

# Code-based, open-source software for teaching interactive data visualisation



Shan-I Lee

Department of Statistics  
The University of Auckland

Supervisor: Dr Paul Murrell

A dissertation submitted in partial fulfillment of the requirements for the degree of BSc(Hons)  
in Statistics, The University of Auckland, 2017.



# Abstract

The software required to implement interactive techniques in data visualisation has become readily available to statisticians with the development of web-based graphics and several open-source R packages. This paper proposes using the R packages, **plotly**, **crosstalk** and **shiny**, to teach a set of powerful techniques for interactive data visualisation. The value of applying linked brushing, identification, scaling, subset selection and tours, to reveal further insight during exploratory data analysis will be demonstrated. Outliers are quickly examined using linked brushing and identification to better understand how they are unusual. Different structures in the data are explored with interactive scaling and subset selection. Dynamic tours provide views of complex multivariate structures and linked brushing enables multiple representations of the data to be explored simultaneously and related together. The additional insights gained from applying interactive data visualisation in comparison to analysis with static plots alone, show that the set of interactive techniques and software identified is a powerful toolbox for statisticians to have.