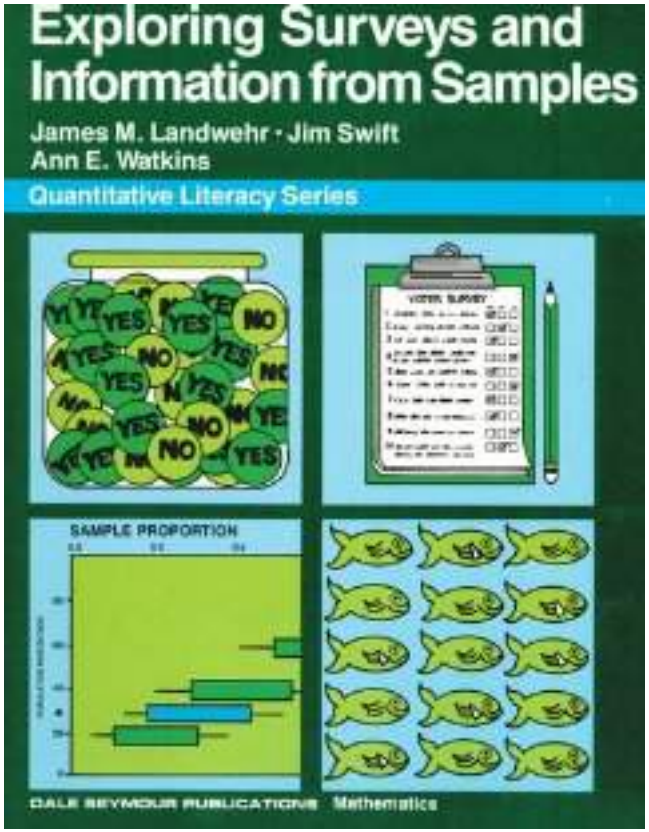
The Practice of Statistics at School: What does Evaluating Evidence look like in the classroom?

Jane Watson

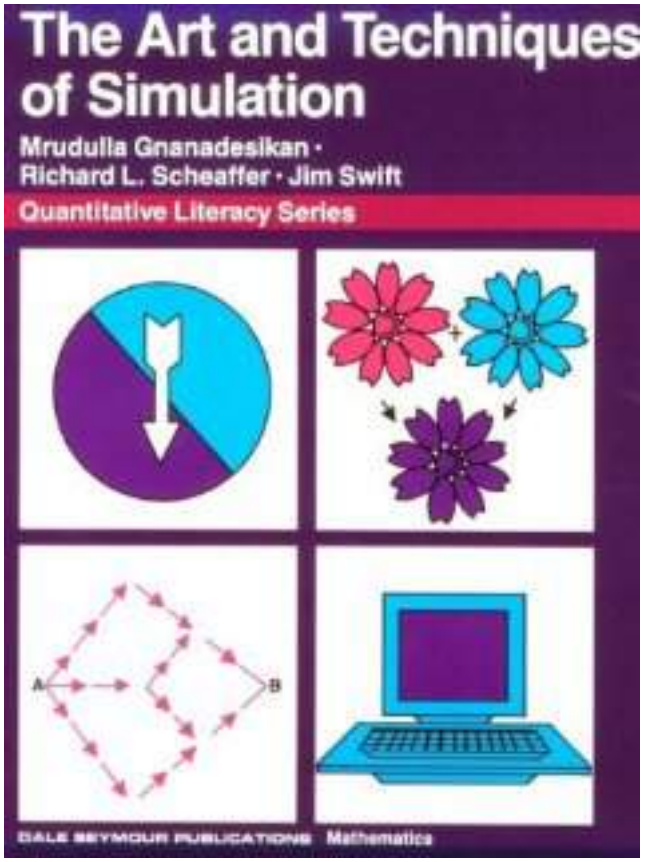
UNIVERSITY of TASMANIA

Background/History

Quantitative Literacy Project

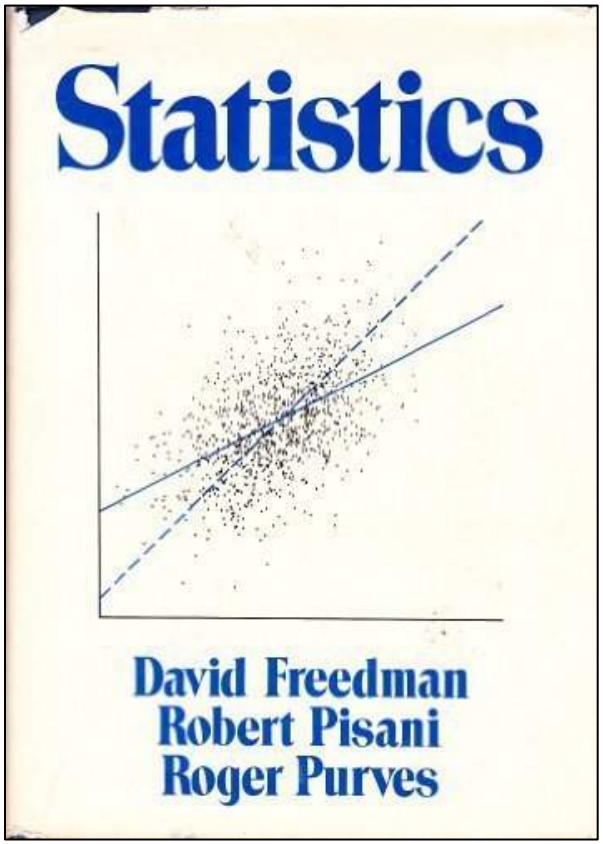


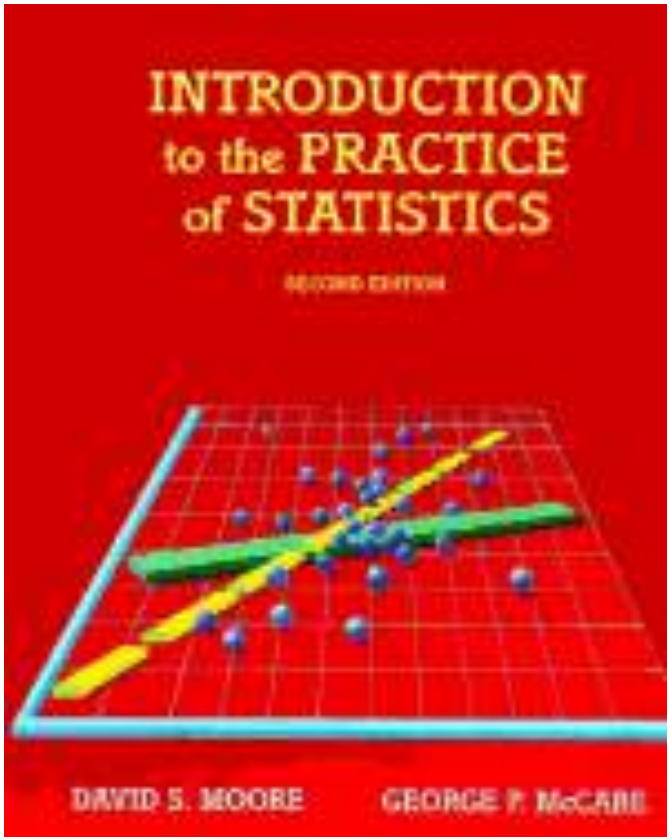
1980s



Background/History

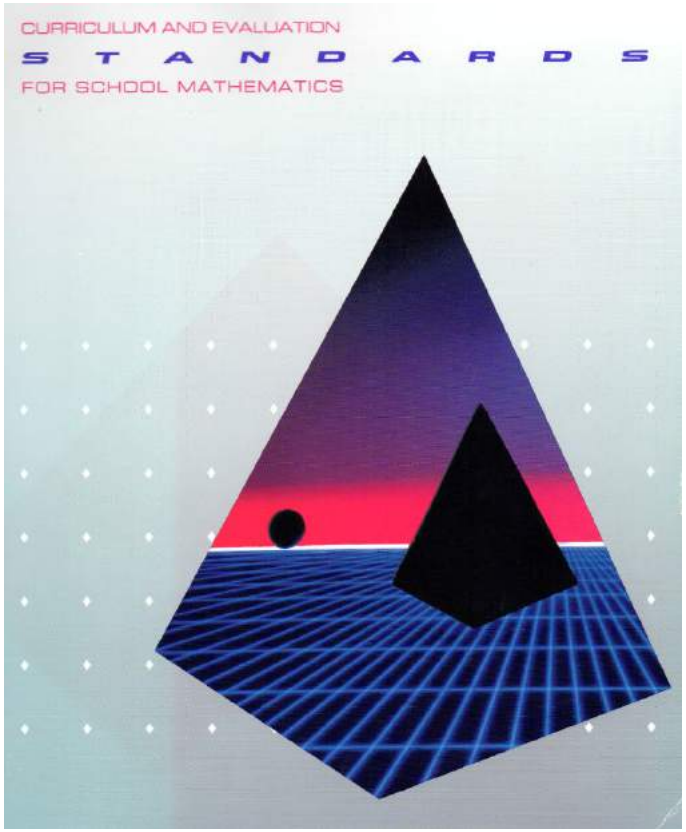
1980's to 1990's Non-traditional Intro to Statistics Textbooks



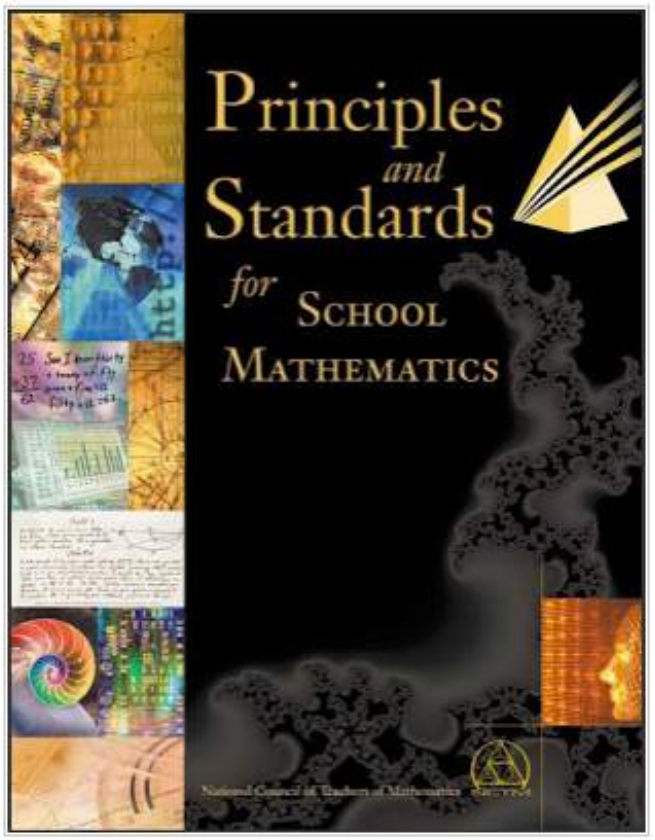


Background/History

NCTM Standards: Include the Probability and Data Analysis Strands



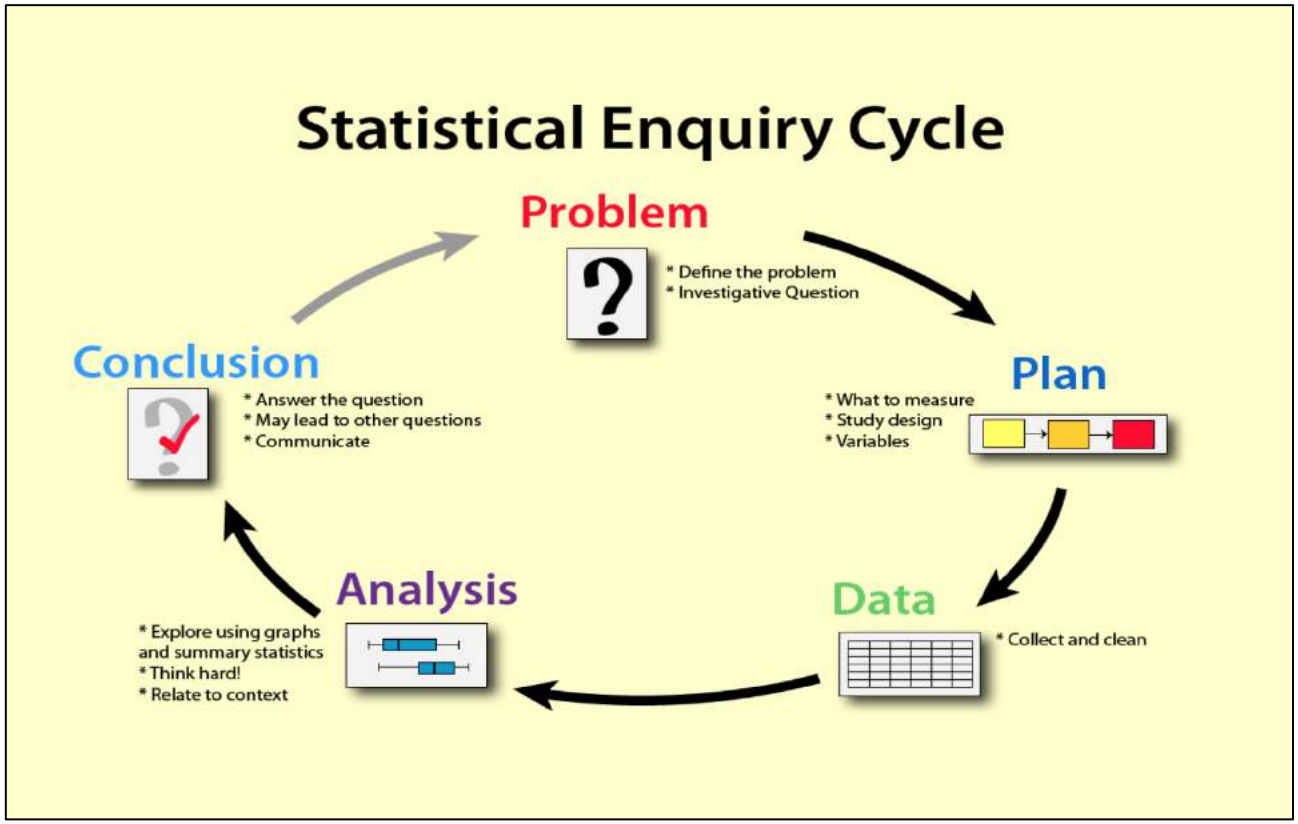
1989



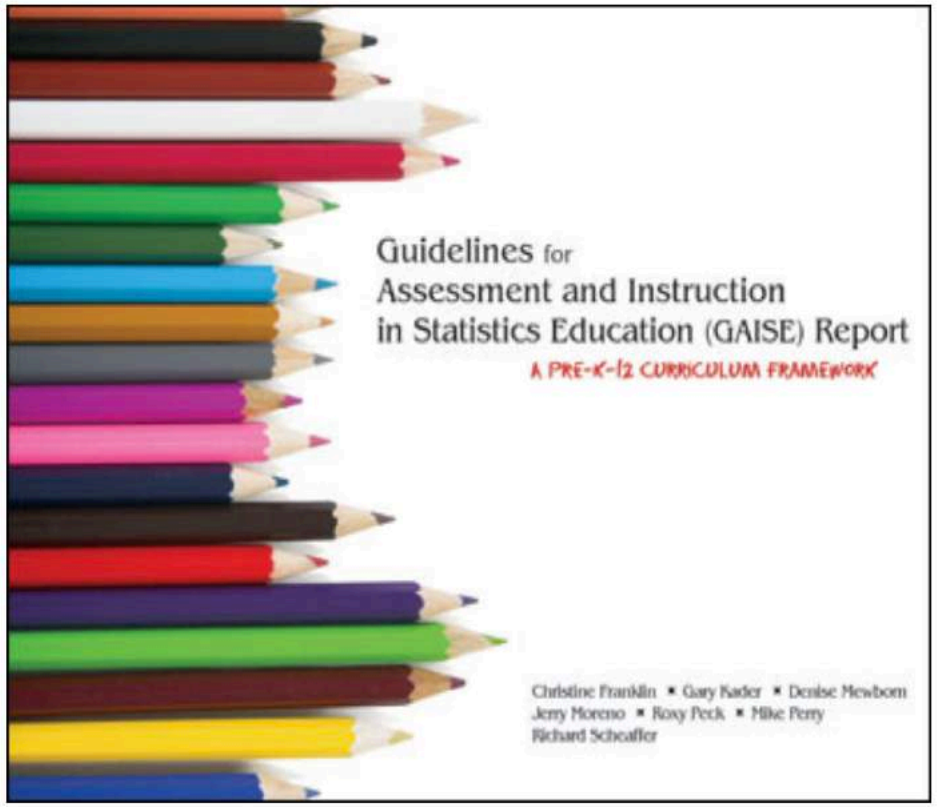
2000

Wild & Pfannkuch, 1999: PPDAC

Based on analysis of the work of their **applied statistician** colleagues



The Practice of Statistics
GAISE (2007) – "statistical problem solving"

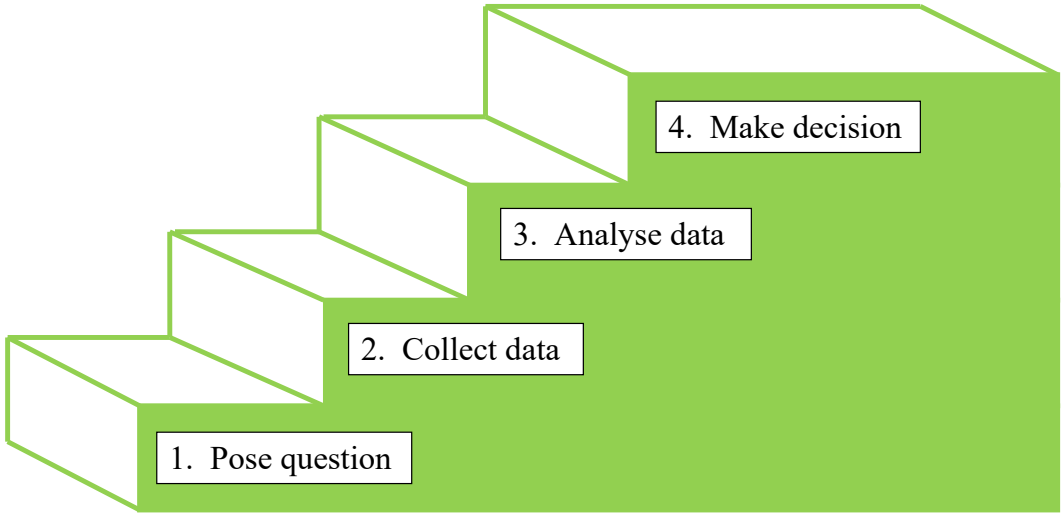


- Formulate Questions - Anticipating Variability;
- Collect Data - Acknowledging Variability;
- Analyze Data - Accounting for Variability;
- Interpret Results - Allowing for Variability.

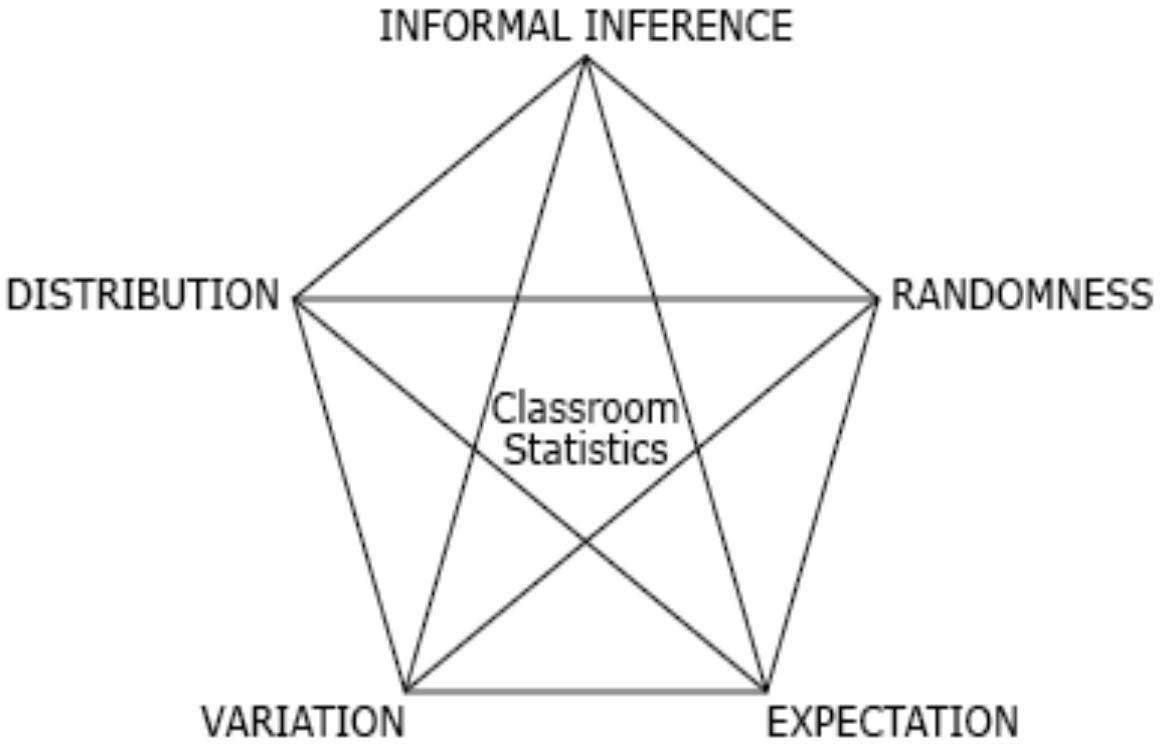
ASA, 2007, Franklin, et al.

Poster in a Primary Classroom

4 steps to making decisions with data



The Big Ideas of Statistics at School



Interrelated Big Ideas underlying statistics

Top Drawer Teachers, <topdrawer.aamt.edu.au/Statistics/Big-ideas>

Importance of Context:

There is no statistics without context. (Rao, 1975)

Importance of Variation:

The ability to deal intelligently with variation and uncertainty is the goal of instruction about data and chance (Moore, 1990)

Importance of Uncertainty:

Decisions about populations based on samples are never totally certain (Makar & Rubin, 2009, top of pentagon)

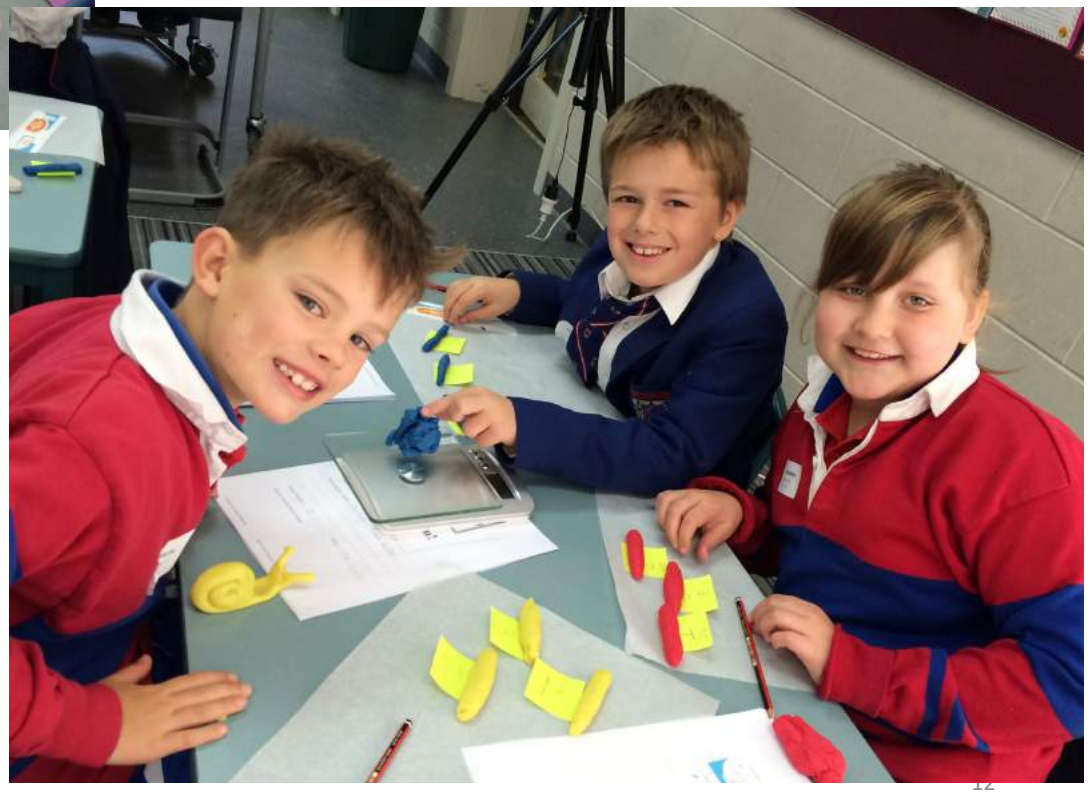
Educational Research at School and the Practice of Statistics

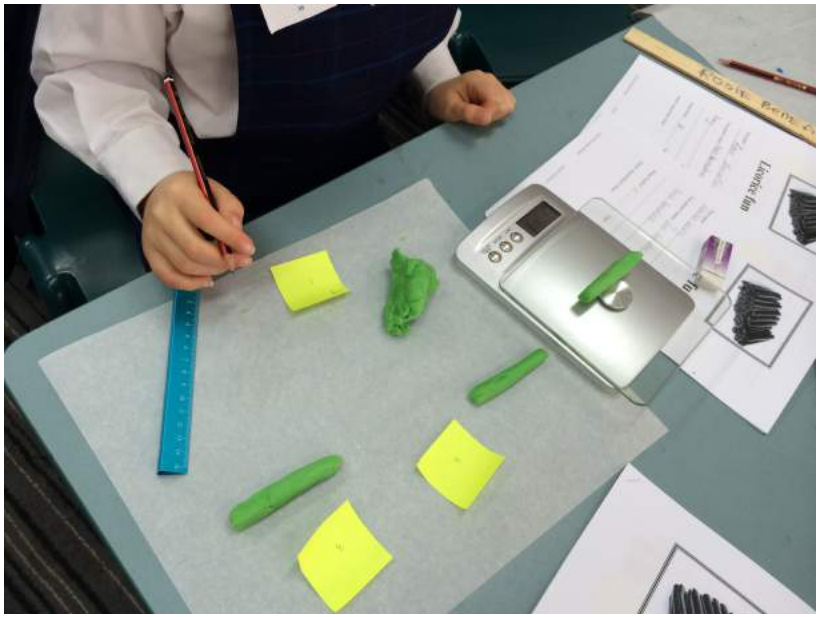
- Building a foundation in early grades: Variation! (Grade 3)
- And Learning PPDAC/GAISE investigative cycle and completing an investigation (Grade 5)
- Resampling for Informal Inference (Grade 10)

Making Licorice Sticks (Grade 3)

- Make by hand – 8 cm long, 1 cm diameter, weigh and record mass.
- Make with a Playdoh “extruder” factory, 8 cm long, 1 cm diameter, weigh and record mass.

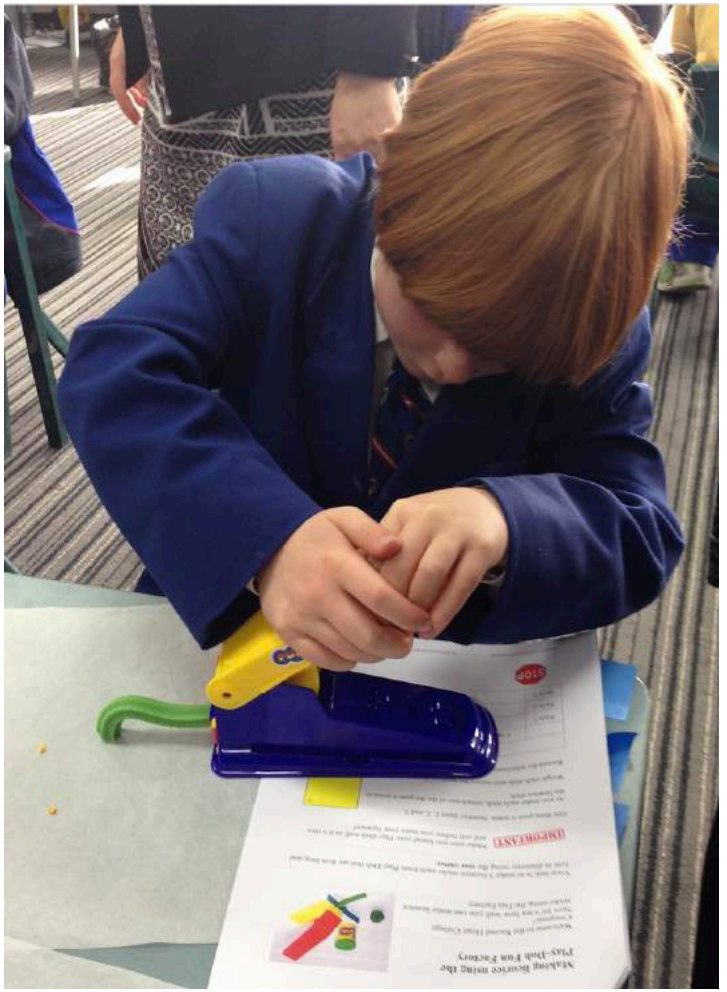
Making Sticks by Hand…





Measuring length and mass

Making Sticks with a Machine (Playdoh Factory)



Comparing Results

… and Making Predictions

Comparing Results

… and Making Predictions

Evaluating Evidence

List any differences between the two plots

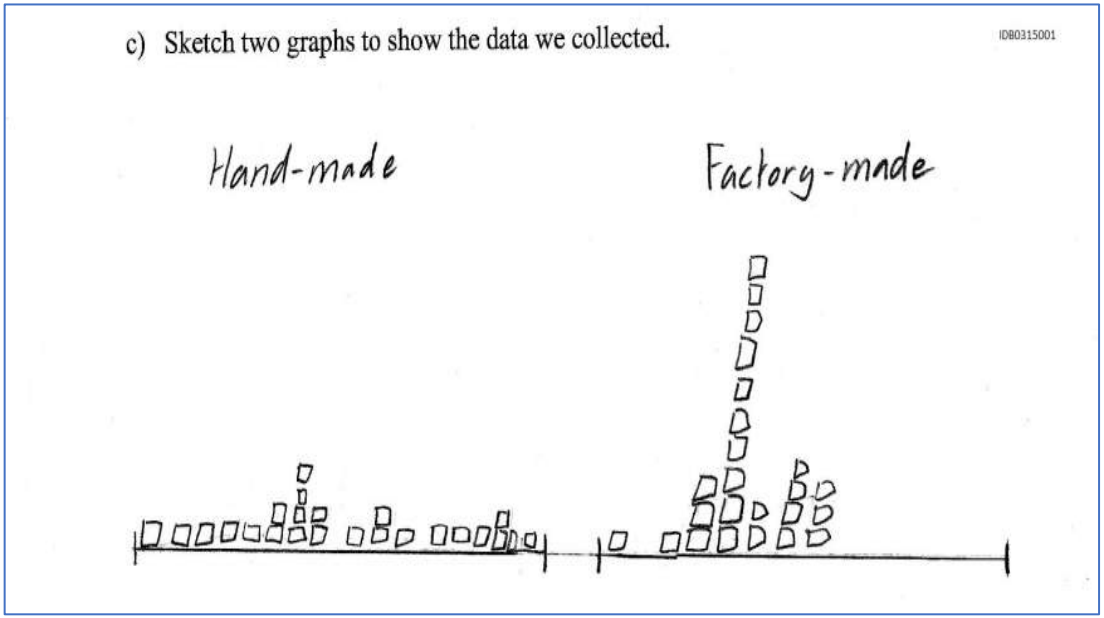
- The ones that we made with our hands were very different but the ones the machine made were a little bit different.
- One was in between 10g and 16g and the other was in between 6g and 28g. The machine is more accurate.
- There are more different weights in the handmade one. Most people have 14g on the machine-made one. Nobody on the machine-made one had 5g and one person on handmade did.

Evaluating Evidence

List any differences between the two plots

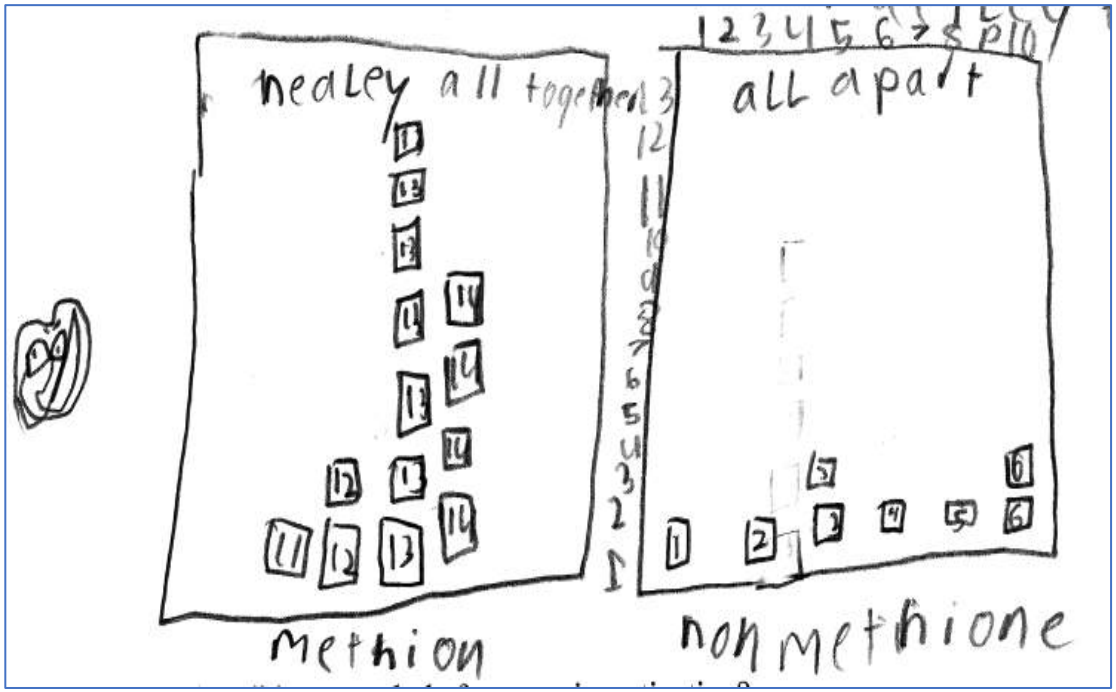
- The handmade is like a city with homes spread out and factory-made is the same except the factory-made is with tall buildings, not many but tall buildings.
- 14 is more common in machine-made.
- Handmade is like a city but machine-made is like a tower.
- Factory-made had a larger typical number. Hand-made had more variation in their mass.

… Recall 4 months later



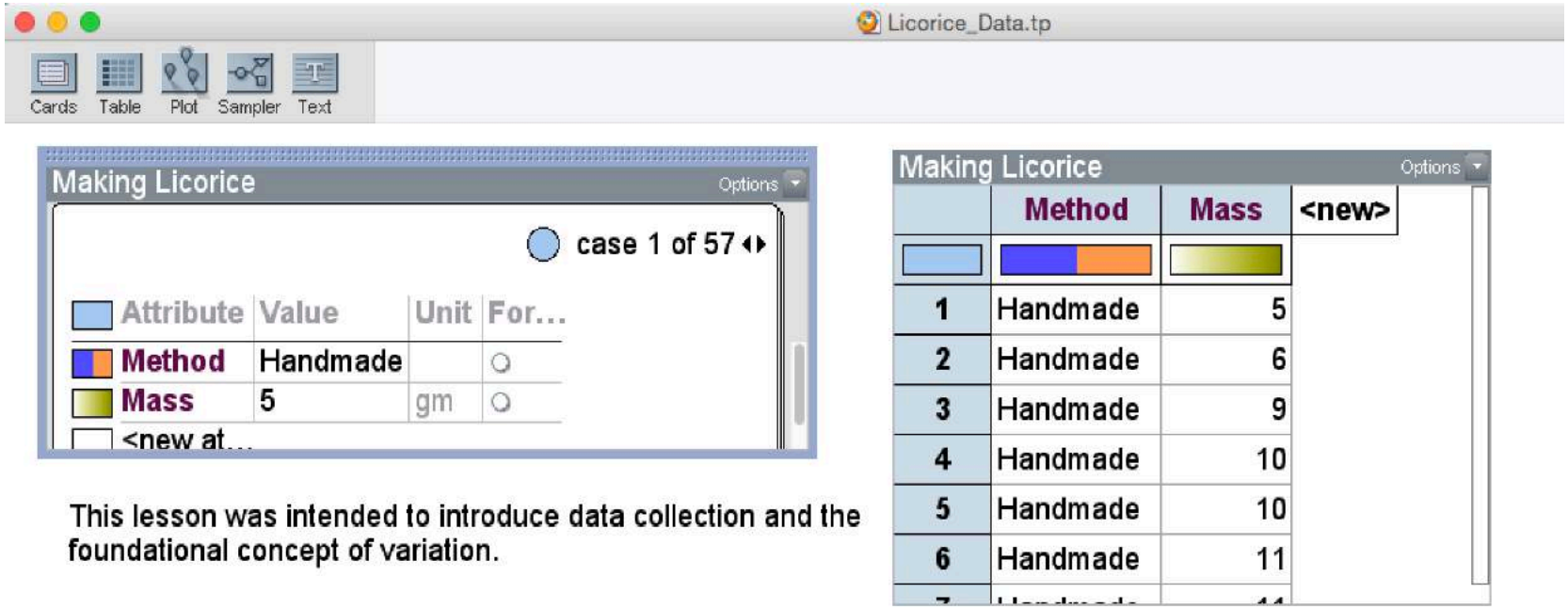
That factory made is more squished together and hand-made is spread out.

… Recall 4 months later



The factory made licorice sticks had less variation than hand-made ones.

Data in *TinkerPlots*



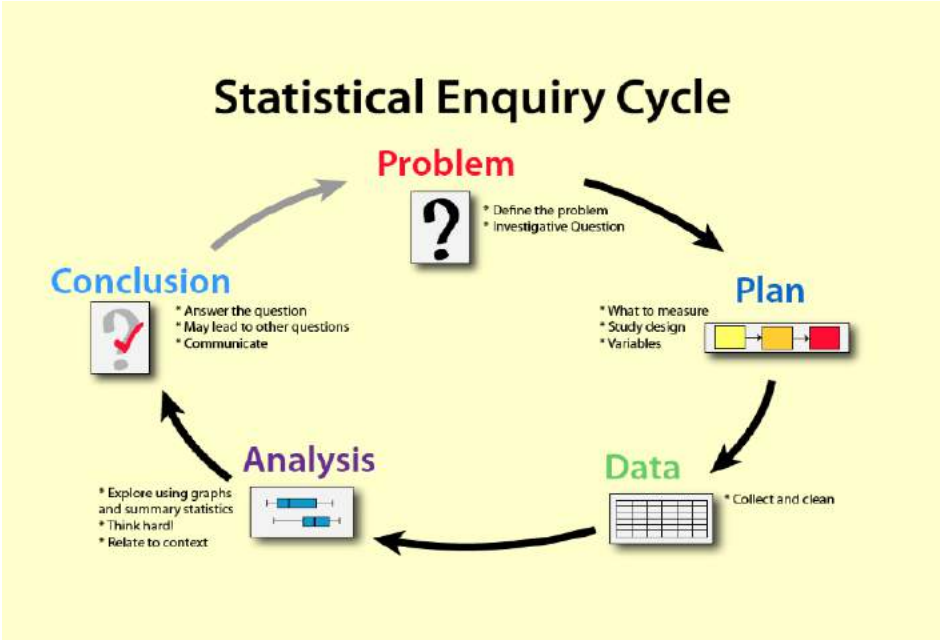
This lesson was intended to introduce data collection and the foundational concept of variation.

In Grade 4 students learned how to plot the Licorice data in *TinkerPlots*.

Link to *TinkerPlots*

The Practice of Statistics In Grade 5

Twice around the Enquiry Cycle



Students had had 3 activities in Grade 4:

- Problem Posing (in the playground)
- Variation (measuring arm length)
- Modelling Uncertainty (tossing two coins)

CONTEXT
The Australian Curriculum
Cross-Curriculum Priority: **Sustainability**

The Sustainability priority provides the opportunity for students to develop an appreciation of the necessity of acting for a more sustainable future and so address the ongoing capacity of Earth to maintain all life and meet the needs of the present without compromising the needs of future generations. (ACARA, 2017)

The Practice of Statistics In Grade 5 – Question

"Are we Environmentally Friendly?"

- Plan data collection.
- Collect data from the class.
- Analyse the data.
- Draw a conclusion for class.
- Ask about all of Australia.
- Collect random sample from large "population".
- Make a decision for Australia.

Survey questions from the ABS Census@School site

Am I environmentally friendly?	Yes	No
Our household has a water tank.		
I take shorter showers. (4 mins max)		
I turn the tap off while brushing my teeth.		
I turn off appliances (e.g., TV, computer, gaming consoles) at the power point.		
My household recycles rubbish.		

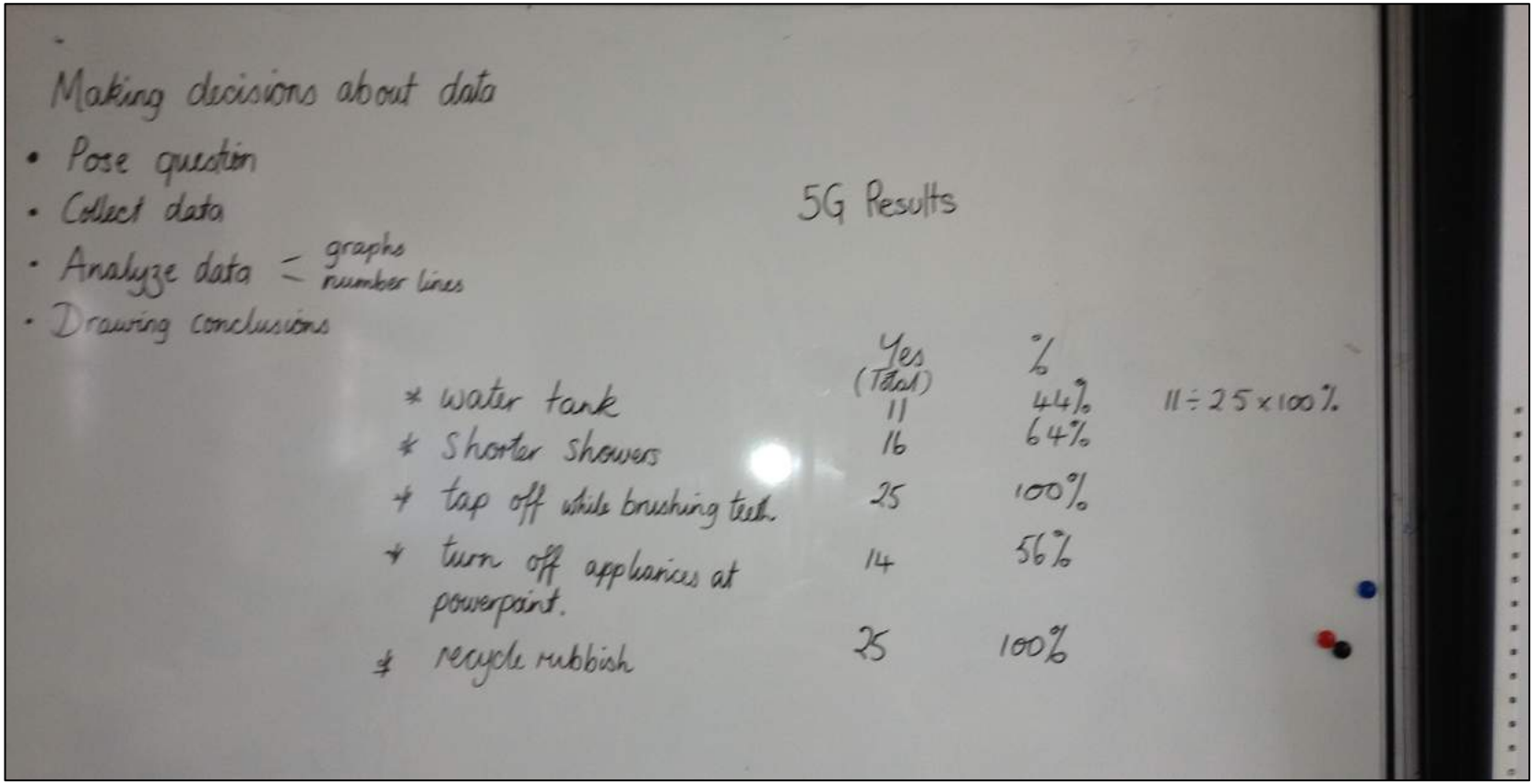
Students have to decide criteria: Percentage “yes” for 5 questions for the class to be environmentally friendly.

Student A:
My rule is 3/5 50% environmentally friendly-ish as I believe all questions are of equal value.

Student B: My criteria:

1. Water tank: 50%
2. Showers: 60%
3. Brushing: 70%
4. Electricity: 90%
5. Recycle: 100%

One classroom's results



Collect Data

Class of Students A and B

Am I environmentally friendly?	Yes Total	Yes %
Our household has a water tank.	10	38.5
I take shorter showers. (4 mins max)	14	53.8
I turn the tap off while brushing my teeth.	23	88.5
I turn off appliances (e.g., TV, computer, gaming consoles) at the power point.	4	15.4
My household recycles rubbish.	21	80.7

Analyse Data and Draw a Conclusion

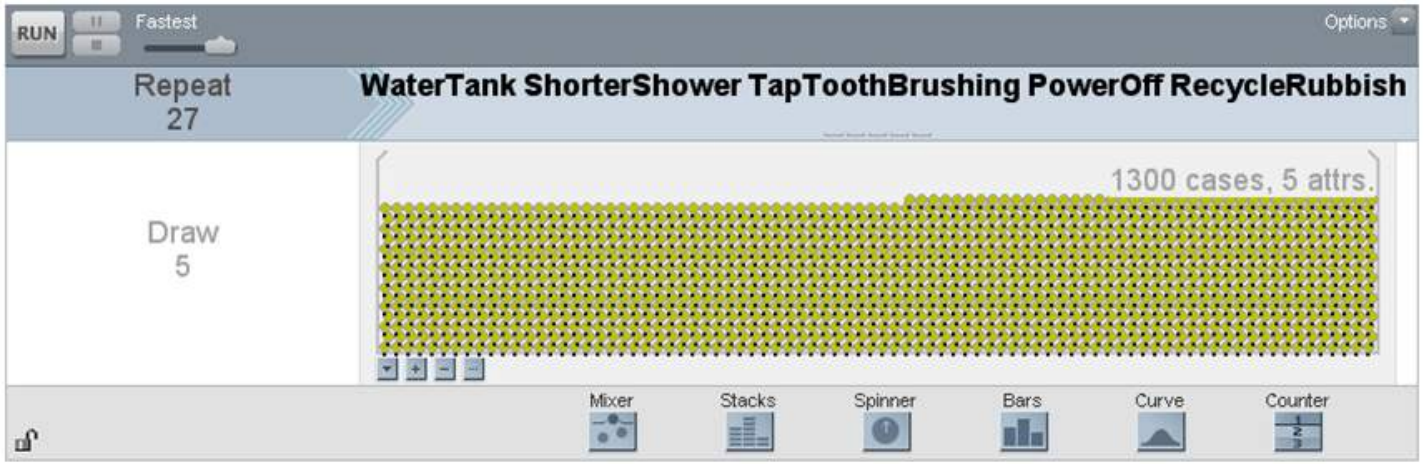
Use the percentages and their criteria to decide if class is environmentally friendly.

Student A: Our class is environmentally friendly-ish because 3/5 of questions are over 50% and that is my criteria.

Student B: Our class is not environmentally friendly because only one question reaches our benchmark. Even though our criteria is harsh rubbish dumps are nearly full and we don't have that much land.

Pose the question for all of Australian Year 5 students

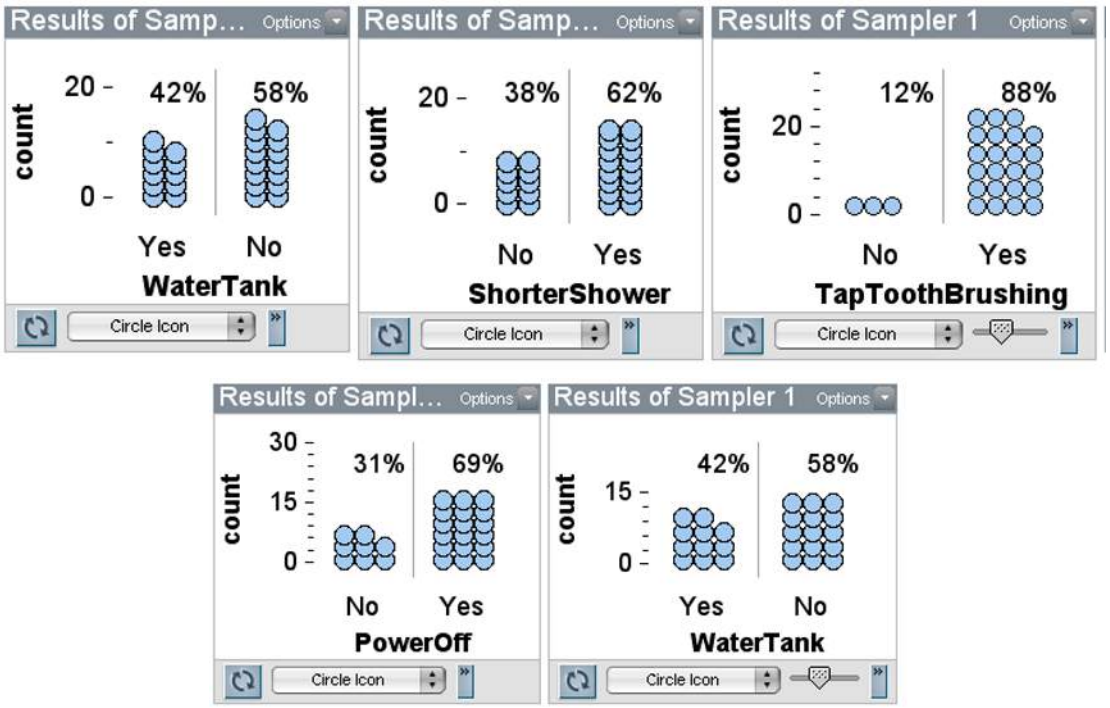
Collect a random sample the same size as the class from a “population” of 1300 Year 5 Australian students.



Results of Sampler 1 — Options

	WaterTank	ShorterShower	PowerOff	TapTo...	RecycleRubbish
1	No	Yes	Yes	No	Yes
2	No	No	Yes	Yes	Yes
3	No	Yes	Yes	Yes	No

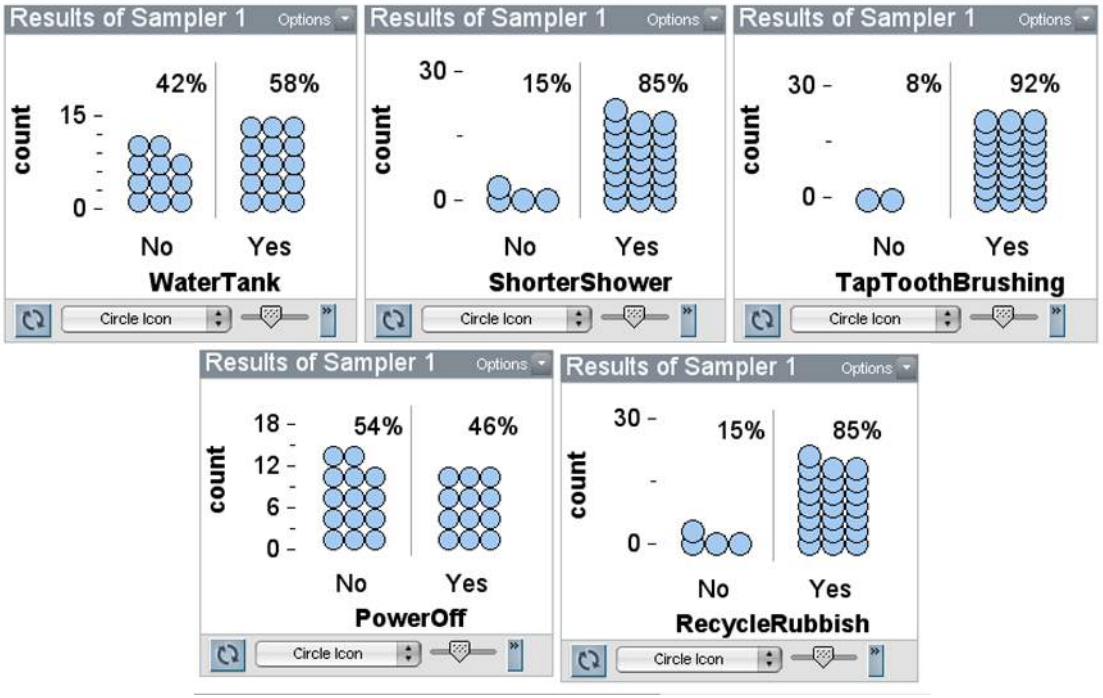
Example: Student A



Student A: *Make a decision for Random Sample*: They are environmentally friendly. Please refer to my criteria for reason. [3/5 over 50%]

Student A: *Certainty for the population?* The sample is too small to conclude with the numbers.

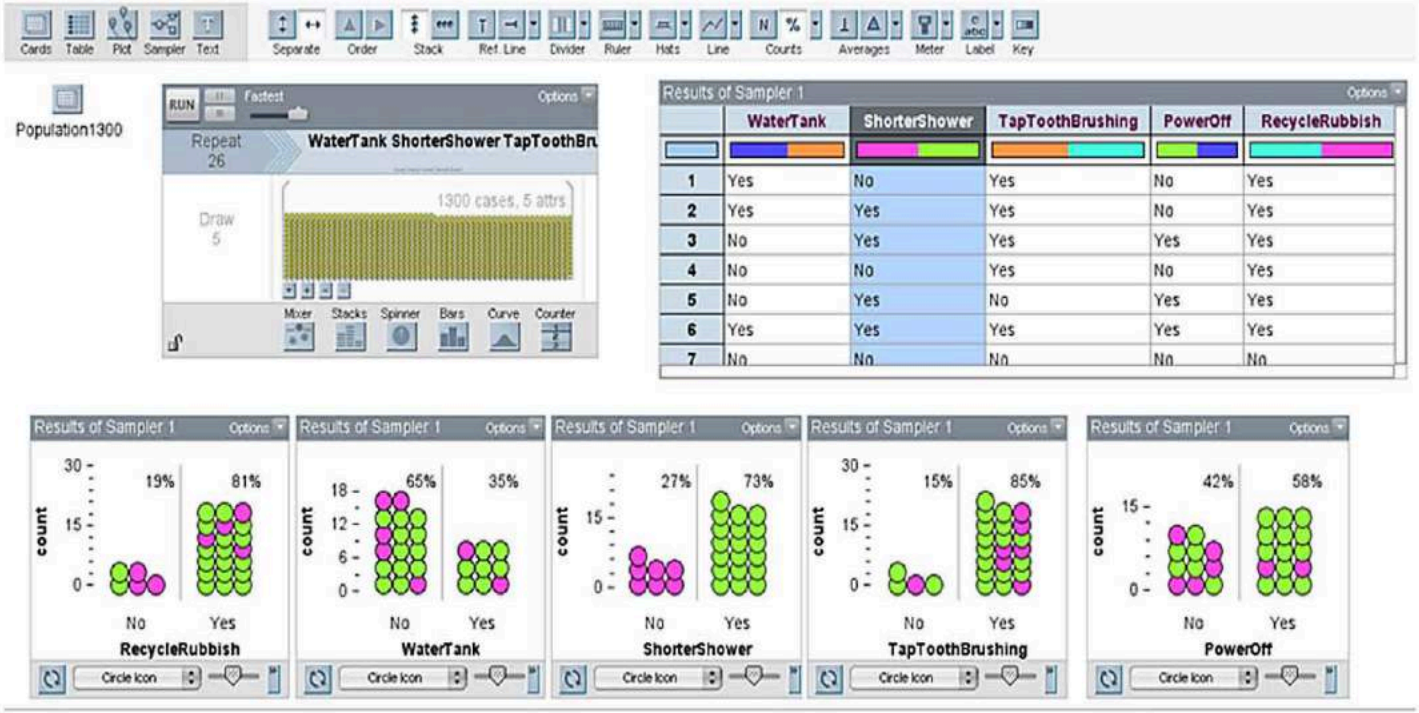
Example: Student B



Student B: *Make decision for Random Sample*: They are not environmentally friendly since they all didn't satisfy our criteria. [Water, 50%; Showers, 60%; Brushing, 70%, Electricity, 90%, Recycle: 100%]

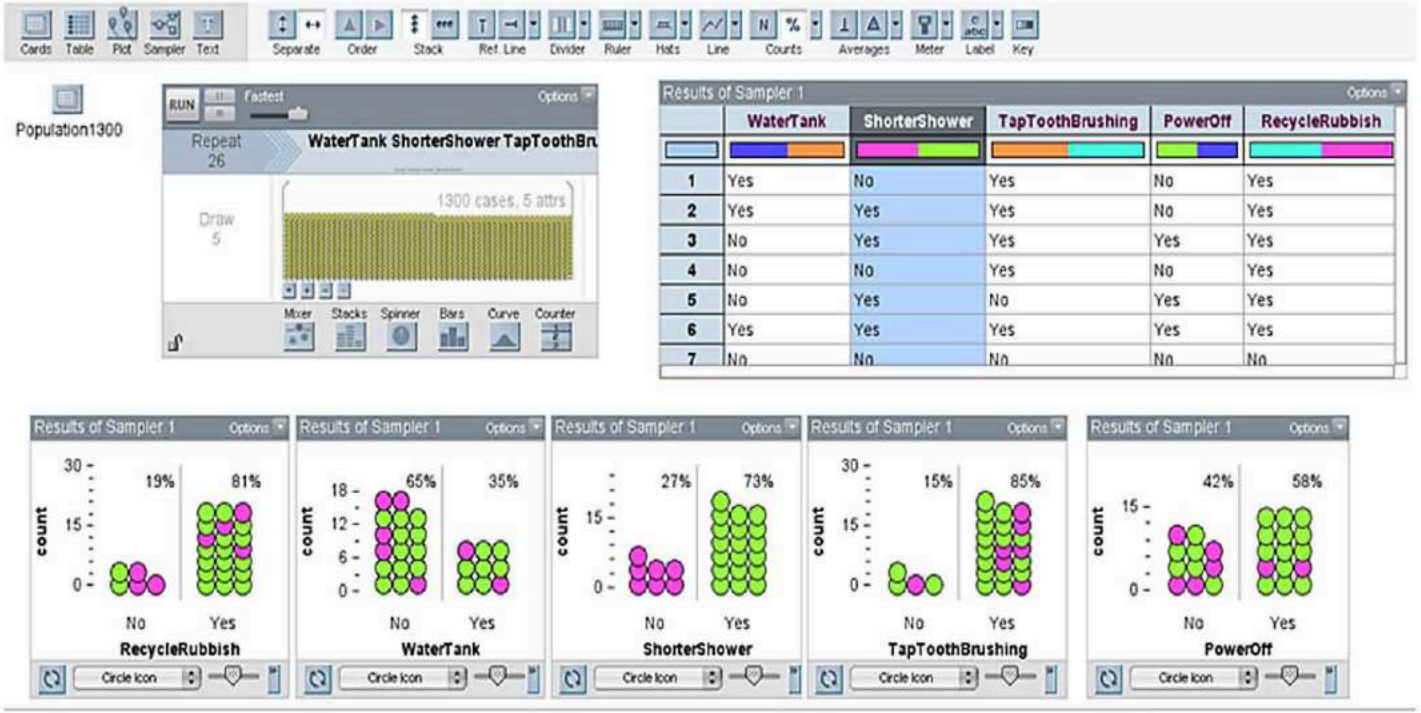
Student B: *Certainty for population?* Not very. This is a small sample and is not a representation for all students.

“Repeated Random Sampling in Grade 5”



2017 Best *Journal of Statistics Education* Paper – Watson and English (2016, 24:1) – Predicting population values from repeated random samples.

“Repeated Random Sampling in Grade 5”



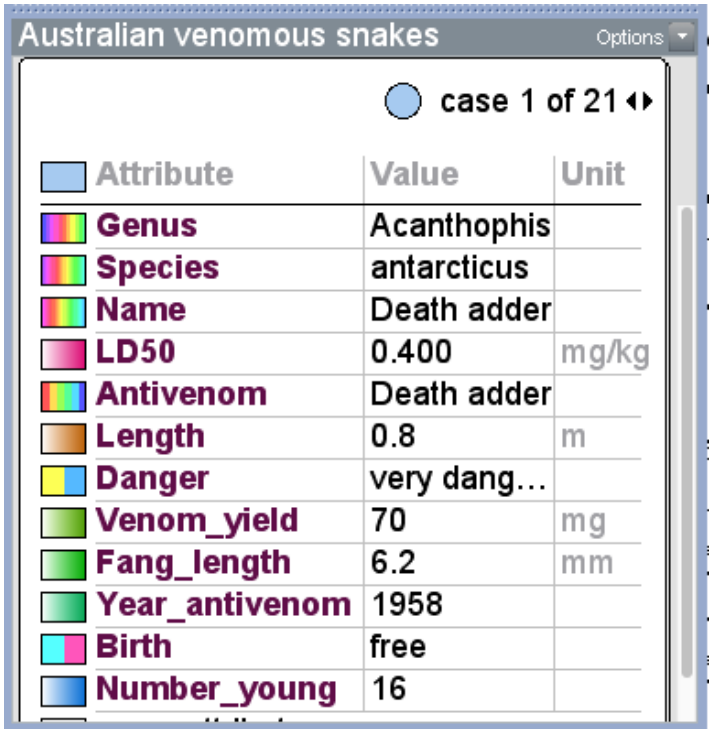
2017 Best *Journal of Statistics Education* Paper – Watson and English (2016, 24:1) – Predicting population values from repeated random samples.

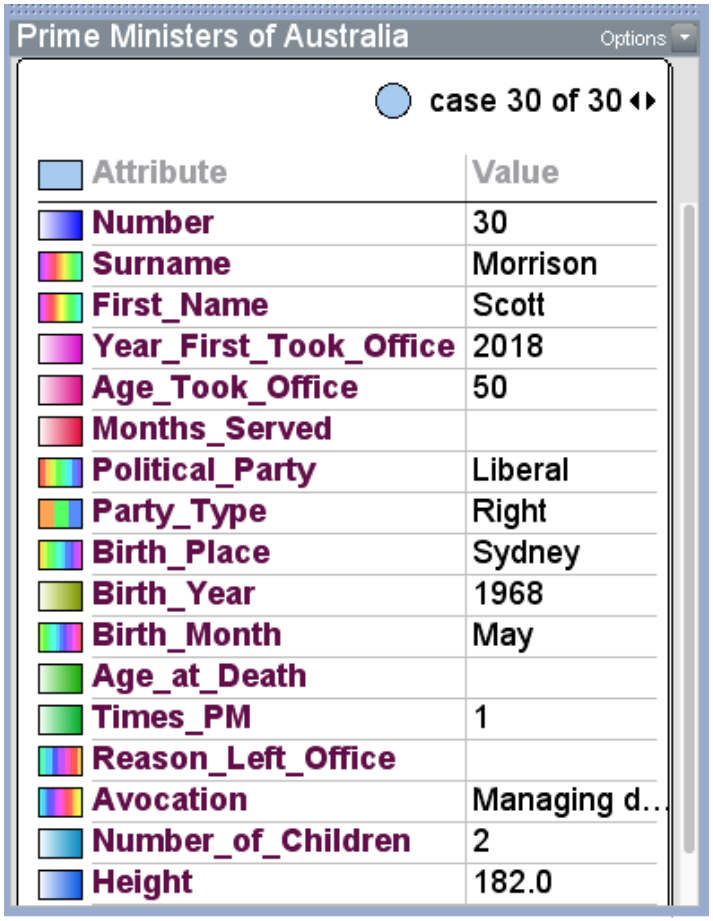
Evaluating Evidence
Using *TinkerPlots* in Grade 10
An 8-lesson Unit

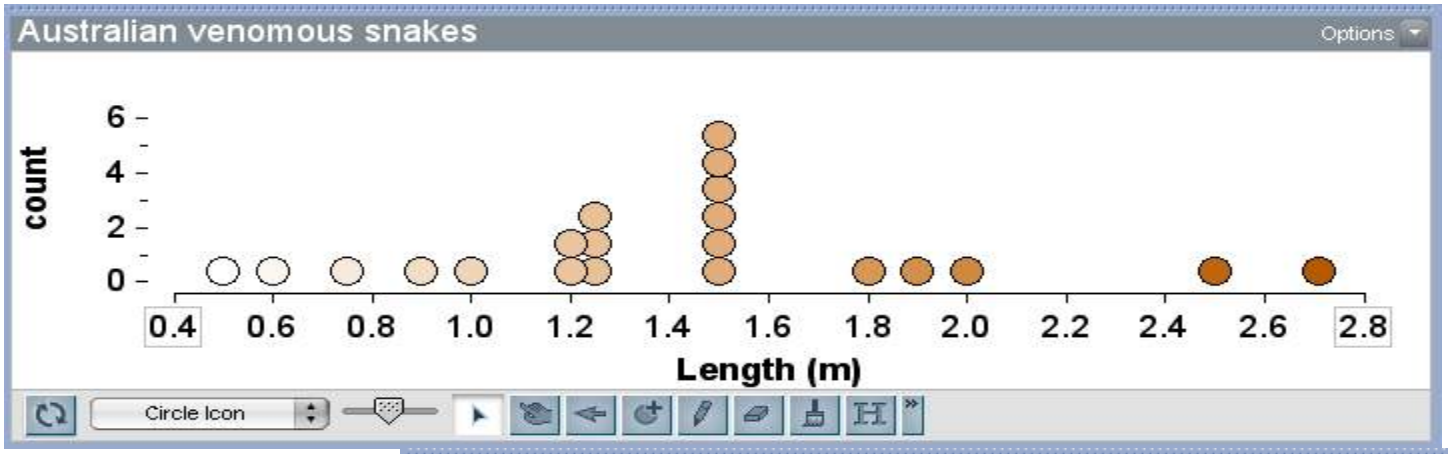
- Lesson 1: *TinkerPlots* Basics and Football
- Lesson 2: Introduce the Sampler and “toss” dice
- Lesson 3: Dividers, percent, and tossing 10 or 30 coins
- Lesson 4: The History tool and modeling the One-son Problem
- Lesson 5: Sampling from a finite population, the First Fleet
- Lesson 6: Introduce resampling, comparing two groups: memorizing nonsense and meaningful words
- Lesson 7: Use resampling for a 2-way table: swimming with dolphins
- Lesson 8: Assessment: Deciding if group differences in reaction times are ”significant” or not.

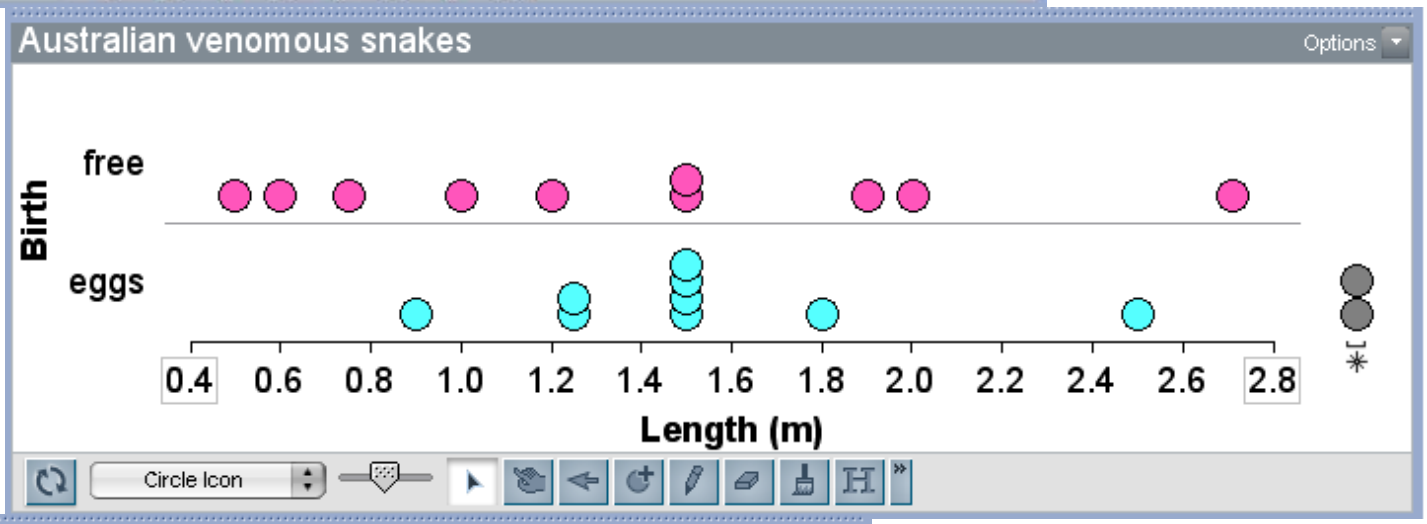
Lesson 1:
TinkerPlots Basics and Football

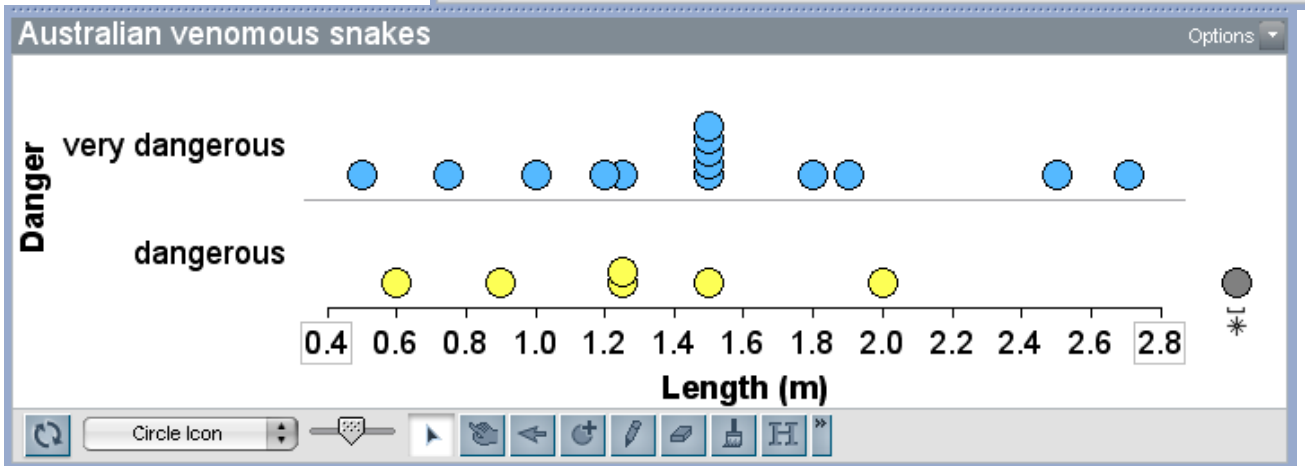
- Data cards and basic plots



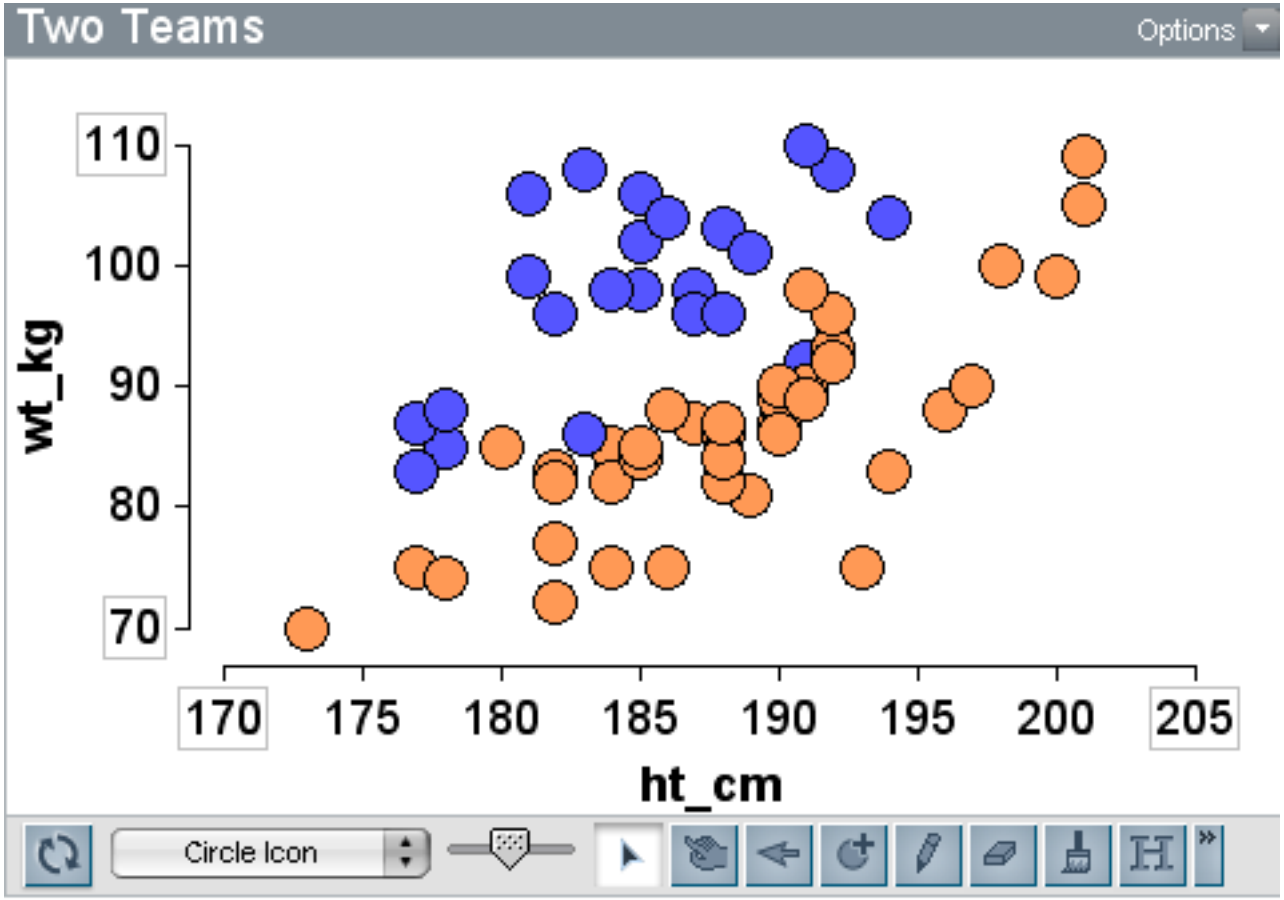



Australian venomous snakes
Options
count
6
4
2
0
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
Length (m)
Circle Icon

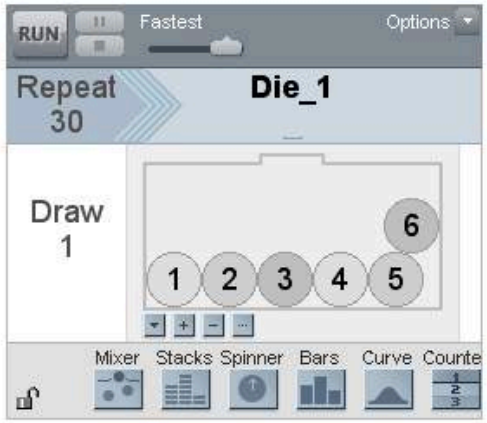

Australian venomous snakes
Options
Birth
free
eggs
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
Length (m)
Circle Icon

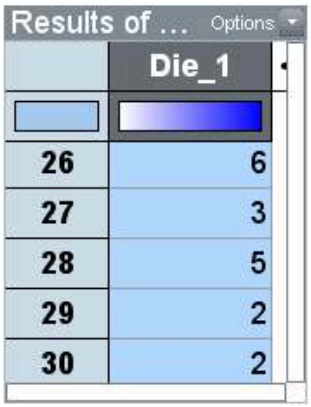

Australian venomous snakes
Options
Danger
very dangerous
dangerous
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
Length (m)
Circle Icon

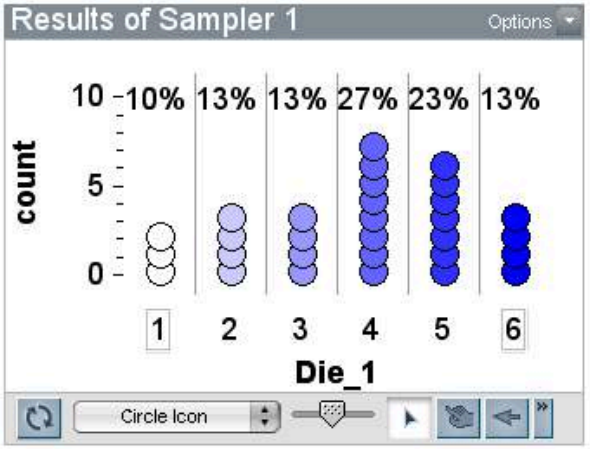
Which team is Rugby League and which is Australian Rules Football?

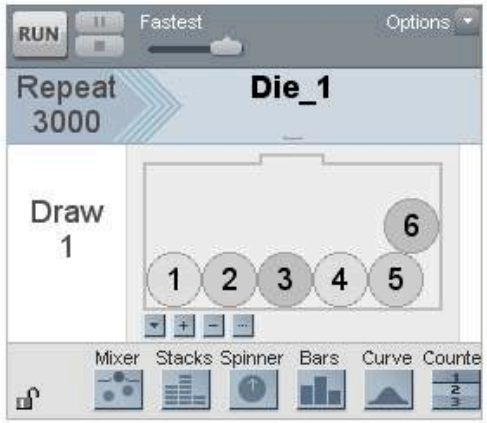


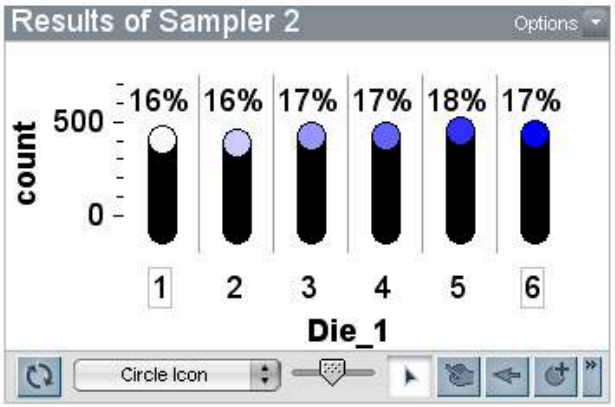
Lesson 2:
Introduce the Sampler and “toss” a die



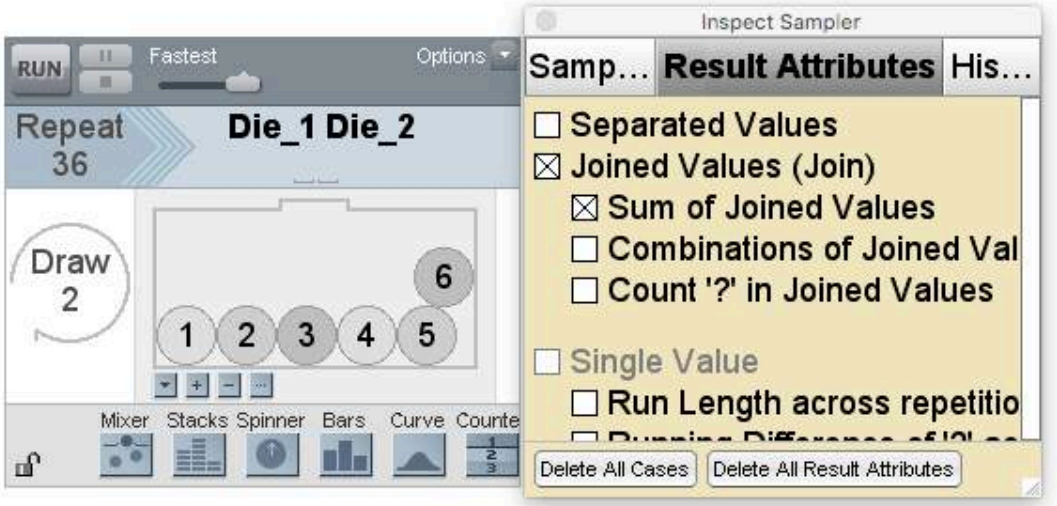


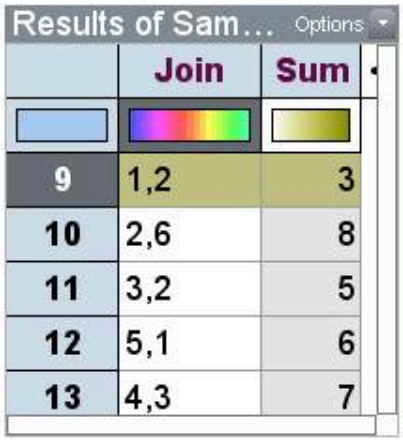


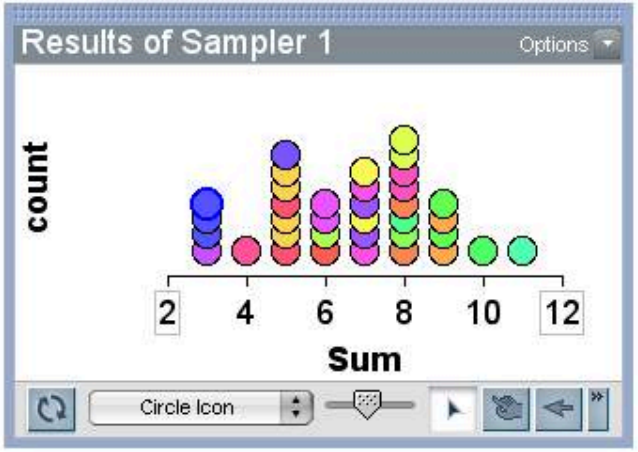




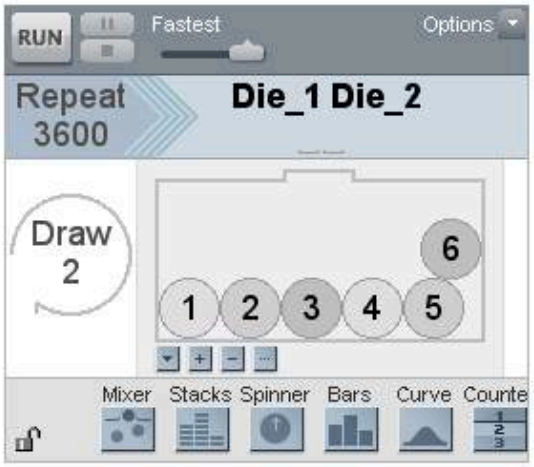
Lesson 2:
Use the Sampler to “toss” 2 dice

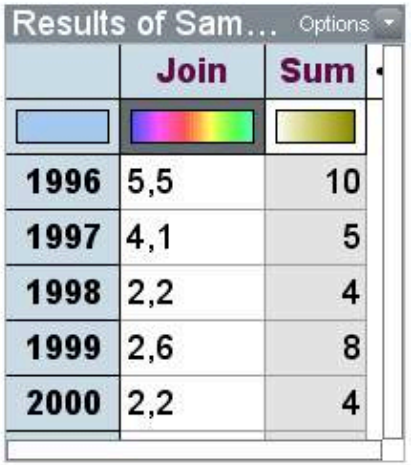


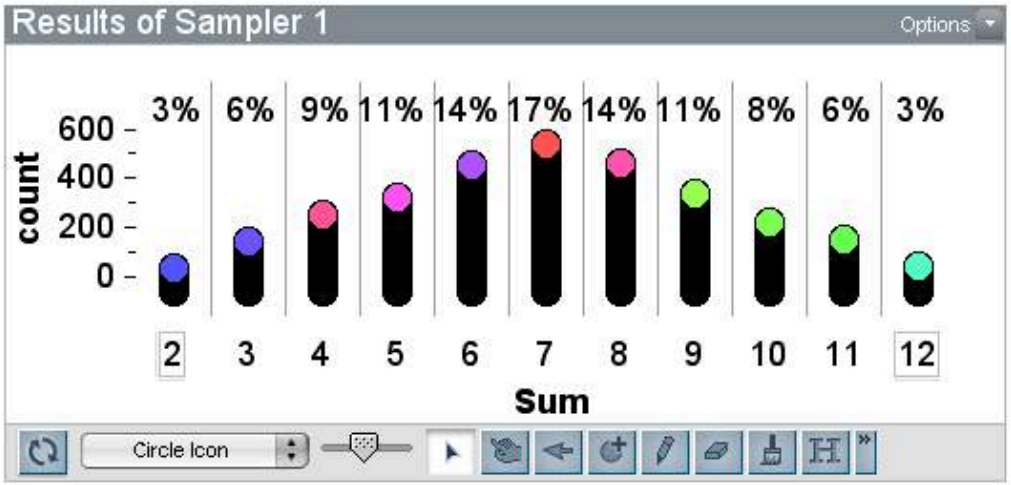




Lesson 2:
Use the Sampler to “toss” 2 dice

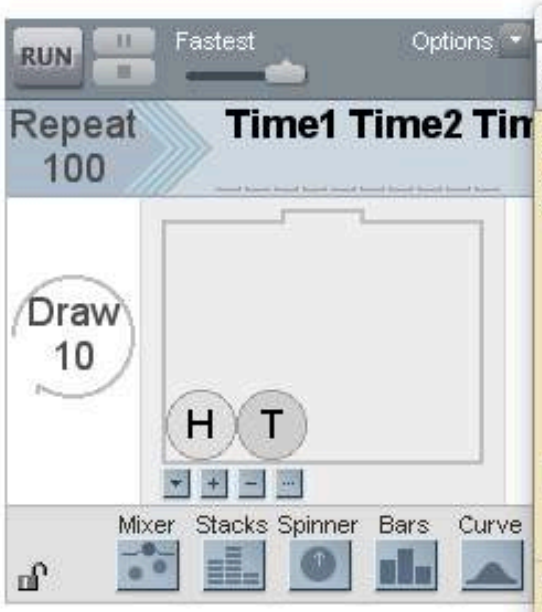


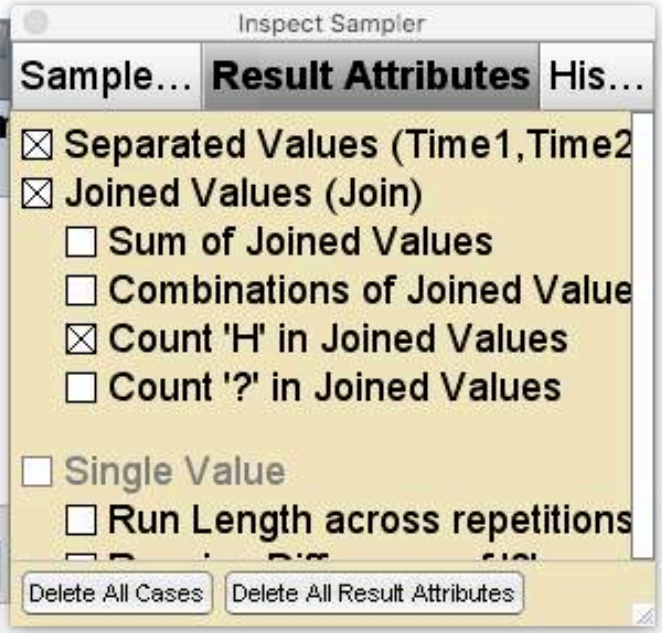


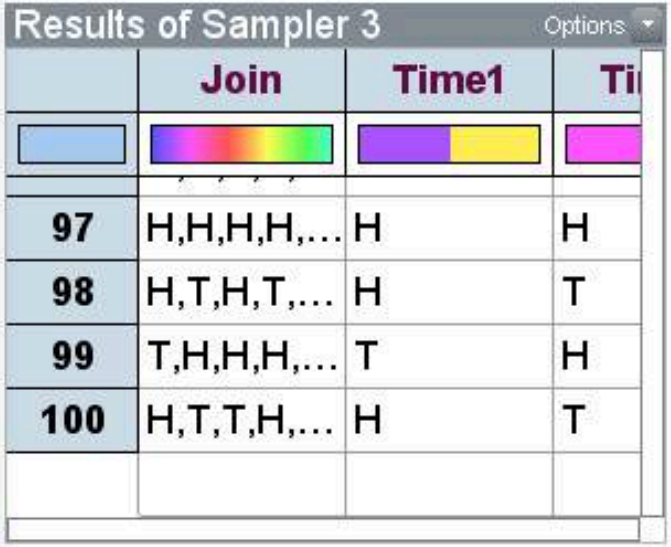


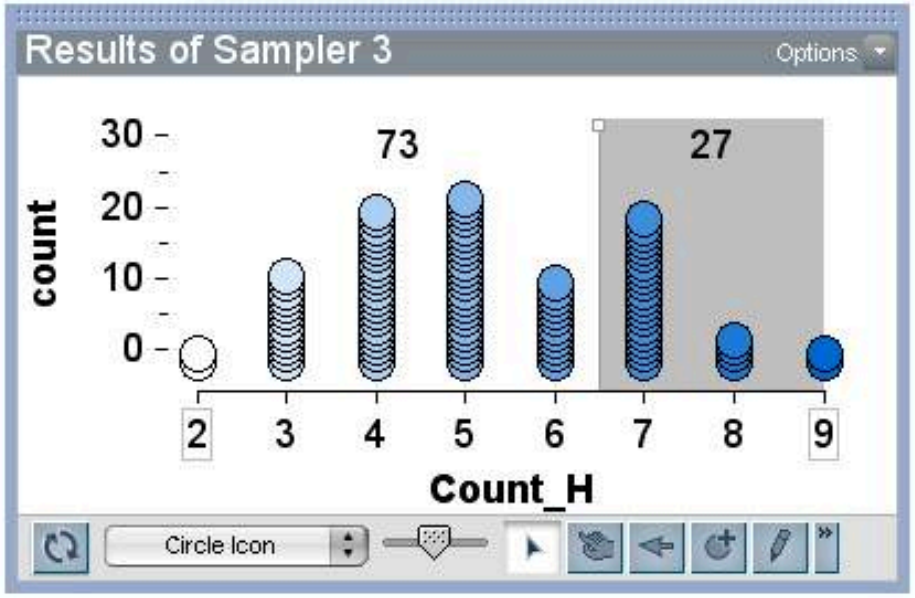
Lesson 3:
The “Hospital” problem with coins

Ted and Jed are each tossing a fair coin. Ted tosses his 10 times and Jed tosses his 30 times. Which one of them is more likely to get more than 60% heads or do they have the same chance? Explain why.

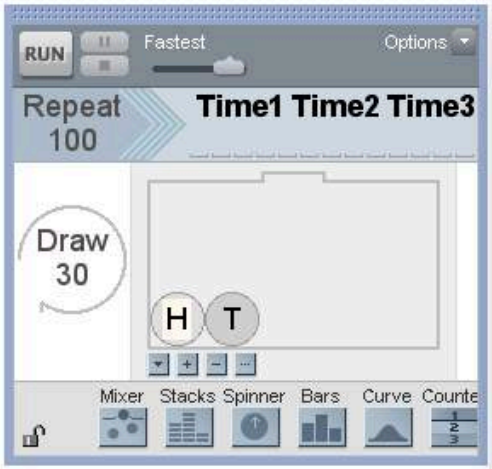

RUN
Fastest
Options
Repeat
100
Time1 Time2 Tin
Draw
10
H
T
Mixer
Stacks
Spinner
Bars
Curve

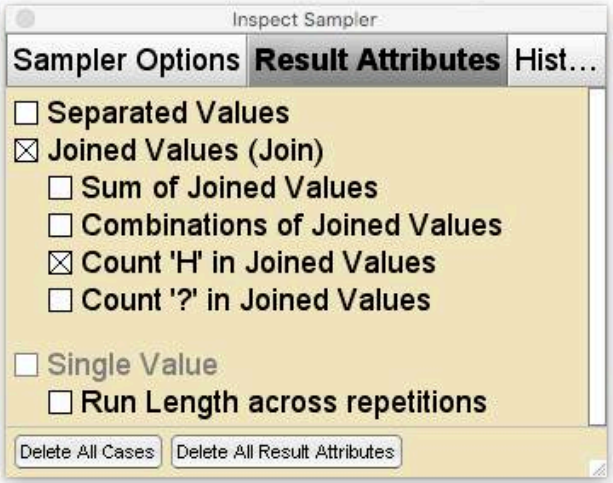

Inspect Sampler
Sample…
Result Attributes
His…
Separated Values (Time1,Time2
Joined Values (Join)
Sum of Joined Values
Combinations of Joined Value
Count 'H' in Joined Values
Count '?' in Joined Values
Single Value
Run Length across repetitions
Delete All Cases
Delete All Result Attributes

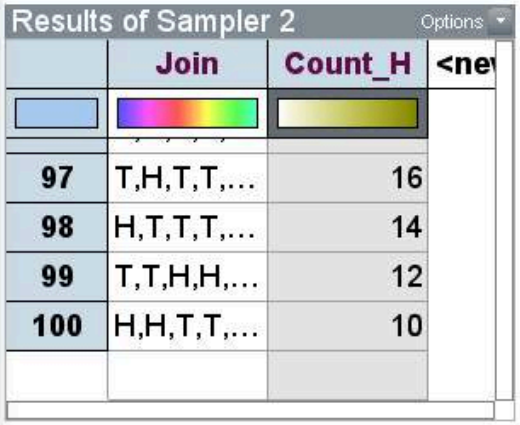

Results of Sampler 3
Options
Join
Time1
Ti
97
H,H,H,H,…
H
H
98
H,T,H,T,…
H
T
99
T,H,H,H,…
T
H
100
H,T,T,H,…
H
T


Results of Sampler 3
Options
30
20
10
0
count
73
27
2
3
4
5
6
7
8
9
Count_H
Circle Icon

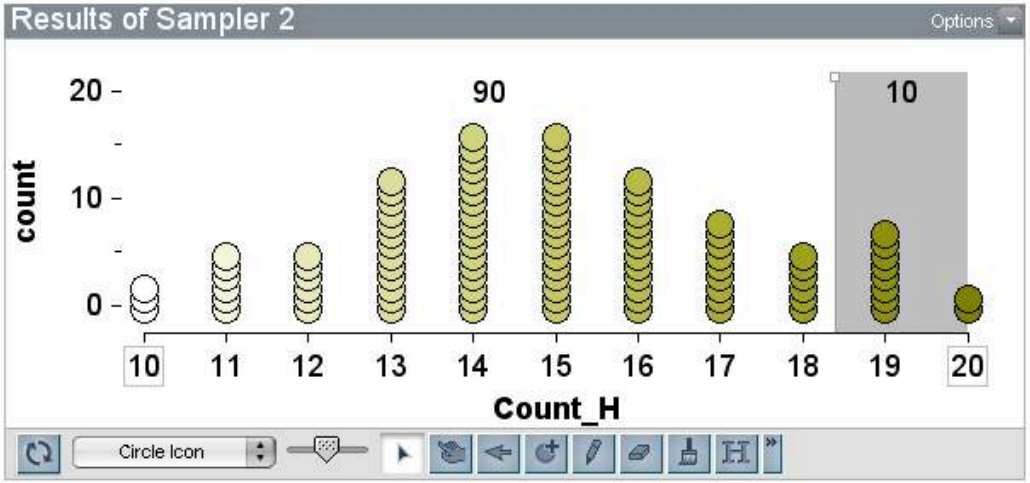
Ted

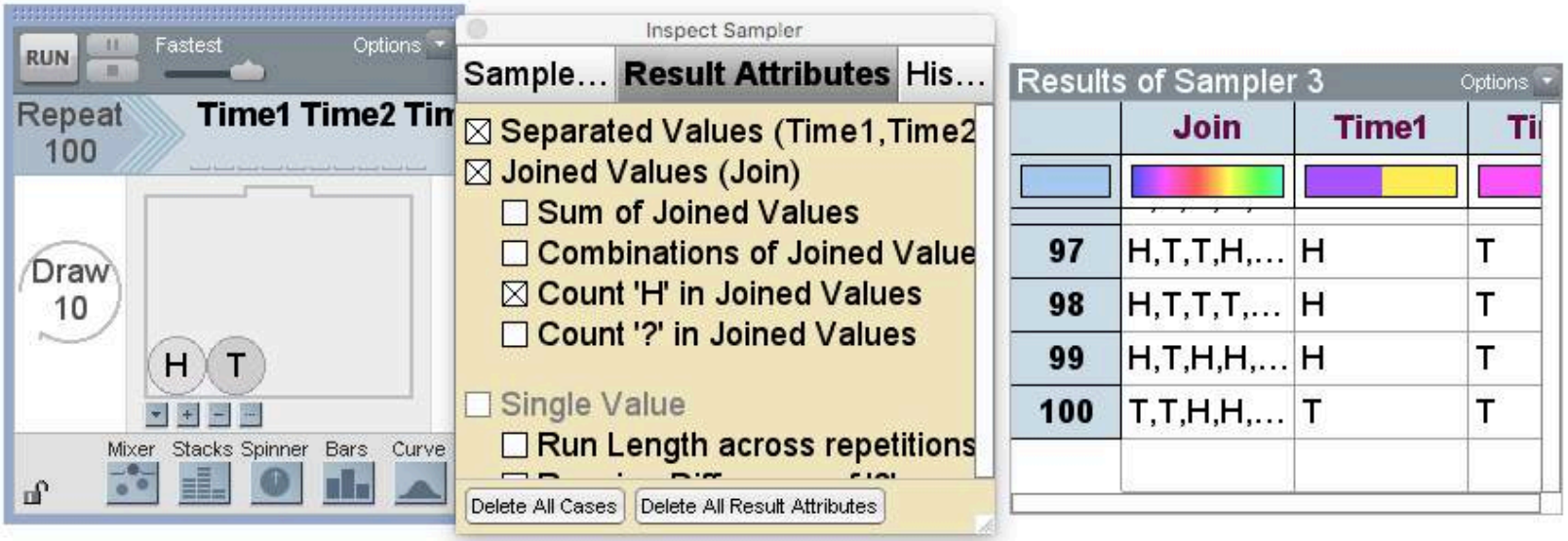

RUN
Fastest
Options
Repeat 100
Time1 Time2 Time3
Draw 30
H
T
Mixer
Stacks
Spinner
Bars
Curve
Counte

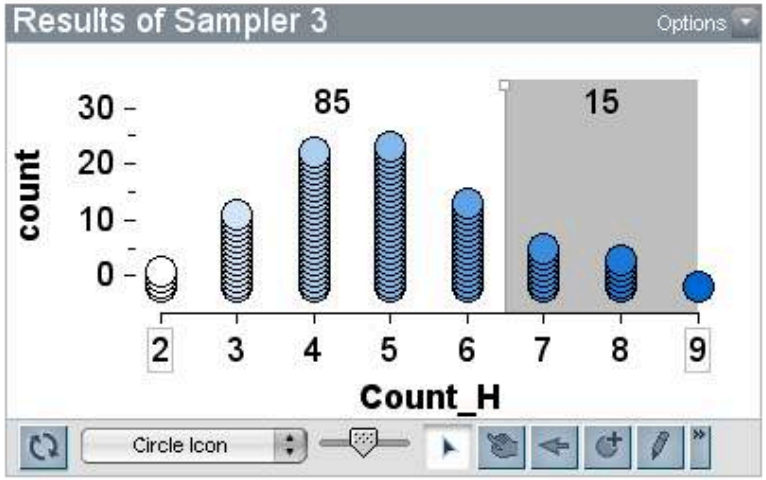

Inspect Sampler
Sampler Options
Result Attributes
Hist…
Separated Values
Joined Values (Join)
Sum of Joined Values
Combinations of Joined Values
Count 'H' in Joined Values
Count '?' in Joined Values
Single Value
Run Length across repetitions
Delete All Cases
Delete All Result Attributes


Results of Sampler 2
Options
Join
Count_H
97
T,H,T,T,…
16
98
H,T,T,T,…
14
99
T,T,H,H,…
12
100
H,H,T,T,…
10

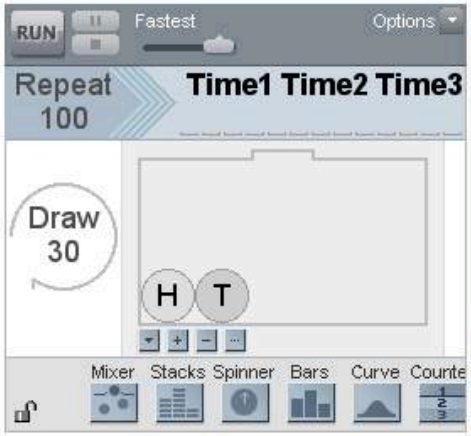
Jed

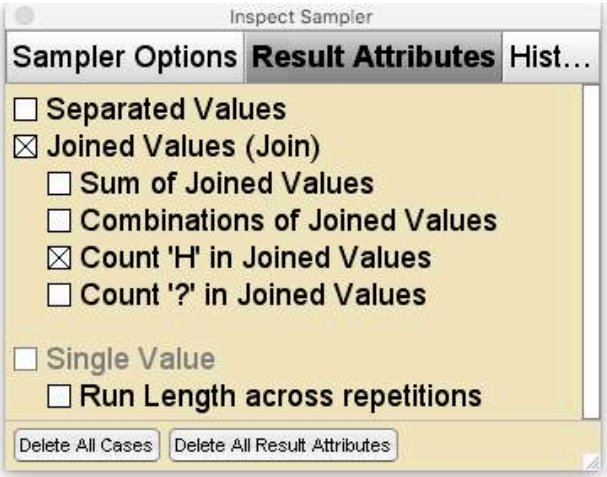

Results of Sampler 2
Options
count
20
10
0
90
10
10
11
12
13
14
15
16
17
18
19
20
Count_H
Circle Icon

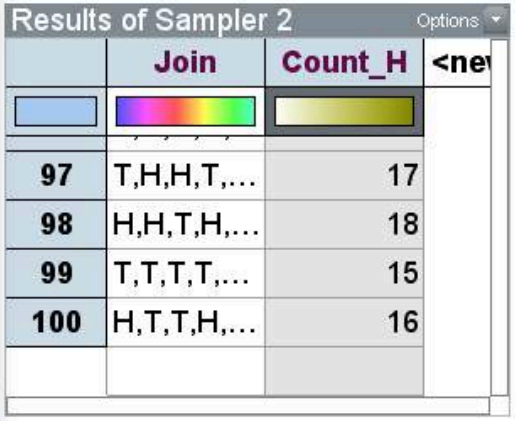

RUN
Fastest
Options
Repeat
100
Time1 Time2 Tim
Draw
10
H
T
Mixer
Stacks
Spinner
Bars
Curve
Inspect Sampler
Sample…
Result Attributes
His…
Separated Values (Time1,Time2
Joined Values (Join)
Sum of Joined Values
Combinations of Joined Value
Count 'H' in Joined Values
Count '?' in Joined Values
Single Value
Run Length across repetitions
Delete All Cases
Delete All Result Attributes
Results of Sampler 3
Options
Join
Time1
Ti
97
H,T,T,H,…
H
T
98
H,T,T,T,…
H
T
99
H,T,H,H,…
H
T
100
T,T,H,H,…
T
T


Results of Sampler 3
Options
count
30
20
10
0
85
15
2
3
4
5
6
7
8
9
Count_H
Circle Icon

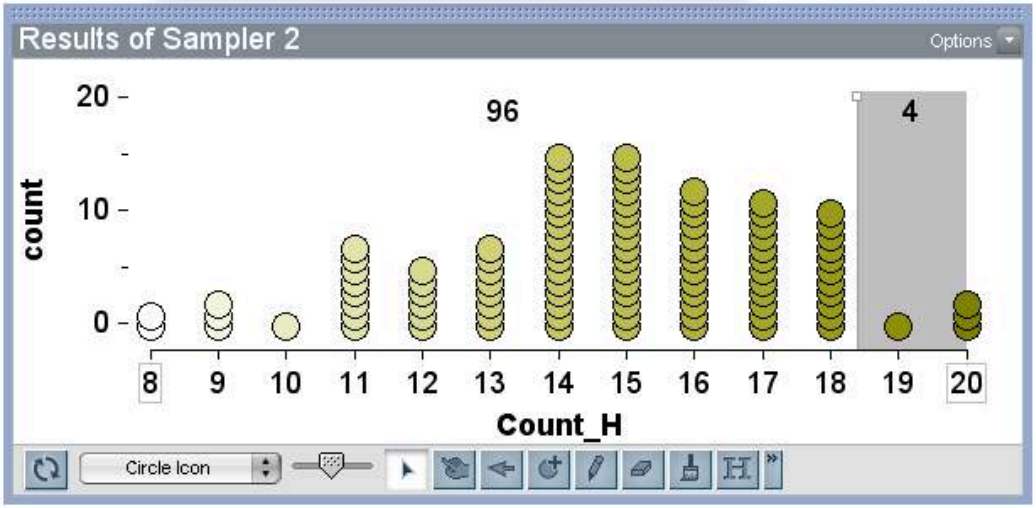
Ted


RUN
Fastest
Options
Repeat 100
Time1 Time2 Time3
Draw 30
H
T
Mixer
Stacks
Spinner
Bars
Curve
Counte


Inspect Sampler
Sampler Options
Result Attributes
Hist...
Separated Values
Joined Values (Join)
Sum of Joined Values
Combinations of Joined Values
Count 'H' in Joined Values
Count '?' in Joined Values
Single Value
Run Length across repetitions
Delete All Cases
Delete All Result Attributes


Results of Sampler 2
Options
Join
Count_H
97 T,H,H,T,... 17
98 H,H,T,H,... 18
99 T,T,T,T,... 15
100 H,T,T,H,... 16

Jed

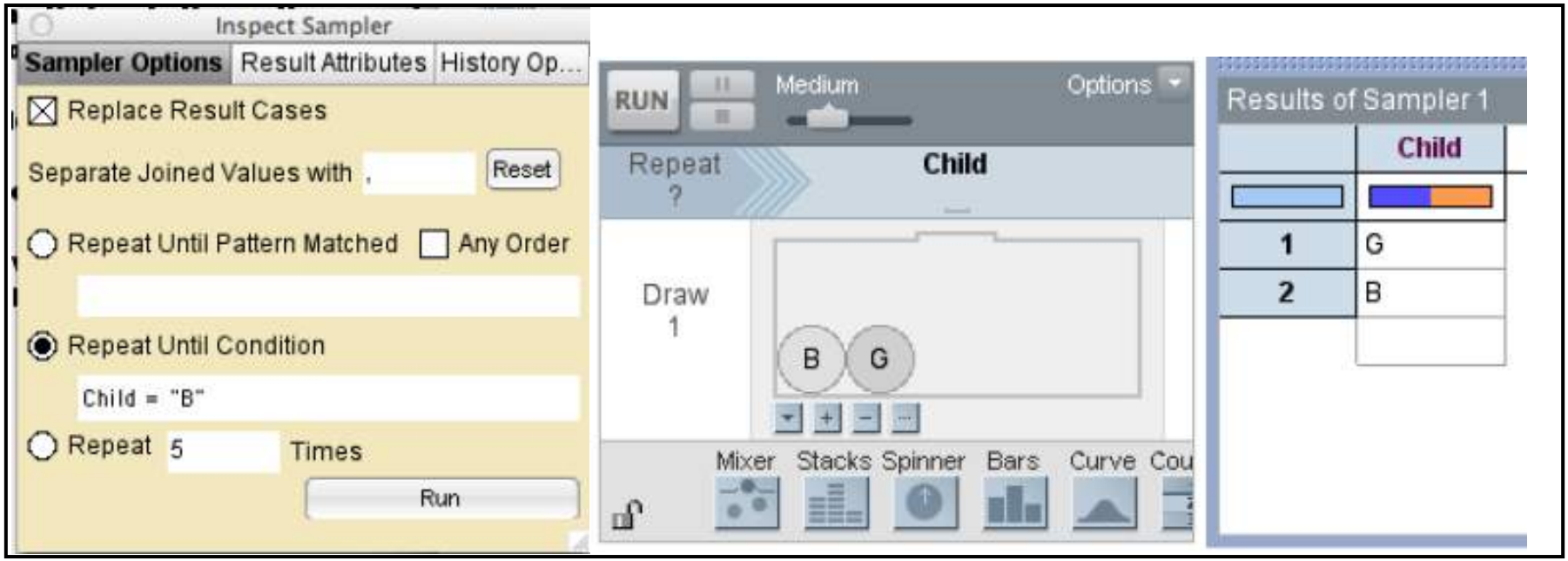

Results of Sampler 2
Options
count
20
10
0
96
4
8 9 10 11 12 13 14 15 16 17 18 19 20
Count_H
Circle Icon

Lesson 4:
One-son problem (Konold, 1994)

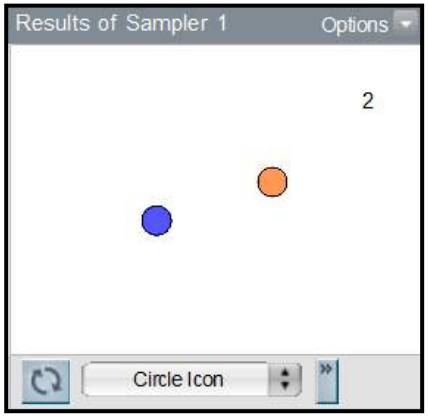
- Some years ago, to limit population growth, the Chinese instituted a “one-child” policy.
- The policy was unpopular because of the desire for the child to be a son.
- Over time this has resulted in an imbalance of gender: fewer girls.
- What if couples were allowed to have children until they had one son?
- Would the population explode? Would there be a different imbalance? Too many girls?

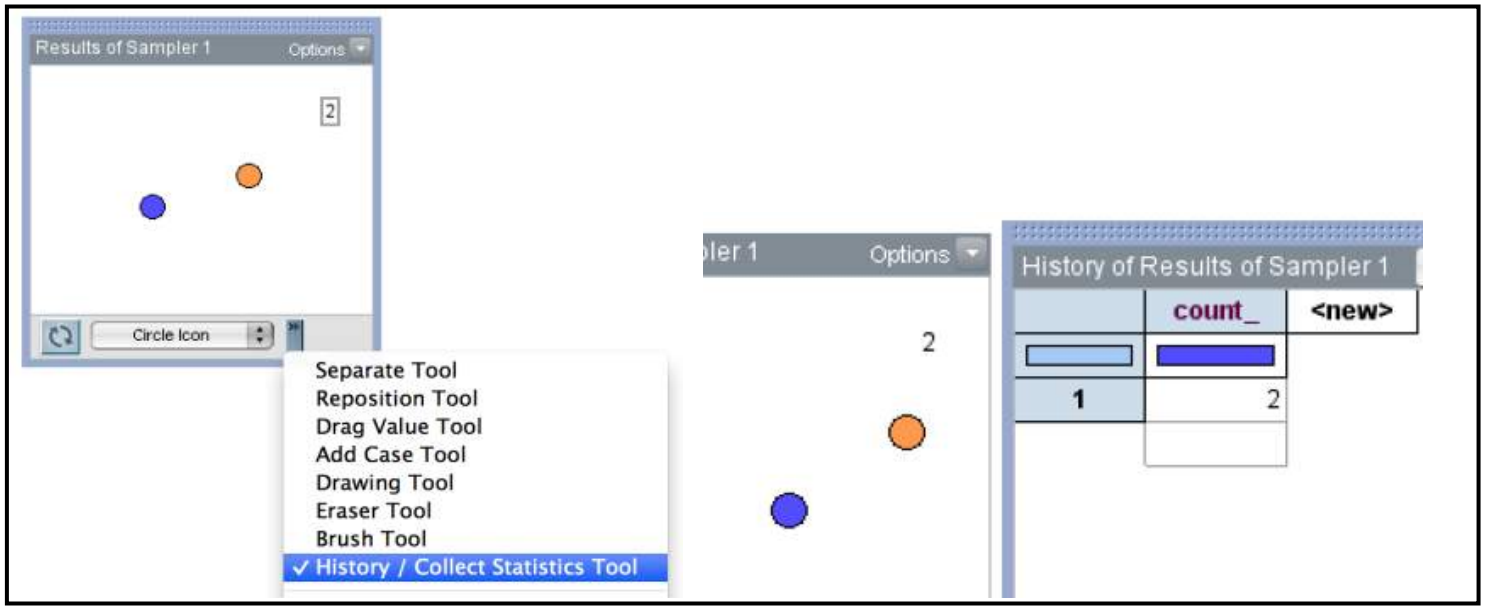
One-son problem (Konold, 1994)

- Hypotheses expressed as questions based on the mathematics involved:
    1. What would the average number of children be in a family?
    2. What would be the ratio of births of girls to births of boys?
- Assume gender of a baby is a random phenomenon. Only single births are considered.
- The Sampler in *TinkerPlots* simulates births based on the binomial model with p=½.
- The pseudo-random design of the Sampler provides the variation required for answering each of the questions.

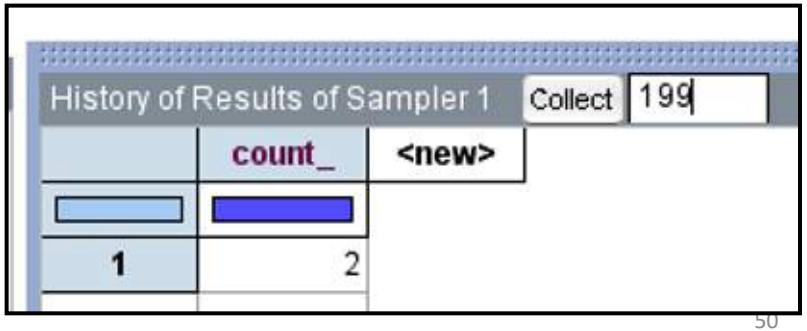


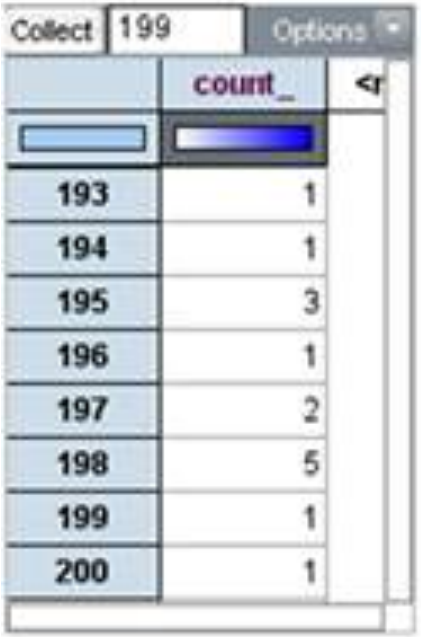
- Set up Sampler with B and G in Mixer, set Draw to 1 and label Attribute “Child”.
- Under Sampler Options in the Options menu (upper right corner), choose “Repeat Until Condition” and insert, Child = “B”. Click Run once.
- Drag down a Plot for the Results.





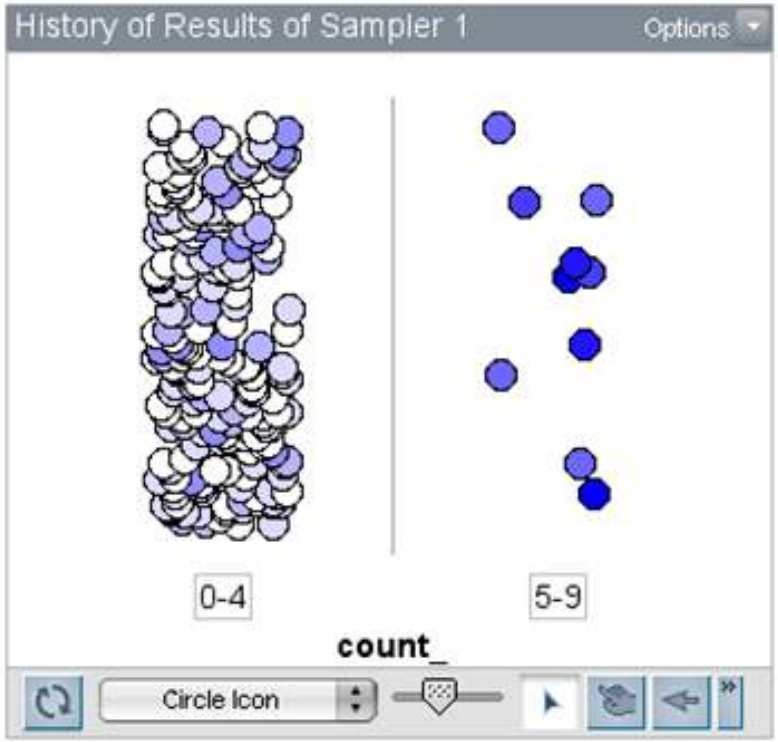
- Click on the History button in the menu below the plot.
- A grey box appears around the number displayed.
- Double click on the box to create a History of Results of Sampler 1 table recording the count from the plot.
- Change Collect field to 199 for a total of 200 “births”.

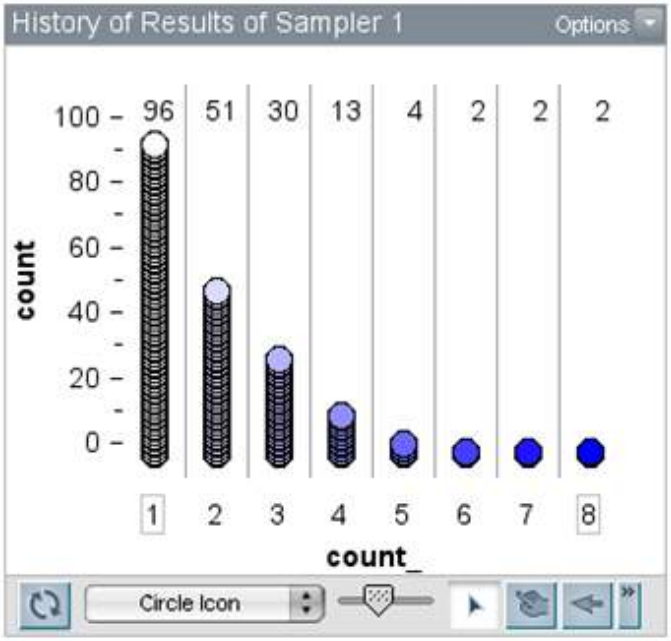




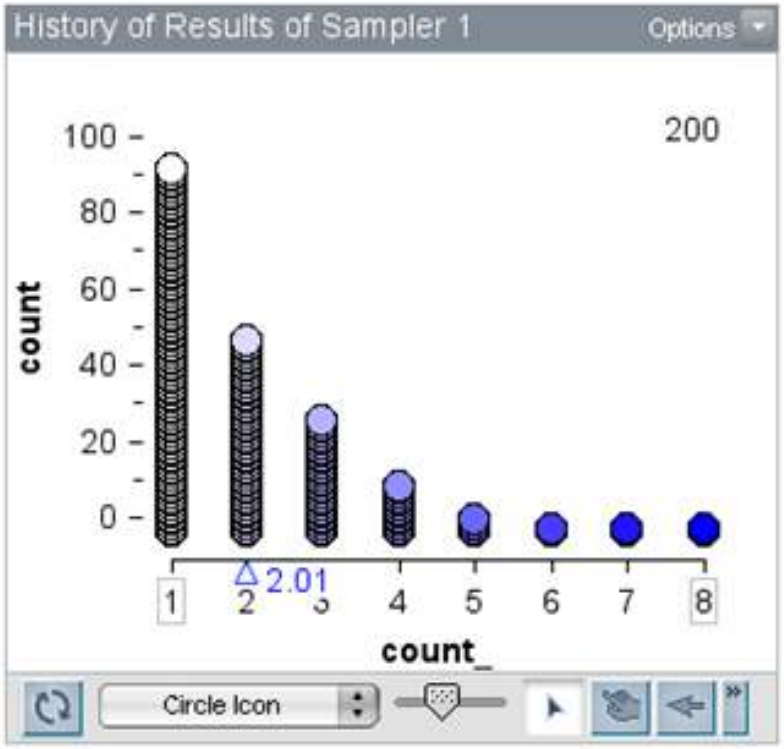
200 sizes of families

- Drag down a plot and then drag **count_** from History of Results of Samples 1 to the horizontal axis.

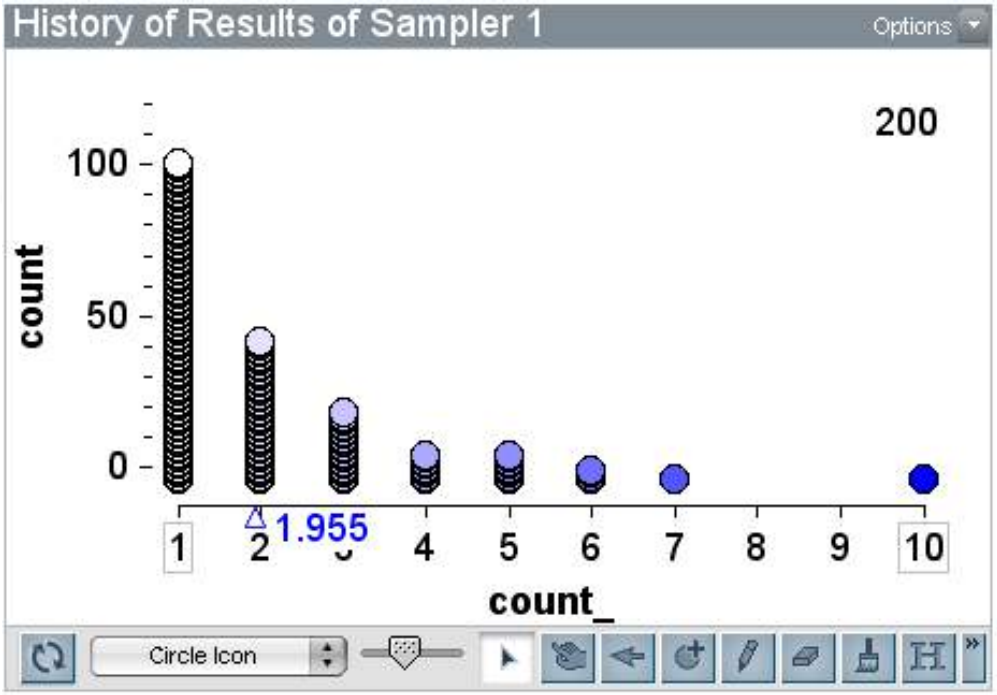




Drag and stack to bins for family size to first boy. Bins show number of each size.

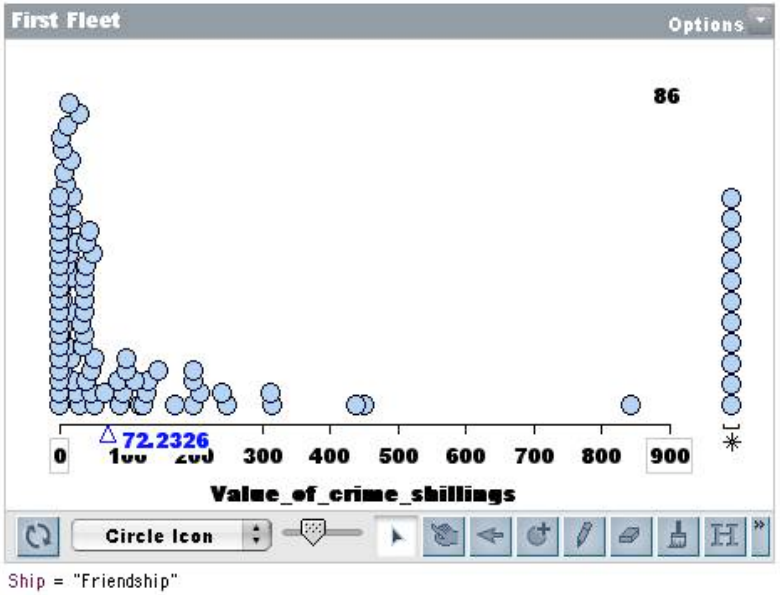


Another sample



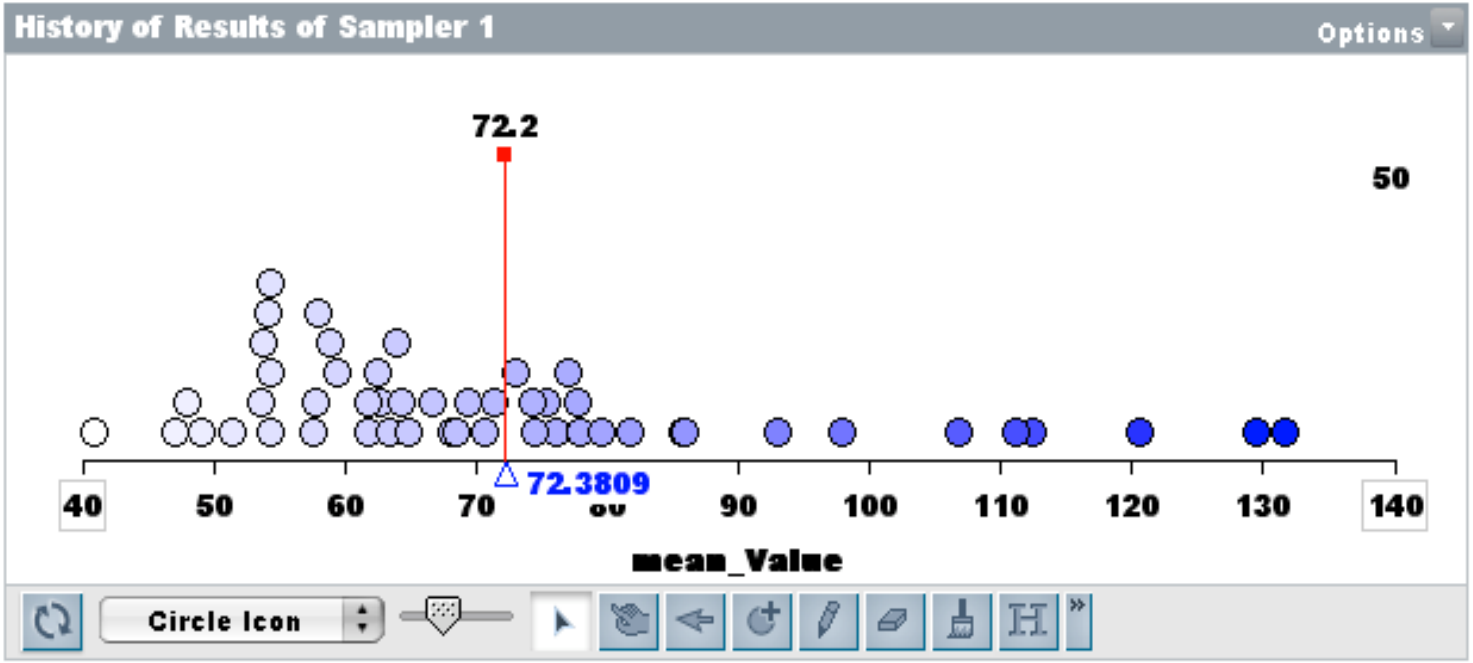
Lesson 5:
Sampling from a Finite Population

- The “population” for Lesson 5 is data for all 780 convicts who were transported to Australia in the First Fleet.
- One purposeful sample is the “Friendship” with 86 convicts.
- One variable, “Value_of_crime_shillings” has a skewed distribution with mean value 72.23 shillings. Is it representative of the entire Fleet?



Lesson 5:
Sampling from a Finite Population

- Again using the History feature to collect the means from 50 samples of size 86 from the entire Fleet, for the variable “Value_of_crime_shillings”, it is seen that this variable, the “Friendship” is typical of the Fleet.

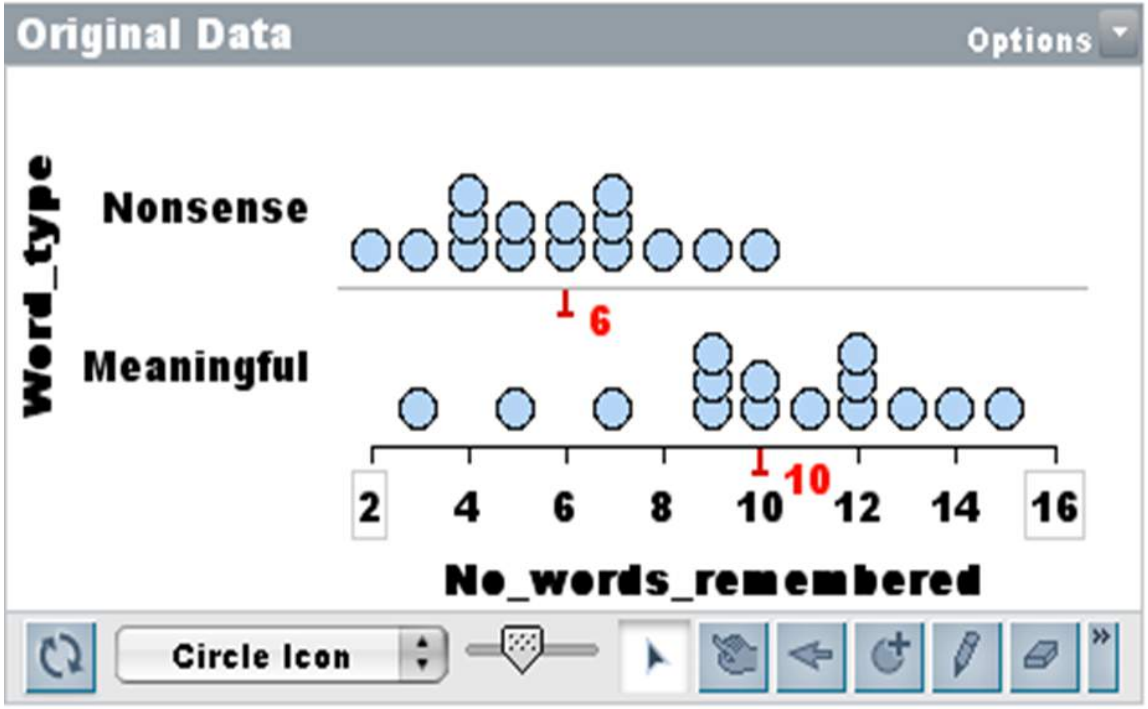


Lesson 6:
Resampling to consider the difference in medians

- Experiment in class: Is it easier to memorize meaningful or nonsense 3-letter words? (Shaughnessy, Chance, and Kranendonk, 2009)
- Students devised experiment given 20 meaningful 3-letter words (e.g., DOG, CAT) and 20 nonsense words (e.g., ATC,ODG): Working in pairs with timers, they had 2 minutes to memorize the words and then 1 minute to write down as many as they could remember (checking at the end). Half of the class started with meaningful words and the other half started with nonsense words. Each student then had two data values to report.

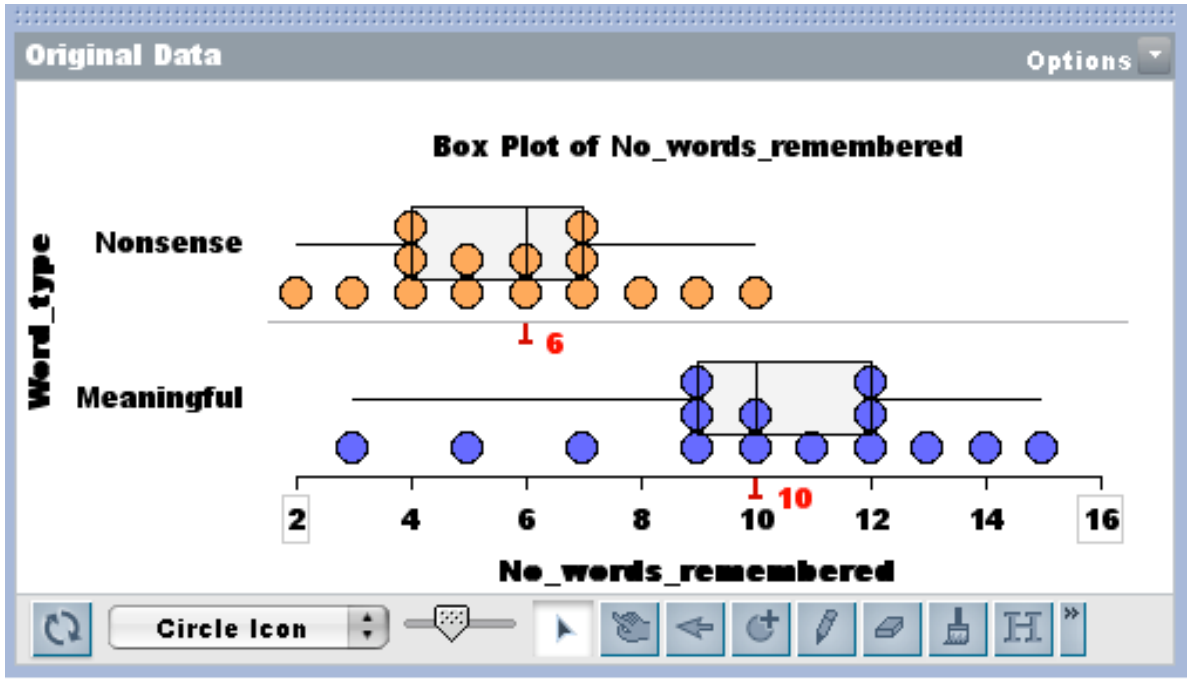
Question: Is it easier to remember meaningful or nonsense words?

- Class data from one Grade 10 class.
- How unusual is the difference?



One type of analysis: box plots

No overlap of the boxes of the box plots means 3/4 of “Nonsense” scores are to the left of 3/4 of “Meaningful” scores. There appears to be evidence for a difference.



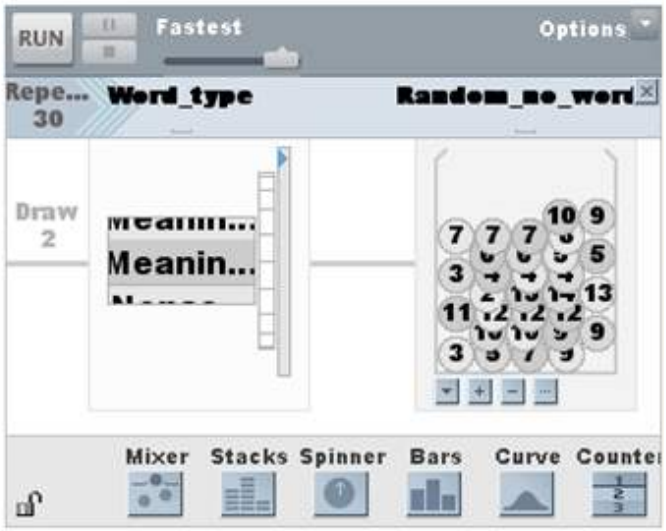
Resampling Analysis

- If there were no difference in memorising the two types of words and the data were reallocated randomly how often would the difference be as great as 4 words?
- In other words, how often would we get a result as large as this by chance?
- Students can reallocate the data randomly by hand.
- Or *TinkerPlots* can be used.

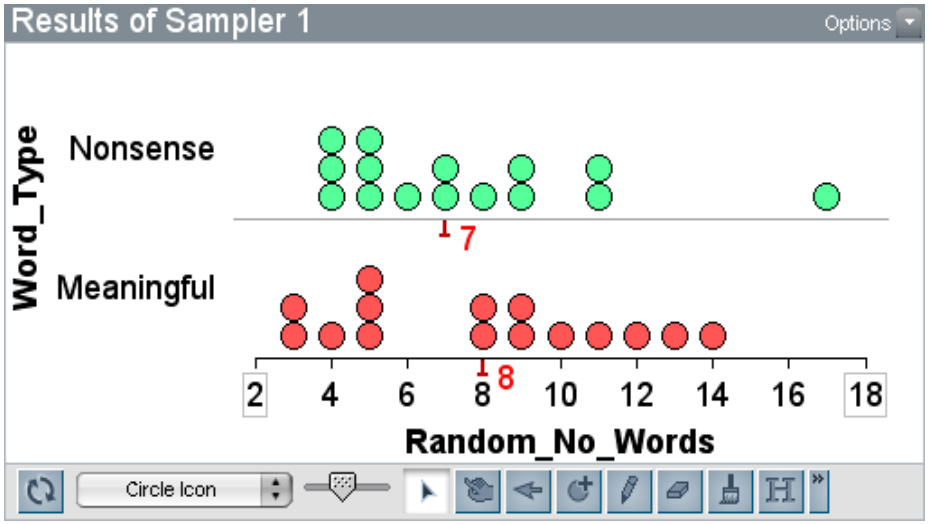
In *TinkerPlots* …

- The Sampler in *TinkerPlots* randomly reallocates the number of words remembered to the two conditions.

Sampler



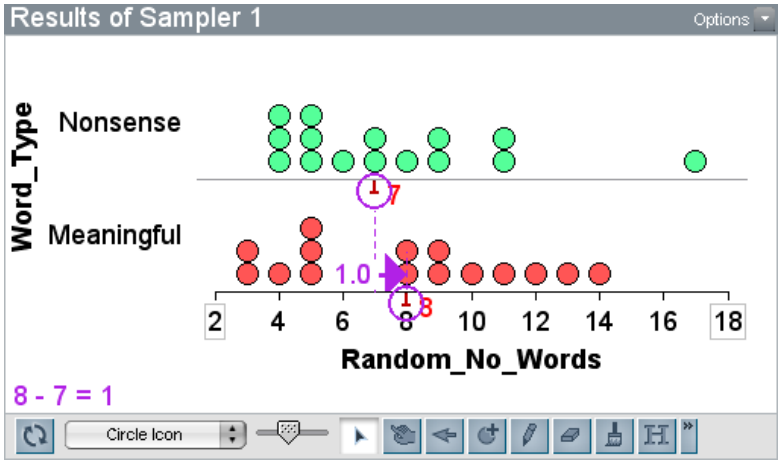
Plot



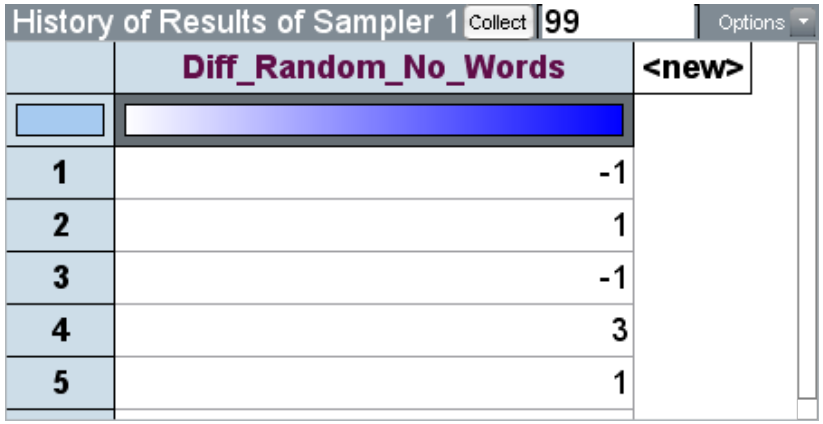
In *TinkerPlots* …

- The Ruler measures the new difference.
- The History button keeps track of many such Samples, say 100 times.

Plot



History



In *TinkerPlots* …

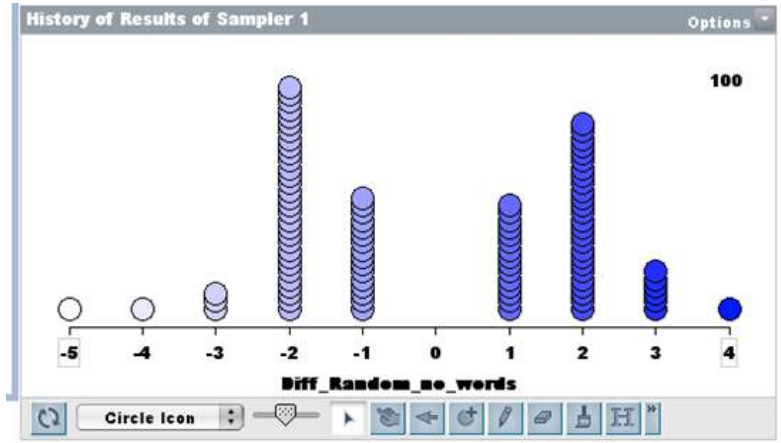
- Plotting the result of the 100 trials shows how many times the difference is 4 or more, as evidence supporting the conclusion that it is more difficult to memorise random words.

History

100 Samples

History of Results of Sampler 1 | Collect | 99 | Options

	Diff_Random_No_Words	<new>
1	-1	
2	1	
3	-1	
4	3	
5	1	

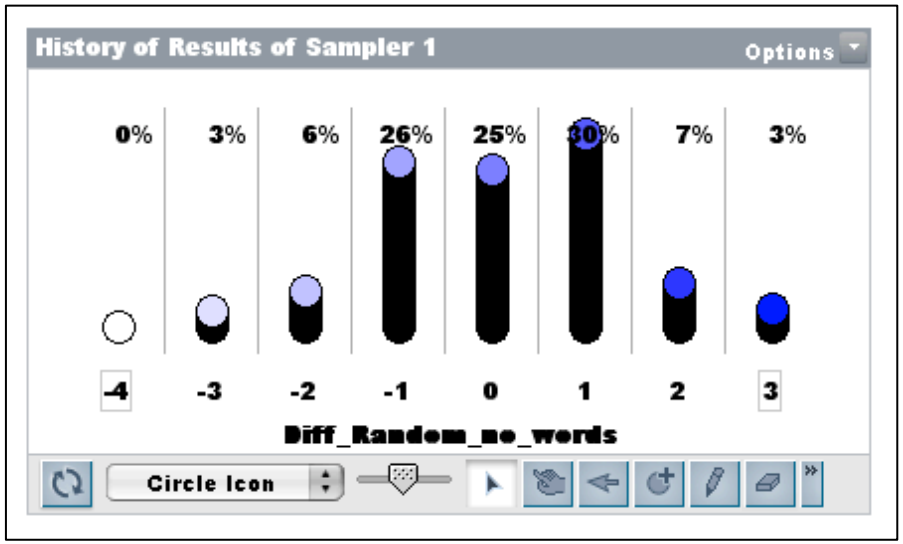


Student report

9. The difference in the original class one was 4 but in the sampler 4 was an extreme. This would of been unlikely to have occured by chance and shows that it was caused rather than just having randomly acured. This show that meaningful words can be remembered more easily then nonsense words.

11. We collected 402 random samples less then 0% were the same as the original samples. This is unlikely that it occured via chance.

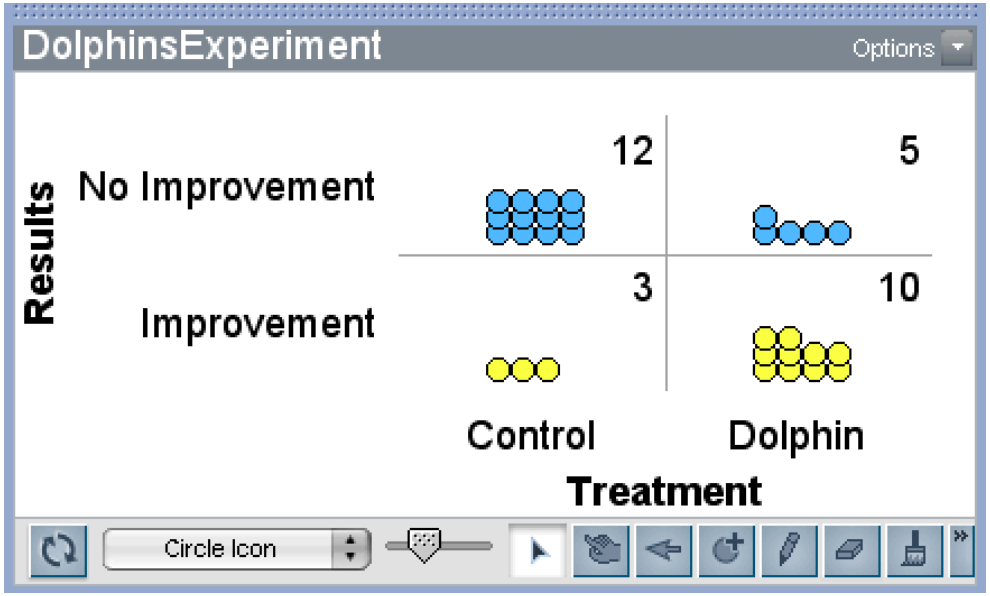
12. We are 99% sure that this didn't occur by chance. This very much confirms our suspicion that meaningful words are more easier to remember then nonsense words.



Lesson 7: Swimming with Dolphins (Rossman, 2008)

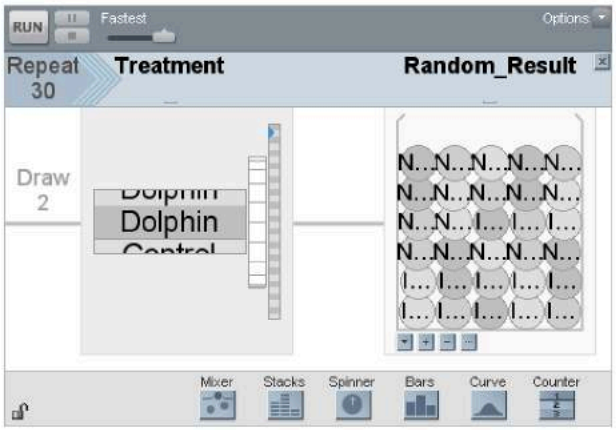
- Data are presented in a two-way table to investigate whether for people with mild to moderate depression swimming 4 hours per day with dolphins produces more improvement in depression that swimming 4 hours per day without dolphins.

- 30 patients were randomly allocated to the two treatments, Control (swimming 4 hours per day in the Caribbean for 4 weeks) and Dolphins (swimming 4 hours per day in the Caribbean for 4 weeks in the presence of dolphins).



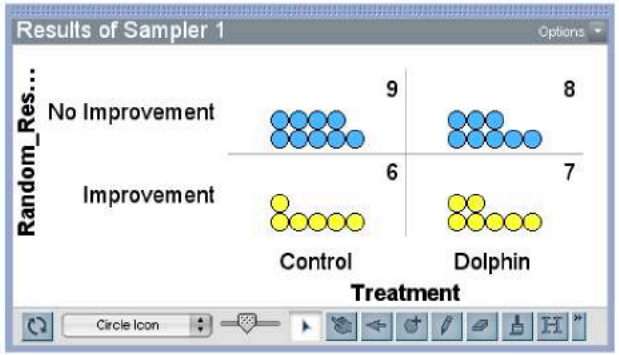
Using *TinkerPlots*…

- The data were placed in the Sampler to randomly allocate one of the Results (in the Mixer) to one of the Treatments (in the Counter), without replacement.
- A RUN of 30 produced the Results of Sampler 1, which are shown.



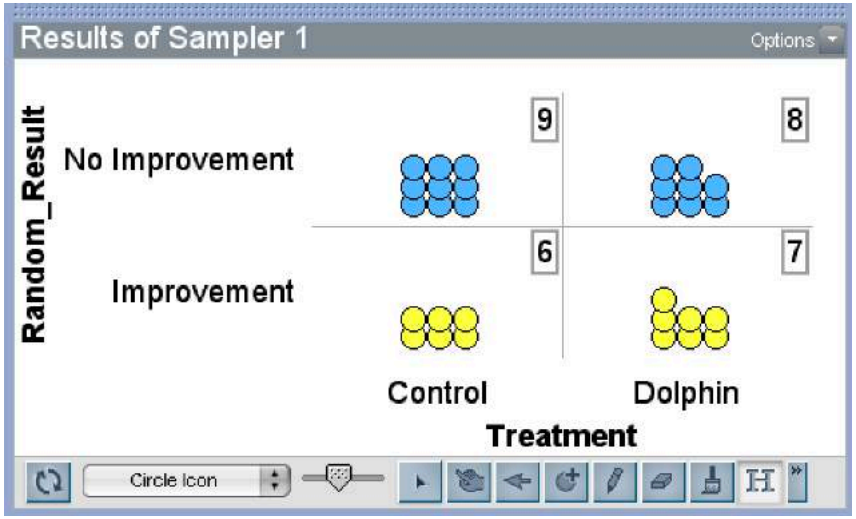
Results of Sampler 1 Options

	Treatment	Random_Result
9	Dolphin	Improvement
10	Dolphin	Improvement
11	Dolphin	No Improvement
12	Dolphin	No Improvement
13	Dolphin	No Improvement
14	Dolphin	No Improvement
15	Dolphin	No Improvement
16	Control	No Improvement
17	Control	Improvement

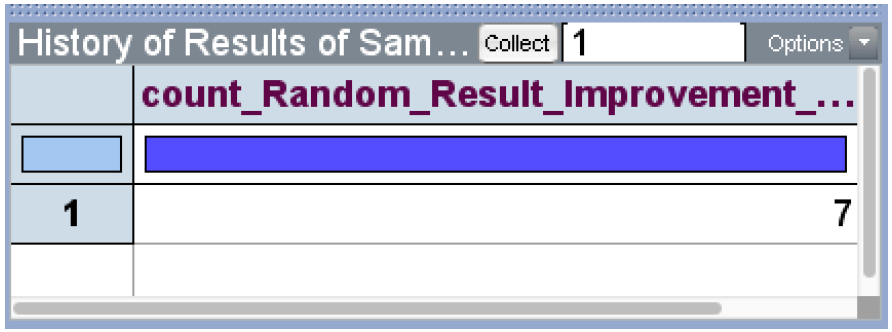


Using the Plot created from the Random Sampler…

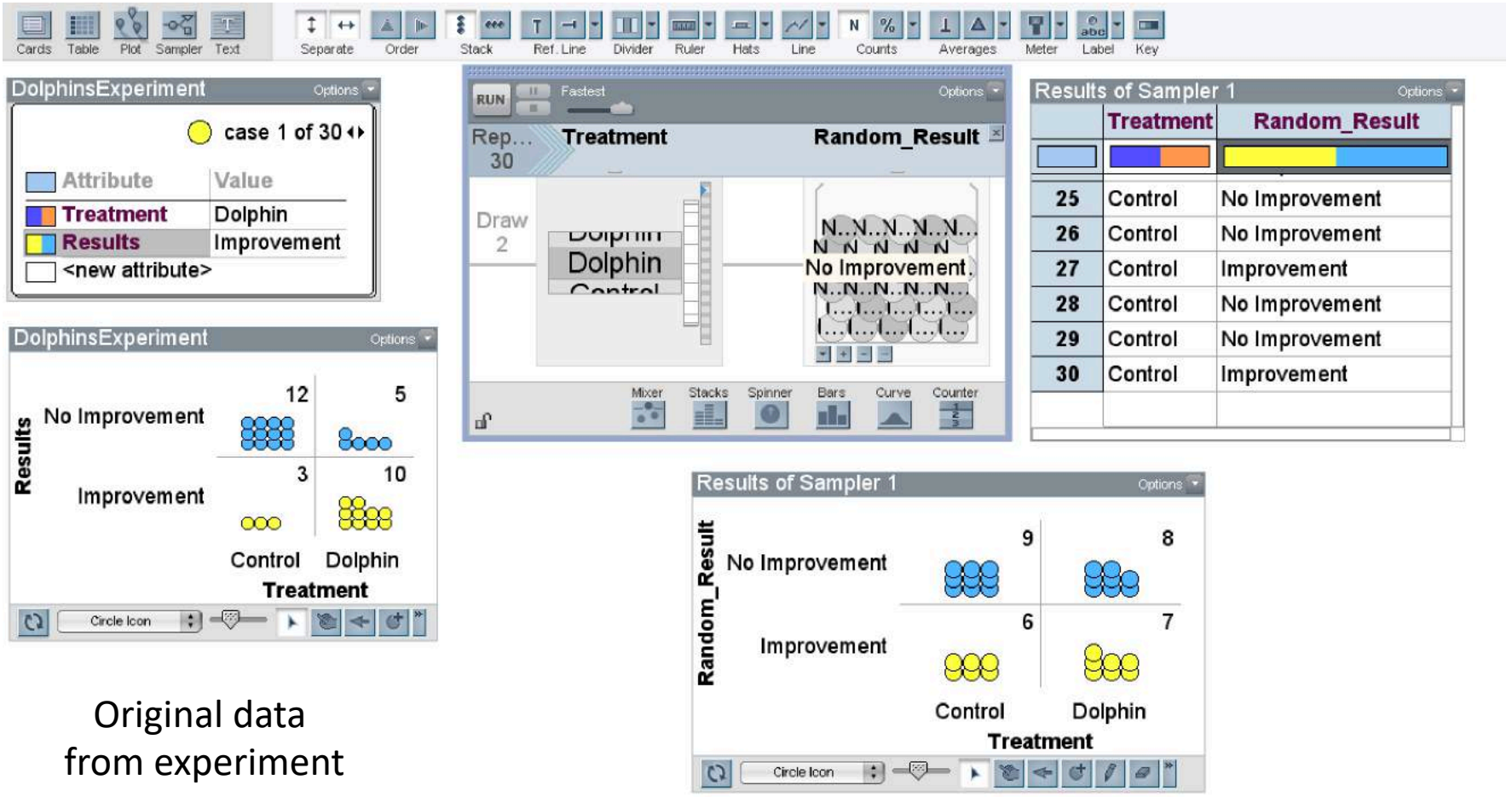
- The History button is used to count the number of people swimming with Dolphins whose Depression Improved (below, the 7).



- The collection of Results for Dolphin/Improvement beings.

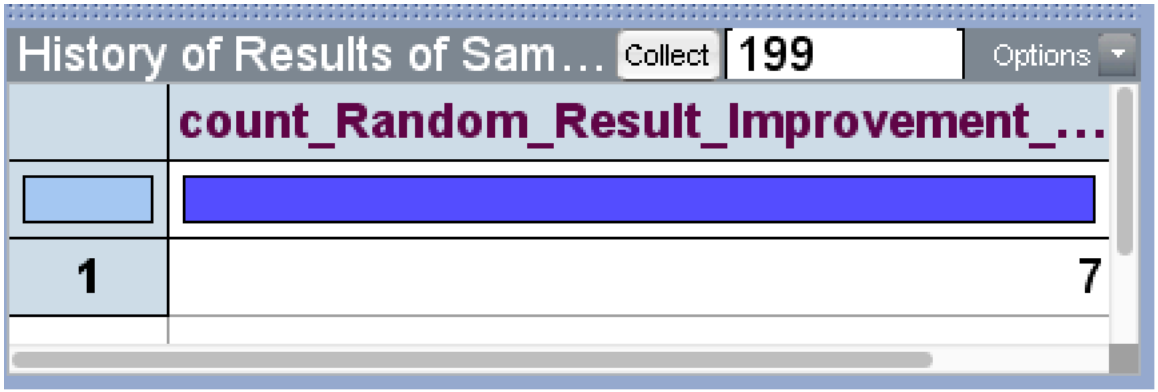


One result from Resampling

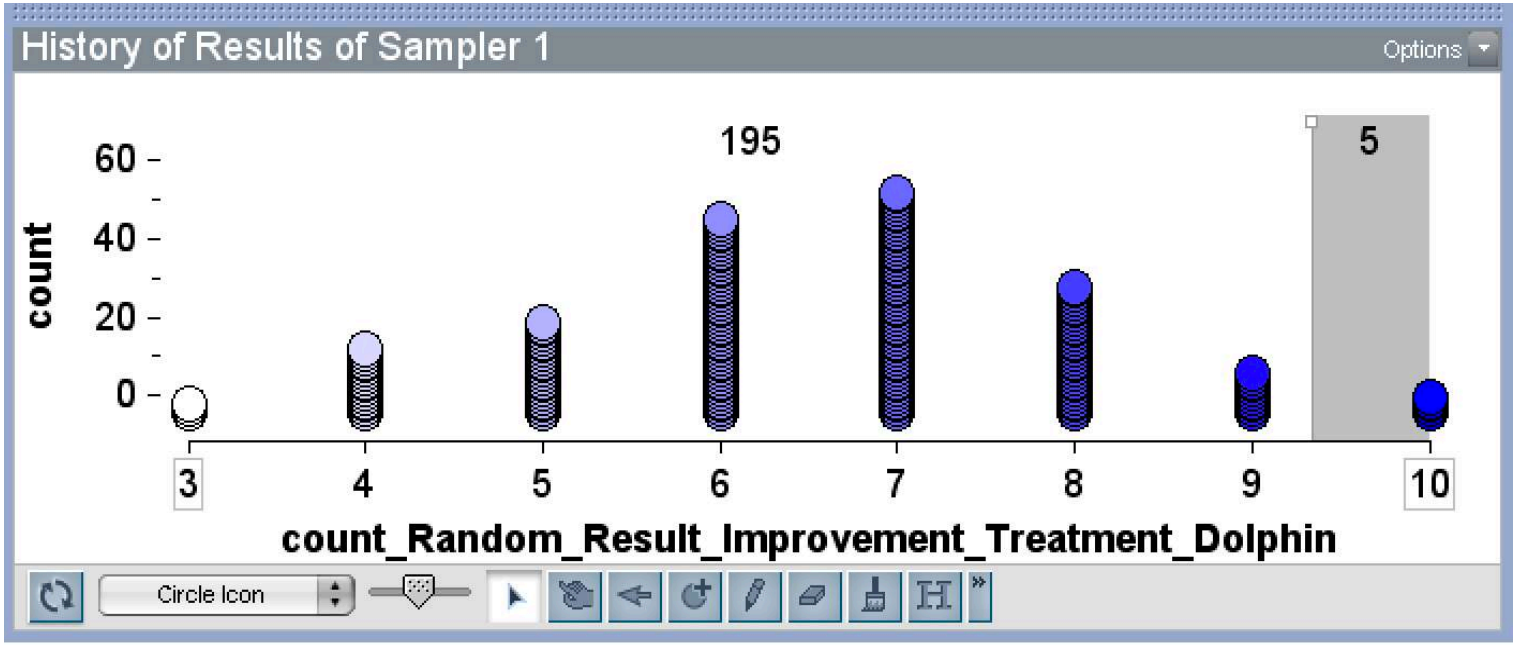


Original data
from experiment

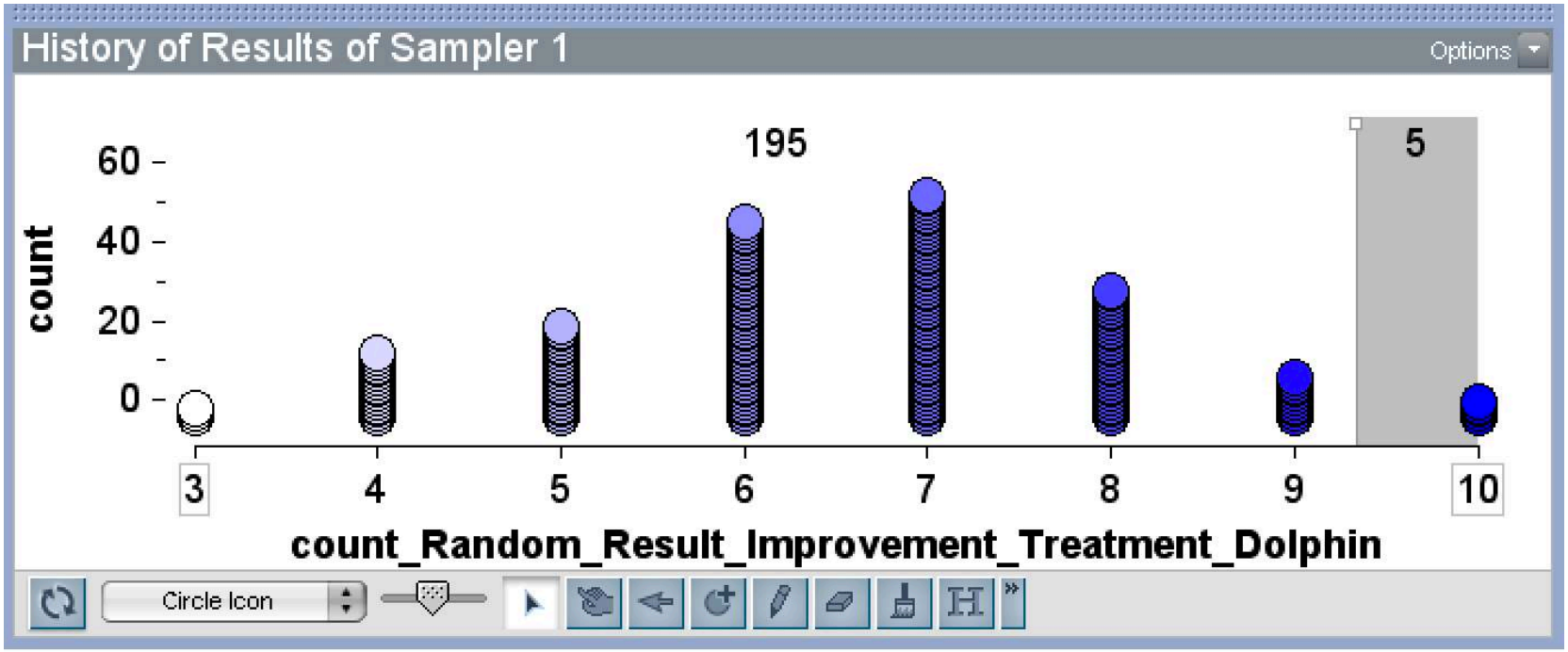
Collecting 199 more random samples



Produces the following results:

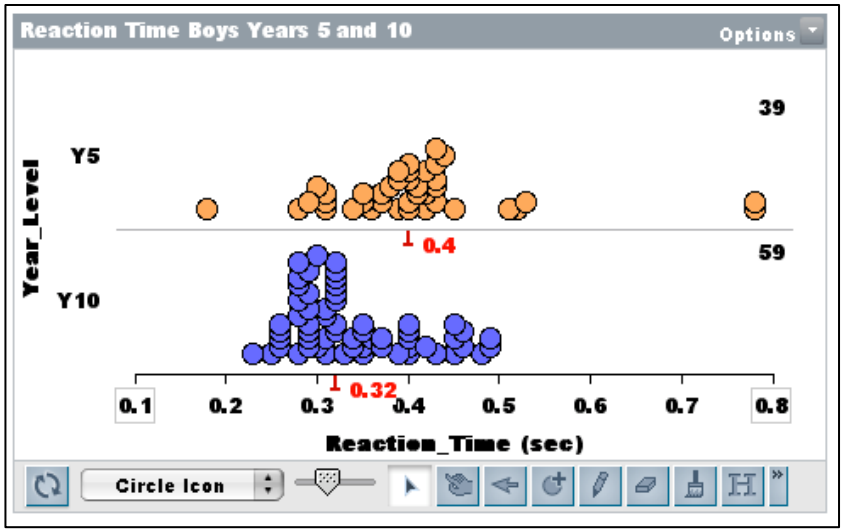


The Results show that only 5 times out of 200 Random reallocations would a result as strong as 10 people Improving after swimming with Dolphins occur.

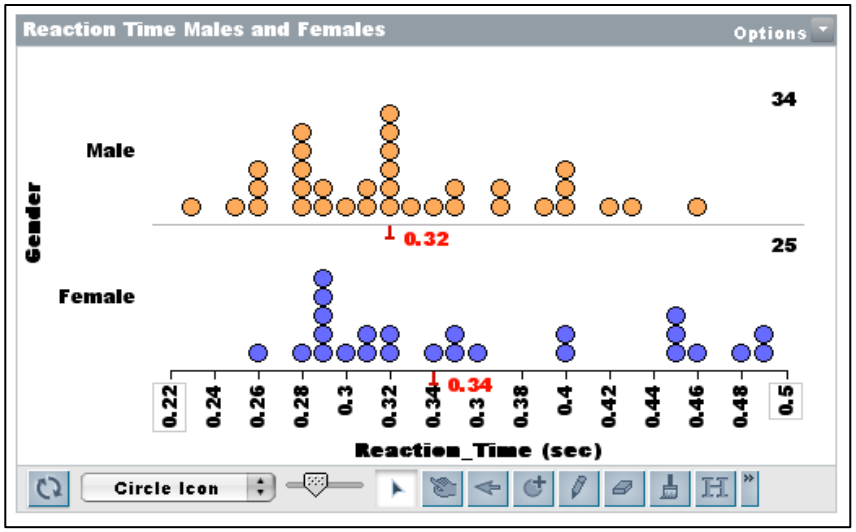


Lesson 8:
Assessment: Deciding if group differences in reaction times are “significant” or not.

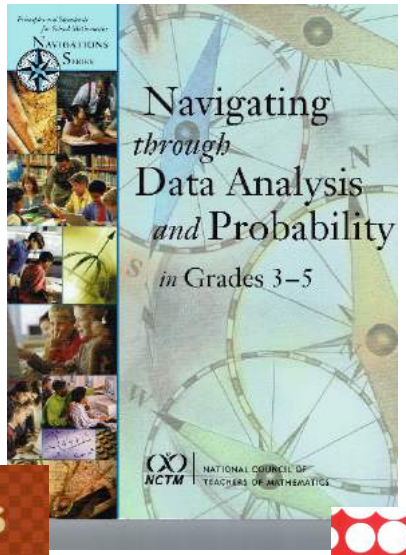
YEAR 10 BOYS & YEAR 5 BOYS

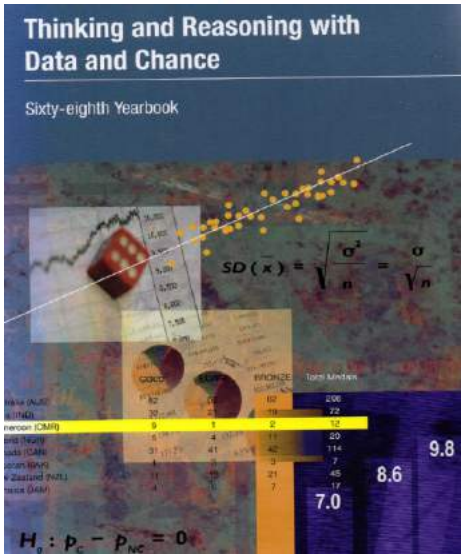


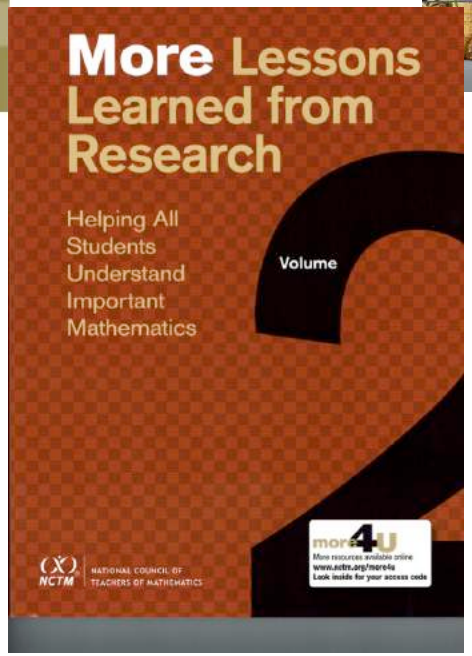
YEAR 10 BOYS & YEAR 10 GIRLS

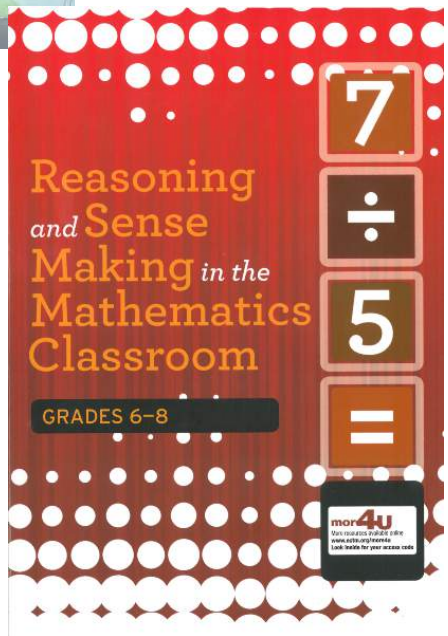


The NCTM is helping teachers









Thank you!!

- Come visit Tasmania