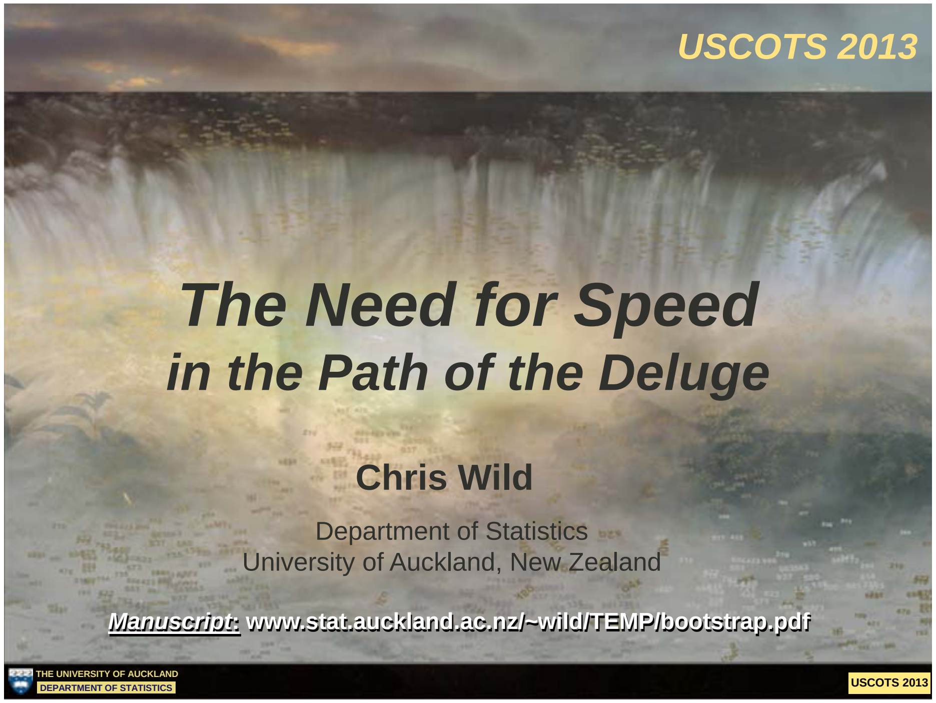


USCOTS 2013

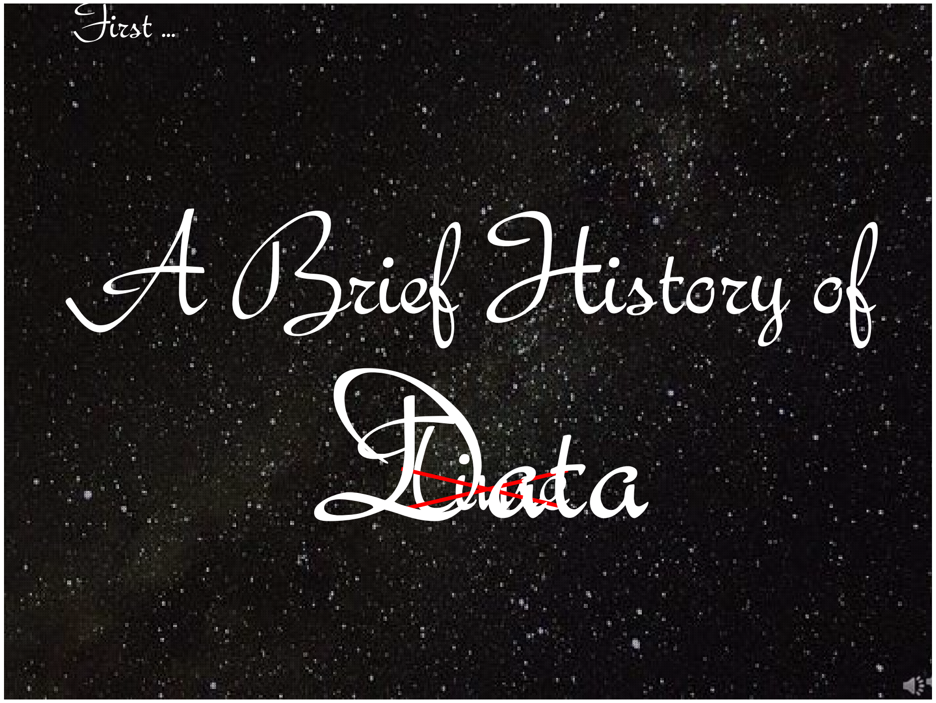
The Need for Speed
in the Path of the Deluge

Chris Wild

Department of Statistics
University of Auckland, New Zealand

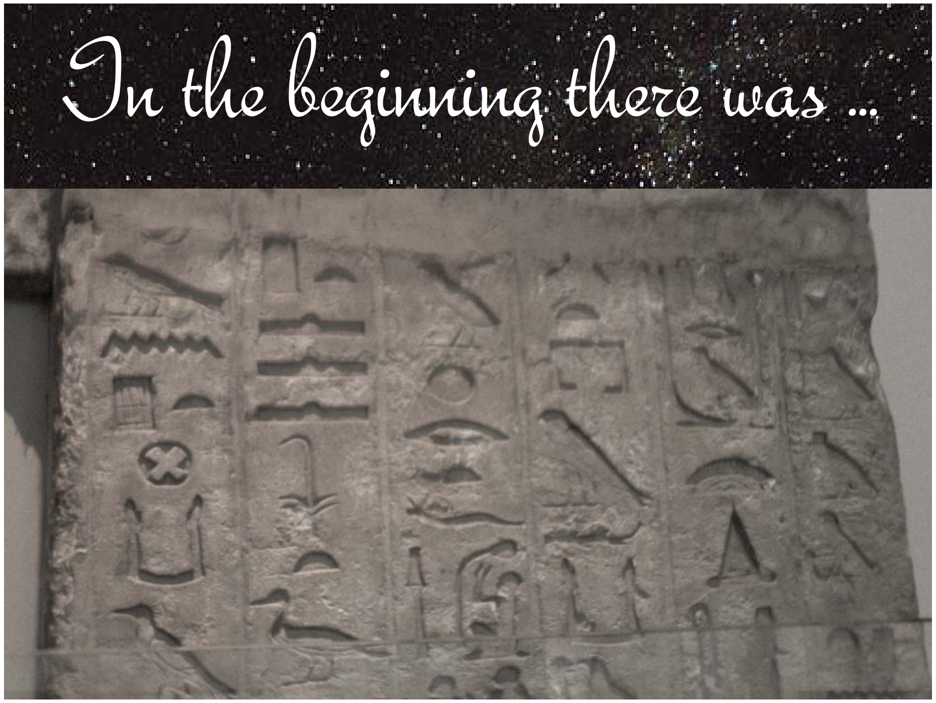
Manuscript: www.stat.auckland.ac.nz/~wild/TEMP/bootstrap.pdf

The Need for Speed ?

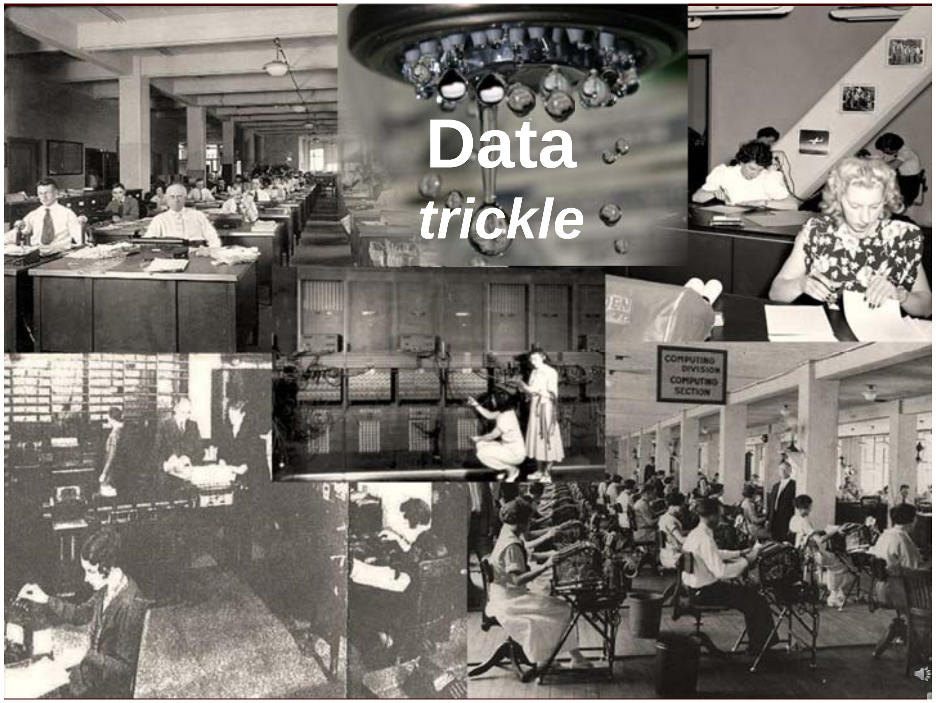


First ...

A Brief History of ~~Time~~ Data

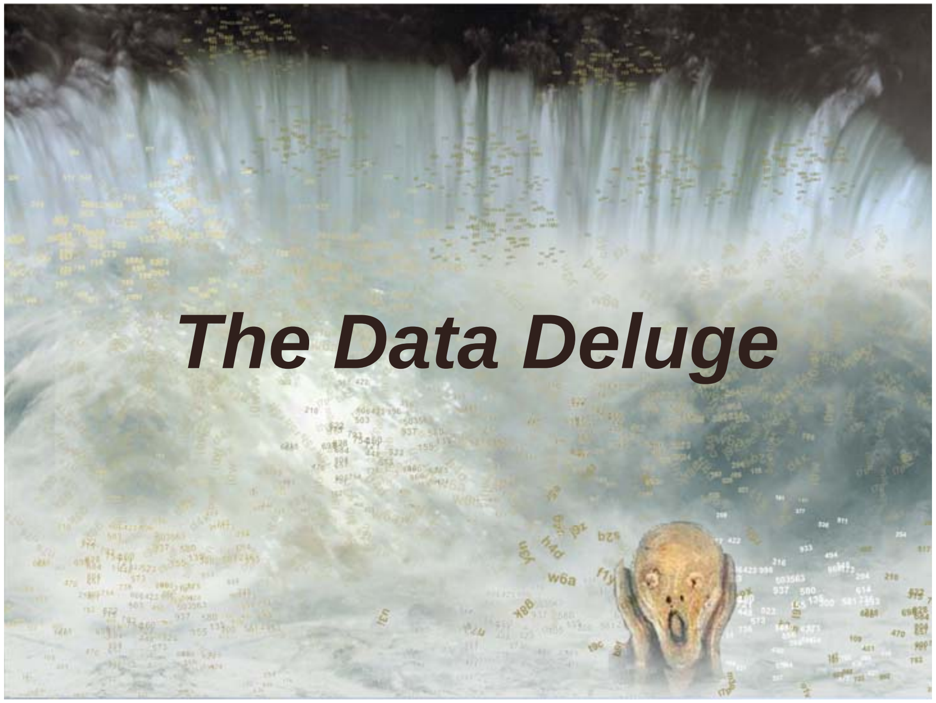


In the beginning there was ...

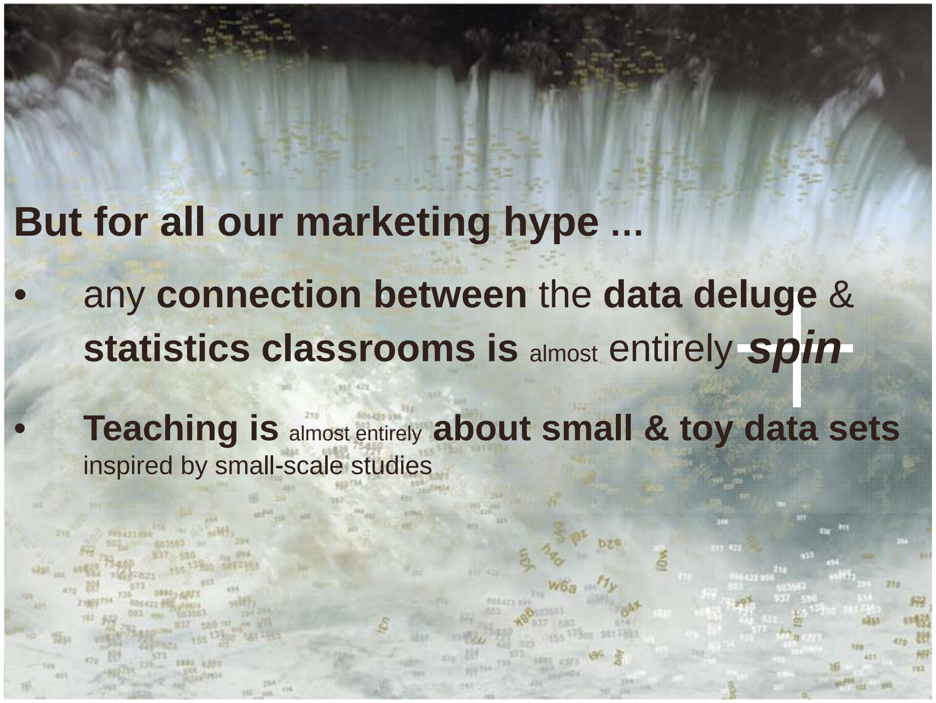


Data

trickle



The Data Deluge



But for all our marketing hype …

- any **connection between** the **data deluge** & **statistics classrooms is** almost entirely ***spin***
- **Teaching is** almost entirely **about small & toy data sets** inspired by small-scale studies

But the data world …

is getting a whole lot bigger

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

But the data world …
is getting a whole lot bigger

- There is an explosion in the …
  - quantities of data being collected
  - conceptions of what constitutes data
  - settings in which it can arise
  - ways of looking at it

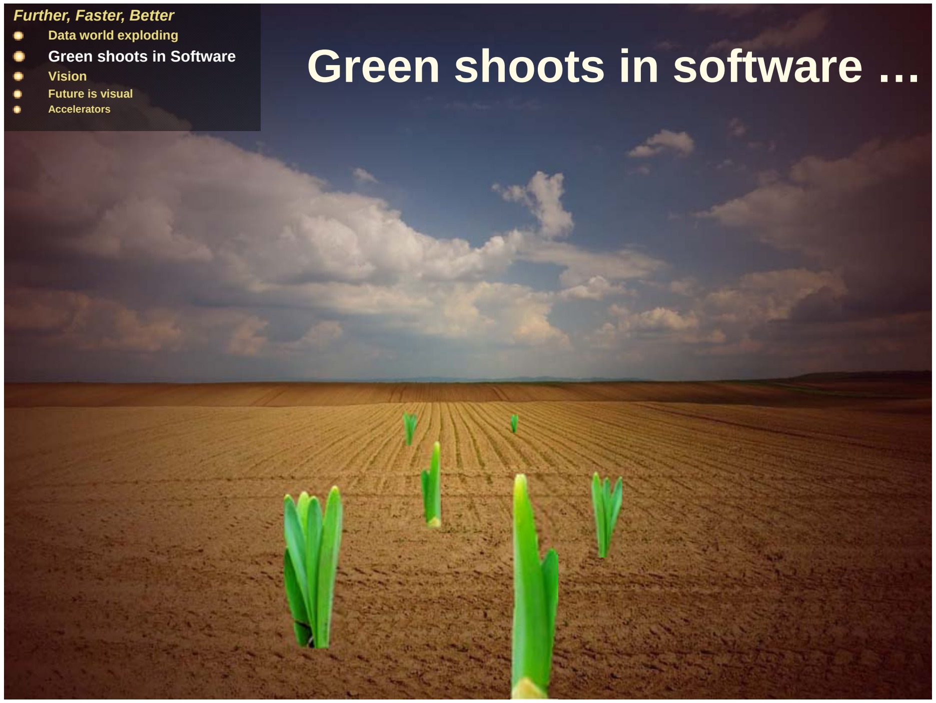
Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

But the data world …
is getting a whole lot bigger

Can't just keep illuminating same small patch

- Need to get much …
  - ***further***
  - ***faster***
  - & with ***better*** *comprehension*



Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

Green shoots in software …

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

Green shoots in software …

- Hans Rossling
- Visualisation generally
  - Tableau

+ a b | e a u
Products
Solutions
Learning
Support
About
Buy
Home · Learning · Whitepapers · Analytics At The Speed Of Thought
hought

ased

ht

"Visualisation is a gateway drug to statistics"

"People who look at visualisations will start asking statistically important questions … even without knowing the jargon."

-- Martin Wattenberg

in interview for New Scientist by Peter Aldhous,5 February 2011, p44.

- AdviseStat

Further, Faster, Better

- Data world exploding
- **Green shoots in Software**
- Vision
- Future is visual
- Accelerators

Leland Wilkinson

AdviseStat, First in the Next Generation of Analytics

AdviseStat is your statistics expert-in-a-box. It does all of the heavy statistical thinking, and you get to take all of the credit!

Simple to Work With

Tell AdviseStat what you want to see, and it will take care of everything else. AdviseStat uses a unique button-tree interface that allows you to command it with a simple sentence:

- "Predict height based on weight."
- "Compare salaries when grouped by gender."
- "Forecast sales."

That's it. AdviseStat will ask a few yes-or-no follow-up questions, and then run the entire analysis itself.

Does the Thinking For You

AdviseStat is the first artificially intelligent statistical assistant. Developed by a statistician, it makes its own observations about the data it sees and then chooses the most appropriate methods. All the transformations are done for you in the background. It makes corrections, like imputing missing data, all on its own. It is the closest thing to having a statistician there with you.

Guides You, and Explains Your Results

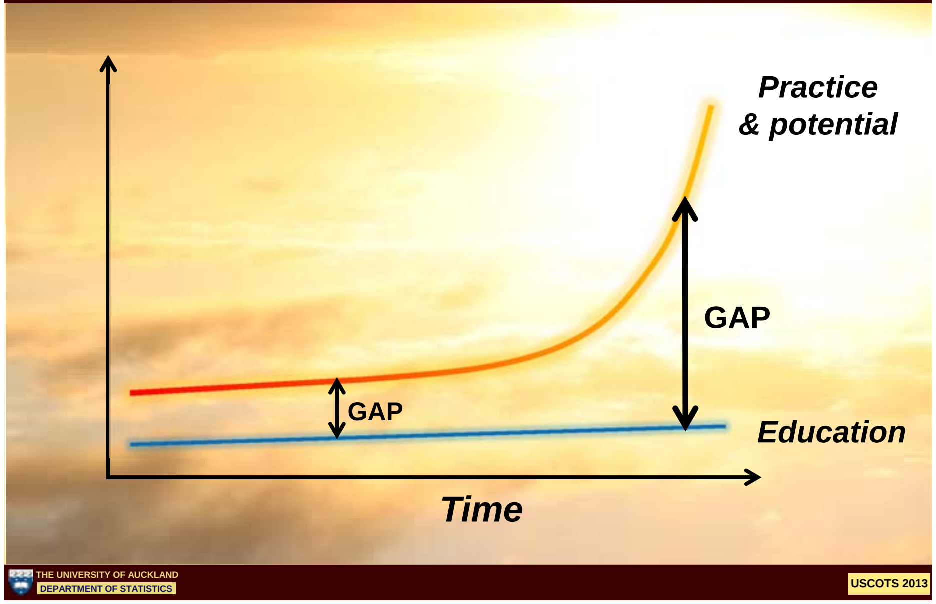
When the analysis is complete, most statistical software will spit out a couple coefficients and graphs, without any explanation or context. AdviseStat gives you a full whitepaper with text that's been customized to your specific findings. It starts by explaining about the type of analysis it chose, then talks about the corrections it made and the significant results it found.

Raises Red Flags

The great thing about an expert statistics program is that it can keep you out of pitfalls you might not even have known were there.

AdviseStat catches subtle statistical mistakes and other lurking factors that even expert statisticians would have a hard time finding, such as:

A Growing Gap



Opening Up the World of data

Need to open it up *wider, faster*

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- **Vision**
- Future is visual
- Accelerators

Uncovering the stories in a sea of data

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

I wonder what's the student experience ?

MEANWHILE BACK AT THE RANCH ...

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

USCOTS 2013

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

Student experience ??

Hate ...

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

Student experience ??

Fear ...

Further, Faster, Better

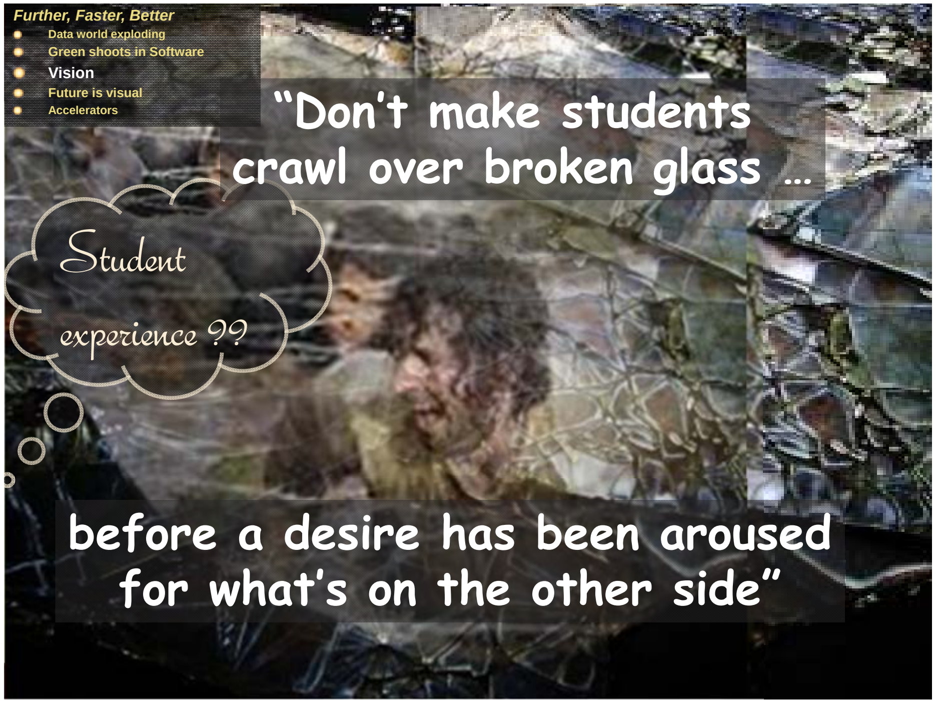
- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

Uncovering the stories
in a sea of data

Student experience ??

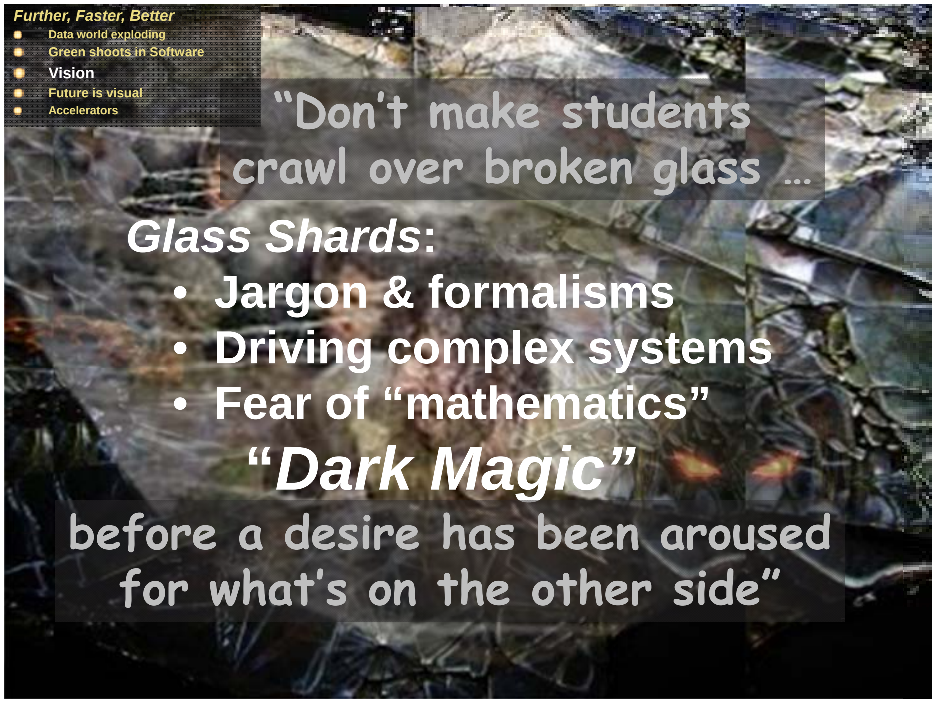
Love ...

Data analysis



"Don't make students crawl over broken glass …

before a desire has been aroused for what's on the other side"

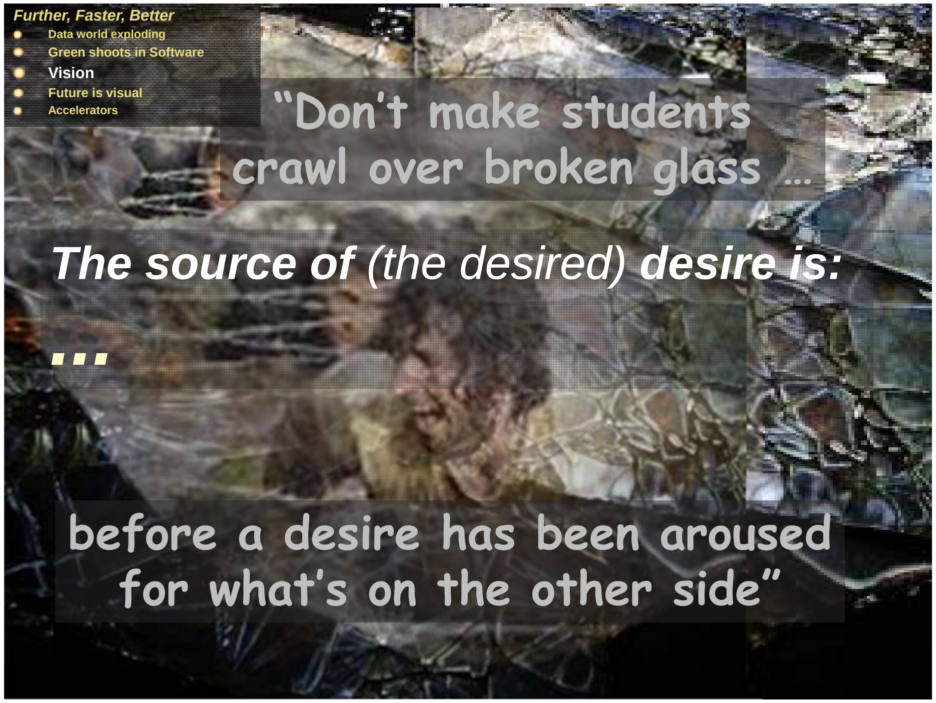


"Don't make students crawl over broken glass …

***Glass Shards*:**
- **Jargon & formalisms**
- **Driving complex systems**
- **Fear of "mathematics"**

"Dark Magic"

before a desire has been aroused for what's on the other side"

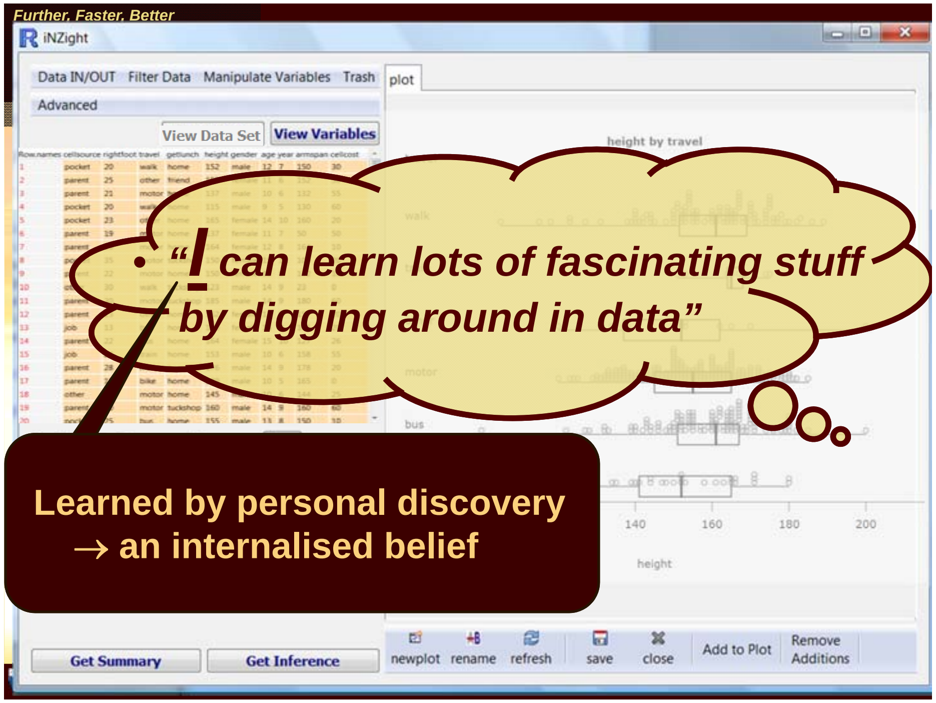


"Don't make students crawl over broken glass …

The source of** (the desired) **desire is:

…

before a desire has been aroused for what's on the other side"



Row.names	cellsource	rightfoot	travel	getlunch	height	gender	age	year	armspan	cellcost
1	pocket	20	walk	home	152	male	12	7	150	30
2	parent	25	other	friend	[illegible]	female	11	6	152	[illegible]
3	parent	21	motor	[illegible]	137	male	10	6	132	55
4	pocket	20	walk	home	115	male	9	5	130	60
5	pocket	23	other	home	165	female	14	10	160	20
6	parent	19	motor	home	137	female	11	7	50	50
7	parent	[illegible]	motor	home	164	female	12	8	[illegible]	10
8	[illegible]	35	motor	tuckshop	150	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	[illegible]	22	motor	home	150	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	other	20	walk	tuckshop	123	male	14	9	23	0
11	parent	[illegible]	motor	tuckshop	185	male	14	9	180	60
12	parent	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	job	13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14	parent	22	[illegible]	home	164	female	15	10	127	26
15	job	[illegible]	train	home	153	male	10	6	158	55
16	parent	28	[illegible]	[illegible]	[illegible]	male	14	9	178	20
17	parent	[illegible]	bike	home	[illegible]	male	10	5	165	0
18	other	[illegible]	motor	home	145	[illegible]	[illegible]	6	144	25
19	parent	[illegible]	motor	tuckshop	160	male	14	9	160	60
20	[illegible]	25	bus	home	155	male	13	8	150	10

- ***"I can learn lots of fascinating stuff by digging around in data"***

Learned by personal discovery → an internalised belief

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- **Future is visual**
- Accelerators

"the future of improved statistical understanding is *visual*"

"The Eyes have it"

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- **Future is visual**
- Accelerators

"the future of improved statistical understanding is *visual*"

improved **understanding of …**

data

inferential concepts

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- **Vision**
- Future is visual
- Accelerators

VISION Statement for Early Statistics

- To ***create excitement*** about
  - **"What I can do with data &**
  - **What data can do for me"**

Further, Faster, Better

- Data world exploding
- Green shoots in Software
- **Vision**
- Future is visual
- Accelerators

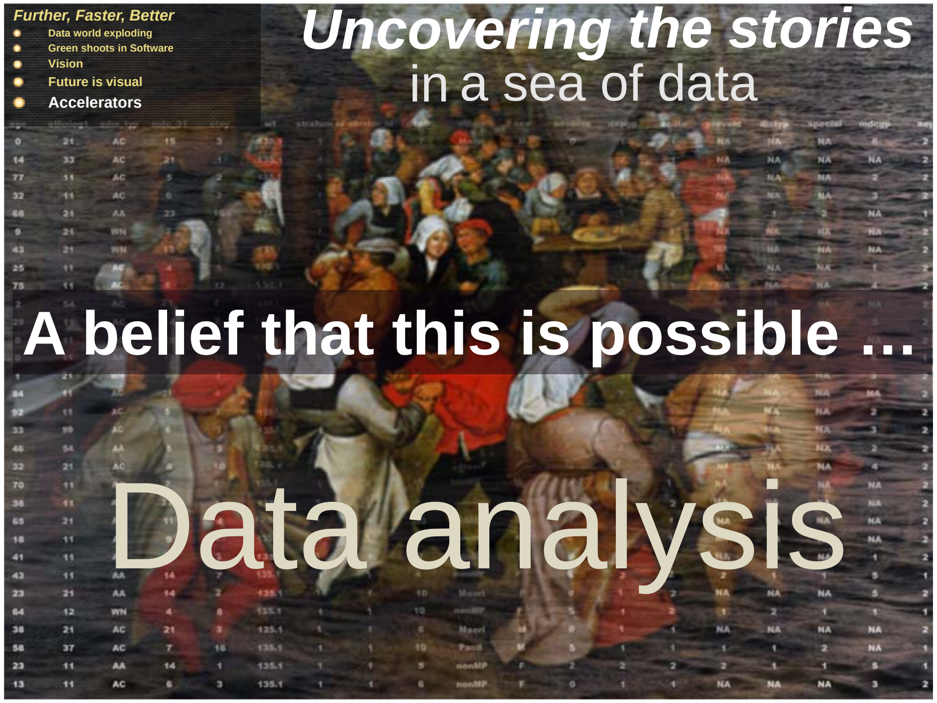
Some things

- **Some things change**
  - Exploding world of data
  - Need to convey more of this more quickly
- **Some things stay the same**
  - Available time
  - Inability to hold more than 4-7 ideas in working memory
- So **something's got to give** !!
  - Details of how we consti ings
    - Segmentation th behind them are

And software (& simulation & visualisation) **provides the key**

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

USCOTS 2013



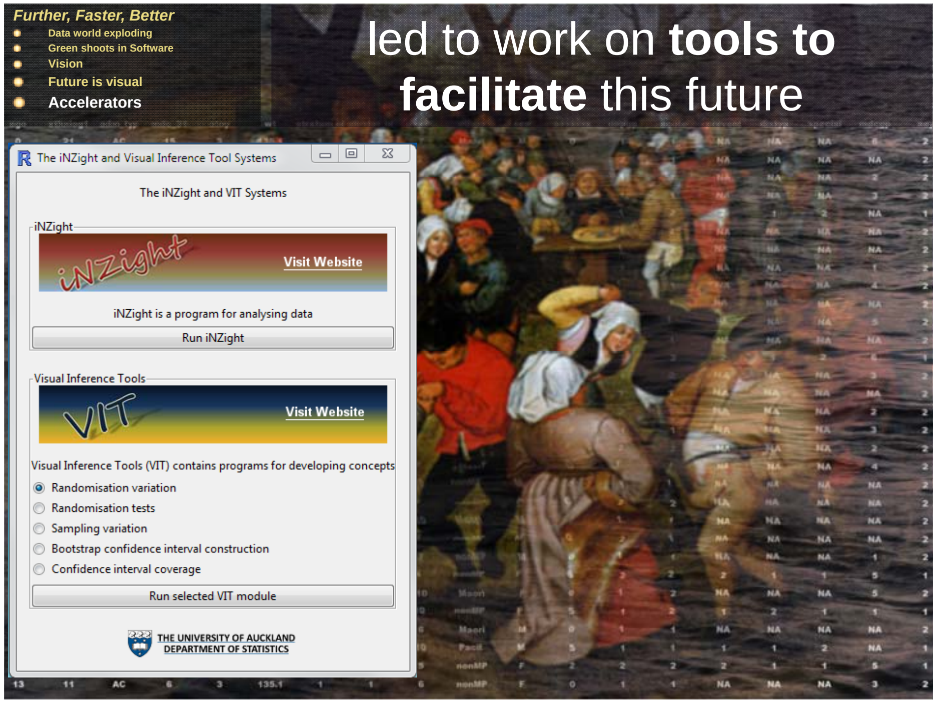
Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

Uncovering the stories
in a sea of data

A belief that this is possible …

Data analysis



Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

led to work on **tools to facilitate** this future

The iNZight and Visual Inference Tool Systems

The iNZight and VIT Systems

iNZight

iNZight is a program for analysing data

Run iNZight

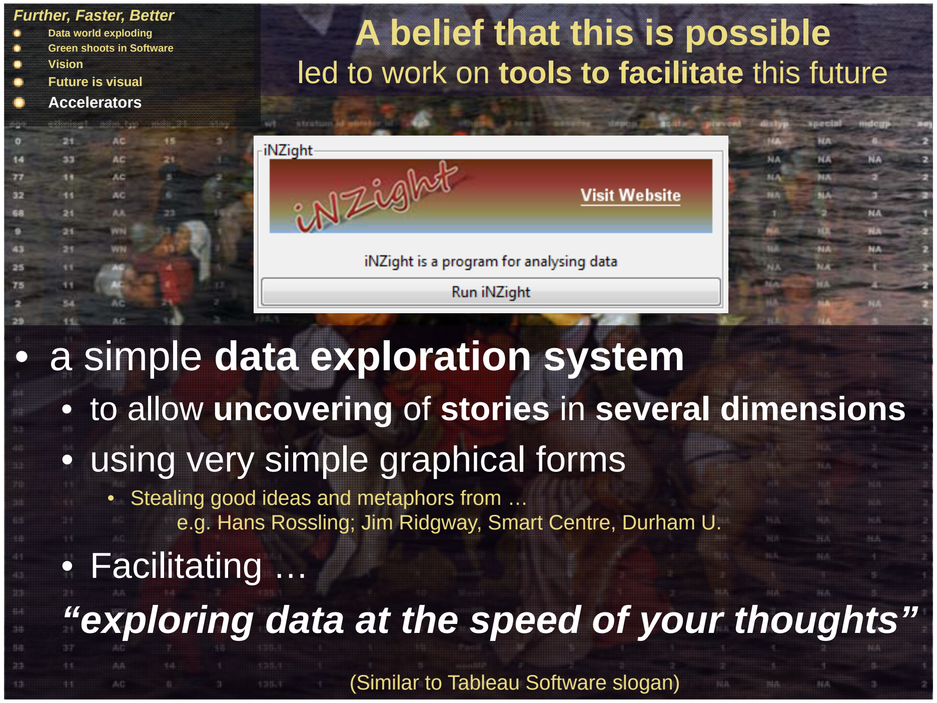
Visual Inference Tools

Visual Inference Tools (VIT) contains programs for developing concepts

- ● Randomisation variation
- ○ Randomisation tests
- ○ Sampling variation
- ○ Bootstrap confidence interval construction
- ○ Confidence interval coverage

Run selected VIT module

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS



Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

A belief that this is possible
led to work on **tools to facilitate** this future

iNZight

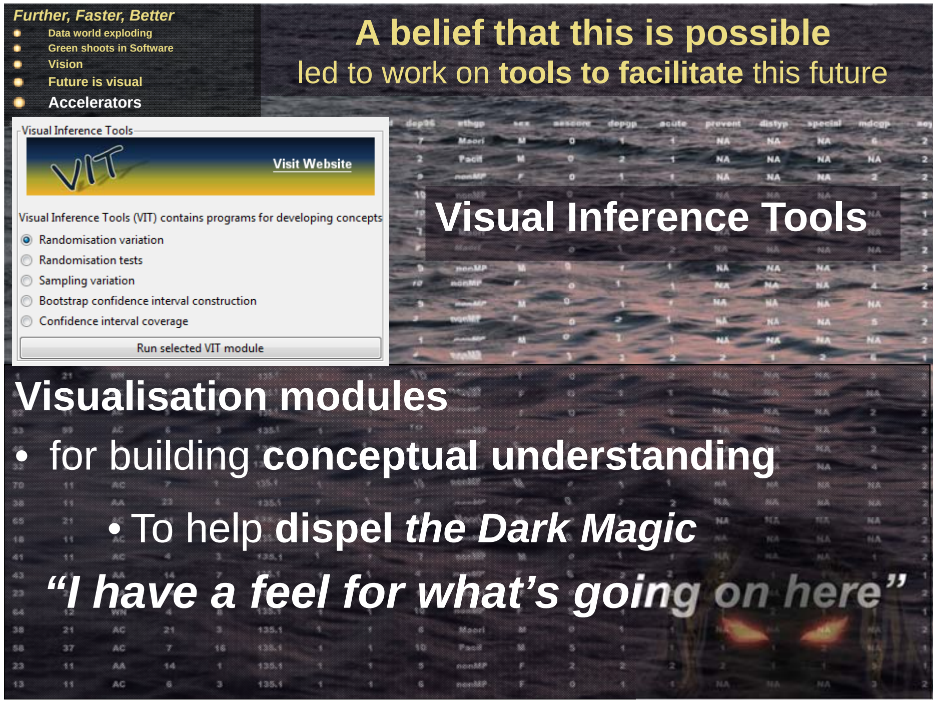
iNZight is a program for analysing data

Run iNZight

- a simple **data exploration system**
  - to allow **uncovering** of **stories** in **several dimensions**
  - using very simple graphical forms
    - Stealing good ideas and metaphors from …
      e.g. Hans Rossling; Jim Ridgway, Smart Centre, Durham U.
  - Facilitating …

"exploring data at the speed of your thoughts"

(Similar to Tableau Software slogan)



Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators

A belief that this is possible
led to work on **tools to facilitate** this future

Visual Inference Tools

Visual Inference Tools (VIT) contains programs for developing concepts

- ● Randomisation variation
- ○ Randomisation tests
- ○ Sampling variation
- ○ Bootstrap confidence interval construction
- ○ Confidence interval coverage

Run selected VIT module

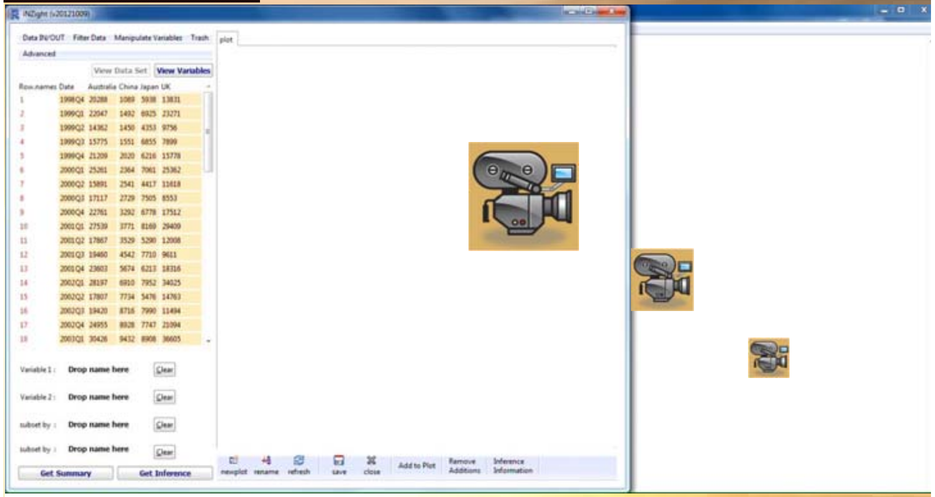
Visual Inference Tools

Visualisation modules

- for building **conceptual understanding**
  - To help **dispel *the Dark Magic***

"I have a feel for what's going on here"

Time Series



iNZight (v20121009)

Row.names	Date	Australia	China	Japan	UK
1	1998Q4	20288	1089	5938	13831
2	1999Q1	22047	1492	6925	23271
3	1999Q2	14362	1450	4353	9756
4	1999Q3	15775	1551	6855	7899
5	1999Q4	21209	2020	6216	15778
6	2000Q1	25261	2364	7061	25362
7	2000Q2	15891	2541	4417	11618
8	2000Q3	17117	2729	7505	8553
9	2000Q4	22761	3292	6778	17512
10	2001Q1	27539	3771	8169	29409
11	2001Q2	17867	3529	5290	12008
12	2001Q3	19460	4542	7710	9611
13	2001Q4	23603	5674	6213	18316
14	2002Q1	28197	6910	7952	34025
15	2002Q2	17807	7734	5476	14763
16	2002Q3	19420	8716	7990	11494
17	2002Q4	24955	8828	7747	21094
18	2003Q1	30428	9432	8908	36605

Variable 1 : Drop name here

Variable 2 : Drop name here

subset by : Drop name here

subset by : Drop name here

Shift of focus

From:

- *"What are all the things I have to do to get the output I'm meant to produce?"*

To:

- *"What are the questions?"*
- *"What can I* ***see****?"*
- *"What does that tell me?"*

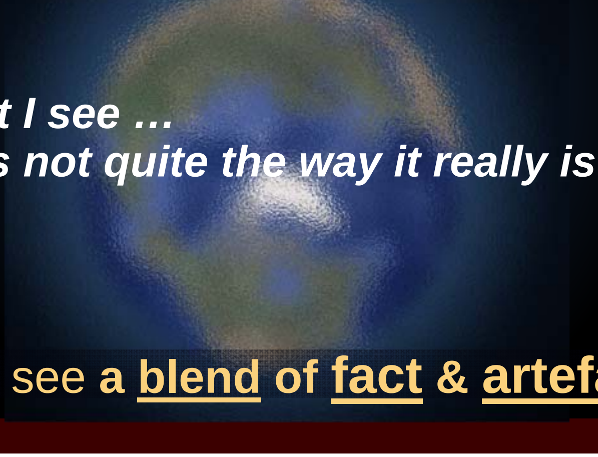
Using ***"enabling software"*** to **"cut out the middle man"**

But …

- **with arousal to the pleasures of discovery …**

- **need fundamental statistical "safe sex" messages**

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"

We see a **<u>blend</u> of <u>fact</u> & <u>artefact</u>**

There are stories …
& there are “stories”

Data analysis

Stories vs “stories”

Fact

versus

Artefact

America's Silent Majority By Joe Klein
TIME
INSIDE TINTIN
Steven Spielberg transforms the classic comic book into an animated adventure
By Lev Grossman

Stories vs “stories”

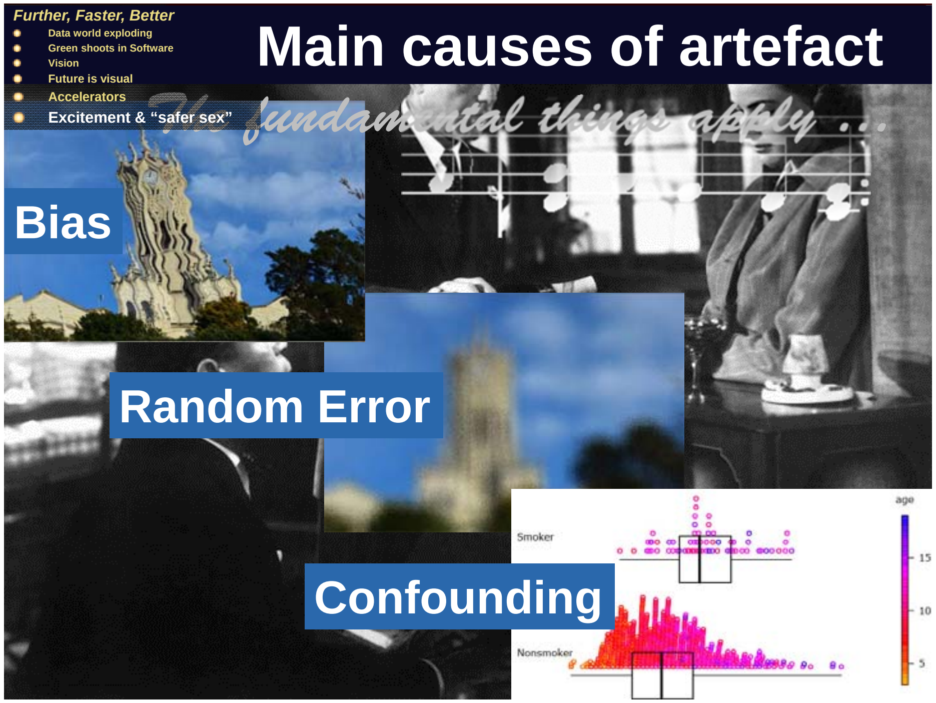
Fact

versus

Artefact

not always an improvement!

America's Silent Majority By Joe Klein
TIME
INSIDE TINTIN
Steven Spielberg transforms the classic comic book into an animated adventure
By Lev Grossman



Further, Faster, Better

- Data world exploding
- Green shoots in Software
- Vision
- Future is visual
- Accelerators
- Excitement & “safer sex”

Main causes of artefact

The fundamental things apply …

Bias

Random Error

Confounding

We see a blend of fact & artefact

- **Statistical study design**
  - employing random sampling or randomised experimentation (in conjunction with other tricks of the trade)

  used to minimize artifact

- **Inference based on randomness theory**
  - most ***obviously relevant*** **and valid** where (all of) the randomness is introduced by the study design
  - other uses are ***based on models***

    that ***assume random mechanisms*** are at work somewhere in the process that generated the data

Hugely more complex & difficult thinking

Start with randomness by design

(The basic concepts can transfer to modelling contexts later)

- Confidence interval ideas
  - arise most directly and simply in sampling contexts
- Significar
  - arise mo settings

> ***“The traditional road to statistical knowledge is blocked, for most, by a formidable wall of mathematics.”***
>
> – Efron & Tibshirani (1993)

- In both cases
  - can motivate and convey all of the essential ingredients ***entirely visually*** (without a formula in sight)

Start with randomness by design

(The basic concepts can transfer to modelling contexts later)

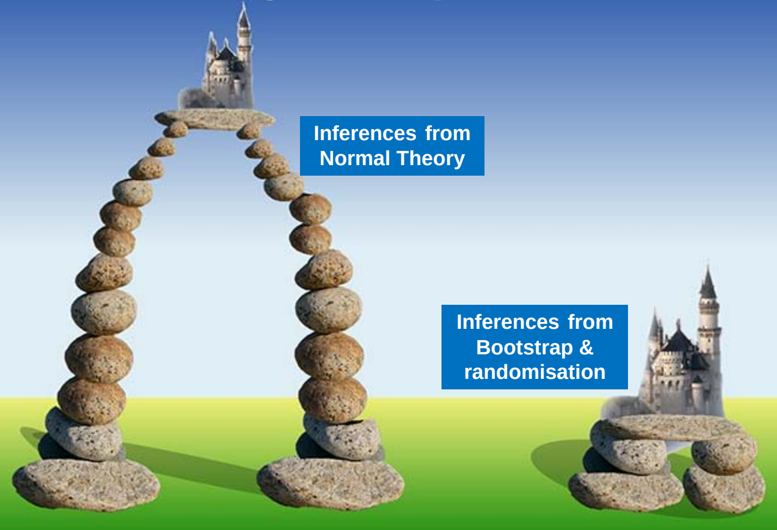
- C
  - xts
- Si
  - nment settings (experiments)

> ***Approach***
> - ***Get basic ideas in place first***
>
>   ***- intuitively, visually***
> - ***Can mathematize later***

- In both cases
  - can motivate and convey all of the essential ingredients ***entirely visually*** (without a formula in sight)

Building conceptual Castle

Building conceptual Castle



Inferences from Normal Theory

Inferences from Bootstrap & randomisation

Principle

The inferential method should mirror the

- I'm going to lead down some sets of conceptual sequences
- It's for you to judge how well they hang together

1. this principle

2. themselves to visual treatments

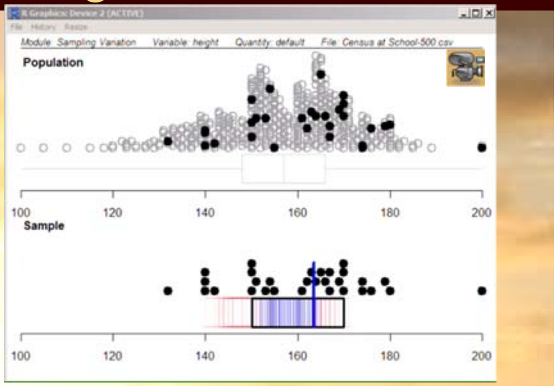
3. Connect better to intuition

Will spend rest of the talk exploring and demonstrating this assertion

Statistical inference & "sampling variation"

Original version

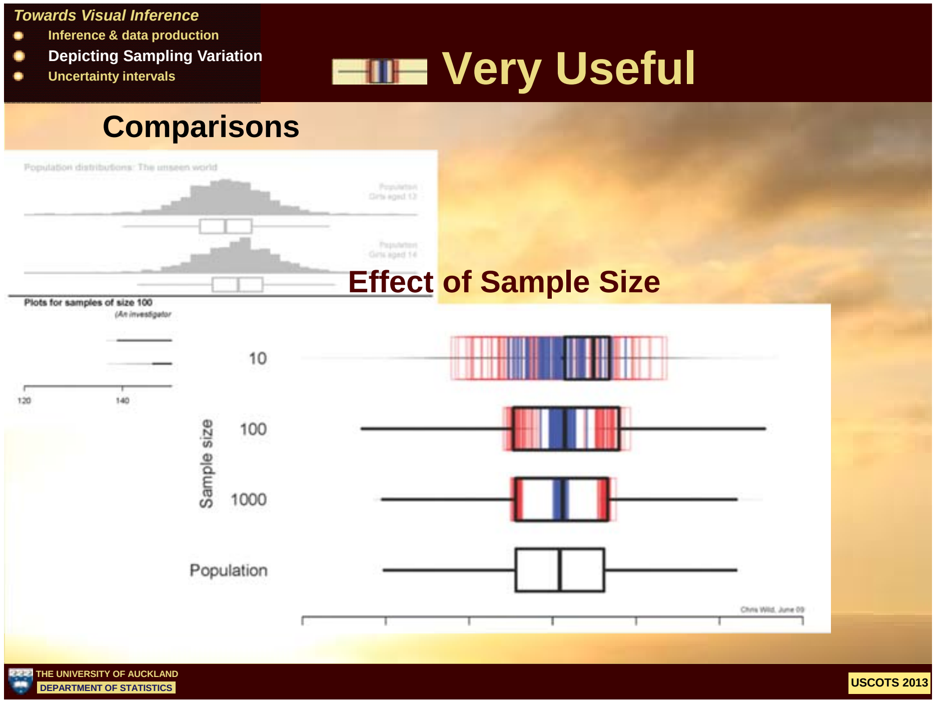
Use ***boxplots***
Track
- medians —
- quartiles —



R Graphics: Device 2 (ACTIVE)

Module: Sampling Variation Variable: height Quantity: default File: Census at School-500.csv

Makes …



Towards Visual Inference

- Inference & data production
- Depicting Sampling Variation
- Uncertainty intervals

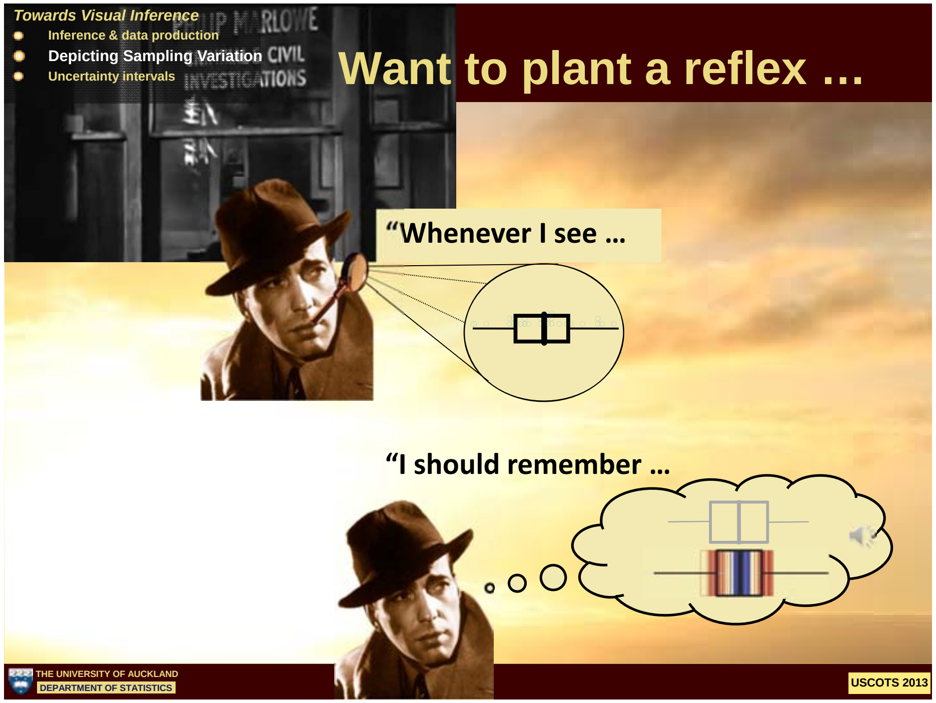
Very Useful

Comparisons

Effect of Sample Size

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

USCOTS 2013



Towards Visual Inference

- Inference & data production
- Depicting Sampling Variation
- Uncertainty intervals

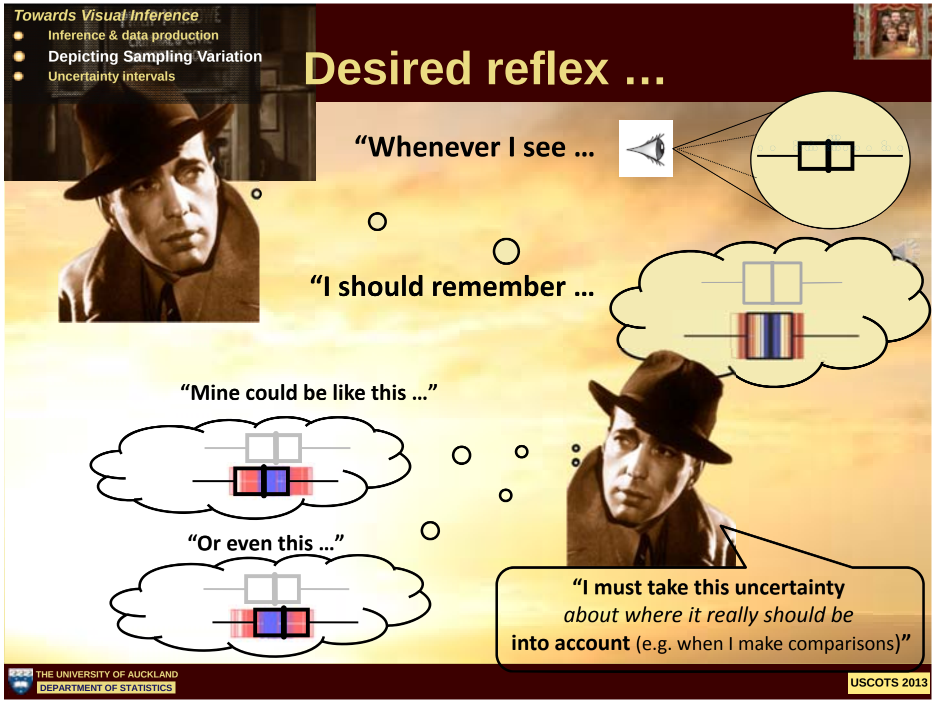
Want to plant a reflex …

"Whenever I see …

"I should remember …

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

USCOTS 2013



Towards Visual Inference

- Inference & data production
- Depicting Sampling Variation
- Uncertainty intervals

Desired reflex …

"Whenever I see …

"I should remember …

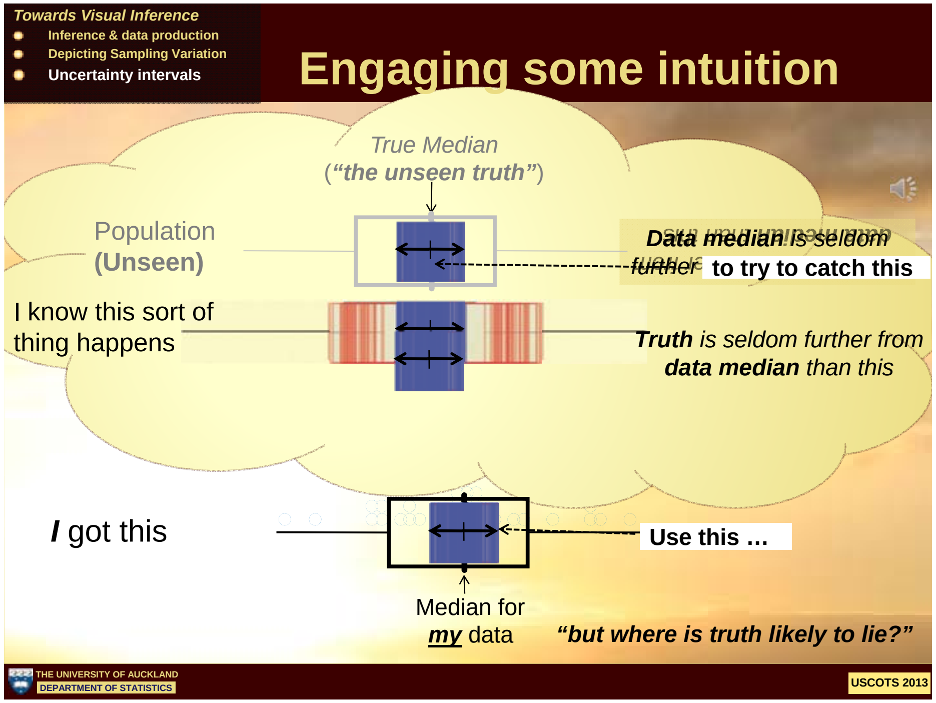
"Mine could be like this …"

"Or even this …"

"**I must take this uncertainty**
about where it really should be
into account (e.g. when I make comparisons)"

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

USCOTS 2013



Towards Visual Inference

- Inference & data production
- Depicting Sampling Variation
- Uncertainty intervals

Engaging some intuition

True Median
(*"the unseen truth"*)

Population
(Unseen)

***Data median** is seldom*
further **to try to catch this**

I know this sort of
thing happens

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

I got this

Use this …

Median for
my data

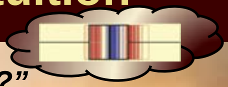
"but where is truth likely to lie?"

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

USCOTS 2013

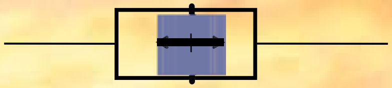
Towards Visual Inference
- Inference & data production
- Depicting Sampling Variation
- Uncertainty intervals

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 ***from the single sample itself***

How ????

Seems impossible!!

Enter Brad Efron & the $_{\wedge}^{\text{simple}}$ Bootstrap

Efron (1979)

(1996 photo)

I wonder if
*"sampling with replacement
from the sample"*
will **mimic** the process of
"sampling from the population"

Bootstrap

- **Bootstrap Sampling with replacement**
  - **What is it?**
  - **What does it do?**
  - **How could we use it?**
  - **Why might it work?**
- **Constructing Bootstrap intervals**
- **Does it work?**

Bootstrap
- Sampling with replacement
  - What is it?
  - What does it do?
  - How could we use it?
  - Why might it work?
- Constructing Bootstrap intervals
- Does it work?

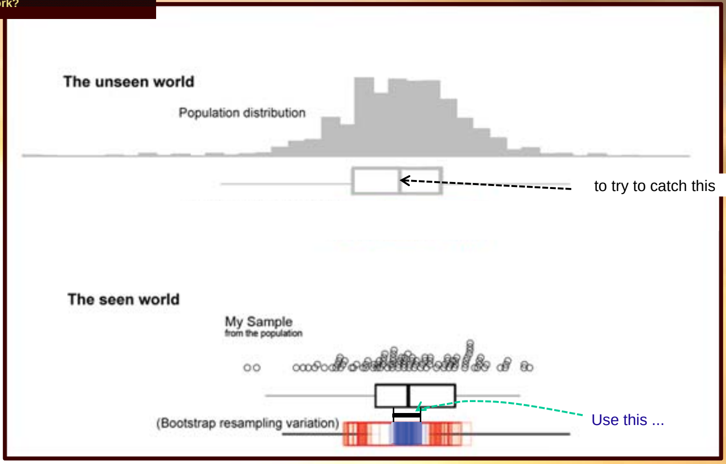
Bootstrap

"Re-sampling with replacement"

- ***What is it?***
- ***What does it do?***

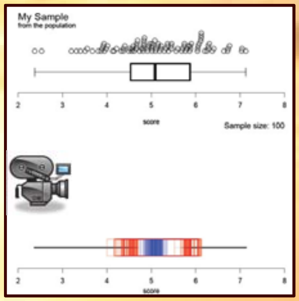
Play

How could we use it?



Bootstrap

How could we use it?

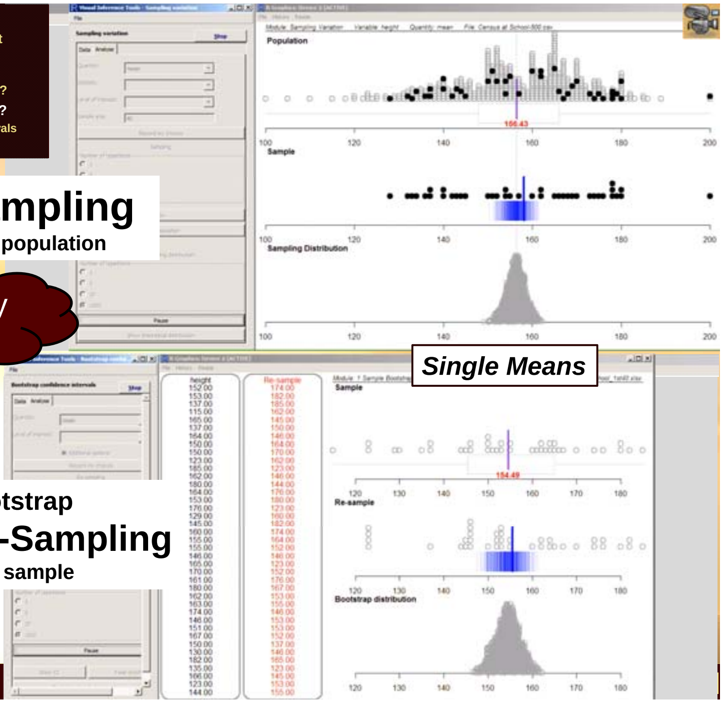


Bootstrap

“Re-sampling with replacement”

- ***Why might it work?***

Plausibility



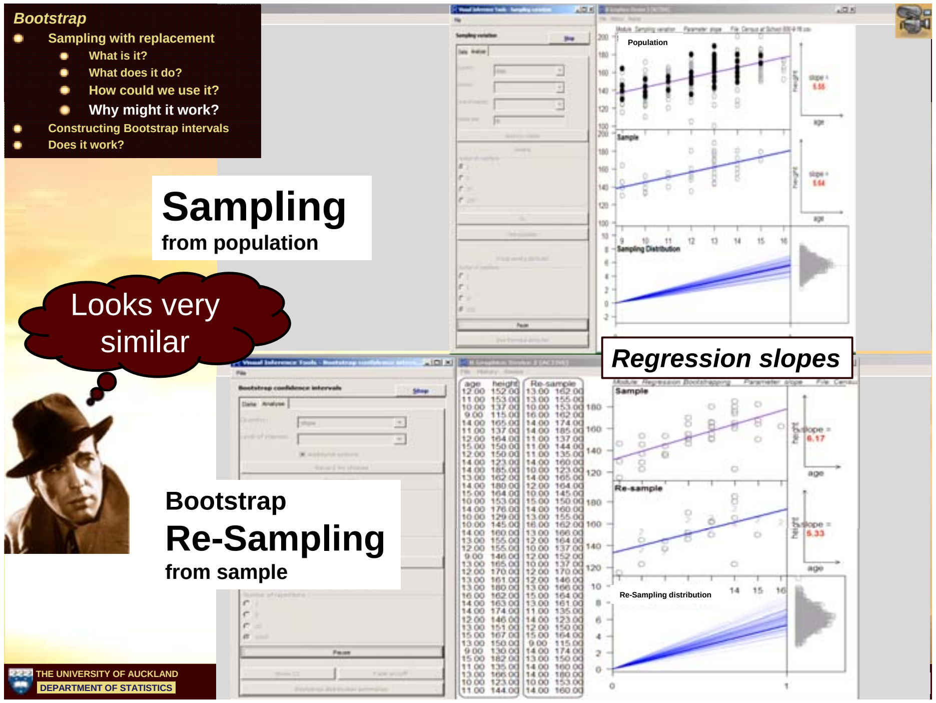
Sampling
from population

Looks very similar

Single Means

Bootstrap
Re-Sampling
from sample

height	Re-sample
152.00	174.00
153.00	182.00
137.00	185.00
115.00	162.00
165.00	145.00
137.00	150.00
164.00	146.00
150.00	164.00
150.00	170.00
123.00	162.00
185.00	123.00
162.00	146.00
180.00	144.00
164.00	170.00
153.00	180.00
176.00	123.00
129.00	160.00
145.00	182.00
160.00	174.00
155.00	164.00
155.00	152.00
146.00	146.00
165.00	123.00
170.00	152.00
161.00	176.00
180.00	167.00
162.00	153.00
163.00	155.00
174.00	146.00
146.00	153.00
151.00	153.00
167.00	152.00
150.00	137.00
130.00	146.00
182.00	165.00
135.00	123.00
166.00	145.00
123.00	153.00
144.00	155.00



Bootstrap

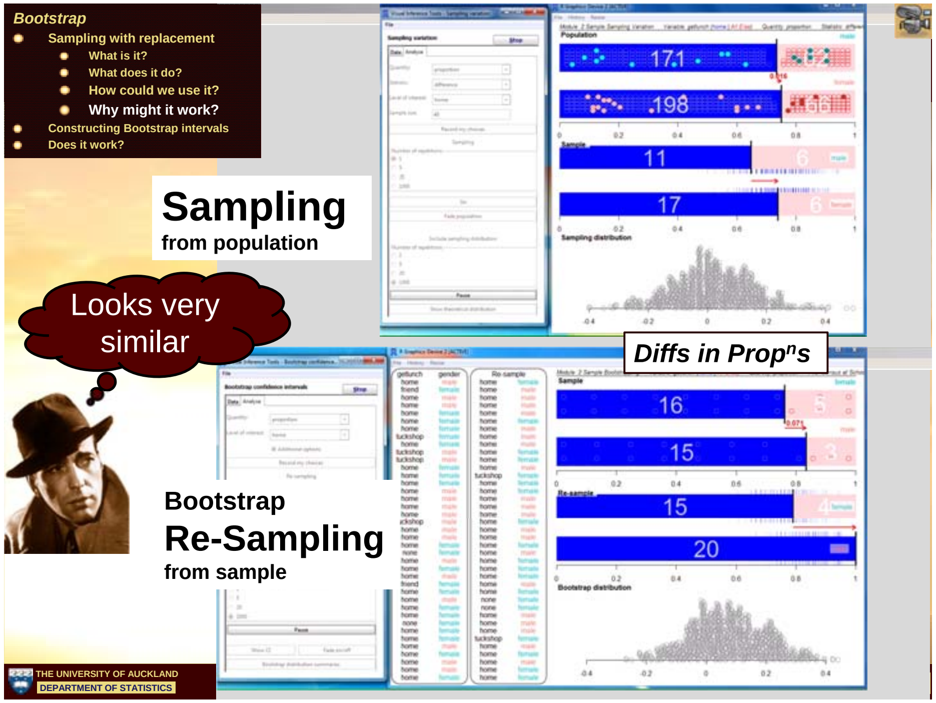
- Sampling with replacement
  - What is it?
  - What does it do?
  - How could we use it?
  - **Why might it work?**
- Constructing Bootstrap intervals
- Does it work?

Sampling
from population

Looks very similar

Bootstrap
Re-Sampling
from sample

Regression slopes



Bootstrap

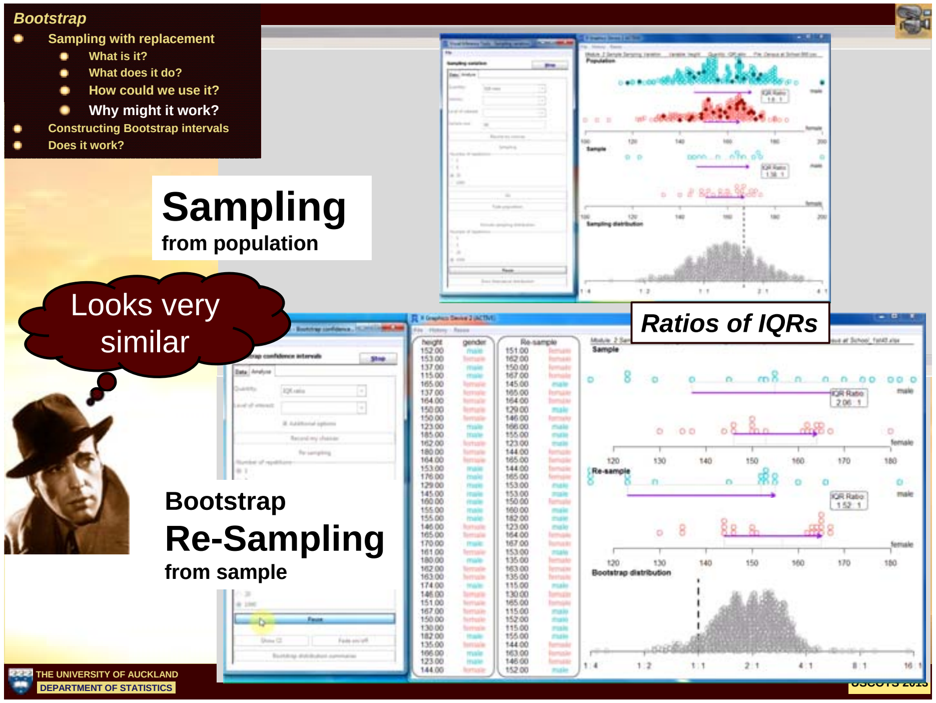
- Sampling with replacement
  - What is it?
  - What does it do?
  - How could we use it?
  - **Why might it work?**
- Constructing Bootstrap intervals
- Does it work?

Sampling
from population

Looks very similar

Bootstrap
Re-Sampling
from sample

Diffs in Propⁿs



Bootstrap

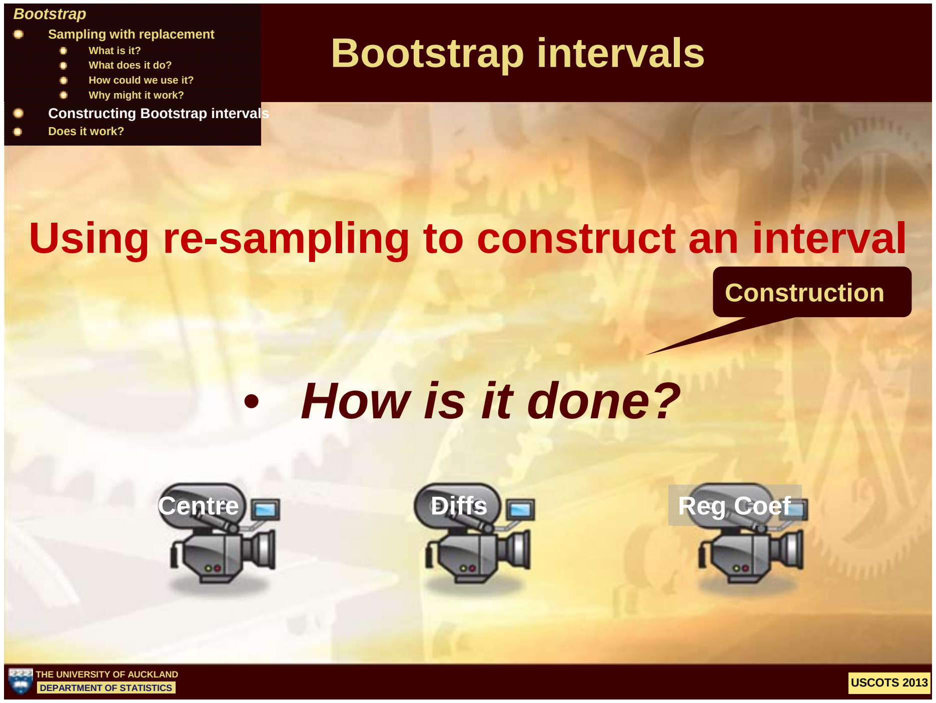
- Sampling with replacement
  - What is it?
  - What does it do?
  - How could we use it?
  - **Why might it work?**
- Constructing Bootstrap intervals
- Does it work?

Sampling
from population

Looks very similar

Bootstrap
Re-Sampling
from sample

Ratios of IQRs



Bootstrap

- Sampling with replacement
  - What is it?
  - What does it do?
  - How could we use it?
  - Why might it work?
- **Constructing Bootstrap intervals**
- Does it work?

Bootstrap intervals

Using re-sampling to construct an interval

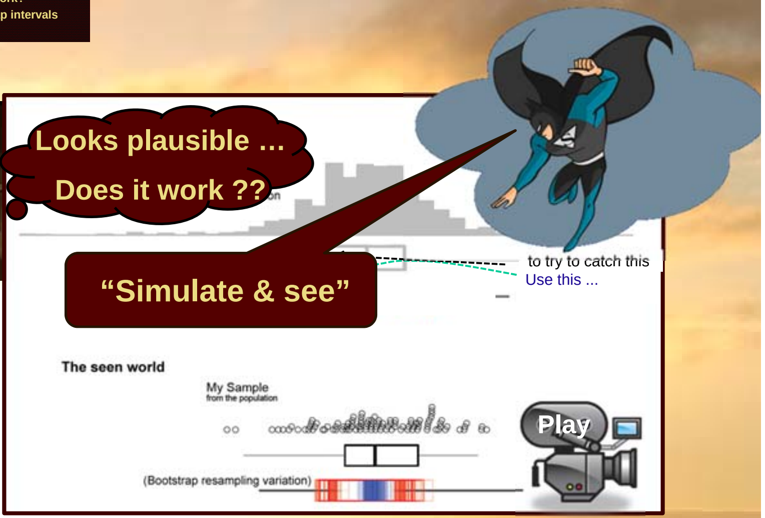
Construction

- ***How is it done?***

Bootstrap

- Sampling with replacement
  - What is it?
  - What does it do?
  - How could we use it?
  - Why might it work?
- Constructing Bootstrap intervals
- Does it work?

Bootstrap intervals

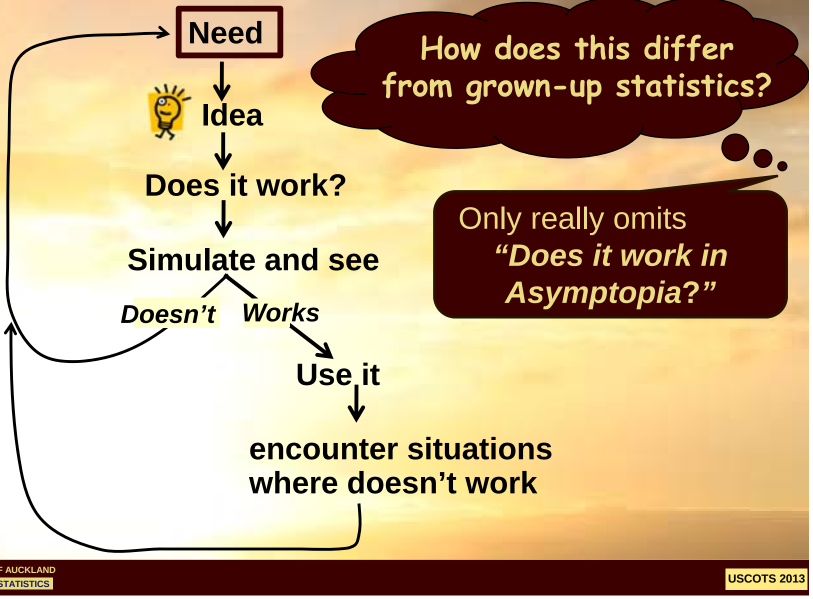


Looks plausible …

Does it work ??

"Simulate & see"

Embed in Discovery learning for Methodology



Need

Idea

Does it work?

Simulate and see

Doesn't *Works*

Use it

encounter situations where doesn't work

How does this differ from grown-up statistics?

Only really omits *"Does it work in Asymptopia?"*

Bootstrap – one basic, widely applicable idea

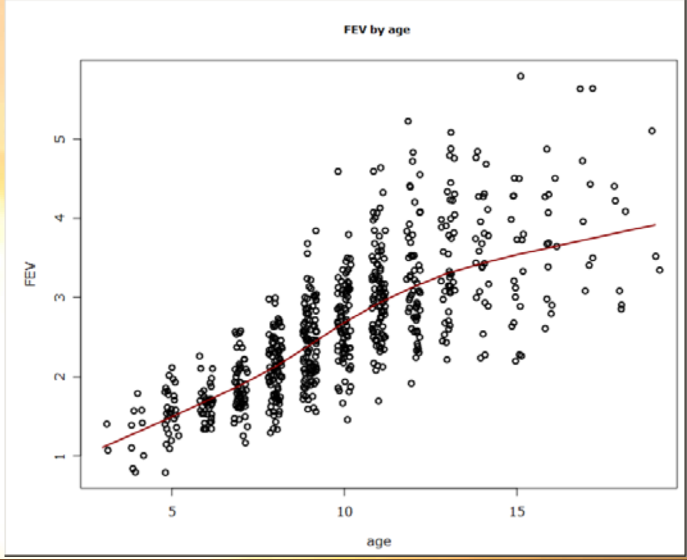
Not a huge step now …

- to accepting that
  - we can & should

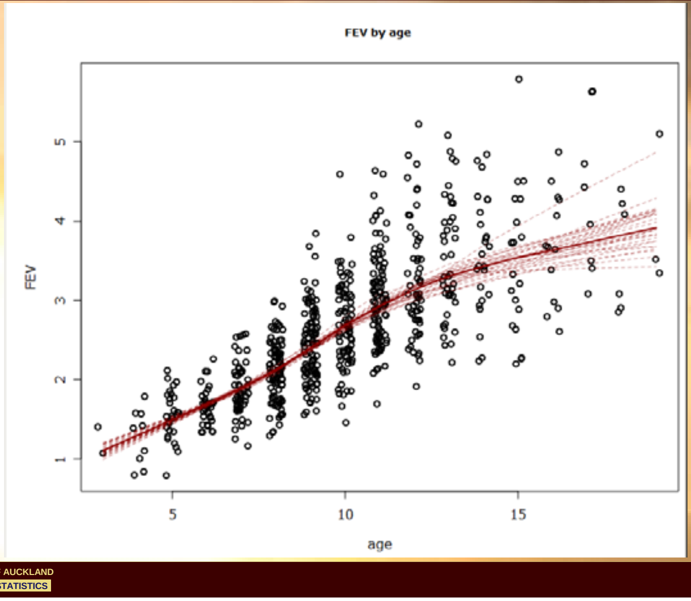
put uncertainty intervals around most everything

- And further …

More complex: Scatterplot with smoother



Bootstrap smooths added to convey uncertainty

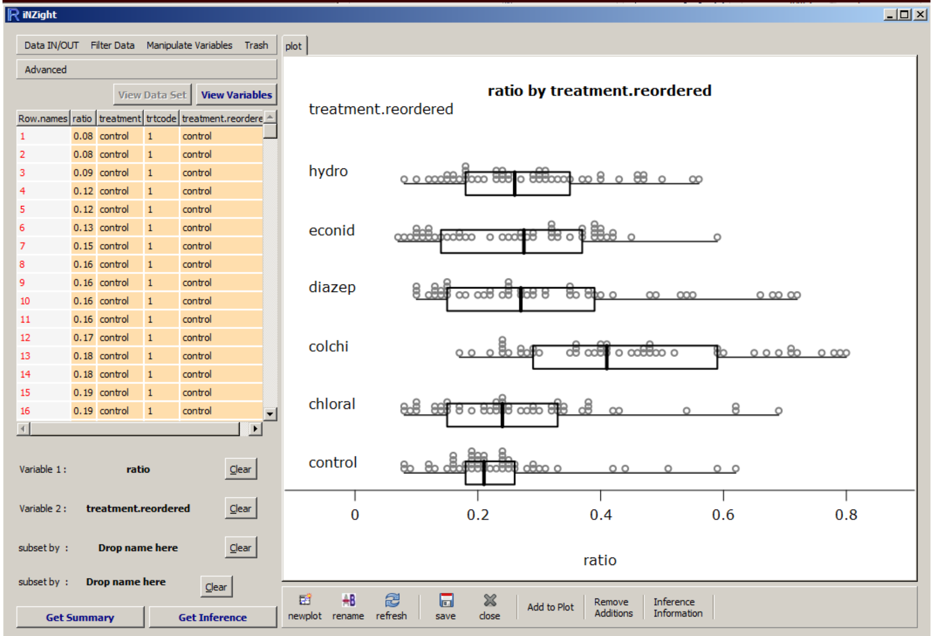


Randomised experiments and Randomisation tests

Cell-ratio data

[Source: Auckland's Cancer Research Unit]

- Experimental units are samples of human blood cells
  - Grown in cell culture and then treated with one of
    - *chloral hydrate, hydroquinone, diazepam, econidazole, and colchicine*
      - some of which are known to be potent carcinogens.
      - The carcinogens act by breaking chromosomes, and thus disrupts cell division
      - Broken fragments of chromosome are left as micronuclei, and the average ratio of the size of a micronucleus to its parent cell nucleus is measured.
        - » The more carcinogenic the chemical, the higher the ratio tends to be
  - A random mechanism determined which samples of cells received which chemical treatment



Row.names	ratio	treatment	trtcode	treatment.reordere
1	0.08	control	1	control
2	0.08	control	1	control
3	0.09	control	1	control
4	0.12	control	1	control
5	0.12	control	1	control
6	0.13	control	1	control
7	0.15	control	1	control
8	0.16	control	1	control
9	0.16	control	1	control
10	0.16	control	1	control
11	0.16	control	1	control
12	0.17	control	1	control
13	0.18	control	1	control
14	0.18	control	1	control
15	0.19	control	1	control
16	0.19	control	1	control

Variable 1 : ratio Clear

Variable 2 : treatment.reordered Clear

subset by : Drop name here Clear

subset by : Drop name here Clear

Experiential Context

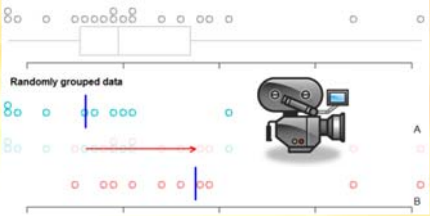
- Simple randomised experiment
  - e.g. drug vs control
- Follows experiences on “why do randomised experiments”
  - Following up message that randomisation is best way we know of balancing groups on unforeseen factors so that, apart from treatment, we are

    ***comparing like with like** (“fair test”)*

Experiential Context

1st new message:

- ***Randomisation is best, but not perfect …***



“*What I see isn’t exactly the way it is*”
Why?: *Randomisation variation*
Randomisation alone can make it look like there is a difference between groups
(the apparent differences result simply from who, by chance, ends up in what group)

We should only be impressed by experimental differences if they are larger than those produced by random labelling alone

Randomisation test visualisations

Can random labelling alone (“*chance alone*”) produce differences as large those as I’m seeing?

2 groups

Play

3 groups

Orig data

Proportions

Epilogue

- **Got a quite a long way with tiny number of ideas**
- Can broaden applications greatly without difficulty
  - so long as stick to …
    - intervals in contexts that emphasize sampling

      (Expand the complexity of sampling, range of quantities estimated)
    - Significance tests in contexts that emphasize random assignment

      (Expand the complexity of design, range of quantities estimated)
  - But when you want to do …
    - interval estimates of effect sizes in experimental contexts
    - significance testing in sampling contexts

    you suddenly have to wheel in many more concepts

Current Module Capabilities

Track/investigate/use behaviour of …		VIT Module Name			
		Motivation of need for method	*Inferential Method*	*Motivation of need for method*	*Inferential Method*
		Randomisation Variation	**Randomisation Tests**	**Sampling Variation**	**Bootstrap Conf. Ints**
		(*random assignment of group labels*)	(*random re-assignment of group labels*)	(*random sampling from Popn*)	(*Random re-sampling from sample*)
1-variable					
	Numeric			Wh	
	Categorical				
2-variables					
	Num | Cat (2 grps)	Diffs in: Means, Medians; 2-sample t-stats; IQR Ratio	Diffs in: Means, Medians; 2-sample t-stats; IQR Ratio	Whole boxplots Diffs in: Means, Medians; IQR Ratio	Diffs in: Means, Medians; IQR Ratio
	Num | Cat (k grps)	Av. Deviation, F "pseudo-F"	Av. Deviation "pseudo-F"		
	Cat | Cat (2 grps)	Diffs in: Proportions	Diffs in: Proport		
	Cat | Cat (k grps)	Av. Deviation, Chisq	Av. Deviation, Chisq		
	Num | Num	Reg Slope (default) Paired diffs (option)			

Columns differ in:
- **"parent data"** being operated on (popn/sample)
- **operation** performed (sampling/assigning grp labels)
- Motivation vs Method

Rows differ in:
- **Quantity/statistic** worked with

Principle:
- **maximise "same-ness" across & down so …**
  - "same-nesses" reinforced
  - visual differences relate to essential differences
    - to improve conceptual transfer

Epilogue

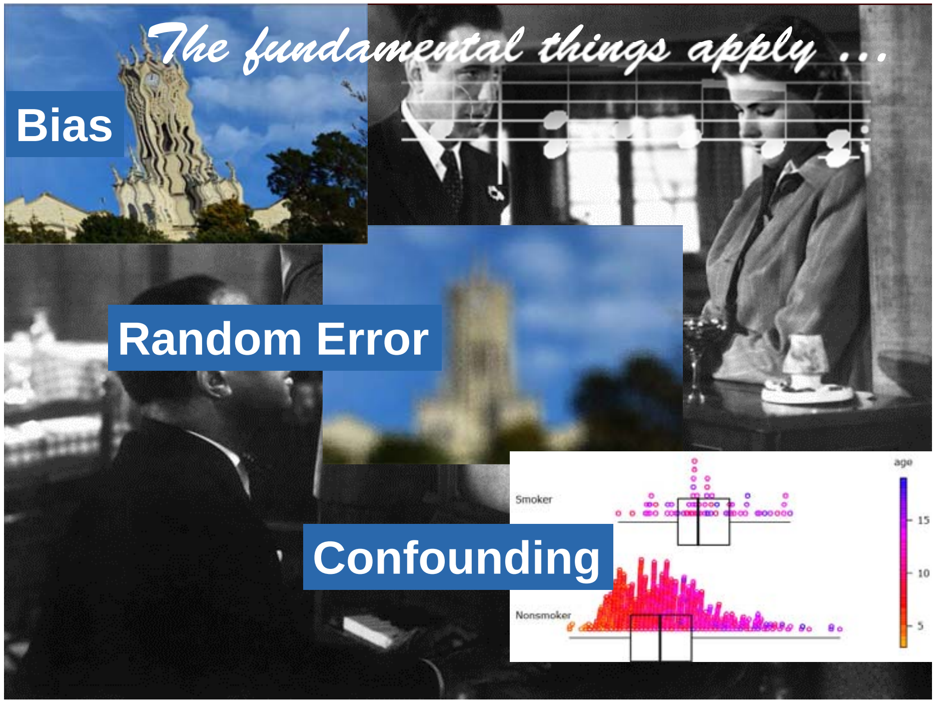
- **What we don't have in this development**
  - Beginner-killing abstractions
    - e.g. null hypotheses, parameters vs estimates, test statistics, formal distributions …
  - Dense clouds of details
  - Dependence on (poorly understood) mathematical ideas
- **What we do have**
  - "Concrete" ideas that make sense in the context
    - e.g. "Can random labelling do this?"
  - Fast access to a wide range of important applications
  - Substantial body of intuition and experience as a foundation to then build abstractions upon

Addressing the Need for Speed

- **iNZight-type software can facilitate a fast, accessible way in to understanding a much wider spectrum of data types**
- **VIT-type software can facilitate a fast , accessible way in to understanding basic inferential conceptions**

My "vision" is

- **initially create an appreciation of a very wide array of data types and what they can tell you**
  - **and only then back fill the details** (for those who need it)



Thank you