ETD: A Design Pattern for Building Web-Based Analytics Dashboards in *R*

Bhaskar Rao[1*], Eric Colson[1], Alan Eng[1],

1. Stitch Fix Inc.
*Contact author: bhaskar@stitchfix.com

Keywords: reports, shiny, analytics dashboards, design pattern, ETD.

ETD, an abbreviation for extract-transform-display, is a design pattern that the Stitch Fix data team observed while building reporting and analytics dashboards in *R*, using the **Shiny** package. Formalizing this pattern reduces the complexity involved in creating web-based dashboards. It also provides a templatized approach for creating dashboards and promotes re-use and encapsulation of *R* and data extraction code.

Developing interactive web-based dashboards typically involves three distinct stages. First, the 'Extract' stage pulls data from a data source - typically a relational database using *SQL*. This extracted data is pulled into an *R* data structure where complex calculations can be applied (e.g. cross-tabulation, cleansing routines, conditional probabilities, complex metric definitions, …etc.). This is the 'Transform' stage. Finally, the transformed information is displayed using standard *R* visualization packages like **ggplot** or **googleVis.** This is the 'Display' stage. This 3-staged workflow is analogous to the extract-transform-load[2] (ETL) pattern prevalent in data warehousing. The important distinction is the final stage where, rather than loading the data for system consumption, we are rendering the information for end-user consumption.

Many data scientists lack the requisite skills to build web-based analytics dashboards. However, packages like **Shiny** provide a layer of abstraction that enables them to build web-based application in *R* without having to learn *HTML, Javascript* and *CSS*. Our ETD design pattern takes it one step further by taming the complexities of **Shiny**'s reactive programming framework and making it possible to templatize the creation of typical analytics dashboards. Using the ETD pattern in the development of Shiny dashboards helps our data scientists build complex web-based dashboards quickly while keeping our *R* code-base modular, clean, and extensible.

References
[1] RStudio (2013). Shiny by RStudio, http://shiny.rstudio.com/articles/
[2] Kimball, Ralph, and Joe Caserta. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Indianapolis, IN: Wiley, 2004. Print.