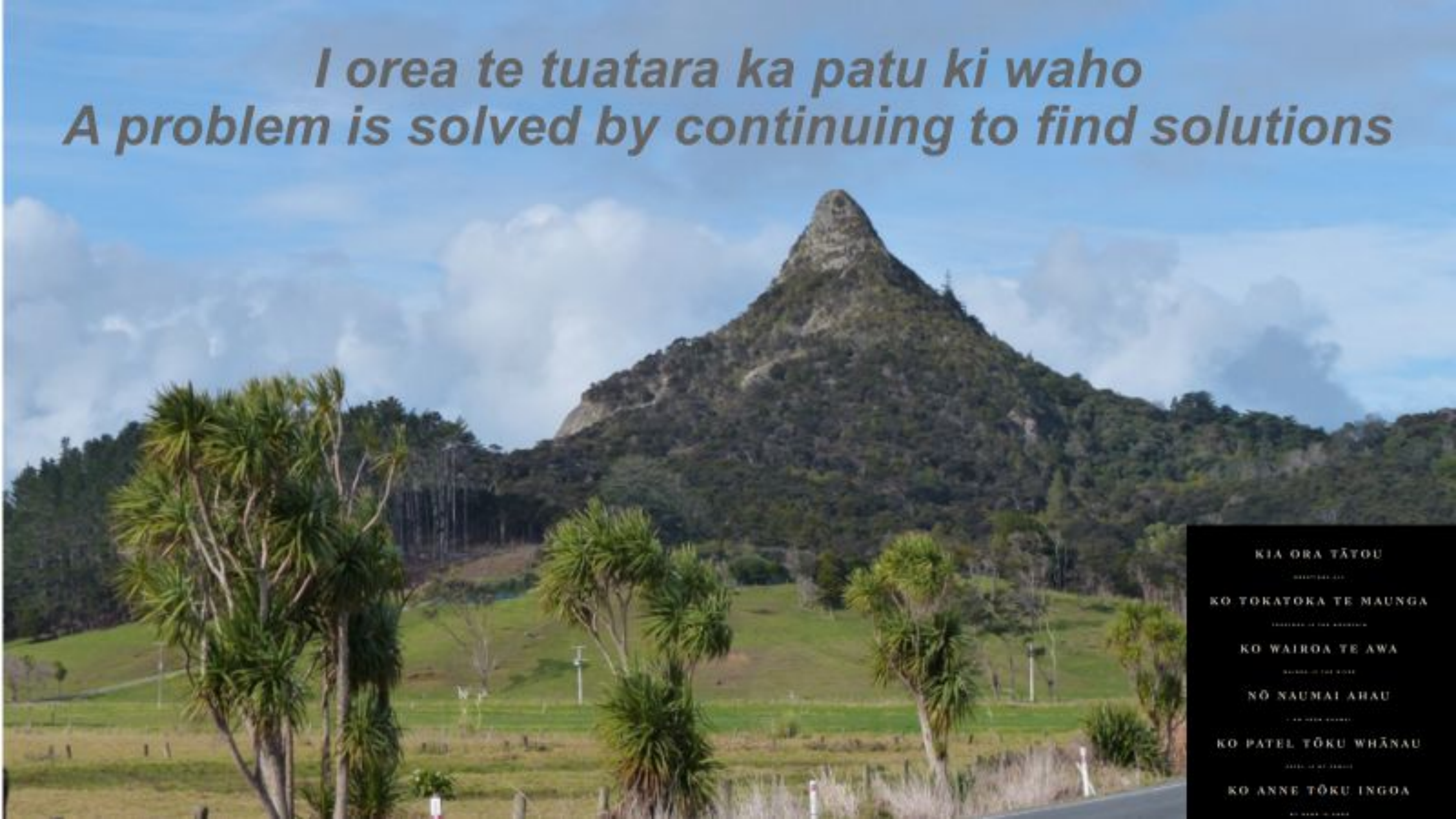
Exploring worlds through data

Curriculum level six Statistics


I orea te tuatara ka patu ki waho
A problem is solved by continuing to find solutions
KIA ORA TĀTOU
KO TOKATOKA TE MAUNGA
KO WAIROA TE AWA
NŌ NAUMAI AHAU
KO PATEL TŌKU WHĀNAU
KO ANNE TŌKU INGOA

There are four probability modelling activities in this session:

The first activity introduces teachers to the CODAP technology.

By building a model that explores random selecting three **equally likely** outcomes (with replacement), then altering the outcomes and number of outcomes in their model they can visualize and explore both **uniform** and **non-uniform** distributions.

Think about these **observable** teaching moments when **visualizing randomness** using a model:

- What does the chance distribution look like?
- How many possible outcomes are there?
- Explore the number of trials needed to be confident all possible outcomes (sample space) appear.
- Where, why/how much variation is there in chance distributions?
- After how many trials is it possible to visualise stabilised **global features** of the distribution?

What do you know about PASSWORDS?

What is the most commonly used password?

Top 25 most common passwords by year according to SplashData

Rank	2011[4]	2012[5]	2013[6]	2014[7]	2015[8]	2016[3]	2017[9]	2018[10]
1	password	password	123456	123456	123456	123456	123456	123456
2	123456	123456	password	password	password	password	password	password
3	12345678	12345678	12345678	12345	12345678	12345	12345678	123456789
4	qwerty	abc123	qwerty	12345678	qwerty	12345678	qwerty	12345678
5	abc123	qwerty	abc123	qwerty	12345	football	12345	12345

CHANCE SITUATION 1: PROTECTING YOUR PASSWORD

Let's Explore generating passwords at random.

For three letters, how likely is it to randomly generate a real word?

Choose 2 consonants and one vowel, which three letters do you want to check out, why those letters?

Who can generate the most real words from only three letters?

More Exploring...

What if... there were 2 vowels?

Supplementary question: Is it possible to make a three letter word with no vowels?

What if... there were two or four or five letters?

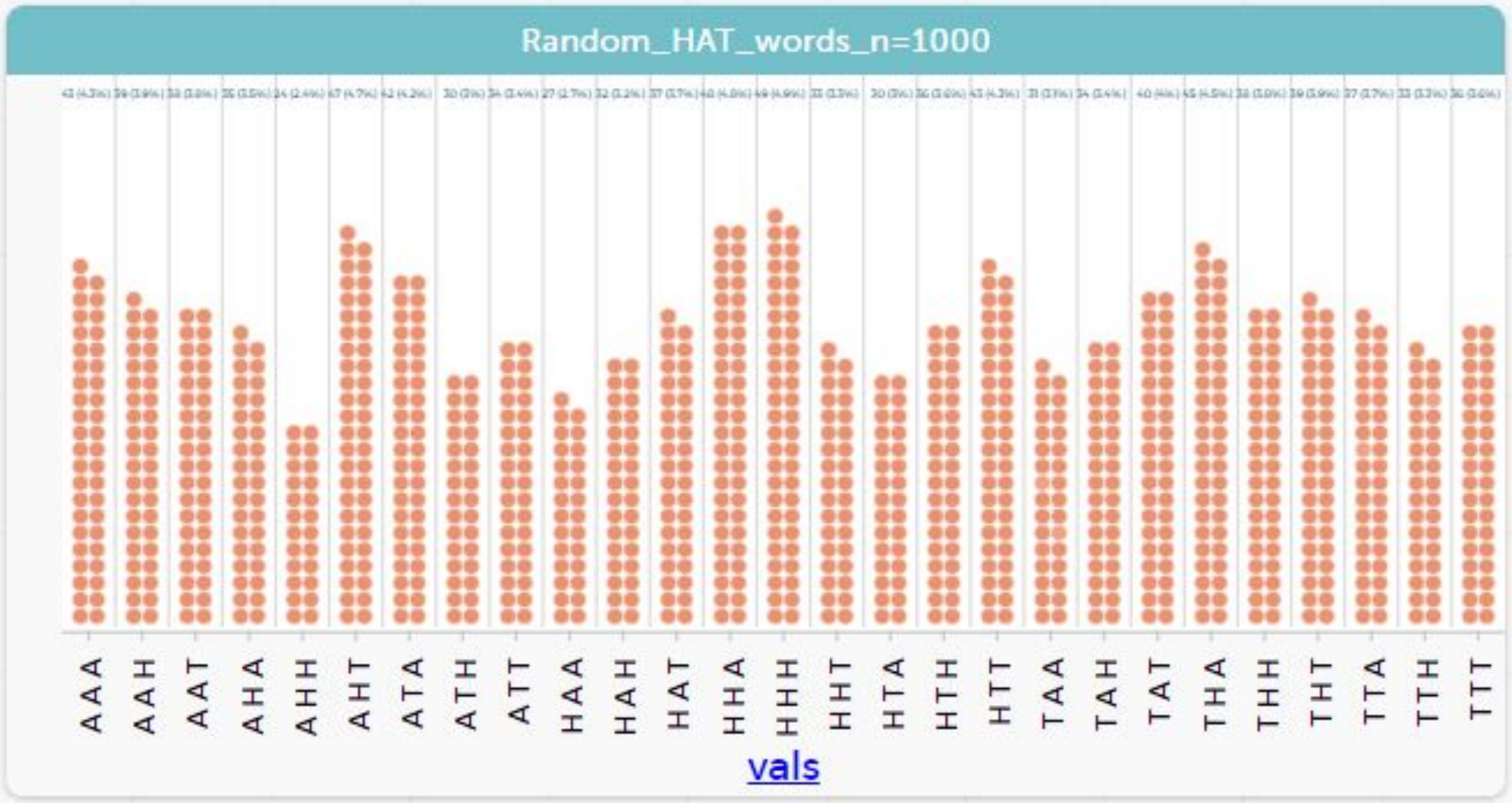
What if a password uses the same letters e.g. Mum, Dad, Gig or Pop?

What if… we explored a different language?

Explore how many combinations of 3 letters are possible in English/Maori/Spanish?

How many combinations of 9 letters are there?

What do you notice, What do you wonder?

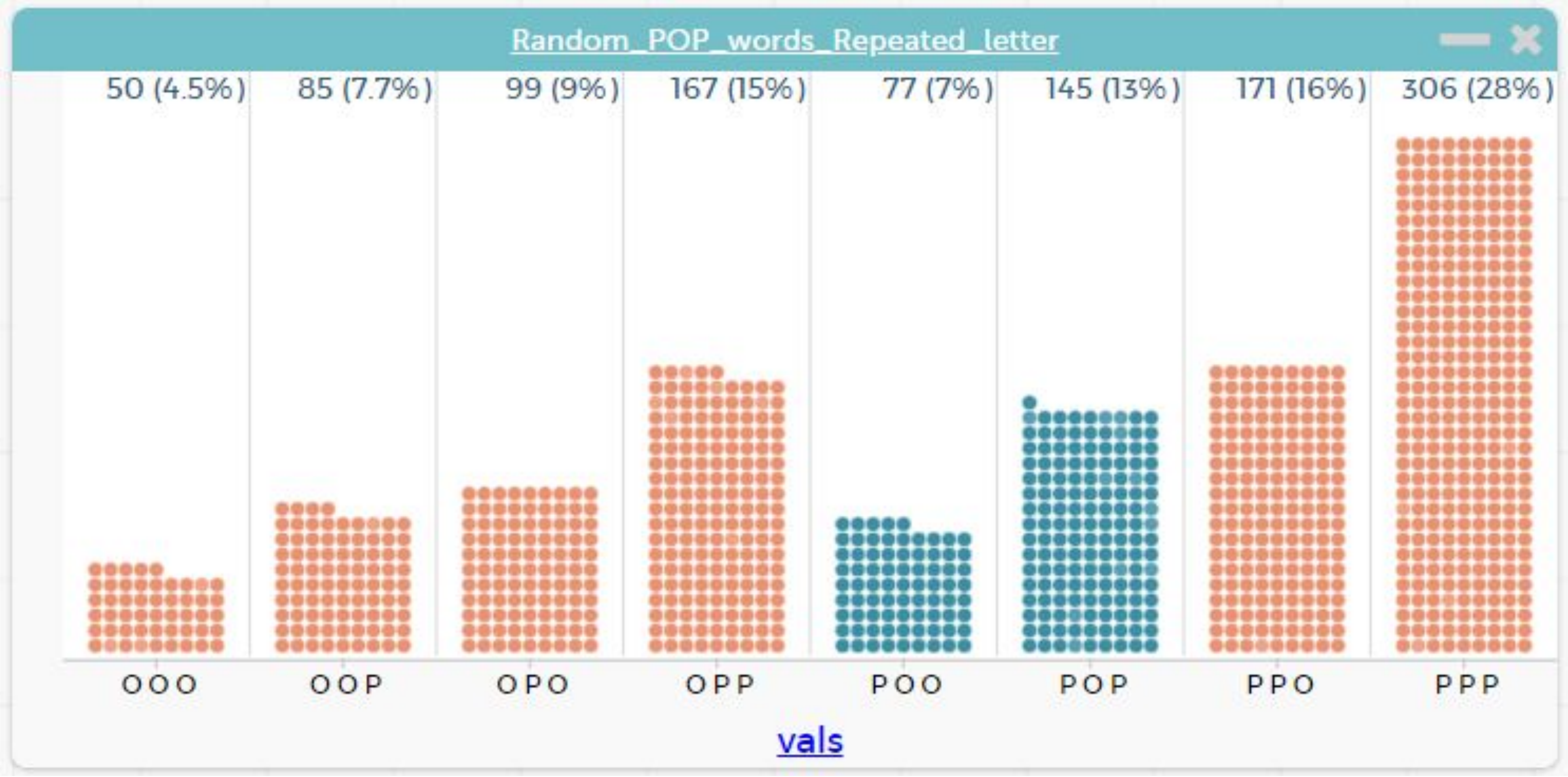


HAT (3.7%)
AAH (2.4%)
HAA (2.7%)
AHA (3.5%)
HAH (3.2%)
TAT (4%)

Total 19.5%

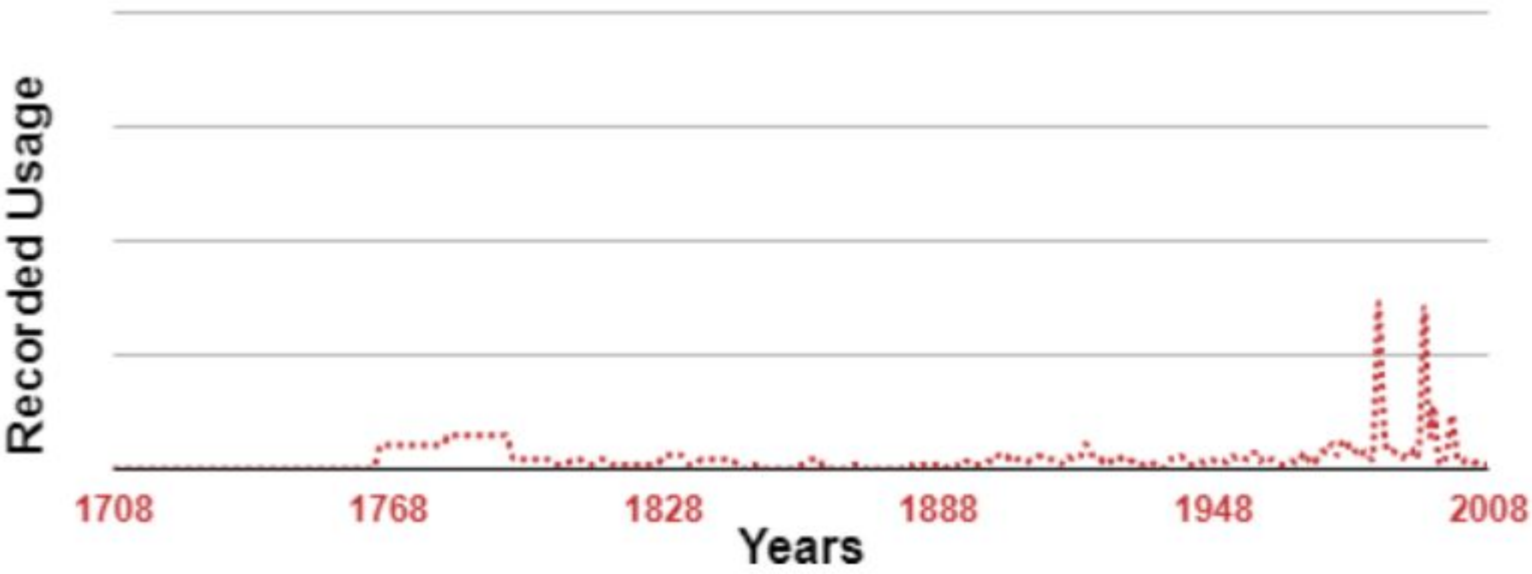
Theory 6/27 = 22.2%

What do you notice, What do you wonder?



HAA

View usage for: All Years



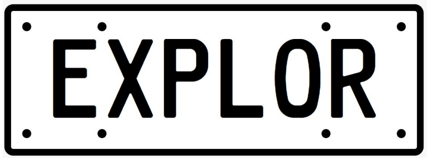
Thinking about themes: Technology/Cyber safety

This article in the Guardian advises that the strongest passwords are three random words.

With a little effort and some knowledge about words and probabilities passwords can be made “stronger”.

Write a letter to the Guardian and include your best three random word password and how you remember it. You should include graphs that help illustrate your reasoning.

Other contexts to explore that require random number generation: licence plates, google doc id, youtube links and cryptography.



The second modelling activity investigates...

...Using an **estimate** of a realistic chance situation (event) to explore the **likelihood** of possible outcomes for a **certain number of events**.

We can construct a chance model of a penalty shootout situation for a given number of shots at goal. Here both the probability (real-world estimate) and the number of goals can be varied to explore how the chance distributions change.

There is also assumptions about the model that need to be considered.

- Will the probability of scoring a goal stay the same?
- Is each event independent?

Chance Situation 2: Penalty Shootout

The House Netball cup is up for grabs!

It will take place during a long lunch. If the game ends in a draw a penalty shootout happens.

A single player from each team is selected and the team with the highest number of goals from 5 attempts will win. If an equal number of goals is scored another 5 attempts take place, by two different players.

Working in groups, taking turns to record data, stand at half the radius of the shooting circle from goal and collect your data - what proportion of your attempts at goal went in?

Calculate your **personal probability (chance)** of getting a goal.

Using this estimate for getting a goal. Explore how many of the five goals you are likely to get in if you are chosen for the penalty shootout.

PROBLEM:

What if... I get selected for the penalty shootout, how many goals am I likely to get?

PLAN:

When you shot for goal one of two things happened, what were they?

Define a trial: One penalty shootout - 5 shots at goal

For 5 shots at goal, five randomisations can be generated from the model.

What are the possible outcomes for each trial?

What outcomes are required in the model? Think about the real situation.

Using my initial estimate for the probability of getting a single goal, I can alter this probability to explore different chance situations, like if I got better or the number of goals was changed.

Is it reasonable to assume the probability will stay the same over time? (Model assumption)

LET'S EXPLORE

While you are exploring this chance situation have a think about the questions you ask can about the chance distributions you generate...

CODAP TECH POINT: When graphing in Codap click on the variable label in the graph and select “**Treat as Categorical**” to use the percentage and count functions

Analysis:

How many goals am I likely to get out of 5?

What is the chance I miss every shot?

What is the likelihood of getting 1 or more shots in?

What is the likelihood of getting 4 or more shots in?

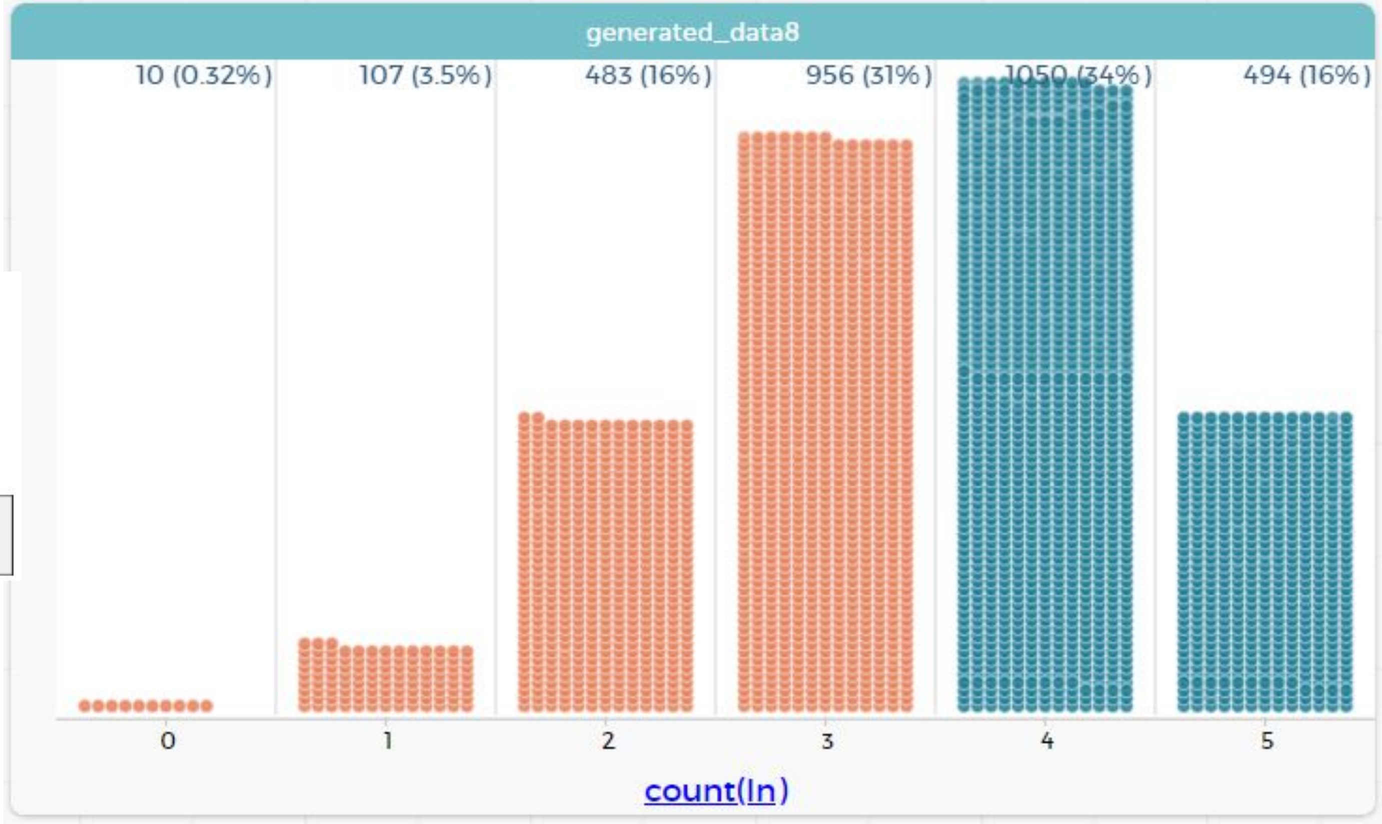
What if...

What would my probability of success have to be to get 4 or more goals at least half the time?

Outcome	In	Out
Weighting	68.1	31.9

Randomise 5 times

In	Out



e-PPDAC!

eXPLORE Identify a situation that involves elements of chance. Can the situation be translated into the model world to answer any questions that arise during the exploration?

PROBLEM Pose an investigative question/purpose. Make a conjecture about what you expect to happen

PLAN Structure/Set up the chance situation.

DATA Generate chance distributions by running and altering the model.

ANALYSIS Describe the chance distributions. Calculate frequencies for events, compare relative frequencies and expected values

CONCLUSION Answer the investigative question/purpose. Is there something you wonder about the chance situation? What if I changed…? What if I looked at…?

SUMMARY: SESSION 1

Dynamic v static approach to chance events

Adding another **representation** (chance distributions) to help students **reason**

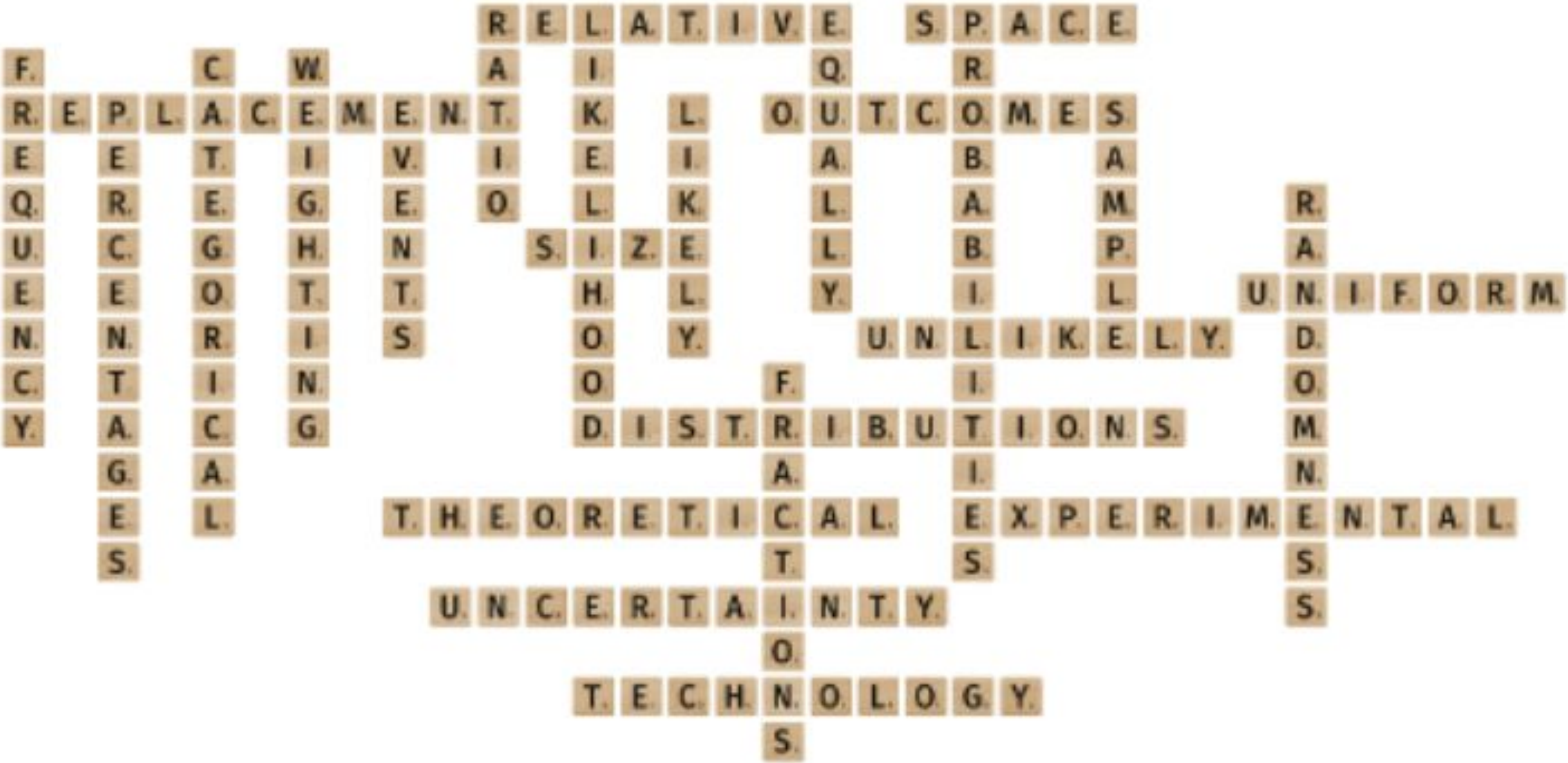
Visualise - changing **relative frequencies** of outcomes that reflect the model and stability of chance distributions in large numbers of trials

Akonga **agency,** transferring reality into a chance model. Appreciation of modelling to explore chance outcomes.

Building and altering chance based models to explore outcomes and relative likelihoods

Creating stronger heuristics around the concept of **randomness** - cannot predict what will happen in the next trial, but able to view stability in chance distribution over many trials.

All possible outcomes (Sample space) visible if you do enough trials!



[WORD SCORE: 371]

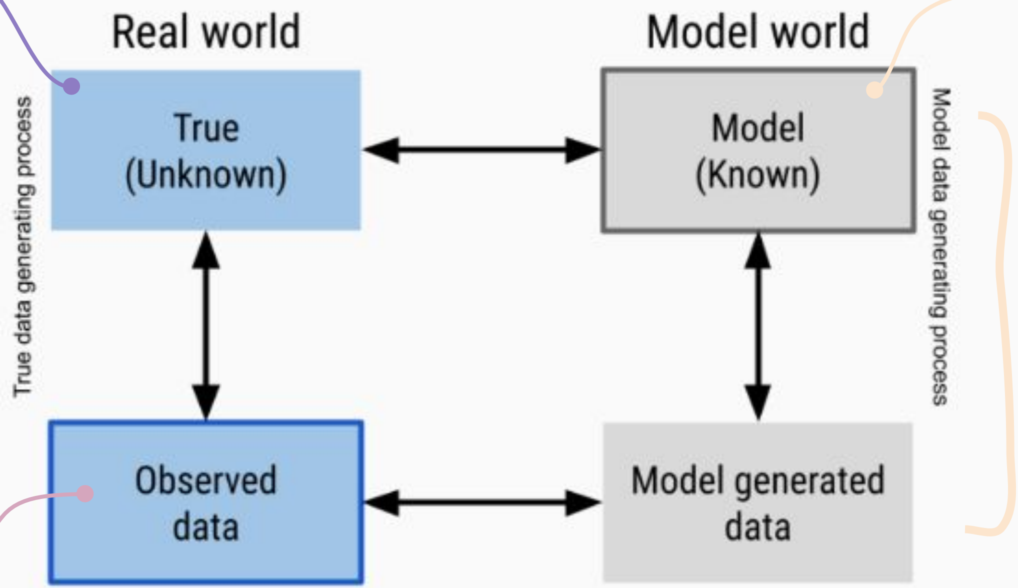
WELCOME BACK

SUMMARY OF FIRST SESSION (5 MINS)

What was one thing you thought about/realised/created in the first session?

Sample to population inference - we don't know the value of the population parameter (an aggregate)

Probability modelling - we don't (yet) know the properties of the models being used



Fergusson & Pfannkuch (2020)

Prediction - we don't know the value of a particular variable/attribute for an individual/case

The third modelling activity investigates...

… Possible combinations (sample space) of **joint events** that do **not** have equally likely outcomes.

Teachers will build models that explore multiple events with outcomes that are **not** equally likely.

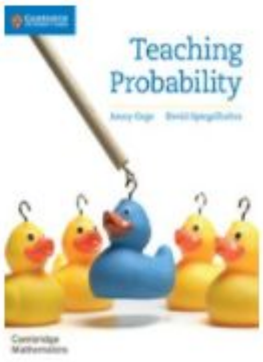
This results in chance distributions with interesting **features**, that can be examined and generalisations made about the likelihoods of joint events.

- Do all possible outcomes occur in a limited number of trials?

Chance situation 3: Which team will win?

Every weekend, Team Yeti & Team Baboon play **2-goal** football – they continue playing until two goals have been scored. Sometimes games take a long time, sometimes they are over in a flash, but either way they finish when two goals have been scored.

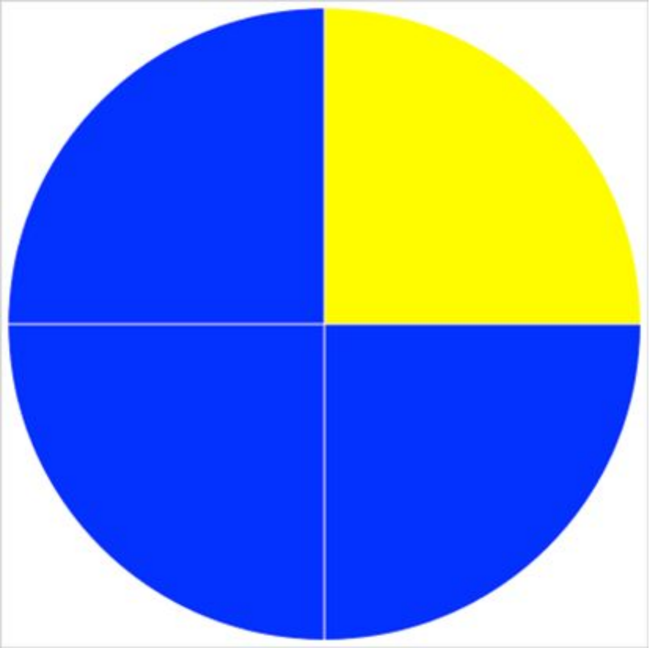
Team Baboon have a new player! This gives them a big advantage over their rivals. But how does their improved goal rate translate into games won? Who will win the season?



Gage & Spiegelhalter page 48, worksheets here http://www.teachingprobability.org/

PHYSICAL MODELS INTO TECH

There are many factors involved in sport, but chance also plays a part…



If the spinner lands on blue this means the Baboons score a goal.

If it lands on yellow this means the Yetis score a goal

Spin the spinner **twice** and record your results for who scored the first and second goal in each match.

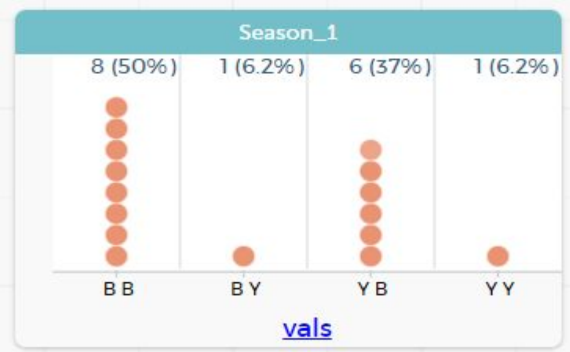
There are only 16 matches each season.

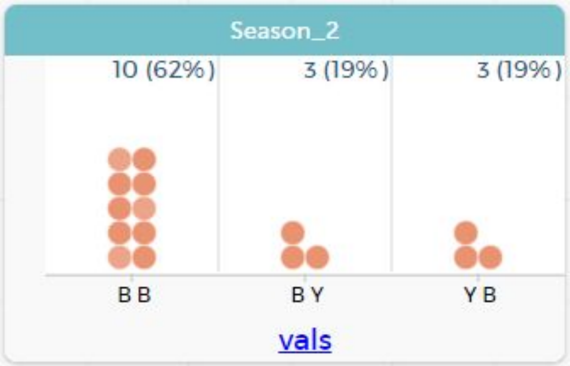
ANALYSIS

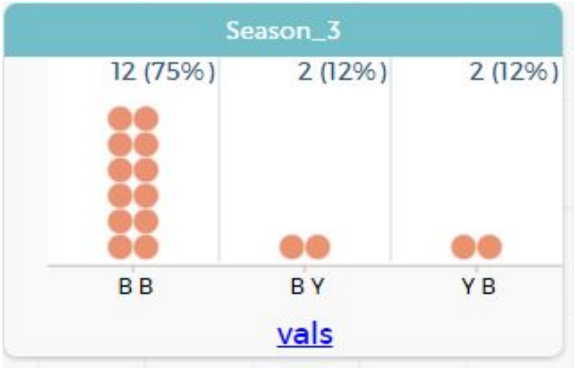
Game	1st goal	2nd goal	Result
1	Y	B	YB draw
2	B	Y	BY draw
3	Y	Y	YY win
4	B	B	BB win
5...			

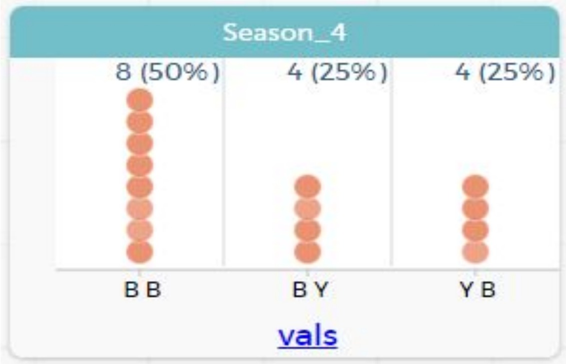
YY wins	///	3
BB wins	~~////~~ /	6
YB draws	~~////~~	5
BY draws	//	2
Total		16

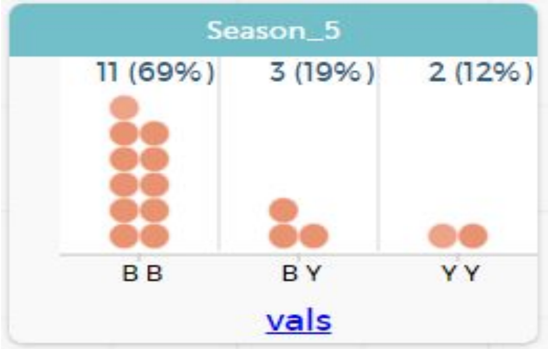
I NEED TO SEE SOME CHANCE DISTRIBUTIONS!

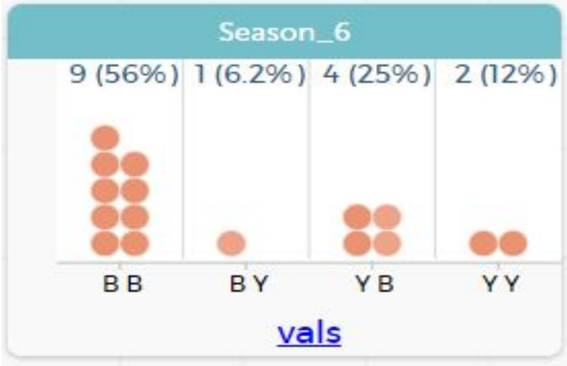












How many times did the Yetis win (YY)?

How many times did the Baboons win (BB)?

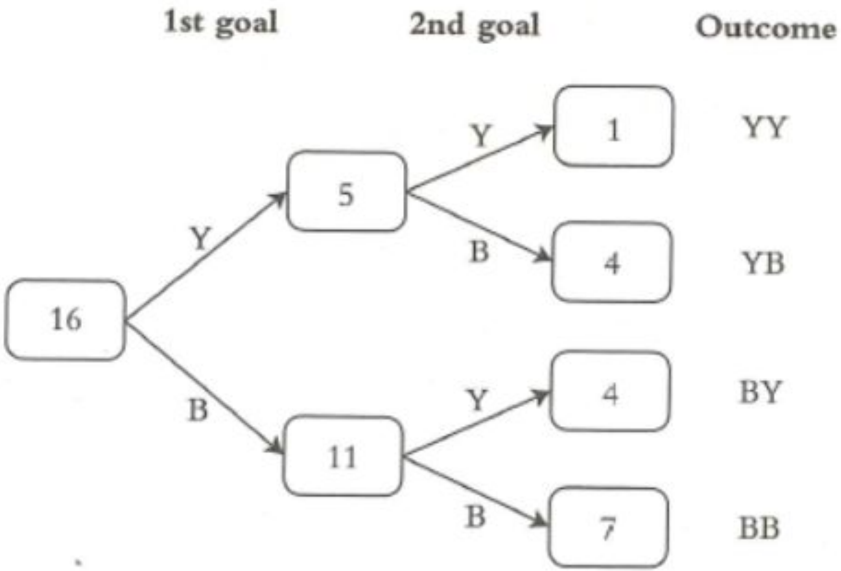
How many times did the Yetis score first, but the Baboons then equalised (YB)?

How many times did the Baboons score first, but the Yetis then equalised (BY)?

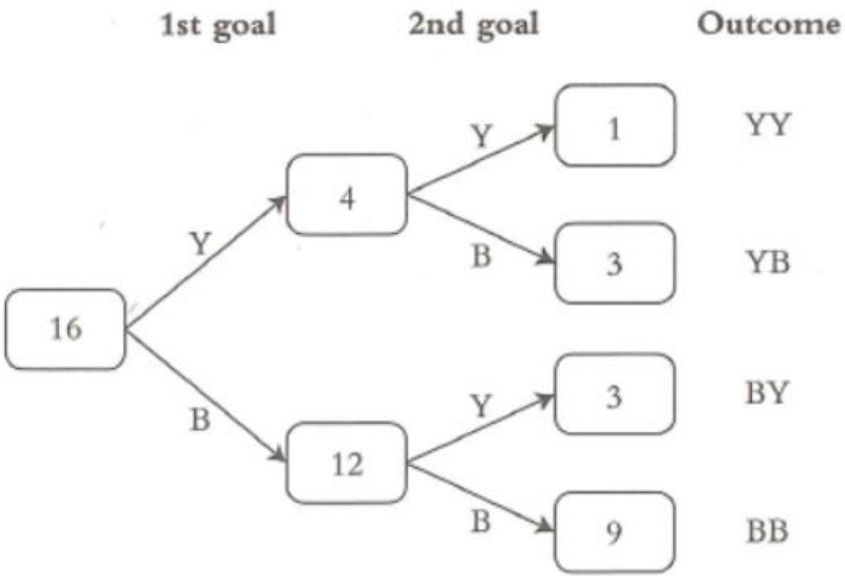
Do any of your results surprise you at all? Why (not)?

The Baboons have more chances to score than the Yetis. What does this tell you about the teams?

ANALYSIS: COMPARE A SET OF ACTUAL RESULTS WITH EXPECTED FREQUENCY TREE



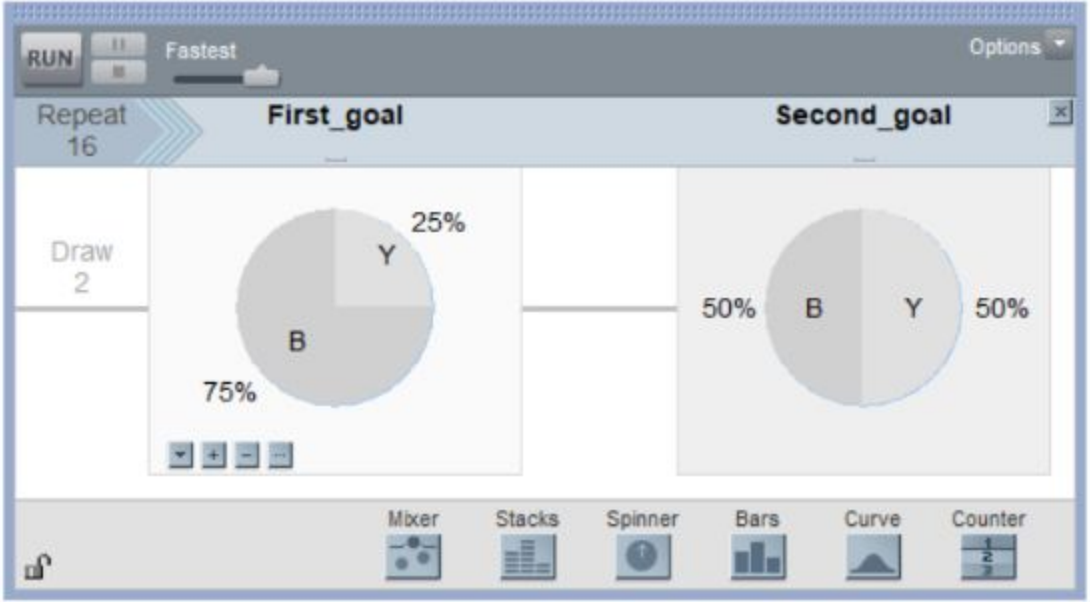
Actual results for Season 1 displayed on expected frequency tree



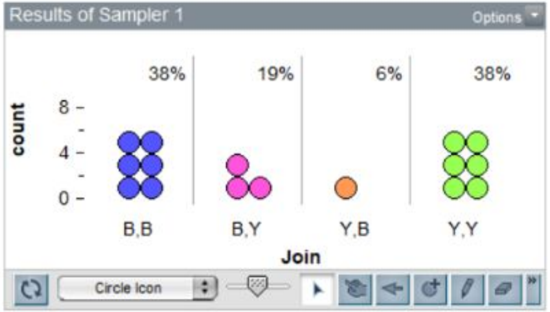
Expected frequency tree for Season 1

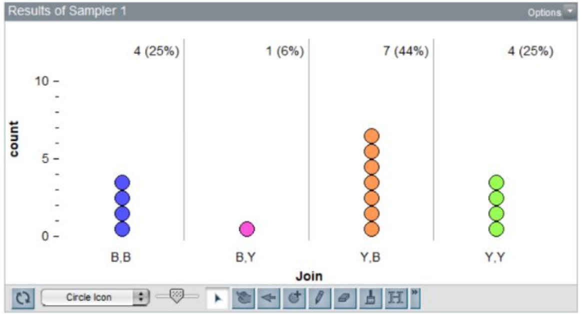
I NOTICE: THREE IMPORTANT THINGS

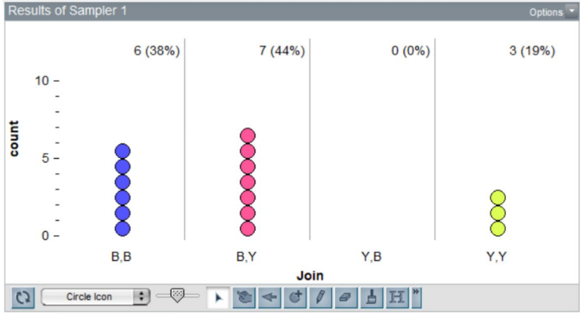
WHAT IF… SEASON 2 AND THE BABOONS STAR PLAYER HAS GONE TO LEAGUE?

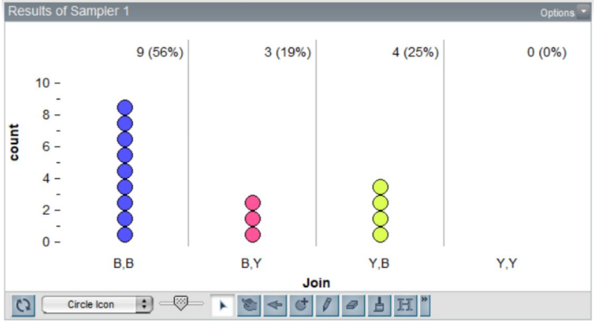


POSSIBLE SEASON TWO EVENTS









WHAT'S UNDERNEATH?

The forth modelling activity involves...

… a mini ePPDAC cycle, putting what we have learned today altogether.

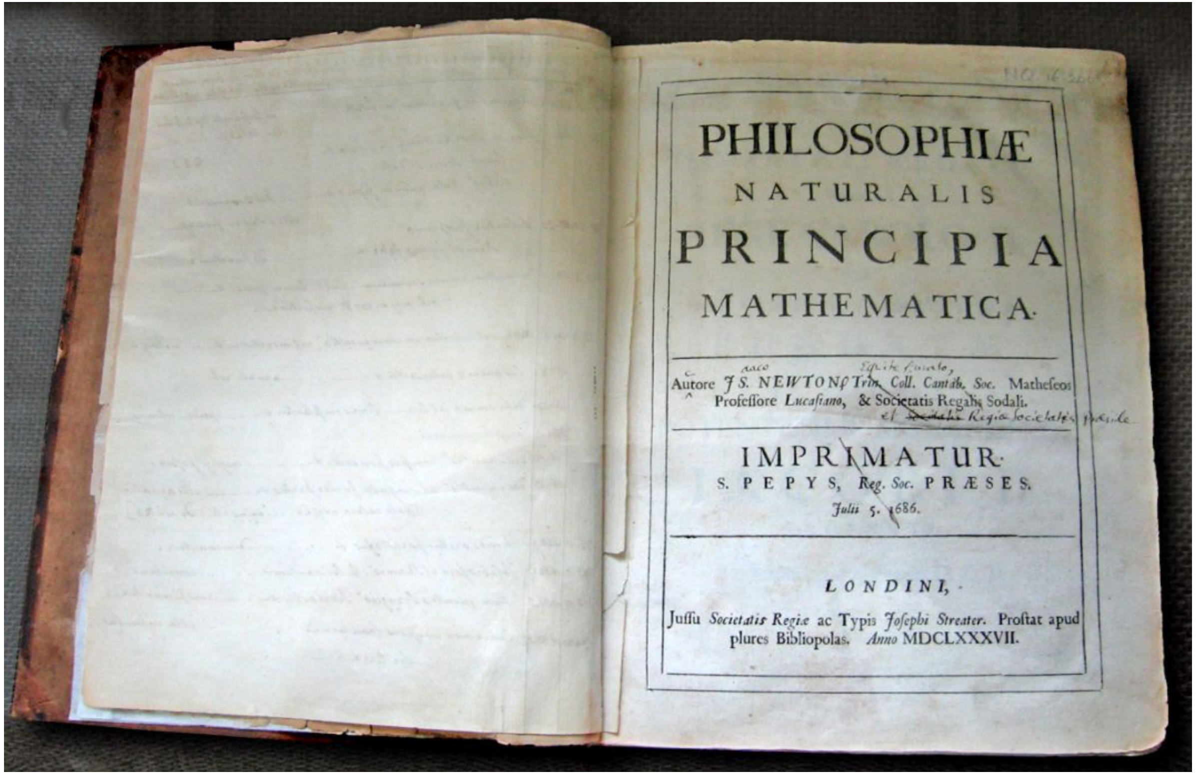
Translating a **real world chance situation** into a **model** to explore **relative likelihoods** in **chance distributions**.

The chance situation (one that was posed to Sir Isaac Newton) requires teachers to build three or more models to investigate what happens when throwing different numbers of fair dice.

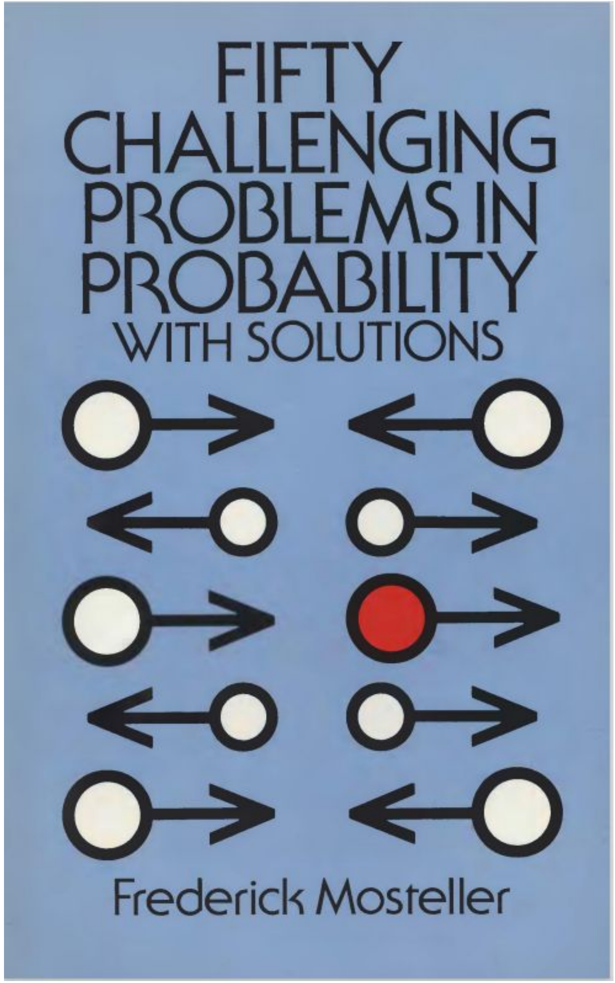
This activity involves proportional reasoning and changing relative frequencies. Can you summarise your understanding of the chance situation using appropriate probabilistic language?

Sir Isaac did not have access to this tech, perhaps if he had, he would not have made his logical mistake…

CHANCE SITUATION 4: ASK NEWTON...



Newton, (1687)



Mosteller, (1965)

MINI ePPDAC INVESTIGATION

Which of the following three propositions has the greatest chance of success?

A. Six fair dice are tossed independently and at least one "6" appears.

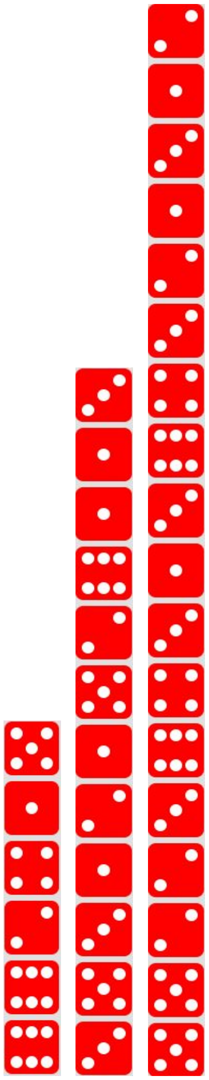
B. Twelve fair dice are tossed independently and at least two "6"s appear.

C. Eighteen fair dice are tossed independently and at least three "6"s appear.

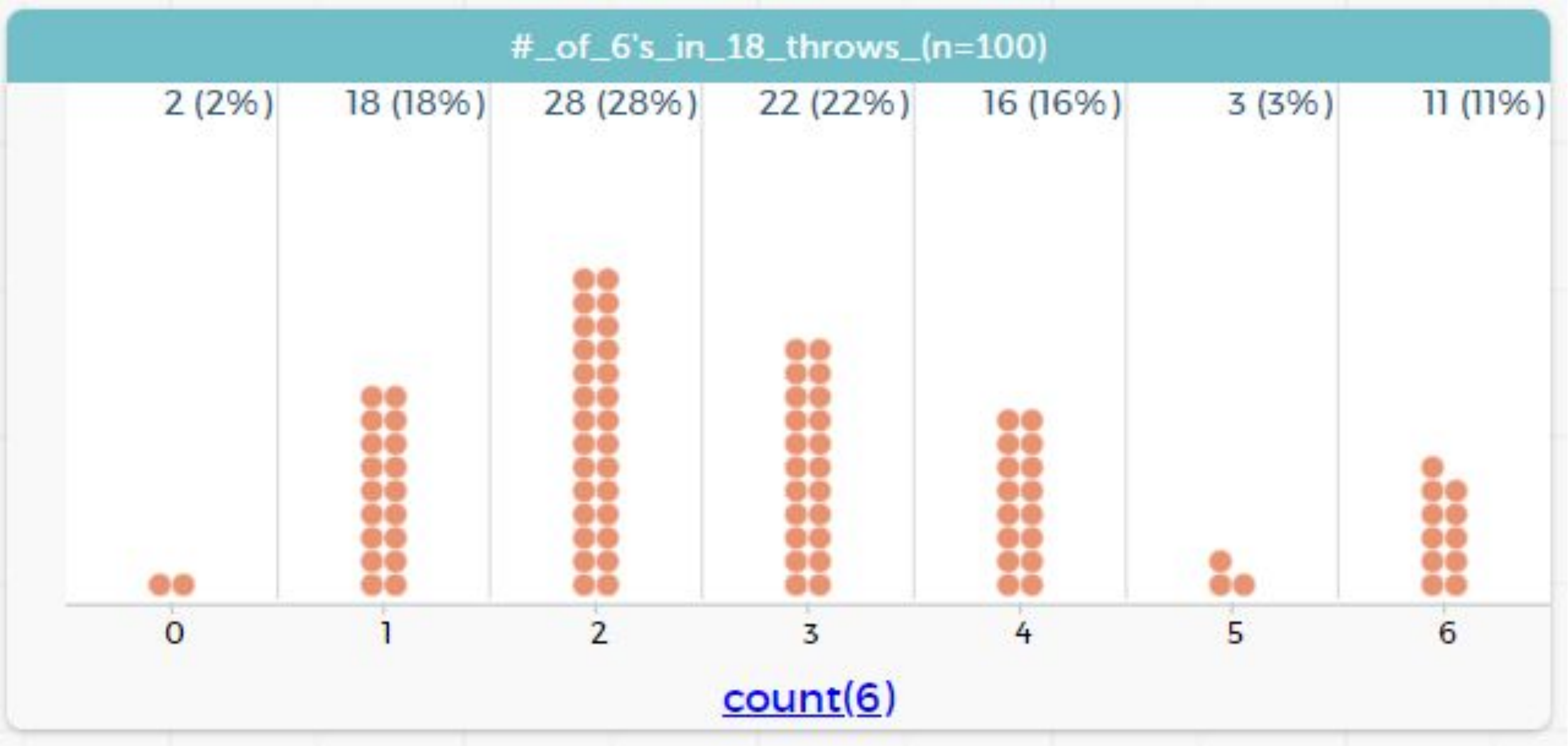
What is your initial prediction?

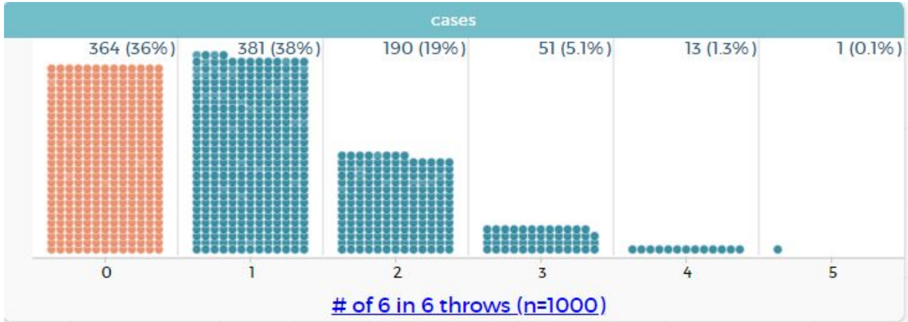
- All three situations have the same likelihood
- A has the greatest chance of success
- B has the greatest chance of success
- C has the greatest chance of success

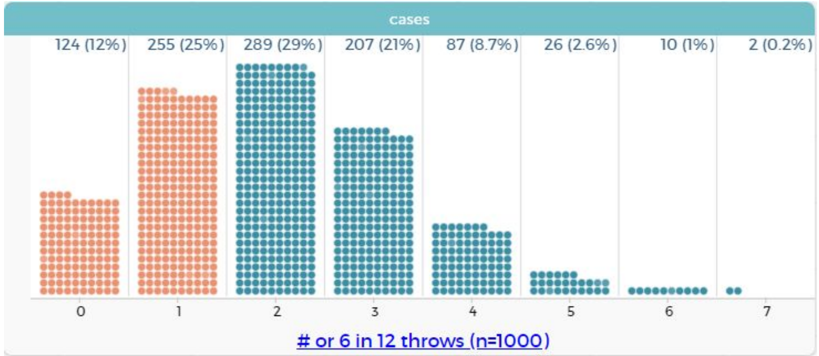
LET'S EXPLORE

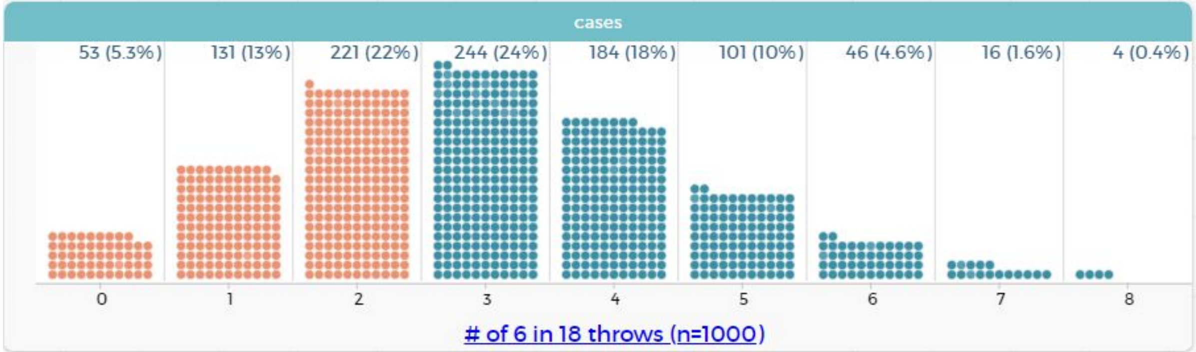


EXPLORING DISTRIBUTIONS

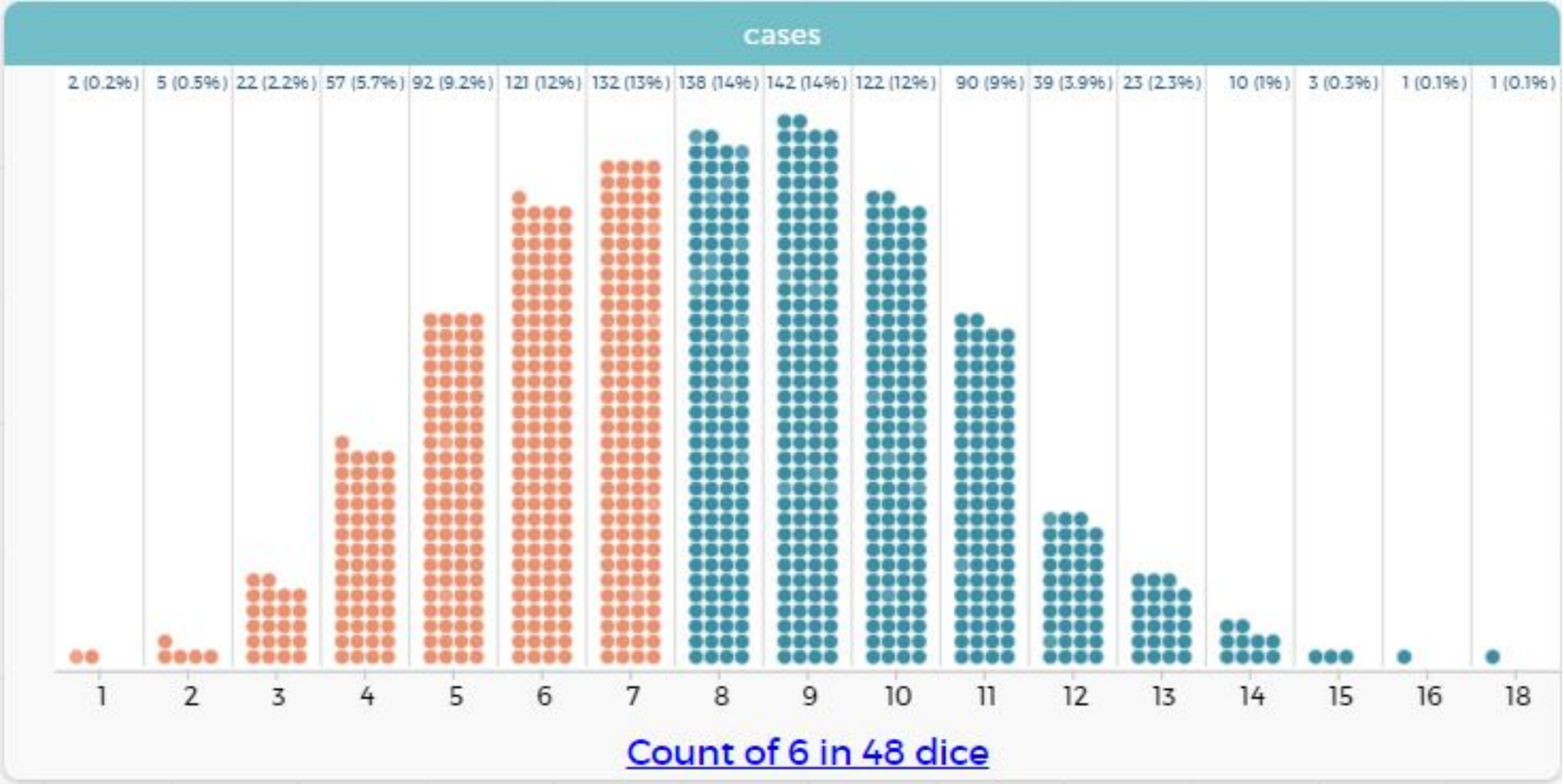



cases
364 (36%)
381 (38%)
190 (19%)
51 (5.1%)
13 (1.3%)
1 (0.1%)
0
1
2
3
4
5
# of 6 in 6 throws (n=1000)


cases
124 (12%)
255 (25%)
289 (29%)
207 (21%)
87 (8.7%)
26 (2.6%)
10 (1%)
2 (0.2%)
0
1
2
3
4
5
6
7
or 6 in 12 throws (n=1000)


cases
53 (5.3%)
131 (13%)
221 (22%)
244 (24%)
184 (18%)
101 (10%)
46 (4.6%)
16 (1.6%)
4 (0.4%)
0
1
2
3
4
5
6
7
8
# of 6 in 18 throws (n=1000)

Just because you can... What do you notice, What do you wonder?



TECH HEADS UP: GENERATING MORE THAN 20 DICE

Generate random data in EXCEL using formulas

Need two formulas: Formulas begin with = sign e.g.

Generate a random number between 1 - 6 `=RANDBETWEEN(1,6)`

Count total number of 6's `=COUNTIF(A12:F12,6)`

Fill across and **fill down:** Black cross in bottom right hand corner of the cell

Importing files (google sheets or excel) files into CODAP they must be **csv** files

CODAP only imports the **first sheet** in a csv file.

So you need to create at different csv files for each situation

Unfortunately, Newton had to work the probabilities out by hand, but we can use the *Tables of the cumulative binomial distribution*, Harvard University Press, 1955. Fortunately, this table gives the cumulative binomial for various values of p (the probability of success on a single trial), and one of the tabled values is $p = \frac{1}{6}$. Our short table shows the probabilities, rounded to three decimals, of obtaining the mean number or more sixes when $6n$ dice are tossed.

$6n$	n	$P(n$ or more sixes)
6	1	0.665
12	2	0.619
18	3	0.597
24	4	0.584
30	5	0.576
96	16	0.542
600	100	0.517
900	150	0.514

Clearly Pepys will do better with the 6-dice wager than with 12 or 18.

Mosteller, (1965, p.35)

EVEN ISAAC NEWTON MAKES MISTAKES IN PROBABILITY…

Pepys was elected a Fellow of the Royal Society in 1665 and served as its President from 1 December 1684 to 30 November 1686. Isaac Newton's *Principia Mathematica* was published during this period, and its title page bears Pepys's name. There is a probability problem called the "Newton–Pepys problem" that arose out of correspondence between Newton and Pepys about whether one is more likely to roll at least one six with six dice or at least two sixes with twelve dice.[59] It has only recently been noted that the gambling advice which Newton gave Pepys was correct, while the logical argument with which Newton accompanied it was unsound.[60]

https://en.wikipedia.org/wiki/Samuel_Pepys#Text_of_the_diary

Augustus De Morgan once wrote, "Everyone makes errors in probabilities, at times, and big ones." (Graves, 1889, page 459)

Isaac Newton as a Probabilist Stigler (2006)

REVIEW:

...it is important to understand that, in randomness apparent disorder, a multitude of global regularities can be discovered. These regularities allow us to study random phenomena using the theory of probability.

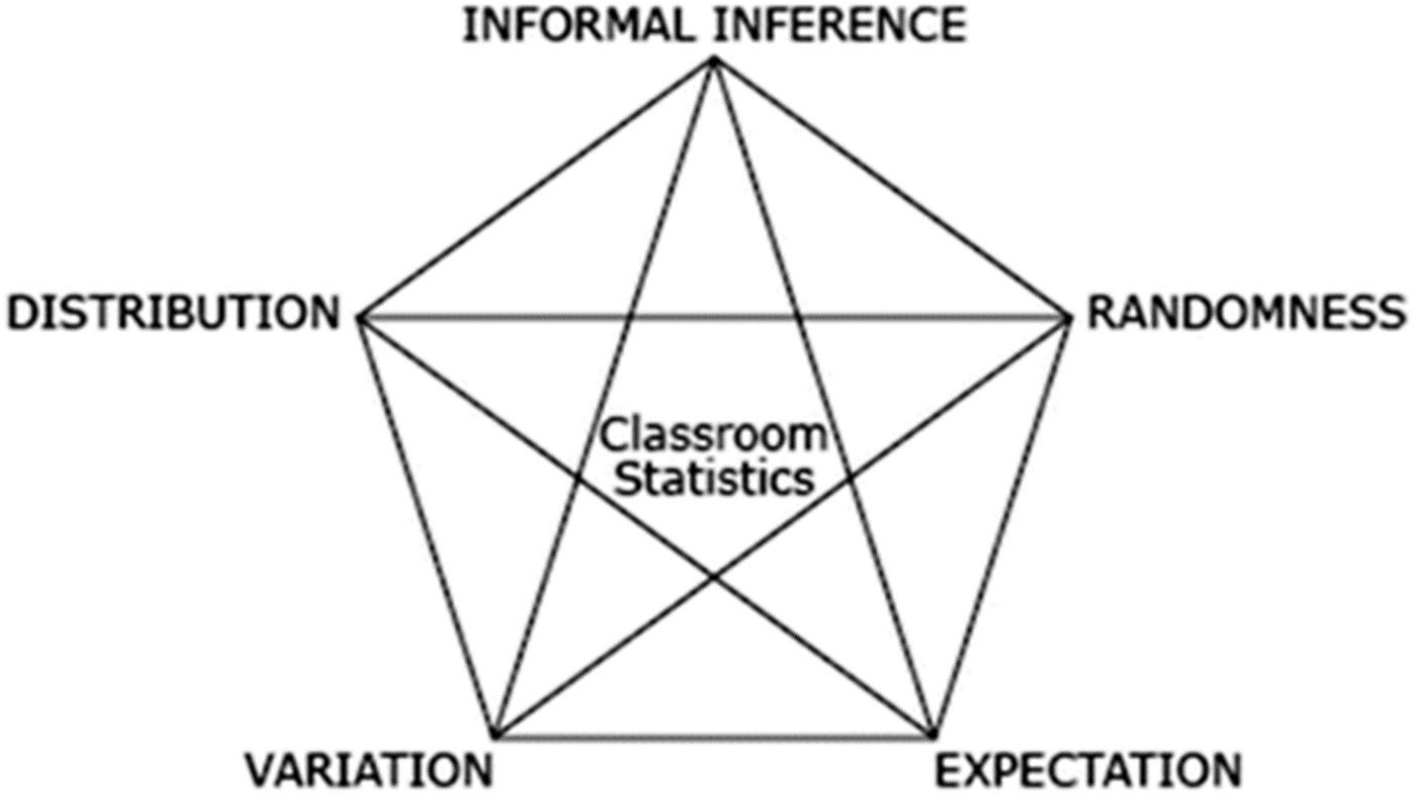
Batanero, Serrano, & Green, (1998)

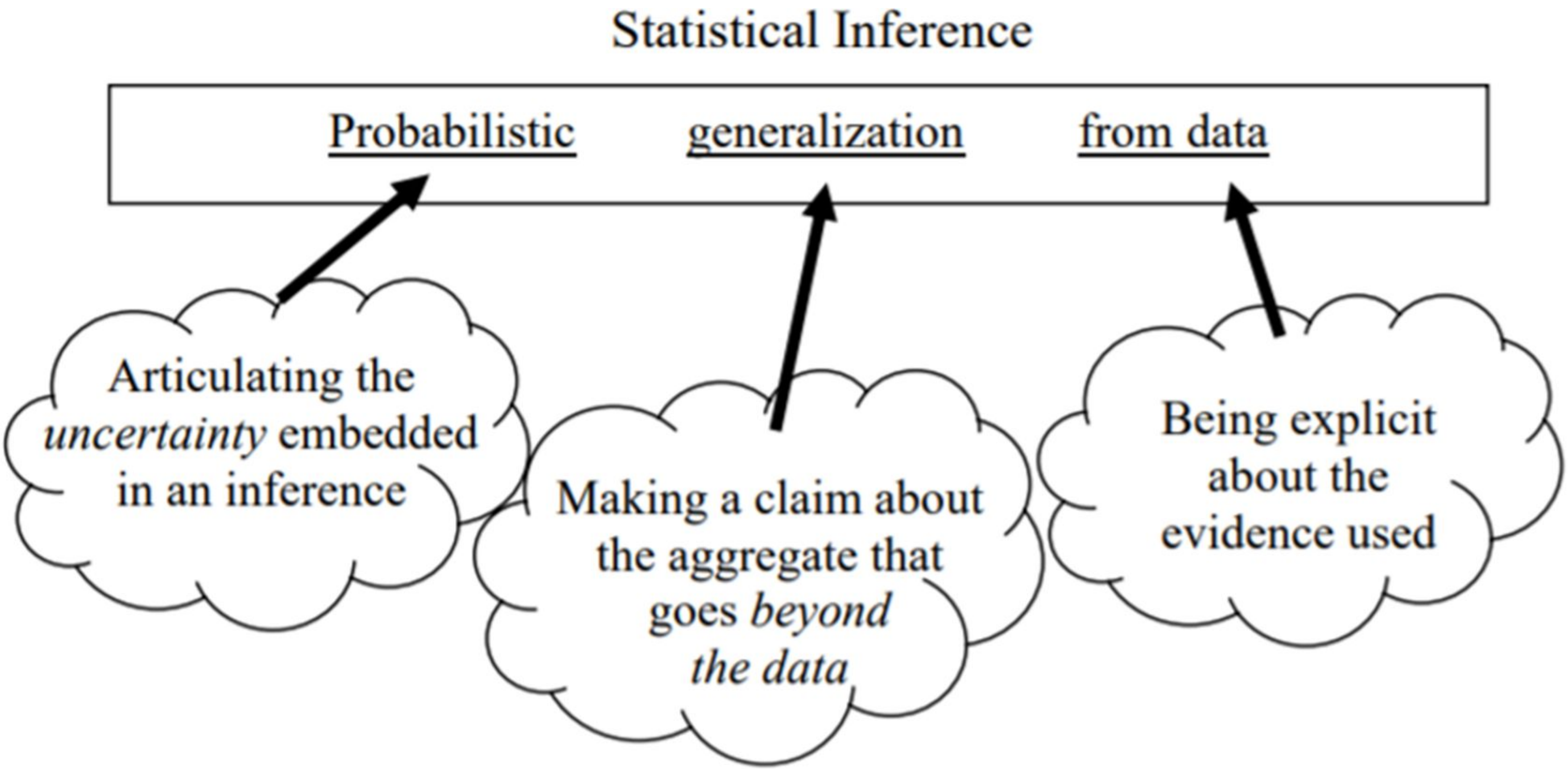
Purposes of a dynamic probability modelling approach:
Can't teach through misconceptions rather use the time available to build better heuristics/understanding about chance/random events and connect the big ideas in statistics…

NB: CL6 does not look at null/empty model or hypothesis testing ideas i.e. confirming how likely/unlikely an event is in the real world.

BIG IDEAS IN STATISTICS

(Watson, Fitzallen, & Carter, 2013)





Makar, K., & Rubin, A. (2009).

BIG THANKS & AROHANUI TO THE PROB MOD TEAM:

Stephanie Budgett, Amy Renelle, Lorraine O'Carroll
Liam Smyth, Aaron Webb, Helen Teal, Katalina Ma & Emma Lehrke

+ Anna Fergusson,
+ Pip Arnold,
+ Maxine Pfannkuch,
+ Chris Wild

[WORD SCORE: 52]

EXTRA TIME: ZHUZHING UP YOUR PROBABILITY LESSONS

Your homework was to record the outcomes of 20 tosses of a dice.

One of your classmates cheated and made up their results

Can you tell who it was?

Student A: T H T H T H T H T H T H T H T H T H T H

Student B: H H T H H H H H T T T T H T H H T T T H

Student C: H H T H T T H T H T H H T T H T H T H T

Explain your reasoning to the person next to you.

INVESTIGATE: MISSING HOMEWORK

Record the outcomes of throwing twenty coins (without actually throwing them) then do it for real...
Do you the outcomes look the same? Why or why not?

Look at the people's outcomes beside you, do they look different? What is the same, what is different?
What can you investigate? Number of Heads, Number of runs, Length of runs.

- Would you be surprised to see less than 6 heads? How likely is this to happen?
- What is the chance of getting exactly 10 heads?

I notice...

I wonder...