CensusAtSchool
NEW ZEALAND

Exploring Our Class Data

2021

LEVEL 4

This activity is designed to support students exploring the class data that the teacher can access following their students undertaking the 2021 CensusAtSchool survey.

Datasets

The teacher gets back several datasets for their class in comma separated variable files (.csv). The order of the data in these individual datasets has been randomised, which means you cannot match it to any individual student. Privacy concerns are the reason for the provision of individual datasets.

One dataset includes questions 8-19, which gives some of the survey questions about you, measurements, and survey questions about school and games.

Q.	Variable	Outcomes	Units	Type
8	eyecolour	blue, brown, grey, green, hazel		Categorical
9	handed	right, left, ambi		Categorical
10	height		cm	Numerical
11	footleft footright		cm	Numerical
12	wrist		cm	Numerical
13	thumb		cm	Numerical
14	travel	walk, motor, bus, train, bike, boat, scooter, skateboard, other		Categorical
15	timetravel		minutes	Numerical
16	bagweight		kg	Numerical
17	memory		seconds	Numerical
18	reaction		seconds	Numerical
19	standleftleg		seconds	Numerical

Table 1: Variable, outcomes, units, and type of variable for questions 8-19 from 2021 questionnaire

Individual datasets are provided for each of questions 20-29, survey questions about activities and opinions.

Q.	Variable	Outcomes	Units	Type
20	screentime		time (hh:mm:00)	Numerical
21	techphone techfacebook techinstagram techtwitter techtiktok techsnapchat techreddit techdiscord techroblox techminecraft techfortnite techyoutubewatch techyoutubeupload technone	yes no		Categorical
22	screentimeopinionphone screentimeopinionsocialmedia screentimeopinionvideogames	too much, about right, too little		Categorical
23	newstv newsradio newsprint newssites newssocialmedia newssearch newspodcasts	often, sometimes, rarely, never		Categorical
24	bedtime waketime sleeptime		24-hour time (hh:mm:00) hours	Numerical
25	favanimatedmovie			Categorical
26	learningathome	much better, better, about the same, worse, much worse		Categorical
27	climatechange	urgent, future, not a problem, don't know		Categorical
28	future	very positive, positive, neutral, negative, very negative		Categorical
29	included	always, very often, sometimes, rarely, never		Categorical

Table 2: Variable, outcomes, units, and type of variable for questions 20-29 from 2021 questionnaire

Statistical Enquiry Cycle

When students are provided with data to explore the statistical enquiry cycle, the PPDAC cycle, is rearranged to accommodate a different way of thinking about the statistical investigation (See Figure 1).

- We start with the **dataset**.
- We then interrogate the data to work out who the data was collected from, what was collected, when it was collected, and where it was collected. In other words, we understand the **plan** the original investigator had. In the plan phase we are identifying what variables we have and who the group of interest is for the dataset we have.
- Knowing what variables are available we can **pose investigative questions** that we can answer with the data we have.
- We then **analyse** the data; we display and then describe the data.
- Finally, we draw **conclusions**, answering the investigative question.

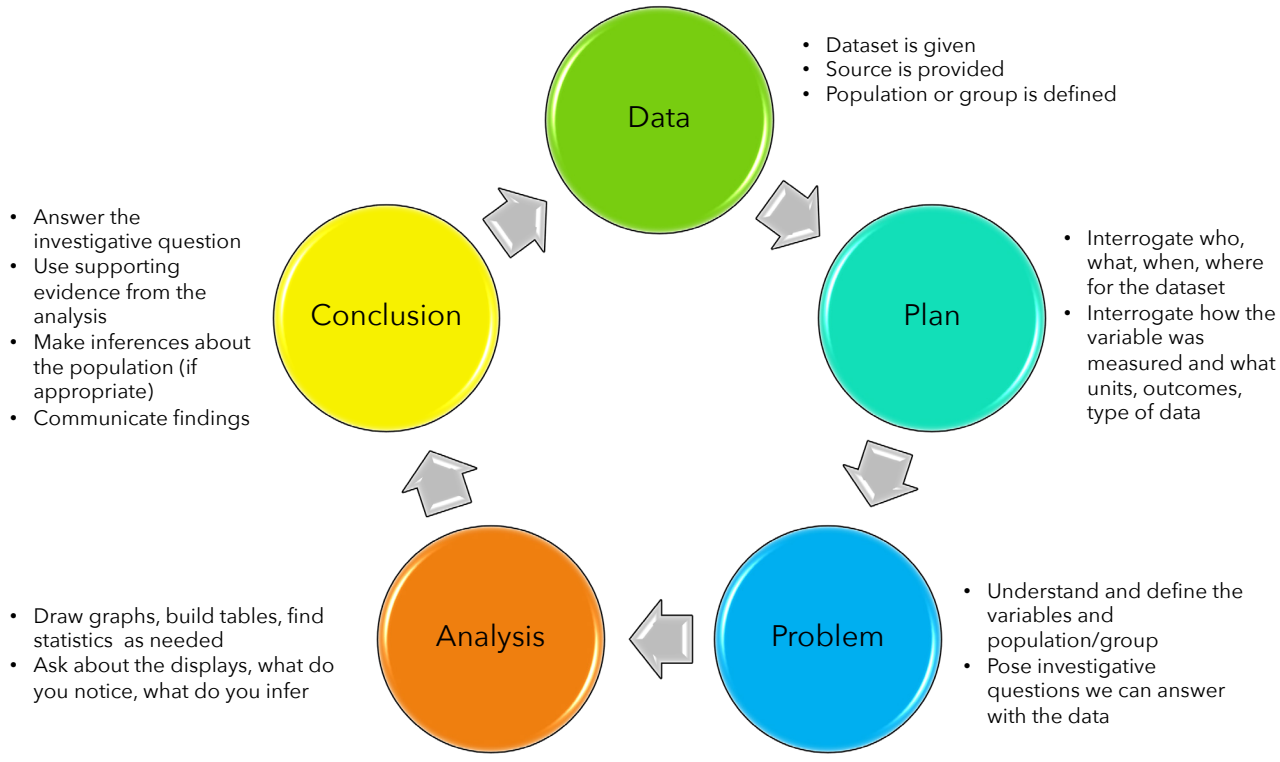

Figure 1: The statistical enquiry cycle for provided datasets (adapted from Arnold, P. (2013). Statistical Investigative Questions – An Enquiry into Posing and Answering Investigative Questions from Existing Data, (Doctoral thesis), Retrieved from https://researchspace.auckland.ac.nz/handle/2292/21305)

Statistical Software

It is assumed that students have access to technology, e.g., a Chromebook, desktop computer or laptop computer. Tablets can be used as well but the aforementioned options are preferable. Statistical software is recommended to create data visualisations. For example, the Common Online Data Analysis Platform (CODAP) https://codap.concord.org/ is a free online tool for students (and teachers) to use. Students in year 5 have used this software successfully following an introduction to the tool.

Pip Arnold created a series of short videos to introduce students to using CODAP during lockdown in 2020. Teachers can make a copy of the Google doc and share it with their students. To access the Google doc see her blog post: **https://karekareeducation.co.nz/codap/getting-your-students-started-with-codap/**.

The teacher can import their .csv files from CensusAtSchool that they received into one or more CODAP documents, for example, Q8-19 into one CODAP document and the remaining .csv files for the individual questions (Table 2) can all be imported into one CODAP document if you want. There will just be many separate tables in the CODAP document (see Figure 2). See this video **https://youtu.be/wdAjZEi7vtk** on how to import data into CODAP.

Once that is done, teachers need to save the CODAP document and share with their students (see **https://youtu.be/8LhcFWVsq5k** on how to share CODAP documents with your students).

Figure 2 is a screenshot of a CODAP document set up with the data for questions 20-29 showing all of them in the one document. To do this, each .csv file is imported into the same CODAP document. In the screenshot all the tables have been changed to case card view and then minimised. Students can open the case card for the question or questions they want to explore. Q22 has been dragged out of the list and then opened to case card view as an example.

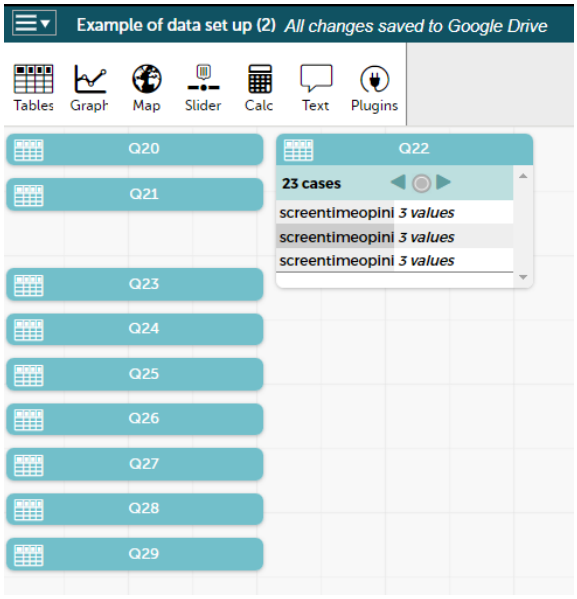

Figure 2. Showing multiple data tables in CODAP, all minimised except for Q22

Data

Share the dataset(s) with your students using the format you have decided on. If you are using CODAP or other statistical software share the link to the data.

Explain that the dataset(s) contains the class data for the CensusAtSchool survey they participated in on [date]. Get students to share what they remember about doing the survey and collecting data for the survey.

Confirm that the data is the class data for the CensusAtSchool 2021 survey.

Plan

With provided data, such as this class data, take the time for students to familiarise themselves with the variables that are available to use. As they will have been involved in the data collection process, they should already be familiar with how the different variables were measured or the possible outcomes. However, it is worthwhile having both the questionnaire available **https://new.censusatschool.org.nz/2021-questionnaire** and the variable list (Tables 1 and 2).

1. In pairs get the students to pick one categorical variable and one numerical variable.
2. For each variable ask them to find the following information:
    a. State the variable
    b. What was the survey question asked to collect the data?
    c. How was the variable measured? E.g., were there specific instructions?
    d. What are the units, if any, for the variable?
    e. What are the possible outcomes for the variable?
    f. What type of data is it? Categorical or numerical?

Problem

At level 4 students are exploring multivariate categorical, measurement, and time-series data to detect patterns, variations, relationships, and trends. They are working with data from the whole group – the class.

Students can explore summary, comparison and relationship investigative

questions with the data that is provided back to the teacher from the 2021 CensusAtSchool questionnaire. You might start with doing one or two investigative questions together as a class and then allow the students to select their own investigative questions to explore.

Example teacher investigative questions

Summary investigative question:

What do the students in our class think about climate change?

Comparison investigative question:

Do students in our class who come to school by car tend to get here faster than those students in our class who walk to school?

Relationship investigative question:

Are our left feet and our right feet the same length?

Predicting class results

Before you start to answer the investigative question get students to predict what they think the class results will be. You could do this by giving students sticky notes to write their guesses on and they can stick their guesses to a chart with the investigative question on it (Figure 3).

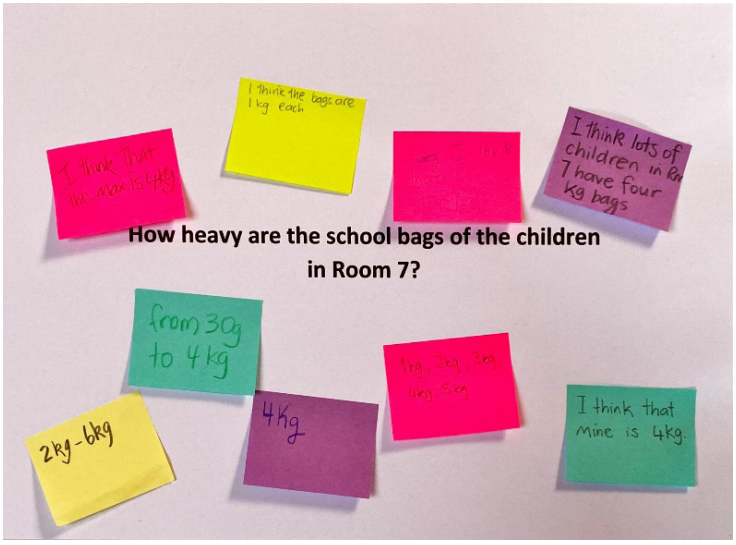

Figure 3. Example of student predictions for an investigative question.

Other possible investigative questions

Summary investigative questions

All the variables can be explored as summary situations. Some of the variables with only yes or no responses may not be as interesting as other variables with three or more outcomes.

Comparison investigative questions

There are limited comparison situations with the class data, the main one being the example given above. Other options if students want to be creative include comparing:

- Bag weight by mode of transport
- How long they stand on their left leg by eye colour

To do more with comparisons you might like to download a sample from CensusAtSchool which will give other variables to compare by e.g., year level, gender, ethnicity, region.

Relationship investigative questions

There are many relationship investigative questions that can be explored. For example:

Numerical variables

- Height and right foot length
- Height and left foot length
- Right foot length and left foot length
- Wrist circumference and thumb circumference
- Height and wrist circumference
- Height and thumb circumference
- Height and bag weight
- Reaction time and memory time
- Reaction time and stand leg time
- Memory time and stand leg time
- Time to school and bag weight
- Bedtime and sleep time

Categorical variables

Relationships between categorical variables can only be explored within a

question for Q20-29. Questions 21, 22 and 23 provide this opportunity. For example:

- Opinion about social media and opinion about video games
- Used in the last week, e.g.:
  - Facebook and Instagram
  - TikTok and Snapchat
  - Fortnite and Minecraft

For Q8-19 the relationship between eye colour (Q8) and handedness (Q9) could be explored.

Analysis

Students create data visualisations using statistical software (e.g., CODAP) to support answering the investigative question. For each data visualisation they should write statements about what they "notice" in the graph. Using the starter "I notice…" is a good writing prompt.

Teachers need to model writing descriptions with their students, which includes "thinking out loud" when you are doing the description. Thinking out loud could be a series of analysis questions that you are asking yourself to support the description. For example:

- `What is the most common opinion for climate change and how many people chose that?`
- `What is the next most common opinion… and now many people chose that?`
- `What is the least common opinion…?`

Including the context in the description is important too. This happens best by including the variable, the group, values and units for numerical variables, outcomes and counts for categorical variables. For example:

- `I notice that 13 [count] students in our class [group] think climate change [variable] is an urgent problem that needs to be managed [outcome].`
- `I notice that the time taken to walk to school [variable] for students in our class [group] varies from 5 [value] minutes [units] to 30 [value] minutes [units].`

Examples of possible data visualisations and accompanying descriptions for each

type of investigative question are given below. Note your class data will very likely look different.

What do the students in our class think about climate change?

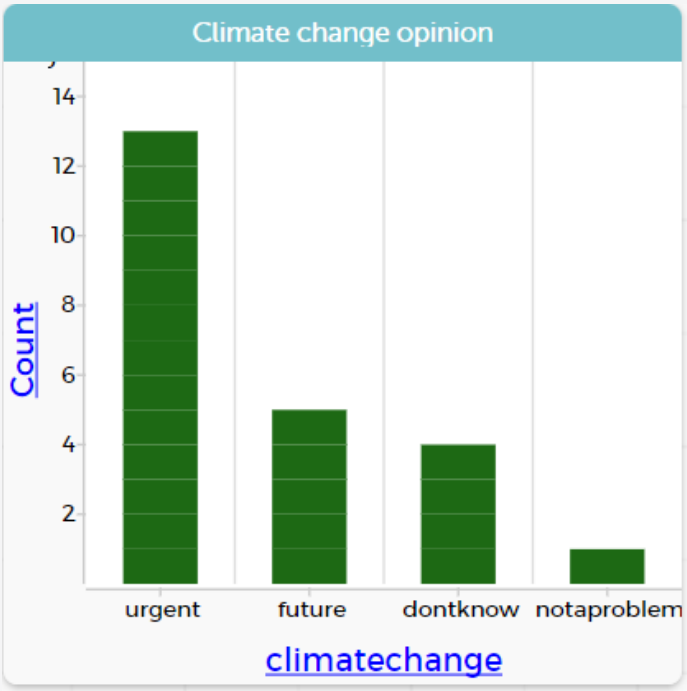

Figure 4. Bar graph showing our class opinions on climate change

I notice… that 13 students in our class think climate change is an urgent problem that needs to be managed now.

I notice… that five students in our class think climate change is a problem that needs to be managed in the future.

I notice … that five students in our class think climate change is not a problem (1 student) or don't know or have no opinion (4 students).

Do students in our class who come to school by car tend to get here faster than those students in our class who walk to school?

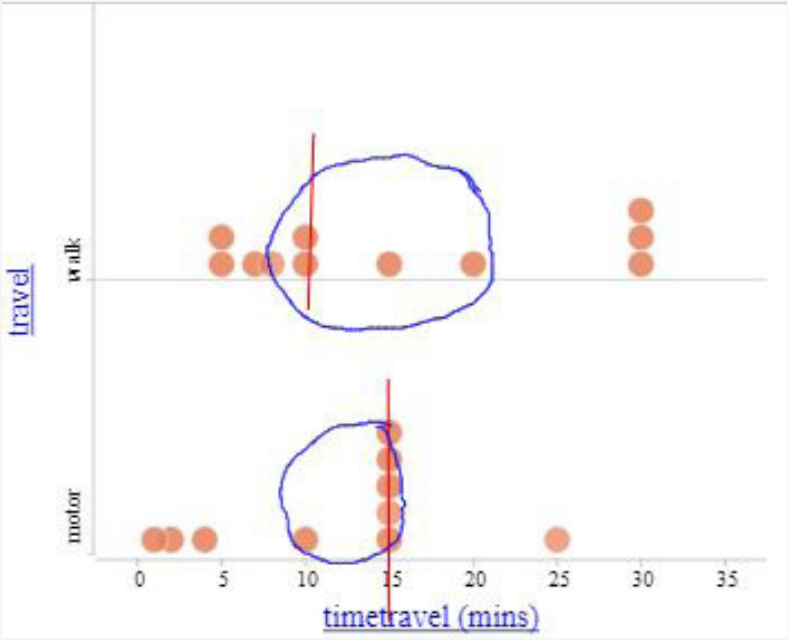

Figure 5. Dot plots showing the travel time to school for our class split into those who walk to school and those who get to school by car. The red line shows the approximate middle value, the blue circle shows the approximate middle 50% of values for time travel to school.

I notice that the time taken to walk to school for students in our class varies from 5 to 30 minutes. The middle time it takes to walk to school is approximately 10 minutes. The middle group of times taken to walk to school is from 8 to 20 minutes, five out of 11 students who walk to school took between 8 and 20 minutes to get to school. Three students take 30 minutes to walk to school, in fact this is the most common time for walking to school.

I notice that the time taken to get to school by car for students in our class varies from one minute to 25 minutes. The middle time it takes to get to school by car is 15 minutes. The middle group of times taken to get to school by car is from 10 to 15 minutes, six out of 10 students took 10 to 15 minutes to get to school by car. Five students in our class who come by car take 15 minutes to get to school. This is the most common time to get to school by car.

The middle time to walk to school (10 minutes) is less than the middle time to get to school by car (15 minutes). The middle group of times to walk to school (8-20 minutes) completely overlaps the middle group of times to get to school by car (10-15 minutes).

Are our left feet and our right feet the same length?

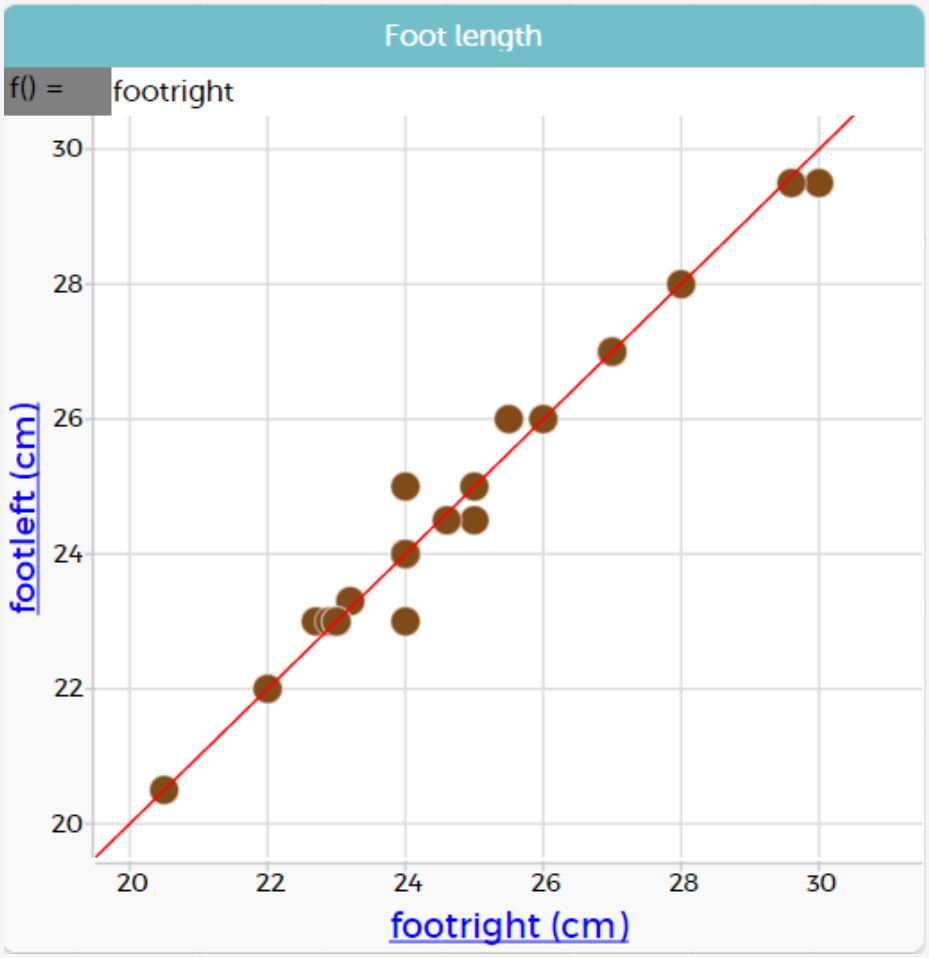

Figure 6. Scatter plot showing left foot length versus right foot length. The line where right foot length and left foot length is equal has been drawn on the graph.

I notice on the graph that 11 students in the class have the same length right foot as left foot. These are the students who are on the line where right foot length is equal to left foot length. Three students have a longer right foot than their left foot, these students are “below” the line. Four students have a longer left foot than their right foot, these students are “above” the line.

However, when I add up these numbers, I get 18 students but there are 23 in the class. I checked the data table and found that there are two students with the same length of 23 cm, four students with the same length of 24 cm, and two students with the same length of 27 cm. This gives me an additional five students who have the same length right foot and left foot. So altogether there are 16 students who have the same left foot length as their right foot length.

Further Support Describing Data Visualisations: Curriculum Level 4

Cars

The activity uses the context of cars to explore paired numerical data. The activity focuses on using the y=x line and describing areas of the scatter plot.
https://nzmaths.co.nz/resource/cars

Measuring Up

In the unit the students will collect statistical data about their own class and school and learn how to compare it to data from students from CensusAtSchool. Students learn about finding the middle and the middle 50% of the data using CODAP. **https://nzmaths.co.nz/resource/measuring**

Crunch the Coach

Amongst other things, Crunch the Coach includes describing distributions (summary - category and numerical, relationships - paired numerical).
https://nzmaths.co.nz/resource/crunch-coach

Conclusion

The last stage of the statistical enquiry cycle is the conclusion, where we answer our investigative question(s). We use evidence from our analysis to answer the investigative question. Examples for each of the investigative questions are given below. Student responses will differ. There is no **one** right answer, but the answers must align with their evidence in the analysis.

What do the students in our class think about climate change?

Most students in our class (13 out of 23 students) think that climate change is an urgent problem that needs to be managed now. Four students didn't know or had no opinion, one thought it was not a problem.

Do students in our class who come to school by car tend to get here faster than those students in our class who walk to school?

It is hard to say if coming to school by car is faster or slower than walking to school. The middle time to walk to school is shorter than the middle time it takes to get to school by car. However, three students take 30 minutes to walk to school, which is five minutes higher than the slowest car ride to school. Looking at the approximate middle values, the times for going to school by car and walking are about the same.

Are our left feet and our right feet the same length?

Our left feet and our right feet are the same length for most of the students in our class. Sixteen students in our class of 23 have the same length right foot as their left foot. Three students have a slightly longer right foot than left foot and four students have a slightly longer left foot than right foot.