THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Technology, pedagogy & outreach

Chris Wild
The University of Auckland

Slides at **http://www.stat.auckland.ac.nz/~wild/08-talks/IASC.wild.pdf**

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Jet Propulsion Laboratory
California Institute of Technology

+ View the NASA Portal

Search JPL

JPL HOME | EARTH | SOLAR SYSTEM | STARS & GALAXIES | SCIENCE & TECHNOLOGY

Home | History | Images | Video & Audio | Downloads | Fast Facts | Events

Links:
http://www.computersciencelab.com/ComputerHistory/History.htm
http://www.jpl.nasa.gov/explorer/captions/computers.php
http://en.wikipedia.org/wiki/Human_computer
http://commons.wikimedia.org/wiki/Image:Human_computers_-_Dryden.jpg

JPL Computers

During the 1940s and 1950s, JPL used the word "computer" to refer to a person rather than a machine. The all-female computer team, many of the members recruited right out of high school, were responsible for doing all the math by hand required to plot satellite trajectories and more.

Image courtesy: NASA/JPL-Caltech

JPL photo number P-163

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Computers as

Human computer

From Wikipedia, the free encyclopedia

Before electronic computers became commercially available, the term "**computer**", in use from the mid 17th century, literally meant "one who computes": a person performing mathematical calculations. Teams of people were frequently used to undertake long and often tedious calculations; the work was divided so that this could be done in parallel.

Contents [hide]

1 Origins in astronomy
2 Wartime computing and the invention of electronic computing
3 Contemporary usage

NACA High Speed Flight Station "Computer Room"

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Business as usual or paradigm shift

- Many things we do with computers just speed up and make cheaper things we have always done
- Want to highlight some things closer to paradigm shifts
  - Where computers open up new ways of conceiving our world and how we “do education”

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

JAVA applets and their cousins

- Simulation-based java applets
  - Goals have been limited, typically targeting
    - an understanding of a few generic concepts such as sampling variation and the operating behaviour of confidence intervals
    - or the patterns that can be seen in data generated by a particular model
- Specialised calculators and analysis engines
  - performing computations, drawing graphs
  - showing how the behaviour of a model is governed by its parameters
- Additionally, we have internet-delivered instructional materials

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

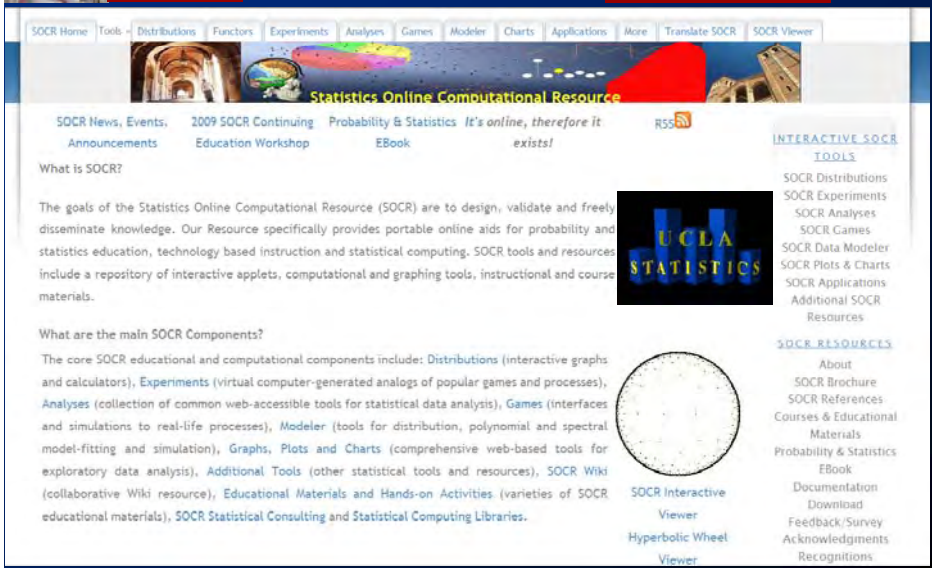
What has been happening at some major applet sites

VERSITY OF AUCKLAND
TMENT OF STATISTICS

http://www.socr.ucla.edu/

Ivo Dinov

Nicolas Christou

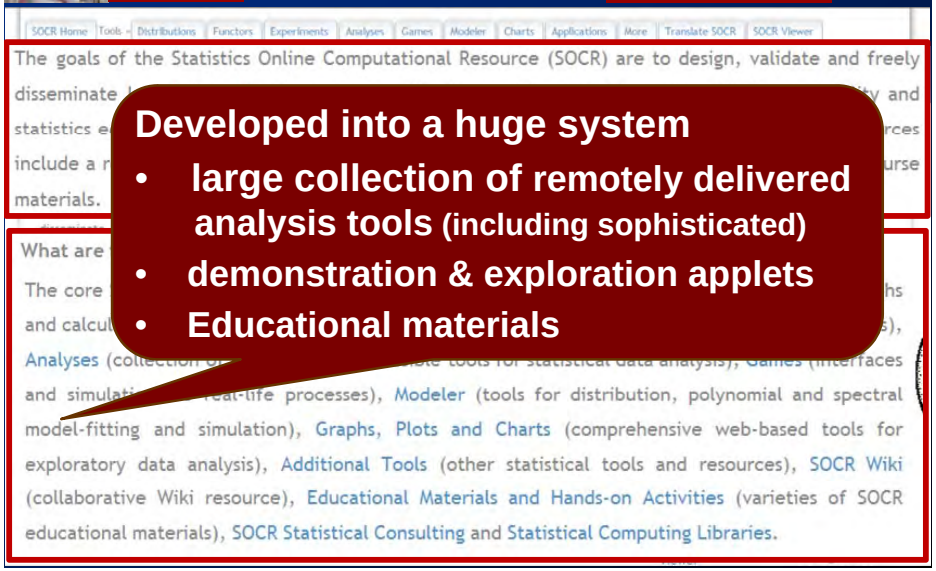


VERSITY OF AUCKLAND
TMENT OF STATISTICS

http://www.socr.ucla.edu/

Ivo Dinov

Nicolas Christou



Developed into a huge system

- **large collection of remotely delivered analysis tools (including sophisticated)**
- **demonstration & exploration applets**
- **Educational materials**

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

http://www.statcrunch.com/

Webster West

StatCrunch
Data analysis on the Web

Home Explore ▾ My StatCrunch ▾

Sign in below

StatCrunch ID:
Password:
Sign in!
Forgot ID/password?
Problems?

Subscribe
Start analyzing data!

Prepay
Purchase bulk subscriptions

License
Run StatCrunch on your server

Testimonials
What folks have to say

Awards and Honors
Kudos for StatCrunch

News
The latest and greatest

Google plotter

Attention StatCrunch users!
Beginning January 1, 2009, StatCrunch will be exclusively hosted and distributed by Pearson Education. While more details about this transition will be available soon, rest assured that you will have continued access to the same StatCrunch software and web site. In fact, Pearson will provide more professional web hosting featuring faster access and live 24/7 technical support.

Welcome to StatCrunch.com
Whether you subscribe or explore, StatCrunch has plenty to offer. As a subscriber:

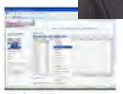
- **Upload** data files from your computer or the Web to your My Data listing
- **Analyze** data using the **extensive list** of numerical and graphical procedures StatCrunch offers
- **Export** data analysis results to your My Results listing for easy access in the future
- **Report** your insights along with attached data sets and analysis results
- **Share** your data, results and reports with the rest of the world or keep them private
- **Comment** on your items or those being shared by other subscribers

As an explorer, browse the numerous data sets, results and reports that StatCrunch subscribers are sharing. Start with the recently shared items shown to the right and use the links at the top of the page to navigate around the site. You can also **link** data sets on the Web to StatCrunch for subscribers to analyze.

What's new in this version?
If you have used StatCrunch in the past, the first thing you'll notice is the site makeover, but the new StatCrunch is more than just a pretty face. Here are a few of the highlights:

- **Create** richer content by providing more information about your items
- **Tag** your stuff to make it easier for others to find
- **Track** how often your items are viewed

A Video In...



Flash - WMV

Recently shared data sets

Responses to StatCrunch election poll

By websterwest
On Oct 24, 2008
View more >

Recently shared results

Results of StatCrunch election poll

By websterwest
On Oct 13, 2008
View more >

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

http://cast.massey.ac.nz/

Doug Stirling

Welcome to CAST

What is CAST?

CAST stands for ***Computer-Assisted Statistics Teaching*** and contains several different textbooks that teach statistical methods. They all take a data-focused approach to the subject and use minimal mathematics.

Unlike conventional textbooks, CAST makes extensive use of interactive diagrams to help explain the statistical concepts. Not only do these interactive diagrams help you to understand statistics, but they also make the ideas more memorable and help to retain interest.

Before starting CAST, you should check that your computer software is recent enough to run it.

Java check

CAST contains many interactive diagrams that are written in the Java programming language. Java is also used for the login procedure. Before starting CAST, check that your computer can run these Java programs. (In particular, many versions of the Microsoft Windows operating system do not have adequate Java support.)

Do you see a coloured rectangle on the left? (It may take a minute to appear.)

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

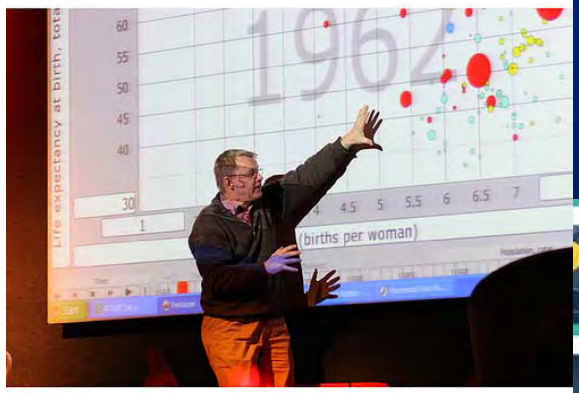
IASC 2008

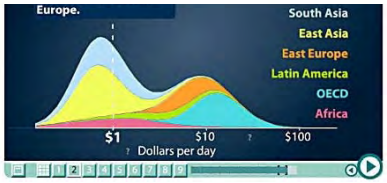
Wonderful aids, but paradigm shifters?

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

IASC 2008

Hans Rosling





THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Hans Rosling

- **Prof. of Global Health** from Sweden
  - Karolinska Institute
- Founder of **visualisation software "Gapminder"**
  - purchased by Google in March 2007

www.gapminder.org

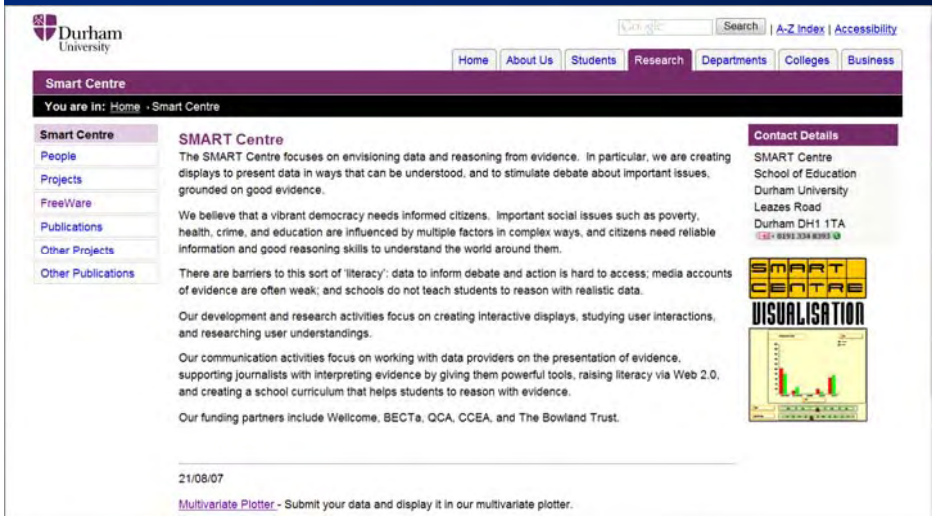
- **Legendary presenter**
  - Keynote Royal Statist. Soc. Ann. Conf. 2007
  - Keynote for next ICOTS

Show movie 1

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

http://www.dur.ac.uk/smart.centre/

Jim Ridgway



THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Best way to see what they do at Durham is to look at some data

- Number of new STI episodes seen at genitourinary medicine clinics by sex, 2000-2004
  - STI data produced by the Health Protection Agency
    - http://www.hpa.org.uk
      - This data set can be explored (along with a number of other data sets) at www.durham.ac.uk/smart.centre/freeware

Fire up **STI_GUM.SWF**

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Some features of the Data

- Differences between sexes in the incidence of Chlamydia
- Different age peaks
- Big increases in Chlamydia over time
- Pattern doesn't appear for Warts, though might expect
  - Why? Because C is asymptomatic in F? [Warts are not!]
- Differences in slopes of incidence over time at different ages

4-way interactions between disease*sex*age*time but still intelligible!

- (cf anova tables with lists of main effects, 1,2,3,4 variable interactions)

Some features of the Data

- Differences between sexes in the incidence of chlamydia
- Different age peaks
- Big increases in C over time
- Pattern doesn't appear for Warts, though might expect
  - Why? Because C is asymptomatic in F? [Warts are not!]
- Interactions between disease*sex*age*time!
  - but still intelligible
    - (cf anova tables with lists of main effects, 1,2,3,4 variable interactions)
- Differences in slopes of incidence over time at different ages

- NOT only can we see these things,
  BUT SO CAN ORDINARY TEENAGERS
  - See two 2007 papers by Ridgway et al.

It may be a simple little piece of software,
but the implications are HUGE …

When the right pictures and animations are used, even children can understand much more complicated data structures than we would ever have thought possible

Similarly with Rosling and Gapminder

TABLE·OF·CONTENTS¶

Volume·75· Issue·3· ·Pages·279· -·438· (December·2007)¶

……¶

Virtual·Experiments·and·Their·Use·in·Teaching·Experimental·Design°(p·281-294)↵
Paul·L.·Darius,·Kenneth· M.·Portier,· Eddie·Schrevens↵

Dynamic,·Interactive·Documents·for·Teaching·Statistical·Practice°(p·295-321)↵
Deborah· Nolan,· Duncan· Temple· Lang↵

Virtual·Environments·and·the·Acceleration·of·Experiential·Learning°(p·322-335)↵
Chris·Wild↵

One·Possible·Frame·for·Thinking·about·Experiential·Learning°(p·336-347)↵
George·W.·Cobb↵

Computer-Aided·Statistical·Instruction—Multi-Mediocre·Techno-Trash?°(p·348-354)↵
Carl·J.·Schwarz↵

On·the·Utility·of·E-Learning·in·Statistics°(p·355-364)↵
Wolfgang· Härdle,· Sigbert· Klinke,· Uwe·Ziegenhagen↵

Statistical·Thinking·in·Computer-Based·Learning·Environments°(p·365-371)↵
Gilberte· Schuyten,· Olivier· Thas↵

……¶

Data generating environments

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

http://lstat.kuleuven.be/env2exp/

Paul Darius

enV2exp

Environments to Experiment

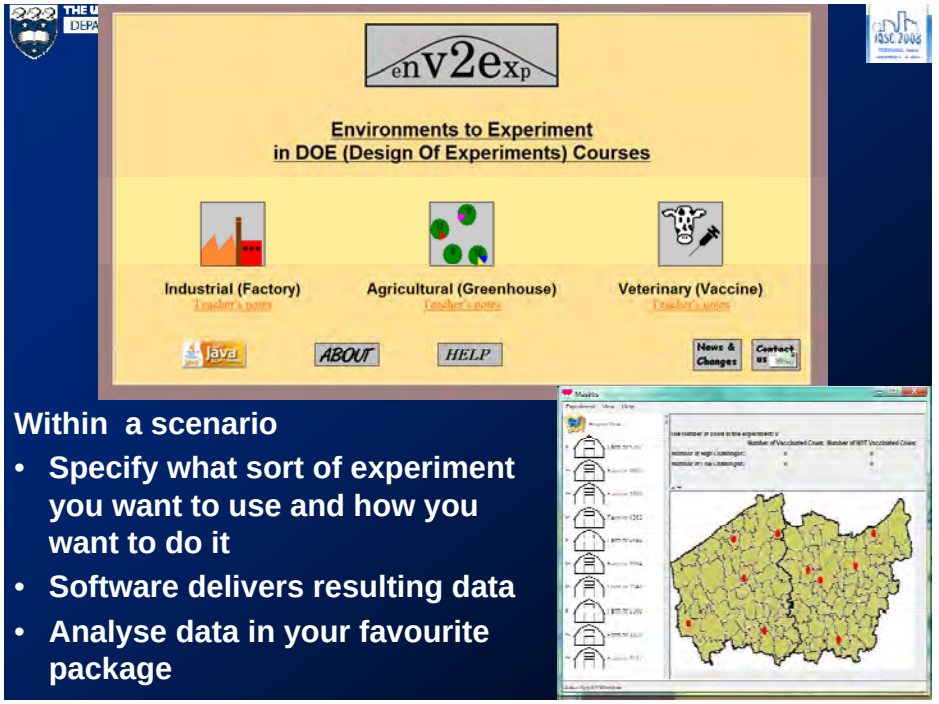
Introductory Statistics Course
Teacher's notes

Sampling Course
Teacher's notes

DOE Course
Teacher's notes

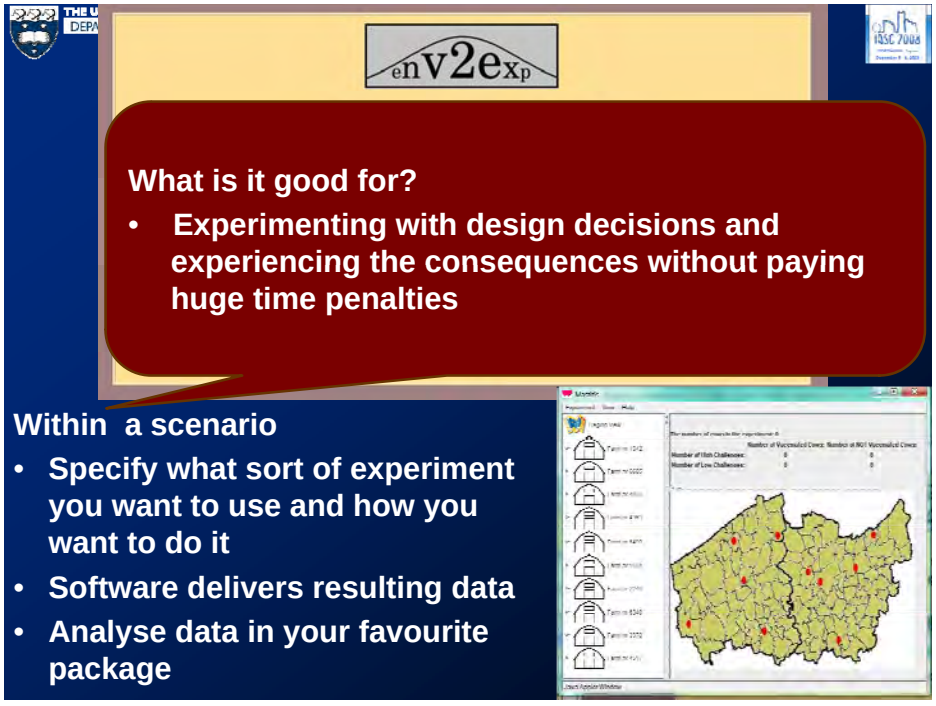
Please note that all the applets on this page require a recent version of Sun's Java Virtual Machine. If you experience any problems, go to http://java.com/java/download and install the most recent version of the Java Runtime Environment (JRE).

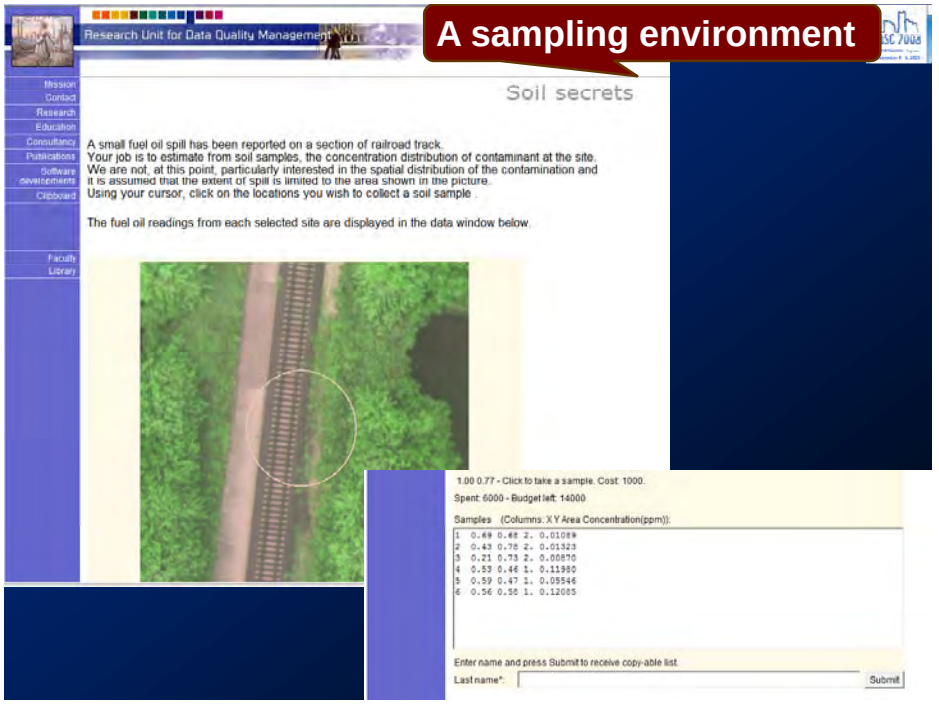
Java ABOUT HELP News & Changes Contact us



Within a scenario

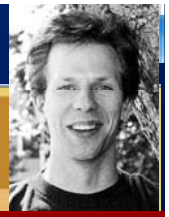
- Specify what sort of experiment you want to use and how you want to do it
- Software delivers resulting data
- Analyse data in your favourite package





THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

http://cgi.sfu.ca/~sblay/cgi-bin/stat/



Derek Bingham

Statistical Experiment Simulations

- Animal Dispersal Model
- Shooting a Cannon
  Completely Randomized Design
- Sampling a school of Salmon
  Mark - Recapture
- Barometric Pressure
  Regression Analysis
- Dragging Sheeps
  Two Factors Experiment Design
- Crop Fertilizers
  Two Factors Split-Plot Experiment Design
- Baking Muffins
  A Model with Randomization Restrictions
- Balls Falling from the Leaning Tower of Pisa

Developed by Sigal [illegible]
Funded by Derek Bi[illegible]
Department of Statistics and Actuarial Science, SFU.

If the Java applets on the left don't run,
try clicking here

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

http://www.stats.uwaterloo.ca/Faculty/Steiner.shtml

Waterloo
Statistics and Actuarial Science

SAS Home
About us
Undergraduate Studies
Graduate Studies
Research
Seminars
Math Home
UW Home

Stefan Steiner

Associate Professor
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada

(519) 888-4567 ext. 36506
Fax: (519) 746-1875

Research Interests

I conduct research in the areas of industrial statistics and operations research. More specifically, much of my work involves the design and analysis of innovative process monitoring and quality improvement techniques for use in industry. Recently my work has revolved around a step by step algorithm for process improvement & variation reduction, called Statistical Engineering.

BMATH (Applied Mathematics), 1987, University of Waterloo
M.Sc. (Applied Mathematics), 1989, University of British Columbia
Ph.D. (Management Science/Information Systems), 1994, McMaster University

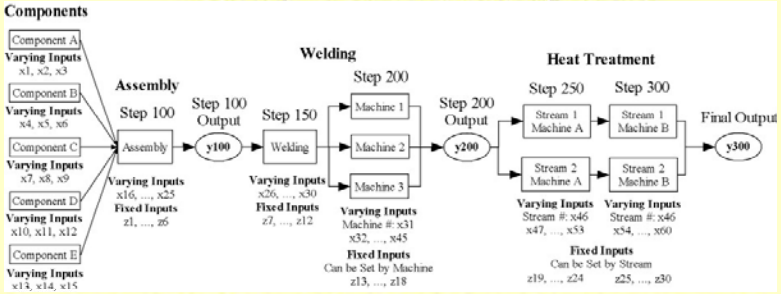
Publications
Consulting Activities
Statistical Engineering
Virtual Manufacturing Process

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Stefan Steiner

Jock MacKay

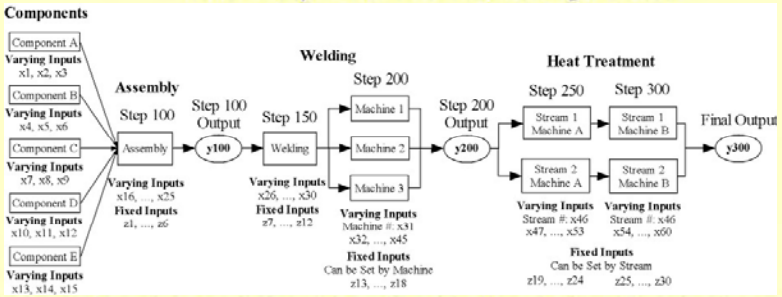
Watfactory Virtual Manufacturing Process



The Watfactory virtual process simulates a manufacturing process for automobile camshafts and has a number of processing steps and many inputs – see process map above. The process has 60 inputs that vary in regular production (x1, …, x60) and 30 inputs are are normally fixed (z1, …, z30). A web interface allows a wide variety of process investigations including many types of observational and experimental plans.

This virtual process has been used successfully for a number of years to teach process improvement and variation reduction techniques and methodologies. In the class or workshop each team (assigned a different version of the virtual process) is required to plan and analyze a series of process investigations to first learn about how the process works (e.g. identify important inputs, find the nature of the output variation, etc.) and ultimately how to reduce variation in the final output (y300). Also, to simulate reality, teams are given a budget and information on costs associated with different investigations. Both university students and industrial personnel find interacting with the virtual process enjoyable and report finding it a very helpful learning aid.

Watfactory Virtual Manufacturing Process



The virtual process is linked to the Statistical Engineering algorithm (Steiner and MacKay, 2005 see http://www.stats.uwaterloo.ca/Faculty/SteinerBook.shtml) for addressing chronic variation problems in manufacturing processes. The virtual process is also well suited to aid in teaching other quality improvement systems such a Six Sigma.

Highlights of teaching with the virtual process:

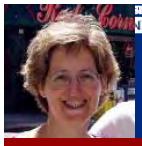
- Allows quick exploration (i.e. during a short course) of process improvement ideas in a realistic environment without risk or cost.
- Forces participants to think about the details of each process investigation, i.e. when and how many parts to select, what characteristics to measure/change, etc.
- Models the iterative and sequential nature of real problem solving
- Encourages the use of process knowledge gained in earlier investigations to help plan sensible new ones.

The software can be accessed at http://services03.student.math.uwaterloo.ca:8080/~stat435/login.htm

To try with standard settings use Login (Student) ID: guest01 & Password (Student Number): 00000001.

For the 5 cause challenge (as described in the rejoinder to my Quality Engineering article "An Overview of the Shainin System for Quality Improvement") use Login (Student) ID: 5cause & Password (Student Number): 55555555.

Waterloo
Department of Statistics & Actuarial Science



Deb Nolan

http://www.statdocs.org/

Duncan Temple Lang

About StatDocs

Index to StatDocs Website

StatDocs Main Page

About StatDocs

How to create a StatDoc

StatDocs Examples

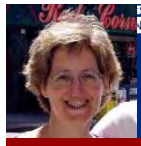
Software

StatDocs are dynamic, integrated, reproducible, interactive, statistical documents.

- A StatDoc document *dynamically integrates* text with data, code, and GUI components, and it supports multiple views of its contents.
- The document can be *reproducible* because the code that produced the figures, tables, and algorithms can be inspected, modified, reused and critiqued by the reader, and because code in the document is represented in a common form that can be translated into different software languages.
- The document may handle *interactive* content in that data can be updated and replotted; and graphical user interfaces can be included to address what-if questions. It can also have *dynamic* structure in that navigation through and views of the document can be determined according to reader responses and actions.

To find out more about the *StatDocs* project, take at look at the following papers.

- Dynamic, Interactive Documents for Teaching Statistical Practice, *International Statistical Review*, Dec 2007.



Deb Nolan

http://www.statdocs.org/

Duncan Temple Lang

What will it be good for?

- **Experimenting with alternate analyses and analysis decisions and experiencing the consequences without paying huge time penalties**
- **and much more besides**

Index to StatDocs Website

StatDocs Main Page

About StatDocs

How to create a StatDoc

StatDocs Examples

Software

StatDocs are dynamic, integrated, reproducible, interactive, statistical documents.

- A StatDoc document *dynamically integrates* text with data, code, and GUI components, and it supports multiple views of its contents.
- The document can be *reproducible* because the code that produced the figures, tables, and algorithms can be inspected, modified, reused and critiqued by the reader, and because code in the document is represented in a common form that can be translated into different software languages.
- The document may handle *interactive* content in that data can be updated and replotted; and graphical user interfaces can be included to address what-if questions. It can also have *dynamic* structure in that navigation through and views of the document can be determined according to reader responses and actions.

To find out more about the *StatDocs* project, take at look at the following papers.

- Dynamic, Interactive Documents for Teaching Statistical Practice, *International Statistical Review*, Dec 2007.

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Statdocs

- Huge potential for creative thinkers to build elaborate interactive documents
- Real potential to be paradigm shifting
  - But the authoring tools that would let me use it are not there yet

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

In praise of quick and cheap solutions

English slang, "Quick and dirty solutions"

- I've just revisited some of these sites one year on
  - Not many signs of progress
- This points to a reality in the way we live and work

In praise of quick and cheap solutions

- These things tend to be add-on, sideline activities for academics
  - Crowded out by "the real job" of teaching, research and administration
  - Built on budgets that range from zero to tiny
  - Thus, hard to achieve ambitious aims
- Points to …
  - a ***need to be able to create something useful in a very short amount of time***

Have bright idea → want it built quickly and cheaply

- even if it is more a prototype than a polished product

Solutions

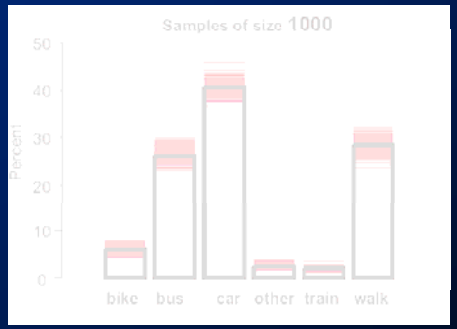
1. **W**ork in ways that let us **leverage off one another**
   - Something like the R project for building statistical learning environments
     - Store houses of building blocks
2. **Leverage off powerful existing systems** we already know how to use
   - For me that is R, and …
     - **I don't really have time to learn new systems**

Examples

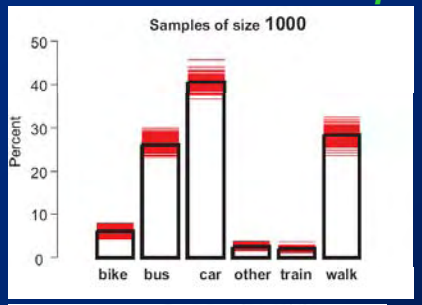
Bar Chart Animations

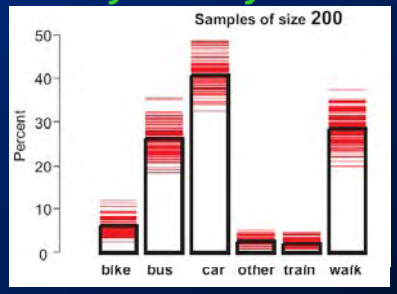
Play

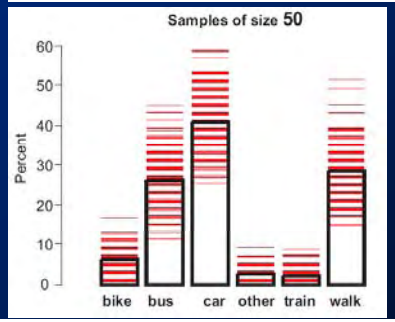
- Samples of 1000
- Samples of 200
- Samples of 100
- Samples of 50
- Samples of 30
- Samples of 30 without jitter

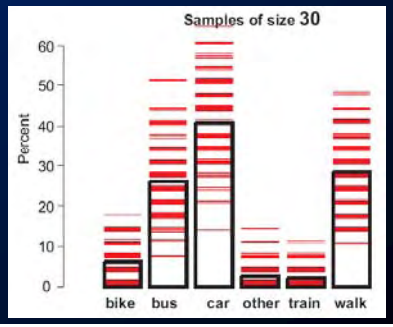


"What I see is not quite the way it really is"

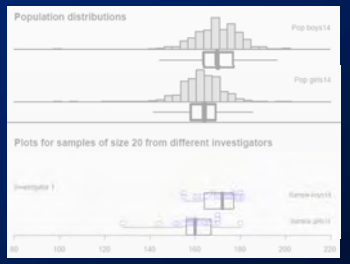








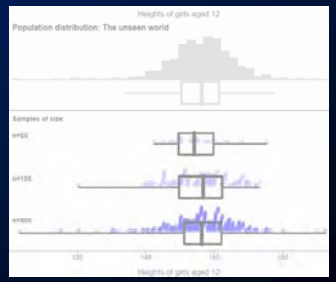
Dot and Boxplot Animations



Play

- Original 2-sample

- Effect of sample size



How to create a sampling variation "movie" in R

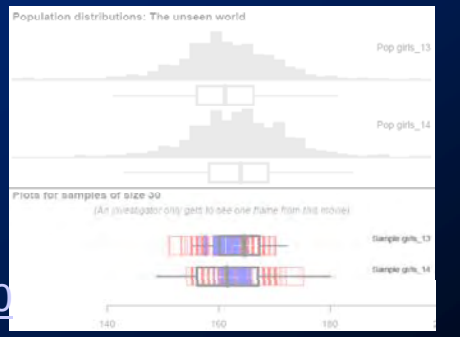
```
pdf(filename)

for (i in 1:n){
  take a sample
   ……
   plot( … )
}

dev.off()
```

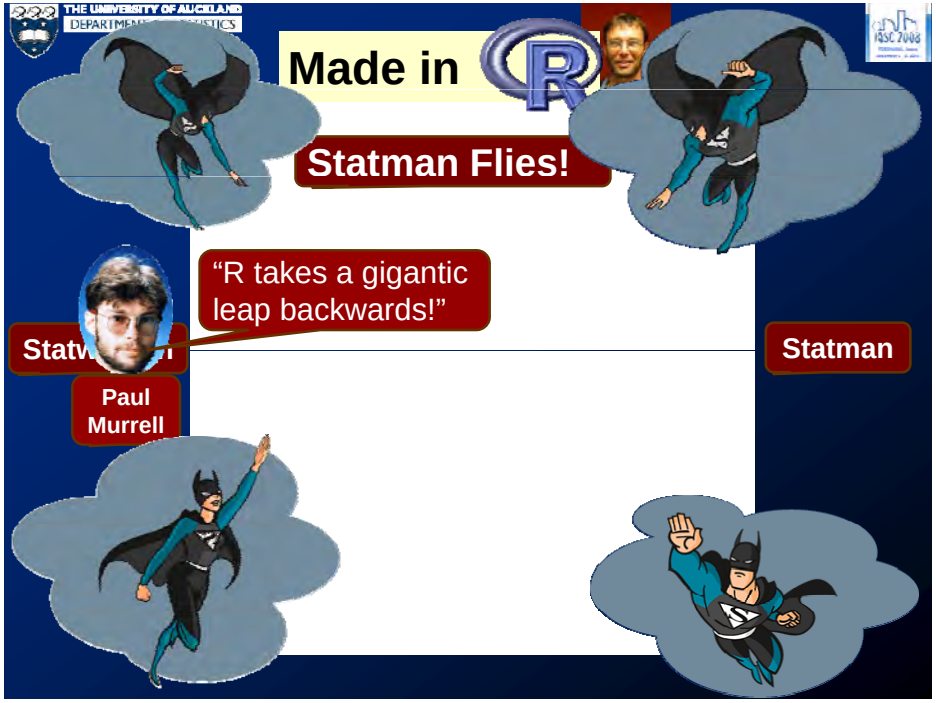
- Then "play" the resulting pdf file!
- If want to get more sophisticated plot to png, transform to animated gif

Boxplots with a Memory II



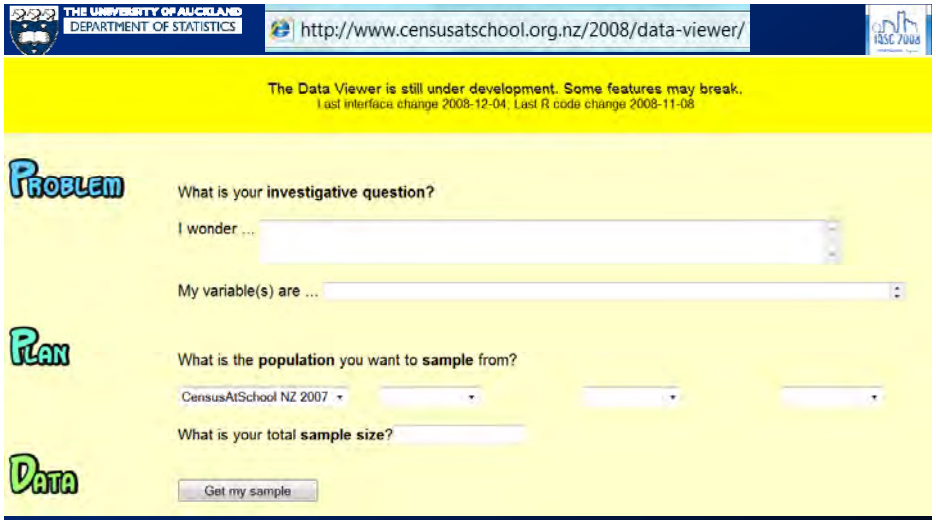
Play

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



Tool needed for Census At School outreach programme

- Want to get some free sampling and analysis software to all schools
  - Help defeat "by hand" mind set
- Stop-gap measure till they all get real software
  - E.g. Tinkerplots, Fathom
- Usual story
  - No budget, no time
- But Department does have a web programmer and I know a bit of R
  - What can we put together quickly ???



http://www.censusatschool.org.nz/2008/data-viewer

Play Movie

This is a very simple tool ….

- with very limited functionality
- But even this has its strengths …
  - The more comprehensive a tool gets the steeper the learning curve
    - Functionality brings complexity

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

How it was built

php web page

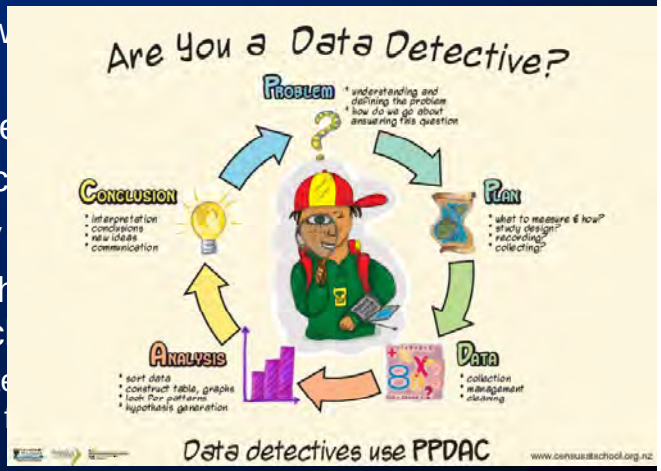
- Extracts information from user
  - Names of variables to sample on (if any)
  - Sample size(s)
  - Names of variables to use in analysis
- Extracts data from databases using SQL query
- **Fires up R**
  - and passes the above information to an R function
  - R does the analysis, drops graphics files **and R dies**
- php page finds and displays output files
- Allows the user to ask for more analyses
  - Every time user clicks “Do Analysis” the fire-up-R step is repeated

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

What is it good for?

- Experiences that stress
  - forming questions
  - identifying w
    from
  - identifying re
  - forming/disc

  within a very
  - So can go th
    in a single c
- The “get data” ste
  and happen in a



THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

What is it good for?

- Experiences that stress
  - forming questions
  - identifying what subpopulations to sample from
  - identifying relevant variables
  - forming/discussing/writing about conclusions

  within a very short time frame
  - So can go through these steps several times in a single class
- The “get data” step and “Analysis” steps are automated and happen in a flash

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

What is missing?

- Good for purposes above
  - Because the questioning, seeing the data and concluding steps are happening close together in time
- But bad …
  - for getting an appreciation of the nature of data collection
  - choosing a methods analysis
    - (e.g. type of graphic)
- Clearly needs
  - To be supplemented by other experiences that address these areas
  - But I’d argue, “establish the big picture before concentrating on details”

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Bottom Line

- **Php ↔ R**
  - is a great way of quickly building free tools for remote delivery to the masses
    - Until, of course, Statdocs gives us easy-to-use authoring tools

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

All that I promised but did not deliver is in …

International Statistical Review (2007), 75, 3, 322–335 doi:10.1111/j.1751-5823.2007.00033.x

Virtual Environments and the Acceleration of Experiential Learning

Chris Wild

University of Auckland, New Zealand. E-mail: c.wild@auckland.ac.nz

THE UNIVERSITY OF AUCKLAND
DEPARTMENT OF STATISTICS

Thank you for your attention

Slides at **http://www.stat.auckland.ac.nz/~wild/08-talks/IASC.wild.pdf**