Exploring worlds through data at curriculum level 6

Sample to population inference

Statistics teachers day | 3 December 2021

He pai ake te iti i te kore

A little is better than none

Presenting team

Pip Arnold - Karekare Education, Auckland

Robyn Headifen - Auckland

Zac Rutledge - Te Kura, Dunedin

Michelle Dalrymple - Cashmere High School, Christchurch

[Mark Hooper - St Peter's, Cambridge]

Sophie Wright - Mt Roskill Grammar School, Auckland

Rachael Ouwejan - Albany Senior High School, Auckland

Marina McFarland - Auckland Girls' Grammar School, Auckland

Maxine Pfannkuch - The University of Auckland

Lucy Edmonds - Green Bay High School, Auckland

Plan for today

Session 1:

1-1 Starter activity - PPDAC cycle

1-2 Context analysis

1-3 Interrogating secondary datasets

Session 2:

2-1 Samples and sampling variability

2-2 Visualising and describing data

2-3 Answering the investigative question

2-4 Wrap up

Materials

Teacher workbook

- Outline of activities
- Links to all the online activities
- Links to other resources
- Key research findings

Link provides you with a copy to save and add your own ideas into.

Workbook: Sample to population inference

1-1 Starter activity

1-1 Starter activity

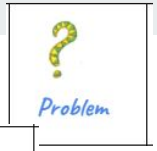
Introduce yourselves, appoint a recorder

PPDAC cycle - page 2 of workbook

Work on the scenario in your breakout room - all have the same theme of bullying

Think about and record how you use the PPDAC cycle for the scenario

- Rooms 1, 4, 7 etc (divide by 3 rem 1) will work with scenario 1
- Rooms 2, 4, 8 etc (divide by 3 rem 2) will work with scenario 2
- Rooms 3, 6, 9 etc (divisible by 3) will work with scenario 3



We are..

Drag this Text box into positio
Overtype to add your descrip
what is happening at this stag
PPDAC cycle for this scenario

Names: ***type your group members names here***

Scenario: In preparation for Bullying-Free week (in May) the PE and Health department have decided to have Bullying-Free as a theme for their current year 10 unit. They have invited the mathematics and statistics department to work with the same year 10 students to undertake a statistical enquiry into bullying at their school and report their results to the school board of trustees.

Students will need to collect data to do this statistical enquiry

INSTRUCTIONS

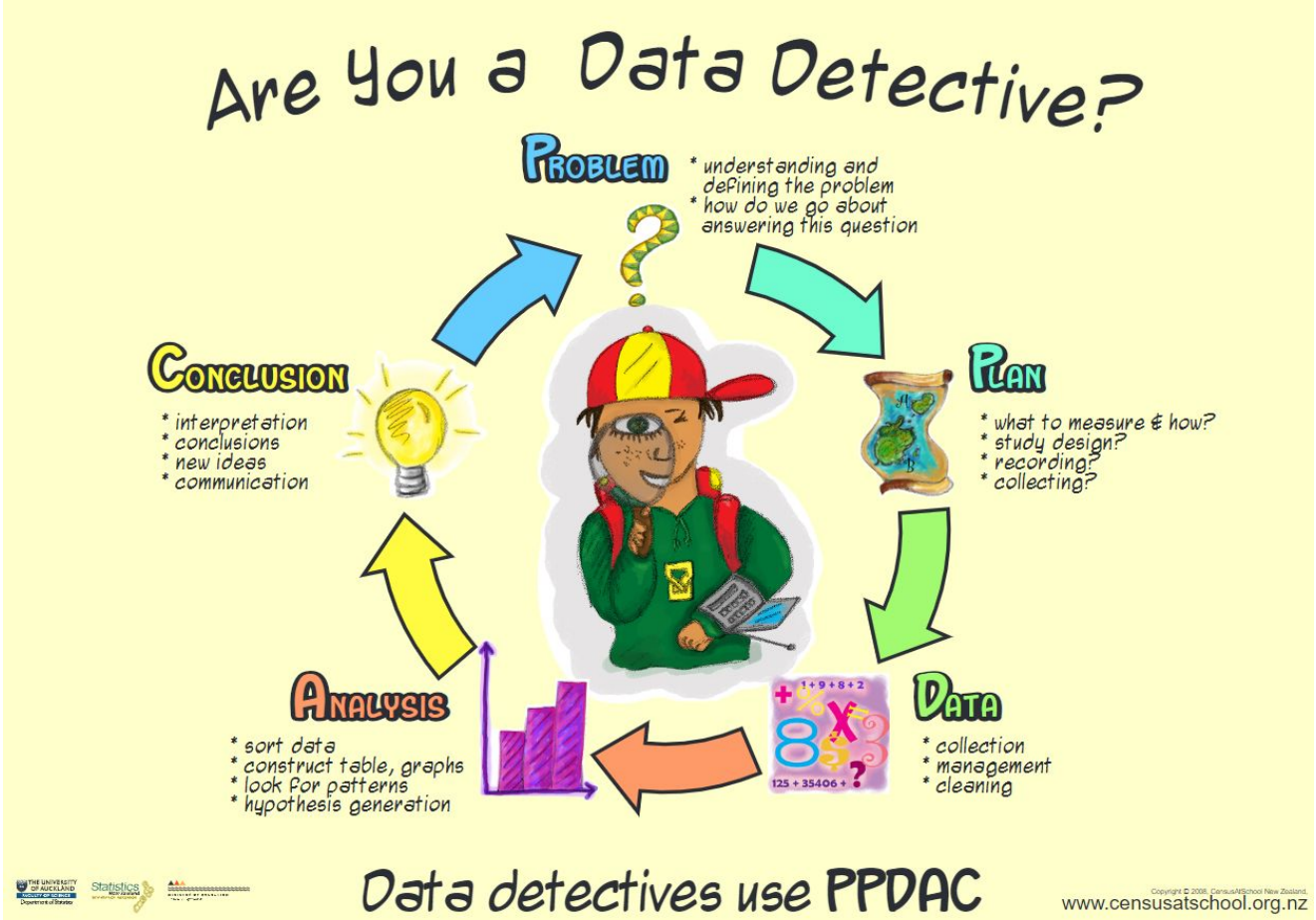
- Introduce yourselves & appoint a recorder
- Add your group names to the box on the left
- Drag the PPDAC cycle cards to reflect how the PPDAC cycle might be used for your scenario given.
- Using the text boxes, describe what is happening in each phase of your PPDAC cycle

1-1 Wrap up

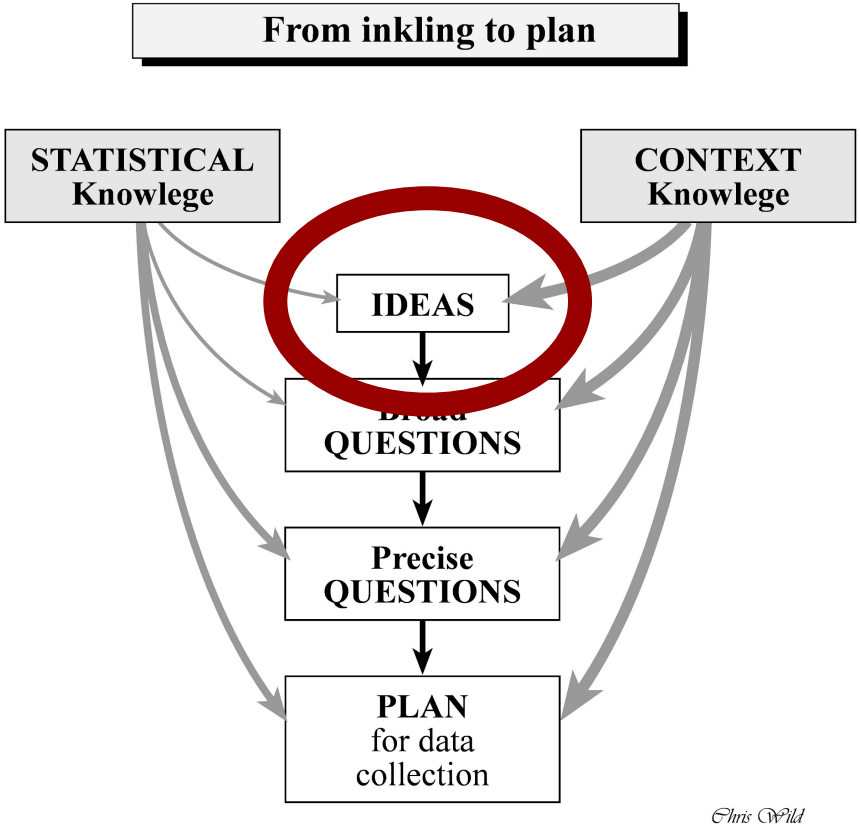
Students may collect, be given or source data to answer an investigative question. The investigative question might come first, or it might evolve through a process of sourcing suitable data to explore a topic of interest.

1-2 Context analysis

Working with primary data

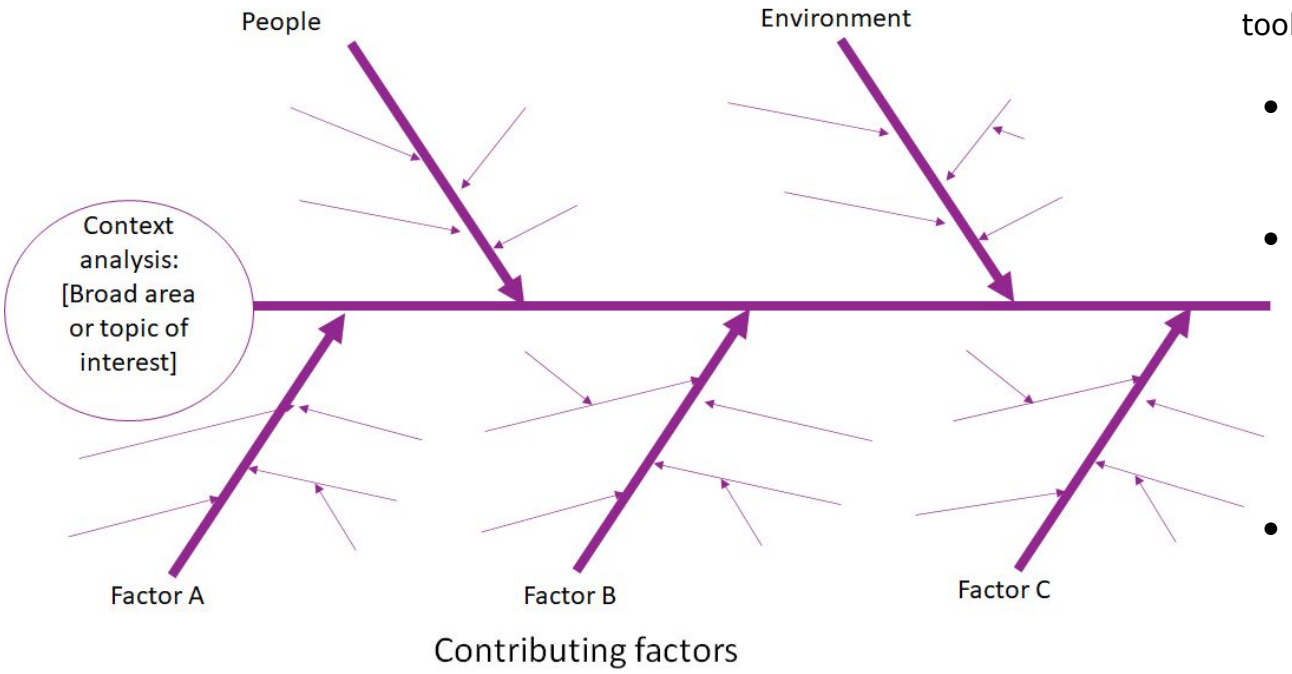


http://new.censusatschool.org.nz/resource/data-detective-poster/ and
http://new.censusatschool.org.nz/resource/te-tuhuratanga-tauanga-he-toa-tuhura-koe/



From inkling to plan (Wild & Pfannkuch, 1999, p. 228)

Context Analysis tool



Created by Pip Arnold and Chris Wild
In the process of writing a book about statistical investigations (Arnold, Trinick, & Pfannkuch, in press).

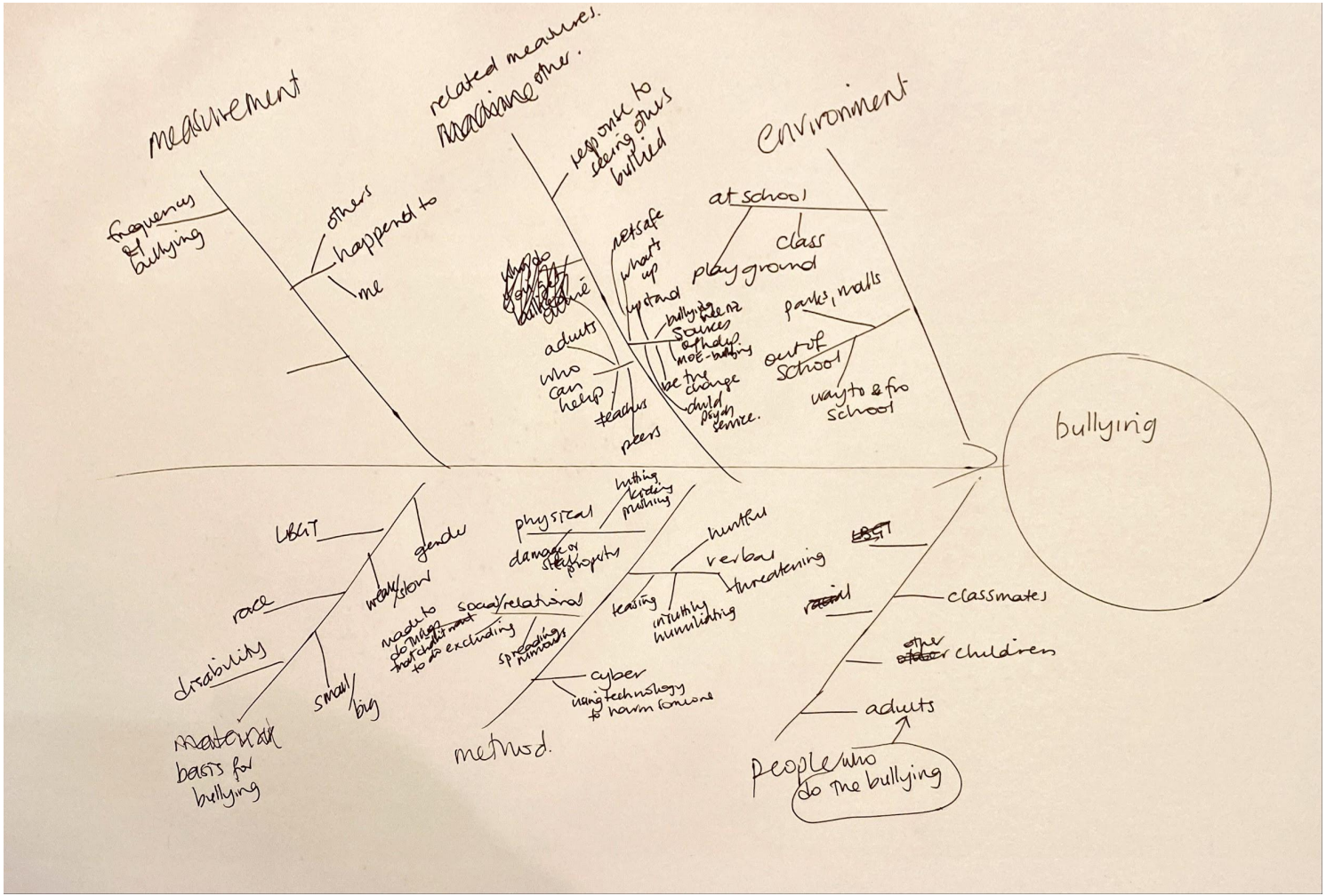
To undertake context analysis using this tool:

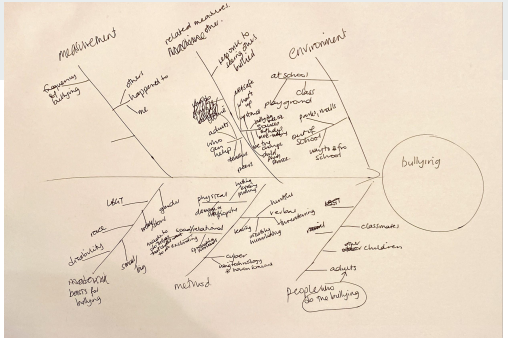
- Identify the broad area or topic of interest.
- Provide suggested starter headings, these can be added to and/or changed. *People and environment are given as two possible factors to explore*
- Add other headings to the diagram that consider contributing factors for the area of interest.
- With the main headings in place, brainstorm ideas and add to the diagram making links where possible.

Modelling the use of this tool

The teacher

- asks the students what things they wonder about bullying in school
- uses the context analysis tool to capture student thoughts about the topic, adding contributing factors (main headings) to the diagram.
- records student ideas under appropriate factor



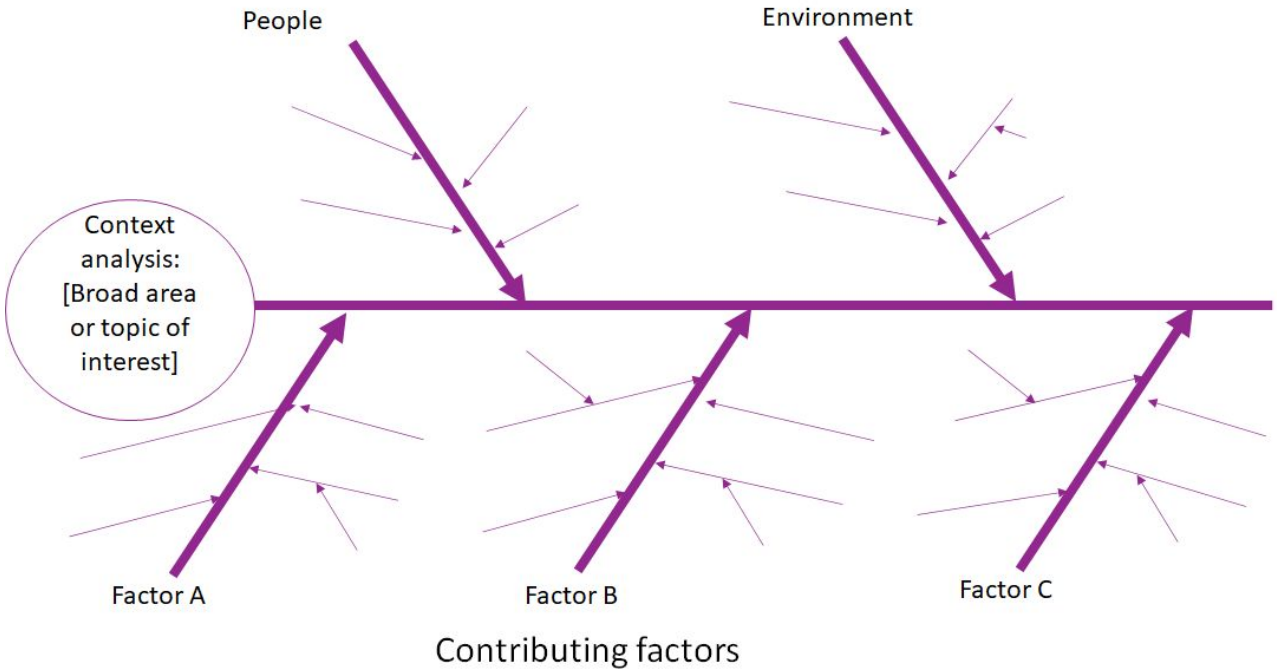


Another view: Bullying

People
• Adults
• Other children
• classmates
		Environment
		• At school
		  ○ Playground, class
		• Out of school
		  ○ Parks, malls, way to and from school

Method of bullying • Physical, hitting, kicking, pushing • Verbal, teasing, hurting, threatening • Cyber • social/relational	**Basis for bullying** • LBGQT • Race • Disability • Gender • weak/slow at doing work • small/big (in size)	**Other** • Frequency of bullying • Who did it happen to, me or others • Response to seeing others bullied • Who can help when someone is bullied

Context Analysis tool



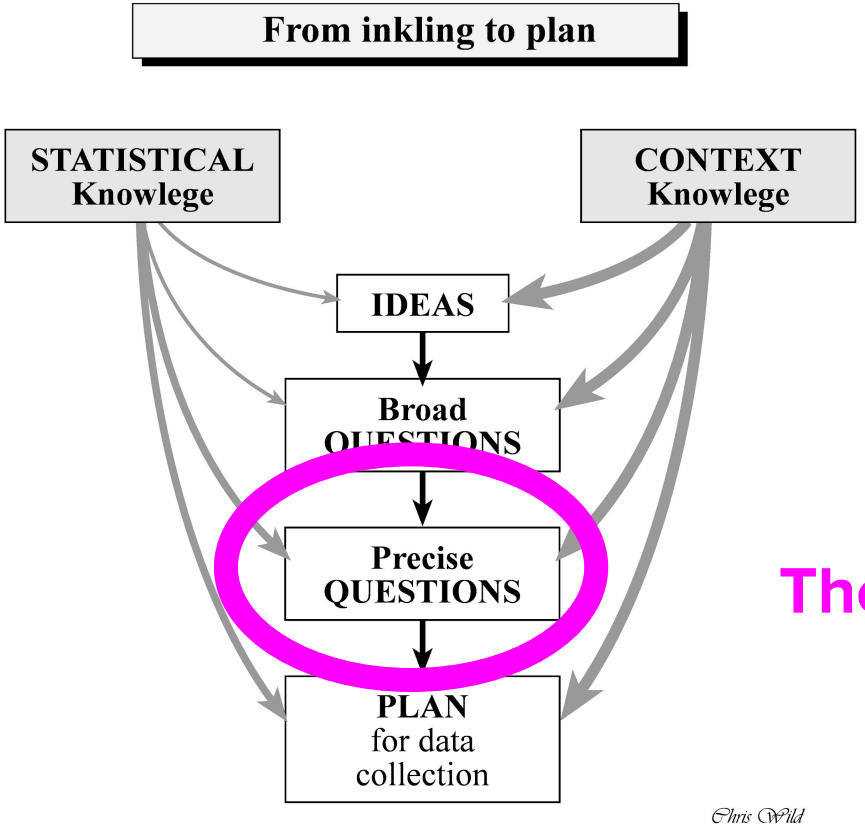
- **Alcohol & Smoking**
- **Climate change**
- **Healthy lifestyle**
- **Natural resources**
- **Rubbish & pollution**
- **Our own Topic**

Created by Pip Arnold and Chris Wild

In the process of writing a book about statistical investigations (Arnold, Trinick, & Pfannkuch, in press).

Breakout rooms

1. Introduce yourselves to your group members
2. Links are page 3 in the workbook
3. Appoint a recorder to complete this table
4. Working together start thinking about people and environment then consider other factors
   [replace the headings Factor A,B,C with your groups chosen headings as they come to light]
5. Brainstorm ideas classifying ideas into columns as you go



The investigative question

From inkling to plan (Wild & Pfannkuch, 1999, p. 228)

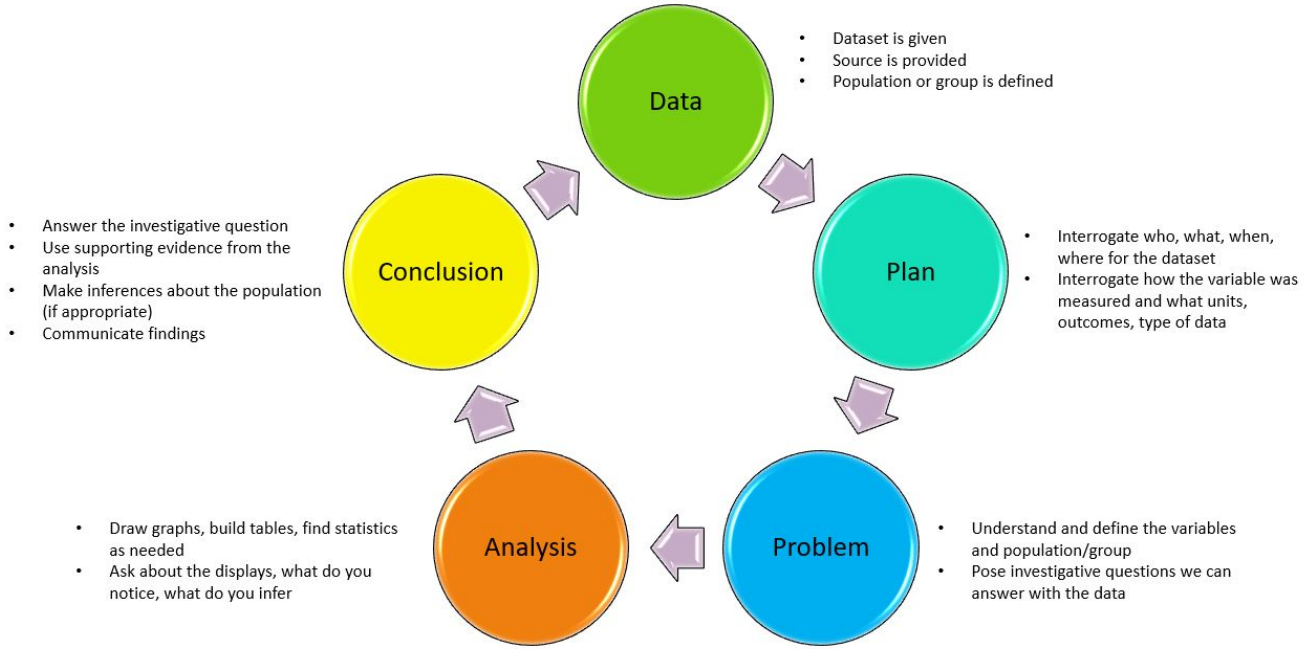
1-2 Wrap up

Students should get to experience collecting data, using the context analysis diagram can support this being a richer experience for students.

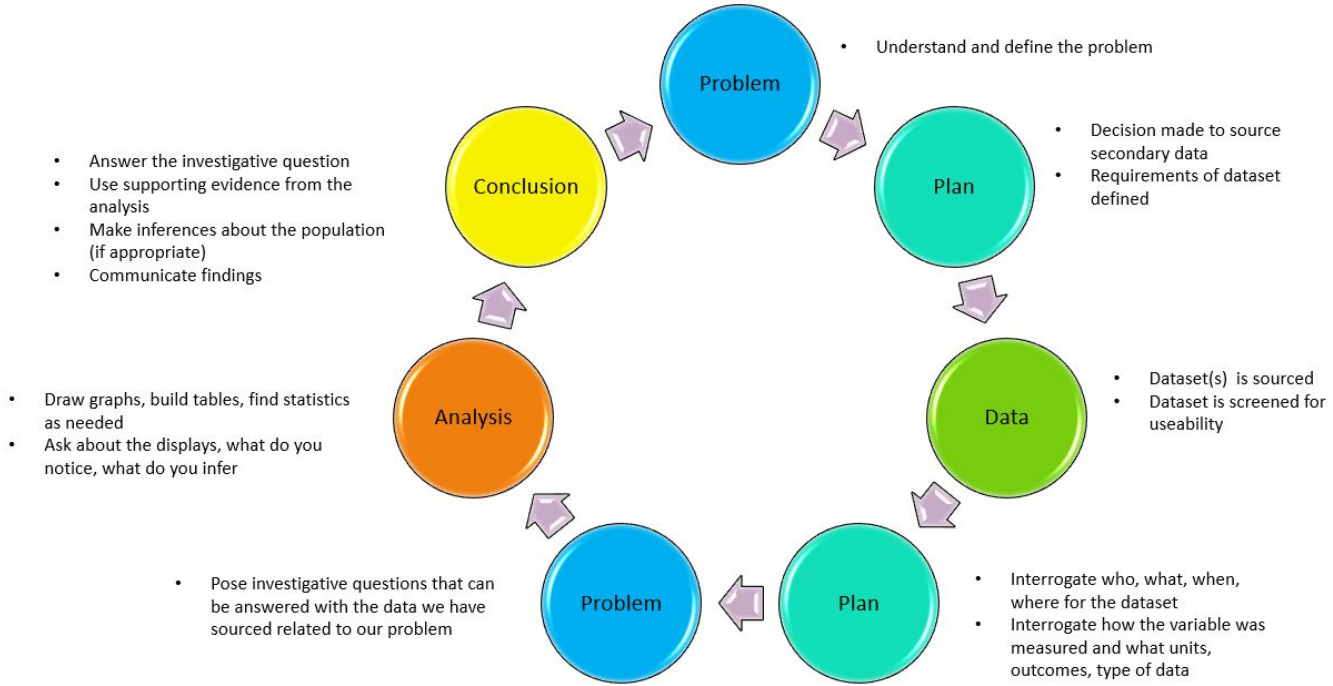
From the context analysis diagram, ideas become broad questions, which are fine tuned to become precise questions - investigative questions. The investigative question drives the rest of the cycle.

1-3 Secondary datasets

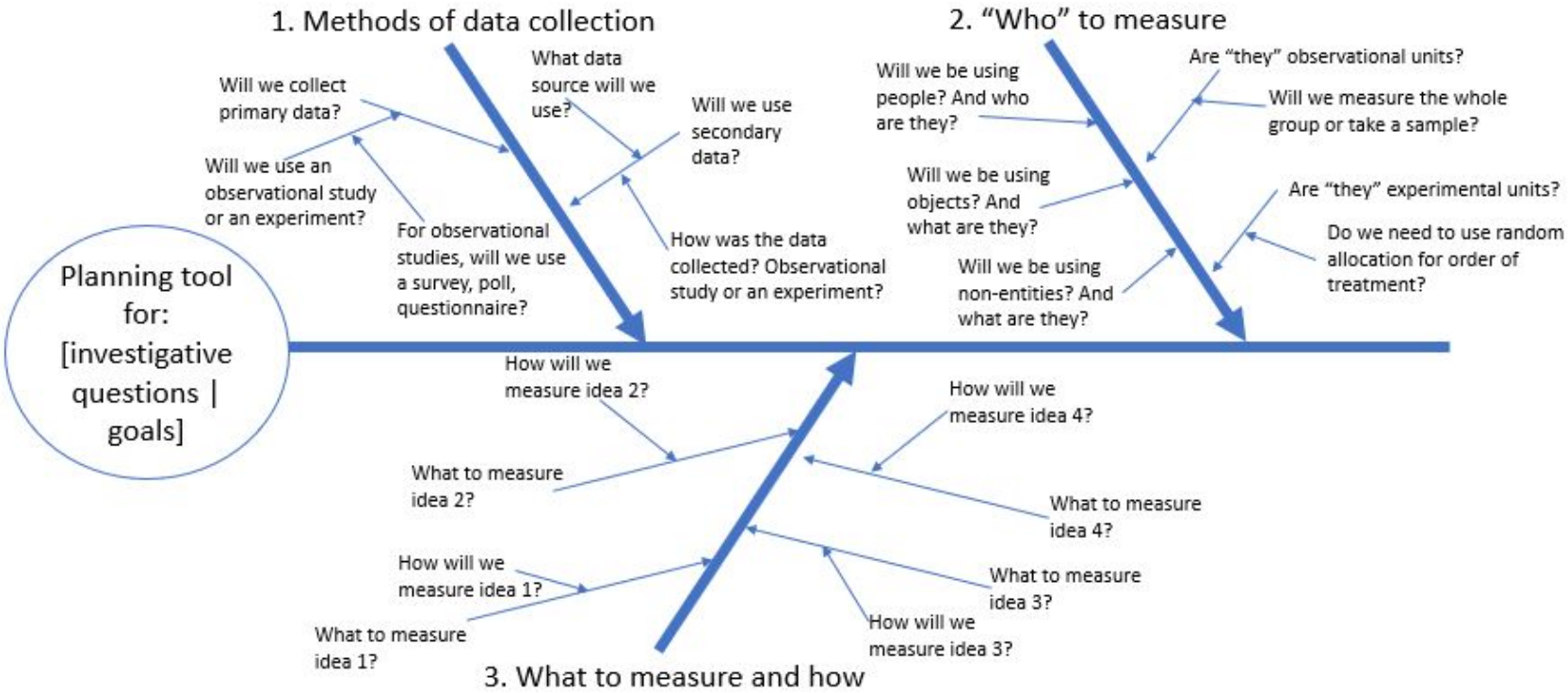
PLAN/DATA



The statistical enquiry cycle for provided datasets (adapted from Arnold, 2013, p.22)



Shuttling backwards and forwards through the statistical enquiry cycle when starting with an overall problem, sourcing data and then fine tuning the investigative question(s) (Arnold, Trinick, & Pfannkuch, in press)



Created by Pip Arnold and Chris Wild

In the process of writing a book about statistical investigations (Arnold, Trinick, & Pfannkuch, in press).

Page 4 in workbook

About the dataset: Climate change

Interrogative questions	Our responses
Was the data collected using an observational study or an experiment (from CL5)? (1. Method)	Observational
Who was the data collected from? (2. “Who”)	NZ students
Who collected the data? (1. Method)	Census at school
When was the data collected? (1. Method)	Vague - 2007-2015
Where was the data collected? (1. Method)	NZ Census at School online survey
What was the purpose for collecting the data? (Initial investigator’s problem/purpose)	Opinion on current local and global issues from students

About variables (3. What and how)

Write your names here:

State the variable	Variable name	Variable name	Variable name
What was the data collection or survey question asked to collect the data?	Slide scale - Region in NZ	How important in global warming to you?	Year level
How was the variable measured?	Provinces/Region	Slide scale	School year level
What are the units, if any, for the variable?	N/A	rating	Year level
What are the possible outcomes for the variable?	Region Names	-100 to 100	5-13
What type of data is it? Categorical or numerical?	Categorical	numerical	Discrete numerical

Our proposed investigation about the population

Our inklings/wonderings about the variable we are planning to explore are:

We think that region will affect ideas about climate change eg NI vs SI.

Our comparison investigative question is:

I wonder if the importance of Global Warming tends to be greater in Canterbury than in Auckland for students in New Zealand.

We think the data might show:

That the south island is more concerned with global warming as it is a more rural environment and nearer to Antarctic which is greatly affected by the global warming impact.

Our proposed investigation about the population

Our comparison investigative question is:

I wonder if the importance of Global Warming tends to be greater in Canterbury than in Auckland for students in New Zealand.

CRITERIA

1. The variable(s) of interest is/are clear and available or can be collected
2. The group (CL1-4) or population (CL5 onwards) of interest is clear
3. The intent is clear (e.g. summary, comparison, relationship, time series)
4. The investigative question can be answered with the data (e.g. sample size sufficient, question is specific, data can be collected, ethics)
5. The investigative question is one that is worth investigating, that it is interesting, that there is a purpose
6. The investigative question allows for analysis to be made of the whole group

Check the investigative question for the variable (Criteria 1) and the group/population (Criteria 2) and the remaining criteria:

- Is the question purposeful or interesting? (Criteria 5)
- Is the question about the whole group/population? Check that the question is not just finding out about an individual or smaller group. (Criteria 6)
- Is the question one that we can collect data for/ or we have the data for (sourced/given)? (Criteria 4)
- Is it clear that the question is a summary/comparison/relationship/time series question? (Criteria 3)

1-3 Wrap up

In year 11 students need to be taking more control of finding the data for statistical investigations.

While some caution is needed, especially if they are needing to explore a comparison situation about a population, students need to understand that just finding a dataset, without the background story, is not sufficient.

Interrogating the dataset to understand the who, what, where, why and how is a key part of statistical enquiry, it is also a key part of statistical literacy.

2-1 “Who” to measure

The whole group or a sample?

Observational studies

Curriculum level 1-4

Summary, comparison and relationship situations

- Whole group

Curriculum level 5-6

Summary and comparison situations

- Whole group
- Use a **representative sample** to answer the investigative question about a population

Relationship situations

- Whole group

Samples from populations

There are a number of reasons why we would use a sample rather than the whole population including:

- it is expensive to take a census (survey all of the population)
- if the whole population is not accessible e.g. doing a biological test on slaters in a particular region
- if using the whole population could be destructive e.g. the breaking strain of steel cables
- when the population is homogeneous e.g. blood type testing for an individual, taste testing wine from a vat
- we are only given a sample to work with
- the technology we are using cannot cope with huge datasets so we have to work with a sample.

(Arnold, Trinick, & Pfannkuch, in press)

Sample size - numerical variables

In our planning we define the population and describe the sample we want, including the sample size. At curriculum level 5 we work with **representative samples** of about 30 for a summary situation and **representative samples** of about 30 for each group for comparison situations. The two samples can be different sizes for comparison situations.

From curriculum level 6 we start to look at sample size and how that impacts on any calls we might make. **Representative samples** of about 100 can be used.

Representative samples

It is usually not possible to make generalizations to a larger group of people unless the subjects for the study are representative of the larger group. The only way to collect a sample that is representative is to collect the objects we study at random.

(Gould, Ryan, & Wong, 2017, p. 45)

Building conceptions of populations and samples

@CL5 - year 10, page 5 in workbook

[Karekare College: building conceptions of populations and samples]

Aims of the Session

- Link between sample and population
- Students need to experience the need to sample.
- Describe sample distributions and then think about the population distributions.
- Predict population distributions.
- Care with language, these boys, these girls.

Sampling variability

Sampling variability is the variation in a sample statistic from sample to sample.

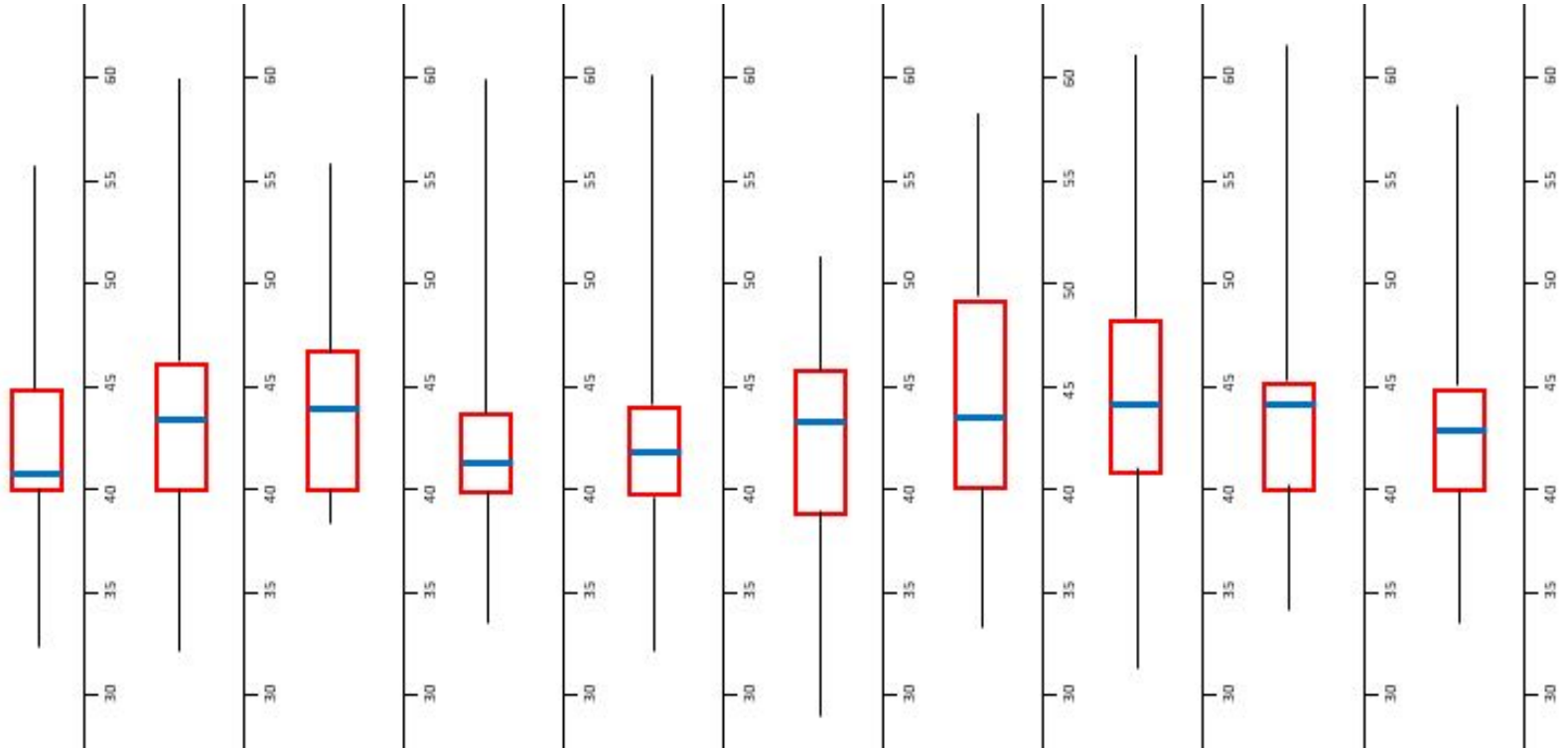
Developing the concept of sampling variability

The students are trying to answer an investigative question about popliteal lengths of the Karekare College population and were making a dot plot using data cards to answer the investigative question. The students got to a point of saturation where they felt that additional data did not change the main signals they were getting about the variable and the shape was remaining relatively constant.

To draw attention to sampling variability the students were asked to walk around the class and observe other groups' graphs. They were asked to see how the graph that they had made with their sample from the Karekare College population was the same and how it was different to the graphs other groups had made.

All of the groups gave an indication of where they felt the middle of their popliteal length data was, and across the class the middle popliteal lengths for the different groups lay within a 3–4 cm band. The students were able to see that the middle popliteal length was similar even though the samples were different (Arnold, 2013). They also noticed that the shape of the distributions was similar, symmetric and unimodal.

Sampling variability with box plots

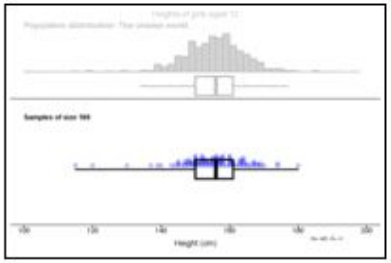


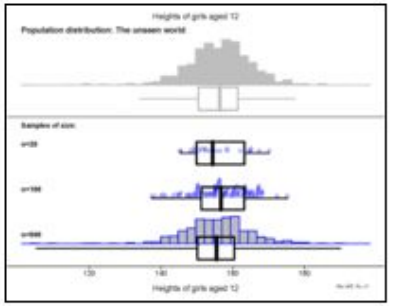
@CL5 Ten samples of 30 from Karekare College for popliteal length, using suggested colouring scheme to highlight median (blue), middle 50% (red), whiskers (black) (Arnold, Trinick, & Pfannkuch, in press)

Animations

@CL6 exploring different sample sizes

[WPRH - Sampling Variation Animations]

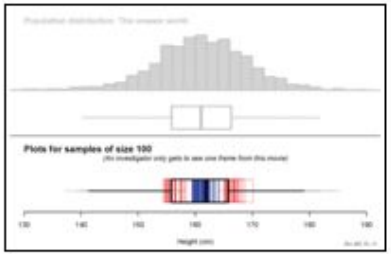


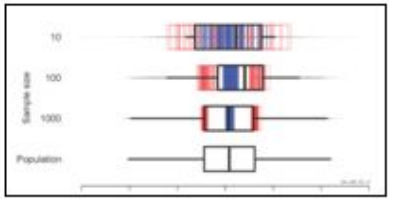


Animated Gifs Click for: n=30 n=100 n=300 Vew Full Size GIF

PDFs Click for: n=30 n=100 n=300 View Full size PDF

1(a) 1(b)





Animated Gifs Click for: n=30 n=100 n=300 View full size GIF

PDFs Click for: n=30 n=100 n=300 View full size PDF

2(a) 2(b)

Enduring ideas

- A big enough representative sample can be used to tell us about a population.
- There is variability in sample statistics from sample to sample (@CL5)
- A sample provides an estimate for population parameters (@CL5)
- Variability in sample statistics between samples reduces as the sample size increases (@CL6)

Resources for teaching and learning

Additional teaching and learning activities are on CensusAtSchool.
Statistical Investigations Part 2

Lesson 5: Introducing sampling

Focus for lesson: • Revisit posing investigative questions, some summary mostly comparison. • Identifying and clarifying the population • Sampling variability – all samples from a given population are different (can't tell exactly) • Distorted view of population – samples give (same) distorted view, similarity among samples • What is a sample? • Why sample?

2-2 Visualising and describing data

Using CODAP to make data visualisations

- Page 6 in the workbook
- Watch the video to find out what to do

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	Clusters Gaps Outliers
Variability	Spread	Range, interquartile range, mean average deviation from the median/mean (MAD) *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 Distance between the medians Middle 50%
	Modal clumps	Peak(s) (local mode) Modal group(s)

Figure 8-16. Distribution description framework

Note: * indicates part of feature listed.

Arnold, P. (2013, p. 217)

2-3 Answering the investigative question

DRAWING CONCLUSIONS & ANSWERING THE INVESTIGATIVE QUESTION

And the big ideas that lead to this

Statistical inference for comparison of two samples of quantitative data at New Zealand Curriculum level 5 & 6			
	Components		
	Generalisation	*Evidence from data*	*Probabilistic*
Verbalisations	Making a claim about the aggregate that goes beyond the data	Being explicit about the evidence used	Articulating the uncertainty embedded in an inference
Underpinning reasoning concepts	Sample (random) Population Distribution	Shift, overlap, overall visual spread Position of medians Decision guide Shape	Sampling variability Uncertainty Sample size (@ CL6)

MAKING THE CALL...

Page 8 workbook

- When students make a claim (call) from the comparison of two box plots they need to:
  1. know what relevant aspects to notice from what they can actually see in the plots
  2. draw on relevant sampling variability knowledge from what they have experienced
  3. invoke contextual knowledge.

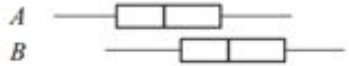
"How to make the call" by Curriculum Level

At all levels:

If there is no overlap of the boxes, make the call immediately
"***B tends to be bigger than A***" back in the populations

Apply the following when the boxes do overlap ...

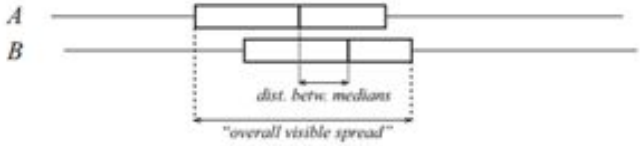
Curriculum Level 5: *the 3/4-1/2 rule*



If the median for one of the samples lies outside the box for the other sample
(e.g. "*more than half of the B group are above three quarters of the A group*")
make the claim "***B tends to be bigger than A***" back in the populations

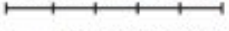
[Restrict to samples sizes of between 20 and 40 in each group]

Curriculum Level 6: *distance between medians as proportion of "overall visible spread"*



Make the claim ***B tends to be bigger than A*** back in the populations
if distance between medians is greater than about ...

1/3 of overall visible spread for sample sizes of around **30**



1/5 of overall visible spread for sample sizes of around **100**

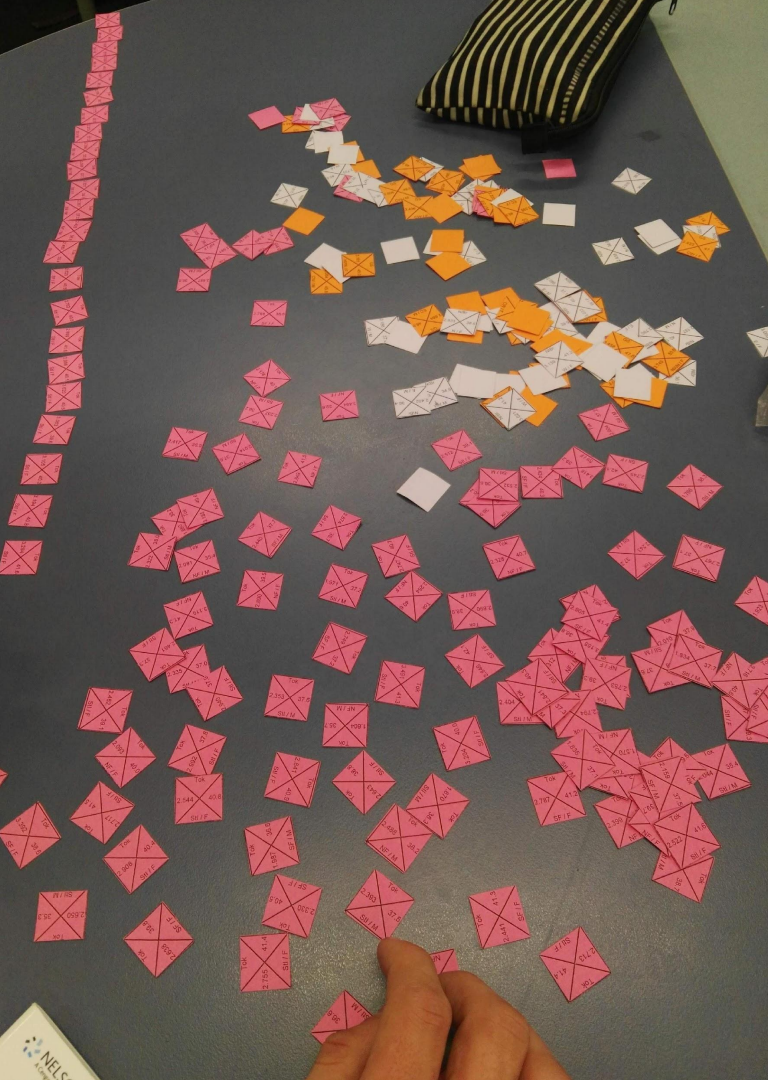
[***Could also use*** 1/10 of overall visible spread for sample sizes of around 1000]

- When students make a claim (call) from the comparison of two box plots they need to:
  1. know what relevant aspects to notice from what they can actually see in the plots
  2. draw on relevant sampling variability knowledge from what they have experienced
  3. invoke contextual knowledge.

- We want students to **UNDERSTAND** why they are drawing this conclusion in this way.

DEVELOPING UNDERSTANDING DEVELOPING AROUND SAMPLING VARIABILITY…

… and what it means about making sample-to-population inferences

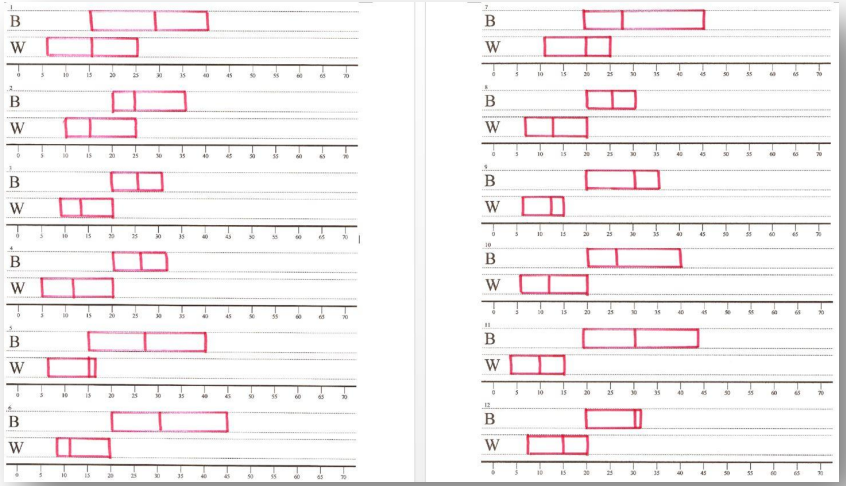


STUDENTS SHOULD EXPERIENCE THESE CONCEPTS MULTIPLE TIMES

Statistical inference is the art and science of drawing conclusions about a population on the basis of observing only a small subset of that population. Statistical inference always involves uncertainty, so an important component of this science is measuring our uncertainty.

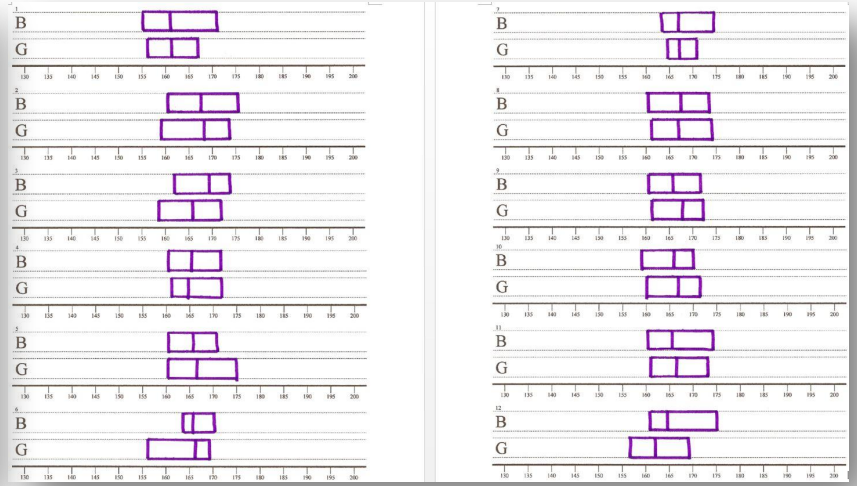
(Gould, Ryan, & Wang, 2017, p.327)

LEVEL 5 LEARNING EXPERIENCES
CONSISTENCY (OR NOT) IN REPEATED SAMPLES



Consistent message from all the samples

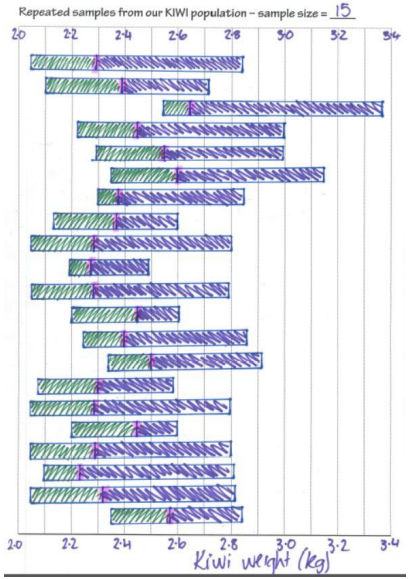
- The samples are far enough apart to see the signal



Inconsistent message from all the samples

- The samples are too close together to see the signal

LEVEL 6 LEARNING EXPERIENCES
SAMPLE SIZE IMPACTS SAMPLING VARIABILITY

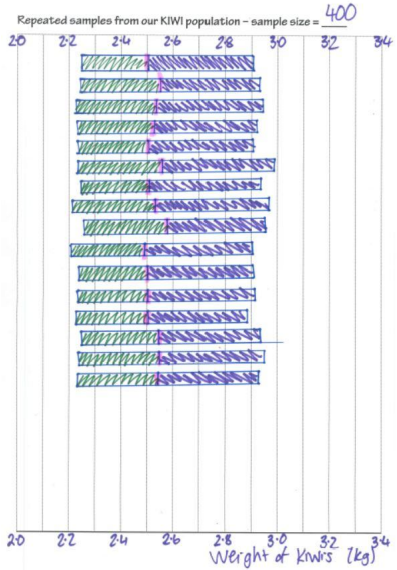


n=15

- There is a lot of variability from sample to sample

n=400

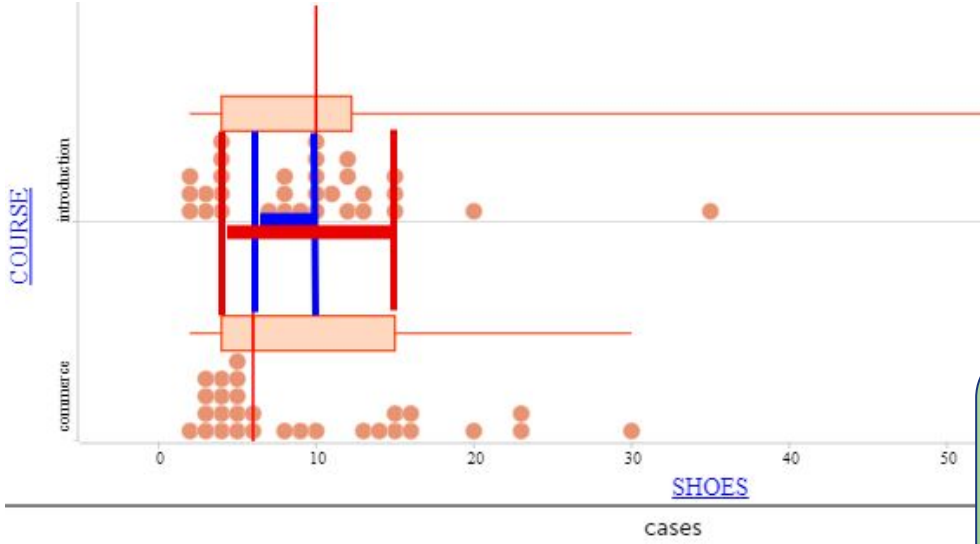
- There is very little variability from sample to sample



BY HAND…

REPEATED SAMPLING □ MEDIANS AND CHANGING SAMPLE SIZE

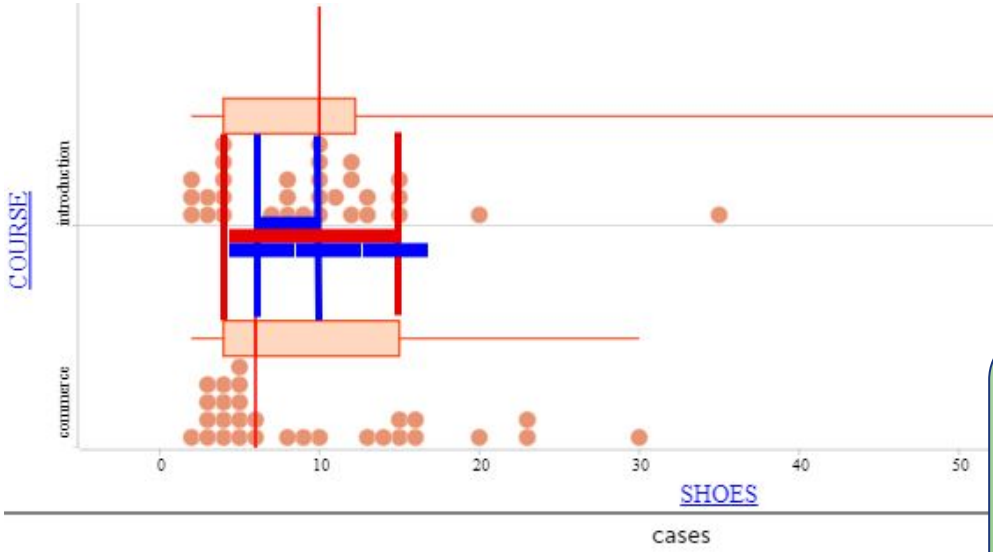
CONCLUSION…



Variable (units)	Number of pairs of shoes owned
Groups	Commerce degree n = 29 Introduction course n = 32
Population	1st year Auckland uni students

- When students make a claim (call) from the comparison of two box plots they need to:
  1. know what relevant aspects to notice from what they can actually see in the plots
  2. draw on relevant sampling variability knowledge from what they have experienced
  3. invoke contextual knowledge.

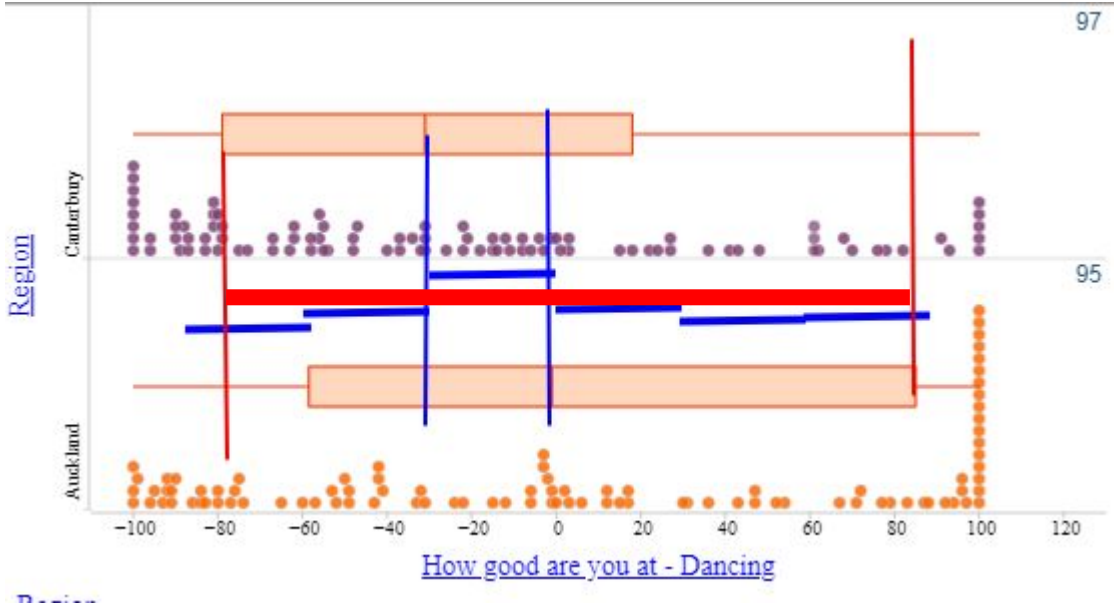
CONCLUSION…



Variable (units)	Number of pairs of shoes owned
Groups	Commerce degree n = 29 Introduction course n = 32
Population	1st year Auckland uni students

- When students make a claim (call) from the comparison of two box plots they need to:
  1. know what relevant aspects to notice from what they can actually see in the plots
  2. draw on relevant sampling variability knowledge from what they have experienced
  3. invoke contextual knowledge.

CONCLUSION...



Variable (units)	Self-rating of dancing ability (score from -100 to 100)
Groups	Auckland (n=100) Canterbury (n=100)
Population(s)	School students from Auckland and school students from

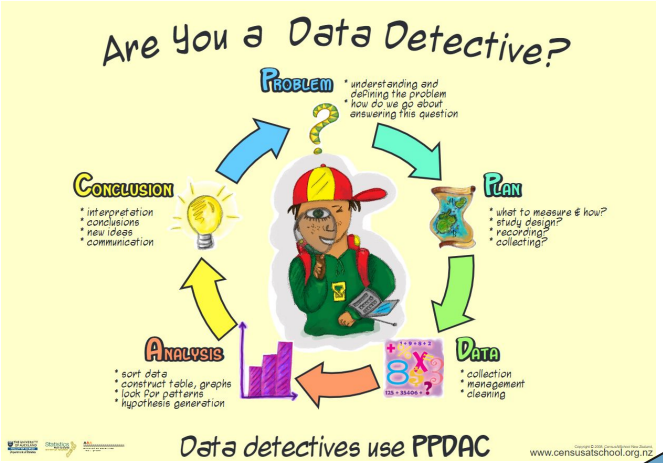
- When students make a claim (call) from the comparison of two box plots they need to:
  1. know what relevant aspects to notice from what they can actually see in the plots
  2. draw on relevant sampling variability knowledge from what they have experienced
  3. invoke contextual knowledge.

2-4 Exploring worlds

Sample to population inference

Knowledge about the world

Existing knowledge about the world. Problem of interest arises.



Undertake a statistical enquiry to address the problem of interest.

New knowledge about the world informed by statistical enquiry.

Concept development

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

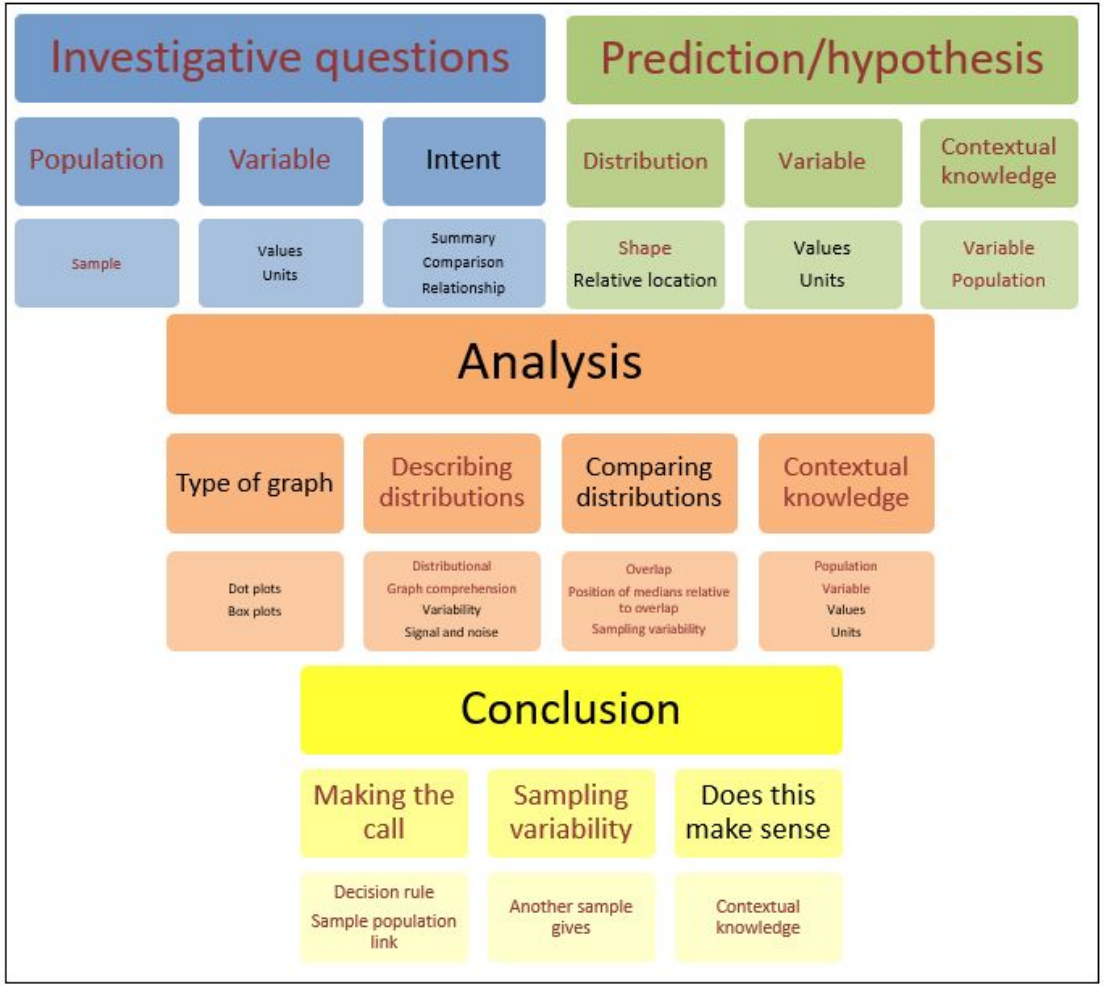


Figure 9-4. Conceptual framework for statistical investigations thread, level 5 of the New Zealand curriculum

Purposeful


1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
SUSTAINABLE DEVELOPMENT GOALS
See all

Connect


In English, students study, use, and enjoy language and literature communicated orally, visually, or in writing.
In the arts, students explore, refine, and communicate ideas as they connect thinking, imagination, senses, and feelings to create works and respond to the works of others.
In health and physical education, students learn about their own well-being, and that of others and society, in health-related and movement contexts.
In learning languages, students learn to communicate in an additional language, develop their capacity to learn further languages, and explore different world views in relation to their own.
In mathematics and statistics, students explore relationships in quantities, space, and data and learn to express these relationships in ways that help them to make sense of the world around them.
In science, students explore how both the natural physical world and science itself work so that they can participate as critical, informed, and responsible citizens in a society in which science plays a significant role.
In the social sciences, students explore how societies work and how they themselves can participate and take action as critical, informed, and responsible citizens.
In technology, students learn to be innovative developers of products and systems and discerning consumers who will make a difference in the world.

Reflect

Jamboard

- One question I still have
- Something I can’t wait to try out with my students
- One thing I want to share with someone at school
- Other thoughts

He pai ake te iti i te kore

A little is better than none

Thank you

- To the team

Robyn Headifen, Zac Rutledge, Michelle Dalrymple, Sophie Wright, Rachael Ouwejan, Marina McFarland, Lucy Edmonds, Maxine Pfannkuch, Mark Hooper

- To Cashmere Maths & Stats Faculty - guinea pigs last week
- To you and the great work you do to make a difference for students
- Anna Fergusson and Emma Lehrke for the planning behind the scenes
- AMA for the work they do to support teachers in Auckland and across Aotearoa New Zealand

Some data

What caused the spike?