# Random samples

Students explore why we can trust sample data and how collection bias can appear when using samples.

## Visible learning

### Learning intention

* To understand how to use sample data to represent a population.

### Success criteria

* I can explain the potential bias that can occur when using a sample.
* I can explain why a variation in measures of centre occur when using a sample.
* I can justify why I can draw valid conclusions from sample data.

### Syllabus outcomes

A student:

* develops understanding and fluency in mathematics through exploring and connecting mathematical concepts, choosing and applying mathematical techniques to solve problems, and communicating their thinking and reasoning coherently and clearly **MAO-WM-01**
* classifies and displays data using a variety of graphical representations **MA4-DAT-C-01**
* analyses simple datasets using measures of centre, range and shape of the data   
  **MA4-DAT-C-02**

[Mathematics K–10 Syllabus](https://curriculum.nsw.edu.au/learning-areas/mathematics/mathematics-k-10-2022/overview) © NSW Education Standards Authority (NESA) for and on behalf of the Crown in right of the State of New South Wales, 2022.

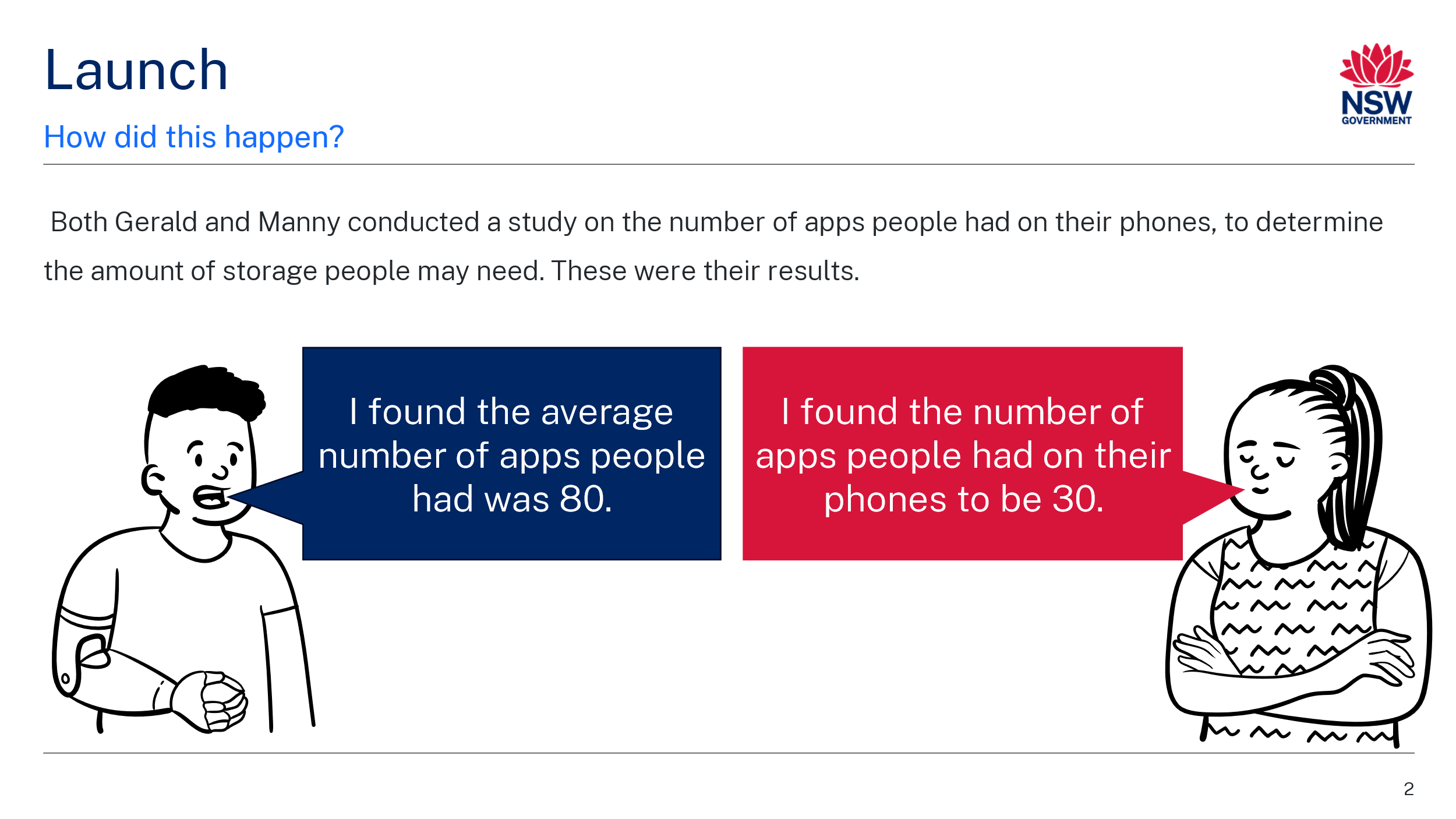
## Activity structure

Please use the associated PowerPoint *Random samples* to display images in this lesson.

### Launch

1. Display slide 2 from the PowerPoint *Random samples.* This slide shows 2 people who arrived at different conclusions from a study they conducted. The slide is also displayed below.

Figure 1: Launch scenario



1. Use a Think-Pair-Share ([bit.ly/thinkpairsharestrategy](https://bit.ly/thinkpairsharestrategy)) for students to discuss why Gerald and Manny had such different results.

Students may discuss if Gerald or Manny used an appropriate measure of centre, if any outliers were affecting their data analysis, if either of their data was skewed, when the data was collected, who they asked to obtain their data or how many people they asked to collect their data.

1. Use a Pose-Pause-Pounce-Bounce (PDF 557 KB) ([bit.ly/posepausepouncebounce](https://bit.ly/posepausepouncebounce)) questioning strategy for students to share their thoughts and reasoning.

### Explore

During this section, students will work in visibly random groups of 3 ([bit.ly/visiblegroups](https://bit.ly/visiblegroups)) to complete a jigsaw activity ([bit.ly/jigsawgroupstrategy](https://bit.ly/jigsawgroupstrategy)).

1. Set up a Jigsaw task with 3 stations. Each station will focus on a different type of bias that can occur when collecting data. The activity for each station can be found in Appendix A ‘Samples’.

For larger classes, it may be beneficial to set up 2 of each station, so groups don’t exceed more than 5 people at each station.

1. Explain to students that each person in their group of 3 is to visit a different station and complete the activity from that station. The activity will be completed at the station and students can discuss answers with other students working at that station.
2. Students at each station should discuss what may have caused the large differences between the measures of centre of the 2 datasets.

Students might notice a difference in the measures of centre in the data which shows the data is located more around different points or they may notice the range is very different as they didn’t ask a large group of people.

1. Inform students that they will be exploring how bias in the data collection could have impacted the data.
2. Distribute the scenarios from Appendix B ‘Data collection scenarios’ and ask one student at each station to read the scenario to the rest of the station. The scenario explains how Gerald and Manny each collected their data.
3. Ask students to discuss at their station, how the data collection could have affected the data.
4. Students should return to their original group and take turns to show their group members the data examined and explain how they believe it created bias in their data analysis.
5. Ask students to consider and propose ways that Gerald and Manny could improve data collection so that it represents the whole population more effectively.
6. Initiate a sharing of responses by selecting random students. For students who do not have something new to add, ask them to revoice or repeat ([bit.ly/classroomtalkmoves](https://bit.ly/classroomtalkmoves)) another student's solution.

### Summarise

1. Use slides 3–5 of the PowerPoint Random samples for a slow reveal graph activity ([slowrevealgraphs.com](https://slowrevealgraphs.com/)). For each slide, students will discuss in a Think-Pair-Share ([bit.ly/thinkpairsharestrategy](http://www.bit.ly/thinkpairsharestrategy)) what conclusions they can draw from the graph.
2. Using a class discussion ask students if they kept or modified the conclusions that they originally drew from the graph on Slide 1. Why or why not?

Students should have modified their conclusions if they originally assumed that the data represented the population.

1. Distribute Appendix C ‘Data displays’ to each pair of students*.*
2. In a Think-Pair-Share, ask students the following questions:

* If they were a company that made dentures, which dataset would they use to base their decisions on? Why?
* What decisions might they make given the data displayed?

1. Students are to create notes to their forgetful future selves ([bit.ly/notesstrategy](https://bit.ly/notesstrategy)) concerning what may cause bias when collecting data. Useful sentence starters include:

* Bias can occur in data collection. Three of the ways it can occur are …
* To know if a dataset has minimised bias, we can check…

### Apply

1. Inform students they will now be asking the class how many apps they have on their phones.
2. Use a class discussion to consider how to record the data for someone if they don’t have a mobile phone.

In preparation for the lesson, students should go home and count how many apps they have on their phones.

If a student does not have a mobile phone, they could enter their data as 0 or, if they prefer not to draw attention to the fact that they don’t own a phone, they could ask a friend or relative and use that number.

1. Assign students into random groups of 5 to create data groups.

This number may change depending on the number of students in your class or if you do conduct this investigation with multiple classes.

1. Students are to find the mean, median, range and mode of the data for their group and display it around the room.
2. The mean, median, mode and range for the whole class should also be calculated, representing the population.

This is best done by collecting each student's data as a single value and using functions on a spreadsheet.

1. Students are to conduct a gallery walk ([bit.ly/DLSgallerywalk](https://bit.ly/DLSgallerywalk)) noticing the different results produced by the different samples. Students should consider whether the samples could be used to represent the population (whole class).
2. Students should select and justify the sample group that best represents the population and most closely reflects the census data.
3. Use a Pose-Pause-Pounce-Bounce questioning strategy to discuss how samples can be used to represent a population and to summarise ways that bias can occur in data collection.

## Assessment and differentiation

### Suggested opportunities for differentiation

**Launch**

* There are no correct answers during the launch and all students should be encouraged to participate and share their thoughts and reasoning**.**

**Explore**

* Challenge students to describe the differences in the datasets using the mean, median, range and mode.
* Ask students if they know of any other types of bias.
* Using a jigsaw activity allows students to be the experts on their station to build their mathematical confidence and develop their communication skills in a group. Students are able to test ideas with other members of their jigsaw group before returning to their original group.
* Extend students by using a spreadsheet to find the mean, median, mode and range.
* In their original groups of 3, students can combine their lists of random numbers to find the mode, median, mean and range of a larger set of data and compare it to each of the original datasets.

**Summarise**

* To challenge students, ask them in what situation would we want to use the other datasets provided.

**Apply**

* Students can conduct the study on a different topic of interest such as the number of siblings or pets each person has.

### Suggested opportunities for assessment

**Explore**

* Monitor student explanations when returning to their groups in the Jigsaw activity to check for understanding.
* When placed in groups of 3, students provide and receive peer feedback on their understanding.

**Summarise**

* Review students’ notes to their future forgetful selves.

**Apply**

* Students will demonstrate their working mathematically skills in discussions and justifications**.**

## Appendix A

### Samples

#### Station 1

Gerald and Manny were interested to know how many apps people had on their smartphones. Each of them completed a survey to find out.

The data collected by Gerald and Manny can be seen in the following stem-and-leaf plots.

|  |  |
| --- | --- |
| Gerald | Manny |
| Stem and leaf plot of 64, 66, 67, 68, 70, 73, 73, 74, 79, 79, 81, 84, 87, 87, 88, 88, 89, 90, 94 and 95. | Stem and leaf plot of 17, 20, 20, 22, 23, 24, 26, 28, 28, 30, 30, 32, 34, 34, 34, 36, 37, 38, 43 and 44. |

1. Find the mean, median, mode and range for each dataset.
2. Describe the shape of the data and which measure of centre would be the most appropriate to use.
3. What has caused the difference in their measures of centre? Discuss in the group.

#### Station 2

Gerald and Manny were interested to know how many apps people had on their smartphones. Each of them completed a survey to find out.

The data collected by Gerald and Manny can be seen in the following stem-and-leaf plots.

|  |  |
| --- | --- |
| Gerald | Manny |
| Stem and leaf plot of 64, 66, 67, 68, 70, 73, 73, 74, 79, 79, 81, 84, 87, 87, 88, 88, 89, 90, 94 and 95 | Stem and leaf plot of 20, 21, 22, 26, 26, 27, 28, 29, 29, 29, 30, 31, 32, 34, 35, 36, 36, 37, 39 and 42 |

1. Find the mean, median, mode and range for each dataset.
2. Describe the shape of the data and which measure of centre would be the most appropriate to use.
3. What has caused the difference in their measures of centre? Discuss in the group.

#### Station 3

Gerald and Manny were interested to know how many apps people had on their smartphones. Each of them completed a survey to find out.

The data collected by Gerald and Manny can be seen in the following stem-and-leaf plots.

|  |  |
| --- | --- |
| Gerald | Manny |
| Stem and leaf plot of 64, 66, 67, 68, 70, 73, 73, 74, 79, 79, 81, 84, 87, 87, 88, 88, 89, 90, 94 and 95. | Stem and leaf plot of 25, 30 and 35. |

1. Find the mean, median, mode and range for each dataset.
2. Describe the shape of the data and which measure of centre would be the most appropriate to use.
3. What has caused the difference in their measures of centre? Discuss in the group.

## Appendix B

### Data collection scenarios

**Station 1**

Gerald collected his data last week. Manny sourced data from a survey completed in 2010 before smartphones became popular.

**Station 2**

Gerald surveyed a wide range of ages. Manny surveyed only people over the age of 70.

**Station 3**

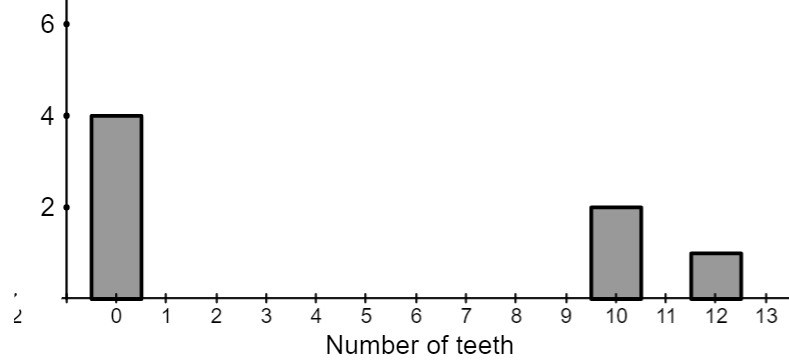
Gerald asked many people. Manny asked their mum, dad and sister.

## Appendix C

### Data displays

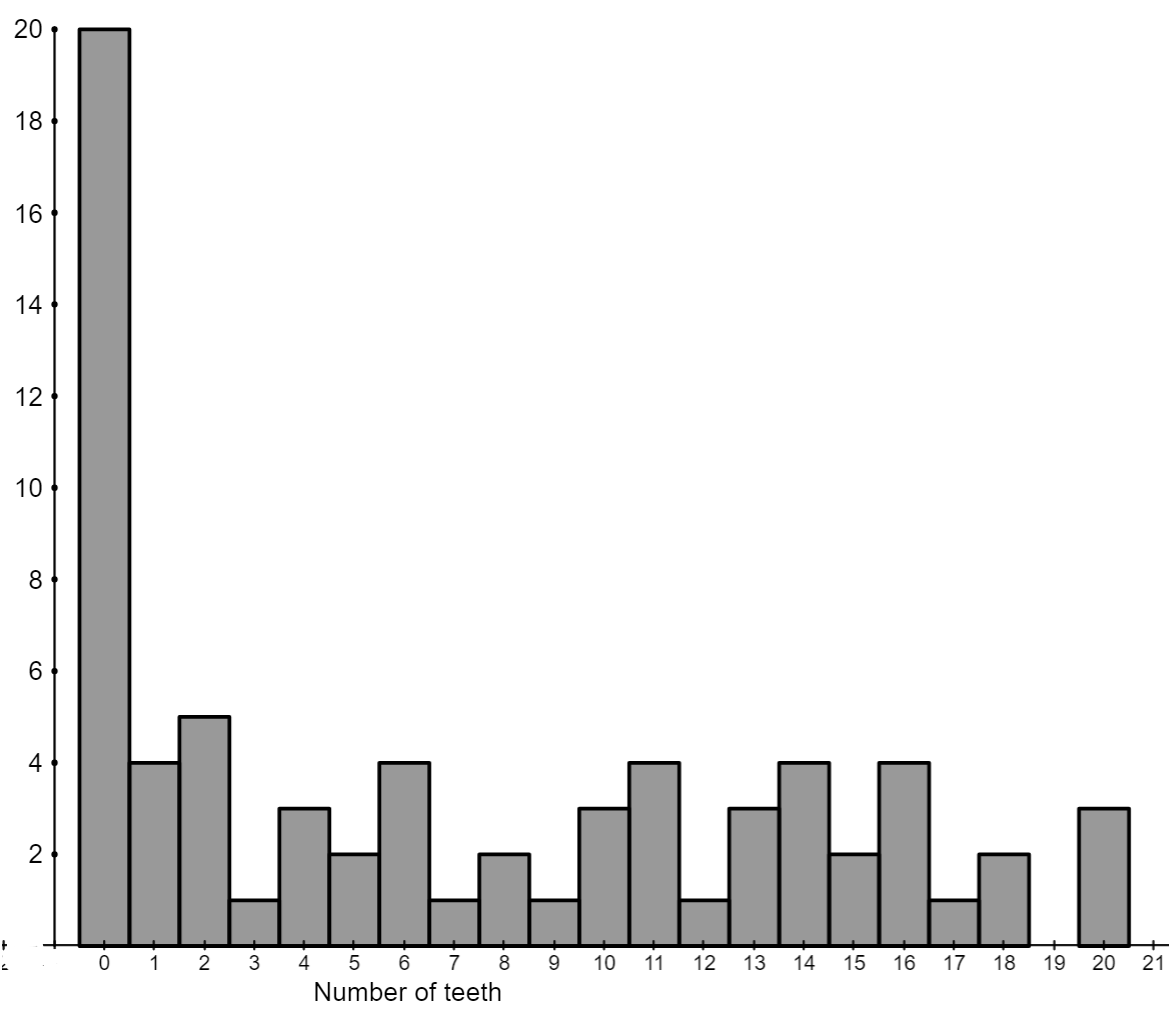
#### Graph 1

##### Number of teeth of people living in a nursing home



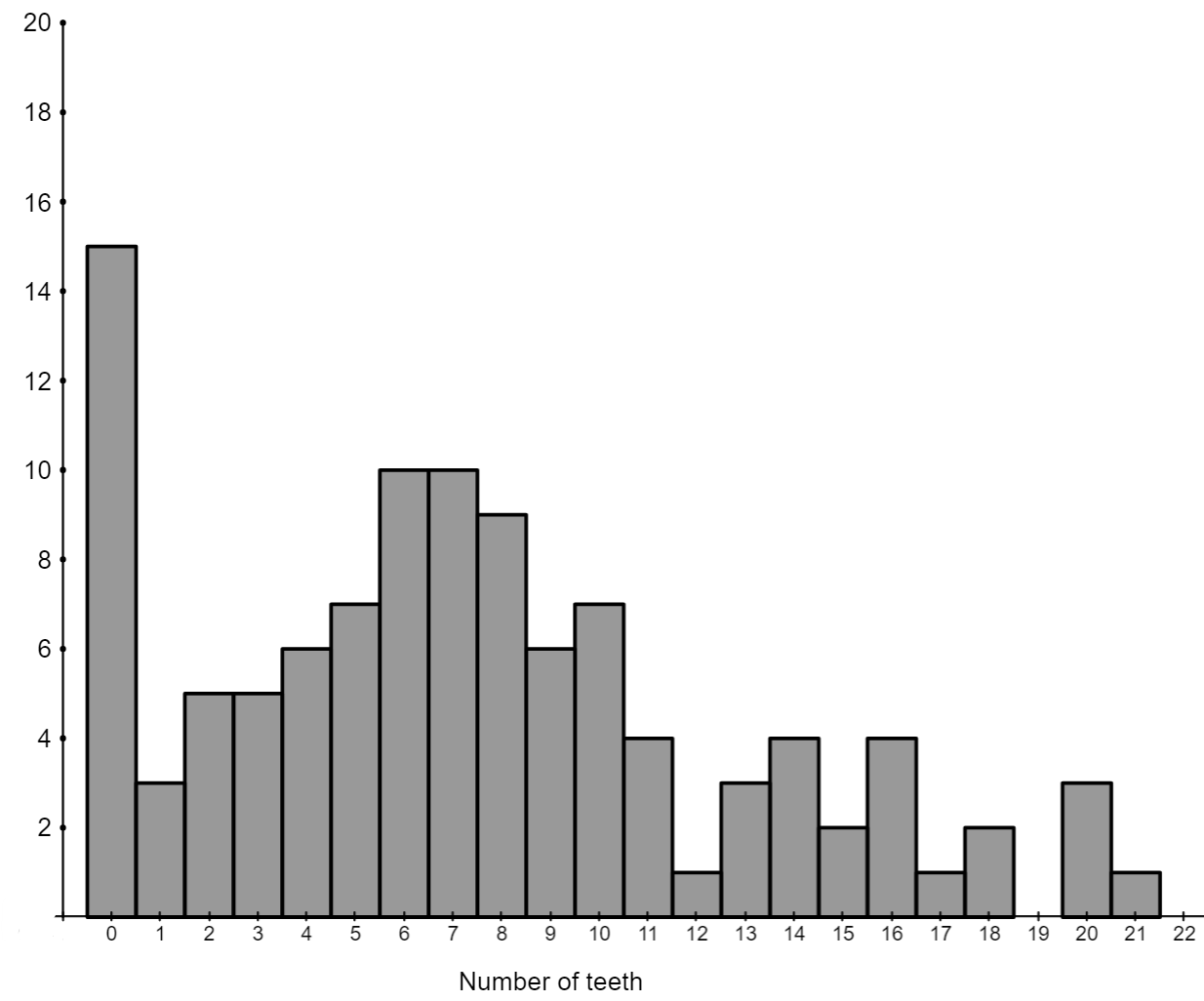
#### Graph 2

##### Number of teeth of children in a daycare centre



#### Graph 3

##### Number of teeth of people living in a nursing home



## Sample solutions

### Appendix A – bias

#### Station 1

|  |  |  |
| --- | --- | --- |
| Person | Summary statistics | Shape of data |
| Gerald | Mean – 79.8  Median – 80  Mode – 73, 79, 87, 88  Range - 31 | The data is symmetrical, which means we could use either the mean or median for our best measure of centre. |
| Manny | Mean – 30  Median – 30  Mode – 34  Range – 27 | The data is symmetrical, which means we could use either the mean or median for our best measure of centre. |

#### Station 2

|  |  |  |
| --- | --- | --- |
| Person | Summary statistics | Shape of data |
| Gerald | Mean – 79.8  Median – 80  Mode – 73, 79, 87, 88  Range - 31 | The data is symmetrical, which means we could use either the mean or median for our best measure of centre. |
| Manny | Mean – 30.45  Median – 29.5  Mode – 29  Range - 22 | The data is symmetrical, which means we could use either the mean or median for our best measure of centre. |

#### Station 3

|  |  |  |
| --- | --- | --- |
| Person | Summary statistics | Shape of data |
| Gerald | Mean – 79.8  Median – 80  Mode – 73, 79, 87, 88  Range – 31 | The data is symmetrical, which means we could use either the mean or median for our best measure of centre. |
| Manny | Mean – 30  Median – 30  Mode – n/a  Range – 10 | The data is symmetrical, which means we could use either the mean or median for our best measure of centre. |

### Appendix C – data displays

I would choose Graph 3 as older people are the clientele for dentures. Children in a daycare are still growing their teeth, so are not the target clientele. Graph 1 does not have enough data to be reliable.

## References

This resource contains NSW Curriculum and syllabus content. The NSW Curriculum is developed by the NSW Education Standards Authority. This content is prepared by NESA for and on behalf of the Crown in right of the State of New South Wales. The material is protected by Crown copyright.

Please refer to the NESA Copyright Disclaimer for more information <https://educationstandards.nsw.edu.au/wps/portal/nesa/mini-footer/copyright>.

NESA holds the only official and up-to-date versions of the NSW Curriculum and syllabus documents. Please visit the NSW Education Standards Authority (NESA) website <https://educationstandards.nsw.edu.au/> and the NSW Curriculum website [https://curriculum.nsw.edu.au/home](https://curriculum.nsw.edu.au/).

[Mathematics K–10 Syllabus](https://curriculum.nsw.edu.au/learning-areas/mathematics/mathematics-k-10-2022/overview) © NSW Education Standards Authority (NESA) for and on behalf of the Crown in right of the State of New South Wales, 2022.

**© State of New South Wales (Department of Education), 2024**

The copyright material published in this resource is subject to the *Copyright Act 1968* (Cth) and is owned by the NSW Department of Education or, where indicated, by a party other than the NSW Department of Education (third-party material).

Copyright material available in this resource and owned by the NSW Department of Education is licensed under a [Creative Commons Attribution 4.0 International (CC BY 4.0) license](https://creativecommons.org/licenses/by/4.0/).

[](https://creativecommons.org/licenses/by/4.0/)

This license allows you to share and adapt the material for any purpose, even commercially.

Attribution should be given to © State of New South Wales (Department of Education), 2024.

Material in this resource not available under a Creative Commons license:

* the NSW Department of Education logo, other logos and trademark-protected material
* material owned by a third party that has been reproduced with permission. You will need to obtain permission from the third party to reuse its material.

**Links to third-party material and websites**

Please note that the provided (reading/viewing material/list/links/texts) are a suggestion only and implies no endorsement, by the New South Wales Department of Education, of any author, publisher, or book title. School principals and teachers are best placed to assess the suitability of resources that would complement the curriculum and reflect the needs and interests of their students.

If you use the links provided in this document to access a third-party's website, you acknowledge that the terms of use, including licence terms set out on the third-party's website apply to the use which may be made of the materials on that third-party website or where permitted by the *Copyright Act 1968* (Cth). The department accepts no responsibility for content on third-party websites.