Interactive Visualization as a Tool to Teach Mathematical Concepts to OHSU Data Science Institute Students

Jessica Minnier, PhD, OHSU-PSU School of Public Health
Ted Laderas, PhD, Department of Medical Informatics and Clinical Epidemiology

Introduction

Students find statistical concepts hard to learn
Visual Explanations are helpful for diverse learners

Goals: Interactive Visualizations as a Tool

Understand relationships in data
Interactive plots allow for exploration of multi-variable relationships

Our Audience

Adult learners for OHSU Data Science Institute (3 day workshop)
Highly heterogeneous backgrounds and levels of quantitative experience

Methods:

Used R, LearnR, and Shiny to develop interactive tutorials
Tutorials don’t require downloading software
Students go to website, can save their progress

Availability

Implemented as a LearnR tutorial, which can be deployed as a website, or on student’s computer (requires R/Rstudio)
https://github.com/laderast/dsiexplore

Results

Students were empowered to learn
Students liked the visualizations
“Very well done and methodical treatment - the sliders were great!”
Students felt engaged with the subject
“Explanation of key statistical concepts was effective and really made me want to learn more.”
Post-workshop survey: 95% of learners gained practical knowledge (n=22)

Impact and Effectiveness

Pros:
- More accessible to beginners
- Makes mathematical concepts more memorable
- Sparks discussion about issues in datasets
- Engages students in the scientific discovery process
- Empowers students to believe they can implement statistical analysis

Cons:
- Advanced students may require development of more challenging activities
- Requires programming skills to implement
- Visualizations must be tested for effectiveness

References- Please refer to handout

Acknowledgements - partially funded with an NLM Training Supplement for Data Science - 3T15LM007088-26S1

Example: Exploring Simpson's Paradox

Proportion of smokers alive is higher for smokers compared to non-smokers
Why is this so? Distribution of age of participants is behind this!
Slider allows students to explore relationship between age and outcome
Older population in cohort has a higher proportion of deaths

Example: Exploring Correlation & Data Artifacts in NHANES

Correlation is strongly linear for weight vs. BMI, but more complex relationship appears for height vs. age
Interactive coding teaches data analysis, visualization in R
Drop down options allow exploration of variable relationships
Sparks questions:
- Why are height/age linearly correlated up to a certain age?
- Why are many subjects age 80?