iNZight Rebuild Workshop Report

Tom Elliott (iNZight Analytics Ltd)

14 November and 5 December 2024

Abstract

The iNZight data analytics tool has been a popular choice for teaching statistics in New Zealand schools and universities. This report summarises the discussions and outcomes of two workshops held to discuss the rebuild of iNZight with statistics lecturers at The University of Auckland and Victoria University of Wellington. The aim of these workshops was to engage with stakeholders over the future direction of iNZight, with a focus on functionality that makes iNZight work for teaching statistics. While the primary focus was on the long term future, discussion inevitably touched on short-term issues that could be addressed in the current versions of iNZight to ensure it continues to provide for our users until such time as the rebuild is complete.

1 Introduction

iNZight is a data analytics tool specialised for exploring data and teaching basic statistics concepts. Developed by a line of students under the supervision of Associate Professor Chris Wild at The University of Auckland, iNZight is available in two versions: the Desktop version built using GTK and an online version built using Shiny[1]. Over the past 5 years, the tools used to build iNZight have become outdated or fully deprecated, as is the case with GTK. The online version is also limited in that shiny is not the best tool for building a large, complex app, and has been plagued by connectivity issues. This means that, to continue providing iNZight to our users, we need to find a new way to build iNZight.

The current project provides us with a unique opportunity not only to rebuild iNZight, but to fully design it with a much broader scope. However, our foremost concern is that we continue to deliver for our education user base, presently our primary stakeholders. We therefore held two workshops with statistics lecturers at The University of Auckland and Victoria University of Wellington to discuss the future of iNZight, with a focus on functionality that makes iNZight work for teaching statistics.

The goal of the workshop was to engage with stakeholders to address their growing concerns over the future of iNZight. We then moved into an open discussion, specifically asking the attendees what features make iNZight stand out from the others, what features perhaps are not as useful, or could be improved, and what features are missing that would make iNZight a more useful tool for teaching statistics. This report summarises the discussions and outcomes of these workshops.

2 Discussion summary

The discussion was mostly unstructured, with attendees encouraged to speak freely and openly about their experiences with iNZight. The following is a structured summary of some of the key points made during the workshops.

2.1 What are your reasons for using iNZight?

- Change is hard: it is firmly built into some courses, and change would require a lot of work updating course notes, etc.

- The interface design supports learners through visualisation; visualisation first is fundamental to data analysis.

- Teaching versus learning versus research, although researchers still need to visualize data - other tools often have more research/industry focus and are less suitable for teaching.

- Immediacy of outputs: something happens immediately after you click a button, change an input, etc.

- The output provided, information displayed is more comprehensive that alternatives (e.g., Jamovi).

- Survey package support is easy to use and intuitive.

- Model fitting module for sample surveys.

- The fact that iNZight uses R makes it easy for students to transition from the UI to coding —- same graphs and outputs to compare.

- Data mangement: rename variables interface is the best UI for doing this; could easily rename variables into your own language, etc.

- It doesn't try to do everything: just a few things really well.

- Import from a URL, this enables some cool teaching ideas.

- Ability to create graphics that users can recreate with code: any changes to this should be considered carefully, discussed idea of a "plot generator" that ensures interaction is always converted into function arguments for reproducibility.

2.2 What's not so good?

- Inconsistencies, idiosyncracies: p-value, rounding not consistent throughout outputs; could be more/better consideration on our output is formatted, rounded.

- Some strange artefacts that mean outputs change over time, so automated teaching/assessment is difficult/needs constant updating.

- Code generation: not necessarily in a logical order (for teaching coding).

- No log of steps taken, reproducibility.

- Some sub-optimal procedures to do common things, when perhaps the default/standard isn't right/the best way to do it (possibly opinionated).

- The plot design for categorical versus numeric — the order of variables *does not* matter, but this does not necessarily translate to teaching modeeling concepts, etc.

- Better handling of errors when importing data, especially when weird characters are present. Robustness is required for teaching, as students import different/live datasets so some will inevitably run into issues.

- Plot types: "spine plots" not universally liked, understood, require more explanation — option to turn off would be good.

2.3 Features that would be nice

- Shareable link (e.g., Codap)

- Better separation of interface/global settings (e.g., font size) versus current chart (plot type, etc). *Note: this may be more of a Lite issue.*

- Option to set/choose the first variable when sharing a link. *Note: this is already a feature, but not documented well.*

- Reordering levels of a factor by drag-and-drop.

- Data cleaning using interactive graphics.

2.4 Research opporunities

- How interface design aids learning.

- Year 12–13, reeducating teachers on using data and tools.

- From spreadsheet to iNZight to R.

2.5 Barriers

The main barrier for iNZight in future is competition from other tools, such as Codap and Jamovi. Codap especially is for data science research and is heavily funded.

2.6   Short-term issues

As well as discussing the long-term future of iNZight, we also discussed some short-term issues that could be addressed in the current versions of iNZight to ensure it continues to provide for our users until such time as the rebuild is complete.

- Two numeric variables: inference has no default (linear trend), user must specify it (unlike the other variable combinations) — but things like this can be valuable teaching moments.

- Ability to set null hypothesis for linear regression.

- Confidence region for linear trend (not just bootstraps).

2.6.1   iNZight Lite (shiny)

- Minor UI fixes to improve responsive design (e.g., for smaller screens).

- Known issues, e.g., trend option disappearing after saving residuals — setting name of new variable doesn't seem relevant, additional clicking and produces the bug.

- Example data: autoload Visualize tab.

3   Conclusion

This section provides concluding remarks and future directions.

References

[1]   Winston Chang et al. *shiny: Web Application Framework for* R. R package version 1.6.0. 2021. URL: https://CRAN.R-project.org/package=shiny.