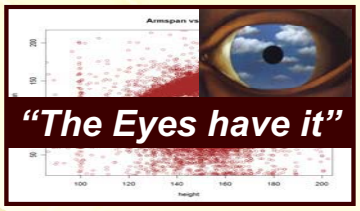
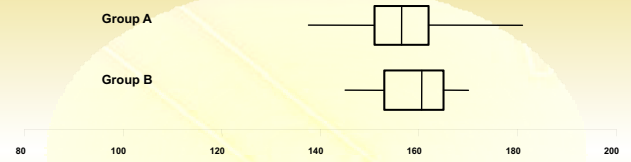
Early Statistical Inferences :



Chris Wild
University of Auckland

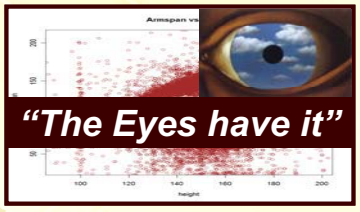
Talk Pre-test: Are these different?



- Are they different? – quick eyeball judgement
- Who automatically interpreted that question inferentially?
- What is the single correct answer to every question in statistics ??
  - **"It depends"**
  - If we interpret the question descriptively, then …
  - If we interpret the question inferentially, then …
    - the display doesn't contain enough information to tell
    - I have read it differently depending on sample size – and "How confusing is that?"

Jock MacKay

Early Statistical Inferences :



Chris Wild
University of Auckland

THE UNIVERSITY OF AUCKLAND DEPARTMENT OF STATISTICS
USCOTS Letting Go To Grow '09

The ideas in this talk have developed ….

through a long series of brainstorming sessions about informal inference with:

Winner, 2009 Waller Award

Maxine Pfannkuch
U of Auckland, NZ

Matt Regan
U of Auckland, NZ

Nick Horton
Smith College, MA, USA

Background to the work we will be talking about

What we are doing for Statistics in K-12 in New Zealand

- **Recently a door opened momentarily**
  - **Just a crack …**

Planets were all lined up

Timing
- New curriculum

Relationships right with
- Educational bureaucracy
- Government Agencies
- Lead Teachers
- Teacher Trainers
- Mathematicians

One chance:
Take it and run with it
or wait another 18 years

What we are doing for Statistics in K-12 in New Zealand

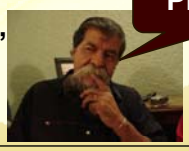
- Recently a door opened momentarily
  - Just a crack
- But enough to slip in a modern new curriculum

History Lesson: NZ Curriculum

What's in a name?

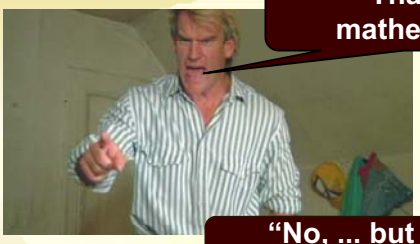
"Mathematics & Statistics" !!

"Mathematics"



Plenty !!!

18?? 1950 2009 20??

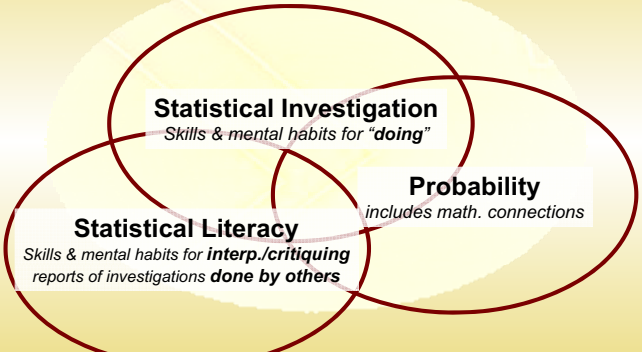


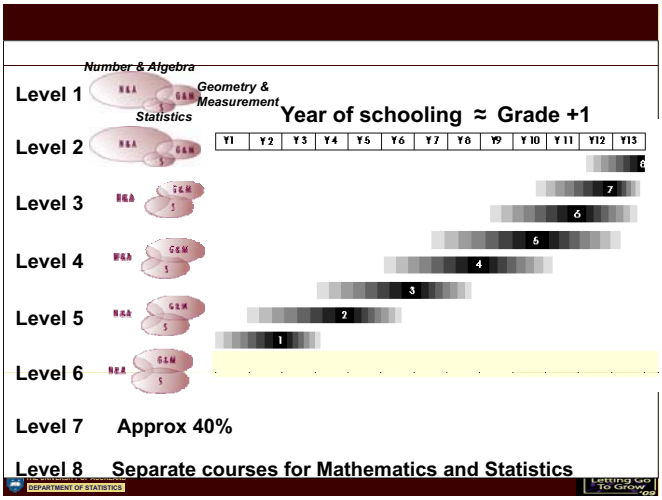
"That's not mathematics"

"No, ... but it *IS* statistics"

What's in a name?

3 strands running through all Levels K-12





One big hole was Inference

- "Doing statistics" without making inferences is boring and largely meaningless
- In NZ, pre-Grade-12 students been encouraged to make inferences about their data for years
  - without any scaffolding or rational basis on which to make them
  - Not built off sampling variation
- New curriculum specifies 3 years of "informal inference" from ages 14-16 to lay the foundations
  - **OOPS !!**

Benefits of panic

- Panic + guilt *is great for spurring action*
- Trying to do "inference" at a stage too young for the usual machinery *unfetters your thinking*
  - an *enforced case* of …

"Statistical inference" – What is it?

Standard statistical usage

- the stuff of conf. intervals, *p*-v… r…
  - ways with accounting for ***u…***
    due to sampling/rando…
    - in well designed and executed stud…s
- This a very narrow area and small part of the whole statistical enterprise
  - We have parallel work focussing on ***"telling data stories"*** which is holistic

Plan of Attack

- Identify and ***work from a minimal set of*** the ***big***gest ***ideas*** of "statistical inference". **Why?**
  - Because students cannot pay attention too many issues simultaneously
    - So make those few connections that matter most
    - Eliminate "busy work"
- ***Put approximate big picture securely in place***, then, over a period of years, …
  - iteratively refine the details, add subtleties, make corrections
- ***Exploit the power of the visual sense***
- Integrate inference within investigative cycle

Building of inferential ideas takes

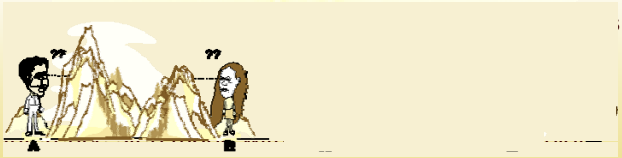
"Notice how the more attention you pay to any one element – picture, story, theory, or rules – the less attention goes to the other three.
In Step … you have to … and the more effectively you do that, the more the story fades into the background."

– George Cobb (1997)

- Whenever students have to struggle with implementation details the big picture gets lost
  - including any consciousness of why they are doing it
  - and what it all means when they have finished

Ideally, …

- want students to have experiences where can
  - ask a question
  - get the data
  - obtain the graphs
  - make descriptive comments
  - and inferential conjectures



Desiderata

- Work from a minimal set of the biggest ideas of statistical inference building off intuitive understandings of the nature of sampling variation
- Mechanics of the inferential step should not be at all demanding
  - Should be able to be performed without students taking their eyes off their graphs
- Should have connections to the more formal methods to be used later (a developmental pathway)

Recipe:

Blend hands-on activities & novel computer animations.
Add spirit of John Tukey and stir. *Bake but do not half-bake*

"Statistical Inference" is a response to …

- ***that particular type of*** **uncertainty** that is caused by having data from random samples rather than the whole universe

(***The nature of the beast***)

- "…
  - not just how samples differ from one another, but how they relate to the parent population(s)/process(es)/distributions

(***The nature of the problem & pointers to solutions***)

It has nothing to say about design & execution issues, data quality, data relevance, importance, …

Corollary

- We should initially approach inference by exploring the nature of sampling variation

But …

"Sampling variation is difficult"
"Students just don't get it"

But maybe that's a consequence of our approaches

- Maybe there is stuff here that

we can **"let go to grow"**

Early Inference: Letting go to grow

- Take a very small set of ideas
  - Run as far as we can with them
    - You'll find that it is quite a long way
- Goals
  - More satisfying & rational ways for beginners to answer interesting real-world questions
  - Building important intuitions
    - Foundations for further learning

Let this lot go to Grow

Letting theoretical distributions go

- ***Build*** inferential ideas in terms of sampling from large finite populations
  - and, in particular, the databases at

census at school new zealand

http://www.censusatschool.org.nz/

(***but use*** them in more general contexts as well)

Letting "null hypotheses" go

Why?

- No one believes their null hypotheses
  - They make no sense with population data
  - Artificial device
- Students find it very hard to think in this way
  - "What would it look like if there was no effect?"

It's a little like looking at the world while standing on your head

Why would anyone do anything so unnatural?

Statistician looks at the world under the Null

Letting "null hypotheses" go

- Causes silly mistakes
  - *"There is no difference"* back in the populations/processes
- Replace by …(back in the populations/processes)
- **either** *making a* ***call*** *on* the ***direction*** of an effect
  or "***can't tell***"

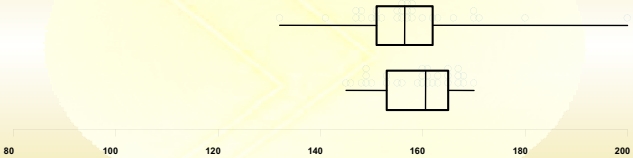
A
B

  - ***"A's tend to be bigger than B's"*** back in the populations

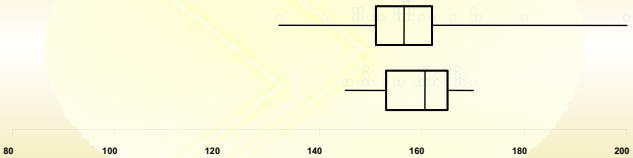
versus

  - ***"I can't tell which is bigger"***
    - "If looked again could well turn out other way around"

Letting conventional reps of sampling variation go



How we usually convey sampling variation

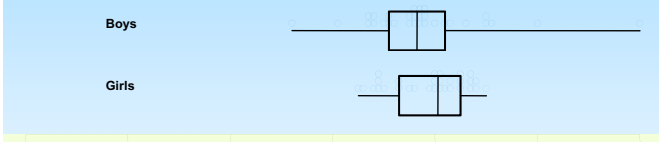


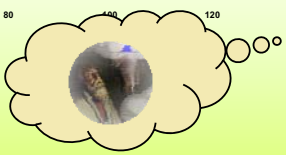
How we usually convey sampling variation

Heights of boys and girls aged 12
from samples of size 30

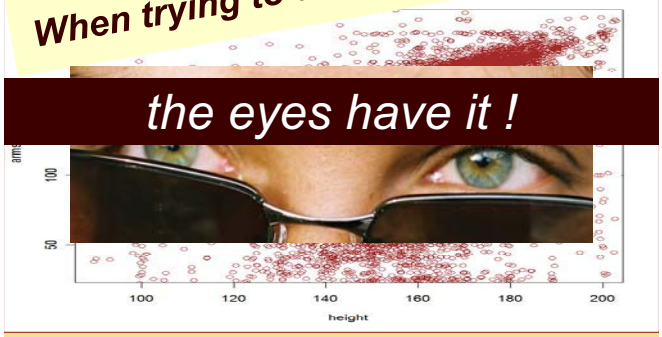
Boys

Girls





When trying to understand data …



the eyes have it !

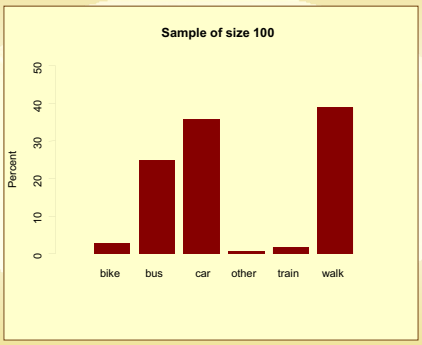
Let's look at some data

from

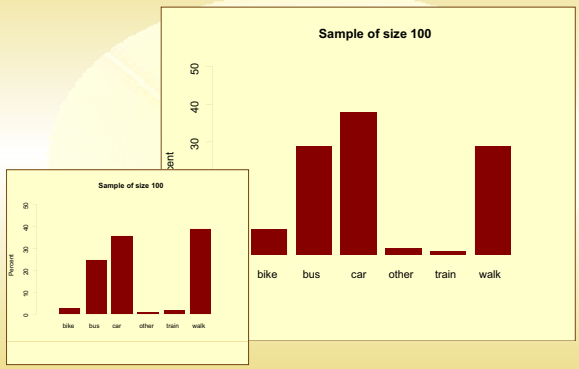
census at school new zealand

http://www.censusatschool.org.nz/

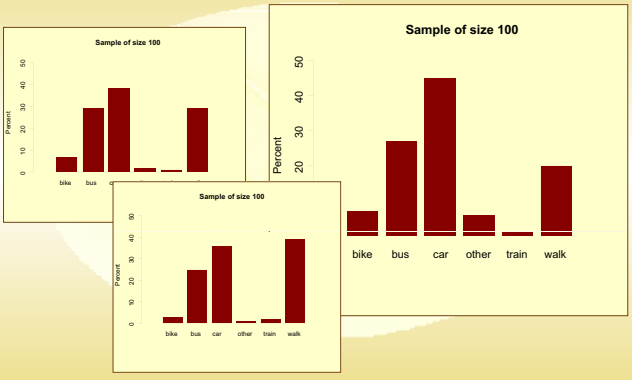
How did they travel to school ?



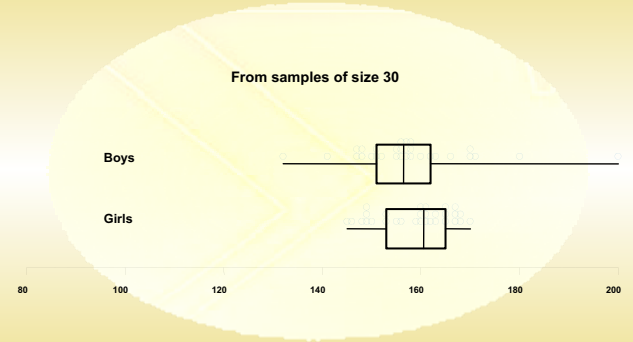
How did they travel to school ?



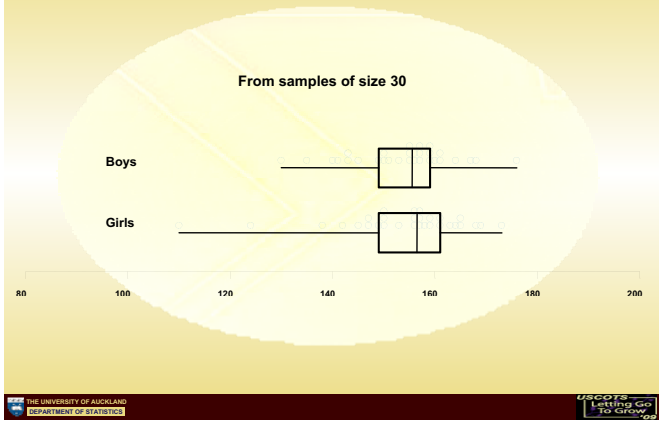
How did they travel to school ?



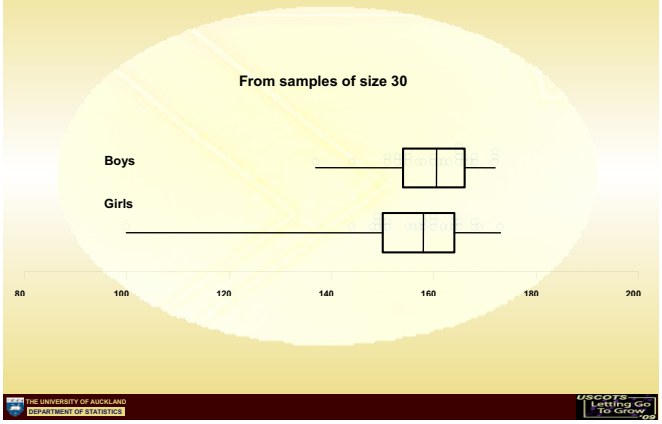
Comparing heights of boys and girls at age 12



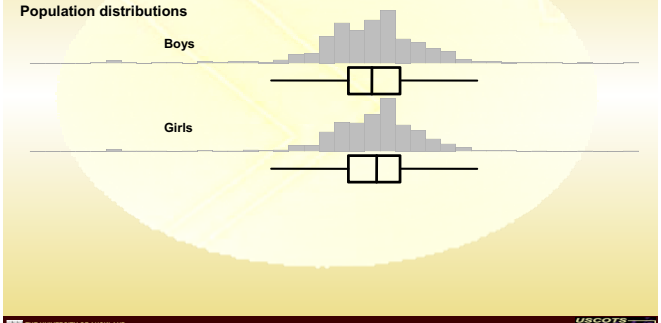
Comparing heights of boys and girls at age 12



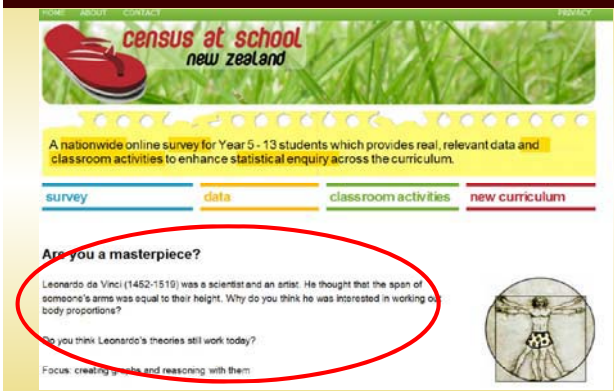
Comparing heights of boys and girls at age 12



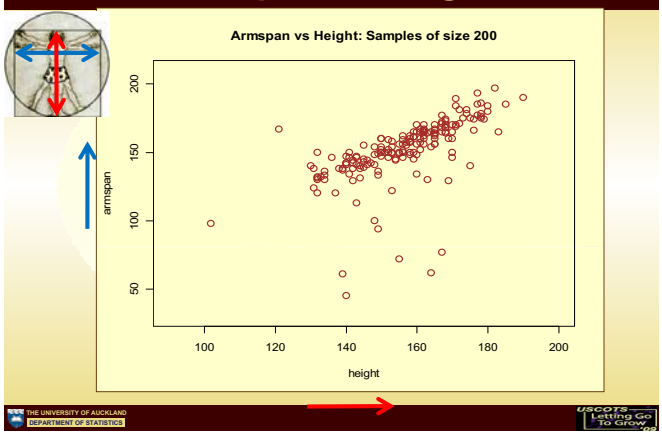
Comparing heights of boys and girls at age 12



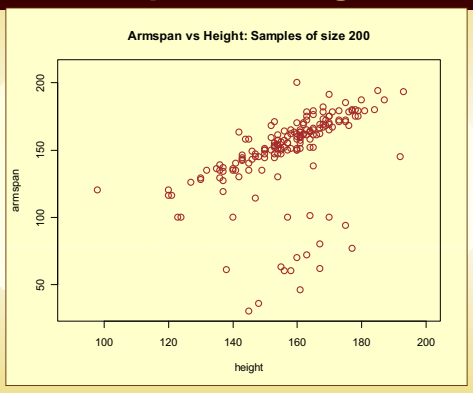
Arm Span vs Height



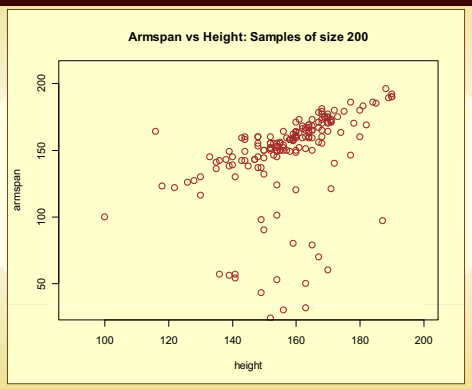
Arm Span vs Height



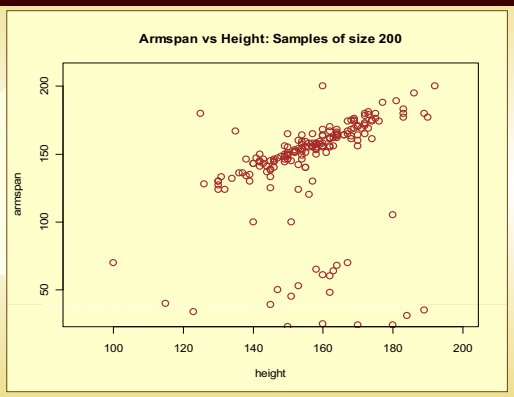
Arm Span vs Height



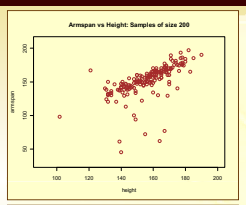
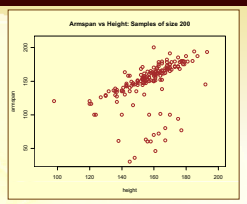
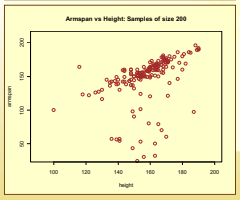
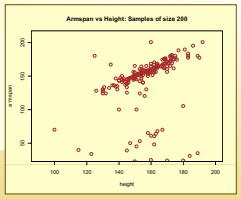
Arm Span vs Height



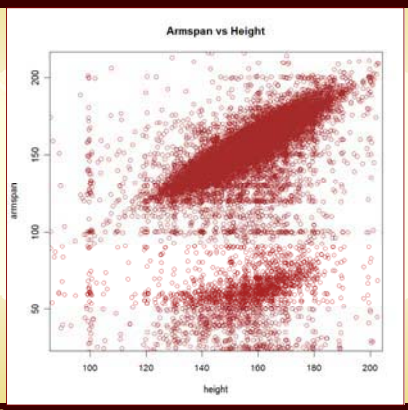
Arm Span vs Height



Arm Span vs Height

Arm Span vs Height



Description vs Inference

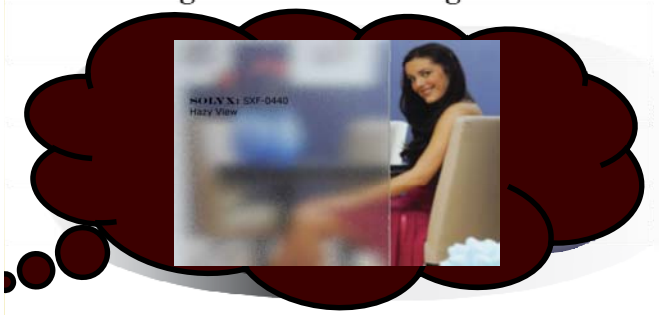
- Description is what I see in the data in hand
  - Theme: "Right here, right now" – Fat Boy Slim
- Inference is what I think is likely to be happening - back in the populations where these data came from
  - Theme: "Back in the USSR" – Beatles
  - We have a natural propensity to move early to inference
  - Many unclear in their thinking & communication when they are describing and when inferring

How do we make inferences?

- Often from coming to believe that something I see in *these* data is a reflection of something occurring back in the populations
- Always know that what we see is, at best, an ***imperfect reflection*** of **the way *it really is*** back in the populations

…

Looking at the world using data



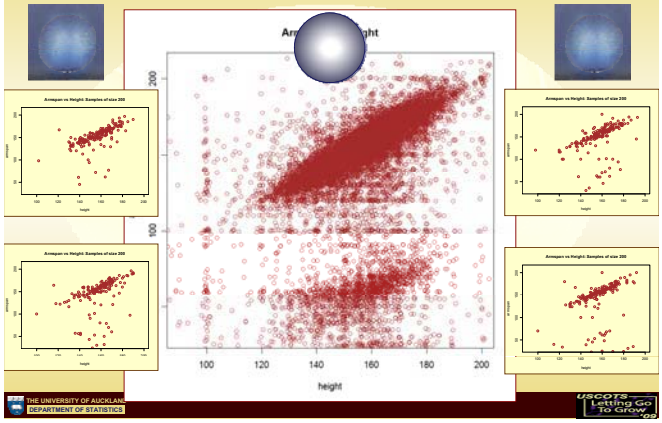
is like looking through a window with ripples in the glass

Looking at the world using data is like

looking through a window with ripples in the glass

"What I see …
is not quite the way it really is"

Arm Span vs Height



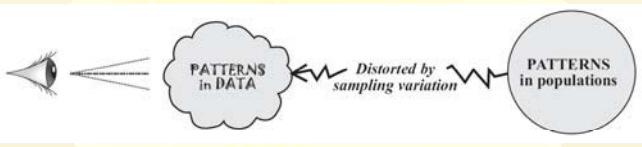
Metaphor to set the stage for statistical inference

Looking at the world using data is like

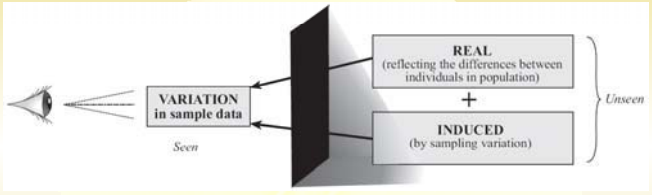
looking through a window with ripples in the glass

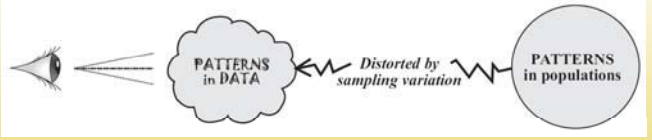
"What I see …
is not quite the way it really is"

"What I see is not quite the way it really is"

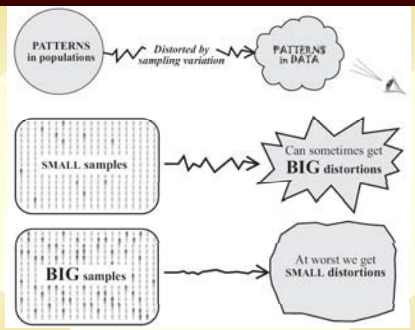


"What I see is not quite the way it really is"





"What I see is not quite the way it really is"

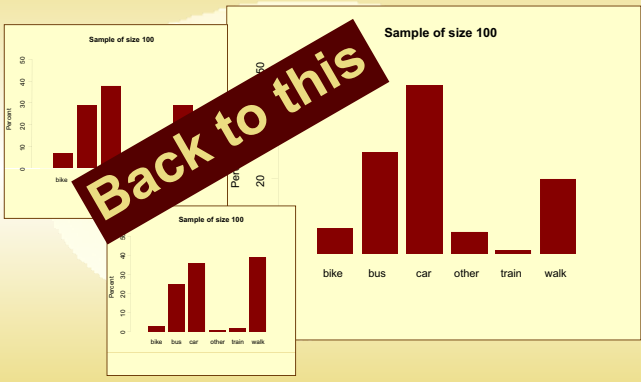


Bigger sample size → *More information* → *Allows me to make more precise claims about what is happening back in the population*

"What I see is not quite the way it really is"

Let's look at some sampling variation

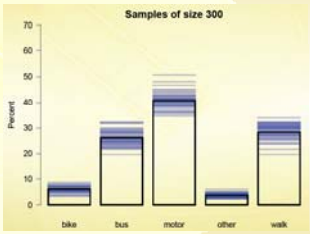
How did they travel to school ?



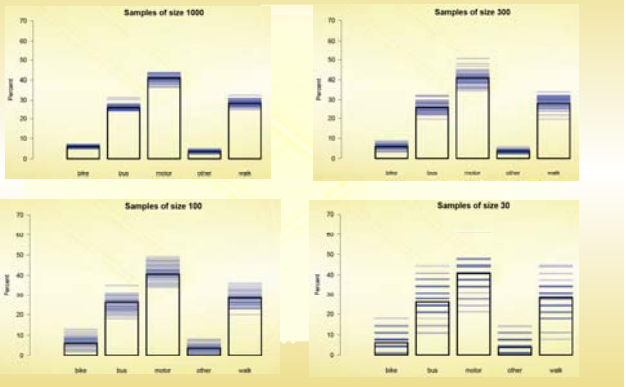
Bar Chart Animations

Play

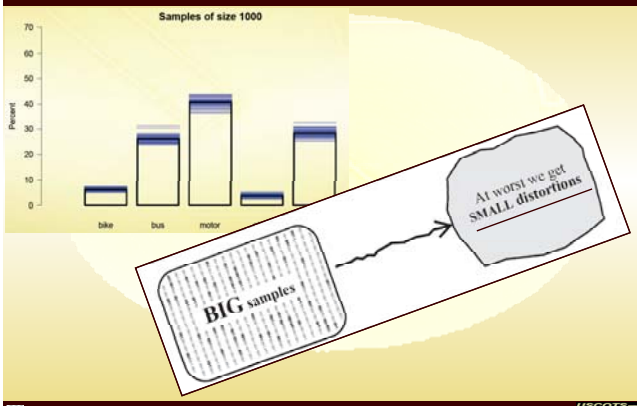
- Samples of 1000
- Samples of 300
- Samples of 100
- Samples of 30
- Old versions 30, 50, 200, 1000



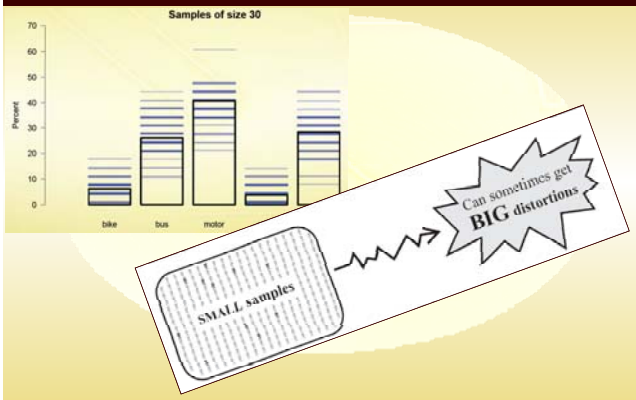
"What I see is not quite the way it really is"



"What I see is not quite the way it really is"

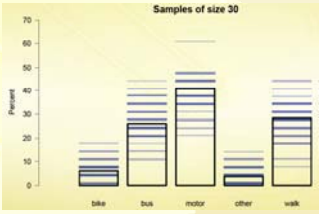


"What I see is not quite the way it really is"

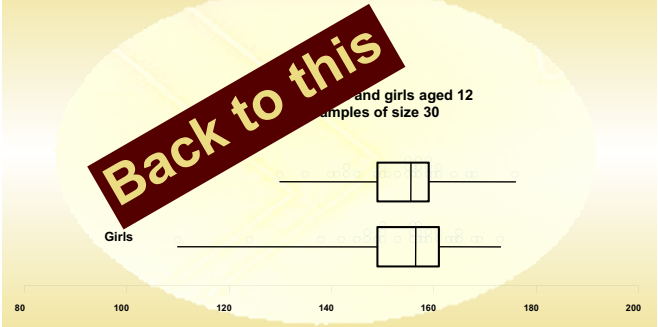


"What I see is not quite the way it really is"

- **What can we learn from proportions taken from samples of size 30?** ("classroom size")
  - Very little !!
- Information content of category data points
  - "Do you fall into this category? Yes/No"
- is very small
- Need very large samples before can say anything very useful
  - Unfortunate fact of life! (Better with measurement data)



Comparing heights of boys and girls at age 12

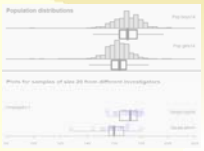


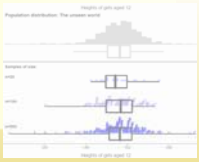
Play movie

Dot and Boxplot Animations

Play

- 2-sample (n=30) , Old
- Effect of sample size

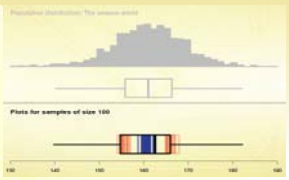


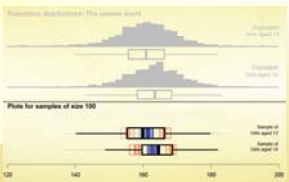


Boxplots with a Memory

Play

- 1-sample build-up, n=30
- 2-sample build-up, n=30
- 1-sample build-up, n=300
- 2-sample build-up, n=300
- 2-sample build-up, n=30 old

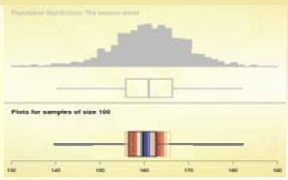


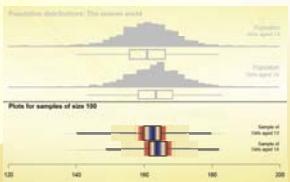


Boxplots with a Memory

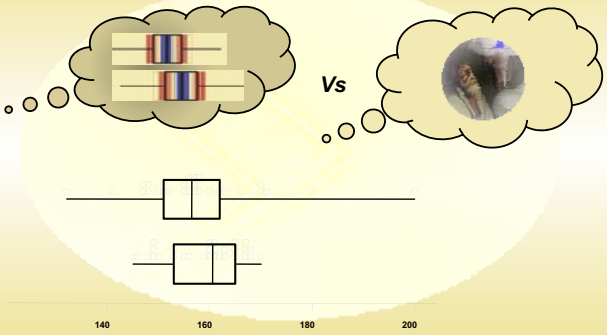
Play

- 1-sample build-up, n=30
- 2-sample build-up, n=30
- 1-sample build-up, n=300
- 2-sample build-up, n=300



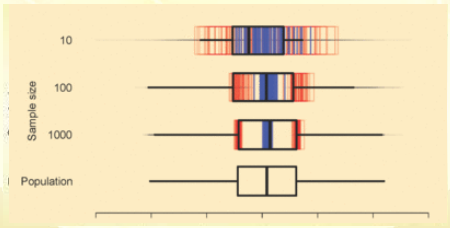


Conveying sampling variation: Contrast …



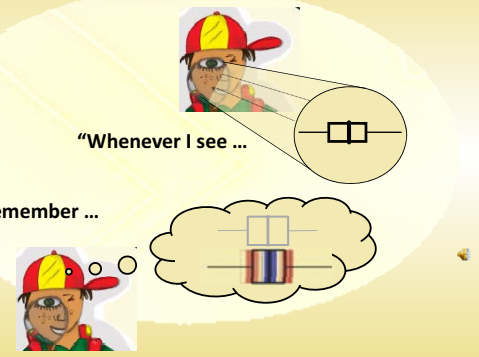
Effect of Sample Size

Effect of Sample Size

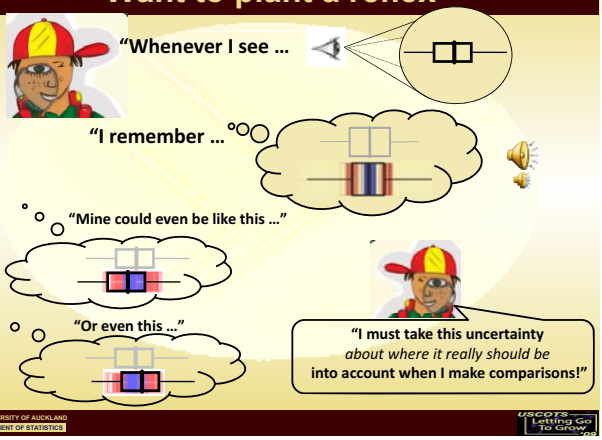


But must ensure students don't just see it as …

Want to plant a reflex of the form …

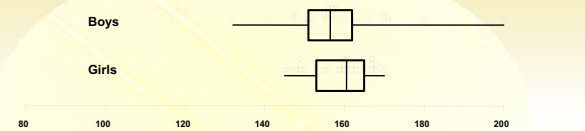


Want to plant a reflex



Back to our data

Heights of boys and girls aged 12 from sample of size 30



What should I say?

Dare I make the claim,
"Girls tend to be taller than boys
back in the populations"

Back to our data

Heights of boys and girls aged 12 from sample of size 30

What should I say?

My boxes aren't quite in the right places

Dare I make the claim,
"Girls tend to be taller than boys
back in the populations"

Back to our data

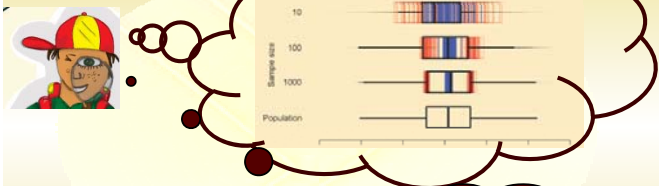
Heights of boys and girls aged 12 from sample of size 30

What should I say?

I can only say if the extent of shift overpowers my uncertainty about where the boxes should be

Dare I make the claim,
"Girls tend to be taller than boys
back in the popula

Also, …



Level of **uncertainty depends on sample size**

Making the Call – the basic idea

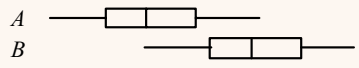
Observed data:	Back in the populations: "Do *B* values tend to be bigger than *A* values?" *My call is ....*	
A B	**B is bigger**	*all sample sizes*
A B	*B* is bigger	
A B	*Claim "B is bigger" if both sample sizes > 20*	*Larger random samples have more information about the populations they came from.* *Thus, with larger random samples, we can make the "B is bigger" call from smaller shifts*
A B	*What's my call here?*	
A B	*What's my call here?*	***But how do we decide?*** *– depends on educational level of students – see next page ..*
A B	*Call "Cannot tell" unless both samples are huge*	
A B	**Cannot tell**	*all sample sizes*

Warning to teachers: avoid doing this sample with sizes smaller than about 20 in each group. Small samples quite often give rise tounstable and often very strange boxplots
To echo the previous diagram, we get very large distortions – see plots for samples of size 10 on page 6

"How to make the call" by Curriculum level

Handout 1, p4 (see website)

At all levels:



If there is no overlap of the boxes, or only a very small overlap make the claim "***B tends to be bigger than A***" back in the populations

Apply the following when the boxes do overlap ...

"How to make the call" by Curriculum level

Handout 1, p5 (see website)

Some notes about the rules

At all levels:

Emphasize the visual, keep the eyes constantly on the plots

- What we are doing here is just one small step in interpreting a comparison
  - It is definitely not "what the statistics module is all about"
- While our depictions are in terms of 2 groups do not hesitate to use more groups
  - The stories uncovered in data by comparing several groups are often much more interesting

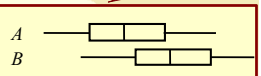
- What we are doing here is just one small step in interpreting a comparison
  - It is definitely not "what the statistics module is all about"

e.g. Handout 2 (see website)

"How to make the call" by Curriculum level

Handout 1, p4

Age 15: *the 3/4-1/2 rule*



If the median for one of the samples lies outside the box for the other sample
("*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]

Majority of one to the right of "**the great whack**" of the other
(Half of one bigger group than ¾ of the other)

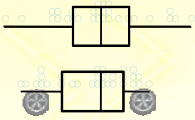
Some notes about the rules

Handout 1, p5

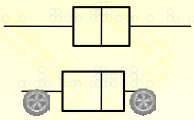
Curriculum Level 5: *the 3/4-1/2 rule*

- The intuitive idea here is "the majority of the *B* group is bigger than the 'the great whack' of the *A* group"
- *Technical aside:* sampling variation alone does not often produce shifts large enough to trigger this rule
  - about 15 times in 100 for samples of size 20, 7 times in 100 for samples of 30, 3 times in 100 for samples of 40, 1 times in 2,500 for samples of size 100.

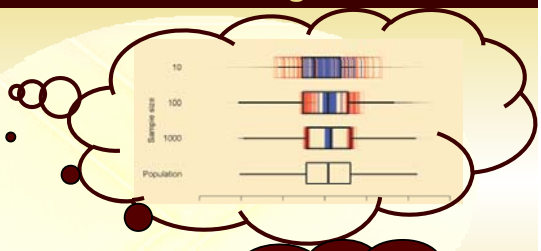
Younger children don't see aggregates and shifts



Younger children don't see aggregates and shifts



Have still to bring in …



Level of
uncertainty depends on sample size

Teacher Aside

What does the 2-sample t-test do?

- It compares
  - the distance between centres
  - to an average measure of within-sample spread
  - down weighted by sqrt(sample size)
- Equivalently it makes the call if
  - the distance between centres as a proportion of within-sample spread
  - exceeds a cutoff which depends on the sample sizes
    - Bigger cutoff-values used for smaller samples

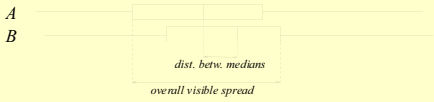
Techo-aside

- Our rules behave pretty well for *normal*, χ_4^2 and t_4 distributions

"How to make the call" by Curriculum level

Handout 1, p. 4

Age 16: *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]

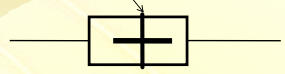
See Tech notes on p. 5

Stress "eye-ball judgements"

Age 17 Intuition

Median for ***my*** data

I got this

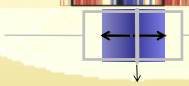


"Where is truth likely to lie?"

I know this sort of thing happens

Truth *is seldom further from* ***data median*** *than this*

Population (Unseen)



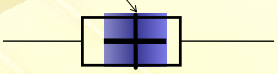
Data median is seldom further from truth than this

True Median
("the unseen truth")

Age 17 Intuition

Median for ***my*** data

I got this



"Where is truth likely to lie?"

Age 17 Intuition

"Where is truth likely to lie?"

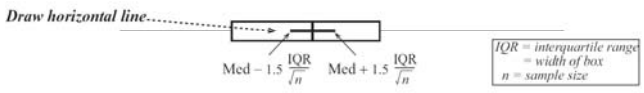
I got this

Truth *is seldom further from* ***my data median*** *than this*

Problem: I don't actually see width of this "uncertainty" band
Why?: I only see **one** frame of sampling variation movie

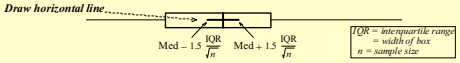
So: We need some sort of estimate of the width of uncertainty band
We know: uncertainty band should be wider for smaller samples

Turns out: The following estimate works pretty well

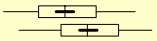


"How to make the call" at Age 17

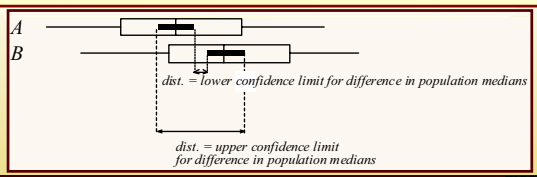
Curriculum Level 7: *based on informal confidence intervals for the population median*



Make the claim ***B tends to be bigger than A*** back in the populations



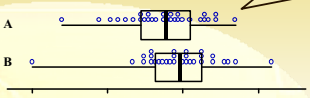
if these horizontal lines (intervals) do not overlap



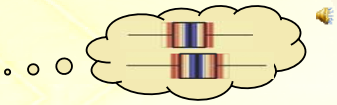
Putting it t

Reminding me that my boxes are not quite in the right places

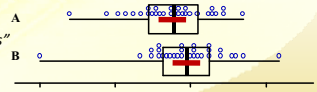
I see …



But
I remember …



Reminding me
I need …
"Uncertainty intervals"



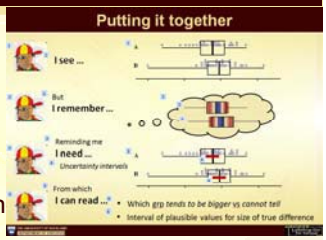
From which
I can read …

- Which group *tends to be bigger* vs *cannot tell*
- Interval of plausible values for size of true difference

Of course there are embellishments…

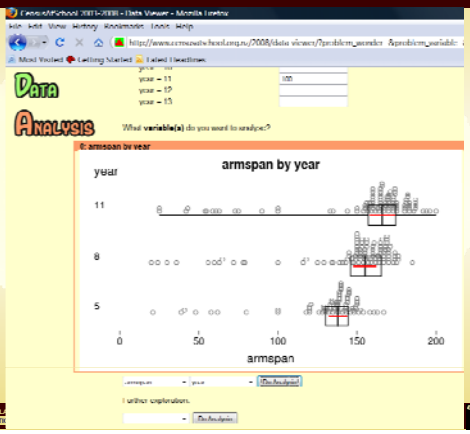
but anyone who even has just this has …

- practically useful tools
- & simple versions of biggest basic ideas of inference to build on



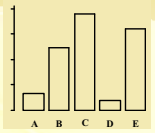
- One embellishment
  - knowledge that these "uncertainty intervals" must get narrower as the sample size gets bigger

From Census At School NZ data-viewer

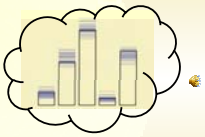


Carry over to category data

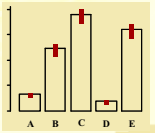
I see …



But
I remember …



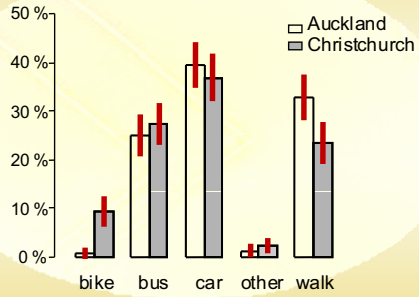
Reminding me
I need …
"Uncertainty intervals"



From which
I can read …

- Which groups *tend to be bigger than which others* vs *cannot tell*
- Intervals of plausible values for sizes of true differences

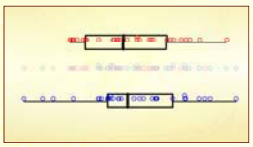
Same game with 2 variables



Software at www.CensusAtSchool.co.nz

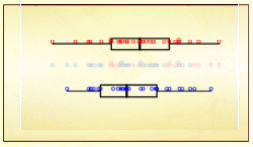
Random Assignment

- **How does that animate?**



Random Assignment

- How does that animate?



Summary

Looking at the world using data is like

"What I see …
is not quite the way it really is"

looking through a window with ripples in the glass

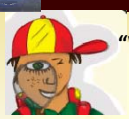
Quick Summary

- ***Description*** is ***what I see*** in the data in hand
- ***Inference*** is what I think is likely to be happening ***back in the populations,*** back *where these data came from*
- In this talk, we have concentrated on inference

Quick Summary

- **Sampling variation alone …**
  - can produce shifts in our box plots
    - **Small shifts** with big samples
    - Sometimes quite big shifts with **small samples**
- **Makes no sense**
  - ***to read meaning*** into shifts in data of a size often produced by sampling variation
- **We have some rules** for signalling when a shift
  - is ***big enough*** that we can make a call on what group gives bigger values

Want to plant a reflex



"Whenever I see …

"I remember …

"Mine could even be like this …"

"Or even this …"

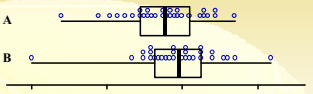
"I must take this uncertainty
about where it really should be
into account when I make comparisons!"

Does the shift we see …..

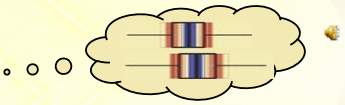
- look bigger than sampling variation would produce?
- The rules
  - Take sample size into account
  - Operated without taking the eyes off the data
  - Get more sophisticated over time
    - Converging towards the tools of formal inference

Putting it together

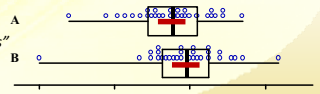
I see …



But
I remember …



Reminding me
I need …
"Uncertainty intervals"



From which
I can read …

- Which group *tends to be bigger* vs *cannot tell*
- Interval of plausible values for size of true difference

If the shift is not big enough …

- then we can't make a call .. on "who is bigger" back in the populations?
  - Simply don't have enough information
- Happens frequently when …
  - the sample sizes are small
    - very little data (very ripply window)
  - differences between the populations are small
    - (looking for fine details rather than gross discrepancies)
- But these are subjects for another talk

Looking at the world using data is like

"What I see …
is not quite the way it really is"

looking through a window with ripples in the glass