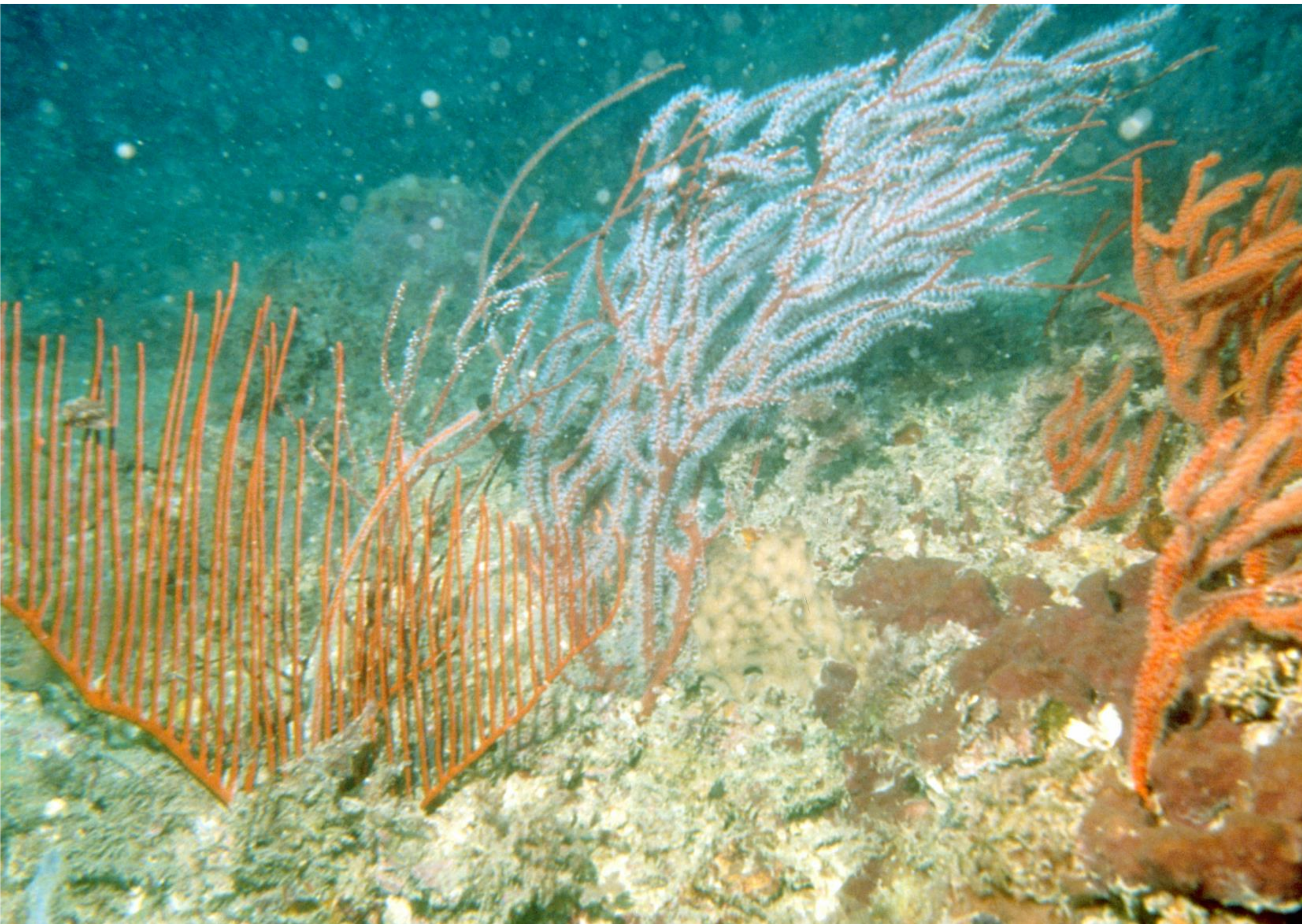
Australia's National Science Agency

Great Barrier Reef Carbon Dioxide Measurements

Educational Datasets Teachers Guide

Year 7-10



Contents

1 Dataset Overview

CSIRO Educational Datasets

CSIRO Educational datasets have been derived from CSIRO research data and adapted for classroom use. They are delivered in three different levels; Novice, Expert and Programmer.

Novice level data has been simplified for the classroom. Potentially confusing outliers and partial entries have been removed from the data for the novice level, to make analysis and comprehension easier. Data labels have been modified to make them easier to understand.

Expert level data retains outliers and partial entries and has not always had the labels adjusted. This may mean that students are required to undertake research into subject language to fully understand what they are looking at. Both Novice and Expert level data contains a limited number of rows to ensure that they can be opened in spreadsheet packages.

Programmer level datasets and activities are intended to be used with more advanced tools and programming languages. This level provides the data in an unmodified format, allowing students to organise and analyse it independently.

Dataset Description

This dataset collects data from the Future Reef MAP project, which is a partnership between Rio Tinto, CSIRO and the Great Barrier Reef Foundation. This project focuses on ocean acidification and uses a custom designed sensor to collect ocean chemistry data along the length of the Great Barrier Reef. The sensor is attached to the Rio Tinto vessel, RTM Wakmatha, which regularly travels the length of the reef, transporting bauxite from Weipa to Gladstone.

The dataset contains key indicators of water health, including absorbed carbon dioxide, temperature, and salinity. In addition to these values, the dataset also contains data to provide context for these measurements, including the latitude and longitude, date and time, atmospheric carbon dioxide, the water flow through the equipment, wind speed and direction.

The recordings in this dataset span a period of 5 years, and are taken from a moving ship, giving researchers a view of the reef's water overall and indicating variations in water quality across the entire reef using only a single recording station. This means, however, that consistency in specific locations cannot be maintained as could be achieved with a single stationary recording station.

The project is ongoing, and as such, more data can be expected to be generated that is not supplied as part of this dataset.

For a link to the original data in the CSIRO Data Access Portal, see Appendix A

Understanding this Dataset

This section relates to understanding this specific dataset. For more general information on understanding and interpreting datasets, see the Educational Datasets Companion document.

	A	B	C	D	E	F	G
1	Group/ship	CruiseID	JD_GMT	date	time	lat	long
2				yyyymmdd	hh:mm:ss	decimal	decimal
3	CSIRO/RTM Wakmatha	WK201503S	77.2416	20150318	5:47:57	-12.671	141.5329
4	CSIRO/RTM Wakmatha	WK201503S	77.2426	20150318	5:49:19	-12.6671	141.5315
5	CSIRO/RTM Wakmatha	WK201503S	77.2435	20150318	5:50:39	-12.6631	141.5301
6	CSIRO/RTM Wakmatha	WK201503S	77.2445	20150318	5:52:01	-12.6592	141.5288
7	CSIRO/RTM Wakmatha	WK201503S	77.2454	20150318	5:53:21	-12.6552	141.5273
8	CSIRO/RTM Wakmatha	WK201503S	77.2463	20150318	5:54:42	-12.6513	141.526
9	CSIRO/RTM Wakmatha	WK201503S	77.2473	20150318	5:56:04	-12.6474	141.5245
10	CSIRO/RTM Wakmatha	WK201503S	77.2494	20150318	5:59:07	-12.6386	141.5213

Figure 1 - This sample from the expert dataset indicates that on the 18th of March 2015, at 5:47:57am GMT (3:47:57pm local time), RTM Wakmatha was at -12.671 latitude, 141.5329 longitude, placing it west of Weipa in Queensland.

The first group of columns describe the readings, outlining the ship they were taken from, the cruise identifier, time, date and location. The group/ship column describes the organisation responsible for the readings and the ship that the readings have been taken from. In the case of this dataset, all the entries have been generated by the CSIRO aboard the RTM Wakmatha.

The CruiseID column details the journey that the ship was on when the readings were taken. Following the initial two characters, the next four digits indicate the year that the readings were taken, followed by two digits that indicate the month. Finally, the CruiseID has a character that indicates the direction of travel. So, for example WK201302S indicates that the reading was taken on a cruise in 2013, in February and the ship was traveling southwards. Similarly, the voyage with a CruiseID of WK201302N was the northbound cruise that took place during Feburary 2013.

It is important to note that time and date are given in UTC (Universal Coordinated Time), which is not the local time of the ship. The date is given in the date column using the format yyyymmdd, with the time given in the time column in the format hh:mm:ss. The time and date is also expressed in the JD_GMT column, in the format ddd.hhhh, where day is represented using the number of the day, counting from the start of the year. The 1st of January will be represented as 1, and the 18th of February will be represented as 049. Hour is in a decimal format, with values from 0000 to 9999. This makes for an interesting departure from more recognisable timestamps and invites potential for comparing the two methods of recording time.

Latitude and longitude are given in decimal form, rather than degrees and minutes. The numeral value is the number of degrees, while partial degrees are recorded using a decimal value.

	H	I	J	K	L
1	xCO2_equilibrator	xCO2_atmospheric	xCO2_atm_interpolated	equilibrator_pressure	pressure_atmospheric
2	ppm	ppm	ppm	hPa	hPa
3	425.09	-999	396.4	1007.36	1008.5
4	431.57	-999	396.4	1007.35	1008.5
5	428.27	-999	396.4	1007.37	1008.5
6	428.55	-999	396.41	1007.38	1008.5
7	426.1	-999	396.41	1007.42	1008.4
8	429.93	-999	396.41	1007.38	1008.4
9	430.41	-999	396.41	1007.35	1008.5
10	429.03	-999	396.42	1007.36	1008.3

Figure 2 - This sample from the expert dataset indicates that the amount of CO_2 detected in the equilibrator is 425.09 parts per million, as compared to the interpolated atmospheric value of 396.4. The atmospheric CO_2 value was not recorded in any of the records here, as atmospheric values are taken as a separate reading, so the interpolated value is the only comparison that can be made.

Carbon dioxide measurements are given here as a mole fraction, in parts per million. The equilibrator column indicates the mole fraction in the equilibrator head space, whereas the atmospheric column indicates the mole fraction in the atmosphere. The atmospheric column also includes the placeholder value of -999. In this dataset, -999 indicates that a variable has not been recorded as part of this reading. In the data above, no records for atmospheric CO_2 were made. Atmospheric CO_2 records were made independently of equilibrator CO_2 records every 3-4 hours, so no rows contain data for both xCO2_equilibrator and xCO2_atmospheric.

The third column takes the values for atmospheric CO_2 and uses linear progression to estimate what those values would be between readings. In the novice dataset, these interpolated values are invaluable, as not all the readings that were used to generate them are included in the dataset.

The last two columns here indicate the pressure inside the equilibrator, and the pressure in the atmosphere.

	M	N	O	P	Q	R
1	equilibrator_temp	sea_surface_temp	salinity	fCO2SW_uatm	fCO2ATM_uatm_interpolated	DFCO2_uatm
2	oC	oC	psu	uatm	uatm	uatm
3	29.46	29.275	32.027	401.8	377.75	24.05
4	29.5	29.272	31.648	407.26	377.76	29.5
5	29.51	29.287	31.867	404.22	377.74	26.49
6	29.6	29.314	31.63	403.51	377.71	25.8
7	29.58	29.313	31.934	401.51	377.7	23.8
8	29.55	29.275	31.725	405.01	377.72	27.3
9	29.5	29.236	31.733	405.66	377.79	27.86
10	29.5	29.202	31.759	403.87	377.75	26.11

Figure 3 - This sample from the expert dataset indicates the temperatures in the equilibrator, surface temperatures of the sea and salinity. The fugacity of CO_2 in sea water under the recorded conditions and an interpolated value for the fugacity of CO_2 in the atmosphere are given here, with a difference between the two calculated in the last column. These records all indicate that CO_2 is more likely to move from the ocean to the atmosphere, with the second record indicating that tendency more than the other records here.

The first three columns of this section are relatively straightforward and indicate the temperature inside the equilibrator, the temperature of the surface of the sea and the salinity of the water.

The values in columns P, Q and R describe the fugacity of CO_2, which refers to its tendency to seek equilibrium in a non-ideal situation and is measured in microatmospheres (µatm). The higher this value, the more likely CO_2 is to escape. fCO2SW_uatm measures fugacity of CO_2 in surface water at the current salinity and temperature, assuming 100% humidity. fCO2ATM_uatm_interpolated measures fugacity of CO_2 in the atmosphere under the same conditions and has been estimated using interpolation. DFCO2_uatm is the calculated difference between these two values. A positive DFCO2_uatm indicates that the ocean is likely to release CO_2 into the atmosphere and a negative value indicates that the ocean is likely to absorb CO_2 from the atmosphere.

	S	T	U	V	W	X	Y
1	LICOR_flow	H2O_flow	wind_speed_true	wind_direction_true	Type	WOCE_QC_FLAG	SUBFLAG
2	mL/min	litres/min	m/s	degrees			
3	103.25	2.26	14	144.3	EQU	2	-999
4	103.31	2.26	14.5	142.5	EQU	2	-999
5	102.53	2.26	15.8	138	EQU	2	-999
6	102.4	2.26	15.5	153	EQU	2	-999
7	103.74	2.26	12.9	141.2	EQU	2	-999
8	102.34	2.26	16	146.8	EQU	2	-999
9	103.55	2.32	11.4	148.7	EQU	2	-999
10	103.01	2.25	14.8	139.3	EQU	2	-999

Figure 4 - This sample from the expert dataset indicates equipment calibration information, such as gas flow rate through the sensor, water flow rate through the equilibrator, the type of reading being recorded and a quality control flag. All of the records here are Equilibrator records, meaning that they have recorded information about the water conditions, including CO_2 saturation.

The first column in this section measures the gas flow through the infrared sensor. Depending on the type of reading, this could be the flow from the equilibrator, the atmosphere, or the standard reference gases. The second column is the flow of seawater through the equilibrator. These measurements are recorded to ensure data quality is maintained.

Another important aspect in modelling the interaction between oceanic CO_2 and atmospheric CO_2 is recording wind conditions. Wind speed is recorded here in meters per second and wind direction is recorded in degrees, with 0 representing North, and 90 representing East.

There are several flags at the end of this dataset, specifically TYPE and WOCE_QC_FLAG. The Type column indicates whether the reading is an equilibrator head space reading (EQU), standard gas readings (STD) or atmospheric reading (ATM). The STD readings are readings of a standard reference gas, which are taken every 3-4 hours to calibrate the equipment. The WOCE_QC_FLAG is used for quality control and has two potential values. In this column, a 2 indicates that the data is reliable, while a 3 indicates that it is questionable.

It is also important to note that the equipment making these recordings is located on a moving ship, so while we can expect the seasons and weather changes to impact the data, that the tracking station moves over time is an additional consideration.

Research Findings

This data is used to create a better understanding of current conditions along the Great Barrier Reef, as well as being used to develop models and tools to help understand how we can expect conditions along the Reef to change in the future.

To date, the project has demonstrated that the ocean chemistry across the Great Barrier Reef remains positive for coral growth and provides an environment for coral to recover from events such as bleaching and cyclones. Ocean chemistry on the Reef was shown to vary seasonally, with summer providing the best conditions for coral growth. Additionally, currents from the Coral Sea were demonstrated to have a higher influence on the ocean chemistry of the Reef than outflow from coastal river systems.

For more information about this research, see Appendix A.

Learning Goals

As with any lesson resources, there are any number of ways this dataset could be brought into the classroom, depending on your approach and personal style. Here you'll find some potential overarching learning goals, most of which address general data literacy, understanding and representation to guide you in introducing this dataset to your students.

Understanding this dataset

Students examine simple ways of exploring datasets to understand them and discuss the positives and negatives of using a specific dataset. In achieving this learning goal, some activities might include:

- Mean vs Median. When taking the mean and the median of a dataset, it's possible to get two different results. What does this mean? Why are they different? Which one is a better indicator of the centre of the dataset? In this case, which value is more useful to us?
- Graphing. What kinds of graphs can we use to represent this data? Are there any subsets of the data that might be useful to compare on a graph?
- Interpolation. Can we trust data that has been created by observing patterns in existing data and estimating? What are some of the methods we can use to estimate an unknown variable, and how reliable are they?

Accurately report findings made from data

Students examine how to best represent their findings from the dataset. How can we display this data so that humans can easily read and understand it? Representing the whole dataset in a single table can make it difficult to identify trends and link related concepts. Using statistical tools, such as using the average, range, median, mode or percentages can help give the audience a better idea of what the data tells us, but some of these values are more useful than others, depending on context. If you're packing for a trip, the range of temperatures for each day is more important than the median temperature for the whole trip. Knowing that the temperature will get as high as 27°C and as low as -2°C is more important than knowing that the median temperature will be 13°C, as it gives you a much better idea of what to pack.

With this dataset, consider if it is useful to display the latitude and longitude records on a line graph. Since this data refers to coordinate data and references a location on the planet, displaying it numerically on a line graph doesn't make it particularly meaningful. It is important to consider the purpose of a visualisation, in terms of the story it presents the viewer.

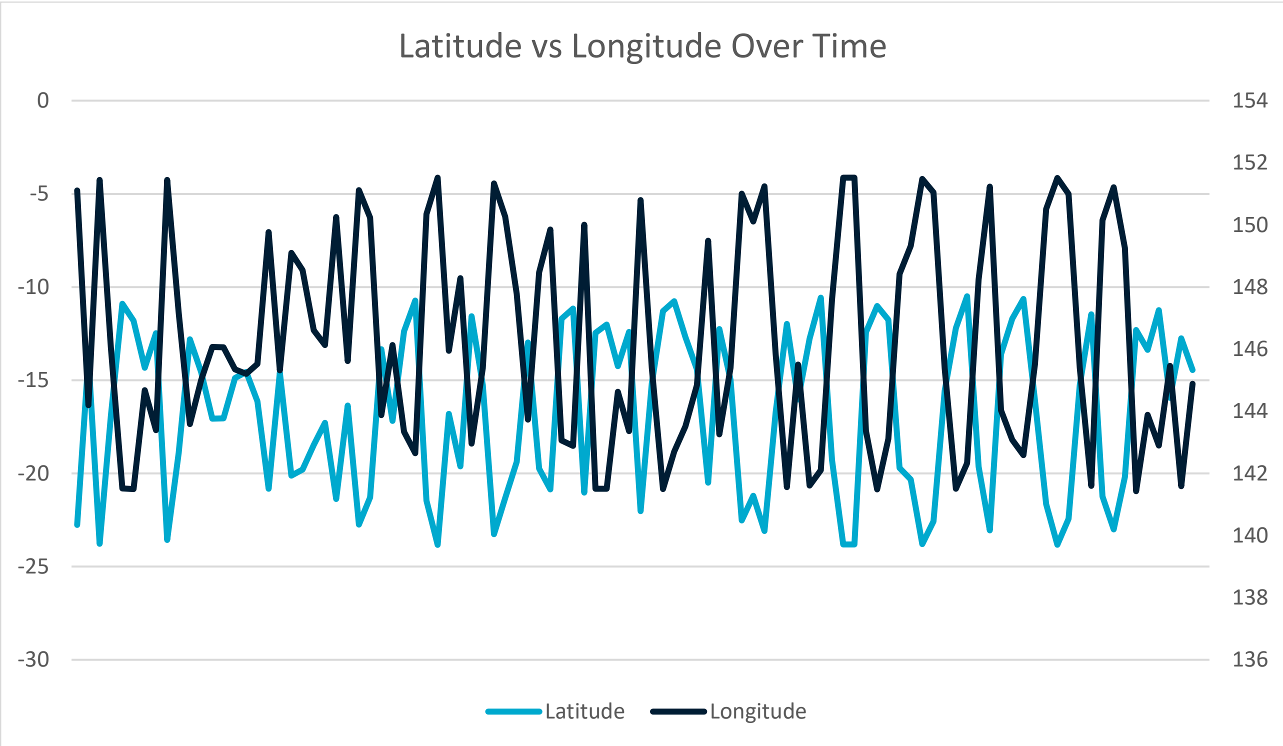


Figure 5 – Line graph comparing longitude values with latitude values over time. This does not help the viewer get a sense of the location of the ship at any given point in time. This would be better plotted as locations on a map.

What else can we do to make sure that the findings we're reporting don't skew the data or misrepresent it? Examine ideas such as considering the whole data set, not just favourable sections of it, or ensuring that when using two graphs to compare data, they use the same start and end points, the same scales, and the same display ranges.

Alternately, you could reverse this lesson idea and ask students to find a way to misrepresent the dataset to distort the truth, without lying about the data. How could you display it so that it looks like it's saying something that it isn't? This could generate discussion about misuse of data in the media and advertising, or ethics in scientific research, and the importance of representing data accurately.

Understanding Outliers

Outliers in data refer to things that fall well outside of the other values observed. They can be legitimate variations in the thing you're measuring, or can be measurement errors, where the reading was not taken correctly for a variety of reasons.

With this lesson goal, students examine the dataset, identifying the average, median and range. Once that is done, students can then identify any outliers, using methods like looking for sharp spikes in readings, or surprisingly high or low values, remove them from the dataset and recalculate their average, median and range, compare the values, and discuss which ones have changed and why, as well as discussing where the outlier values may have come from.

For more information on Outliers, see the Educational Datasets Companion document.

Identifying the Right Visualisation

As the idiom goes, a picture is worth a thousand words and there are lots of ways we can take data and make it visual. Some of the more common methods of creating visualisations are pie charts, line graphs and bar graphs. Depending on the data set, other visualisations may be appropriate to give the audience a better illustration of the data and the trends and patterns it contains.

For this dataset, since it consists of measurements taken aboard a travelling ship, maps can help give a better sense of exactly where each reading was taken. For numerical data, most graphs are acceptable, especially line graphs and scatter plots, as they can indicate overall trends over time.

It's important to remember that while students can generate visualisations for data using digital resources, there's also the opportunity with smaller datasets to create these visualisations by hand, using printed maps or sketches.

For more examples of data visualisations, see the Educational Datasets Companion document.

Spreadsheet and Numeric Skills

Spreadsheets and numerical skills are embedded across the curriculum, and this data offers an opportunity for students to put their skills to work on real-world scientific measurements. While a wide range of mathematical skills and spreadsheet skills can be applied, some key examples are:

- Developing spreadsheet formulae. Look for places in the data that an automatically calculated total or average might be useful. We might need to check the average pollutant level across the set time periods addressed in the standards. Formulas could be used to determine the minimum and maximum value of a list of data, to identify changes in surface temperature, or the difference between each day's minimum and maximum values, both in numerical form and as a percentage.
- Graphing. Consider the different types of charts that your spreadsheet software can make. How can we modify the settings of a graph to display data appropriately? What is an appropriate title? What are labels and value ranges should be used for its axes? Students could construct graphs on paper, to build manual graphing skills.

- Conditional formatting. Create a set of rules so that when the salinity or temperature reaches a certain value it is highlighted in the data.
- What-if calculations. Students can use the real-world data to make predictions. If one of the recorded values for xCO2_equilibrator changed by 10%, how would that impact the interpolated values? What other values would that impact?
- Non-digital numerical skills. Students can manually take averages of sets of readings, examine other statistical quantifiers such as median and range or identify the standard deviation.

Programming

Many of the files in this dataset can be opened and manipulated in a variety of programming languages. CSV files are very easy for most programming languages to work with, since they are simple text files which use commas to split data points. Python has a specific module (csv) that adds additional functionality when working with these files.

Teaching programming with this dataset gives students an opportunity to practice skills relating to reading and writing data to and from files directly and incorporates string manipulation so they can directly access specific pieces of data. Students can investigate data structures such as lists, dictionaries and objects, assessing their usefulness in storing this data, and utilise control structures to perform calculations on the data, or organise it in a manner appropriate for output.

Depending on the prior understanding students have of programming principles, this can lead to activities ranging from calculating averages automatically and outputting them to the screen, to searching for potential outliers and removing them from the dataset before outputting it as a separate file, to creating interactive visualisation tools for the dataset.

Subject Links

This dataset can be linked to the Australian curriculum learning areas of Mathematics, Science (Earth Sciences and Environmental Science), Technologies (Digital Technologies), and Humanities and Social Sciences (Geography.)

For a more detailed list of subject links, content descriptions and year levels, see Appendix B.

2 Lesson Materials

Required Understanding

A list of the existing skills students will require to work effectively with each level of this dataset can be found in the table below. This dataset can also be used as a tool to develop these skills.

The Novice data collects the records from a single pair of voyages (one southbound, one northbound.) Activities listed for use with the Novice data can be achieved with the Expert data.

SPREADSHEET NOVICE	SPREADSHEET EXPERT	PROGRAMMER
• Spreadsheet software and the relevant key terminology, such as cell, row, column, sheet, data, cell reference and cell range • Developing spreadsheet formulas • Creating charts in spreadsheet software packages • Basic mathematical statistical concepts, such as averages, range and median values.	• Spreadsheet software, including appropriate formatting skills and relevant key terminology, such as cell, row, column, sheet, data, cell reference and cell range. • Developing spreadsheet formulas • Creating charts in spreadsheet software packages • Basic statistical concepts, such as averages, range and median values.	• Basic understanding of commands for a specific programming language • Understanding of data structures and file input/output • Understanding of programming control structures, such as sequence, selection and repetition • Basic statistical concepts, such as averages, range and median values

Content Engager

Use these resources to introduce the importance of measuring ocean acidification.

- CSIRO – Monitoring Ocean Acidification on the Great Barrier Reef
- YouTube – Future Reef
- Vimeo – Rio Tinto Alcan – Future Reef MAP
- SBS News – Coral reefs may dissolve in acidic oceans
- ABC News – Acidic oceans could slow coral reef growth by a third

Some questions that you can use to start discussion about this topic and activate students' prior knowledge include:

- Have you heard of ocean acidification?
- Why is it important to maintain coral growth on a reef?
- What could happen to a reef ecosystem if the coral dies off completely?
- Where does carbon dioxide in the oceans come from?
- Is it normal to have carbon dioxide in the ocean? If so, how much is normal?
- Why is the Great Barrier Reef of significance to Australians?
- Why is the Great Barrier Reef of significance to the world?

Introductory Description

To introduce students to this dataset, consider reading the following paragraph to them, or something similar.

‘Today we’re going to be looking at data gathered as part of the Future Reef project, which is a partnership between CSIRO, the Great Barrier Reef Foundation and Rio Tinto. The project aims to get a broader picture of the health of the reef, with a focus on oceanic acidification. Our oceans are particularly good at absorbing Carbon Dioxide from the atmosphere, with an estimated 25% of human carbon dioxide output being absorbed by our oceans every year[1]. But beyond getting Carbon Dioxide out of our atmosphere, this absorption increases the acidity level of the water which limits the ability of key organisms such as coral to develop their shells and skeletons.’

[1] – Future Reef – Great Barrier Reef Foundation

Thinking Time

Once students have an idea of the dataset’s content, give them 5 minutes to brainstorm questions they’d like to try and answer using this data. Try not to lead students too much during this time. There is a high chance that students will develop questions which cannot be answered by the data. This creates an opportunity to explore why those questions cannot be answered.

Activities

Spreadsheet Novice

1. Determine the minimum and maximum latitude and longitude at which records were made.

   This can be achieved by sorting the latitude column and the longitude column. Using the minimum and maximum formulae in a spreadsheet would achieve this without needing to sort the columns repeatedly.

2. In this dataset, a value of -999 is used as a placeholder for a record that does not exist in a given reading. Why is -999 used instead of 0? What does that mean when we try to graph this data? How can we take that into account when graphing the data?

   The graph is dramatically skewed as it includes -999s as a data point. Using a 0 as a placeholder might result in the 0 being mistaken for a valid record, whereas there should be no way that -999 is a true data point in this data set. Removing these readings completely would make the graphs easier to interpret.

3. Determine the range of CO_2 concentration across all readings. What is the percentage difference between the minimum reading and the maximum reading? What does the percentage difference mean?

   The base level of CO_2 is high, so more CO_2 does not change the percentage much. The absorption represented by the percentage change is significant, but the percentage change makes it look very low. Examining the change as a raw value can give more information in cases like this.

4. Determine the median value and quartile boundaries for CO_2 fugacity in sea water. Are the quartiles evenly distributed? Where does the median sit inside the range? How does the interquartile range compare to the range?

   The interquartile range is approximately 34 for this column (fCO2SW_uatm). The median sits slightly closer to the lower quartile and almost the centre of the overall range. The interquartile range is approximately a quarter of the total range. This tells us that while there is a large range, most of the results are tightly clustered and the large range is due to a few low readings (notably an outlier of 322.81).

5. These readings were taken by equipment that is located on a moving ship. Are they more valuable or less valuable than readings taken by equipment that records in a single location? Why?

   There can be a lot of answers here. Positive responses could include that using moving equipment allows us to cover a larger section of the reef with one piece of equipment and allows us to see how readings change by location, which can help identify areas that are most at risk. Negative responses could include that with a single location we could better track changes over time without having to account for differing locations.

Spreadsheet Expert

1. Looking at the latitude and longitude at the start and end of each voyage, use a map to determine which ports the ship travels between. Using the internet, research those ports and explain why the ship travels this route, and what its main cargo is.

   The ship travels between Weipa and Gladstone in Queensland. Weipa is home to a major bauxite mine. There are facilities in Gladstone capable of refining bauxite into alumina and aluminium metal. The ship transports bauxite ore from the mine at Weipa to Gladstone for refinement.

2. The date format here is one that is difficult to read at a glance. Using appropriate formulas for your spreadsheet package, convert this format to a format that is easier to read.

   In most spreadsheet packages, the =date function can be used. This function will allow you to specify where the year, month and day can be found. If the date is stored using text, as it is here, tools such as left, mid and right will splitting the characters into groups and allow them to be entered into the date formula.

3. Looking at all the available data, what is the average sea surface temperature for each season? Which season has the highest average temperature? Which one has the lowest? Is that what you expected? What impact does the temperature have on ocean currents?

   This is harder than it appears at first glance. It requires either complex formulas to extract the month from the date and using an averageif formula to average only the values that meet the criteria or using a pivot table to group the data. Summer is the highest average temperature, with winter as the lowest. The relationship between temperature and ocean currents is not obviously discernible from this data and requires additional research.

4. Examine the salinity readings across each year. When are the high/low points? Do we have enough data to identify a clear seasonal trend in salinity? What might cause salinity to change over the course of a year?

   The salinity steadily rises from March to around November and falls again after that point. The lack of months we have available for analysis is very clear here. We do not have enough data to identify a clear trend. Salinity may change over the course of the year due to currents, environmental factors, and movement of the ship.

5. Use this data to establish a graph of sea temperature averages over the course of the readings. Is this graph accurate enough to make trend conclusions from regarding time? What might need to be changed about the graph to make the data more accurate? How can we use interpolation here to get a better picture? Will using interpolation give a clearer picture, or will some aspects of the data still be unaccounted for?

   Data is missing for entire months between trips, which includes December and February towards the end of the recordings. Missing an entire summer month of recordings near the end of the data will impact the average dramatically. Taking a linear interpolation might give us a good idea of the missing data if only a small amount is missing, but if we're missing the peak winter or summer month, a lot of high or low values would be ignored by simply connecting the two points we have.

6. Examine the maximum and minimum values for CO_2 fugacity in sea water over a single day. Considering this range, graph this value using a range with a minimum value of 0. What information can we get from this graph? What range would be most appropriate for graphing this data?

   Using the 19th of March 2015 as an example, the maximum and minimum values are 456.29 and 402.44 (There is also a -999 which will need to be accounted for here.). Graphing this day with a range starting at 0 will make the pattern difficult to read as the data will take up 10% of the total height of the graph, with 90% being blank space. Using a range of around 400-460 would be appropriate, as it makes the finer detail more apparent.

7. Establish the mean and median for each variable. How do these values compare? Where do they sit within the range? How useful are these measures when assessing each variable?

   Mean and median sit very close for most variables in this data set. They generally sit in the middle of the range, though they sit high in the range for temperature and low in the range for fugacity of CO_2 in surface water. Some of the variables in this data set, such as wind speed, wind direction, longitude and latitude are not well suited to assessing them in this manner.

8. Create a pivot table in a separate sheet, grouping the readings as an average per hour, with columns for each day of recordings. Graph these averages, representing each day using a different data series. What do you notice about the general trend?

   Depending on the variable being graphed, these trends can vary. Examining fugacity of CO_2 in surface water, the trend for a single day tends to be minimal variance across the day. There is no specific time-related trend, and when values rise and fall, they often take some time to do so rather than presenting as sharp spikes.

9. Add another series to your graph to represent another day of readings. Can you see a specific trend across both days? Is there a trend that becomes noticeable with ten days of readings on a single graph?

   This depends on the variable being graphed. In the case of fugacity of CO2, there should not be any major noticeable trends beyond days tending to remain relatively stable around initial values, despite being at different levels of initial readings.

10. Plot five of the sets of coordinates on a Cartesian plane. Use the distance formula to work out the distance between each set of points. With a tool like Google Maps, check the distance between those points in the real world and create a formula that can estimate real world distance from coordinates, using your examples. Test your formula on two random coordinates. Does it work everywhere in the world? Why?

Using the distance formula and converting it to reflect kilometres will work for distances similar to the one that was used to develop the formula, but as it starts to get used on varied distances, it will start to reveal flaws in the method. This is because the basic distance formula does not account for Earth's curvature.

11. Using a map (either a physical map, or a digital tool such as nationalmap.gov.au) and the latitude and longitude data, plot the locations that data was collected on a single voyage.

    – Plot the journey using a colour scale to indicate CO_2 fugacity at each collection point.

Simplifying the dataset and creating a file with just the data that is needed for mapping will make the use of digital tools much easier.

12. Using these data points and a map, plot the course of the ship as accurately as possible.

    – Plot all the available voyages, using a different colour for each journey

Simplifying the dataset and creating a file that only contains latitude, longitude and a voyage identifier will make this much easier to achieve.

13. Coral bleaching events can be triggered by several conditions, but the most common trigger is increased ocean temperatures. In summer of 2016, a significant bleaching event was noted on the Great Barrier Reef. Compare temperatures during March of 2016 to recordings made in March of other years. Is there a notable difference? Do we have enough data in this dataset to link temperature to this event?

    – The most severe bleaching of this event was between Cape York and Port Douglas. Isolate only the readings taken between these points

The 2016 March sea surface temperature average is much higher than 2015 but is not much higher than 2017. The maximum sea temperature in 2016's March, however, is roughly 2 degrees higher than 2015 and 1 degree higher than in 2017. Given that this dataset is missing the months during which the event occurred, we do not have enough data in this dataset to link temperature. The latitude coordinate for Port Douglas is -16.4834 and Cape York is -10.7000. Restricting the data to being between these two values will give a notably higher temperature in 2016. Though it should be noted, that some of those recordings were made west of Cape York, which is not considered to be part of the reef and should be accounted for in analysis. There are additional files included as part of the Programmer package which can be added to the Expert data for further analysis

14. With the understanding that high surface temperatures and high salinity are linked to coral bleaching events, use this data to identify areas on the reef that are potentially vulnerable to coral damage.

This is a matter of identifying the coordinates linked to high salinity levels and high temperatures. It's important to note that of the 50 highest salinity recordings, 38 of them were recorded on the same day which would indicate a stronger link between time and salinity than location, but with that said, the other recordings from different dates in this group were all in within 70km of each other. Generating a heatmap using these aspects of the data would be a good way to examine trends.

Programmer

Write a program using your chosen programming language to perform the following tasks:

1. Store the data in an appropriate data structure.
2. Convert the windspeed from knots into metres per second.
3. Convert the date format in the original files to a more human readable format. Convert the month to a written month and add the appropriate suffix to the date number so that the date appears in the format '18th March, 2015'. Output the file with the modified date.
4. Using coordinates in latitude and longitude as input, estimate what the DFCO2_uatm would be at that point based on the available data. Add the ability to input a date to weight voyages near that date more heavily than those at different times.
5. Graph the data for your predictions, along with the historical data.
6. Take random samples from the file, retaining 75%, 50% and 25%, creating a new file for each. Compare the maximum and minimum recorded DFCO2_uatm values for the new files to the original. Map one of the voyages using the original data. Map the same voyage using the sampled files. Compare the maps. How accurate are the sampled files?

Open Inquiry

In addition to the activities listed above, this dataset can be used for student-centred open inquiry projects. Using open inquiry, students generate research questions and design investigations to answer those questions. Students can use this dataset to support their independent research and investigation in a range of areas. There are additional readings from this project available on the CSIRO Data Access Portal

Examples of inquiry questions that could be explored using this data include:

- What impact do large cities have on the water quality of the surrounding reef?
- What impact does seasonal change have on oceanic CO_2 absorption?
- What major currents impact the water quality of the Great Barrier Reef?
- Does wind coming from major cities have a greater impact on oceanic CO_2 absorption than wind coming from the ocean?

Assessment

Assessment items for this dataset could include:

- A map indicating the path of the ship and visualising the data generated along one voyage.
- A poster explaining oceanic acidification and the factors that contribute to it.
- A report outlining the areas of the reef most at risk of coral degradation due to oceanic acidification.
- Code to estimate values of CO2 fugacity, given the recorded values.
- Heatmaps indicating oceanic temperatures by location.

Appendix A References

Educational Dataset:	Great Barrier Reef Alkalinity and Carbon Dioxide Measurements
Original Dataset:	Underway fCO2 measurements collected onboard the RTM Wakmatha on the Great Barrier Reef, Australia 2012 • December (Northbound), December (Southbound) 2013 • February (Northbound), February (Southbound), March (Northbound), March (Southbound) 2014 • April, May, June, July, August, September, October, December 2015 • January, March, May 2016 • March, May, July, September, November 2017 • January, March, May, June
Published Papers:	Ocean Acidification: Ecological Response Analysis, and cross-platform comparison of the available carbonate chemistry data for the Great Barrier Reef
Supporting Information:	CSIRO – Monitoring ocean acidification on the Great Barrier Reef Great Barrier Reef Foundation – Ship of Opportunity Rio Tinto – One of our ships is monitoring the Great Barrier Reef

Appendix B Australian Curriculum Guide

STRAND	YEAR 7	YEAR 8	YEAR 9	YEAR 10
Mathematics				
Number and Algebra	Given coordinates, plot points on the Cartesian plane, and find coordinates for a given point (ACMNA178) Investigate, interpret and analyse graphs from authentic data (ACMNA180)		Find the distance between two points located on the Cartesian Plane using a range of strategies, including graphing software. (ACMNA214) Graph simple non-linear relations with and without the use of digital technologies and solve simple related equations (ACMNA296)	Substitute values into formulas to determine an unknown (ACMNA234) Solve problems involving linear equations, including those derived from formulas (ACMNA236)
Statistics and Probability	Identify and investigate issues involving numerical data collected from primary and secondary sources (ACMSP169) Construct and compare a range of data displays including stem-and-leaf plots and dot plots (ACMSP170) Calculate mean, median, mode and range for sets of data. Interpret these statistics in the context of data (ACMSP171) Describe and interpret data displays using median, mean and range (ACMSP172)	Investigate techniques for collecting data, including census, sampling and observation (ACMSP284) Explore the practicalities and implications of obtaining data through sampling using a variety of investigative processes (ACMSP206) Explore the variation of means and proportions of random samples drawn from the same population (ACMSP293) Investigate the effect of individual data values, including outliers, on the mean and median (ACMSP207)	Compare data displays using mean, median and range to describe and interpret numerical data sets in terms of location (centre) and spread (ACMSP283)	Determine quartiles and interquartile range (ACMSP248) Use scatter plots to investigate and comment on relationships between two numerical variables (ACMSP251) Investigate and describe bivariate numerical data where the independent variable is time (ACMSP252)

<table>
<tr><th>STRAND</th><th>YEAR 7</th><th>YEAR 8</th><th>YEAR 9</th><th>YEAR 10</th></tr>
<tr><td colspan="5">Science</td></tr>
<tr><td>Science Understanding</td><td>Mixtures, including solutions, contain a combination of pure substances that can be separated using a range of techniques (ACSSU113)</td><td></td><td>Ecosystems consist of communities of interdependent organisms and abiotic components of the environment; matter and energy flow through these systems (ACSSU176)</td><td>Global systems, including the carbon cycle, rely on interactions involving the biosphere, lithosphere, hydrosphere and atmosphere (ACSSU189)</td></tr>
<tr><td>Science as a Human Endeavour</td><td colspan="2">Solutions to contemporary issues that are found using science and technology, may impact on other areas of society and may involve ethical considerations (ACSHE120) & (ACSHE135)
People use science understanding and skills in their occupations and these have influenced the development of practices in areas of human activity. (ACHSE121) & (ACSHE136)</td><td colspan="2">Values and needs of contemporary society can influence the focus of scientific research (ACHSHE228) & (ACSHE230)</td></tr>
<tr><td>Science Inquiry Skills</td><td colspan="2">Identify questions and problems that can be investigated scientifically and make predictions based on scientific knowledge (ACSIS124) & (ACSIS139)
Construct and use a range of representations, including graphs, keys and models to represent and analyse patterns or relationships in data using digital technologies as appropriate (ACSIS129) & (ACSIS144)
Summarise data, from students' own investigations and secondary sources, and use scientific understanding to identify relationships and draw conclusions based on evidence (ACSIS130) & (ACSIS145)
Reflect on scientific investigations including evaluating the quality of the data collected, and identifying improvements (ACSIS131) & (ACSIS146)
Use scientific knowledge and findings from investigations to evaluate claims based on evidence (ACSIS132) & (ACSIS234)
Communicate ideas, findings and evidence based solutions to problems using scientific language, and representations, using digital technologies as appropriate (ACSIS133) & (ACSIS148)</td><td colspan="2">Formulate questions or hypotheses that can be investigated scientifically (ACSIS164) & (ACSIS198)
Analyse patterns and trends in data, including describing relationships between variables and identifying inconsistencies (ACSIS169) & (ACSIS203)
Use knowledge of scientific concepts to draw conclusions that are consistent with evidence (ACSIS170) & (ACSIS204)
Evaluate conclusions, including identifying sources of uncertainty and possible alternative explanations, and describe specific ways to improve the quality of the data (ACSIS171) & (ACSIS205)
Critically analyse the validity of information in primary and secondary sources and evaluate the approaches used to solve problems (ACSIS172) & (ACSIS206)
Communicate scientific ideas and information for a particular purpose, including constructing evidence-based arguments and using appropriate scientific language, conventions and representations (ACSIS174) & (ACSIS208)</td></tr>
<tr><td colspan="5">Technologies: Digital Technologies</td></tr>
<tr><td>Processes and Production Skills</td><td colspan="2">Analyse and visualise data using a range of software to create information, and use structured data to model objects or events (ACTDIP026)</td><td colspan="2">Develop techniques for acquiring, storing and validating quantitative and qualitative data from a range of sources, considering privacy and security requirements (ACTDIP036)
Analyse and visualise data to create information and address complex problems, and model processes, entities and their relationships using structured data (ACTDIP037)</td></tr>
</table>

STRAND	YEAR 7	YEAR 8	YEAR 9	YEAR 10
Geography				
Geographical Knowledge and Understanding	Economic, cultural, spiritual and aesthetic value of water for people, including Aboriginal and Torres Strait Islander Peoples and peoples of the Asia region (ACHGK041) Causes, impacts and responses to an atmospheric or hydrological hazard (ACHGK042)			Human-induced environmental changes that challenge sustainability (ACHGK070) The application of geographical concepts and methods to the management of the environmental change being investigated (ACHGK074) The application of environmental economic and social criteria in evaluating management responses to the change (ACHGK075)
Geographical Inquiry and Skills	Evaluate sources for their reliability and usefulness and select, collect and record relevant geographical data and information, using ethical protocols from appropriate primary and secondary sources (ACHGS048) & (ACHGS056) Represent data in a range of appropriate forms, for example climate graphs, compound column graphs, population pyramids, tables, field sketches and annotated diagrams, with and without the use of digital and spatial technologies (ACHGS049) & (ACHGS057) Represent spatial distribution of different types of geographical phenomena by constructing appropriate maps at different scales that conform to cartographic conventions, using spatial technologies as appropriate (ACHGS050) & (ACHGS058) Interpret geographical data and other information using qualitative and quantitative methods, and digital and spatial technologies as appropriate, to identify and propose explanations for spatial distributions, patterns and trends, and infer relationships. (ACHGS051) & (ACHGS059) Apply geographical concepts to draw conclusions based on the analysis of the data and information collected (ACHGS052) & (ACHGS060) Present findings, arguments and ideas in a range of communication forms selected to suit a particular audience and purpose; using geographical terminology and digital technologies as appropriate (ACHGS053) & (ACHGS061) Reflect on their learning to propose individual and collective action in response to a contemporary geographical challenge, taking account of environmental, economic and social considerations, and predict the expected outcomes of their proposal (ACHGS054) & (ACHGS062)		Evaluate sources for their reliability, bias and usefulness and select, collect, record and organise relevant geographical data and information, using ethical protocols, from a range of appropriate primary and secondary sources (ACHGS064) & (ACHGS073) Represent multi-variable data in a range of appropriate forms, for example scatter plots, tables, field sketches and annotated diagrams, with and without the use of digital and spatial technologies as appropriate (ACHGS065) & (ACHGS074) Represent spatial distribution of geographical phenomena by constructing special purpose maps that conform to cartographic conventions, using spatial technologies as appropriate (ACHGS066) & (ACHGS075) Interpret and analyse multi-variable data and other geographical information using qualitative and quantitative methods, and digital and spatial technologies as appropriate, to make generalisations and inferences, propose explanations for patterns, trends, relationships and anomalies, and predict outcomes (ACHGS067) & (ACHGS076) Apply geographical concepts to synthesise information from various sources and draw conclusions based on the analysis of data and information, taking into account alternative points of view (ACHGS068) & (ACHGS077) Present findings, arguments and explanations in a range of appropriate communication forms, selected for the effectiveness and to suit audience and purpose; using relevant geographical terminology, and digital technologies as appropriate (ACHGS070) & (ACHGS079) Reflect on and evaluate findings of an inquiry to propose individual and collective action in response to a contemporary geographical challenge, taking account of environmental, economic, political and social considerations; and explain the predicted outcomes and consequences of their proposal (ACHGS071) & (ACHGS080)	

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiroenquiries@csiro.au
csiro.au

For further information

CSIRO Education and Outreach
1300 136 376
+61 2 9490 5588
education@csiro.au
csiro.au/education

Cover Image Credit

CSIRO Marine Research