Describing distributions

In this workshop participants will actively engage in the learning process for developing the concept of distributional shape. Distributional shape is the foundation for describing the distribution of data for a specific variable. Onto distributional shape we can attach other descriptors such as signal and noise, contextual knowledge and variability.

A teaching and learning sequence for developing the concept of distributional shape, the distribution description framework and criteria for judging the quality of distribution description based on SOLO taxonomy will be shared with participants.

All materials that will be used and shared in the session were developed as part of Pip's PhD research: *Statistical investigative questions. An enquiry into posing and answering investigative questions from existing data*

Dr Pip Arnold
Karekare Education

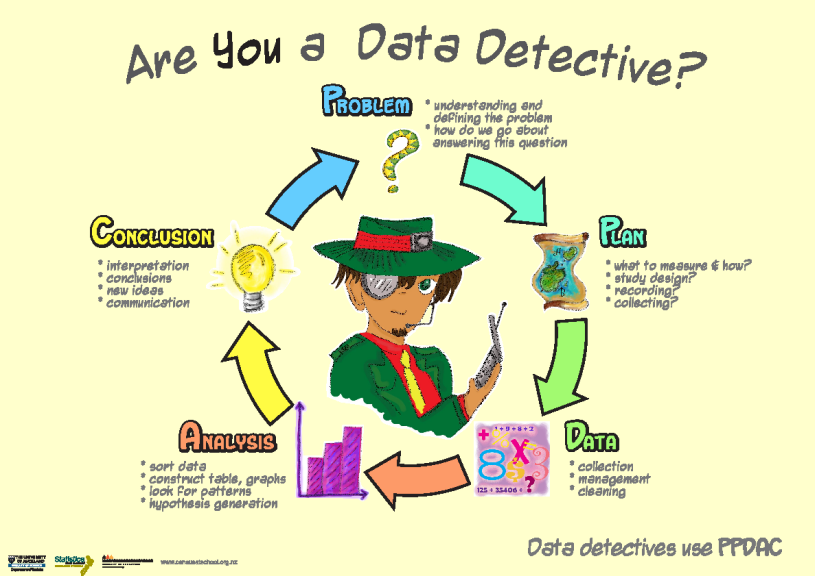
Statistics Day November 2018

Contents

Workshop plan 2
Sketching shapes 3
Activity 3
Predicting graphs 6
Activity 6
Connecting graphs and contexts 7
Activity 7
Describing distributions framework 9
Describing distributions 10
Activity 10
Using SOLO 11
Activity 12
Appendices 13
List of the variable and populations for the 15 graphs 13
Statistical terms for describing shapes of distributions 13
Actual graphs 14
Contexts for graphs 15
Teacher “sketched” graphs 16
Lessons on shape 17
Mix’n’match activity 21
Shape descriptors 22
Example of student work – building a context library 23
Dot plot 23

Workshop plan

Activity	Resources
Sketching shapes	PowerPoint Statistical language Sketching shapes
Predicting graphs	
Connecting graphs and contexts	Graphs Contexts
Describing distributions	



http://new.censusatschool.org.nz/resource/data-detective-poster/

Sketching shapes

Activity

Sketch the shape of the distribution as it is displayed on the slide. Looking for overall shape rather than fine detail. (See lesson on shape in appendices for use with students)

Resource needed: PPT - http://new.censusatschool.org.nz/resource/developing-the-language-of-statistical-graphs-shape/ sketching shapes link

1.	2.	3.
4.	5.	6.
7.	8.	9.
10.	11.	12.

13.	14.	15.

Sort the graphs into groups of similar shapes.

Write down the graph numbers in their groups and describe the shape of the graphs in each group

Sketch the different shapes that we have discovered with their statistical description

Match the statistical terms with graphs.

Predicting graphs

Activity

What do you think the graph of height in cm for year 5–10 students would look like? Describe and then sketch.

What do you think the graph of attendance percentage half days for year 7–8 students would look like? Describe and then sketch.

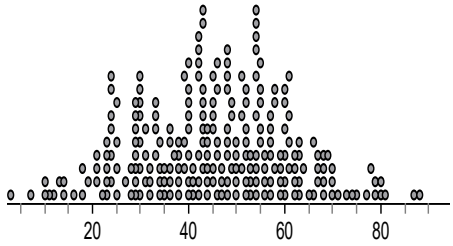
What do you think the graph of hair length in cm for year 4–13 students would look like? Describe and then sketch.

Connecting graphs and contexts

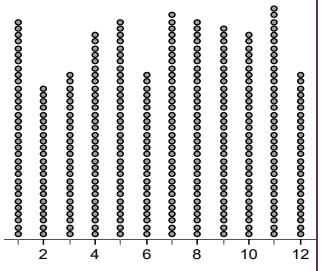
Activity

Connect the contexts with the right graph

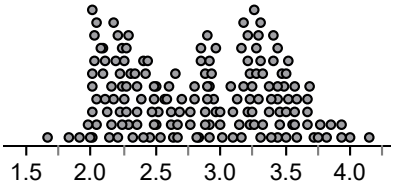
1.



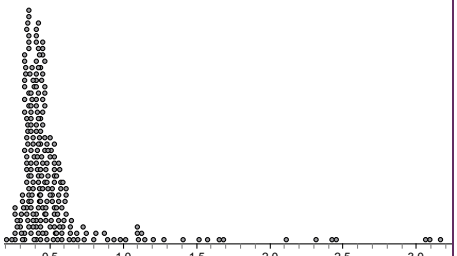
2.



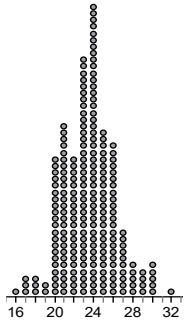
3.



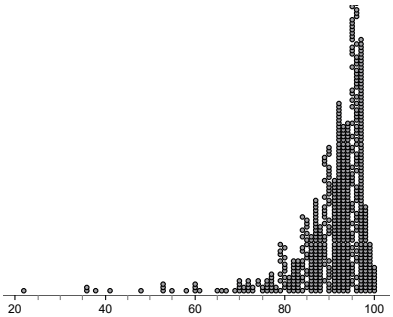
4.



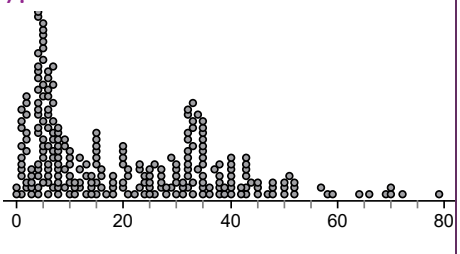
5.



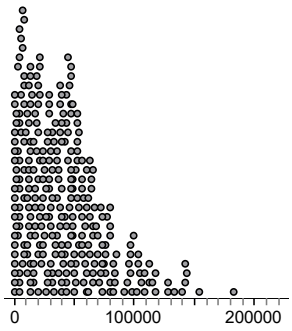
6.



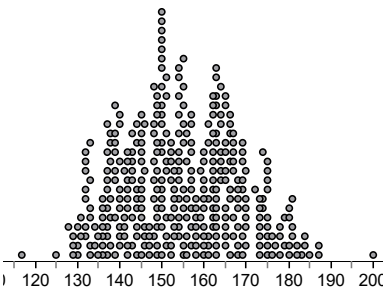
7.



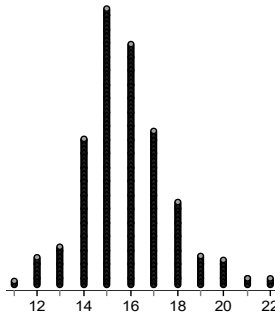
8.



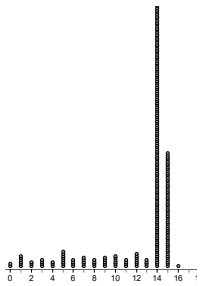
9.



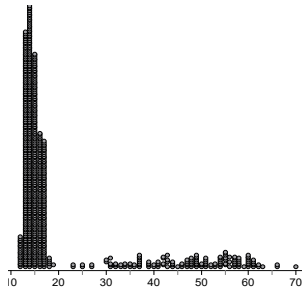
10.



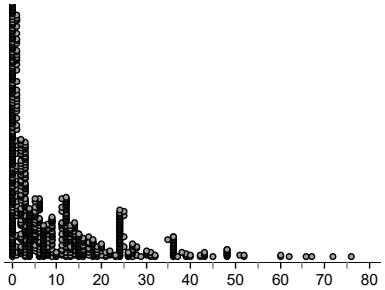
11.



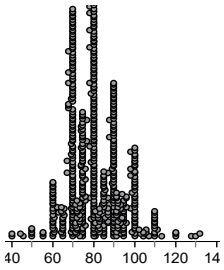
12.



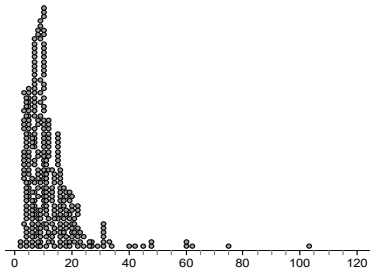
13.



14.



15.



Possible contexts

1. Age years: Everyone at a high school
2. Number of years living in NZ: C@S Yr 10 students
3. Hair length – cm: 2007 C@S Yr 4–13 students
4. Time to school – mins: 2009 C@S Yr 7–8 students
5. Reaction time – secs: 2007 C@S Yr 4–13 students
6. Household debt – $: Synthesised Unit Record File based on NZ data
7. Wrist length – cm: 2009 C@S Yr 7–8 students
8. Index finger length – mm: 2009 C@S Yr 7–13 students
9. Right foot length – cm: 2003 C@S Yr 5–10 students
10. Number of skips in 30 secs: 2003 C@S Yr 5–8 students
11. Attendance – percentage half days: Yr 7–8 students
12. Cell phone ownership – months: 2009 C@S Yr 9–13 students
13. Birth month: 2003 C@S Yr 5–10 students
14. Weight – kg: Kiwi Kapers Great Spotted Kiwi
15. Height – cm: 2003 C@S Yr 5–10 students

Describing distributions framework

Overarching statistical concepts	Characteristics of distribution	Specific features measures/depictions/descriptors
Contextual knowledge	Population	*Target population* (e.g. New Zealand year 5–10 students) Other acceptable population (e.g. year 5–10 students)
	Variable	Variable Units
	Interpretation	*Statistical feature described in contextual setting* (e.g. interpreting right skew as very few high test scores, with most test scores between 20 and 50 points)
	Explanation	*Possible reason for a feature* (e.g. bimodal due to gender for kiwi data)
Distributional	Aggregate view	General shape sketched Hypothesis and prediction
	Symmetry	Overall shape
	Modality	*Modality*
	Skewness	Position of majority of the data (to the left or the right)
	Individual cases	*Highest and lowest values*
Graph Comprehension	Decoding visual shape	Overall shape **Parts of the whole* (splitting the distribution into parts and describing the parts as well as the whole) *Modality*
	Unusual features	Gaps Outliers
Variability	Spread	Range, interquartile range **Interval for high and/or low values (*may be describing a tail)
	Density	Clustering density Majority (mostly, many) Relative frequency
Signal and noise	Centre	Median, mean
	Modal clumps	*Peak(s) (local mode)* *Modal group(s)*

Figure 8-16. Distribution description framework for curriculum level 5 (ages 13–15) (Arnold, 2013, page 217)

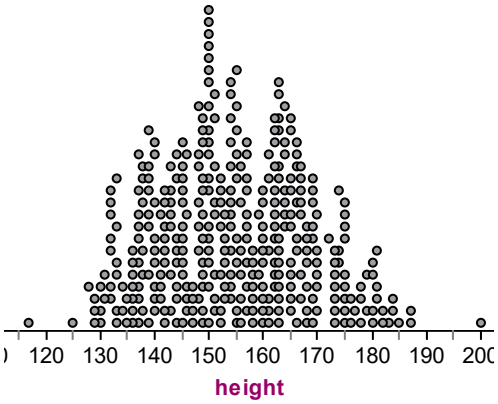
When students are describing statistical distributions they need: (1) to invoke contextual knowledge, (2) to know what relevant characteristics of distributions they can actually see in the plots and therefore describe, and (3) to be explicit about the evidence for specific features. In other words, students need to be able to identify which features are evident in a particular plot, name and provide evidence (values) for the features, and to interlace these with contextual information such as the population, variable and units. (Arnold, 2013, page 217)

Data are numbers with a context

Remember: **VARIABLES, VALUES, UNITS**

Describing distributions

Activity

Problem 	I wonder what typical heights of year 5–10 New Zealand students are.
Plan/Data 	Data collected from the 2003 Census At School database. http://new.censusatschool.org.nz/tools/random-sampler/
Analysis 	
	I notice… I notice… I notice…
Conclusion 	

Data are numbers with a context

Check your "I notice" statements and your conclusion for the context.

Remember: **VARIABLES, VALUES, UNITS**

Actively reflect on your statements, make corrections – this is a working document…

Using SOLO

SOLO taxonomy level	Grade	Description of evidence	Example of student responses with *commentary*
No response or idiosyncratic	0	• No response.	No examples.
Pre-structural	1	• Context and/or evidence missing.	Bunched up between 20 and 50? (2011 student, pre-test). • *no context*
Uni-structural	2	• Students give one correct* piece of evidence in simple context (usually variable only; e.g. heights, scores). OR • Evidence for multi-structural without any context.	The results are quite spread out. There is a high from 36–38 and 40–42. Then there are fewer heights from 35–36 and 42–43 (2011 student, pre-test). • *one piece of evidence density, and simple context (just)*
Multi-structural	3	• Students have identified a simple context. • They have correctly described two features. OR • Evidence for relational without any context.	The most common height is between 36 cm–37 cm. The heights range from 35 cm–43 cm (2011 student, pre-test). • *two correct features (mode and range) and simple context* The distribution for the scores of the AB is approx right skewed with a tail to the right. It peaks at approx 20. It seems like the 2 groups, one tightly together from 20–45 and another from 60–100. It ranges from 0–100 (2011 student, post-test). • *three correct features (shape, mode, modal clump) although range incorrect; acceptable close population and variable* • *missing the context connected throughout, i.e. apart from first sentence there is no context in rest of description, no units given; therefore is not a relational level description*

SOLO taxonomy level	Grade	Description of evidence	Example of student responses with *commentary*
Relational (see significance of parts of the whole, can sketch the graph with shape and two pieces of evidence)	4	• Students have identified the context including the variable and either the target population or acceptable close population (see Figure 8-16). • They have connected the context throughout most of the description through use of the variable and units. • They have correctly described the overall shape and at least two other features.	The distribution of the heights (cm) of NZ Yr 5–10 students is approx symmetrical and unimodal. The heights range from 110–200. There is a peak at 150. A large group is located from 130–180 cm. If another sample was taken, these plots could change (2011 student, post-test). • *four correct features (shape, range, mode, modal clump); context (variable and units) connected throughout most of the description*
Extended abstract (integrating the statistical and contextual, also can be seeking explanation or interpretation)	5	• Students have identified the context including the variable and the target population. • They have connected the context throughout the description through use of the variable and units. They have included the population more than once. • They have correctly described the overall shape and at least three other features. • They may have included some explanation or interpretation of data to the context (e.g. explanation – these two groups might mean the two different genders).	The distribution of the heights of Tokoeka kiwis is approx symmetrical and bimodal. The heights range from 35–43 cm. The middle Tokoeka kiwi height is 39 cm. The heights peak at around 36.5 and 40 cm. The heights are tightly grouped in two groups one between 36.1–39 cm and another between 39–42 cm. These two groups might mean the two different genders (2011 student, post-test). • *five correct features (shape, range, median, modes, modal clumps); context connected throughout description (variable, units and population); possible explanation given*

Figure 8-21. Criteria for a good description of distributions (Arnold, 2013, pages 224-225)

Note: *means the value(s) is given for the feature and it is the correct value(s)).

Activity

Have a look at your description and grade it using the criteria for a good description of distributions.

Use another colour and update your description to an extended abstract description.

Appendices

List of the variable and populations for the 15 graphs

Read left to right and then down as with the masters.

Number of skips in 30 secs: 2003 C@S Yr 5-8 students	Birth month: 2003 C@S Yr 5-10 students	Weight-kg: Kiwi Kapers Great Spotted Kiwi
Reaction time-secs: 2007 C@S Yr 4-13 students	Right foot length-cm: 2003 C@S Yr 5-10 students	Attendance-percentage half days: Yr 9-13 students
Hair length-cm: 2007 C@S Yr 4-13 students	Household debt-$: Synthesised Unit Record File based on NZ data	Height-cm: 2003 C@S Yr 5-10 students
AsTTle test results: Yr 9 reading	Number of years living in NZ: 2005 C@S Yr 10 students	Age-years: Everyone at a high school
Time to school-mins: 2009 C@S Yr 9-13 students	Index finger length–mm: 2009 C@S Yr 9-13 students	Cell phone ownership-months: 2009 C@S Yr 9-13 students

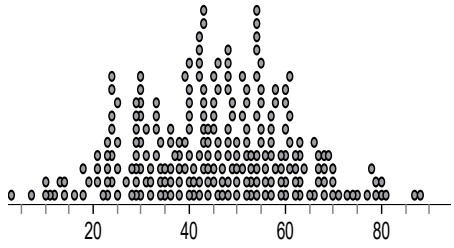
Statistical terms for describing shapes of distributions

symmetrical	bimodal
trimodal	unimodal
uniform	long tail to the right
long tail to the left	bell shaped
normal curve	right skew
left skew	negatively skewed
positively skewed	

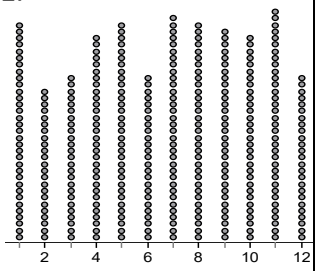
Actual graphs

for students print 2 pages per A4 sheet; for teachers enlarge to at least A3 and laminate.

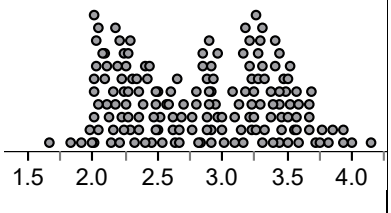
1.



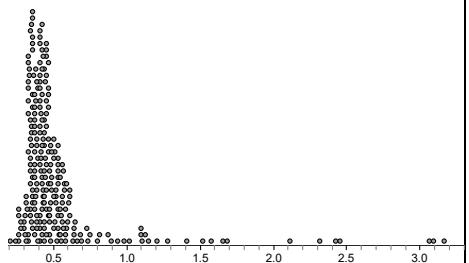
2.



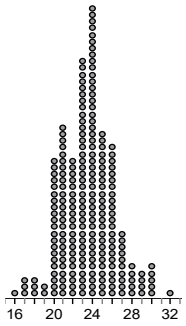
3.



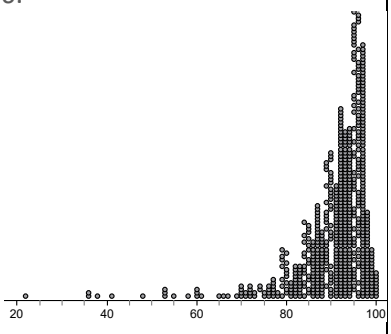
4.



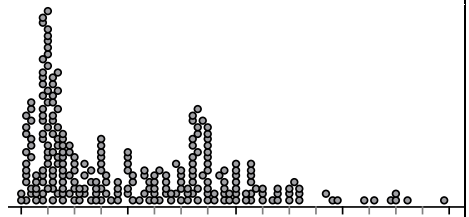
5.



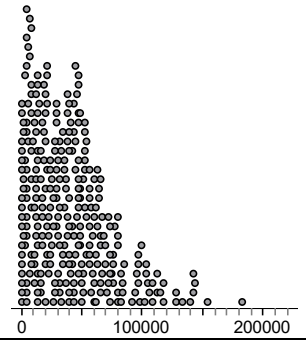
6.



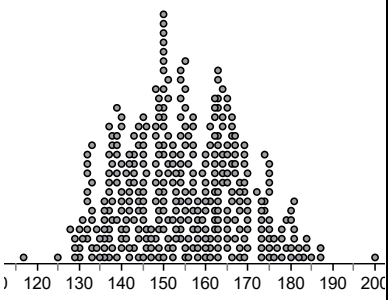
7.



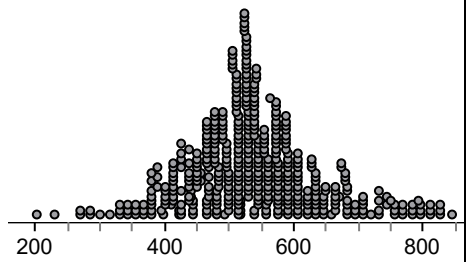
8.



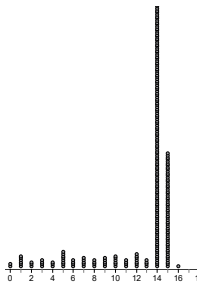
9.



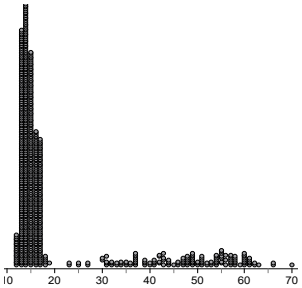
10.



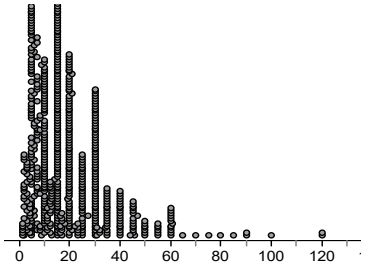
11.



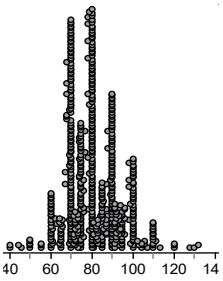
12.



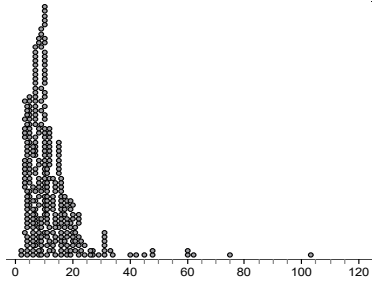
13.



14.



15.



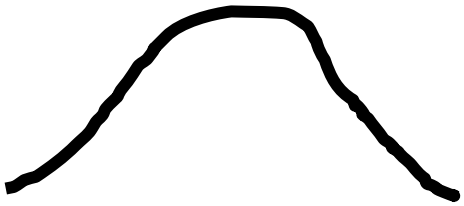
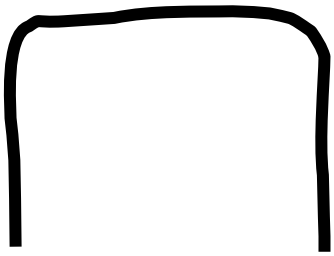
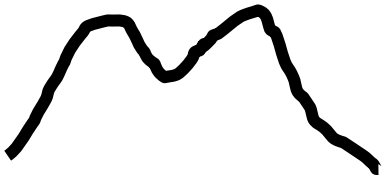
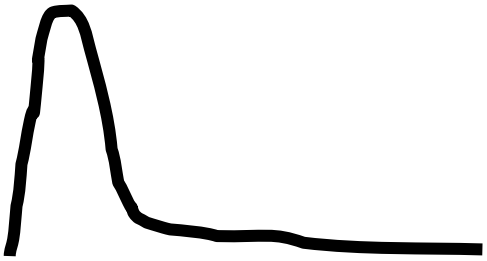
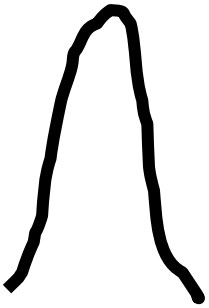
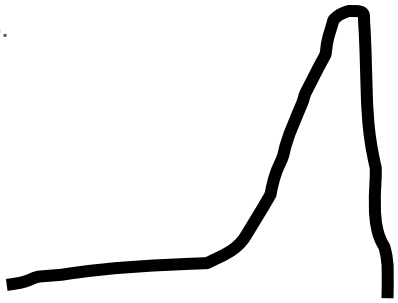
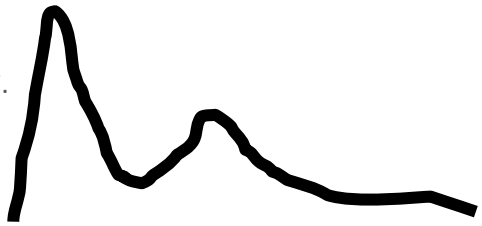
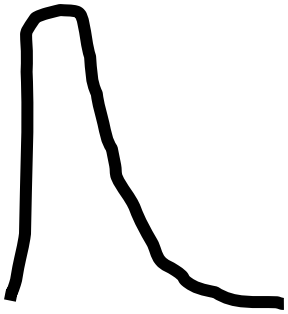
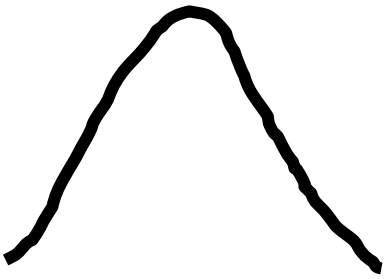
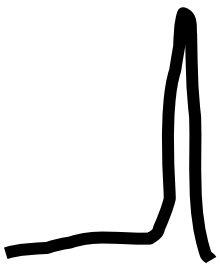
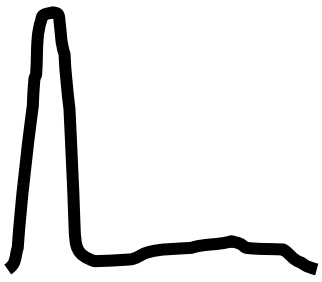
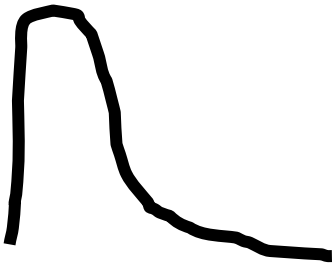
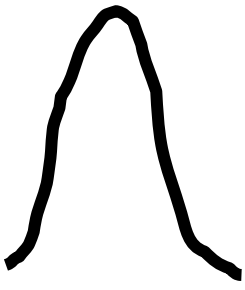
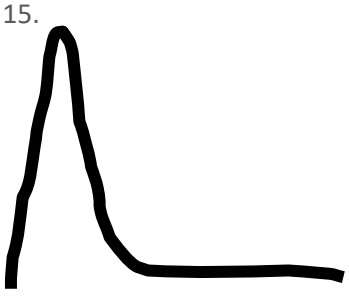
Contexts for graphs

for students to use copy 2XA4 sheet and give to pairs;

Age-years: Everyone at a high school
Number of years living in NZ: C@S Yr 10 students
Hair length-cm: 2007 C@S Yr 4-13 students
Time to school-mins: 2009 C@S Yr 9-13 students
Reaction time-secs: 2007 C@S Yr 4-13 students
Household debt-$: Synthesised Unit Record File based on NZ data
AsTTle test results: Yr 9 reading
Index finger length-mm: 2009 C@S Yr 9-13 students
Right foot length-cm: 2003 C@S Yr 5-10 students
Number of skips in 30 secs: 2003 C@S Yr 5-8 students
Attendance-percentage half days: Yr 9-13 students
Cell phone ownership-months: 2009 C@S Yr 9-13 students
Birth month: 2003 C@S Yr 5-10 students
Weight-kg: Kiwi Kapers Great Spotted Kiwi
Height-cm: 2003 C@S Yr 5-10 students

Teacher "sketched" graphs

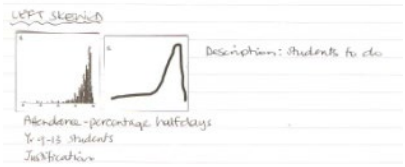
for teachers enlarge to at least A3 and laminate.

1. 	2. 	3. 
4. 	5. 	6. 
7. 	8. 	9. 
10. 	11. 	12. 
13. 	14. 	15. 

Lessons on shape

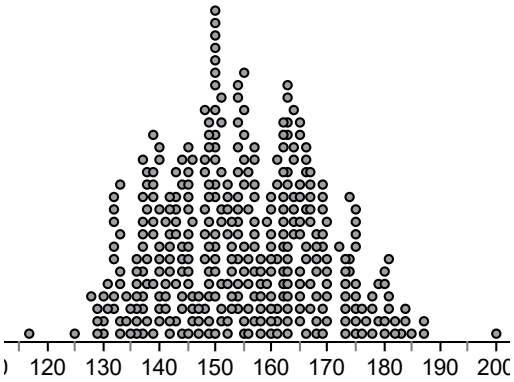
Figure H-1 is split into two columns. The first column outlines the lesson detail that was developed for the teachers to use in class. The second column describes the background to or purpose of the particular aspect of the activity. In addition the aspect of distribution that was being attended to ((1) the notion of distribution; (2) shape of distributions; (3) predicting distributions; and (4) contextual knowledge) is noted.

Lesson Number	Activity detail	Background/purpose to the particular aspect of the activity
Lesson Two	1. Students make 15 "squares" of paper from an A5 sheet. Get students to label their 15 "squares" of paper with the numbers 1-15 in the top left hand corner. This is to help with identification later on.	Organisation/preparation
	2. Using the prepared PowerPoint presentation, show each of the 15 graphs for a very short time, 1-2 seconds and get students to sketch the shape they see in the quick glimpse.	Looking at the gross shape of the data rather than specific detail. Use bigger sample sizes so that the shape was clearer. Large population/sample to help make the shape more "obvious". **(2) shape of distributions**
	3. When all 15 graphs are drawn students should check with their neighbour and compare what they have sketched for each graph. At this stage the teacher can put up the 15 "teacher" sketched shapes and they can compare against these as well.	Students to compare their sketches with the view that they might decide that one was better than the other, and also compare this with the teacher graph and what the teacher graph might offer that theirs doesn't. A more generalised shape. **(2) shape of distributions**
	4. Get pairs of students to sort one set of graphs into similar shapes. Collate responses from the class and arrive at a consensus as to which shapes are similar. Use the teacher shapes on the board.	Grouping the shapes was about trying to see the patterns that are there. Generally statistical graphs fall into a limited group of patterns, there is not an infinite number of patterns. **(2) shape of distributions** *Notes: Symmetric LS RS uniform*
	5. For each group get students to describe the shape they see using words that they are comfortable with. Note these words under each of the groups of graphs.	Starting with student language for the shapes that they see to give a foundation for the development of statistical language. **(2) shape of distributions**
	6. Introduce the statistical words used for describing graphs – teacher prepared resources. Have a good discussion with the students about what they think the different words mean both in everyday and statistical sense. Get the students to suggest which words might best go with which group of graphs.	Gave students the language to see what they do with the language, see how much of it is intuitive. **(2) shape of distributions** *Then have the conversation with at the end about what the words mean.*

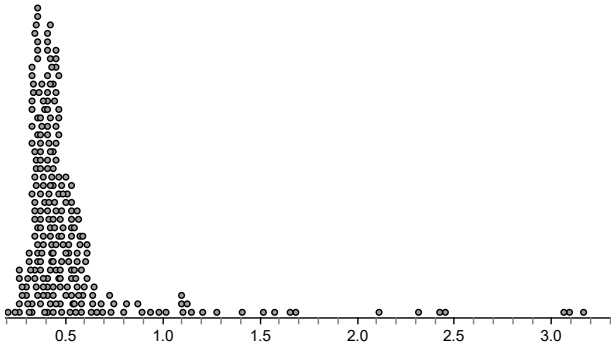
Lesson Number	Activity detail	Background/purpose to the particular aspect of the activity
Lesson Three	7. Hand out strips of graphs. Get students to cut and paste the graph and their sketch into their book under each of the description words. Allow room for the variable, justification, other examples and the description. Suggested layout below. Need about six pages in double spread. This will become a reference resource for students	 Organisation of graphs, but also to start a "library" of contexts that are similar shapes. Building their contextual knowledge library. **(4) contextual knowledge**
	8. Put up the list of variables that made the graphs. Before students match them with the graphs get them to predict what shape they think the graph of the variables will be and why. Discuss as a class. Collect ideas on the board.	Consideration of what the data might look like; want students to think about data before they sketch it. That is to get students to think about what might be sensible values for a particular variable. The prediction is also about thinking about the shape of the data and using contextual knowledge to decide on what the shape might be. Understanding when data is incorrect, cleaning data. Getting the students to start to think about the context a bit more, building their contextual knowledge. **(3) predicting distributions** *Note: E.g. right foot length, reaction time, attendance, birth months*
	9. Get students to match the context with the graph – get them to use the mix and match labels initially and record the final context in their book with their justification. Add the variable and the unit to the graph.	Organisation, but also using their contextual knowledge and information from previous activity to see what makes sense. **(4) contextual knowledge**
	10. Once this is finished get students to look back at their graphs from the previous lesson and decide which "shape" they are. Add these contexts in the appropriate space.	Organisation, but also building their contextual knowledge library for different variables with the same shape. **(2) shape of distributions** **(4) contextual knowledge**
Lesson Four	A. Review activity: Mix and match – statistical graphs and shape descriptors Resource: mix'n'match activity – shape descriptors • Students place the statistical graphs under one of the headings. There may be different numbers of graphs under each of the headings. • Add the contexts (and paste the graph) to the other examples in the work done previously.	Further work on deciding on shapes. Adds to the "library". Opportunity to check use of language, especially with skewed graphs. **(2) shape of distributions** **(4) contextual knowledge**

<table>
<tr><th>Lesson Number</th><th>Activity detail</th><th>Background/purpose to the particular aspect of the activity</th></tr>
<tr><td rowspan="3"></td><td>B. Describing distributions

Discuss with students what key features of a graph to describe are.

● Put the challenge out if they had to draw the graph from the description only what info would they need. Collect ideas from the class.
○ Suggest may be: shape, description of range, median/centre, middle group, and peak(s) – there may be other features, discuss as a department first.</td><td>About developing what makes a good description.
(1) notion of distribution
(2) shape of distributions
(4) contextual knowledge</td></tr>
<tr><td>● Model for #9 and #4. Model this process for the students.
○ Talk out loud your thinking and get them to contribute.
○ Eg. What shape is the graph? Write the first sentence explaining the use of approximately and the use of the variable and the group we are talking about.
○ What values do the heights range from and to?
○ Write the next sentence and so on. The questions should be around the features you decided on with the class.
○ Remember to include the CONTEXT. Variable, values and units.
○ Use active reflection that is making descriptions correct and complete.</td><td>Modelling the thinking process for students when writing a description, also modelling the language to use, including the intertwining of the context throughout the description including especially the variable, values and units.
(1) notion of distribution
(2) shape of distributions
(4) contextual knowledge</td></tr>
<tr><td colspan="2">Examples:
#9 Graph is: heights in cm of Yr 5-10 students
 

The distribution of heights for these year 5-10 students is approximately symmetrical and unimodal. The heights range from 116cm to 200cm. The median height is about 155cm and the middle group of heights is between 142cm and 167cm.</td></tr>
</table>

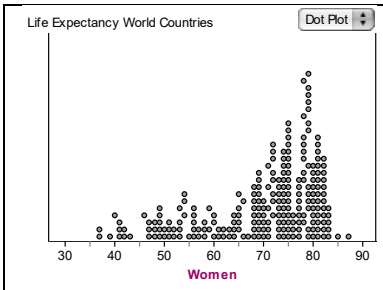
<table>
<tr><th>Lesson Number</th><th>Activity detail</th><th>Background/purpose to the particular aspect of the activity</th></tr>
<tr><td></td><td colspan="2">#4 Graph is: reaction times in secs of yr 4-13 students
 

The distribution of reaction times for these yr 4-13 students is right skewed. Nearly all of the reaction times are tightly bunched between 0.2 and 0.6 secs. There are some reaction times slower than 0.6 secs and they spread out to 3.15 secs. The graph of reaction times peaks at about 0.4 secs and is approximately symmetrical between 0.2 and 0.6 secs.</td></tr>
<tr><td></td><td>● Students to do rest of the descriptions for homework, one per night over the next few weeks. Review these at the beginning of the following lesson, remembering to model good practice (see above).</td><td>To continue to develop their descriptive skills over the whole unit of work, to keep the focus in this area and provide plenty of practice at writing descriptions, a new skill to be developed.
(1) notion of distribution
(2) shape of distributions
(4) contextual knowledge</td></tr>
</table>

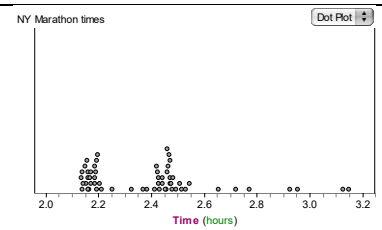
Figure H-1. Detailed lesson planning for teaching experiment four (Arnold, 2013)

Mix'n'match activity

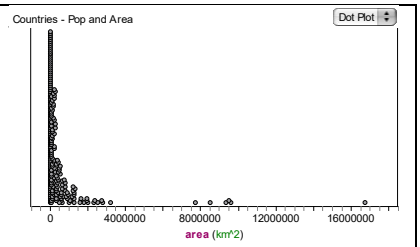
copy one between 2 and cut up



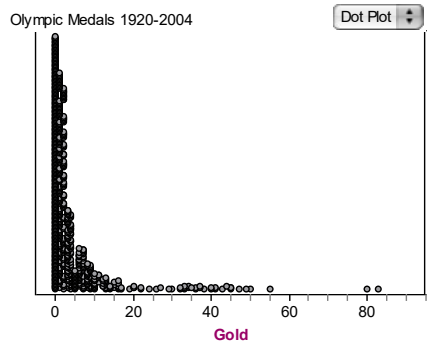
Life expectancy for women at birth for different countries



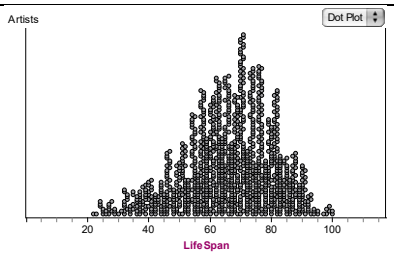
New York Marathon winning times



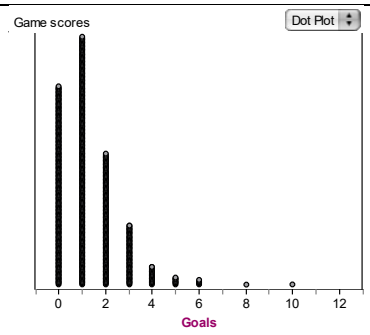
Land area of different countries



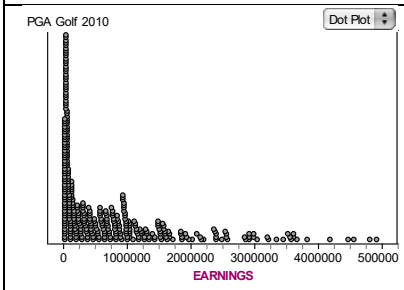
Number of gold medals by country at the Olympics from 1920-2004



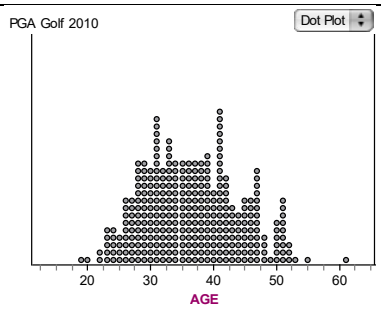
Life span in years of 1146 artists



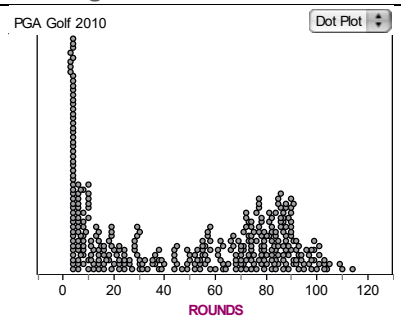
Number of goals scored during standard play time in World cup soccer games from 1978-2002



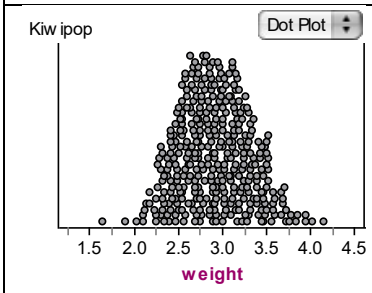
Earnings – top 360 golfers 2010 PGA tour



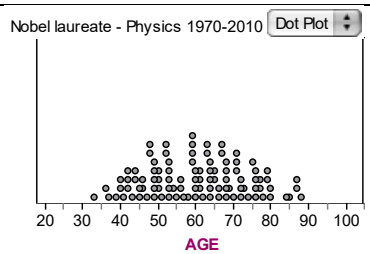
Age of golfers - top 360 golfers 2010 PGA tour



Number of rounds of golf played – top 360 golfers 2010 PGA tour



Weight in kg of female kiwis – Kiwi kapers



Age in years of Nobel laureate in Physics 1970-2010

Unemployment rates for males and females in different countries

Shape descriptors

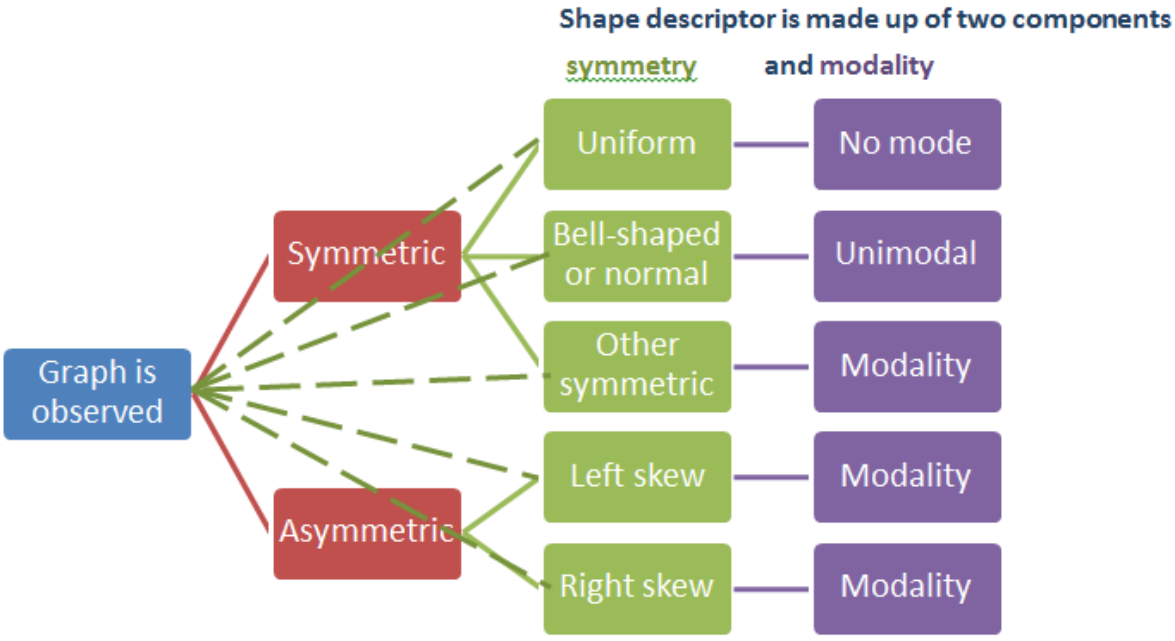
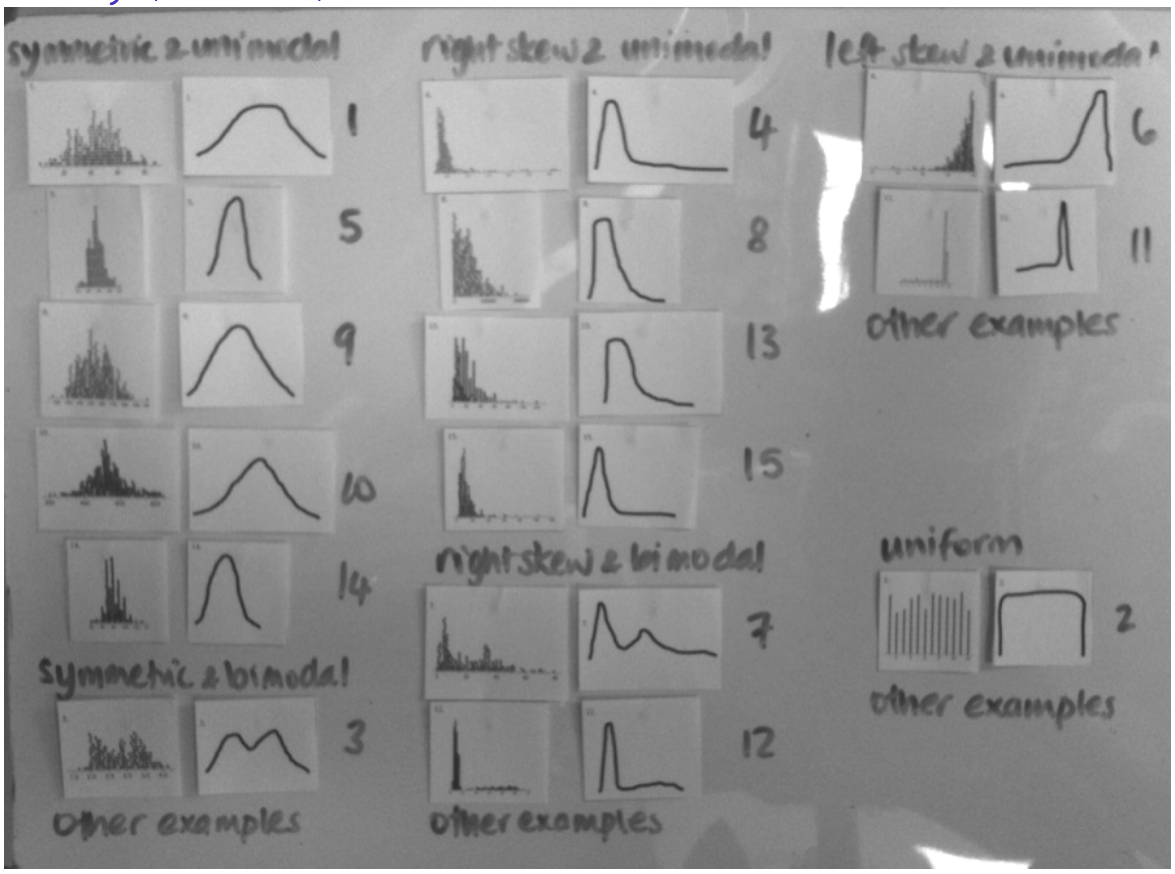


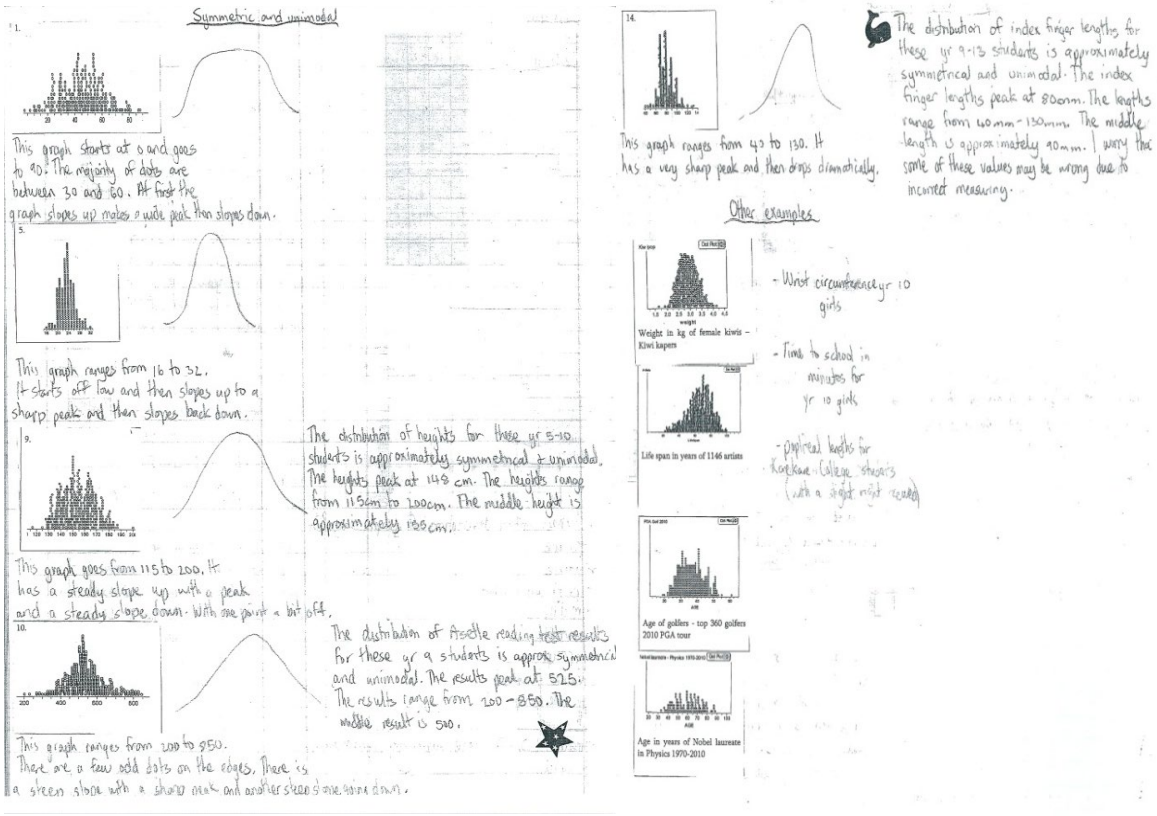
Figure 8-14. Components of shape descriptors

Arnold, P. (2013). *Statistical investigative questions: An enquiry into posing and answering investigative questions.* Doctoral thesis https://researchspace.auckland.ac.nz/handle/2292/21305

Sorted graphs and shapes



Example of student work – building a context library



Dot plot

http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Glossary/Glossary-page-D#dotPlot

A graph for displaying the *distribution* of a *numerical variable* in which each dot represents a value of the *variable*. For a *whole-number variable*, if a value occurs more than once, the dots are placed one above the other so that the height of the column of dots represents the *frequency* for that value.

Dot plots are particularly useful for comparing the distribution of a numerical variable for two or more categories of a *category variable*; this is shown by displaying side-by-side dot plots on the same scale. Dot plots are particularly useful when the number of values to be plotted is relatively small. Dot plots are usually drawn horizontally, but may be drawn vertically.

Example

The actual weights of *random samples* of 50 male and 50 female students enrolled in an introductory statistics course at the University of Auckland are displayed on the dot plot below.