SEP 14, 2023

DOI:
dx.doi.org/10.17504/protocol
s.io.n92ldmr5nl5b/v1

External link:
https://delaney.shinyapps.io/
Spectrum_Data_Plotter/

Protocol Citation: Joe R
Delaney 2023. Spectrum
Data Plotter: web based
violin, box, dot, and average
with error overlaid rich data
plots. **protocols.io**
https://dx.doi.org/10.17504/p
rotocols.io.n92ldmr5nl5b/v1

🌐 Spectrum Data Plotter: web based violin, box, dot, and average with error overlaid rich data plots

Joe R
Delaney[1]

[1]Department of Biochemistry and Molecular Biology, Medical University of
South Carolina, Charleston, SC, United States of America

Joe Delaney

ABSTRACT

Data scientists have been releasing advanced graphical tools in R packages for
many years. Unfortunately, individuals with no coding knowledge are unable to
access such tools. One particularly well-used package, ggplot2, produces graphical
outputs which scientists in many disciplines find useful for their data sets. Here, we
release Spectrum Data Plotter, a web-based integration of standard ggplot2 features
optimized for general scientific data plots. Spectrum Data Plotter is able to produce,
overlaid or one-by-one, user-customized dot plots, average and standard deviation or
error plots, boxplots, and violin plots. While options are not comprehensive to the
ggplot2 library, many customizations including line type, color, transparency, and
sizes are available to the user using simple text inputs, pull-down menus, or check
boxes. Given the limitations of paid software, the intent is for users to be able to
generate beautiful plots which fairly and wholly communicate users' hard-earned
data.

Protocol status: Working
We use this protocol and it's
working

Created: Sep 14, 2023

Last Modified: Sep 14,
2023

PROTOCOL integer ID:
87783

Keywords: ggplot2, shiny,
violin plot, dot plot, boxplot,
web software

Background and use cases

1   Quick link to tool:
    https://delaney.shinyapps.io/Spectrum_Data_Plotter/

    Interpretation of data via rich plots is almost as important as the data itself for the
    advancement of science. As big data becomes more prevalent in scientific studies, there is a
    diverse need of graphical outputs which may be inaccessible in commonly used software
    such as Excel or Sheets. While paid software exists in many formats, this can also be
    inaccessible to students and other trainees due to costs. Data scientists and
    bioinformaticians often embrace open source code and sharing of advanced software to
    remove such barriers and advance fields more rapidly. Now, R packages are released by the
    thousands each year and others are updated with generous features. The most widely used
    package for generating plots is ggplot2, with tens of thousands of monthly downloads.

2   Big data can be better displayed using features such as histogram binning to avoid
    overwhelming readers with thousands of observations on a single plot. The violin plot, a
    version of a histogram, is growing in usage in the literature for its succinct, visually appealing
    representation of large datasets. Boxplots are used to add calculated metrics including the
    median and interquartile range, often layered on top of violin plots. While these two plots are
    ubiquitous in the literature, they are not readily accessible in free software.

3   Shiny is an R package which allows other R packages to be utilized with a web-based user
    interface. Here, we describe the Shiny app Spectrum Data Plotter, which formats ggplot2
    violin, boxplot, dotplot, and average +/- error plot functions in a user-friendly web interface.

Rich data plots from dozens to thousands of observations   6m

4   Input Data.   `1m`

Spectrum data plotter is intended to complement other commonly used graphing software by specializing in visualization of larger scale data sets. Data is input into Spectrum Data Plotter either by copying-and-pasting from a spreadsheet into a text field or by uploading a table of data from the

user's computer (**Figure 1**).
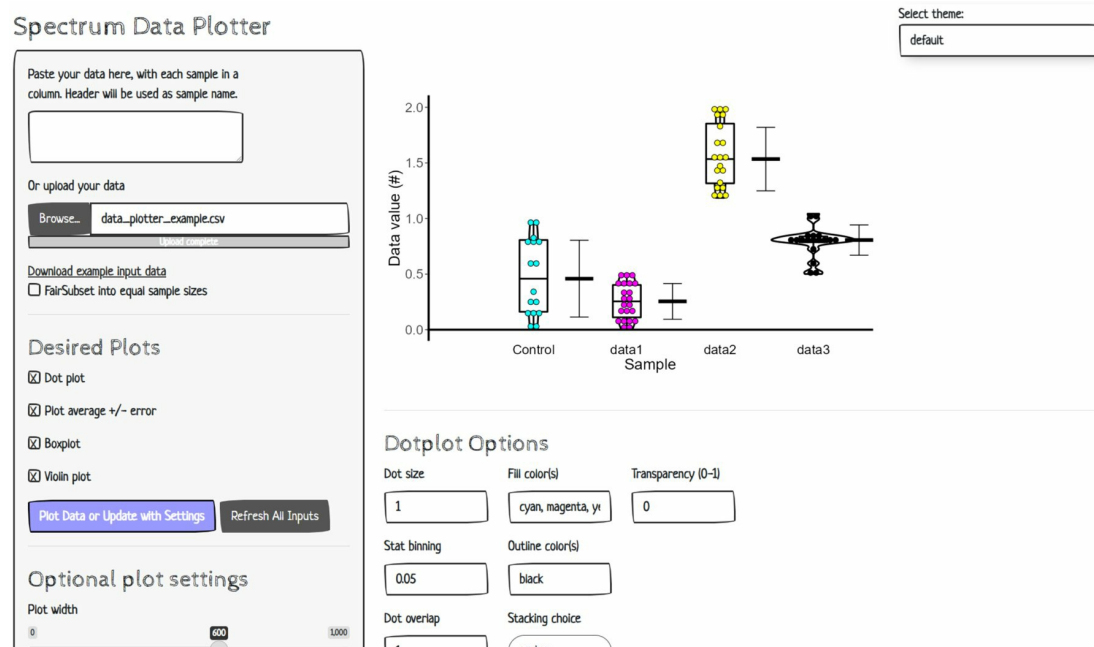
Figure 1. Landing webpage for Spectrum Data Plotter.
Displayed example utilized the provided example data, loaded into the data upload field in the upper left.

5   Data formatting.   `5m`

Each column represents a different sample and requires a header of sample names. Each column of data may contain different numbers of observations; each sample need not be the same size. Each sample may be colored differently for any field in Spectrum Data Plotter labeled with "Color(s)". Colors must be separated by commas and may be common terms like "red, blue" or include hexadecimal code such as "#008080, #C2C2C2, black".

Dot plots

6   One method of observing both quantity of observations while also describing the range and distribution of data is through dot plots. To demonstrate this described functionality, publicly

available data regarding the incidence of rainbows observed by date and sun angle was downloaded and formatted for input into Spectrum Data Plotter. The output graph shows the distribution of sun angle according to each rainbow in the dot plot with colors associated with time periods (**Figure 2**). Note that the differing number of samples is clear in the plot and indicates what the authors of the data originally described in their publication: an increasing number of observed rainbows over time [1]. There are some cases in which it is more desirable to plot exactly the same number of observations per sample, in which case Fairsubset [2] is integrated and can be used with Spectrum Data Plotter.
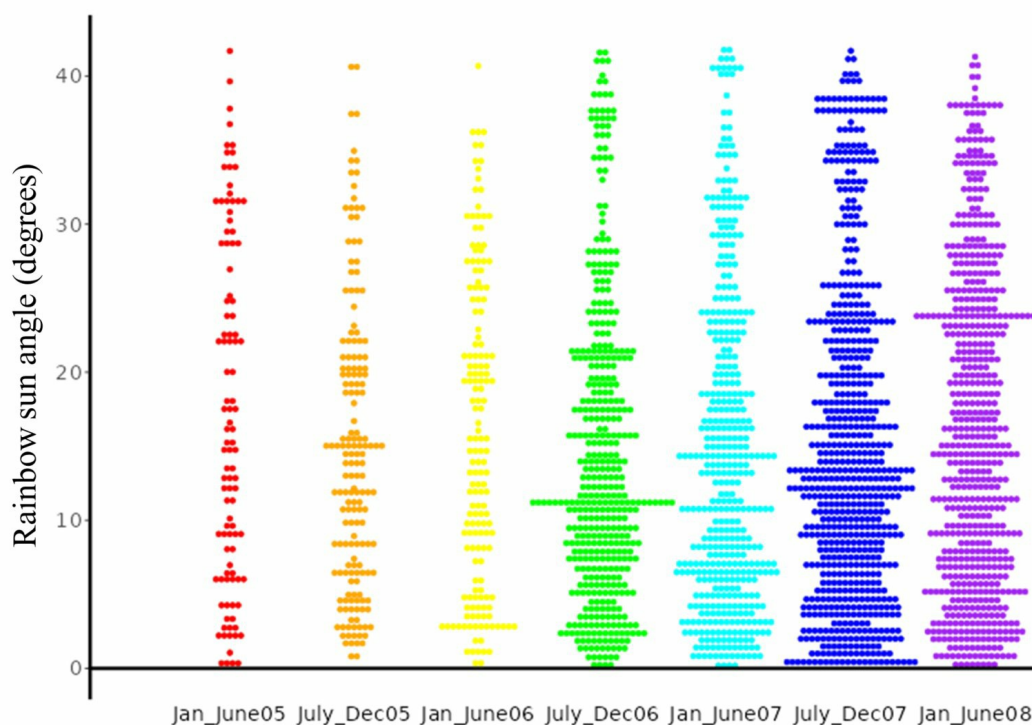

Figure 2. Spectrum Data Plotter colored dot plot example.
Rainbow observations by sun angle during different time periods.

Violin plot, box plot, and overlays

7   One plot feature of general interest is to overlay multiple types of plots to allow deeper understanding of the data characteristics. Spectrum Data Plotter can overlay plots generated from numerical data from any field of science. The field of cancer genetics has expanded over the last decade. Information regarding somatic aneuploidy and segmental copy-number alterations has better informed tumor biology. Samples number in the hundreds to thousands in these studies, which are best displayed using histograms such as violin plots and/or boxplots. Individual datums are difficult to display coherently in this case. Previously published copy-number alteration network analysis data of the autophagy molecular recycling

pathway from gynecologic cancers [3] were used as input data from Spectrum Data Plotter. Overlaid violin and boxplot graphs demonstrate that the distribution of the autophagy copy-number alteration networks are on average negative, that is in a suppressive direction, but some tumors are near zero or above zero and not suppressed (**Figure 3**).
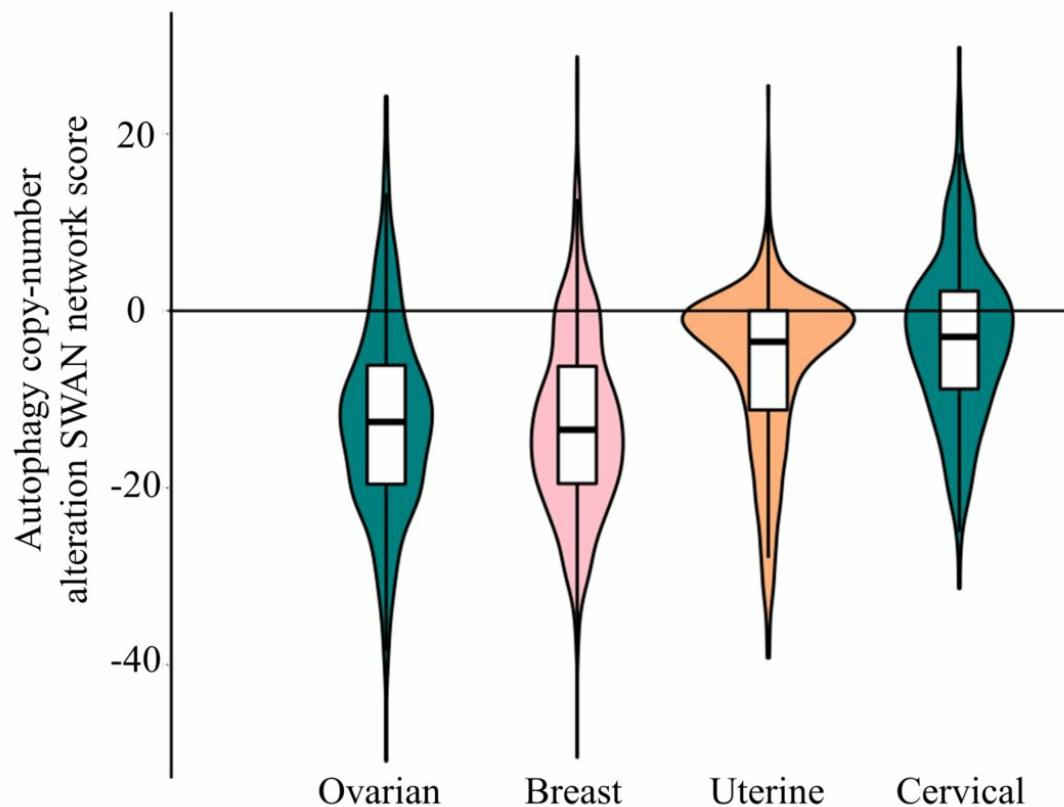

Figure 3. Overlaid violin and boxplots. Copy-number alteration network scores in gynecologic cancers.

Staggered means and error with associated dot plots

8

Another aspect of multi-layered plots giving the same data different visual representations is a desire to represent both statistical measures (eg, means and standard deviation) alongside the raw data (eg, a dotplot of all observations). This is common in scientific publications for animal data in which each animal's observed data should be displayed to assess for any concerning outliers. To demonstrate such a situation, data from a well-cited study evaluating aspects of rapamycin treatment on the aging process were analyzed from a published figure panel regarding fecal pellets associated with microbiome changes[4]. Pixel counts are

displayed per pellet and graphed alongside medians and standard deviation using Spectrum Data Plotter "Shift x" offset feature for the "average +/- error" plot option (**Figure 4**).
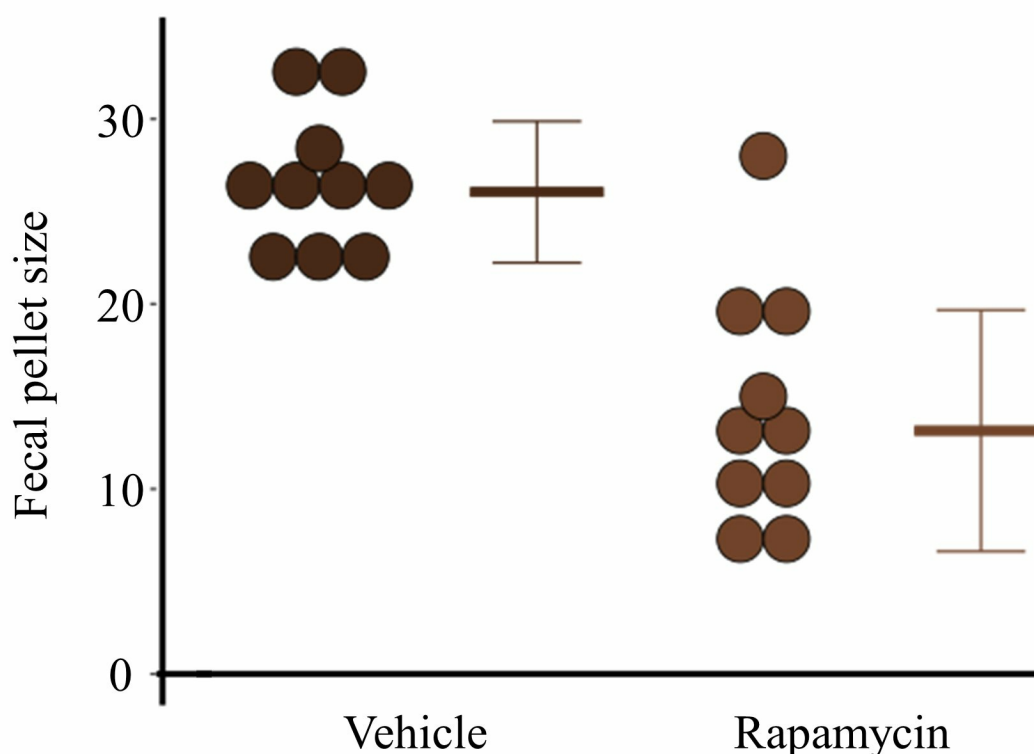
Figure 4. Dot plot and average plus error offset feature. Fecal pellet sizes with or without rapamycin treatment.
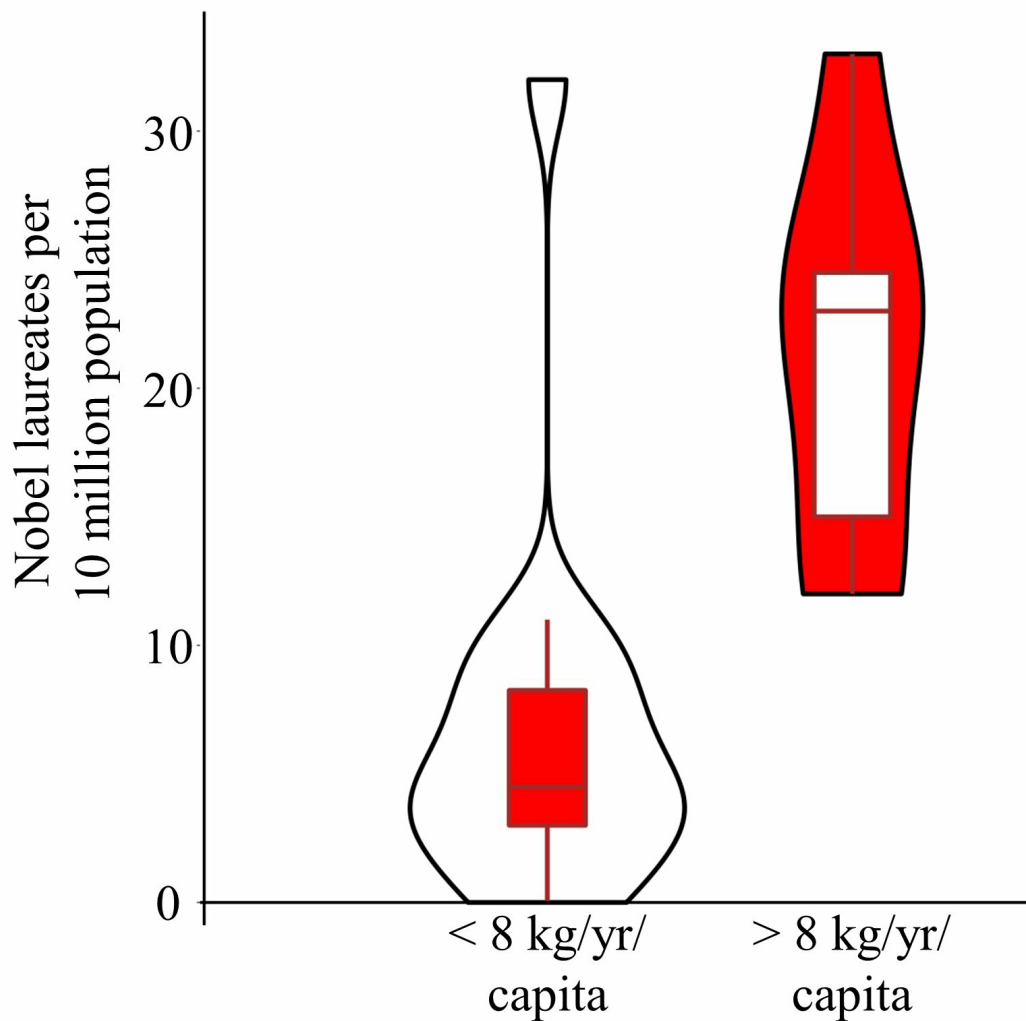
Limitations

9

A number of limitations are currently associated with the release of Spectrum Data Plotter. First, ggplot2 is a powerful R package with far more graphical outputs available than those currently wrapped into Spectrum Data Plotter. Second, most scientific plots indicate significance directly on the plot itself. This feature is not present in Spectrum Data Plotter. To visualize an example of this limitation, data comparing chocolate consumption by country to Nobel laureates, normalized by population size, was downloaded from a previous publication [5]. Data were categorized into high chocolate consumption countries and compared to lower chocolate consumption countries. While Spectrum Data Plotter can display the association, it cannot determine statistical significance nor plot the significance on the plot output (**Figure**

5). It is recommended that the user utilize other tools to assess significance, and manually draw relevant significance annotations onto the Spectrum Data Plotter plots when appropriate.

Figure 5. Chocolate consumption and Nobel laureate prevalence. A limitation of Spectrum Data Plotter is the inability to calculate or display statistical significance on the plot.

Summary

10   Here, Spectrum Data Plotter is released as a web application and associated code is available from GitHub under a GNU Public License V3. This app may be considered for generation of

violin, boxplot, dot plot, and average +/- error multi-layered plots when paid software is unavailable, or when free software is onerous or unable to produce publication-quality plots for scientific data.

Resource links

11   To access raw code and set up on local hardware:
https://github.com/jrdelaney/Spectrum_Data_Plotter/

12   To use the free web portal version:
https://delaney.shinyapps.io/Spectrum_Data_Plotter/

13   To use associated basic science tools:
http://www.delaneyapps.com/