GAISE 2016: ANALYSIS AND RECOMMENDATION

Milo Schield
Visiting Professor, New College of Florida

Abstract:
The GAISE 2016 update marked a major change in the recommendations for statistical education by including multivariate thinking and confounding: two topics that were totally absent from prior recommendations. The justification for these topics is analyzed along with their subsequent impact. Emerging trends summarized by Burrill and Pfannkuch are reviewed. It seems obvious that no single course can encompass the breadth and depth involved in the GAISE 2016 recommendations and the four emerging trends. Based on the needs of our students, future GAISE updates should consider supporting three flavors of introductory statistics: consumer statistics (observational studies and confounding), producer statistics (traditional statistical inference) and data science statistics (computational thinking).

1. GAISE 2016 Update

GAISE is the acronym for Guidelines for the Assessment and Instruction in Statistical Education. GAISE (2005) introduced these guidelines. GAISE 2016 provided the first update to the guidelines.

The GAISE 2016 update contained two new ideas:

- Multivariable thinking: This new goal was presented as an aspect of "Teach statistical thinking".
- Confounding: This new goal was presented as an aspect of multivariable thinking in Appendix B.

Neither of these ideas were mentioned in GAISE 2005. Neither were mentioned in the McKenzie (2005) list of the top 30 topics in statistical education.

The justification for introducing multivariable thinking was presented as follows:

> "Give students experience with multivariable thinking. We live in a complex world in which the answer to a question often depends on many factors. Students will encounter such situations within their own fields of study and everyday lives. We must prepare our students to answer challenging questions that require them to investigate and explore relationships among many variables. Doing so will help them to appreciate the value of statistical thinking and methods."

The connection between multivariable thinking and confounding was presented as follows:

> In this appendix we describe simple examples where a third factor clouds the association between two other variables. Simple approaches (such as stratification) can help to discern the true associations. Stratification requires no advanced methods, nor even any inference,

Appendix A of this paper summarizes the highlights of the GAISE 2016 update. Appendix B presents the 20 instances of 'confound'. Nine of the 20 are found in Appendix B of the 2016 GAISE update entitled "Multivariable Thinking". Appendix C of this paper summarizes the frequency of selected words and phrases such as thinking. Of the 80 lines that contain an instance of "thinking," 67 are preceded by adjectives: statistical (40), multivariable (18), inferential (4), computational (2), collective (1), critical (1) and mathematical (1).

2. GAISE 2016 Update: Analysis and Impact

Although the GAISE 2016 doesn't give a definition of statistical thinking or multivariate thinking, it does identify many of the elements. Arguably, the biggest weaknesses in the GAISE 2016 update are these:
(1) What is the connection between multivariate thinking and confounding. Multivariate thinking is necessary to understand confounding, but confounding is not necessary to teach multivariate thinking.

One can teach multivariate thinking without addressing confounding. One can teach predictive analytics without ever mentioning confounding. The fact that adding an additional variable to a multivariable regression may change the sign of an existing predictor is all but irrelevant so long as R-squared increases without overfitting. (2) What is the motivation for statistical educators to address confounding given their motivation to teach multivariable thinking? The unwillingness of statistical educators to focus on confounding may reflect their allegiance to mathematical (deductive) reasoning. Schield (2013). Schield (2017) argues that the GAISE 2016 recommendations endorse a confounder-based statistical literacy.

GAISE 2016 has had some positive impact. One example is the 5th edition of Statistics by Agresti, Franklin and Klingenberg. In chapter 3, they analyze the association between two categorical variables, the relationship between two quantitative variables, and linear regression. In the last section they present Cautions in Analyzing Associations. They finish that section with Correlation Does Not Imply Causation, Simpson's Paradox, The Effect of Lurking Variables on Association, and Confounding.

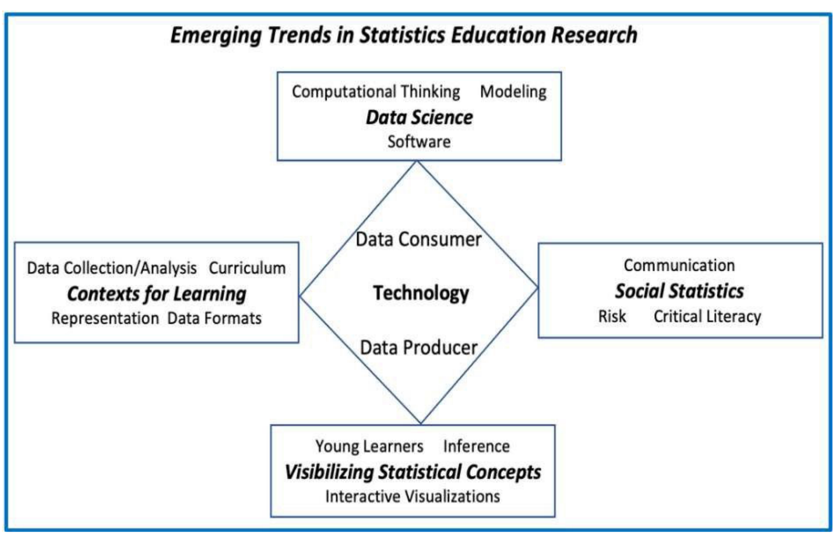
Their succinct presentation provides real-world examples of how a statistical association can be confused (confounded) by a lurking variable (a measured confounder). Their presentation follows the advice of the 2016 GAISE update in using selection to illustrate the influence of a lurking variable.

Note that the GAISE 2016 update never used the phrases "Correlation does not imply causation" or Association does not imply causation" thereby bypassing the ambiguity in the word imply. Whereas Agresti et al use the mathematical interpretation of "imply" which means "is sufficient to conclude". This mathematical usage ignores the everyday interpretation of "imply" which is "supports". Students may mistakenly conclude that statistical educators are claiming that a statistical association can never give any support for inferring causation. Students know that this claim is false and may dismiss everything involving the relationship between association and causation in observational studies.

Agresti et al (2023) return to confounding in section 10.5 "Adjusting for the Effects of Other Variables" where they introduce "statistical control."

3. Emerging Trends

Burrell and Pfannkuch (2023) did a masterful job of soliciting emerging trends from 24 leading statistical educators (along with 800 related references) and summarizing these emerging trends into four categories shown as follows:



The emphasis on computer-related topics (computational thinking, data science, etc.) is readily apparent.

4. Analysis

It seems obvious that there is no way a single course can cover the GAISE 2016 recommendations (for multivariate thinking and confounding) and include the subsequent emerging trends involving data science. Indeed, the idea that one course could satisfy the statistical needs of most college graduates seems idealistic.

Rossman (2007, slide 11) said, "You simply can't achieve these [GAISE 2005 statistical literacy] goals in one course if you also teach a long list of methods. Most students would be better served by a Stat 100 [statistical literacy] course than a Stat 101 [statistical methods] course."

5. Recommendation

Future GAISE updates should support offering at least three different introductory courses based on the statistical needs of three different groups of students.

- Consumer statistical literacy: designed to meet the needs of students in non-quantitative majors (majors that don't require a particular quantitative course) and in quantitative majors that focus primarily on observational studies (sociology, social work, and social epidemiology, as well business management and marketing). This course should be broadly based: association vs. causation (disparity vs. discrimination), and experiments versus observational studies. It should focus primarily on observational studies since these are most common in the everyday media. As such, it should address those things that control or ward off confounding such as effect size, study design (RCT), selection, ratios, and standardization (weighted averages). Computer-based regression should be optional. Standardization allows students to calculate what it means to take something into account quantitatively. Doing multivariate regression with computer software should be optional. Since many of these students will still encounter sample surveys, this course must cover the highlights of random sampling, expected error (margin of error), and since they will encounter the phrase statistical significance, teach that as the absence of overlap in confidence intervals.
- Producer statistical inference: designed to meet the needs of students in the social sciences where random assignment is involved (psychology, etc.) or who may be involved in conducting or analyzing studies or surveys. As such, it should focus on randomness and expected error: standard error, margin of error, confidence intervals and hypothesis tests. This may involve analytic solutions or the use of computer-based resampling.
- Data-science statistics: designed to meet the statistical needs for students in data science, computer science and business analytics. In addition to data acquisition and data wrangling, these students should use a programming language (R-studio, etc.) to generate descriptive statistics, tables of counts and percentages, multivariate regression, confidence intervals and hypothesis tests (analytically and using resampling). They should generate appropriate data displays.

Statistical educators should support offering three different introductory courses based on the different needs of our students in order to help achieve the McNaughton (1996) goal for statistical education: "to help students appreciate the vital role of the field of statistics in empirical research."

Note: This paper was submitted for publication in the 2023 ASA Proceedings of the Section on Statistics and Data Science Education. It was not published since the title and content were substantially different from what was proposed in spring 2023 and presented at JSM. It is self-published as a technical report.

Bibliography

ASA Guidelines for Assessment and Instruction in Statistical Education (GAISE 2005). College report: https://www.amstat.org/docs/default-source/amstat-documents/2005gaisecollege_full.pdf

ASA GAISE Update (2016). 2016 Update to the ASA GAISE Guidelines. Copy at https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf

Burrill, Gail and Maxine Pfankuch (2023). Emerging Trends in Statistical Education. https://www.isi-web.org/webinar/iase-emerging-trends-statistics-education-2017-2022-what-next

De Veaux, R. (2015), "What's wrong with Stat 101?" USCOTS. Available at www.causeweb.org/cause/sites/default/files/uscots/uscots15/presentations/DickOpening.pptx

McKenzie (2004). Conveying the Core Concepts. ASA Copy at www.statlit.org/pdf/2004mckenzieasa.pdf

McNaughton, Donald (1996). Goals of Your Introductory Statistics Course. Copy at https://www.matstat.com/teach/p0002.htm

Moore, D. (1997). Statistical Literacy and Statistical Competence in the 21st Century. ASA. Slides only. Copy at www.statlit.org/pdf/1997MooreASAslides.pdf

Moore, D. (1998). Statistical Literacy and Statistical Competence in the 21st Century. Making Statistics More Effective in Schools of Business. Abstract at www.statlit.org/pdf/1998MooreMSMESB.pdf

Rossman, Alan (2007). Seven Challenges for the Undergraduate Statistics Curriculum. USCOTS Plenary Address. Slides at www.statlit.org/pdf/2007RossmanUSCOTS6up.pdf

Schield, M. (2013). Statistics Education: Steadfast or Stubborn? 2013 ASA *Proceedings of the Section on Statistical Education*. P. 1005-1017. Copy at www.StatLit.org/pdf/2013-Schield-ASA.pdf

Schield, M. (2017). GAISE 2016 Promotes Statistical Literacy. *Statistics Education Research Journal*, 16(1), 46-50, http://iase-web.org/Publications.php?p=SERJ

Wasserstein, R. L., and Lazar, N. A. (2016), "The ASA's Statement on p-Values: Context, Process, and Purpose," *The American Statistician,* 70 p. 129-131. http://dx.doi.org/10.1080/00031305.2016.1154108,

Wasserstein, Ronald (2016). "ASA Statement on Statistical Significance and P-values." *The American Statistician,* 70, P. 131-133. http://dx.doi.org/10.1080/00031305.2016.1154108,

Wasserstein, R., Schirm, A. and Lazar, N. (2019) Editorial: "Moving to a World Beyond '$p < 0.05$'", *The American Statistician* 73(S1): 1-19. DOI: http://dx.doi.org/10.1080/00031305.2019.1583913

Appendix A: Excerpts from the GAISE 2016 Update

Executive Summary

"The revised recommendations are:

1. Teach statistical thinking.
   _ Teach statistics as an investigative process of problem-solving and decision making.
   _ Give students experience with multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning."

"We also created some new appendices to provide details about the evolution of introductory statistics courses; examples involving multivariable thinking; and ideas for implementing the GAISE recommendations in a variety of different learning environments."

"In addition to these six recommendations, which remain central, we suggest two new emphases for the first recommendation (teach statistical thinking) that reflect modern practice and take advantage of widely available technologies:

a. Teach statistics as an investigative process of problem-solving and decision-making. Students should not leave their introductory statistics course with the mistaken impression that statistics consists of an unrelated collection of formulas and methods. Rather, students should understand that statistics is a problem-solving and decision-making process that is fundamental to scientific inquiry and essential for making sound decisions.

b. Give students experience with multivariable thinking. We live in a complex world in which the answer to a question often depends on many factors. Students will encounter such situations within their own fields of study and everyday lives. We must prepare our students to answer challenging questions that require them to investigate and explore relationships among many variables. Doing so will help them to appreciate the value of statistical thinking and methods."

Goals for Students in Introductory Statistics Courses

"The desired result of all introductory statistics courses is to produce statistically educated students, which means that students should develop the ability to think statistically. The following goals reflect major strands in the collective thinking expressed in the statistics education literature. They summarize what a student should know and understand at the conclusion of a first course in statistics.

1. Students should become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted.
2. Students should be able to recognize questions for which the investigative process in statistics would be useful and should be able to answer questions using the investigative process.
3. Students should be able to produce graphical displays and numerical summaries and interpret what graphs do and do not reveal.
4. Students should recognize and be able to explain the central role of variability in the field of statistics.
5. Students should recognize and be able to explain the central role of randomness in designing studies and drawing conclusions.
6. Students should gain experience with how statistical models, including multivariable models, are used.
7. Students should demonstrate an understanding of, and ability to use, basic ideas of statistical inference, both hypothesis tests and interval estimation, in a variety of settings.
8. Students should be able to interpret and draw conclusions from standard output from statistical software packages.
9. Students should demonstrate an awareness of ethical issues associated with sound statistical practice."

Give students experience with multivariable thinking. (P. 14)

"To illustrate the power of multivariable thinking and modeling, consider an example that shows how accounting for the percentage of students taking the SAT exam in a state completely changes the conclusion that would be drawn about the relationship between average SAT score and average teacher salary in the state."

"De Veaux (2015) has also challenged statistics educators to think about how to improve introductory courses by, among other things, emphasizing the multivariate nature of the discipline. He calls for the motivation of univariate questions to arise from more complex models, and he illustrates how this can be done with examples that highlight (a) the relationship between diamond price and color, and how this relationship changes when carat weight is taken into account, and (b) the relationship between the presence or absence of a fireplace and the price of a home in New England, and how this relationship also changes markedly when square footage is taken into account."

"Kaplan's, Horton's, and De Veaux's examples illustrate that instructors do not need to go into detail about multivariable modeling in order to provide students with an appreciation for the need to consider how multiple variables interact. Students can explore and investigate such relationships by being presented with interesting questions from rich datasets and then producing appropriate graphical displays. These examples also give rise to discussions of how confounding plays an important role in determining the appropriate scope of conclusions to be drawn from such data."

Appendix B: All references in the GAISE 2016 update to "confound"

There are 20 uses of "**confound**" (bold) or <u>equivalent</u> (underscore) in the GAISE 2016 update.

P. 11: "Goal 9: Students should demonstrate an awareness of ethical issues associated with sound statistical practice. "With large datasets containing many variables, especially from observational studies, understanding of **confounding** and multiple testing false positive rates becomes even more relevant."

P. 15: "These examples also give rise to discussions of how **confounding** plays an important role in determining the appropriate scope of conclusions to be drawn from such data."

P. 34: "GAISE Update Appendix B: Multivariable Thinking"
"The 2014 ASA guidelines for undergraduate programs in statistics recommend that students obtain a clear understanding of principles of statistical design and tools to assess and account for the possible impact of other measured and unmeasured **confounding** variables (ASA 2014). An introductory statistics course cannot cover these topics in depth, but it is important to expose students to them even in their first course (Meng 2011). Perhaps the best place to start is to consider how <u>a third variable</u> can change our understanding of <u>the relationship between</u> two variables.

"In this appendix we describe simple examples where <u>a third factor</u> clouds the association between two other variables. Simple approaches (such as stratification) can help to discern the true associations. Stratification requires no advanced methods, nor even any inference, though some instructors may incorporate other related concepts and approaches such as multiple regression. These examples can help to introduce students to techniques <u>for assessing relationships between more than two variables</u>. Including one or more multivariable examples early in an introductory statistics course may help to prepare students to deal with more than one or two variables at a time and examples of observational (or "found") data that arise more commonly than results from randomized comparisons."

P. 99. Smoking on Wickam

P. 36: "This example represents a classic example of <u>Simpson's paradox</u> (Simpson 1951; Norton and Divine 2015). For all of the subjects, smoking appears to be "protective," but within each age group smokers have a higher probability of dying than non-smokers."

P. 37: What should we conclude? After <u>controlling for</u> age, smokers have a higher rate of mortality than non-smokers in this study.

P. 38: "Simple methods such as stratification can allow students to think beyond two dimensions and reveal effects of **confounding** variables. Introducing this thought process early on helps students easily transition to analyses involving multiple explanatory variables."

P. 38: SAT Scores and Teacher Salaries.
"But the real story is hidden behind one of the "other factors" that we warn students about but do not generally teach how to address! The proportion of students taking the SAT varies dramatically between states, as do teacher salaries. In the Midwest and Plains states, where teacher salaries tend to be lower, relatively few high school students take the SAT. Those that do are typically the top students who are planning to attend college out of state, while many others take the alternative standardized ACT test that is required for their state. For each of the three groups of states defined by the fraction taking the SAT, the association is non-negative. The net result is that the fraction taking the SAT is a **confounding** factor."

P. 38: "This problem is a continuous example of Simpson's paradox. Statistical thinking with an appreciation of Simpson's paradox would alert a student to look for the hidden **confounding** variables. To tackle this problem, students need to know that multivariable modeling exists but not all aspects of how it can be utilized."

P. 39: "Within an introductory statistics course, the use of stratification by a potential confounder is easy to implement. By splitting states up into groups based on the fraction of students taking the SAT it is possible to account for this **confounder** and use bivariate methods to assess the relationship for each of the groups."

P. 39: "It's important to have students look for possible **confounding** factors when the relationship isn't what they expect, but it is also important when the relationship is what is expected. It's not always possible to stratify by factors (particularly if important confounders are not collected)."

Closing Thoughts
P. 41: "Multivariable thinking is critical to make sense of the observational data around us. This type of thinking might be introduced in stages:
1. learn to identify observational studies,
2. explain why randomized assignment to treatment improves the situation,
3. learn to be wary of cause-and-effect conclusions from observational studies,
4. learn to consider potential **confounding** factors and explain why they might be confounding factors,
5. use simple approaches (such as stratification) to address **confounding**."

P 42: "Multivariable models are necessary when we want to model many aspects of the world more realistically. The real world is complex and can't be described well by one or two variables. If students do not have exposure to simple tools for disentangling complex relationships, they may dismiss statistics as an old-school discipline only suitable for small sample inference of randomized studies."

P. 105 Footnote 50: " For an observational study which assessed the association between coffee drinking and cancer, smoking status could mask (or "**confound**") the relationship, since smoking could be associated with both coffee drinking and cancer."

P. 113: "Use the language of statistics to critique the statement by Dr. Koenig and the claim, suggested by the article, that religious faith and practice help people fight depression. You will want to select some of the following words in your critique: observational study, experiment, blind, double blind, precision, bias, sample, spurious, **confounding**, causation, association, random, valid, and reliable."

P. 120 "Item 36:
1. Is this an observational study or experiment? Defend your answer.
2. What are the explanatory and response variables?
3. Identify a potential **confounding** variable in this work.
4. Is this a matched-pair design? Defend your answer."

P. 122, ITEM 40: When conducting a randomized experiment, the original randomization of units to treatment groups breaks the association between
1. the explanatory variable and the response variable.
2. the explanatory variable and **confounding** variables.
3. the response variable and confounding variables.

P. 122, ITEM 41: When conducting a randomization test, the simulated re-randomization of units to treatment groups breaks the association between [the same three choices as Item 40].

Appendix C: References to Selected Keywords and Phrases

THINKING:
There are 80 lines that contain an instance of "thinking." Of these, 67 are preceded by adjectives: statistical (40), multivariable (18), inferential (4), computational (2), collective (1), critical (1) and mathematical (1). Of the remaining 13, four involve prepositions (of), four involve possessives (your), two involve appositives (the), one involves a conjunction (and), one is a verb and one starts the sentence.

STATISTICAL THINKING:
There are 42 instances of statistical thinking. Statistical thinking appears to be a higher-level concept than decision-making or multivariable thinking since it is the only one of these three phrases that appears in the six highest-level goals in prior GAISE reports and in this 2016 update. There is no attempt to define statistical thinking.

DECISION MAKING:
There are just five references to (six instances of) "decision-making". "Decision-making" appears as a subtitle in the 2016 highest-level summary. Of the six instances (bold added), four are identical: "Teach statistics as an investigative process of problem-solving and decision-making." (Pages 4, 6, 13 and 14). Pages 6, 9 and 14 introduce more details.

P. 6: "Teach statistics as an investigative process of problem-solving and **decision-making**. Students should not leave their introductory statistics course with the mistaken impression that statistics consists of an unrelated collection of formulas and methods. Rather, students should understand that statistics is a problem-solving and **decision-making** process that is fundamental to scientific inquiry and essential for making sound decisions."

P. 9: "While many questions do not have simple yes or no answers, knowing how to obtain or generate data that are relevant to the goals of a study is crucial to providing useful information that supports **decision-making** in the sciences, business, healthcare, law, the humanities, etc."

P. 14: "Statistics involves an investigative process of problem-solving and **decision-making**, which makes it a fundamental discipline in advancing both scientific discoveries and business and personal decisions."

MULTIVARIABLE AND MULTIVARIATE:
There are 30 instances of "**multivaria**" in the 2016 GAISE update: two to "multivariate" and 28 to "multivariable." Of these 28, 18 are instances of "multivariable thinking" while seven are instances of "multivariable model" (root of multivariable models and multivariable modeling)." There is one instance of "multivariable relationships," one instance of "multivariable examples", and one instance of "multivariable methods."

Aside from being the most common, "**multivariable** thinking" is arguably the most important of these five instances of "multivariable", since that is the only one of these that appears in the highest level summary, the only one that is the heading of a sections (p. 14), and the only one that is used as the title to one of the five appendices (P. 34). The reference on page 6 gives a justification for why students should experience multivariable thinking.

P 6: "Give students experience with **multivariable** thinking. We live in a complex world in which the answer to a question often depends on many factors. Students will encounter such situations within their own fields of study and everyday lives. We must prepare our students to answer challenging questions that require them to investigate and explore relationships among many variables. Doing so will help them to appreciate the value of statistical thinking and methods.

PROBLEM SOLVING:
Problems solving is mentioned 11 times. Of these 11, six appear in "problem solving and decision-making."

P 12: "We urge instructors of statistics to emphasize the practical **problem-solving** skills that are necessary to answer statistical questions."

P. 14: "Statistics involves an investigative process of **problem-solving** and decision-making, which makes it a fundamental discipline in advancing both scientific discoveries and business and personal decisions.

CONCEPTUAL
There are 19 instances of "conceptual." In all 19, "conceptual" is used as an adjective. "Conceptual" modifies "understanding (

CONCEPT
The root "concept" is included in 116 instances.

RELATIONSHIP

There are 51 sentences that include the word "relationship". Of these 11 involve "relationships".

Of the 11 instances of "relationships," three involve "relationship among" while two involve "relationships between."

Appendix E: ASA Statement on Statistical Significance and P-values

The 2016 update to the GAISE guidelines stated the following:

> "ASA's Statement on p-Values, which puts forward several important principles about hypothesis testing based on consensus among those in the statistical community, in an effort to improve the ways in which the statistical results of scientific studies are reported and interpreted." (Wassertstein and Lazar, 2016).

The Wassertstein-Lazar (2016) article was titled: "The ASA's Statement on P-values: Context, Process and Purpose." It was immediately followed by Wasserstein's (2016) paper: "ASA Statement on Statistical Significance and P-values." Notice the subtle difference. ASA is possessive in the first but not the second.

This second paper (Wasserstein 2016) gives the ASA position. It was signed at the bottom as "Edited by Ronald L. Wasserstein, Executive Director. On behalf of the American Statistical Association Board of Directors." It carried this acknowledgement: "The ASA Board of Directors thanks the following people for sharing their expertise and perspectives during the development of the statement."

In 2019, there was a third paper involving Wasserstein. This paper was more assertive: "it is time to stop using the term 'statistically significant' entirely." However this paper carried this disclaimer: "The editorial was written by the three editors acting as individuals and reflects their scientific views not an endorsed position of the American Statistical Association."