The Big Earth Data Project

This set of activities explores the hole in the Ozone layer from 1980-2023 featuring daily and monthly data for the ozone hole and various atmospheric and weather measurements. The lessons cover aspects of the statistics element of A level Mathematics.

There is a very accessible guide to the hole in the ozone layer at https://discoveringantarctica.org.uk/oceans-atmosphere-landscape/atmosphere-weather-and-climate/the-ozone-hole/

All the lessons use screenshots CODAP which is free to use data exploration software. At the end of each lesson there is a CODAP file which you can use with your students to explore the data further.

We would value yours and your students' feedback. If time allows, please could your students fill in this feedback form (it does not have to be every student in your class) https://forms.office.com/e/UEGwYcanaz

The teacher feedback form is here, and also available on the website

https://forms.office.com/e/NGq9yr01pe

Overview of the lessons

Lesson 1: Interpret, analyse and compare distributions of data

Lesson objectives:

- Be able to interpret diagrams for single variable data including box plots and histograms
- Understand why there are different measures for average and spread
- Apply statistics to describe a population

Activity overview:

- Introduction to the ozone layer and how satellite data is used in observing the hole.
- Seasonal variation – how do different statistical calculations and graphs tell us how the ozone hole changes each year
- Exploration – what other features might affect the ozone hole?

SB 09/05/2022
Version 2

Lesson 1: Interpreting averages and measures of variation

Lesson objectives

- Be able to interpret diagrams for single variable data including box plots and histograms
- Understand why there are different measures for average and spread
- Apply statistics to describe a population

Interactive data

Resource:
https://codap.concord.org/app/static/dg/en/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2FVOwx7hZDB6xwTyybe671%2Ffile.json

Lesson Plan

Exploring the importance of understanding a context to solve a statistical problem

Introduce/remind students of the context

- If this is the first time using the context then use the opening section to talk to students about the importance of using context within statistics and use the slides to inform students about the ozone layer and the discovery of the ozone hole.
    - If you have completed some of the other Big Earth Data Project activities already with your class, you may want to briefly remind them of the context.

Exploring the ozone hole for different seasons

- Talk about the importance of visualising data in order to spot patterns
- Show the dot plot for the change in the hole over the course of the year
    - Ask the students to discuss the questions on the screen before feeding back
    - They should be able to that Spring has the largest ozone hole areas – notice that they can do this without doing any calculations but the calculations can give more information about the magnitude of the differences.
    - These seasons are for the Southern hemisphere so Spring is September, October and November
    - Information about why this happens in the Spring is in the link from discovering Antarctica.
- Talk about what the averages are and the different calculations
    - You might want to discuss why you wouldn't use mode for the average height, or why we talk about median wage rather than mean.
    - Often students will mention the range as an average so you may want to discuss the difference between averages and measures of spread
- Show the averages on the dot plot diagram and get the students to discuss the three questions on the slide.
    - The Spring data has a higher mean than median because the large values for the ozone hole area "drag" the mean higher. You could talk about the skew of the data here

SB 09/05/2022
Version 2

  - The mode would likely be 0 or close to 0 for all seasons and therefore wouldn't tell you about any differences
- Discuss measures of spread with your students, they may know both range and interquartile range
  - You may want to mention that there are more types of both average and measure of spread here but that we only study a few of them in schools
- Show the dot plot with the box plots laid over the top and get students to discuss the questions on the slide
  - You might notice that the interquartile range for summer is less than that for Autumn but it's the other way round for the range

Histograms and bar charts

- Talk about the fact that the diagrams used so far still aren't perfect for comparing populations of data
  - It's possible to have two sets of data which have very similar averages and measures of spread but can still look very different
- Show the diagram of the ozone hole area in Spring.
  - Note that in most parts of the world, this would be called a histogram even though the vertical axis is a "count" rather than a frequency density. For the purposes of this lesson, the students are invited to look at the shape of the data rather than the actual values themselves
  - You may need to go back to the box plot for the ozone hole for Spring to compare.
- You may want to show histograms for the other seasons. This might be an opportunity to show the CODAP file and demonstrate finding different histograms at this point.
  - There are some videos demonstrations if you are more comfortable with showing these.

- Summarise the main points about box plots and comparing distributions with averages and measures of spread.

SB 09/05/2022
Version 2

Lesson 2: Time series

Lesson objectives

- Understand what a time series shows and its uses
- Be able to use time series to describe trends and

Interactive data

Resource:
https://codap.concord.org/app/static/dg/en/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2FVOwx7hZDB6xwTyybe671%2Ffile.json

Lesson Plan

Exploring the importance of understanding a context to solve a statistical problem

Introduce/remind students of the context

- If this is the first time using the context then use the opening section to talk to students about the importance of using context within statistics and use the slides to inform students about the ozone layer and the discovery of the ozone hole.
    - If you have completed some of the other Big Earth Data Project activities already with your class, you may want to briefly remind them of the context.

Time series

- Briefly explain what a time series is for and how it is plotted
- Use the slide to talk about how time series can be used to answer the questions on the screen
- Show the overall time series
    - Invite the students to discuss the questions on the screen
    - Students should recognise that the there is a repeating pattern of peaks and troughs in the data
    - Students should notice that in the 1980s the peaks are lower than later in the graph
    - Largest values are early 2000s or middle of the 2010s
    - There are a few suggestions of how to improve the graph in the notes on the slides
- Use the slide to talk about smoothing out the time series by only plotting one point per year
    - There are so many values close to 0 in the data that most years may have a median of 0 which would mean the graph would be unusable
    - In addition, we should be worried about when the ozone hole is large which would be ignored by using the median – the maximum value might be more useful (this is shown later)
    - Trend from the 1980s should be more apparent with the yearly averages
- Show the two graphs side by side if needed to encourage students to spot what is more useful about each of the two graphs
    - Note that one criticism of some visualisations is that ‘you can’t see all the data’ but sometimes it’s better to remove some of the data to more easily see trends
- Show the graph of the maximum and the mean at the same time – is this more useful?
    - The shapes are fairly similar but there are some subtle differences such as the maximum increasing from 2020 to 2021 but the mean decreases.

SB 09/05/2022
Version 2

Seasonal variation

- Repeating patterns within a year on time series are often called seasonal variation.
    - It's hard to tell on this graph but the large values always start appearing towards the end of the year
- Show the overlay of all the years
    - This is much easier to see that the ozone hole appears around the 220th day of each year (8th August). This is when the Southern hemisphere is at the end of its Winter.
    - The year which has a much lower peak of ozone hole area is 2019, there were different weather patterns in that year which resulted in a smaller ozone hole. A good explanation of this can be found XXXXXXXX
- Ask the students which of the features in the data set may be causing the ozone hole to form
    - There's a good explanation in the links in the first section. Essentially the CFCs cause the ozone hole to form when the temperature is high enough and there is sunlight.
    - You can demonstrate putting other features onto the graph using the link.
    - There are some videos demonstrations if you are more comfortable with showing these.

SB 09/05/2022
Version 2

Lesson 3: Scatter graphs

Lesson objectives

- Be able to interpret associations between two variables using a scatter graph
- Be able to use a line of best fit to make predictions
- Understand when predictions are likely to be reliable

Interactive data

Resource:

https://codap.concord.org/app/static/dg/en/cert/index.html#shared=https%3A%2F%2Fcfm-shared.concord.org%2FKugDtm4SZwjLfLNCVRIS%2Ffile.json

Lesson plan

Exploring the importance of understanding a context to solve a statistical problem

Introduce/remind students of the context

- If this is the first time using the context then use the opening section to talk to students about the importance of using context within statistics and use the slides to inform students about the ozone layer and the discovery of the ozone hole.
    - If you have completed some of the other Big Earth Data Project activities already with your class, you may want to briefly remind them of the context.

Scatter graphs

- Briefly explain what a scatter graph is and what it is for
- Use the slide to talk about how scatter graphs can be used to answer the questions on the slide
- Show the two scatter graphs
    - Students should quickly be able to spot that there is probably a relationship between the ozone hole area and the column minimum but not between ozone hole area and the surface air pressure.
    - Take this opportunity to point out how useful this diagram is compared to a table of data when spotting relationships.
    - Ozone column minimum is the amount of ozone between the surface of the Earth and the top of the atmosphere and is measured in Dobson units (DU). Typical values are in the region of 300 DU
- Show the scatter graph between ozone hole area and column minimum.
    - You may want to talk about negative and positive correlation at this stage
    - Students should be able to associate the words ‘negative correlation’ with a description of ‘as this goes up, this goes down’
    - Notice the flat area of the graph higher than 220DU. Any model we make for predicting data should include this.
- As you may have already seen in previous activities, the ozone hole is mainly visible in the spring.
    - Summer appears to have 0 ozone hole area for all months
    - Both Winter and Autumn have some non-zero values and these appear to follow a similar pattern to the Spring data

SB 09/05/2022
Version 2

Lines of best fit

- Briefly explain what a line of best fit is for
- Use the handout to get students to draw their own line of best fit on the data
    - Again the 0 values for the ozone hole area probably won't fit with many of students lines.
- Lines of least squares are calculated by minimising the total vertical distance between the points and the line. The line will have tried to include the 0 values as well.
    - It's perfectly acceptable for students to think that their line is better than the calculated one (probably because they ignored the non-zero values)
- Show the equation of the calculated line
    - You could use this opportunity to talk about y = mx + c and how to find the gradient and intercept
- Complete the example to show students how to substitute into the line and calculate an estimate of the ozone hole area
- Ask the students to complete the questions
    - One of the answers will be below 0. Students need to recognise that they are calculating an area so this is an invalid estimate
    - Encourage students to think about how they could adapt their model to fit more data
    - You could put a limit on values to substitute into the equation and say anything more than 220 DU should give an area of 0
- Summarise the work including the fact that not all relationships are a straight line and that estimates beyond the end of the data set (extrapolation) is potentially unreliable
- Use the interactive data to draw scatter diagrams for the other features.
    - There are some videos demonstrations if you are more comfortable with showing these.

SB 09/05/2022
Version 2