The State of Computing in Introductory Statistics

Chelsey Legacy, Andrew Zieffler, Elizabeth Fry, & Laura Le

Department of Educational Psychology
University of Minnesota
Driven to Discover®

St. Catherine University

School of Public Health
University of Minnesota

INTRODUCTION

Computing with data is fundamental to contemporary statistical practice and scientific inquiry. The proliferation of data and the increased demand for a data-literate workforce has led to several calls for reforming the introductory statistics curriculum to give students broader experiences with computation and modern data structures (American Statistical Association, 2014; Horton, 2015; National Academies of Sciences, Engineering, and Medicine, 2018; Nolan & Temple Lang, 2010).

RESEARCH QUESTIONS

1. To what extent are ideas of statistical computing being integrated into the introductory statistics curricula?
2. Are students receiving experiences with modern data structures in the introductory statistics curricula?

Many introductory statistics courses are not providing students experiences with computation and data structures essential for modern scientific inquiry.

Take a picture for references and additional information

METHODS

- The Statistics Teaching Inventory (STI; Zieffler et al., 2012) was modified to better align with current recommendations for teaching introductory statistics (ASA, 2016) and to measure the extent to which computing and ideas of computational thinking were being embedded in the introductory statistics curriculum (Weintrop et. al, 2016).
- Think-aloud interviews were conducted using the modified instrument with three statisticians/statistics educators, which informed revision on several items.
- In Fall 2019, we sent email messages to three statistics education listservs inviting tertiary-level statistics instructors to complete the STI.
- The results of the 293 participants that responded are summarized in this poster.

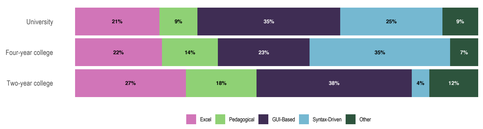
LIMITATIONS

- The voluntary, convenience sample employed in the study makes generalizations tenuous. The audience in the email listservs are all members of the statistics education community. We suspect that these results are positively biased.
- The item non-response might also positively bias the results.

FUTURE WORK

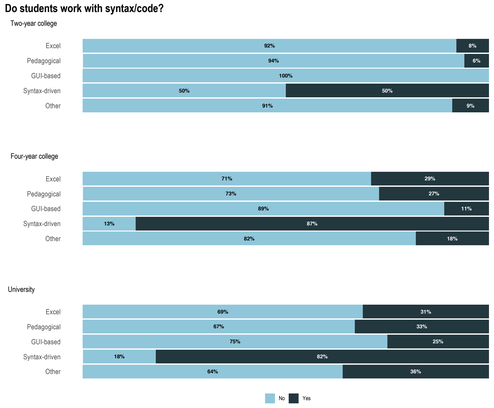
- We will compare results to previous administrations of the STI to gain insight into how the courses have changed over time.
- Administration of the STI will be extended to include secondary and graduate level statistics instructors.

What percentage of instructors use each type of software?



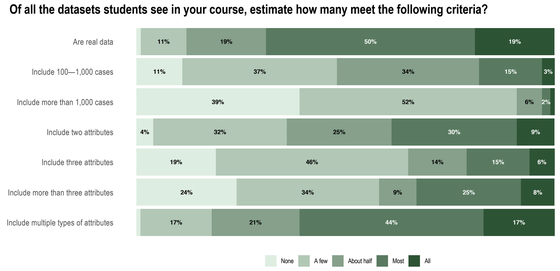
Excel and GUI-based software are popular choices across institution types. Syntax-driven softwares are more commonly adopted in four-year colleges and universities than in two-year colleges.

Do students work with syntax/code?



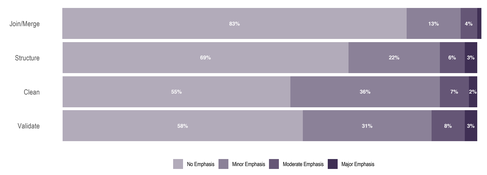
Coding is not commonly taught in introductory statistics courses. Instructors who adopt syntax-driven software are the ones primarily teaching coding, but not all of them.

Of all the datasets students see in your course, estimate how many meet the following criteria?



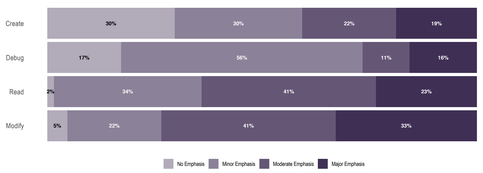
Most instructors use real data as recommended by GAISE (ASA, 2016). The majority of datasets include multiple types of attributes (e.g., categorical and quantitative attributes), but tend to be small (less than 1,000 cases, fewer than three attributes).

How much emphasis is placed on having students do each of the following with data?



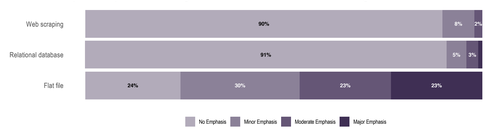
Manipulating data to get it into a useable form is not emphasized, yet it is an important part of data analysis.

How much emphasis is placed on each of the following when students work with syntax/code?



Instructors who teach coding tend not to emphasize debugging nor creation of syntax—higher-order skills associated with deeper and more critical thinking (e.g., DeLiema at al., 2020; Weintrop et al. 2016).

How much emphasis is placed on having students work with each of the following data structures?



Students encounter flat files (e.g., CSV) more often than relational databases and web scraping.