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)

2.2. Creating graphs

- plot: central to the interface -> *look first*
- drag and drop variable onto prominent boxes (or use dropdowns)
- **iNZight** draws appropriate 'default' graph based on variable type - TABLE
- subsetting boxes allow easy faceting of 1 or 2 variables, slider to view specific subsets; numeric variables cut into 4
- advanced modification avaialable from Add to Plot menu -> colour/size/alternative graphs/lines/axes/labels/etc

2.3. Summaries and inference

- again: *look first*
- Get Summary and Get Inference buttons produce output based on chosen vars (in graph)
- summary includes summary stats: mean, quantiles, variance, etc
- inference provides inference information: confidence intervals for means/differences, etc
- And hypothesis tests: shows a list of tests applicable to the chosen vars/graph
- Normal theory and Bootstrap (using the **boot** package, Canty and Ripley 2020) methods

2.4. Data wrangling

- often a simple exploration of raw data not enough
- **iNZight** provides collection of data wrangling and variable manipulation methods (including most from *R for Data Science*, Wickham and Grolemund 2017) such as filtering, aggregating, stacking, reshaping, merging, changing variable types, relabelling factors, etc
- each a separate GUI window with intuitive prompts for users to fill out, often with a preview

2.5. Special data types

Complex survey designs

- native handling of surveys: set and forget

Time series

Maps

Other types

- multiple response data
- multivariate data (addon module, section 4)
- frequencies
- extensible to others too

2.6. Code writing for getting started with R

- displays code for current plot (if enabled)
- can be edited -> GUI updates
- can be saved to 'script'
- **iNZight** keeps history of all actions as R code -> can copy and paste into R, edit manually, share, etc

[[show code history in Appendix for the demo session; note: plots not saved unless requested]]

3. Technical details

- based on **gWidgets2** (Verzani 2019) and **gWidgets2RGtk2** (Verzani 2020), which call to **RGtk2** (Lawrence and Temple Lang 2010); cross-platform
- like **gWidgets2**, we used reference classes to describe each component of the GUI [[FIGURE of GUI overlaid with object names?]]
- object oriented programming (OOP) makes sense for a GUI
  - panel has buttons, methods; belongs to another panel/window/GUI
  - inheritance: some windows have similar methods, just a few differences
- each major panel calls a single function from another packages
  - plots: `iNZightPlots::inzplot()`
  - import data: `iNZightTools::smart_read()`
  - filter: `iNZightTools::filterData()`

- wrappers: many custom-made packages in separate **iNZight*** packages; make easy interface between GUI and R (inputs = arguments)
  - enables code-writing: wrapper includes R code for its call
- describe sub-packages (TABLE: plots/regression/etc)
- diving deeper: data stored in *Documents*, each with a dataset/info about it; plots settings, variable choices, etc
  - users can switch between documents (data sets)
  - or merge them

3.1. Usage

- its an R package
- GUI is a single reference class object, '`iNZGUI`'

```
R> library(iNZight)
R> ui <- iNZGUI$new()
R> ui$initializeGui()
```

- however, include a wrapper to make easier startup:

```
R> iNZight()
```

- For development, first option is better since it provides direct access to the objects:

```
R> ui$getActiveData()
R> ui$ctrlWidget$V1box$set_value("height")
```

4. The add-on system

- **iNZight** comes with several "advanced" modules
- also a (newer) add-on system allowing anyone to install new addons from a 3rd party (or from our repo)
- a single file (currently) with a single refernence class object inheriting from '`CustomModule`'
- **iNZight** lists all modules installed to a specific modules directory on user's computer (either using manager or manually) in the Advanced menu
- when opened, shows up in the left-hand panel; has access to data, graphics device, menu items (can change these)
- developer can use their own methods, controls, etc

- Demo: Bayesian demographic modelling module (under development)

5. Installation and availability

- can install as an R package from combination of our repository and CRAN (still working towards publishing final few packages to CRAN)

5.1. Operating system specific requirements

- windows: will be prompted to install GTK binaries on first run
- macOS: very difficult, need to install XQuartz, gtk2+ framework, and compile **RGtk2** manually ...
- Linux: install system dependencies, then install as usual

5.2. Windows installer

- difficult to expect novice users/students to install R, etc
- we have a bundled version which comes as a `.exe` installer, unpackaging into chosen directory (default `Documents` `iNZightVIT`) which includes a copy of R and package library
- also some shortcuts to launch R in a subdirectory containing a `.Rprofile` which loads **iNZight** and launches the GUI

Updating

- updater included with installer: makes it easy for non-R users to update packages periodically

5.3. Docker image

- Docker (**?**) lets developers bundle apps for distribution
- containiner based on Linux and works for Linux and macOS hosts
- less speedy (performance) but doesn't require huge install; currently the only way to run **iNZight** on macOS without manually compiling lots of things

- still requires a little work from the user to enable X11 forwarding (so the interface windows display on the host screen)

5.4. Online shiny version iNZight Lite

- for users who can't install (e.g., macOS, tablet)
- most of the same functionality, separate but parallel development
- call to same packages: same results across systems
- runs on Amazon Web Services (AWS) server: some performance issues

6. Summary and future work

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. We also thank the technical support of the University of Auckland IT services for providing hosting services for our repository and Lite servers.

References

Canty A, Ripley BD (2020). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-25.

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.

Lawrence M, Temple Lang D (2010). "RGtk2: A Graphical User Interface Toolkit for R." *Journal of Statistical Software*, **37**(8), 1–52. URL http://www.jstatsoft.org/v37/i08/.

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/.

Verzani J (2019). *gWidgets2: Rewrite of gWidgets API for Simplified GUI Construction*. R package version 1.0-8, URL https://CRAN.R-project.org/package=gWidgets2.

Verzani J (2020). *gWidgets2RGtk2: Implementation of gWidgets2 for the RGtk2 Package*. R package version 1.0-7.1, URL https://github.com/jverzani/gWidgets2RGtk2.

Wickham H, Grolemund G (2017). *R for Data Science*. O'Reilly Media. ISBN 978-1491910399. URL https://r4ds.had.co.nz/.

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