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. Graphing examples are shown using Google Sheets, Excel, LibreCalc, Matlab, Mathematica, gnuplot, R Studio, GNU Octave, Matplotlib, Pandas, Gluviz, and ggplot. 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.

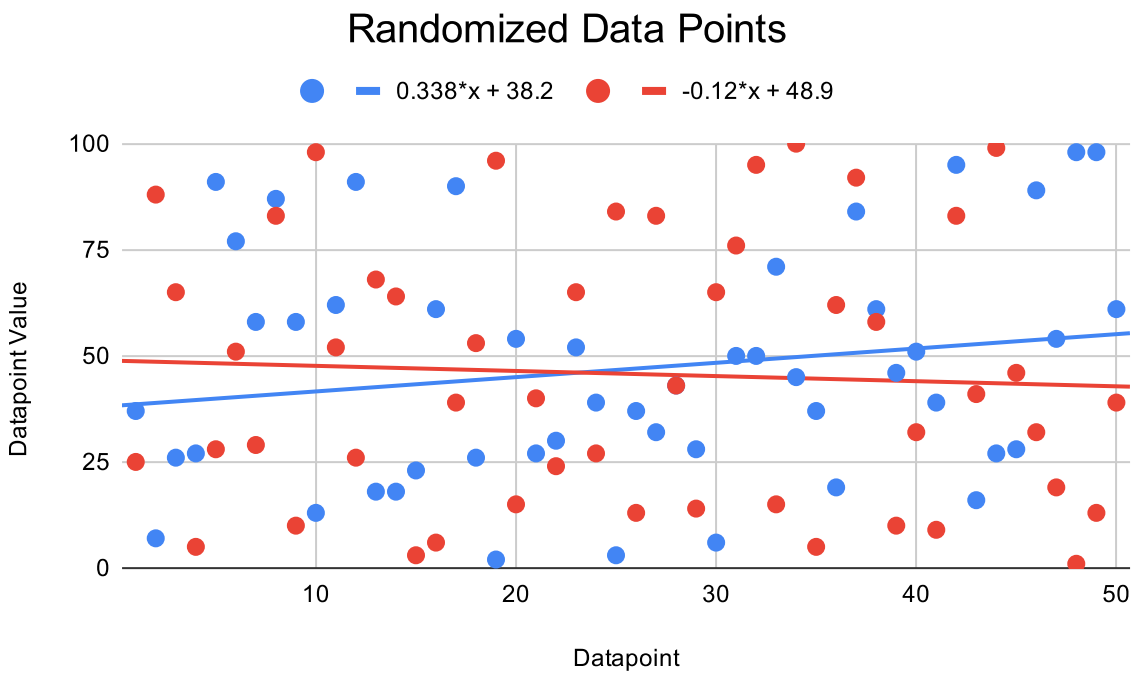

Figure 1: Google Sheets Scatterplot with Trendlines and Labels

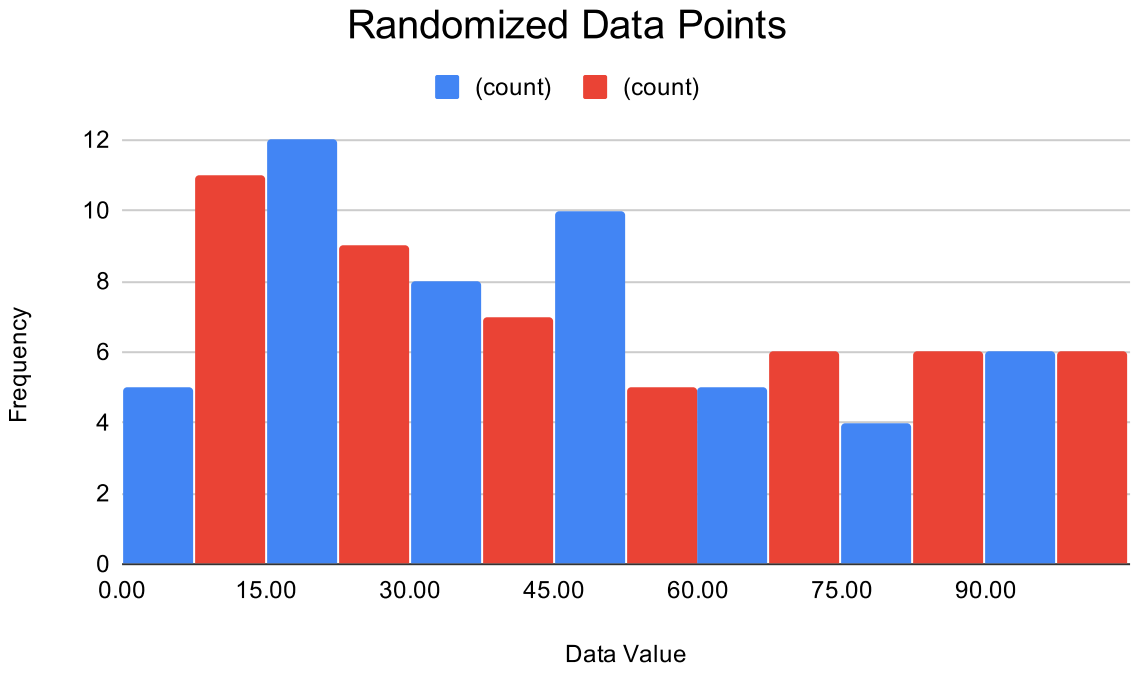

Figure 2: Google Sheets Histogram

EXCEL

The Google Sheets dataset was downloaded as a .xlsx file for use in Microsoft Excel

```
=RANDBETWEEN(1,100)
```