Mathematics in the New Zealand Curriculum
Second Tier

Strand: Statistical Thinking Thread: Statistical Investigations Level: Three

Achievement Objectives: Conduct investigations using the statistical enquiry cycle by:
- Gathering, sorting and displaying multi-variate category and whole number data, and simple time series data to answer questions;
- Identifying patterns and trends in context, within and between datasets;
- Communicating findings, using data displays.

Exemplars of student performance:
Exemplar One: The Subject of Exercise
The student (C) gets a set of 40 completed data cards from other people in her class. Each card has this information:

Name:
Gender:
Age:
Height in centimetres:
Favourite subject:
Sports:
Favourite role model:
Hours of exercise taken last week:

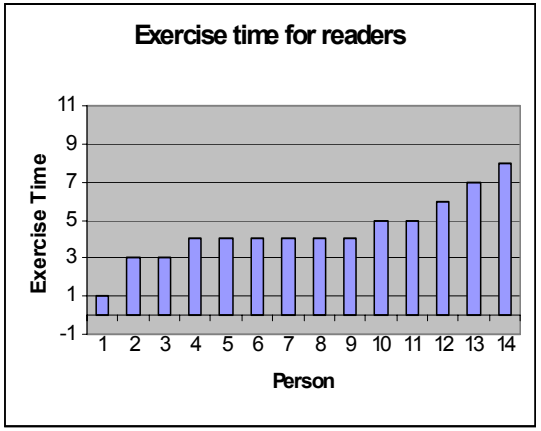
She groups and regroups the cards according by establishing criteria for the variables,
e.g. Height by these criteria; short (less than 130 cm), medium (between 130 cm and 150 cm), tall (over 150 cm).
Role model by type; sportsperson, singer/musician, TV or movie actor, politician, family/friend, etc.
Sports/Hobbies by these criteria; Rugby, Soccer, League, Netball/Basketball, Hockey, Athletics, Swimming, Cricket, Tennis, Cycling, Skating/Skiing/Surfing, other, none.

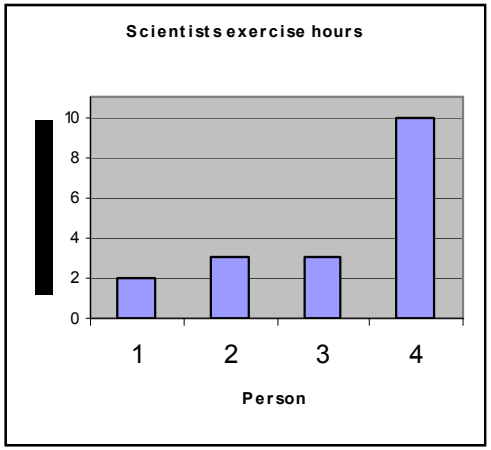
C generates a set of questions to investigate, like;
1. Do boys have different favourite subjects to girls?
2. Do younger students have different role models to older students?
3. Do tall students play different sports to medium and short students?

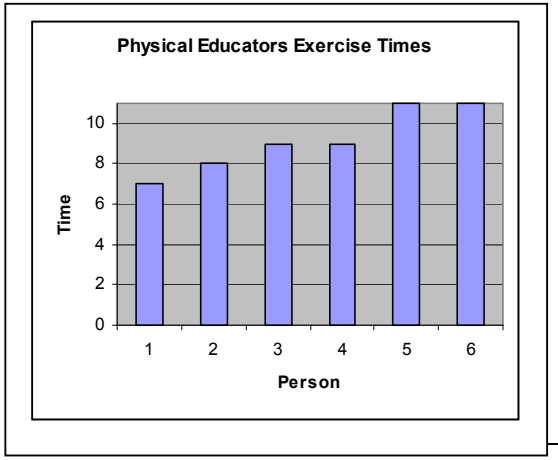
She selects this question to investigate, "Do students with reading as their favourite subject take less exercise than those students who prefer other subjects?" She sorts the cards by favourite subject then records the number of hours exercise for the students in each subject grouping, using a table.

Favourite Subject	Hours exercise taken in last week
Reading	4, 3, 5, 1, 7, 4, 4, 5, 8, 4, 3, 6, 4, 4
Mathematics	5, 3, 4, 7, 6
Science	2, 3, 3, 10
Music	1, 6
Physical Education	10, 8, 7, 9, 11, 9
Art	4, 3
Other	2

C uses computer software to order the scores in each category, creates graphs to seek patterns and selects bar graphs to highlight the differences between groups. She produces a bar graph for each favourite subject group. This requires her to ensure each vertical axis has the same scale. She aligns the graphs as shown below.







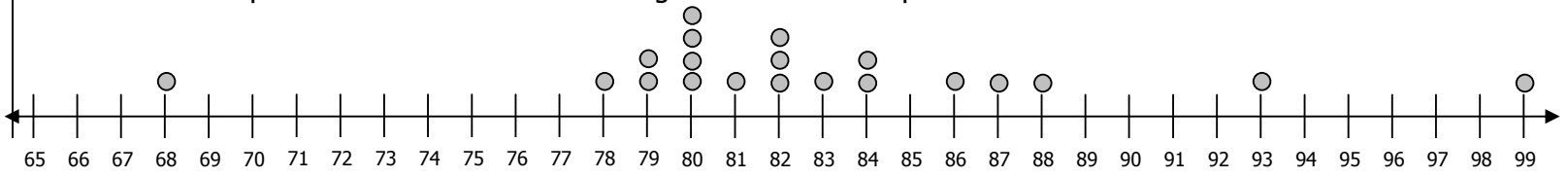
C concludes that students who prefer Physical Education do more exercise each week than other groups, students who prefer Science do the least exercise, and that most students who prefer Reading do about around four hours exercise each week.

This exemplifies Level Three because C has sorted and resorted data to find relationships. The dataset is multivariate and the student establishes relationships between exercise time (one variable) and the category of favourite subject. He/she compares multiple displays (bar graphs) to establish the relationships.

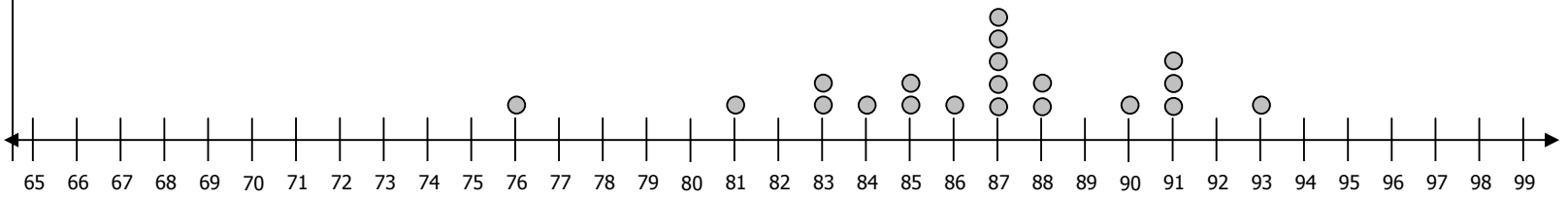
Exemplar Two: Fast Fries

The student (K) poses this question, "Which fast food store gives the best value for packets of fries?" He considers a number of measures for "best value" including price of packets of different sizes, weight and length of the fries, and the number of fries in a packet. The student notes that the prices are the same and that measuring fries will make the data gathering very inconvenient. He opts for number of chips as the measure of value.

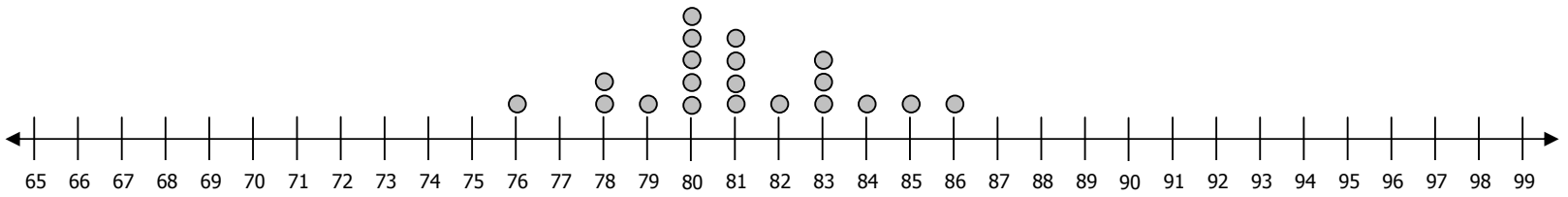
He asks classmates to count the number of chips they get in their large packets when they eat at the mall. This involves getting data on twenty packets from each of the McDougals, Burger Queen, and Tennessee Fried Chicken stores. As the data arrives over a two-week period he collates the data using an individual dot plot for each outlet



Number of chips in McDougal's Large Packets



Number of chips in Burger Queen's Large Packets



Number of chips in TFC's Large Packets

K concludes that the number of chips in a large packet varies a lot, even from the same store. He also suggests that there are usually more fries in a large packet from Burger Queen than the other two stores, though sometimes this may not be true. In reflecting on his data gathering K maintains that number of chips is a good measure of value, noting that the variation in chip size seemed similar for each store.

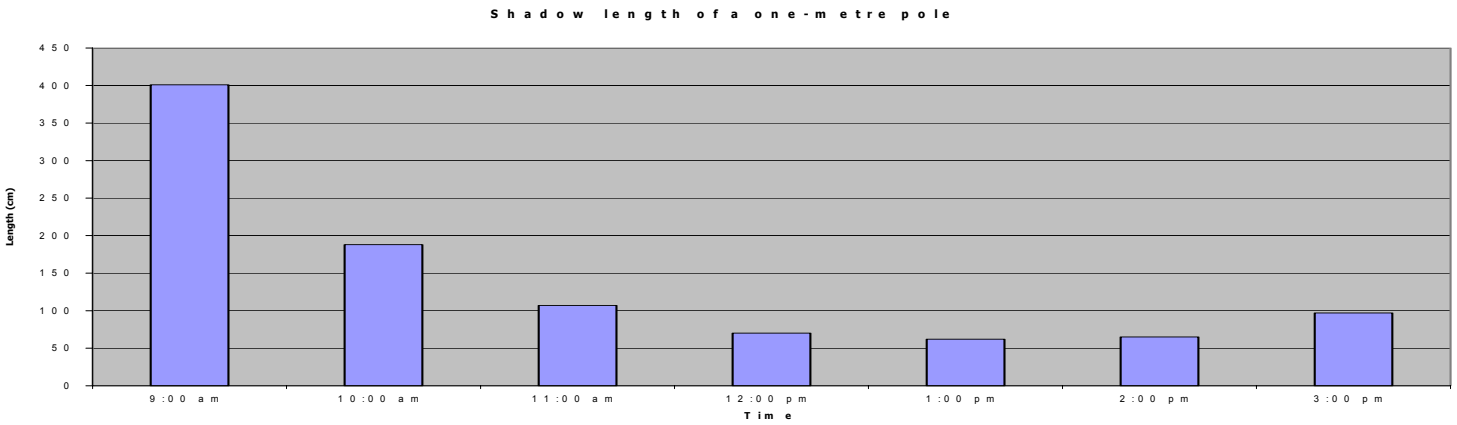
This exemplifies Level Three as K has identified patterns within and between the datasets from each fast food outlet. He uses appropriate data displays (dot plots) to show the discrete numeric data (number of chips). K uses these displays to compare how the data from each store is distributed and forms a conclusion based on the data.

Exemplar Three: Shadows of Time

The student (L) carries out an experiment as part of her Science Fair project about the sun. This experiment involves measuring the length of shadows cast by a one-metre pole at different times of the day. She decides to measure every hour and records the results in a table.

Time	9:00 am	10:00 am	11:00 am	12:00 pm	1:00 pm	2:00 pm	3:00 pm
Shadow length (cm)	401	188	107	70	62	65	97

From her table L notices that the shadow length decreases then increases and decides to graph the data using a bar graph:

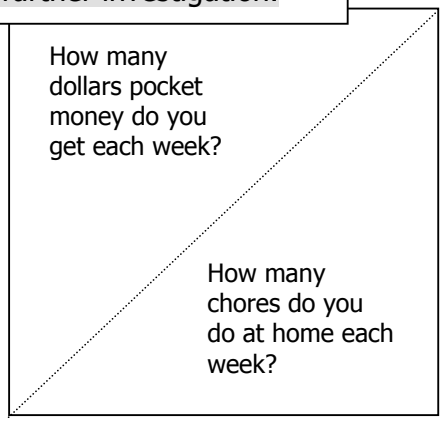


L suggests that the shadow length is longest early and late in the day because the sun is not very high in the sky and that short shadows occur at the hottest time of the day when the sun is highest. She wonders if the pattern would look the same for the other four seasons and if her friend in Darwin would get similar data.

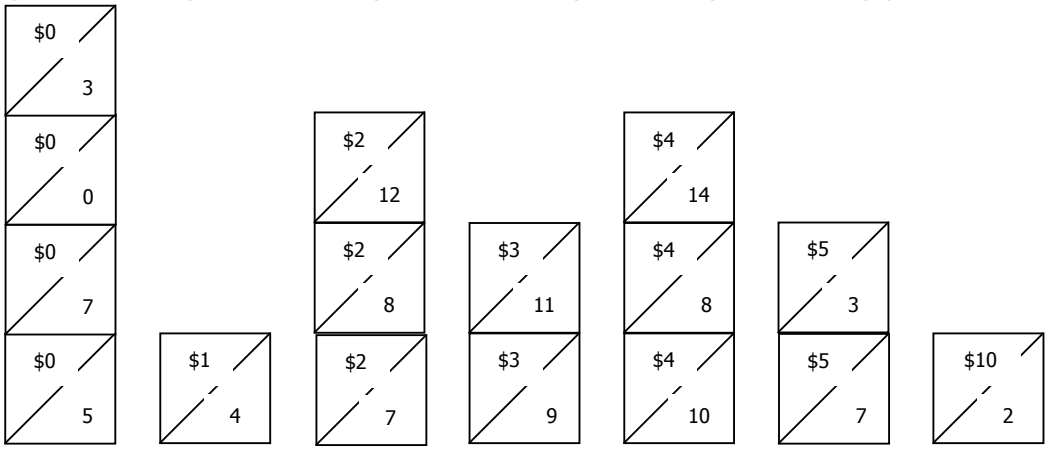
This exemplifies Level Three because L is finding trends in simple time series data (data gathered over time). She is able to detect a cyclic trend in the data and uses an appropriate bar graph to communicate her findings. L shows increasing independence in the Statistical enquiry Cycle by asking "I wonder" questions for further investigation.

Exemplar Four: Time is Money

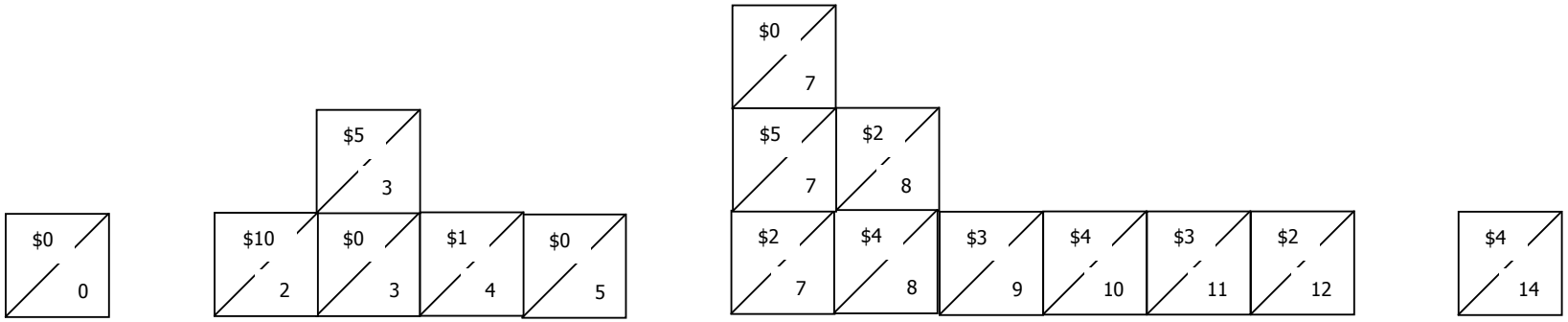
Students (M and N) investigate this statement, "People who do more chores at home get paid more pocket money."
They survey all their classmates asking them to answer two questions on a sorting card or post-it note. M & N find that the chores question is ambiguous so they make up a list of common chores, like doing dishes, making the bed, folding washing, to help their classmates to answer the question.



They sort the response cards by one variable (univariate), at first, by pocket money:



M & N notice that their classmates got between zero and ten dollars pocket money per week and that zero dollars pocket money is the most common. From this graph they notice that their classmates in the "middle" seem to do the most chores and wonder if this is true. The students confirm the relationship by resorting the cards, this time by chores:



M & N conclude that their classmates who get two, three, or four dollars in pocket money tend to do the most chores. They also notice that students who get no pocket money or a lot of pocket money do not do many chores and wonder if the number of children in the family has an effect on chores and pocket money.

This exemplifies Level Three as L and M are identifying patterns in bivariate data (number of chores and amount of pocket money). They sort and resort the data and display it in bar graphs to detect the relationships between these variables. Their conclusion relates directly to the original question and the data gathered, and they seek to explain the results in context.

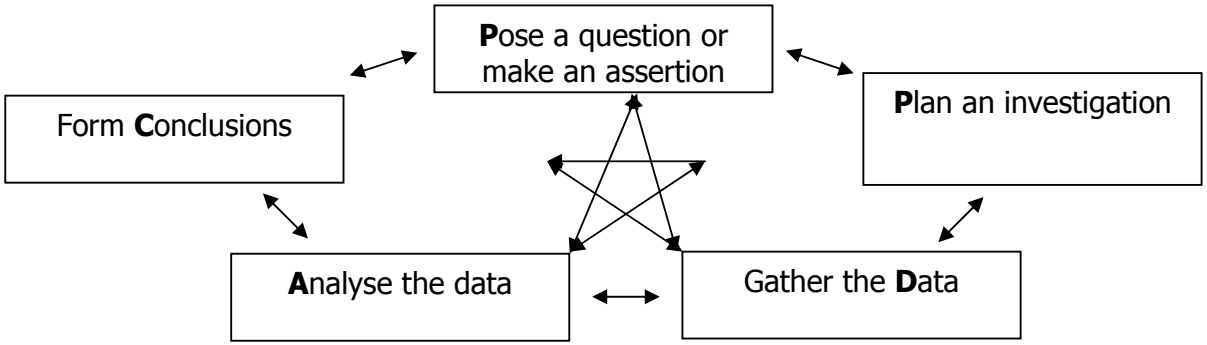
Important teaching ideas (working at):

Students at level three are learning to investigate multi-variate datasets rather than restricting their investigations to univariate data. Multi-variate data contains many variables while univariate data has one. For example, if data was gathered about students' age, gender, height, usual bedtime, and pulse rate, this would create a multi-variate dataset that involved five variables. It would be possible to find relationships between the variables, like "Are boys taller than girls?" or "Do older students have later bedtimes than younger students?" If data was only collected about pulse rate, this would be univariate data.

Some of the data in the set given above was measurement data, like height, while some of the variables were categoric. Category data involves belonging to a collective group, for example with gender students are either male or female, and for eye colour they have blue, green, brown, or hazel coloured eyes. The data on age can be termed discrete numeric data as only whole number values were possible, e.g. Age 10 years.

Students at level three need to have ways to process category, discrete measurement (numeric) and simple time series data. Time series data is data that is gathered over time, like the air temperature at different times of the day. The measurements are whole number units like time at every hour and temperature in a whole number of degrees.

As with all levels of the statistical investigations thread the emphasis is on students implementing the statistical enquiry cycle with increased sophistication. This cycle contains five connected steps, as shown below:



Posing Questions

Students at level 3 must be posing their own questions and forming their own assertions. They should be doing so through either investigating an existing dataset, as with www.censusatschool.org.nz, or by creating the need to gather data that is appropriate by asking a question or making an assertion. Both data driven and question driven types of investigation are valid. As the emphasis is on multi-variate data the questions can be of three main types:

1. Summary questions, e.g. What is the height of an ten year old person?
2. Comparative questions or assertions, e.g. "Are eleven year old girls taller that eleven year old boys?"
3. Relationship questions, e.g. "Are people's foot-lengths related to their handspans?"

Lessons need to capitalise on the interests of their students in providing contexts that will engage them. Sport, entertainment, families, food, travel, and animals are good examples of such contexts. Students should also be encouraged to pose questions that will involve measurement data rather than restricting themselves to category data. Science investigations frequently involve measurement data", e.g. "Do heavier marbles roll further than lighter marbles?"

Planning an investigation and Gathering the data

Key questions for students to ask themselves in planning an investigation from a question or assertion are:
What information will help me answer this question?
Who/what shall I ask/measure to get this information?
What questions will I ask? or What measurements will I take?
How will I organise the information as I collect it?
It is important for students to predict difficulties in the data collection phase. Typical examples of this are:

a. ambiguity in the question being asked, e.g. How many people are in your family?
b. complexity of the raw data, e.g. What is you favourite television programme?. Be aware that removing the complexity by creating categories of response can destroy potentially useful information.
c. appropriate measures, e.g. What measure should be chosen for goal shooting ability? [This is not formally required until level four]
d. missing data and duplication, i.e. What systems will ensure that each person is asked and only asked once?
e. security of the data so it is not lost or tampered with.

During the data gathering phase students need to be systematic in accounting for the whole sample. For example, in surveying their class they may use a class list to ensure that every student is asked once and only once. Students need to develop recording methods to process the data as it is gathered. These methods will most commonly involve using tally charts and/or frequency tables, for example:

Left-handed	Right-handed	Ambidextrous
\\\	𝍸 𝍸 𝍸 𝍸 \\\	\

Left-handed	Right-handed	Ambidextrous
3	23	1

Note that tally charts and tables are data displays in their own right. Some graph types allow the data to be organised and displayed as it is collected. In particular dot plots, stem and leaf graphs, and line graphs (see below) provide a

picture of trends and patterns as the data accumulates.

Analysing the Data

One focus at level three is the use multi-variate datasets, that is, datasets that have contain many measures of attributes. For example, a multivariate dataset of students in Room 4 might include their age, gender, height, pulse rate, favourite sport and television programme, and their normal bedtime. Students at level three need to develop increasing independence in considering how measures or frequencies are spread and clustered. This is spread and clustering is known as distribution. With multi-variate datasets distributions can be compared in simple ways, e.g. "Do boys goto bed later than girls?"

The ability of students to sort and resort the data becomes more significant than in previous levels. Resorting implies that a student can reclassify the data items by selecting a different variable whilst looking for relationships between variables. For example, the data on a class of students might be sorted by gender, then resorted by age, then by height. This resorting might prompt observations of pattern, e.g. "The tall students seem to play mostly rugby, basketball and netball in the winter.", or further questions, e.g. "Why do taller students play rugby, basketball and netball?"

Computer technology makes this sorting and displaying efficient. Software such as Tinkerplots™ and MS Excel™ allows students to redefine the variables and makes sorting a rapid process. Students must attend to patterns detected through the resorting and decide what displays are most appropriate in detecting patterns.

There is a greater emphasis on discrete numeric at level three. Discrete numeric data is made up of whole numbers that come from counting and measuring, e.g. "How many beats does a heart make in one minute?", or, "What is the length of a shadow in centimetres?" Technically measurement data is continuous. That means that the measures can take up any value in a range, e.g. Z's height is 1.64 metres. At level three the measurements are taken in whole numbers, e.g. 164 cm, so the values are always rounded. Level Three is the point where students explore simple time series data, that is data gathered over time, e.g. Temperature at different times of the day.

A progression from level two is for students to identify patterns and trends. A pattern may exist within a dataset and can be about comparative frequencies or proportions, e.g. "A lot more people are right handed than are left handed" (frequencies), or, "About one tenth of people are left handed (proportional)". At level three the emphasis is on detecting frequency comparisons with a view to helping students to reason proportionally. Note the connection between the development of multiplicative thinking in the number and algebra strand and the ability to reason proportionately in statistical analysis.

Patterns also exist between distributions. Students need to compare distributions by creating separate graphs and looking between the displays. For example, suppose two classes compare their travel times to school (in minutes) and present the results in two stem and leaf graphs.

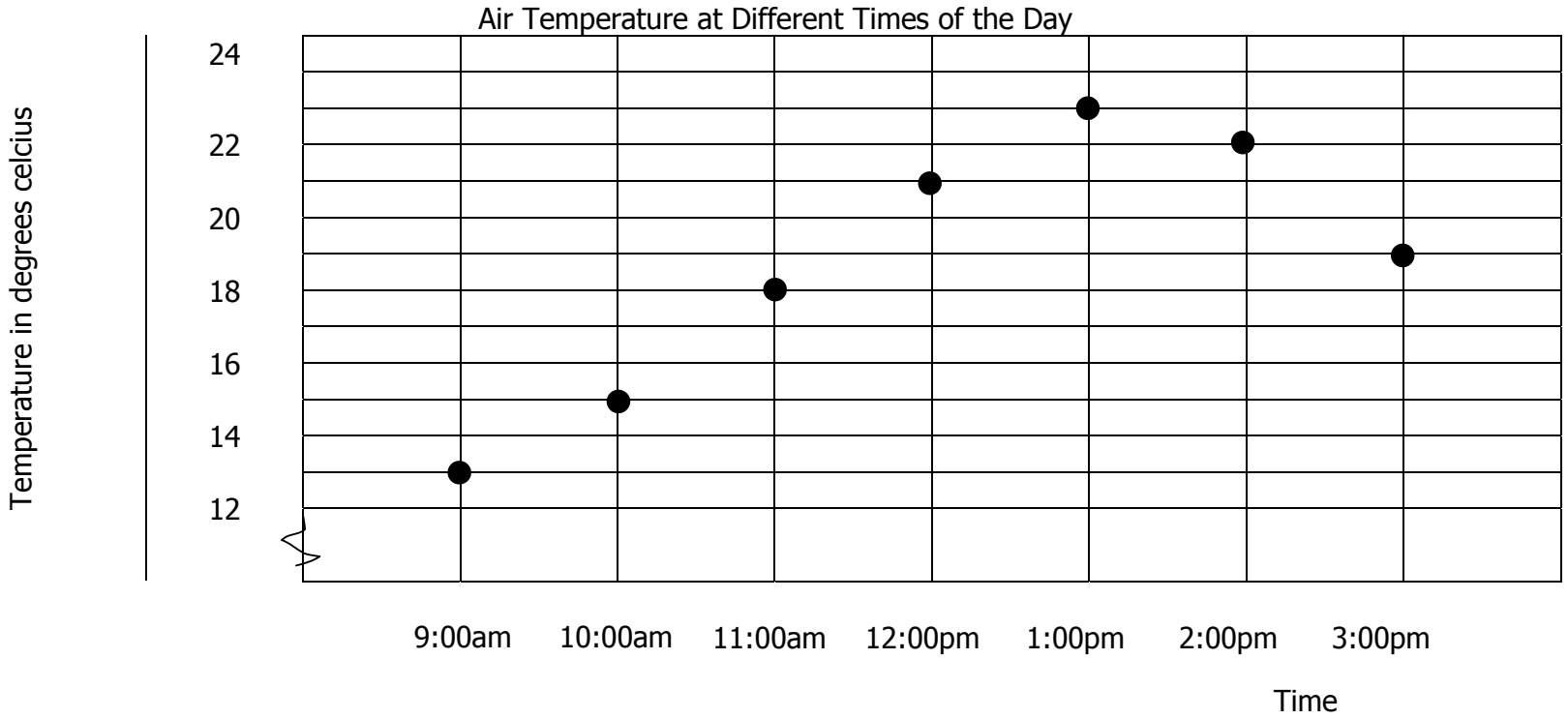
Travel Times to School

0	1	2	5	5	8	9		**0**	4						
1	0	0	0	1	4	5	9	**1**	2	6	8	9			
2	0	5						**2**	0	2	3	5	5	5	
3								**3**	0	5	7				
4								**4**	0						

7 year olds 11 year olds

Looking within the graphs the students notice that the main clustering of travel times for 7 year olds is in the teens of minutes whereas for 11 year olds it is in the twenties of minutes. Looking between the graphs students might notice that older students take longer to get to school. They might wonder what methods of transport are being used that might explain this pattern.

Trends usually apply to time series data. Trends are patterns within the dataset that allow prediction of measurements outside the dataset. For example, a student measures the air temperature at each hour of the school day, and plots the data on a graph:

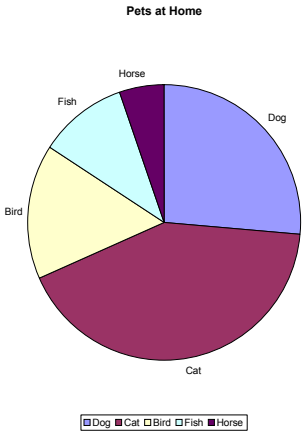


A broad description of the trend is that the temperature increases steadily until 1:00 pm then begins to decline. Interpreting a trend also involves making within and outside predictions, e.g. The temperature at 9:30 am was about 14°C and at 5:00pm is predicted to be about 15°C. Within interpretation is known as interpolation, outside interpretation is known as extrapolation.

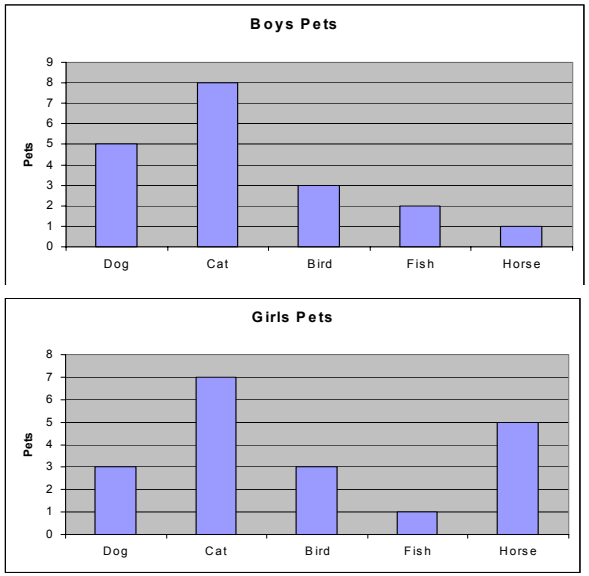
It is important that students at level three broaden the display types they have available and choose critically about how best to display data for the purposes of analysis and communication. The most common types of display are:

Category data: bar graph (comparing samples by related bar graphs), strip graphs, pie charts, tally charts, two-way tables.

Strip Graph



Pie Chart



Comparitive Bar Charts

Tally charts…

	Right-handed	Left-handed	Ambidextrous
Female	☺☺☺☺☺☺☺☺☺☺ ☺☺☺☺☺☺	☺	☺
Male	☺☺☺☺☺☺☺☺☺☺ ☺☺☺☺	☺☺☺☺	

leading to Frequency Tables

	Right-handed	Left-handed	Ambidextrous
Female	**16**	**1**	**1**
Male	**14**	**4**	

Pie Charts and strip graphs are excellent for showing proportions while tally charts and bar graphs highlight differences in frequency (number in each category). Frequency tables are a more abstract representation.

Numeric data: Stem and leaf graphs (Two different groups)

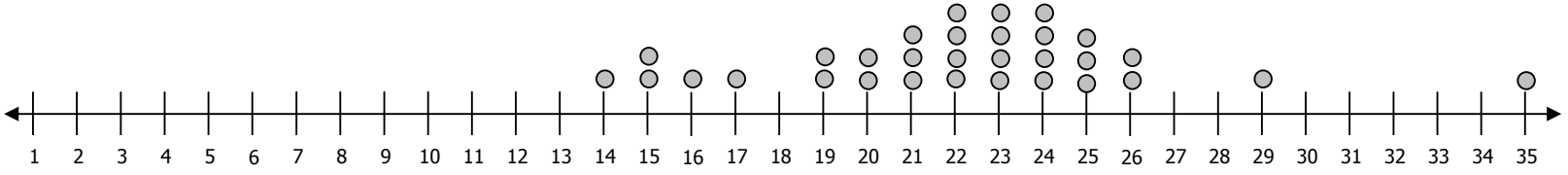
Pulse Rates

5	3							**5**								
6	0	4	5	7	8	8	9	**6**	4	7	8	9				
7	1	2	4	6	6			**7**	0	2	2	3	4	6	8
8	0	4						**8**	0	0	1				
9	0							**9**							

Teachers | Students in Room 6

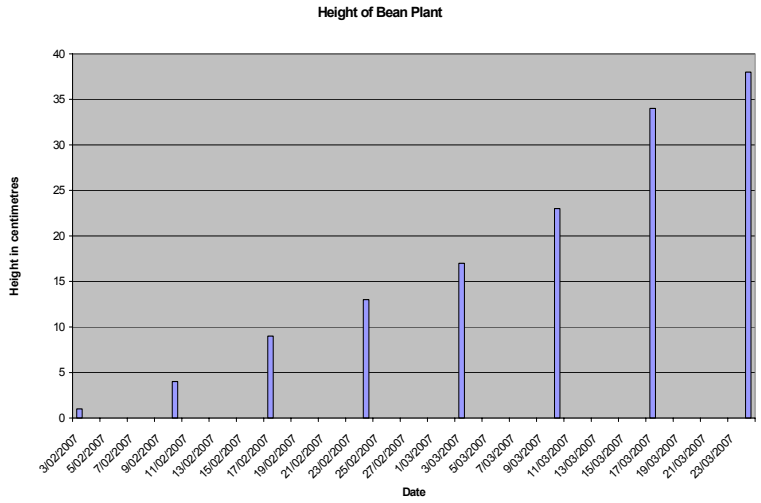
Dot Plot

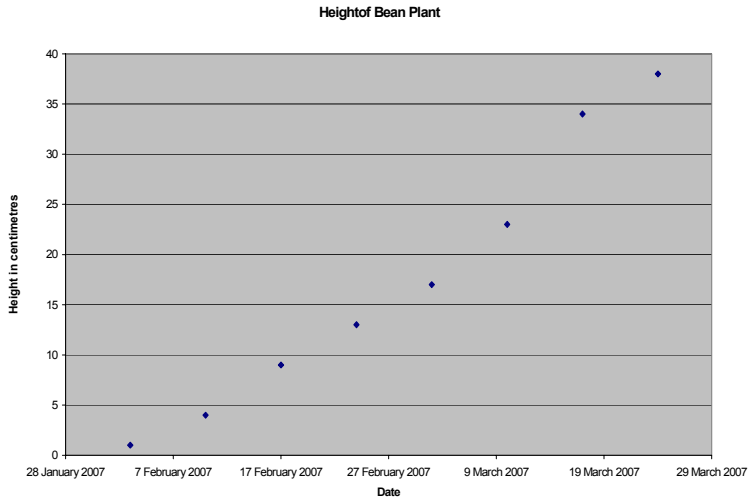
Weight of Collie Dogs at Show



Weight in Kilograms

Time series: Bar Graphs or Number Planes





Note that in both displays the data is in whole number form meaning that there are discrete (individual) data points.

Forming Conclusions

It is important at this phase of the cycle that students recognise that they are communicating their conclusions to an audience. Students need to consider who might be interested in their findings and consider this audience when reporting. Some outline of the statistical process used is important background for any audience. Documentation of the questions

posed (or assertions made), and the data collection and analysis methods is useful for the audience.
At this level it is expected that students will make a critical choice from the data displays they have available. This involves selecting the display that best communicates the patterns and/or trends they have identified. Students should also report what "I wonder…" questions have resulted from their investigation that may start the investigation cycle again. This can involve hypothesising about why the patterns and trends occur.
It is expected that students at level three will use the correct terms for graph types, give graphs a suitable title, choose sensible whole number scales, and label the axes appropriately. It is not expected that they will use statistical measures like mean, median, and range to communicate their findings until level four at the earliest.

Useful resources
Figure It Out, Level 2-3 Statistics, Pages 1-16
Figure It Out, Level 3 Statistics, Pages 1-16
[Comprehensive Teacher notes are provided for each student book. These notes have been distributed to schools and can also be accessed through http://www.tki.org.nz/r/maths/curriculum/figure/index_e.php]
Numeracy Project Book 9: Teaching Number through Measurement, Geometry, Algebra, and Statistics, pages 41-52.
nzmaths.co.nz units:
http://www.nzmaths.co.nz/node/155 (Statistical Investigations: Sports)
http://www.nzmaths.co.nz/node/156 (Statistical Investigations: Fridge Pickers)
http://www.nzmaths.co.nz/node/157 (Statistical Investigations & Length: Paper Planes L3)
http://www.nzmaths.co.nz/node/158 (Statistical Investigations: Carrots)
http://www.nzmaths.co.nz/node/159 (Statistical Investigations: Planning an Investigation L3)
http://www.nzmaths.co.nz/node/160 (Statistical Investigations: Data Squares L3)

CensusatSchool (This website is sponsored by Statistics New Zealand and The University of Auckland)
http://www.censusatschool.org.nz/
Digital Learning Objects (These are accessed through the Ministry of Education Digi-Store and are the result of a collaborative project run by The Learning Federation, Australia)
http://www.nzmaths.co.nz/learningobjects/317/3
Other Website links:
http://illuminations.nctm.org/WebResourceList.aspx?Ref=2&Std=4&Grd=0
http://peabody.vanderbilt.edu/depts/tandl/mted/Minitools/Minitools.html
http://nlvm.usu.edu/en/nav/category_g_2_t_5.html