Comparing Code and Results of Basic Plotting across Programs

David M Vermillion
davidmvermillion.com
July 26th, 2020

Abstract

This project exists as a quick reference guide for the work required to achieve simple graphs while showing how common programs develop those graphs. Some graphs are created using a GUI and some using command line coding. More customization is possible than is shown in this document. This document shows the approximate workload and default or near-default design philosophies of each program. Graphing examples are shown using Google Sheets, Excel, LibreCalc, Matlab, Mathematica, ggplot, Gluviz, GNU Octave, gnuplot, Matplotlib, Pandas, Past, R Studio, and Sagemath. The dataset used for this project was created in Google Sheets using the code snippet `=RANDBETWEEN(1,100)` across a swath of cells 2 rows wide and 50 columns deep. The column to the left was set from 1 to 50 to label the counts of each data series, resulting in a file with 3 columns and 50 rows. This document was created using LibreOffice Writer in an .odt file before being exported as a PDF for distribution to maintain formatting across viewing devices.

Sheets

Using the originating spreadsheet, the first graph (**Figure 1**) was a default scatterplot created using Google Sheets (Insert > Chart) with two trendlines added (Edit Chart > Customize > Series > Trendline > Linear > Line Opacity = 100%) and downloaded as an SVG. Following the path (Customize > Chart & axis titles), the chart title was set as displayed under the chart title drop-down. In the same drop-down, the horizontal and vertical axes were set as displayed. Under the Series tab, after selecting the individual series set, (Label > Use Equation) was chosen as the name for the lines. All other default settings were kept. All graph programming was made using the GUI.

Figure 2 is obtained by tweaking the existing chart to create a histogram as shown. Under Setup, the Chart Type drop-down list contains a histogram option. Under Data Range, the data was changed from `A1:C50` to `B1:C50` to filter out the existing count column. Under (Customize > Histogram), the bucket size was manually set to 15. The vertical axis title (same path as the last chart) was set to Frequency. The horizontal axis title was reset to Data Value. This clearly shows a skewing toward the bottom values for both sets of *random* numbers. No further customizations were made. All changes made were executed through the GUI.

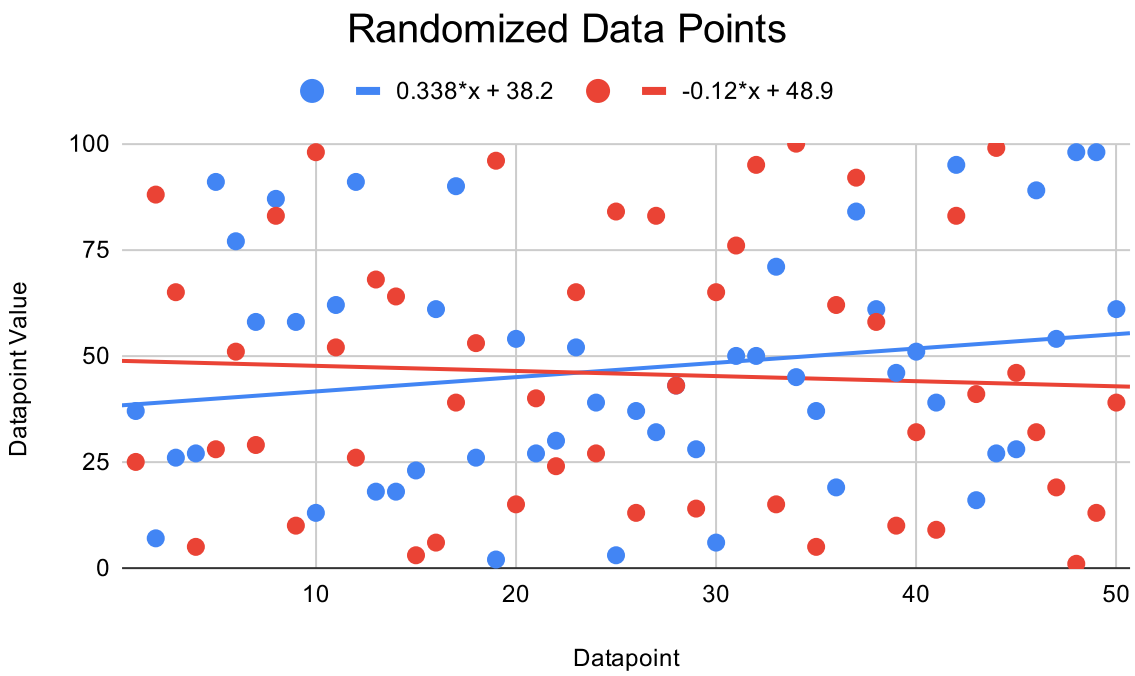

Figure 1: Google Sheets Scatterplot with Trendlines and Labels

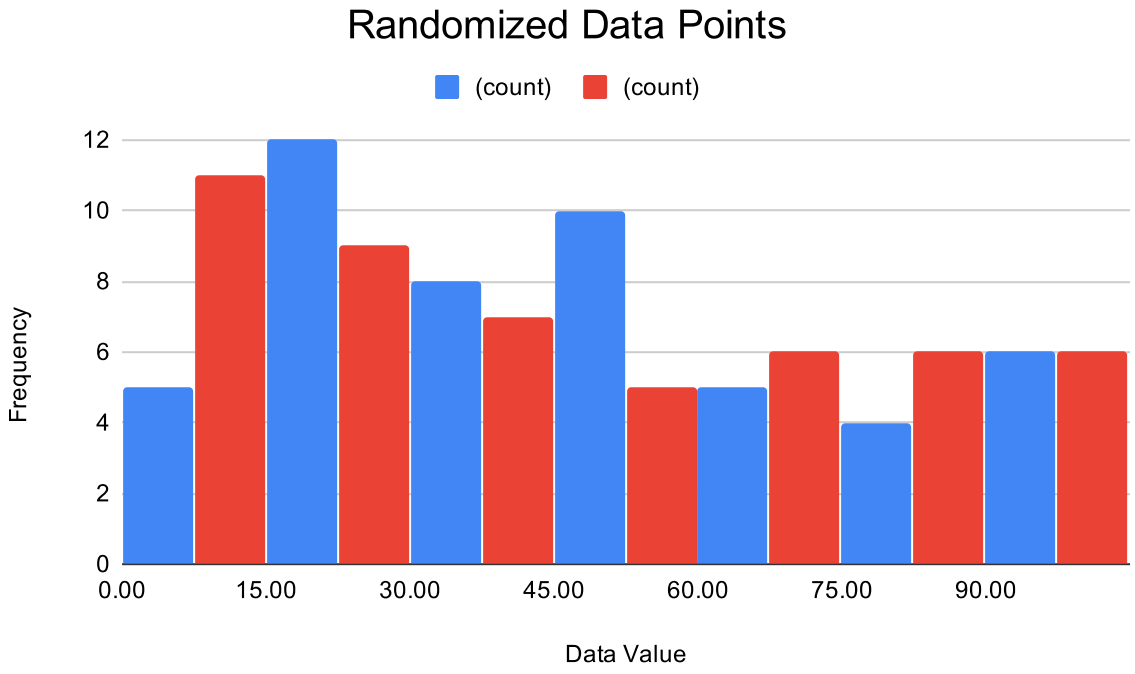

Figure 2: Google Sheets Histogram

EXCEL

The Google Sheets dataset was downloaded as a .xlsx file for use in Microsoft Excel. Following the menu options (Ribbon Bar > Insert > Scatter) populated a scatterplot from the selected range. Double-clicking the default chart title allowed it to be changed as shown. Clicking the GUI to add a chart element and selecting Axis Titles populated default titles which could then be changed through a double-click action. On the Ribbon Bar element for Chart Design, the following path inserted trendlines. (Add Chart Element > Trendline > Linear). The process was repeated for Series 1 and Series 2. Selecting the trendlines and going to the Format Trendline Window, the button was selected to display the equation on the chart. The default placement had to be changed for readability. Under label options in the same window, a side button opens for Text Options. Color selection is then possible. This was changed from the default color (same as other labels) to match the series colors. The formula bar displayed `=SERIES(,Data!$A$1:$A$50,Data!$C$1:$C$50,2)` as the series selection while checking Series 2 on the graph. No further changes were made. All settings were completed using the GUI. Excel does not allow directly exporting graphs. Therefore they must be copied/pasted through the computer clipboard to a different program. For this demonstration, **Figure 3** was pasted into Inkscape to save as an SVG. The dotted trendlines display in Excel and Inkscape as ~ 20% of the series dot size. They currently show as ~ 80% of the series dot size in this document.

Selecting all three columns, a histogram was created by the same path of (Ribbon Bar > Insert > Histogram). This time the graph was proportionally scaled by clicking the right top corner and dragging while holding down the shift key. Axis titles were inserted as before. The data series had to be readjusted by following the path (Ribbon Bar > Chart Design > Select Data). Column A was renamed as Positions, Column B as Series 1, and Column C as Series 2. Series 2 data selection was coded as `=Data!$C$18:$C$50` as displayed during the selection process. Positions was also placed into the Horizontal (Category) Axis Labels. After selecting the horizontal axis, going to the window with (Format Axis > Axis Options > Axis Options) the bin width was manually adjusted to 15. This resulted in **Figure 4**. It therefore appears that default histogram plotting of multivariate data is not currently an option in Excel. However, the workaround is explained in the following paragraph.

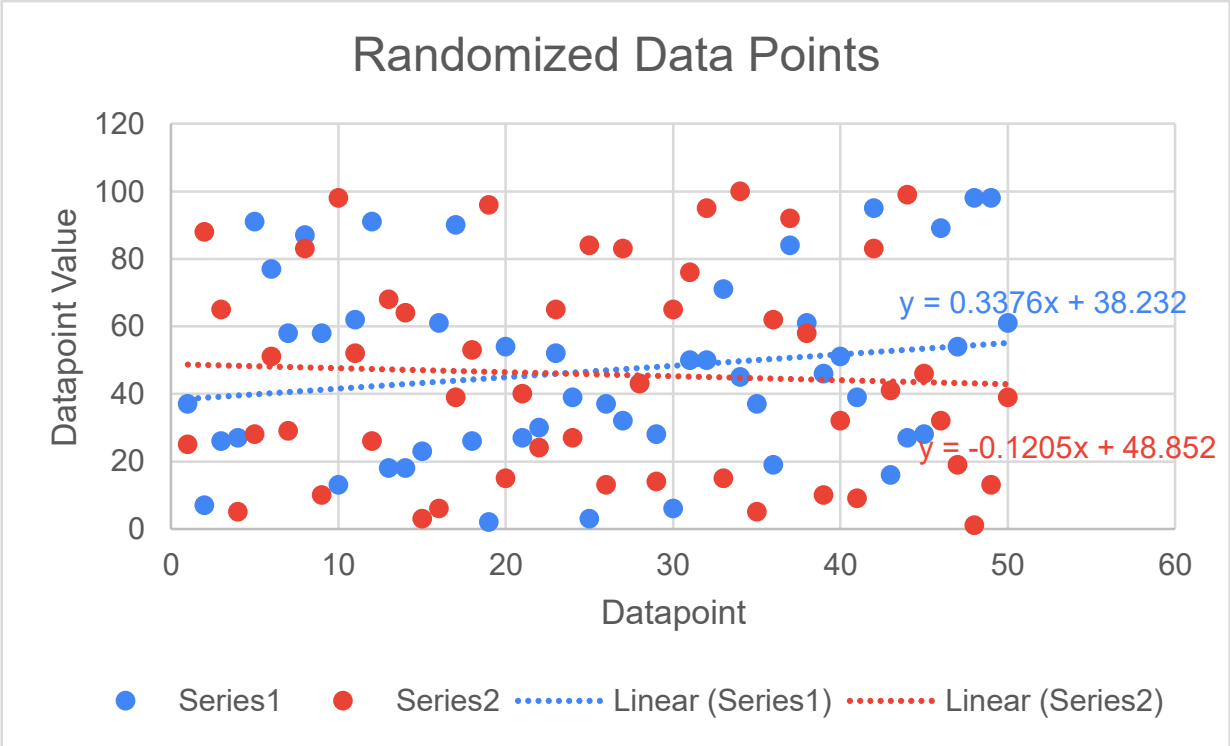

Figure 3: Excel Scatterplot

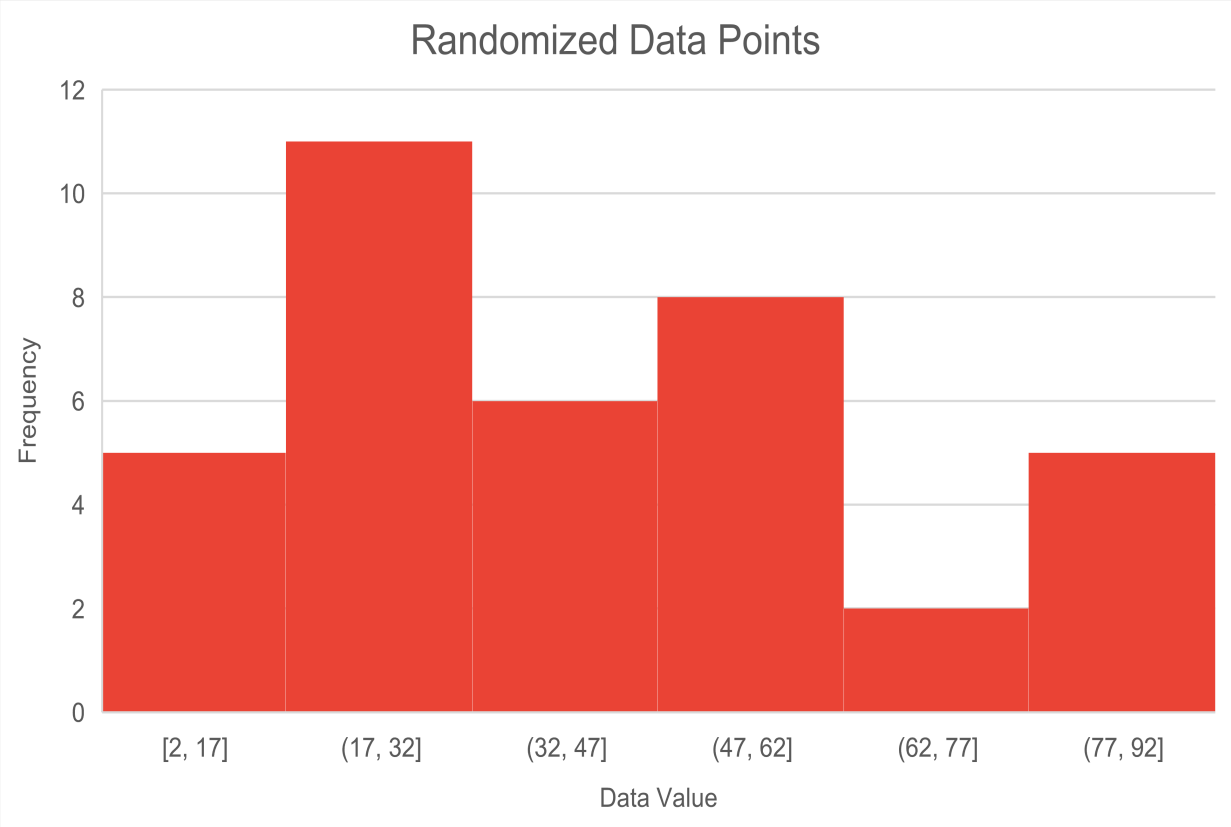

Figure 4: Excel Default Histogram