.

Learning demo

1. Depending on your students’ skills and familiarity with spreadsheeting software such as Excel for Windows, Numbers for iOS systems, or Google Sheets using an internet browser, you may decide to scaffold their introduction to the dataset or let them explore the dataset based on a question of interest.

Girls in focus: Girls often have a poor self-concept as mathematicians, believing commonly-held stereotypes that boys are naturally better than girls at maths. If you observe a number of girls showing reluctance or a lack of confidence with using spreadsheets, consider offering an additional session to practice and improve their skills. Support a growth mindset by praising their effort, strategies and behaviours.

1. Provide students with the Humpback whale: student task presentation slide deck, which provides a guide to the student inquiry.
2. This inquiry provides an opportunity to consider the issues of scientific data collection in the marine environment, and to extend scientific understanding by designing new experiments to fill gaps in existing data and knowledge.

Explain that in many scientific endeavours the data collected is a proxy for the data we really want. In this case, we have data that counts the observed number of whales but not the actual number of whales in the bay. The latter is, of course, almost impossible to count with complete accuracy. It’s also impossible to have people observing all day and night. So we try to approximate numbers in the bay by observing for short periods of time and extrapolating.

Learning construction

1. Enable students to work with the data in their preferred spreadsheet package to explore ways of making sense of the csv file. Open the Pender Bay data in a spreadsheet program such as Google Sheets, Excel or Numbers.

Point out that there are 7203 rows and 42 columns of data, which is really difficult to make sense of at first glance. Using the slide deck, the students will work through the question: How do we understand the data?

1. Discuss the mindset needed to investigate large datasets. Integral to the mindset is being organised and logical. Ensure that students first duplicate the sheet in their spreadsheet file, and save a new copy of the data, so the original remains untouched. Rename one sheet ‘original’, and one ‘copy’. Work in the copy sheet, not the original. This ensures that students can always go back to the untouched data. This is good practice when doing data analysis – never edit your original dataset!
2. For those that require assistance, help them start by working across the columns and looking at all of the titles. Categorise the titles into groups. For example, we have details about:
	* the observation sessions themselves: date, time (start and finish)
	* number of observers present
	* numbers of whales: adult, calves and total
	* a set of behaviours such as breaching and lobtailing
	* a collection of observations about the weather and related conditions
	* boats sighted (yes or no, type, and number)
	* tidal info (incoming/outgoing/low/high, and spring/neap).

Which groups of information are going to be most useful to determine whether Pender Bay is an important whale habitat?

1. Reiterate that, ‘We want to know whether whale numbers are changing’. Help students through this process if needed.

Start by sorting the data by the total number of whales. When were the most whales observed? Insert a bar chart of the total whales. What extra information do you need to add to the graph in order to determine when the most whales were observed? The top 49 entries for most whales observed were all August 2009, but how different are the numbers for other years? What might be the easiest way to find that out?

1. Highlight the ‘Total number of whales’ column and insert a chart of that data. It should look like this:

Image: Bar chart of total number of whales observed over time

1. Discuss as a class what information you need to add in order to be able to identify the months when there are more whales. Add the month column as the X-axis of the graph. Make sure you include axis labels and a graph title.



Image: Bar chart of total number of whales observed over time against month

1. Try graphing the number of adults and calves against the months. To do this, highlight all three columns using the ‘Ctrl’ or command key and insert chart. You may need to go into the chart settings and make the months the X-axis, and the adults and calves the Y-axis. Point out that numbers have decreased significantly since the first year of observations, but increased a little in the final year of recorded data.



Image: Bar chart of number of adults and calves sighted, by month

Note the sharp increase in numbers in the final year of data, in contrast to the slow rise in previous years. Discuss what happened here. Go back to the spreadsheet and note the year in which each month’s sightings began. How does this change your ability to compare the data from year to year?

1. Make separate sheets for each year and copy the data across. To do this, select the first row for the year by clicking on the row number, then hold down the shift key and scroll down to the last value for that year (there are over 1500 entries for each year so this will take a while!). Then hit control-C (or command-C on a mac). Calculate the average number of adults, calves and totals for each year. Do the same for each month. Use the average function. How has the number of whales changed over time?

Copy the averages into a new column in a new sheet, together with the years they relate to, and graph the changes. Does this make it easier to see how the numbers have changed? Is it easier to read as one graph or with separate graphs for adults, calves, and totals?

Learning demo

1. Write a proposal to study whale numbers in Pender Bay to determine the impact of sonar and other noise pollution. List the ideal setup for your study, including details such as timing of observations, number of observers, and types of information to record. Consider how technology might be used in monitoring whale populations, for example some projects have piloted the use of satellite technology.
2. Not a lot is known about the response of humpback whales to sonar. What kind of experiments could you conduct to determine the impact of sonar activity in Pender Bay?

What are the ethical considerations of running that kind of experiment?

1. As a result of their inquiry, students are asked to write a proposal to run their experiment. Ensure students include the reasons why the experiment is important, as well as the details of what the experiment is and how it should run. In their proposal, they are asked to think about:
	1. the ethics of their experiment. What will they do if it becomes clear it is harming the whales or scaring them away?
	2. the resources required to conduct their experiment as well as how many people they will need, and for how long
	3. why should we be concerned about protecting whale habitat? They should be able to argue their point of view against others in the class.
2. As a team, students share their experiment designs. Each group decides on a suitable way to present their experiment proposal, such as a 1–2 min video, brief report or infographic. Decide on the best experiment, and explain what criteria you have used to judge them.

Girls in focus: Research suggests that girls are motivated when they are given opportunities to approach projects their own way, exercising their personal preferences and creativity. Engaging with creative problem solving also encourages students to embrace failure as part of the learning process, building resilience.

Discussion

Digital Technology focus

* Consider how technology might be used in monitoring whale populations, for example some projects have piloted the use of satellite technology.

*As satellite imagery technology improves, better quality images are possible. Now with the high-resolution images it is possible to identify whales captured by satellites.*

*The tradeoff is the cost of gathering this data and the process is time-consuming. Work is being done to develop a computerised scanning system that can spot the whales within the black-and-white satellite images.*

* In what ways can you use a spreadsheet to analyse data?

*Once the data is cleaned up, it can be sorted and filtered according to the question being answered. Typically the data is sorted by columns based on the cell’s value or text. Sorting allows for grouping entries by a common attribute which helps identify patterns or trends in data.*

* What skills did you develop through your data analysis?

*Manipulating spreadsheets to help answer a question using sorting and filtering*

*Choosing a suitable graph to visually present data*

*Labeling a graph*

*Developing a mindset of questioning (can this data help me answer a question of interest, what are its limitations?)*

* Consider the data collection technique. What impact might it have on your calculations? For example:
	+ The number of observers varies. What impact might this have on the number of whales spotted?
	+ Different days have different numbers of observation sessions at different times. Can you compare days? Months? Years? What kind of comparisons might be valid?
	+ What can you compare from day to day without skewing your results?

(Hint: Some values to consider include average number of whales per sweep, average number of whales per day and maximum number of whales in one sweep in a day.)

* + What if it’s the same whales from sighting to sighting, and whales that happen to hang out within sight of the shore get counted on every sweep
* Consider the Limitations of the data

Bear in mind that the number of whales observed is not the same as the number of whales present, so you won’t get absolute numbers.

Like many forms of data collection, the data you get (how many whales the observers saw each time they observed) is a proxy for the data you actually want (how many whales actually use the area).

**Science focus**

* What factors affect the humpback whale populations?

**Whaling:** There is currently a suspension of commercial whaling. However, in the future there may be increased pressure to resume whaling.

**Climate change:** Impacts of climate change may include increasing sea surface temperatures, decreasing sea ice cover, rising sea levels, changes to ocean circulations, ocean acidification and changes in salinity. These changes may lead to changes in migration, species distribution and prey availability.

**Overharvesting of prey:** Lower numbers of Antarctic krill through over-harvesting may be a potential future threat.

**Noise interference:** Noise sources that are potential problems include seismic exploration, industrial noise, shipping noise, and sonar systems.

**Habitat degradation:** Modification in areas of importance to the humpback whale may result in reduced occupancy, compromised reproductive success and even mortality.

**Entanglement:** Entanglements in nets can cause serious injury and distress to whales, and in some cases lead to the death of the animal.

* Data is available from previous scientific studies (2009–12); why would it be useful to follow the same surveying methodology and approach?

To ensure meaningful comparisons can be made from historical data, it is important that the same methodology be used and consistent data is collected using the same parameters.

Why is this relevant?

**Acquire data**

* Acquire data examines how we collect and access data from a variety of sources.
* Students can generate data of various types through their own experiments and investigations.

**Store data:** Record data in a format that allows it to be easily accessed or obtained. Students can describe how the data they have acquired can be stored in different ways using different representations and/or software. It is important to select the most suitable representation.

**Organise data:** Organise data explores the ways we order, sort and arrange data to assist us with interpretation in different contexts.

**Interpret data:** Use data and its characteristics, properties and patterns to form a conclusion or derive meaning from it. Students can work with data that requires some simple processing using software. This could be in the form of simple spreadsheet calculations or using data in code. They draw conclusions about the data as a result of this processing.

Resources

* Whale data
* Whale: student task presentation slide deck
* Whale quiz

### Humpback whale information

* Humpback whales in Pender Bay: short video about the WAMSI study of humpback whale Distribution in the Kimberley
* Humpback whale distribution: more information about the WAMSI humpback whale project
* Marine Mammal Laboratory: NOAA information about baleen and other whales
* Species Profile and Threats Database: Megaptera novaeangliae — Humpback Whale: scientific background about humpback whales

### Video tutorials

**Spreadsheeting tips:**

Please refer to the online lesson plan on the DT Hub to access all website links and additional resources.

References

* Australian Curriculum, Assessment and Reporting Authority (ACARA) 2016, Cross-curriculum priorities (especially Sustainability)
* Australian Government Department of the Environment and Energy, Whales, dolphins and sound
* Australian Museum: Humpback Whale
* Choi, C 2011, Amazing Navigation Skills Seen in Humpback Whales , Live Science (website)
* Woods Hole Oceanographic Institution 2012, ‘Understanding of hearing in baleen whales amplified: New information about how baleen whales use sound ’, Science Daily 17 April 2012

Assessment

Negotiate assessment with the class. Use the following marking guide as a starting point.

Discuss how teams may use the negotiated marking guide as a self-evaluation tool.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experimental Design and Data Analysis | 1 | 2 | 3 | 4 |
| Research into humpback whales (ICT Capability) | I didn’t use any references.  | I viewed up to three references and recorded brief notes.  | I viewed up to three references and recorded notes in a table to organise the information.  | I viewed more than three references and recorded detailed notes in a table to organise the information.  |
| Whale dataset and using a spreadsheet(Digital Technologies) | I didn’t use the whale dataset. | I sorted the whale dataset to help answer a question. | I created a chart to present the data visually. | I was able to draw conclusions from the data and create information used in the experimental design.  |
| Science (Scientific Inquiry Skills) | I did not use scientific ideas.  | I designed a new experiment that doesn’t collect the data needed. | I designed a new experiment that collects some of the data needed. | I designed a new experiment that collects most of the data needed. |
| Cross-curriculum capabilities (Sustainability) | I did not consider sustainability in my experimental design. | I considered sustainability or ethics but not both in my experimental design. | I considered both sustainability and ethics in my experimental design. | I adapted my experimental design to be more ethical and to collect multiple types of data needed to assess the sustainability of development in Pender Bay. |

Australian Curriculum alignment

## Digital Technologies

Years 7–8

* Acquire, store and validate data from a range of sources using software, including spreadsheets and databases (AC9TDI8P01)
* Analyse and visualise data using a range of software, including spreadsheets and databases, to draw conclusions and make predictions by identifying trends (AC9TDI8P02)

## Science

Years 7–8

* Plan and conduct reproducible investigations to answer questions and test hypotheses, including identifying variables and assumptions and, as appropriate, recognising and managing risks, considering ethical issues and recognising key considerations regarding heritage sites and artefacts on Country/Place (AC9S7I02 ) / (AC9S8I02 )
* Examine how proposed scientific responses to contemporary issues may impact on society and explore ethical, environmental, social and economic considerations (AC9S7H03)/ (AC9S8H03 )
* Explain how new evidence or different perspectives can lead to changes in scientific knowledge (AC9S7H01 ) / (AC9S8H01 )
* Analyse data and information to describe patterns, trends and relationships and identify anomalies (AC9S7I05 ) / (AC9S8I05 )
* Construct evidence-based arguments to support conclusions or evaluate claims and consider any ethical issues and cultural protocols associated with using or citing secondary data or information (AC9S7I07 ) / (AC9S8I07 )
* Analyse methods, conclusions and claims for assumptions, possible sources of error, conflicting evidence and unanswered questions (AC9S7I06 ) / (AC9S8I06 )
* Write and create texts to communicate ideas, findings and arguments for specific purposes and audiences, including selection of appropriate language and text features, using digital tools as appropriate (AC9S7I08 ) / (AC9S8I08 )

Mathematics

Years 7–8

* Describe relationships between variables represented in graphs of functions from authentic data (AC9M7A04 )
* Acquire data sets for discrete and continuous numerical variables and calculate the range, median, mean and mode; make and justify decisions about which measures of central tendency provide useful insights into the nature of the distribution of data (AC9M7ST01 )
* Investigate techniques for data collection including census, sampling, experiment and observation, and explain the practicalities and implications of obtaining data through these techniques (AC9M8ST01 )