Introducing GAISE II: A Guideline for Precollege Statistics and Data Science Education

Christine Franklin^{1,2} Anna Bargagliotti³

¹Department of Statistics Franklin College of Arts and Sciences, University of Georgia, Athens, Georgia, United States of America,

²American Statistical Association, Alexandria, Virginia, United States of America,

³Department of Mathematics, Seaver College of Science and Engineering, Loyola Marymount University, Los Angeles, California, United States of America

Published on: Oct 29, 2020

DOI: <https://doi.org/10.1162/99608f92.246107bb>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Column Editor’s Note: In this column Anna Bargagliotti and Christine Franklin, two of the authors of the new *preK–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)*, provide an introduction to the new report. The new report highlights the importance of data science education already at the pre-college levels. Future articles in the *Minding the Future* column will showcase examples of data science activities that are suitable for school children at different stages of their education, from the earliest primary levels and up through secondary. In a companion column, a current high school student shares her perspective on the importance of data science courses at the pre-college level.

The Pre-K-12 Report discussed in this column is accessible [HERE](#).

Introduction

We are excited to introduce the Pre-K–12 (i.e., precollege) Guidelines for Assessment and Instruction in Statistics Education II: A Framework for Statistics and Data Science Education report (GAISE II). This report presents a set of recommendations toward data literacy at the elementary, middle, and high school level. GAISE II (Bargagliotti et al., 2020) incorporates enhancements and new skills needed for making sense of data today while maintaining the spirit of the original Pre-K–12 GAISE (Franklin et al., 2007), which we shall review briefly. Now more than ever, it is essential that all students leave secondary school prepared to live and work in a data-driven world, and this report outlines how to achieve this goal.

Never have data and statistical literacy been more essential than in 2020. Processing information associated with global issues such as the COVID-19 pandemic, extreme weather conditions and a changing planet, economic upturns and downturns, and important social issues such as the Black Lives Matter movement or the refugee crisis are a few cases in point. Data are presented through visualizations (whether interactive or not), reports from scientific studies (such as medical studies), and results from statistical models for the purpose of revealing patterns, predicting, and decision making. A statistically literate high school graduate needs to be able to critically evaluate stated conclusions from data for legitimacy and applicability. Steve Levitt, co-author of *Freakonomics*, addressed the necessity of statistical and data literacy with this quote from an October 2, 2019 podcast:

I believe that we owe it to our children to prepare them for the world that they will encounter—a world driven by data. Basic data fluency is a requirement not just for most good jobs, but also for navigating life more generally, whether it is in terms of financial literacy, making good choices about our own health, or knowing who and what to believe.

Driven by the digital revolution, data are now readily available to help gain insights and make recommendations on how to deal with world issues. Data can be invaluable, but only if used appropriately and in a proper context. Now is a critical time for the development of data literacy. Students need to gain the ability

to understand data, to arrive at plausible and reasonable decisions based on the data available, and to be responsible and ethical with the analytic tools available. As eloquently stated by Keller et al. (2020),

(A)s data are becoming the new currency across our economy, the University of VA research team emphasizes the obligation of data scientists to enlighten decision-makers on data acumen (literacy). The need to help consumers of research to understand the data and the role it plays in problem solving and policy development is important, as is building a data-savvy workforce...

Due to the importance of data in society, the development of data literacy should begin early in a person's education (Martinez & LaLonde, 2020). Today, many disciplines at the undergraduate level as well as many sectors of the economy and even specific jobs are increasingly reliant on data skills (Business Higher Education Forum, 2018). In addition, good data sense is needed to simply interpret the news we encounter and be able to participate in society in an educated manner. In the upcoming issues of *HDSR*, selected examples in GAISE II from each of the three developmental levels A, B, and C (as described following) will be presented to illustrate some of these essential concepts.

History, Overview, and Goals of GAISE II

The Guidelines for Assessment and Instruction in Statistics Education: A PreK–12 Curriculum Framework ([GAISE I](#)) was published in 2007. A seminal and visionary report, it advocated for the necessity of data and statistical literacy from the earliest grades. It provided a framework of recommendations for the evolution of statistical concepts and the development of foundational skills for statistical reasoning of students across the school years described as three developmental levels: A, B, and C. These levels are maintained in GAISE II and are roughly equivalent to elementary, middle, and high school, respectively. Since its initial publication, the original GAISE has significantly impacted the inclusion of statistics standards at the state and national level in the United States, and internationally. The report has been used internationally as a reference point for statistics education at the precollege level, including a [Spanish translation](#) of the first Pre-K–12 GAISE.

Google Scholar documents about 800 citations (as of this article's writing) for GAISE I in scholarly works; this in addition to numerous National Science Foundation grant projects and other professional science, technology, engineering, and mathematics (STEM) educational organizations' reports using the GAISE I recommendations.

GAISE I primarily focused on traditional data types from both quantitative to categorical variables and study designs using smaller data sets of samples from a population. Thirteen years later, data types have expanded beyond being classified as quantitative and categorical, thus necessitating the acquisition of different and new statistical skills. Today, for example, data can also be text posted on social media or highly structured (or unstructured) collections of pictures, sounds, or videos. Data are vast and readily available. Data are multidimensional, and data representations and visualizations often need to display many variables simultaneously.

GAISE II addresses the evolution of data types and skills needed to make sense of the wealth of data that confronts us. The new updates include:

1. The importance of questioning through each stage of the statistical problem-solving process (formulating a statistical investigative question, collecting or considering data, analyzing data, and interpreting results) and how this process remains at the forefront of statistical thinking;
2. The consideration of different data and variable types and the importance of carefully designing how primary data are collected or secondary data are considered to answer a statistical investigative question, the process of collecting and cleaning data, the interrogation of data, and the analysis of data;
3. The inclusion of multivariate thinking throughout all levels of K–12 education;
4. The role of probabilistic thinking in quantifying randomness used by students throughout all school levels;
5. The shifts and deepening of the role of technology throughout Pre-K–12 (school level) education;
6. The importance of how now, more than ever, statistical information is communicated.

Data in Our Lives Today in 2020

“It is critical that statisticians, or anyone who uses data, be more than just data crunchers. They should be data problem solvers who interrogate the data and utilize questioning throughout the statistical problem-solving process to make decisions with confidence, understanding that the art of communication with data is essential.” (GAISE II)

In early 2020, the world was suddenly disrupted by a new virus named SARS-CoV-2 that soon resulted in the COVID-19 pandemic. Our lives were turned upside down as, overnight, social isolation became mandatory, with people staying home from work and school, from exercising at the gym, eating in restaurants, and traveling. Suddenly the media was publishing article after article about the statistical models predicting the exponential number of cases and the number of deaths expected if mediation measures were not taken. GAISE II stresses the importance of communicating with data. It was stressed that these mediation actions were important to ‘flatten the curve,’ which was often depicted as illustrated in Figure 1. The goal was to delay the number of cases so the health care system would not be overburdened.

Figure 1. The Importance of Flattening the Curve.

National Forecast

Figure 2. Number of Deaths in USA from COVID-19 as of June 1, 2020.

From Centers for Disease Control (CDC) website (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html>).

The Centers for Disease Control (CDC) in the United States publishes graphs on a regular basis representing the number of deaths in the United States as a result of COVID-19. Figure 2 displays graphs as of June 1, 2020, forecasting deaths from this virus based on models developed by different groups of researchers. Included on the graph are 95% prediction intervals. Such graphics in the news underscore the importance of knowing how to ask questions about the information displayed in the graph. GAISE II emphasizes the statistical investigative process, which includes formulating questions, considering data, analyzing data, and then interpreting results. An important contribution of GAISE II is the articulation of how important questioning is throughout the process. For example, while considering and analyzing data, we might ask: What data were utilized and what

was known about their quality, and what assumptions were made to derive these intervals and how should these intervals be interpreted?

Key to reasoning with statistical models is understanding that these models require probabilistic thinking. Random variability in the data means these models will not yield a deterministic or definitive answer. Also, in developing statistical models, assumptions are made about the contextual situation, both statistical (such as distributional) and nonstatistical (such as individual behavior). These assumptions do not necessarily stay constant. As we look at the prediction intervals, the band of plausible values incorporates the expected variability condition upon the assumptions made. In this context, as the model is predicting the number of deaths will continue to increase, our hope is that the model will prove to be incorrect. This could occur if the individual behavior risk for contracting the virus is less than what the model assumed. This change in nonstatistical assumptions could also then affect statistical assumptions previously made.

Interrogation of data using questioning is essential (see no. 1 in the list provided earlier). How are data measured? Are data reliable? Are data reported or represented in misleading ways? As a consumer, can you recognize where corrections are needed? As an example, consider a graphical representation published by the Georgia Department of Public Health (Figure 3).

Figure 3. Top 5 Counties in GA with COVID-19 Over Time.

From [https://www.businessinsider.com/graph-shows-georgia-bungling-coronavirus-data-2020-](https://www.businessinsider.com/graph-shows-georgia-bungling-coronavirus-data-2020-5)

The time-plot representation (current when this article was written, but always a changing situation with COVID-19) shows Georgia's top five counties most impacted with the greatest number of COVID-19 cases over 15 days. At first glance, one observes that there is a decreasing trend over the days; in fact, a perfect decreasing trend for the five counties within each day. Upon closer examination, it is noticed that on the horizontal axis the days listed are not in consecutive order (the first day April 26 is listed after May 7) but jumbled. The categories within each day are not listed consistently in order by name but instead by the highest frequency to lowest frequency to always show a decreasing trend. Fortunately, consumers reading media reports of the published graph quickly caught the mistakes, and a corrected graph with apologies was issued hours later.

Not only is it important to ask questions about data and data visualizations, but also questions about how studies are conducted and reported in the media and professional journals. As this article was being written, decisions were not only being made about how our world will move forward from the pandemic but also about important issues such as global warming, social movements related to equity, and economic development. GAISE II provides a wealth of examples guiding students through the important questions to ask about study design.

A Future Driven by Data

An enormous amount of personal data is generated each day by an individual. Examples include surfing the World Wide Web, sending emails, posting to and browsing social media, paying bills online, using a grocery store shopper card, uploading photos, using Google Drive or Microsoft Teams, texting, using fitness devices, and the list goes on. It has been [estimated](#) that an individual generates on average 1.145 trillion MB of data each day. How can individuals' data be used to better society and quality of life? What are the ethical considerations with data collected on individuals? With such an abundance of data available for consideration and the technological tools available, what are the skills needed to organize the data, to clean the data, to make judgments about the appropriateness of the data for answering statistical investigative questions, and to make predictions?

As stated in GAISE II,

Data are used to tell a story. Statisticians see the world through data—data serve as models of reality. Statistical thinking and the statistical problem-solving process are foundational to exploring all data.

The vision GAISE II aims to convey is one in which students should feel confident in statistical reasoning, in making sense of data and in maintaining a healthy dose of skepticism to question the validity of evidence or information they receive. GAISE II presents a framework of essential concepts and 22 examples that support all students to gain an appreciation for the vital role of statistical reasoning and data science, and to acquire the essential life skill of data literacy. Some of these examples will be presented in upcoming issues of *HDSR*. It should be noted that subsequent recommendations for the college introductory statistics course are made in the

Guidelines for Assessment and Introduction in Statistics Education (GAISE) College Report (Carver et al., 2016). GAISE II lays the conceptual foundations for students to become practitioners and consumers of statistics and data science by the time they graduate high school. Following these recommendations provides a path forward for students to thrive in employment and in postsecondary schooling.

Disclosure Statement

Christine Franklin and Anna Bargagliotti have no financial or non-financial disclosures to share for this article.

References

- Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) report II*. American Statistical Association and National Council of Teachers of Mathematics.
- Business Higher Education Forum (BHEF). (2018). *The new foundational skills of the digital economy*. http://www.bhef.com/sites/default/files/BHEF_2018_New_Foundational_Skills.pdf
- Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., ... & Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report 2016*. American Statistical Association. <https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx>
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) report: A pre-K–12 curriculum framework*. American Statistical Association. <https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx>
- Keller, S. A., Shipp, S. S., Schroeder, A. D., & Korkmaz, G. (2020). Doing data science: A framework and case study. *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.2d83f7f5>
- Levitt, S. (2019, October 2). *America's math curriculum doesn't add up* [Audio podcast]. Freakonomics. <https://freakonomics.com/podcast/math-curriculum/>
- Martinez, W., & LaLonde, D. (2020). Data science for everyone starts in kindergarten: Strategies and initiatives from the American Statistical Association. *Harvard Data Science Review*, 2(3). <https://doi.org/10.1162/99608f92.7a9f2f4d>
-

©2020 Christine Franklin and Anna Bargagliotti. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Preview image for this article illustrated by Brenna Bastian.