

2014 SCHILD eCOTS 1

**Panel:  
TEACHING BIG DATA**

---

**Comments**

by  
**Milo Schild**  
Augsburg College, USA  
Electronic Conference on Teaching Statistics  
(E-COTS)  
May 20, 2014.  
[www.StatLit.org/pdf/2014-Schild-eCOTS2-Slides.pdf](http://www.StatLit.org/pdf/2014-Schild-eCOTS2-Slides.pdf)

2

**Milo Schild:  
Position and Background**

---

I think that making programming a dominant part of an undergraduate statistics major or minor is misguided.

I'm not opposed to programming. I am chair of an MIS major within Business. I spent 10 years in actuarial at a national insurance company using APL and SAS. I spent 5 years working with data streams from satellites.

Inserting more programming means that something else must be omitted. We need to reflect on what is missing.

3

**Do Others Disciplines  
Feature Programming?**

---

Other quantitative disciplines (Actuarial Science, Management Science, Physics) don't list programming as a dominant part of their major or minor.

In Colombia's MS in Actuarial Science\*, none of the eight required courses involve programming. Only two of the 32 electives involve programming (S-Plus).

Actuaries are more likely than statisticians to have to program and work with large amounts of data.

\*Source: <http://ce.columbia.edu/actuarial-science/courses>

4

**Statistical Practitioners:  
Big Data & Computing**

---

2014 ASA Conference on Statistical Practice.  
Fraction of talks that involve big data or computing:

**MINORITY: 5-20%**  
Posters: PS1 (2/16), PS2 (1/17), and PS3 (1/18)  
Concurrent Sessions: 7/43

**MAJORITY: 60-100%**  
Tutorials: 2/3  
Short Courses: 6/9  
Practical Computing Exhibits: 3/3  
[www.amstat.org/meetings/csp/2014/onlineprogram/index.cfm](http://www.amstat.org/meetings/csp/2014/onlineprogram/index.cfm)

5

**Computing Trumps  
Causation and Confounding**

---

Search ASA website – [www.amstat.org](http://www.amstat.org) – May, 2014:

Matches	Search Word or Phrase
34,500	Data
8,050	Computer/computing
2,710	programming
2,320	Causal/causality
1,080	"Observational studies"
940	"big data"
352	"directed graphs", "directed acyclic graphs"
220	"Structural equation modeling" or SEM
182	Coincidence
175	Confound/confounder/confounding

6

**Programming Trumps  
Causation and Confounding**

---

2014 Draft Guidelines for Statistical Science\*

*Required Undergrad topics: Statistics, **Programming**, Data, Mathematics and Communications.*

Data is mentioned 55 times [Programming: 8 times]  
Computer/computing is mentioned 15 times.

Causal inference is one of 7 topics under Modeling.  
Confounding is one of 8 topics under DOE.  
Not enough distinction between major and minor.

\* Draft dated May 11, 2014 via ISOSTAT. [www.amstat.org/education/curriculumguidelines.cfm](http://www.amstat.org/education/curriculumguidelines.cfm)

7

**What are We Ignoring???**

---

**Judea Pearl sponsors ASA Causality in Statistical Education Award. A \$10,000 award in 2014.**

Why is this award necessary? Because we don't want to talk about causation in observational studies.

Most students taking intro stats or a statistics minor are in majors that use observational data to identify causes.

*Randomization, Programming and Big Data* are important topics that allow statistical educators to ignore the three 'elephants': **Causality, Confounding and Context.**

8

**Recent Papers/Talks on Causality & Confounding**

---

PAPERS:

Pearl, J. (2014). Statistics and Causality: Separated to Reunite. Commentary. Health Service Research.

Tintle et al. (2013). Challenging the state of the art in post-introductory statistics. ISI

TALKS/POSTERS:

Deveaux, R. and D. Kaplan (2014). Statistics for the 21st Century: Are we Teaching the Right Course? ECOTS

Schild, M and D. Kaplan (2011). Modeling in Context: Teaching Confounding & Adjustment ... USCOTS