Data Warehousing for Interactive Visualization of Student Data

Kim Speerschneider^{1,2,*}

1. University at Albany, SUNY

2. Excelsior College

*Contact author: kimkspeer@gmail.com

Keywords: Data Warehousing, Reproducible Research, Visualization

This talk will explore the intersection of data warehousing and data visualization. Data warehousing allows us to collect a depth of information about longitudinal trends while also retaining student-level details. In this way, the interactive graphics we share with our colleagues can be manipulated to serve specific needs, but the process remains transparent and reproducible. I will describe how, at our institution, we warehouse student data and in turn share this data with our colleagues. Specifically, I will describe some of the warehousing functions in our privately used package, ecir, and why we created a package specific to our own institution and data. Additionally, I will share some of the visualizations that we share throughout our institution that make use of ggplot2 (Wickham, 2009), shiny (RStudio, Inc., 2014), and googleVis (Gesmann et al., 2013) and explain how data warehousing allows us to tell a more complete story of our students' experiences. The figure below shows one such visualization. Although applications here are in the field of Education, this talk is relevant to anyone interested in how data warehousing can expand how data is shared, particularly with highly interactive graphics.

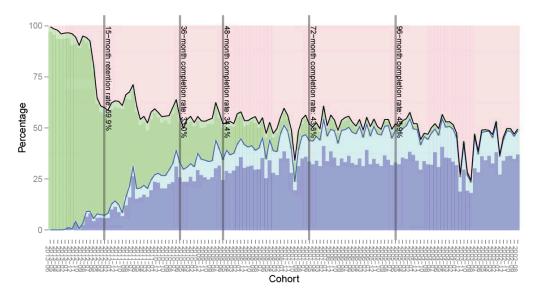


Figure 3.2: Institutional Beginning Retention by Cohort

References

Gesmann, M, de Castillo, D., & Cheng, J. (2013). googleVis: Interface between R and the Google Chart Tools. R package version 0.4.7

RStudio, Inc. (2014). shiny: Web Application Framework for R. R package version 0.9.1

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer New York.