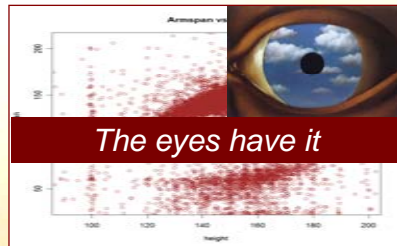


Statistical Informal \wedge Inference *Revisited*



Chris Wild

Dept of Statistics, University of Auckland
New Zealand

The ideas in this talk have developed

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



Maxine Pfannkuch
U. of Auckland, NZ



Matt Regan

U. of Auckland, NZ



Nick Horton

Smith College, MA, USA

“Informal statistical inference”

- important new element of the new curriculum

What is it?

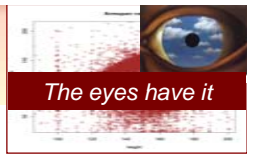
- plain old statistical inference, but ...
 - ***operated simply*** enough for young students

“Informal statistical inference”

We will ...

- Start with the big ideas of statistical inference
- Describe simple methods for students to apply ***when looking at their own data***
 - Minimise steps that lead students to take their eyes off the data
 - ***“Exploit the power of the visual sense”***

When trying to understand data ...



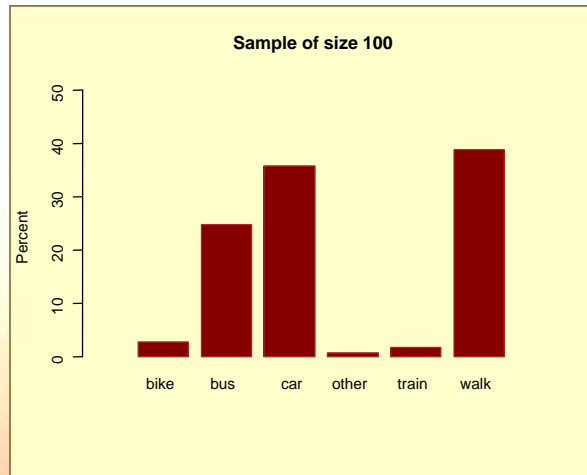
The eyes have it

Let's look at some data
from

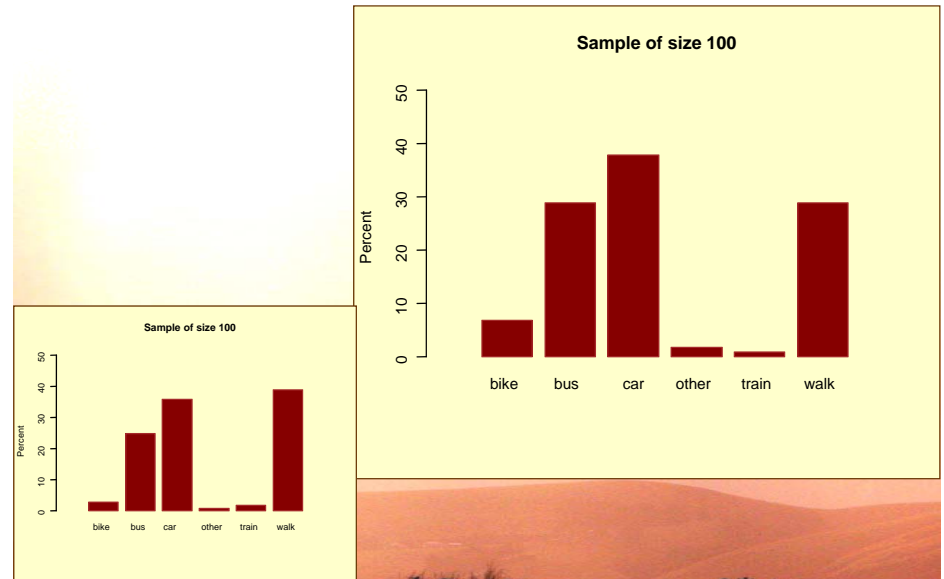


<http://www.censusatschool.org.nz/>

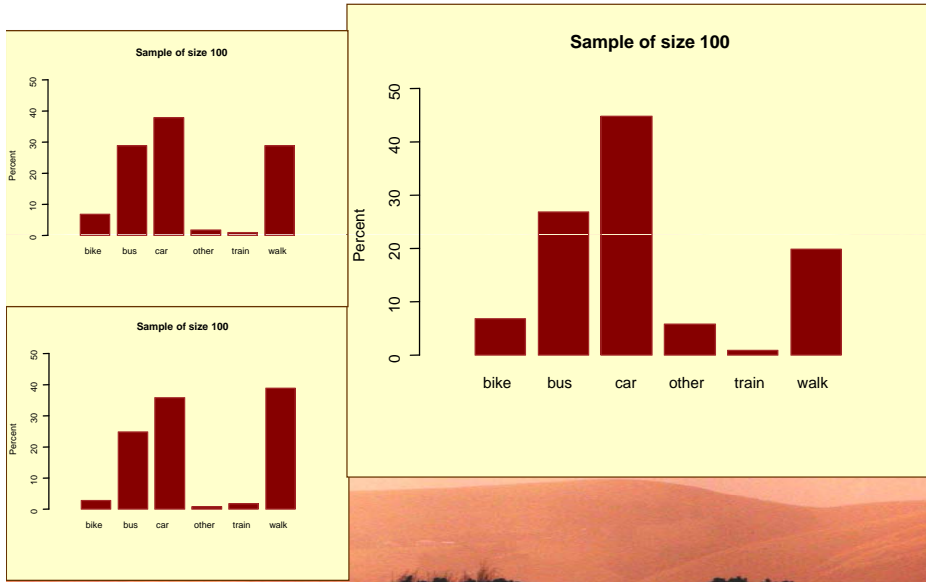
How did they travel to school ?



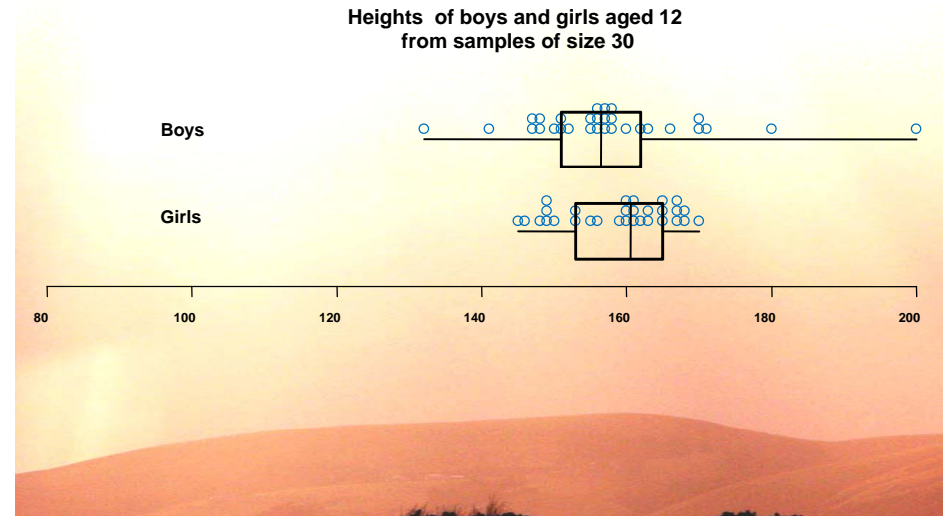
How did they travel to school ?



How did they travel to school ?



Comparing heights of boys and girls at age 12



Comparing heights of boys and girls at age 12

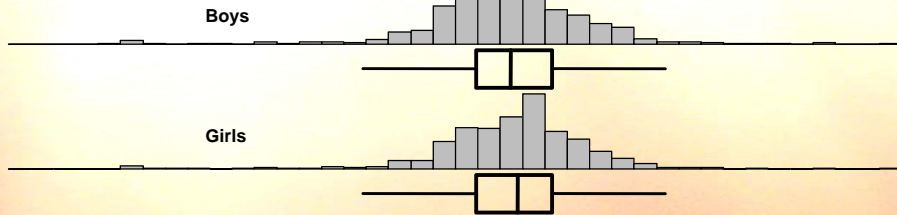


Comparing heights of boys and girls at age 12



Heights of boys and girls aged 12

Population distributions



A nationwide online survey for Year 5 - 13 students which provides real, relevant data and classroom activities to enhance statistical enquiry across the curriculum.

survey

data

classroom activities

new curriculum

Are you a masterpiece?

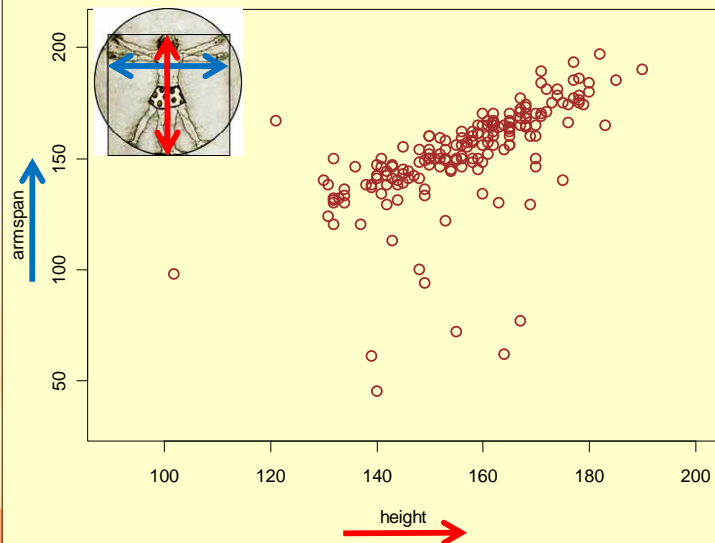
Leonardo da Vinci (1452-1519) was a scientist and an artist. He thought that the span of someone's arms was equal to their height. Why do you think he was interested in working out body proportions?

Do you think Leonardo's theories still work today?

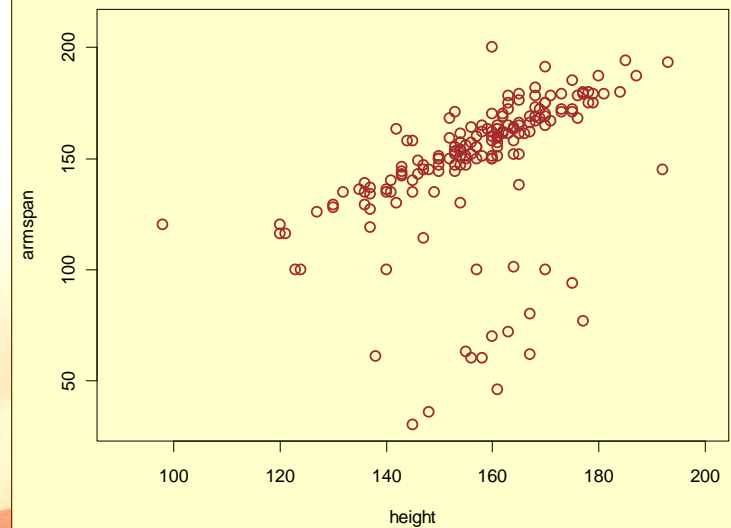
Focus: creating graphs and reasoning with them



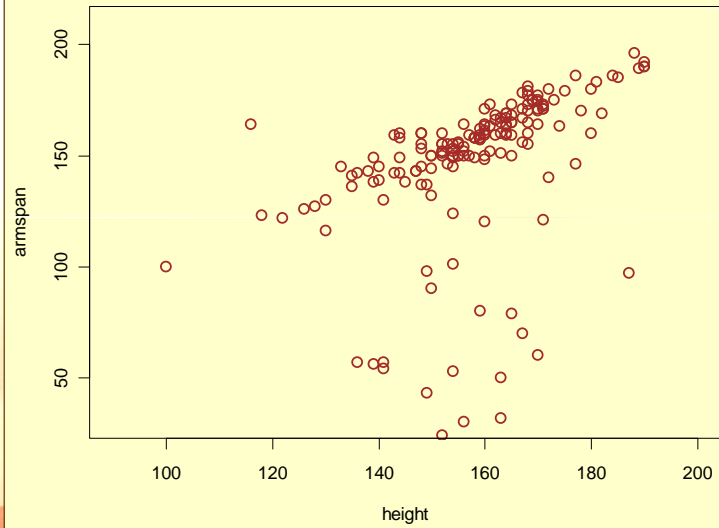
Armspan vs Height: Samples of size 200



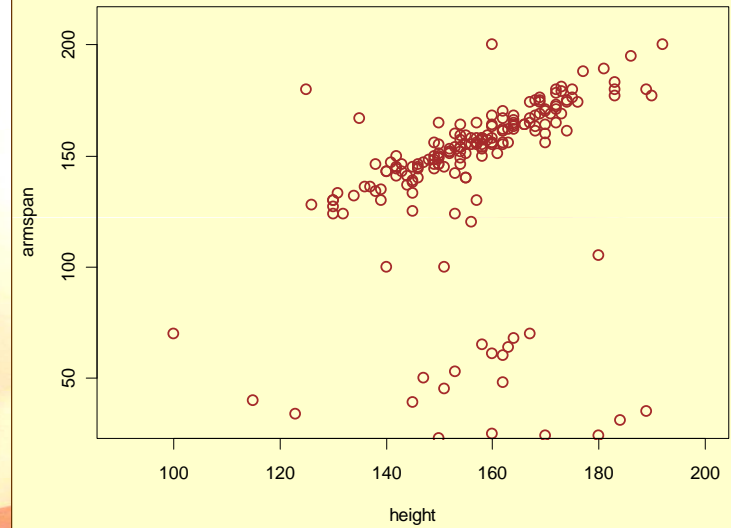
Armspan vs Height: Samples of size 200



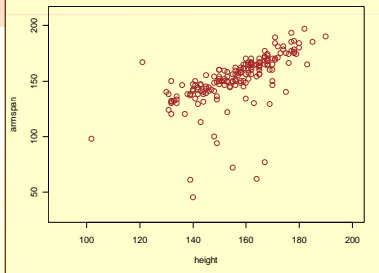
Armspan vs Height: Samples of size 200



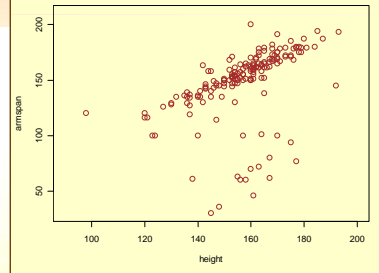
Armspan vs Height: Samples of size 200



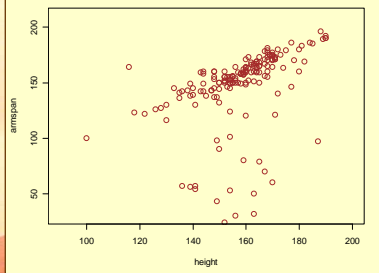
Armspan vs Height: Samples of size 200



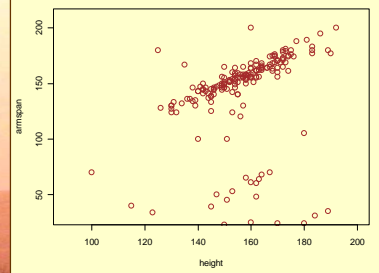
Armspan vs Height: Samples of size 200



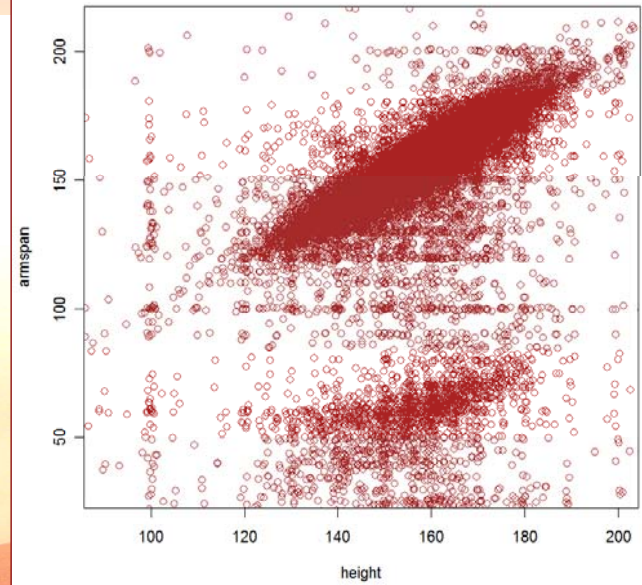
Armspan vs Height: Samples of size 200



Armspan vs Height: Samples of size 200



Armspan vs Height



The nature of statistical inference

Description versus 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*, back 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

Description theme 📢

Inference Theme 📢

We will be concentrating on inference, but ...

To see the richness of the interplay between description and inference at work

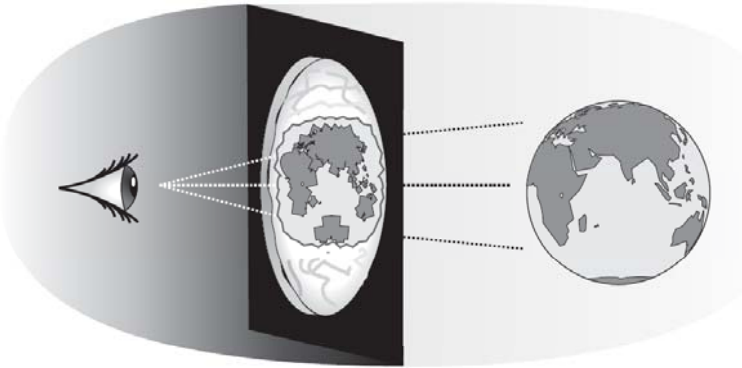
see Handout 2 (on the website)

How do we make inferences?

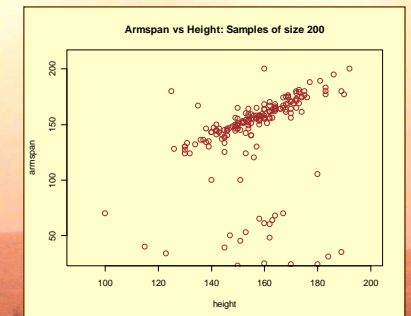
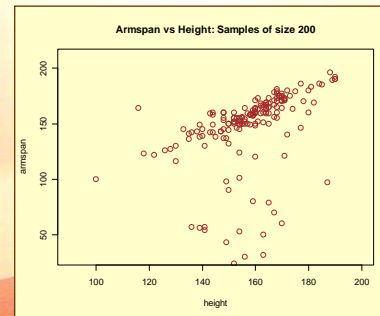
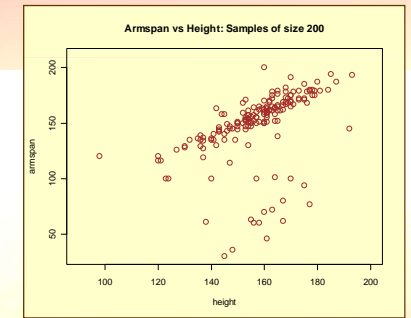
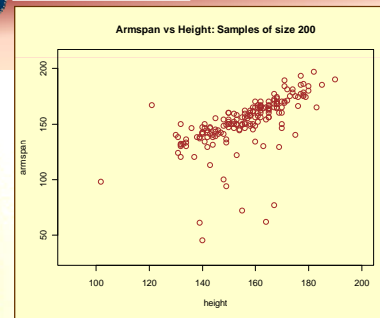
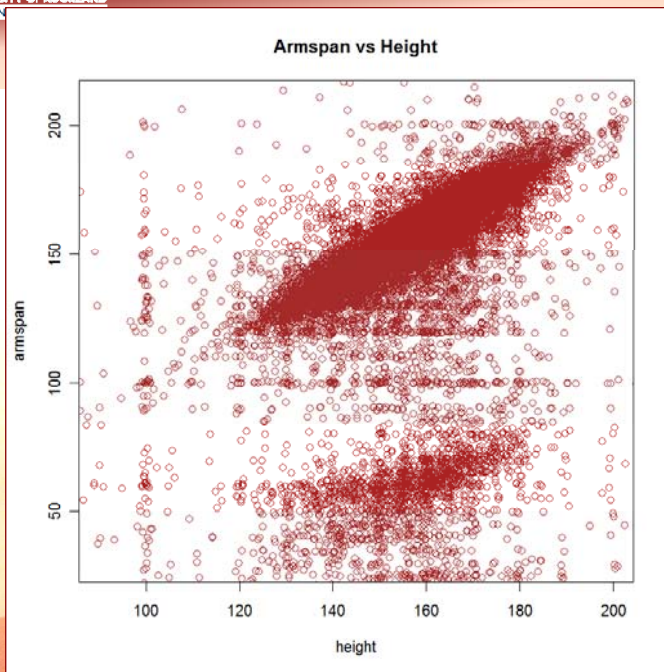
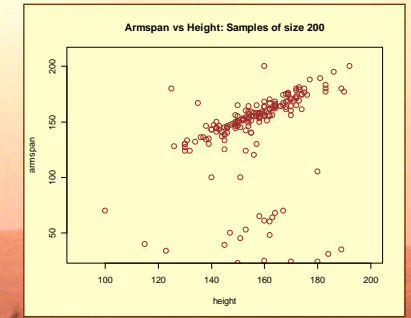
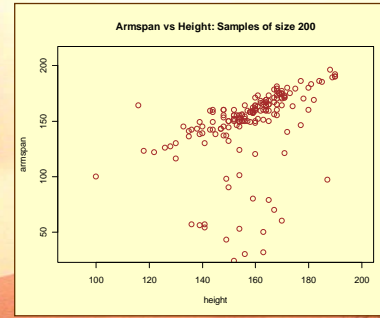
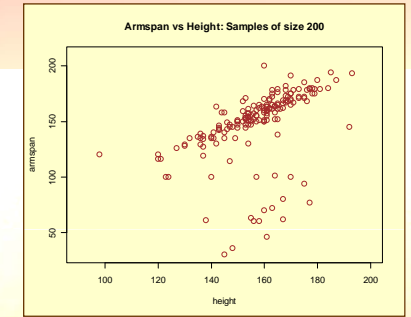
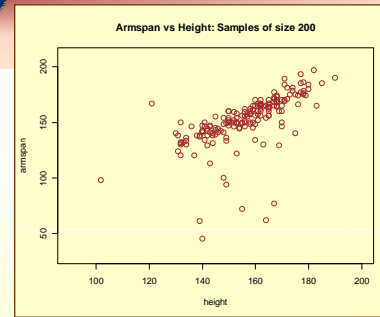
- Often from coming to believe that something I see in *these* data is a reflection of something occurring back in the populations
- Always know that what we see is, at best, an **imperfect reflection** of the way it really **is** back in the populations

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

Looking at the world using data

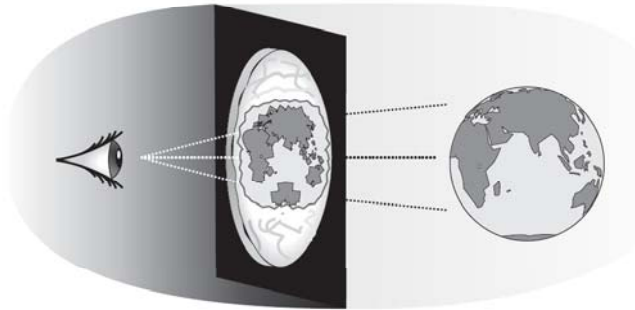


is like looking through a window with ripples in the glass



Metaphor to set the stage for statistical inference

Looking at the world using data



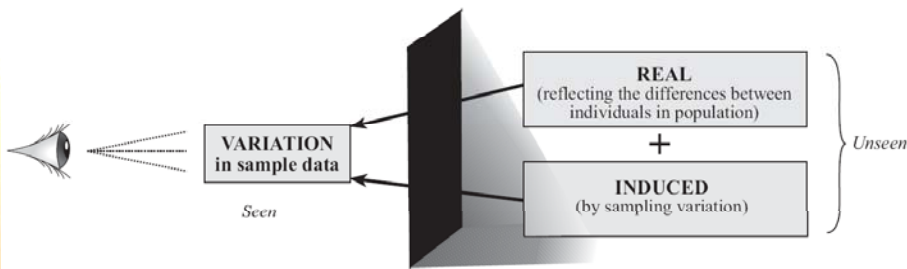
is like looking through a window with ripples in the glass

**“What I see ...
is not quite the way it really is”**

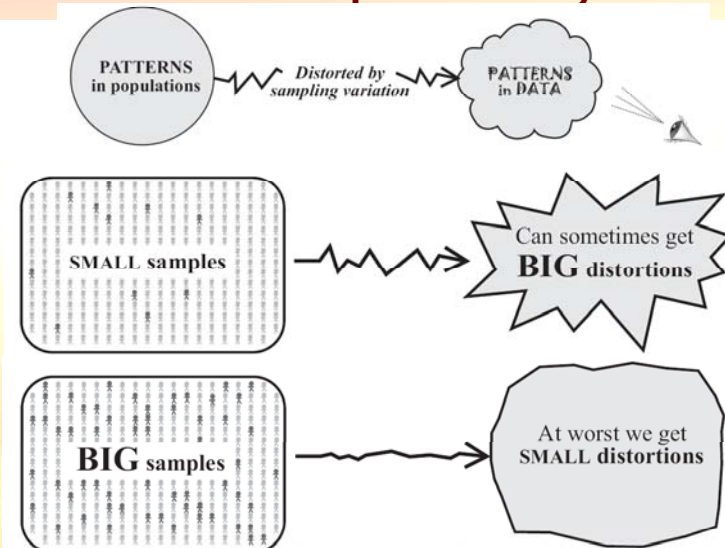
“What I see is not quite the way it really is”



“What I see is not quite the way it really is”



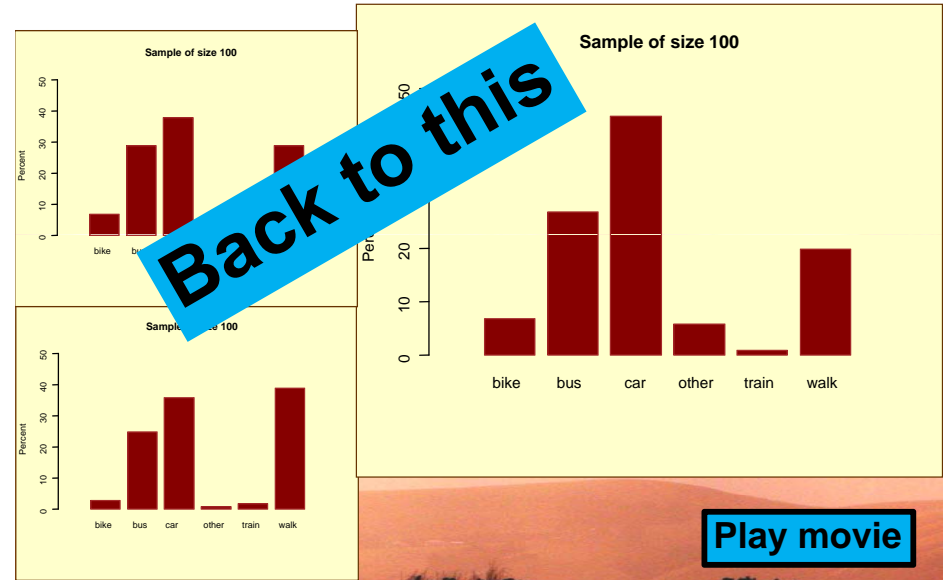
“What I see is not quite the way it really is”



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

Let's look at some sampling variation

How did they travel to school ?



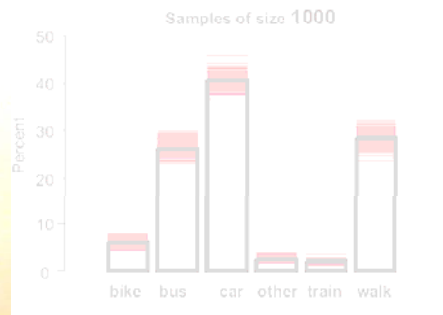
Play movie

Bar Chart Animations

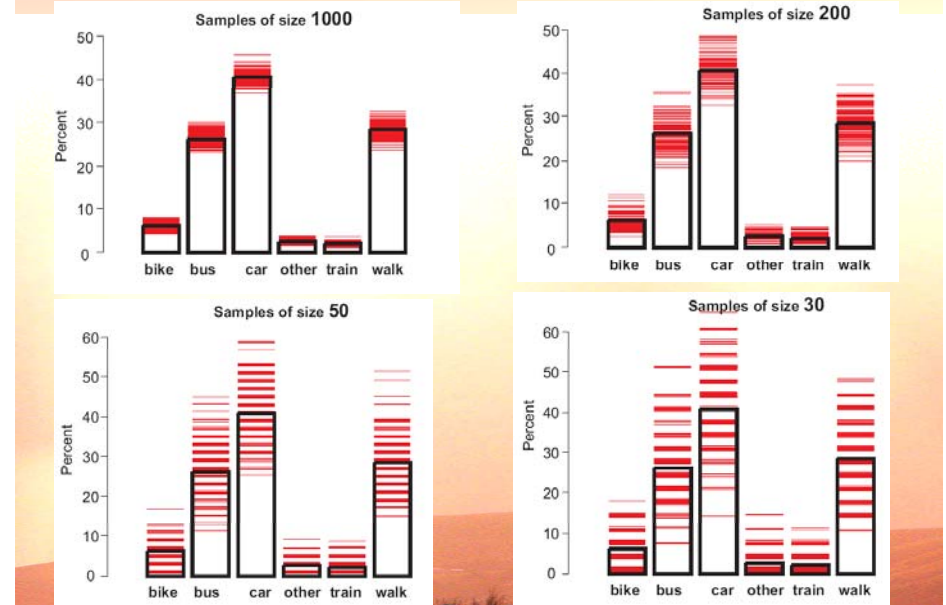


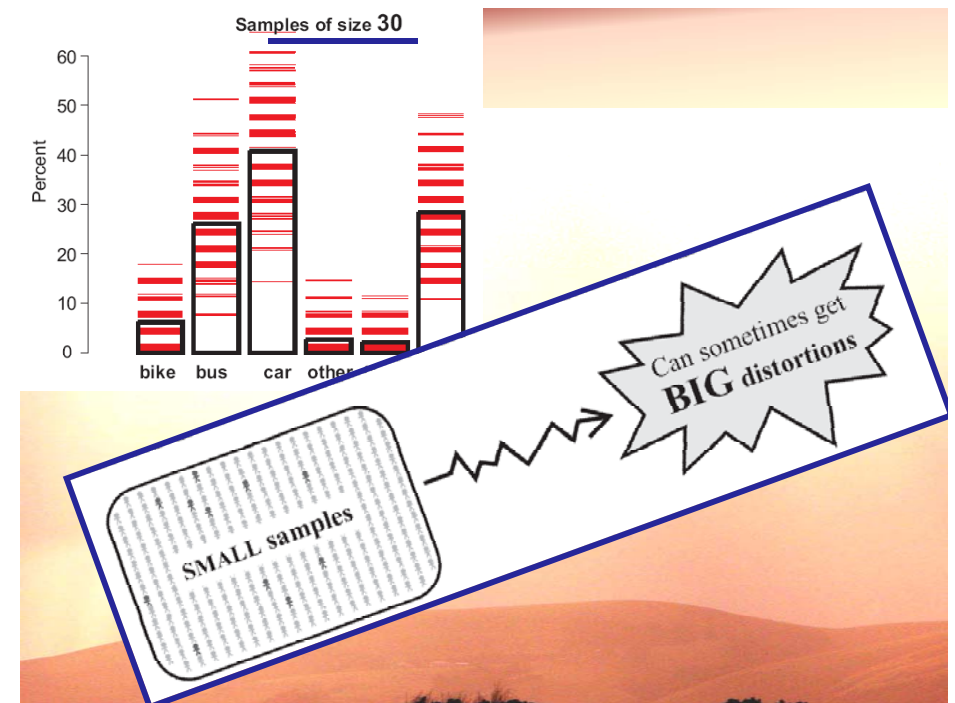
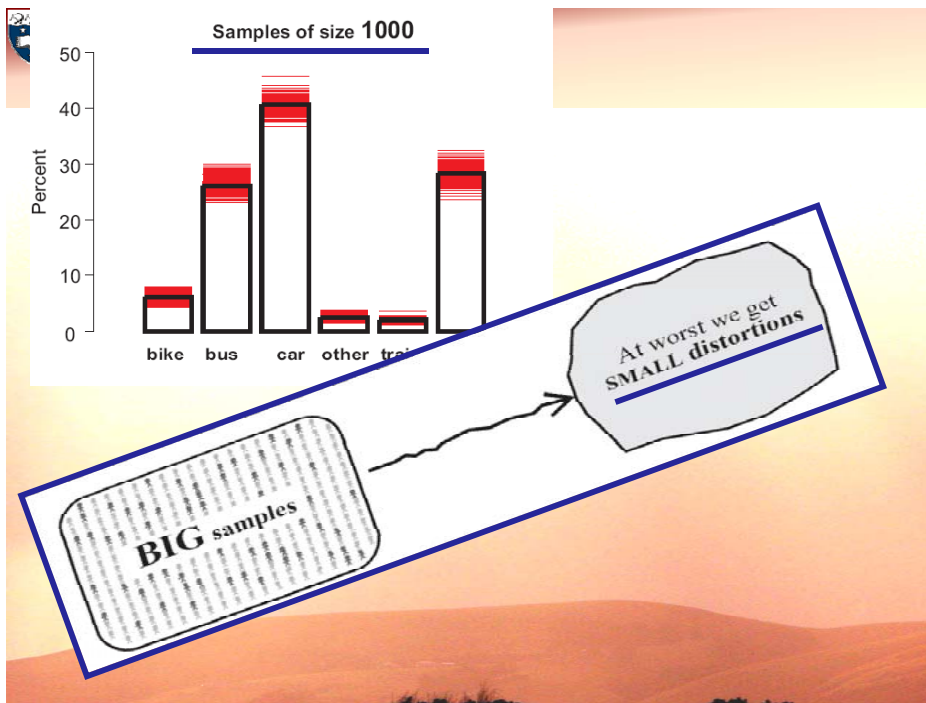
Play

- Samples of 1000
- Samples of 200
- Samples of 100
- Samples of 50
- Samples of 30
- Samples of 30 without jitter



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





THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

“What I see is not quite the way it really is”

- **What can we learn from proportions taken from samples of size 30?**
 - 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!
 - Situation better with measurement data

Samples of size 30

Percent

bike bus car other train walk

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Comparing heights of boys and girls at age 12

Back to this

Boys and girls aged 12
samples of size 30

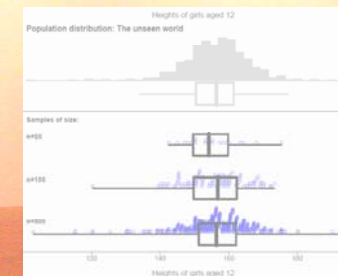
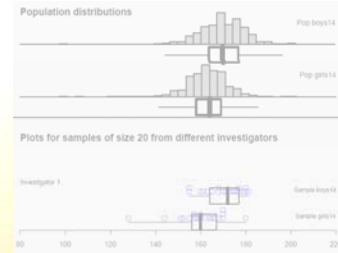
Boys

Girls

80 100 120 140 160 180 200

Play movie

Dot and Boxplot Animations



Play

- Original 2-sample
- Effect of sample size

Boxplots with a Memory I



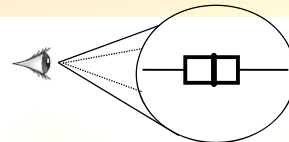
Play 1-sample build-up, n=30



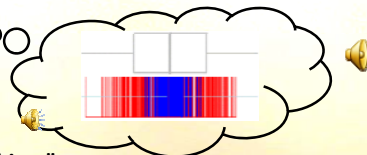
Want to plant a reflex



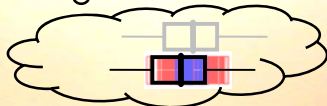
"Whenever I see ..."



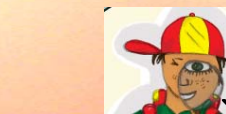
"I remember ..."



"Mine could even be like this ..."



"Or even this ..."



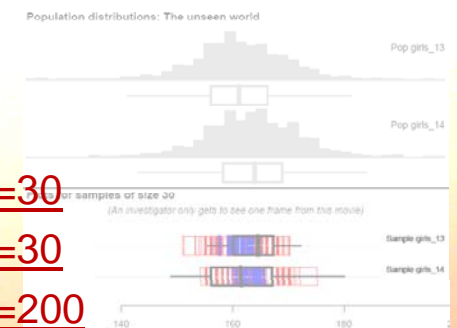
"I must take this uncertainty about where it really should be into account when I make comparisons!"

Boxplots with a Memory II



Play

- 1-sample build-up, n=30
- 2-sample build-up, n=30
- 1-sample build-up, n=200
- 2-sample build-up, n=200



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



Must securely anchor
to something real
and believable

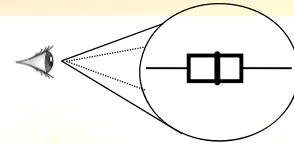
-- Maxine & Pip have great ideas



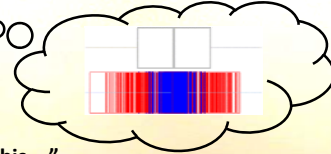
Want to plant a reflex



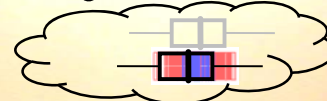
"Whenever I see ..."



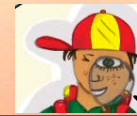
"I remember ..."



"Mine could even be like this ..."

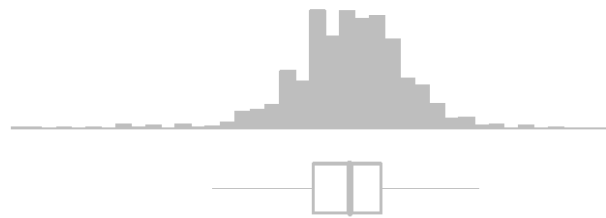


"Or even this ..."



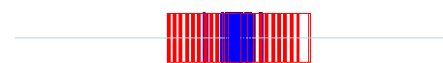
"I must take this uncertainty
about where it really should be
into account when I make comparisons!"

Population distribution: The unseen world



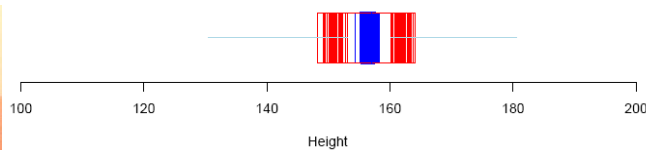
Plots for samples of size 30

(An investigator only gets to see one frame from this movie)

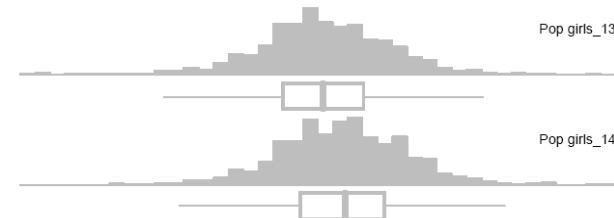


Plots for samples of size 200

(An investigator only gets to see one frame from this movie)

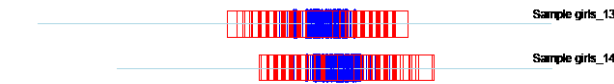


Population distributions: The unseen world



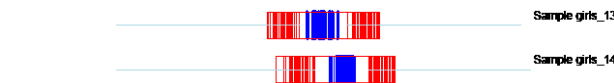
Plots for samples of size 30

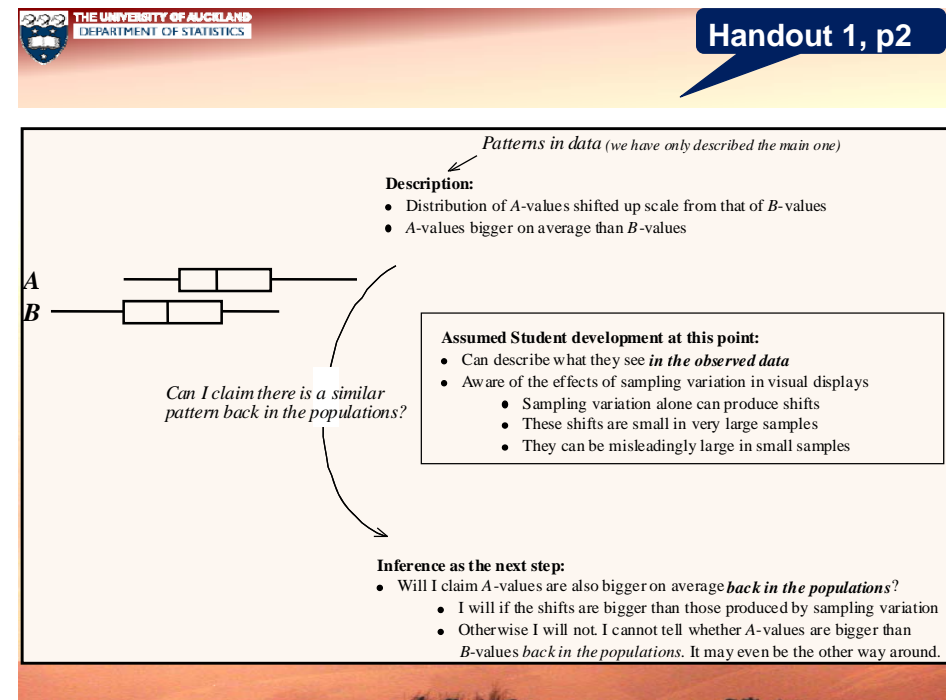
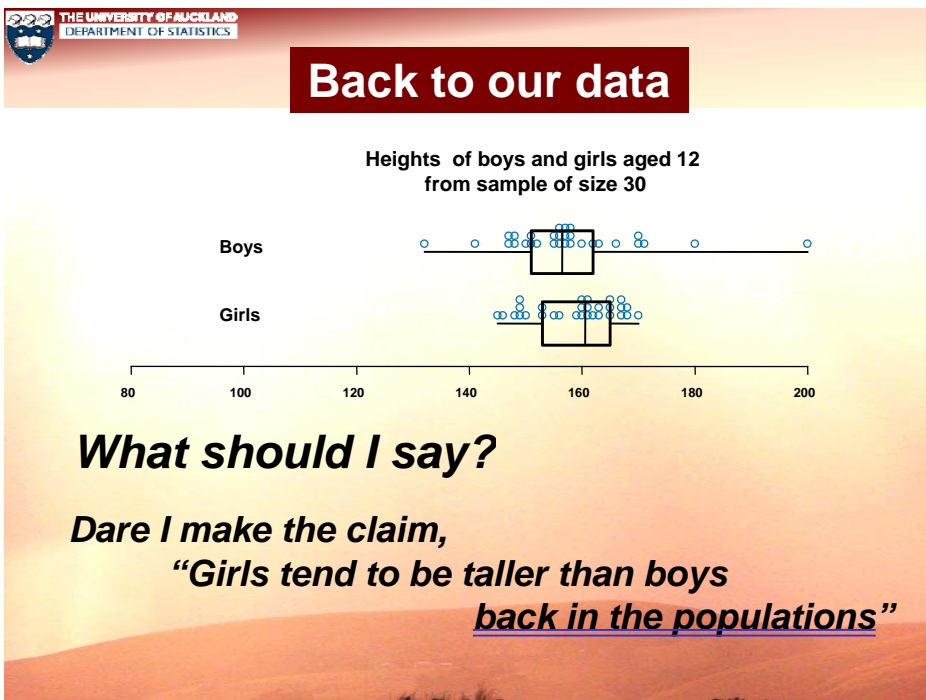
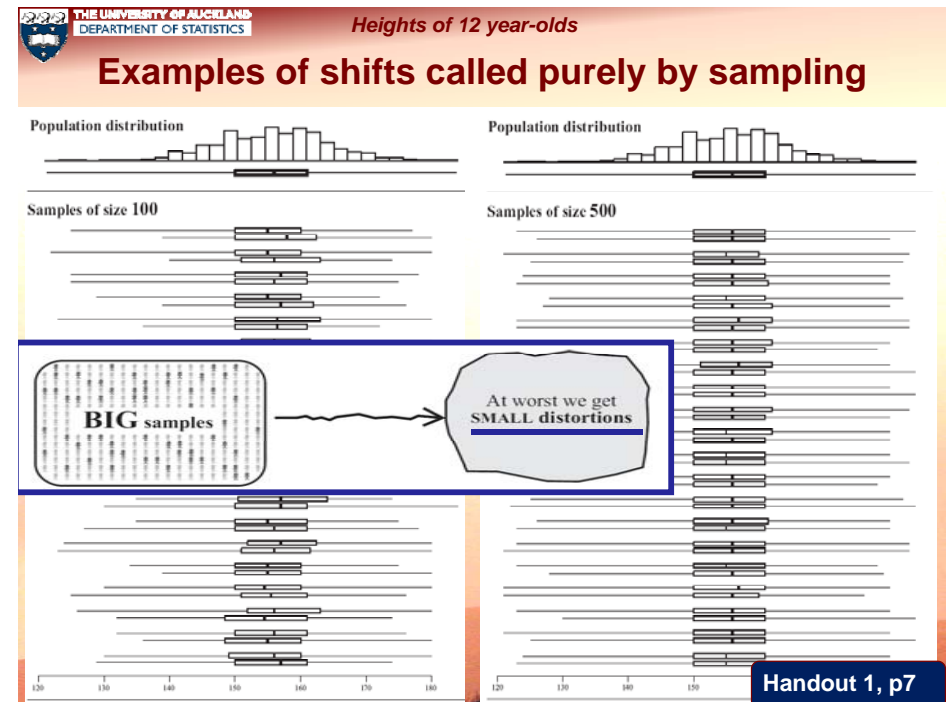
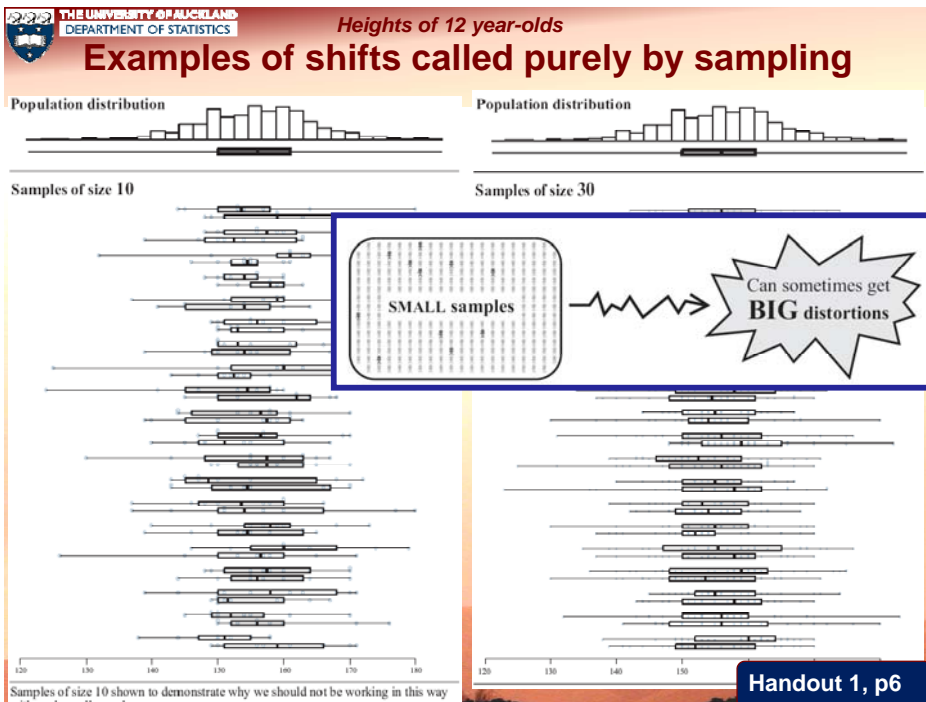
(An investigator only gets to see one frame from this movie)



Plots for samples of size 200

(An investigator only gets to see one frame from this movie)





Making the Call – the basic idea

Observed data:

Back in the populations:

Making the Call – the basic idea

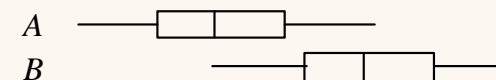
		B is bigger	} all sample sizes	Handout 1, p3
		B is bigger		
		Claim "B is bigger" if both sample sizes > 20	} Larger random samples have more information about the populations they came from.	
		What's my call here?		
		What's my call here?	} Thus, with larger random samples, we can make the "B is bigger" call from smaller shifts	
		Call "Cannot tell" unless both samples are huge		
		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 to untrustworthy 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

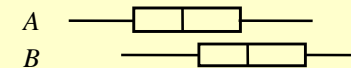
- What we are doing here is just one small step in interpreting a comparison
 - It is definitely not "what the statistics module is all about"

e.g. **Handout 2** (see website)

"How to make the call" by Curriculum level

Handout 1, p4

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



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

Some notes about the rules

Handout 1, p5

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

- The intuitive idea behind the 3/4-1/2 rule is that if the B group is bigger than the "the great whack" of the A group
- Technically, the 3/4-1/2 rule states that if the median of the B group is to the right of the 3/4 mark of the A group, then we can claim "B tends to be bigger than A" back in the populations. This rule is applied 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.

See handout 1, p5 for discussion

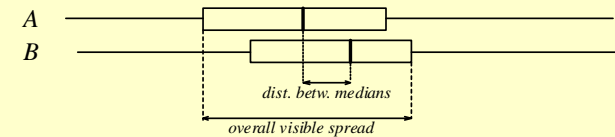
Teacher Aside

- 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 $\sqrt{\text{sample size}}$
- 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

“How to make the call” by Curriculum level

Handout 1, p. 4

Curriculum Level 6: distance between medians as proportion of “overall visible spread”



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

$\frac{1}{3}$ of overall visible spread for sample sizes of around 30

$\frac{1}{5}$ of overall visible spread for sample sizes of around 100

[Could also use $\frac{1}{10}$ of overall visible spread for sample sizes of around 1000]

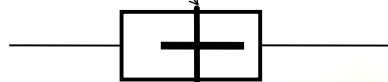
See Tech notes on p. 5

Stress “eye-ball judgements”

Level 7 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 *my data median* than this

Population (Unseen)



True Median (“the unseen truth”)

Level 7 Intuition

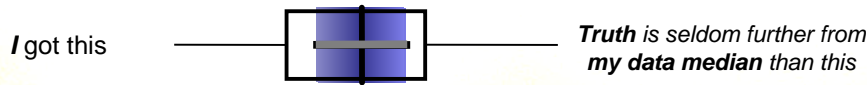
Median for *my* data

I got this



“Where is truth likely to lie?”

“Where is truth likely to lie?”



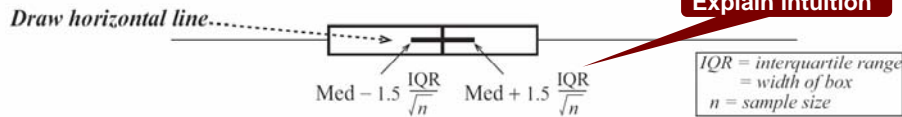
Problem: I don't actually see width of this “uncertainty” band

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

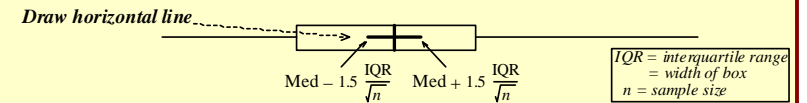
Turns out: The following estimate works pretty well



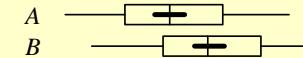
“How to make the call” by Curriculum level

Handout 1, p. 4

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



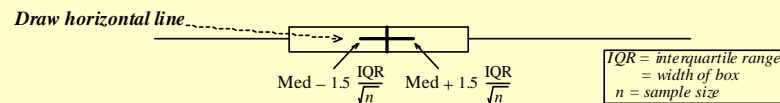
if these horizontal lines (intervals) do not overlap

See Tech notes on p. 5

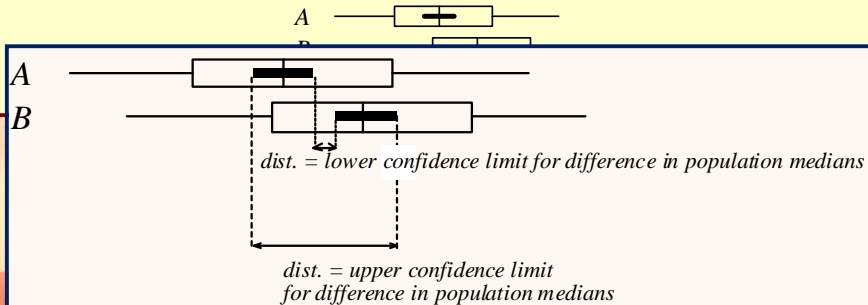
“How to make the call” by Curriculum level

Handout 1, p. 4

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



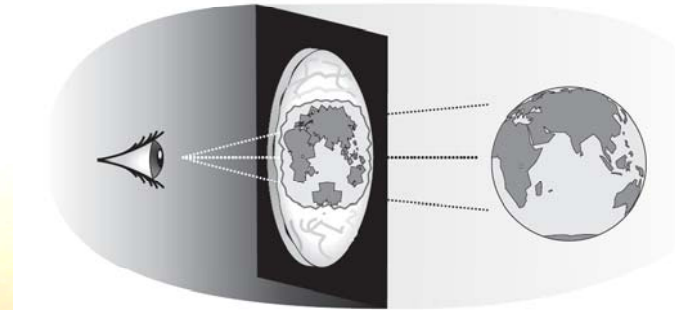
“How to make the call” by Curriculum level

Curriculum Level 8: on to formal inference

Informal Inference Summary

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”***

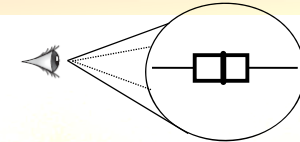
Quick Summary

- **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

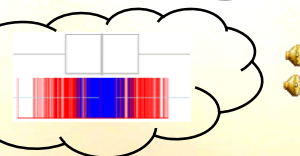
Want to plant a reflex



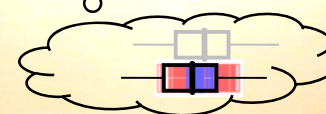
“Whenever I see ...”



“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!”***

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

Does the shift we see

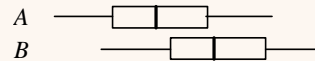
look bigger than sampling variation would produce?

- **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

Is the data shift big enough?

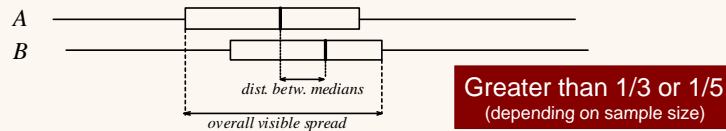
(for us to claim *B* bigger than *A* back in the populations)

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



Majority of one to the right of “the great whack” of the other

Curriculum Level 6: distance between medians as proportion of “overall visible spread”



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



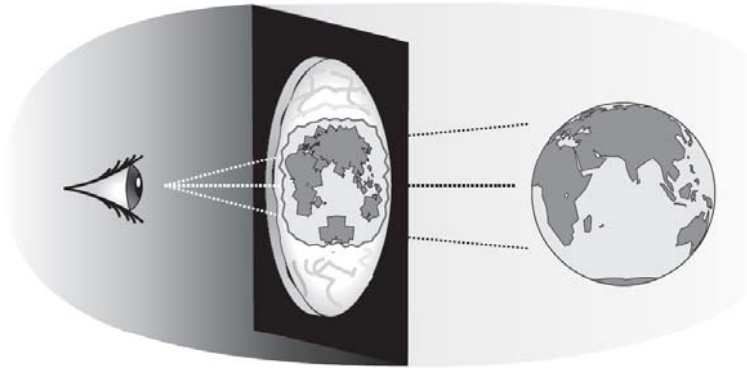
Separation (no overlap) of constructed intervals

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 looking through a window with ripples in the glass