Statistics learning for the next decade

Helen MacGillivray
President, International Statistical Institute
Principal Fellow, Higher Education Academy
Editor, Teaching Statistics
Chair, UN Global Network of Institutions for Statistical Training

Thank you to Institute of Statistical Science
Possible sub-titles

- What’s new, what’s not
- Looking back, looking forward
- Lessons from the past feeding into the future
- The good, the bad, the ugly
Outline

- Data literacy
- Data science
- Big data

- Developments over past few decades
- Some challenges
  - university, across disciplines, school, postgraduate
- Elaborations on some of the challenges
- Some ways forwards
- Bit of history
- Opportunities
Recent comments from data scientists

- **Data science is everywhere and not new**
  - A label for work being done for years
  - What’s changed is recognition of what analysts + technology can do and bringing this out of the back room
  - Need to know what can & can’t be done with data
  - Methods and technology not biased – data can be

- **Data Science is not a person – it’s a team. Diversity essential**
  - Need specialists: statisticians, data engineers/computer scientists, machine learning optimiser, subject matter person
  - Need: curiosity in problems; problem-solving; bring together data, data problem-solving, technical; challenge team
  - But all members need some statistical foundation

- **Data science gives what, statistics gives why/understanding**
  - Data scientists are statisticians who make meaning from data

Need statistics in data science & data science in statistics
How much coding/programming in data science education? Reminders of long-lasting questions re maths in stats
Absence of explicit recognition of value of statistical and technical (and mathematical) skills not new

- Advice for decades to job-seeking graduates: *look for skills in ads; look for ‘analyst’.*

- Student portfolios: identification of skills and awareness of broad, methodological & technical skills

- More than two decades ago, I set up double degree in maths/stats and IT.
  - Those graduates went everywhere
  - Feedback included:
    - IT needs to be enabler to tackle anything in IT
    - Value of statistical learning which reflects the *practice* of statistics
    - Foundations for further learning

Employers also emphasize do not want production line of hybrid graduates: want sufficient diversity of graduates with balance+specialisation to work in effective teams.
What is statistics?

- **Statistics**: science of variation, data, uncertainty, questioning of models, assumptions and interpretations

- Critical importance lies in:
  - pervasiveness
  - universality of concepts and thinking
  - power in specific contexts – across disciplines, business, industry, government and society
  - can be a driver, partner or servant, but from the most theoretical to the most applied, its roots lie always in real problems.
“Data science” “Data literacy” “Big data”

Statistics and its teaching even MORE important

- Views ‘data science’ and ‘statistical sciences’: recap
  - *Data science is everywhere: brings together statistics, computer science, engineering, context knowledge*
  - Need statistics in data science; data science in statistics
  - Need ownership, problem-solving, curiosity, thinking, critiquing

- Diversity of perceptions of relationships between ‘data literacy’ and ‘statistical literacy’
  - “everyone knows they’re different”
  - “everyone knows they’re the same”

- ‘Big data’ – complex and rich
  - Multivariate, variable diversity, and/or many cases
  - Data quality, high level technological data management
Some recent descriptions of data literacy

- **Data literacy is the ability to read, create and communicate data as information and has been formally described in varying ways.**

- **The desire and ability to constructively engage in society through and about data** [http://datapopalliance.org/item/what-is-data-literacy/](http://datapopalliance.org/item/what-is-data-literacy/)

- **Data literacy: ability to interpret, evaluate, and communicate statistical information…how statistical information is created, encompassing data production**

Some descriptions of statistical literacy

- Good “statistical citizens”: able to consume information that they are inundated with on a daily basis, think critically about it, and make good decisions. Rumsey (2002)

- People’s ability to interpret and critically evaluate statistical information and data-based arguments appearing in diverse media channels, and their ability to discuss their opinions regarding such statistical information (Gal 2000)

- Become much more critical about the way data are produced, the way data are presented and the way data are interpreted.

- Statistics is a thinking science

Why is ‘data literacy’ trying to reinvent ‘statistical literacy’?

- Lack of message penetration
- New contexts: data sources; UN SDG’s; citizen empowerment; technology
Learning through reflection and analysis

Learning in teaching

- Reflective practice: analysing, evaluating, synthesizing
- Evidence of effectiveness/impact and non-effectiveness/non-impact
- **Listening and observing** students, reading students’ work
- Asking and exploring why

Higher Education Academy (HEA) criteria, standards & requirements

---

As we do with students, we need to do with users, refuters, collaborators, authors, dissenters, across all disciplines

*listen, observe, read, discuss, reflect, analyse, devise, evaluate, revise*
Some recent sources

WSC’s, RSC’s, ICOTS, OZCOTS
Principal Fellow, HEA

Co-Editor, 2014-2016
Editor, 2017-

Teaching Statistics first appeared 1979, published three times a year. Published by the Teaching Statistics Trust. Teaching Statistics is intended for all those who teach statistics to students aged 9-19 years. The emphasis is on good practice in teaching statistics and statistical thinking in any context, whether in statistics subjects/courses/modules or in other disciplines such as economics and business, biology and health sciences, technology, psychology, mathematics and any area which uses statistics.

• UN World Data Forum: 2017, 2018
• UN Global Network of Institutions for Statistical Training (GIST)
Less recent sources

1972-2011:
- University: all sizes, levels, disciplines, curricula, texts, mentoring …..
- Schools: teachers, schools, enrichment, curricula, resources, texts……..
- Learning support……..
- International, national: Reviews, IASE, SSA, RSSCSE, ICOTS, OZCOTS, HEA
- National senior fellow

Organising computing labs for introductory “service” statistics 1980’s

[Image of cartoon with the text: "There’s a wait... Everyone wants a window"]
Developments over past few decades

- During 1980’s and 1990’s, many statisticians and statistics educators worldwide initiated & implemented variety of changes
  - in teaching statistics at university, particularly introductory levels across disciplines, & at school level.
  - in workplaces & community.
  - In statistics education research.

- Much reported in papers, at conferences, particularly since 1990 (ICOTS3)
  - ICOTS, IASE satellites & roundtables
  - SERJ (started 2002) JSE (ASA), Teaching Statistics, ISR & statistics journals (American Statistician, JRSS, etc)
Developments over past few decades

- Advocacy of
  - Data-driven concepts and statistical thinking
  - Real, ‘large’ contexts and data: simple within complex
  - Statistics in its own right (maths is servant)
  - Technological and data systems know-how
  - Student ownership and constructivism
Long-time advocacy from statisticians


- Vic Barnett (1986)
  - “we see, tied up together, the role of the statistician as consultant, consultancy as the stimulus for research in statistics, and consultancy as the basis for teaching statistics”.

- Authentic experience of full statistical investigation process
  - Cameron (2009) builds on Chambers’ (1993) ‘greater statistics’* comments that “such training is an appropriate foundation for most statisticians wherever they may be employed.”
    - Note: part of the pyramid model

- * Donoho (2017) ‘greater data science’
Long-time advocacy from statisticians

- Authentic experience of full statistical investigation process
  - Kenett & Thyregod (2005) 5 steps in statistical practice
    - “important to take part in collection of data, or at least have the opportunity to watch data being collected or generated.”
    - “encourage academic courses to cover the full 1–5 cycle....especially steps 1, 2 and 5.”

- 1. Problem elicitation & preparation for tackling statistically
- 2. Preparing data (including planning, collecting, sourcing, identifying, organizing, validating........)

- 5. Presentation of findings

Similar to advocacy for data science
Data investigation process

- Descriptions can depend on context

- SQC – Shewhart 1939
- Deming-Ishikawa: PDCA – gap analysis for problem finding

- Hadley Wickham: input – tidy – transform – communicate

- All descriptions
  - emphasize importance of everything before analysis and everything after
  - emphasize cycle: building solutions to improve understanding of issues/problems

- Statistical analysis is essentially exploratory
- Need to teach communication of assumptions and findings
  - “Solution” ≠ the answer
Authentic learning of data investigations

- 1994-2011: semester-long free-choice full data investigation embedded in large introductory statistics courses in engineering, all sciences, IT and mainstream statistics programs
  - “Set” data and contexts, no matter how real, can’t provide experience of \textit{setting up, investigating, reporting}
  - \textit{Motivation to find tools}
  - \textit{Ownership} of data and context engagement

Student “wow!”

- * ownership
- * visualisation + exploration
- * tool empowerment within complex (>5 variables)
- * student judgement + communication

Entice
Excite
Empower
Grab
Keep
Maintain

Similar to advocacy from data science
Student choices: > 5000 projects!

Just a few!
- The three minute pop song
- Length of corporate employee phone calls
- 24 hours in a service station
- Lift or stairs?
- Aircraft noise levels
- Go go go!
- Human curiosity
- Death by statistics
- Holding breath
- Where are all the single people?

Egg strengths

Human curiosity

Crash testing stubbies
Many effects on learning and teaching

- Choices of topics illustrate types of examples in which students want to see how statistical thinking and techniques can help
- Improved overall results
- Past students remember their projects - as do staff
- Discovered what students need
- Discovered what engages students.... “get students to the sexiest, most useful techniques faster & more effectively…” Wild, 2006

- Significant curriculum re-development to better reflect
  - learning needs
  - real statistical problems
  - modern statistics
  - statistical practice

Some challenges: particularly for introductory tertiary, school, other disciplines

<table>
<thead>
<tr>
<th>Penetration insufficient within and across disciplines and levels. Important to reflect on the why .....</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature, size &amp; pervasiveness of Statistics</td>
</tr>
<tr>
<td>Dynamic nature of Statistics: responds to data, technologies, disciplines, workplaces</td>
</tr>
<tr>
<td>Technology: resources, use &amp; how much to learn</td>
</tr>
<tr>
<td>Need real, complex, many-variable datasets</td>
</tr>
<tr>
<td>Visualisation: still too much focus on measures</td>
</tr>
<tr>
<td>Assessment fears</td>
</tr>
<tr>
<td>Workload</td>
</tr>
<tr>
<td>Open-ended</td>
</tr>
<tr>
<td>Students “won’t do it right”</td>
</tr>
</tbody>
</table>
Some challenges

- Too much focus on new ways of learning old content & old sequencing
- Domination of 1 and 2 variables
- Not enough understanding/emphasis assumptions and models
- Need more on identifying variables (& types), cases
- Leftovers past their use-by-date
  - Return ‘population’ to its proper meaning
- “surface” referencing
- What tools can & can’t do
  - Histograms, boxplots
Some challenges

- Lack of coherent development
- Non-authentic experience of statistical investigation process
- Rigid, discipline-embedded approaches, top-down case studies
- Can’t build on shaky foundations
- Perpetuation of norms
- Reclaim and reform learning of probabilistic thinking
- Research hypotheses vs statistical investigation

- Reflect on overall
  - Sometimes digging just produces a hole, and digging deeper gives mud
  - Sometimes climbing and looking around shows way forward
Some challenges

- ‘The’ question & ‘the’ answer

- Not enough of the initial exploration/framing of issues, what data and what variables

- Too much rush to force into ‘desired’ form or get to ‘desired end’.
Some ways forward

- **Authentic experience**
  - “What goes on in head?”
  - Students have to **experience** it.
  - “Empathy” - cultivate by role model: “let’s see what we’ve got”

- **Too much training for research: statistics and other disciplines**

- **Real data and real contexts but**
  - Contexts must not dominate statistical learning
  - Contexts must be familiar/readily accessible to students
  - Staff research interests must be controlled
  - Beware teacher-centred, top-down or context-complex case studies

- **Must use technology as used in practice of statistics**
- **Authentic learning and assessment**

---

All also relevant to data science!
Some comments about postgraduate training

- “funding for doctoral training is primarily about ensuring a growing supply of well-trained researchers to help exploit the potential benefits of the new knowledge economy.”

- In some countries as few as five percent of PhD graduates find permanent academic positions.

- “Need for PhD programs to include learning to teach”.

- Many PhD graduates find themselves in non-academic, non-research positions.

- “More attention to the more generic and transferrable skills and knowledge that research students develop and the need to pay more explicit attention to their development”.

Some comments about postgraduate training

- Reports on HDR training emphasise general research skills, analytic and critical thinking skills and many increasingly highlight the importance of statistical and data analysis skills.
  - Transferable skills … closely linked to the process of research training yet valuable to a range of other professions (for example, critical thinking, project management and statistical analysis).

  - Skill sets in data analysis, predictive modelling and decision-making are also highly sought after and there was consensus that this demand is expected to continue to increase.

- Some reports specify data visualisation & analysis techniques without explicit reference to ‘statistics’ or ‘data analysis’, but also emphasize findings that:
  - Knowledge about designing and undertaking research, and about analysing information or data played a significantly larger role than did knowledge of their PhD disciplinary area.
Some statistical challenges in other disciplines

- Foundational understanding and content pedagogy knowledge insufficient across disciplines and educational levels
  - Can’t build on shaky foundations
  - Perpetuation of norms in other disciplines

- Example: Tragic case of Sally Clark included
  - Lack of identification of issues and context
  - Inappropriate data for estimates of probabilities
  - Misunderstanding of conditional probabilities and incorrect multiplication of probabilities
  - More misunderstanding of conditional probabilities - ‘Prosecutor’s fallacy’
  - Withholding of (pathology) data/information
Reclaim & reform probability learning

- Language & visualisation paramount
  - Use probability diagrams with probabilities represented by areas or lengths
  - Extensive student experience of conditioning language

- Conditional probability BEFORE independence
  - All probabilities are conditional
  - Use data, estimates, beliefs…….
  - \( P(A \text{ and } B) = P(A|B)P(B) \). Ban term ‘multiplication rule’

- There are different ways of assigning probabilities, NOT different types of probabilities
  - Estimate
  - Model
  - Belief
  - Part of cycle of data investigation and models
Some more challenges

- Incorrect use of types of data
- Understanding discretization & effects

- Essentials of hypothesis testing are natural
- Multiple procedures and forcing into norms
  - overuse of t

- Lack of identification, questioning and visualisation of assumptions

- Over-analysis of old instead of reflection on what and why
  - Forcing the new into the old

- Simulating the boring
Curriculum design

A design process

A statistical consulting job

A statistical and data investigation process

Clients are students, staff, teachers, administrators, bean counters.......

Examples/anecdotes

- What did MBA & electrical engineering have in common?
  - Learning to comment; discrete before continuous
- School: old habits die hard
- Postgraduates across disciplines
  - Variable types; role of statistics in research
- Introductory across disciplines
  - Get to multivariable & real empowerment as soon as possible
Assessment design - for learning

Reflect what is of value

- **Workload fears**
  - Can balance open-ended + multiple choice

- **“Doing it right” fears**
  - Need authentic student experience

- **“Must be useful”**
  - Students learn best in contexts that matter to them

- **Multiple choice questions**
  - Naturally course-specific
  - Tend to be highly dependent on local culture/conditions

- **Criteria and standards for investigations**
  - Tend to be more universal
  - Need exemplars
Bit of history: ISI & IASE

- In 1948, ISI President Stuart Rice set up ISI Education Committee, increasing ISI's mandate to undertake educational activities and collaborate with UNESCO and other UN agencies.

- UNESCO grant to ISI for govt statistical training: ISEC set up in India, 1950, by P.C. Mahalanobis, has trained > 1500 from >80 countries.

- In 1970’s, ISI increased attention to promoting statistics education in schools and universities. ISI Education Committee established task forces.


- Task Force on Teaching Statistics at School Level (TOTSAS), led initially by Vic Barnett
Bit of history: ISI & IASE

- TOTSAS group established regular newsletter (International Statistical Education newsletter). This lead to Vic and Joe Gani setting up the Teaching Statistics Trust to establish the journal *Teaching Statistics* in 1979
  - *Teaching Statistics* planning special issue on Data Science and Statistics for early 2020

- Warren Gilchrist & Vic established the first (UK) Centre for Statistical Education in 1982 with its first Director, Peter Holmes, now sponsor of TS prize to highlight excellence in motivating practical classroom activity.

- International Association for Statistics Education (IASE) established 1992, 1 of 7 ISI Associations. (Vere-Jones, 1994).

- In 1994, ISI committee to stimulate spread of quantitative skills around the world. In 2000 IASE invited to oversee it; called International Statistical literacy Project (ISLP) from 2002. In 2009, current structure of ISLP set up, including IAOS involvement
Currently: ISI and all Associations involved in various ways in learning and teaching of everything relating to data, statistics, data science

The learning and teaching of statistics is relevant to every statistician

- Schools, university, workplace, users, adult, citizens.....
- ISLP poster competition for two age groups of school students:
  - In 2017 more than 12,000 students participated, from 23 countries
  - Extended to university students for 2019
- New ISI-Esri competition for advanced undergraduates
- IASC competition in data
- New IASC journal in Data Science
- UN Sustainable Development Goals (SDG’s)
  - NSO’s
  - Across ISI: 11 IPS in WSC 2019 from different groups
  - GIST
- Awards, competitions
- Short courses, workshops

Need more: mentoring young academics in teaching; greater voice in curricula
Opportunities

- Lessons from decades of work in statistics education
  - biggest challenges lie in the nature and pervasiveness of statistics
  - universality of educational needs – statistics and data science
  - dynamic nature of statistics and data science in responding to data, technologies, disciplines, and workplaces
  - lessons from statistics for data science

- Challenges are as big as statistics and data science but every contribution and effort make a difference
  - Challenges are ongoing

Variation, continuum within and across countries and disciplines
Opportunities: the how and collaboration
Observe, listen, communicate, reflect, collaborate

- Data science gives opportunities to renew push for authentic learning that reflects practice of ‘greater statistics’ and ‘greater data science’
  - Parallels and commonalities must be constantly and strongly emphasized

- Enable coherent development
- Authentic working with other disciplines
- ASSESSMENT is key
  - Authentic and balance for efficient effectiveness
  - Real contexts, real data, complex data
  - Technology resources for learning and assessment
- Authentic collaboration & sharing

Thank you and here’s to statistics and data science!