Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *doi: 10.18637/jss.v000.i00*

iNZight: A Graphical User Interface for Visualisation and Exploration of Data with R

Tom Elliott
Victoria University of Wellington

Chris Wild
University of Auckland

Daniel Barnett
University of Auckland

Andrew Sporle
University of Auckland

Abstract

Getting started with data science is a daunting task, particularly when it requires a large amount of coding before you can even start looking at data. Graphical user interfaces (GUIs) have often been used as a way of proving novice users the ability to interact with complex systems without the need for coding. However, many of these themselves have steep learning curves to understand how to make the software do what's needed, and do not provide a pathway to more standard and flexible methods, such as coding. **iNZight** is a GUI based tool written in R that provides students of statistics and data science the opportunity to interact with data and explore without first learning to code. The tool is designed to be easy to use, with logical interactions and clever defaults. However, it also provides some more complex features to manipulate and analyse data, and further provides a code history of the actions performed, creating a pathway between GUI and learning to code for those interested in progressing into the more open and exciting world of data science.

Keywords: GUI, statistical software, statistical education, R.

1. Introduction

The R programming environment (R Core Team 2020) is used throughout statistics and data science due to it being open source, easy to learn, and backed by a huge package repository to solve even the most unique of problems.

- several graphical user interfaces (GUIs) have been developed over the years to ease

access to advanced features of R: graphs, hypothesis tests, etc

- eg: Fox (2005), Fellows (2012)
- these tend to work by asking users to first choose an action, then fill in the fields (including variable choice)
- this requires some higher level of understanding to get useful information out (i.e., need to know what a "t-test" is)

An alternative approach is to work variable-first, such that users choose variables they are interested in, and then choose from an automatically curated list of options to perform. **iNZight** uses this approach, and presents users with a exploration-focussed interface.

- focus is on visualising, removing the need for any basic understanding of statistical procedures to get started with data visualisation
- like other GUIs, there's a code component: however, **iNZight**'s is more *behind-the-scenes*; useful for seeing a history of what you've done, with a little emphasis on editing a command to see how it changes things
- goal is to develop data exploration skills before moving them on to coding (high-level to low-level)

Due in part to its ease of use, **iNZight** has been adopted throughout New Zealand's statistical education program. Final year high school students are introduced to basic statistical concepts using **iNZight**, including a foray into time series analysis. Universities accross the country have also begun to use **iNZight** in both introductory and some advanced statistics courses. This paper provides an overview of some of the main features of **iNZight**, along with technical details, an introduction to its *Add-on* system, and description of the install process.

2. A tour of iNZight's features

- novice-oriented
- attempt to make it intuitive (drag-and-drop, etc)
- variable-first approach

2.1. Loading data

- R reads many file types—many different functions in a range of packages
- often different syntax/arguments for each
- **iNZight** uses file extension to guess file type and load it automatically
- e.g., load data window in FIGURE (with data preview)

Acknowledgments

iNZight is a free to use, open source software. The work would not have been possible without the support of the University of Auckland, Census at School, ..., Statistics New Zealand, and the Australian Bureau of Statistics.

References

Fellows I (2012). "Deducer: A Data Analysis GUI for R." *Journal of Statistical Software, Articles*, **49**(8), 1–15. ISSN 1548-7660. doi:10.18637/jss.v049.i08. URL https://www.jstatsoft.org/v049/i08.

Fox J (2005). "The R Commander: A Basic Statistics Graphical User Interface to R." *Journal of Statistical Software*, **14**(9), 1–42. URL https://www.jstatsoft.org/article/view/v014i09.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Affiliation:

Tom Elliott
School of Health
Victoria University of Wellington
Wellington, New Zealand
and
Department of Statistics (Honorary)
University of Auckland
Auckland, New Zealand
E-mail: tom.elliott@auckland.ac.nz
URL: https://people.wgtn.ac.nz/tom.elliott

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:10.18637/jss.v000.i00 *Accepted:* yyyy-mm-dd