"I have a dream …"

They too had dreams!

“I have a dream of students spellbound by the broad vistas of the data landscape

I have a dream of their flying on magic carpets that enable them to swoop effortlessly over this landscape exploring its nooks and crannies in search of its hidden treasures

I have a dream of students empowered to look at data, explore analysis systems and educational environments designed so that, like Alice in Wonderland, they keep crying “Curiouser and curioser!” and have the ability and confidence to go where that curiosity leads.
I have a dream of educational and analysis environments designed to leverage the power of “I wonder …?” to draw students in to discovering more and more –

the power of “I wonder why that is?”, the power of “I wonder what happens if …?”, the power of “I wonder what that does?”, the power of “I wonder what’s around the next bend or just over the horizon?”

I have a dream of software that finesses away the mundane, the mind-numbing and the soul-destroying difficulties.

Beliefs/assumptions

• Missionary for statistics and “statistics for the masses”
  • Beyond just “literacy”
  • “Courtship” not just “what is good for you”
  • “It’s all far too slow”
• Primacy of graphics
• Desirability of “Middleware”
• Coding can’t cut it

Early phases …

• Think “courtship” not “what’s good for you”
• Entice / En-trance / Empower / Educate
• Target …
  • “What I can do with data & what data can do for me”
• Build …
  • broad “awareness” and “empowerment” simultaneously
  • sense that “this is important”
    – but not just for some high-powered PhDs somewhere
    – “This will be useful for me”
  » It also has to seem “accessible to me”
Early phases …

• Conveying a sense of empowerment in a limited amount of time comes from
  • isolating a small number of easily understood metaphors
  • that can combine in flexible, powerful ways
  • are easy and fast to use

NEXT …

The data world …
has gotten a whole lot bigger

Can’t just keep illuminating same small patch

• Need to get much …
  • further
  • faster
  • & with better comprehension

“It’s all far too slow”

• Statistics education is far too slow
• Need ….
  • to be able to experience much more of what the data world has to offer enormously faster
  • to leap to warp speed!

• Coding is a barrier
Primacy of graphics
• Key to accessibility, acceleration, conceptual development
  • serve up graphics first and by default
  • everything else has to be asked for
  • you can’t actually get anything without looking at graphics first

Desirability of “Middleware”
(Not in the technical sense)
• software aimed at
  • allowing student to experience
    • as much as possible of “discovery in the data world”
    • in the least possible time
      – May also be useful for professionals but only as a fortunate by-product, not a design feature
  • Designed in conjunction with the educational experiences that are the main driver

Developing in concert ...
Software
   Educational Experiences
Desired Capabilities
(Trade-offs everywhere)

What should early experiences of statistical data exploration look like?
Early experiences of data analysis should be like this … (animated “movie”)

NOT like this …

Like this …

For lots of exposure quickly

• Two sides of the coin …
  1. Deciding what to expose
     Seek opportunities to serve up powerful capabilities where the output graphics are easy to comprehend
     • If it takes a long time to set up & explain, don’t go there
  2. Delivery mechanisms that are very fast & intuitive to operate
     • Can get a long way without having to know/remember the names of things
       – Context awareness & maximising what is done by default
Data Meets Viz

Developing in concert …

Software

Educational Experiences

Desired Capabilities

(My projects … (Visual Inference Tools)

iNZight

Data Analysis

Conceptual development

VIT

Research

MOOC

“Data to Insight” (prototype for “getting further faster”)

(Trade-offs everywhere)

https://www.stat.auckland.ac.nz/~wild/iNZight/

(animated GIFs on iNZight homepage)
Why??

BYOD

("Bring your own device"-driven need for tools to work on tablets and even phones)
Week 1: Introduction; gee whiz; software; data
Week 2: Boot Camp (Basic Training) – you'll see bits of this
Week 3 & 4: Relationships between variables
  • Relationships between categorical variables – you'll see bits of this
  • Relationships between numeric variables
    • Trend, scatter & outliers; Clusters
    • Prediction with uncertainty
    • Association & Correlation
    • Trends: Lines, curves & smoothers
    • Large data-set problems and solutions
    • Overprinting, jitter & transparency, granularity & point size, running quantiles
    • More variables with size, colour and subsetting

Week 5: Why “what I see is never quite the way it really is”
  • Measurement and “selection” bias
  • Sampling error and sampling variation
  • Causation and confounding
Week 6: Estimation with confidence via bootstrap
Week 7: Designed Experiments and randomisation tests
Week 8: Time series stressing seasonal series with forecasting & comparing related series
Middleware for Middle Earth

- If only have one bite, prioritise creating a sense of possibility and potential
  - Liberate imagination
  - Foster “there is value here for me”
- So that the “one bite” is actually only the first of many bites
  - So they will want to come back for more somehow, somewhere, sometime.
  - And with increasingly flexible education ...

Initial experiences should be like this!

Thank you