17:610:554:90 INFORMATION VISUALIZATION AND PRESENTATION

Spring 2022 Term Project:
A Prototype Web-Based Data Visualization Tool

John Killian

Introduction

This project is a prototype for a general-purpose online interactive tool for data visualization. It takes user-supplied data in the form of a CSV file and generates a table and graph of that data in a form and format based on selections made by the user.

While designed to be a general-purpose tool and not specifically for any one domain, for the purposes of this paper, I will examine the use of this prototype with financial data and elicit domain expert feedback from a professional in that area.

Original Conception and Design Evolution

My original vision for this tool was for it to be something akin to a CodePen or JSFiddle for data visualization. For those unfamiliar with these two tools, their function is provide a self-contained front-end development environment for creating and testing web pages and applications that resides on a single web page and works in any modern web browser. My idea was to apply this concept to data visualization and create an (admittedly stripped-down) version of offline applications such as Tableau or Excel, that would require no software installation other than a browser and a basic text editor. Users would be able to choose a data source from their local file system, select a chart type and generate an interactive graph based on their selections.

One aspect of my original conception that I was forced to jettison was a feature that allowed the user to download the finished visualization as an SVG file. While it would seem that this would be easy to accomplish, SVG being a simple text file after all, there is currently no way to save an SVG file from within a web page that works for all common browsers. The alternative of requiring the user to select the SVG portion of the page by using the browser's developer tools was unacceptable, so the sacrifice had to be made.

With what would have been a fairly important feature eliminated, I decided to shift the focus of the prototype somewhat to be more of an instructional tool, an environment that teaches the user about creating data visualization as they create a data visualization.

As the user goes through the process of creating their graphic, text panels would describe the type of graph, how it is used and what type and format of data is required. These cues would appear progressively, one at a time, as the user went through the step-by-step process described below.

Tools and Resources

The project was coded entirely in HTML, CSS and JavaScript and makes use of the D3.js library.

Having decided on a web-based prototype, the obvious choice for developing the data-driven elements of the project was the D3.js library.

- **Pros of using D3**
  - As a robust, actively maintained project, D3 has become the de facto standard visualization library for web applications. As such, tutorials, forums and sample code are readily available and were immeasurably helpful in the building of this prototype.
  - On a personal note, part of the reason I wanted to take this course was to improve my very limited knowledge of D3 and choosing to use the library for my project certainly helped accomplish that goal.
- **Cons of using D3**

- The primary con to using the D3 library is its somewhat steep learning curve. It can take time and much trial and error to get one's brain around the D3 way of doing things, particularly with regard to interactions between elements within and outside of the visualization.
- Another negative is its lack of interoperability between versions, which are frequently updated. This is especially irksome when using online resources such as tutorials or forums and finding that a certain answer cannot be applied to a problem because it was written in a different, incompatible version.

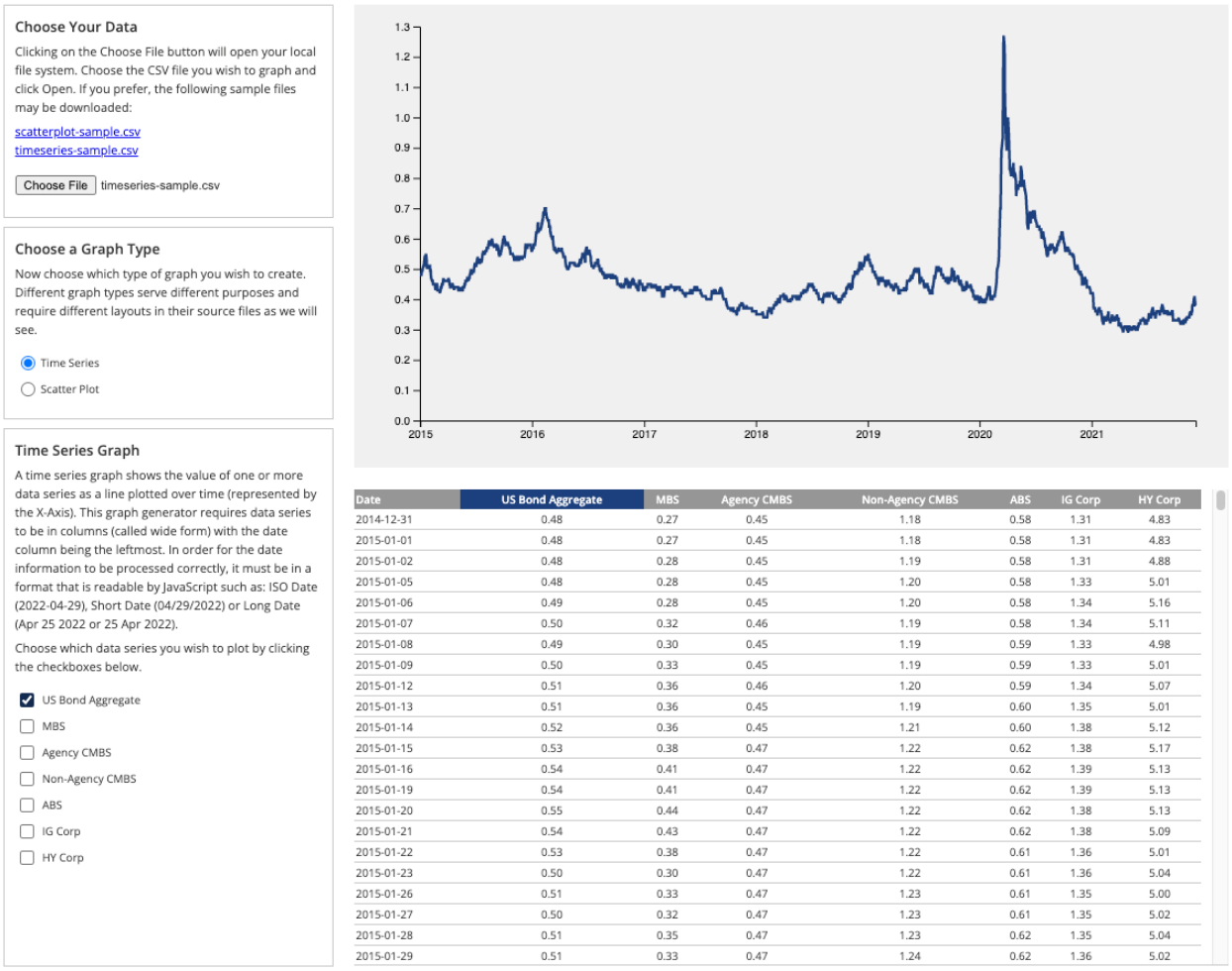
Design Approach and Data Visualization Principles Applied

My approach to this project was to build a visualization tool that is simple and intuitive, that guides the user step by step through the creation of a data visualization and in the process teaches a little something about data visualization itself.

This project required the application of data visualization principles in two separate areas, the data vis and the user interface. Of these areas, the more critical in my opinion is the user interface. The reason for saying this is that the data vis itself uses time-tested, standard graphic elements (line chart and scatter plot) that would be familiar to most users. That said, attention was paid to using good design principles in that area as well, particularly where interactions between the visualization and the UI take place.

Screen Layout

The prototype's screen is divided into two sections, on the left, the Control area and on the right the Display area.



Date	US Bond Aggregate	MBS	Agency CMBS	Non-Agency CMBS	ABS	IG Corp	HY Corp
2014-12-31	0.48	0.27	0.45	1.18	0.58	1.31	4.83
2015-01-01	0.48	0.27	0.45	1.18	0.58	1.31	4.83
2015-01-02	0.48	0.28	0.45	1.19	0.58	1.31	4.88
2015-01-05	0.48	0.28	0.45	1.20	0.58	1.33	5.01
2015-01-06	0.49	0.28	0.45	1.20	0.58	1.34	5.16
2015-01-07	0.50	0.32	0.46	1.19	0.58	1.34	5.11
2015-01-08	0.49	0.30	0.45	1.19	0.59	1.33	4.98
2015-01-09	0.50	0.33	0.45	1.19	0.59	1.33	5.01
2015-01-12	0.51	0.36	0.46	1.20	0.59	1.34	5.07
2015-01-13	0.51	0.36	0.45	1.19	0.60	1.35	5.01
2015-01-14	0.52	0.36	0.45	1.21	0.60	1.38	5.12
2015-01-15	0.53	0.38	0.47	1.22	0.62	1.38	5.17
2015-01-16	0.54	0.41	0.47	1.22	0.62	1.39	5.13
2015-01-19	0.54	0.41	0.47	1.22	0.62	1.39	5.13
2015-01-20	0.55	0.44	0.47	1.22	0.62	1.38	5.13
2015-01-21	0.54	0.43	0.47	1.22	0.62	1.38	5.09
2015-01-22	0.53	0.38	0.47	1.22	0.61	1.36	5.01
2015-01-23	0.50	0.30	0.47	1.22	0.61	1.36	5.04
2015-01-26	0.51	0.33	0.47	1.23	0.61	1.35	5.00
2015-01-27	0.50	0.32	0.47	1.23	0.61	1.35	5.02
2015-01-28	0.51	0.35	0.47	1.23	0.62	1.35	5.04
2015-01-29	0.51	0.33	0.47	1.24	0.62	1.36	5.02

This screenshot shows the basic layout of the prototype. The left quarter of the screen is the Control area and is made up of panels containing HTML input elements of various kinds. The area on the right is the Display area composed of the Graph section and the Table section.

- **Control Area**

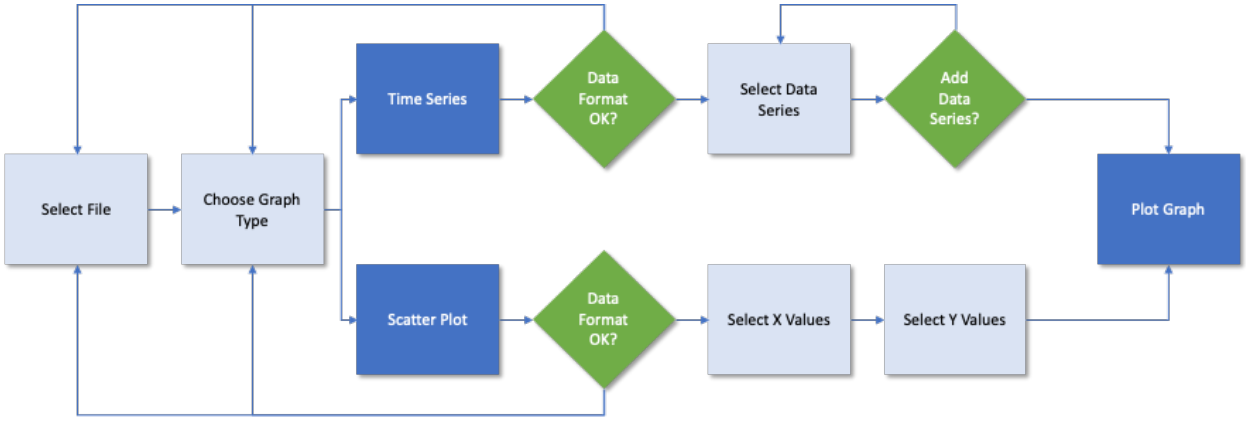
  As its name implies, this is the area in which interaction will mainly occur. Each interaction with Control elements will update elements in both areas as the application cycles through its various states.

- **Display Area**

  This area is subdivided into a Graph section at the top (a gray rectangle) and a Table section beneath it. These two sections show the results of user interactions.

A Step-by-Step Approach

In the prototype, the user interface builds on itself in a progressive manner. UI elements appear on screen only when needed preventing accidental (out of order) interactions. At each step, the user has an opportunity to revisit a previous step or reset and start over (Nielsen UI Heuristics #3)



This flow diagram illustrates the prototype's step-based approach to interaction.

- **Proximity / Common Region**

  The separation of Control items into separate panels conforms to the Proximity and Common Region Gestalt principles. The panels are clearly delineated and the top to bottom arrangement is easy and natural to follow.

- **Common Fate and Similarity**

  The simultaneous appearance of visual elements within the Control and Display areas creates a connection between them in the user.

- **Similarity**

The consistent appearance of the Controls and the consistent use of color in the UI and graphics are examples of the Gestalt principle of Similarity and leverage pre-attentive perception

Prototype Walk-Through

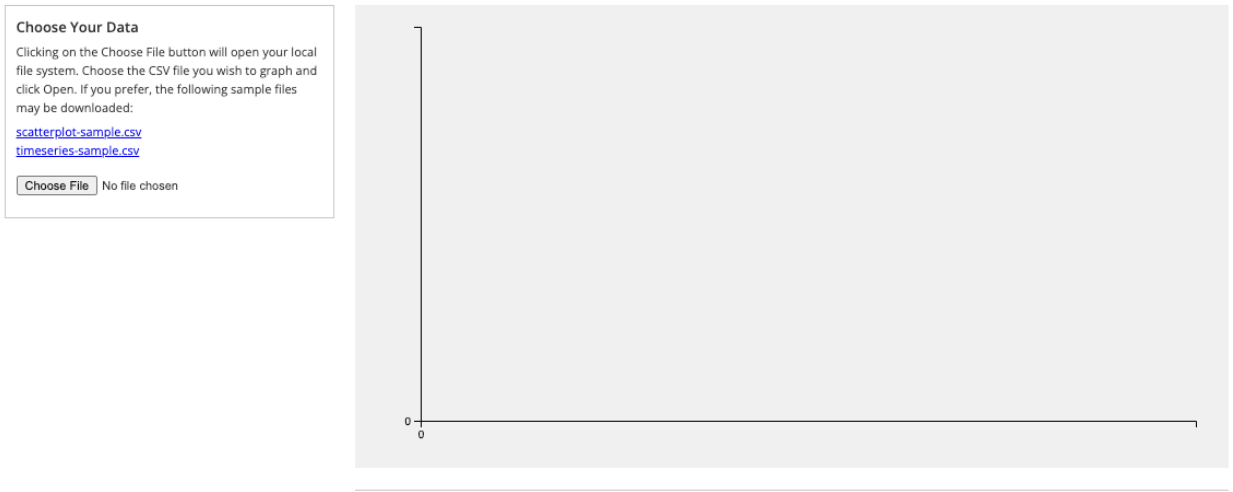
The following walk-through will examine the effects on both Control and Display areas as the user interacts with the prototype.

Interaction States

Interaction with the prototype progresses through three main states: 1 the Initial (Reset) State, 2 the File Loaded State and 3 the Interactive Graph State

State 1 – Initial (Reset) State

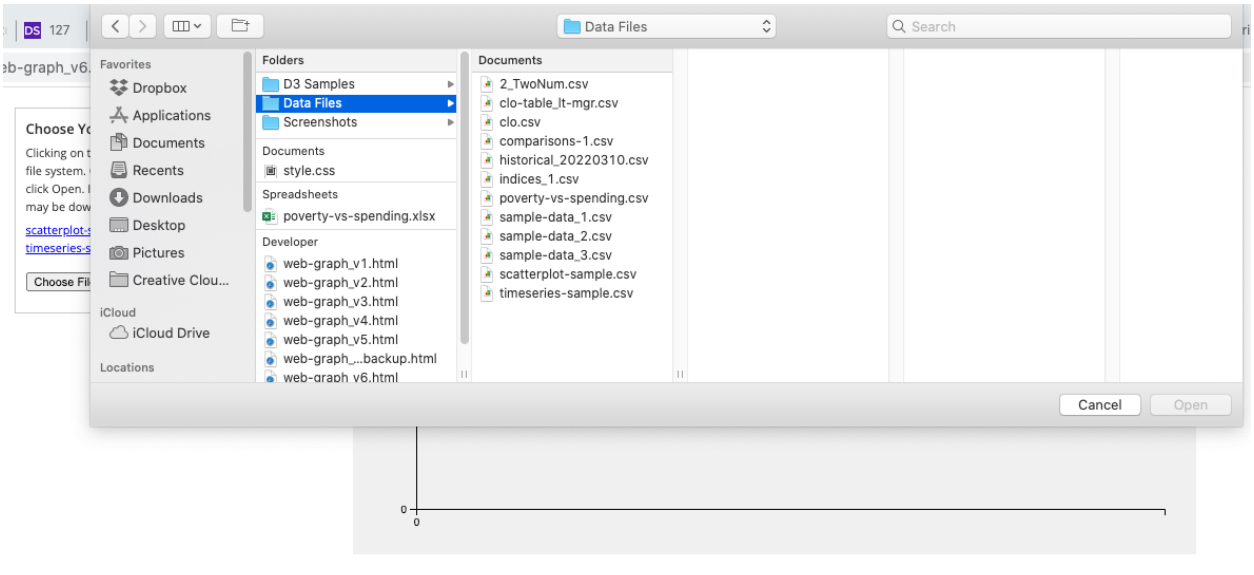
At this point, the app is essentially a blank slate.



- **State 1: Control Area**

  In its initial state, the Control area consists of one panel, the File Selection Control containing an instruction paragraph, links to sample data files and an HTML "File" input button.

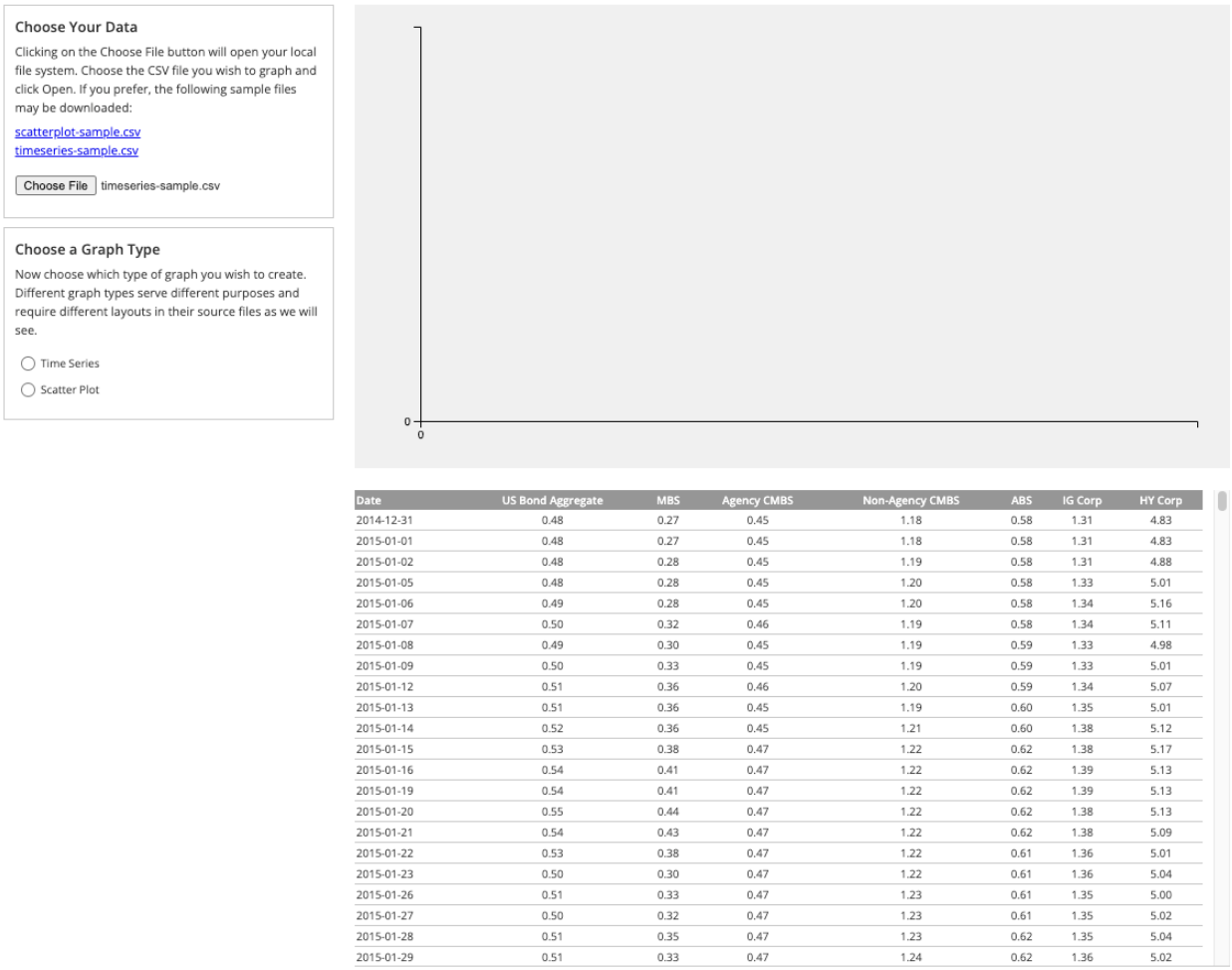
  Clicking on this button opens a dialog box to select a file from the user's local file system.



- **State 1: Display Area**

  The Graph (top) section of the Display area holds only unmarked X and Y axes. The table section has yet to make its appearance.

State 2 – File Loaded



Date	US Bond Aggregate	MBS	Agency CMBS	Non-Agency CMBS	ABS	IG Corp	HY Corp
2014-12-31	0.48	0.27	0.45	1.18	0.58	1.31	4.83
2015-01-01	0.48	0.27	0.45	1.18	0.58	1.31	4.83
2015-01-02	0.48	0.28	0.45	1.19	0.58	1.31	4.88
2015-01-05	0.48	0.28	0.45	1.20	0.58	1.33	5.01
2015-01-06	0.49	0.28	0.45	1.20	0.58	1.34	5.16
2015-01-07	0.50	0.32	0.46	1.19	0.58	1.34	5.11
2015-01-08	0.49	0.30	0.45	1.19	0.59	1.33	4.98
2015-01-09	0.50	0.33	0.45	1.19	0.59	1.33	5.01
2015-01-12	0.51	0.36	0.46	1.20	0.59	1.34	5.07
2015-01-13	0.51	0.36	0.45	1.19	0.60	1.35	5.01
2015-01-14	0.52	0.36	0.45	1.21	0.60	1.38	5.12
2015-01-15	0.53	0.38	0.47	1.22	0.62	1.38	5.17
2015-01-16	0.54	0.41	0.47	1.22	0.62	1.39	5.13
2015-01-19	0.54	0.41	0.47	1.22	0.62	1.39	5.13
2015-01-20	0.55	0.44	0.47	1.22	0.62	1.38	5.13
2015-01-21	0.54	0.43	0.47	1.22	0.62	1.38	5.09
2015-01-22	0.53	0.38	0.47	1.22	0.61	1.36	5.01
2015-01-23	0.50	0.30	0.47	1.22	0.61	1.36	5.04
2015-01-26	0.51	0.33	0.47	1.23	0.61	1.35	5.00
2015-01-27	0.50	0.32	0.47	1.23	0.61	1.35	5.02
2015-01-28	0.51	0.35	0.47	1.23	0.62	1.35	5.04
2015-01-29	0.51	0.33	0.47	1.24	0.62	1.36	5.02

- **State 2: Control Area**

  After the user chooses a file and that file is loaded, a second panel appears beneath the File Selection Control. The Graph Type Control contains an instruction paragraph followed by an HTML radio button with graph types for the user to choose from.

- **State 2: Display Area**

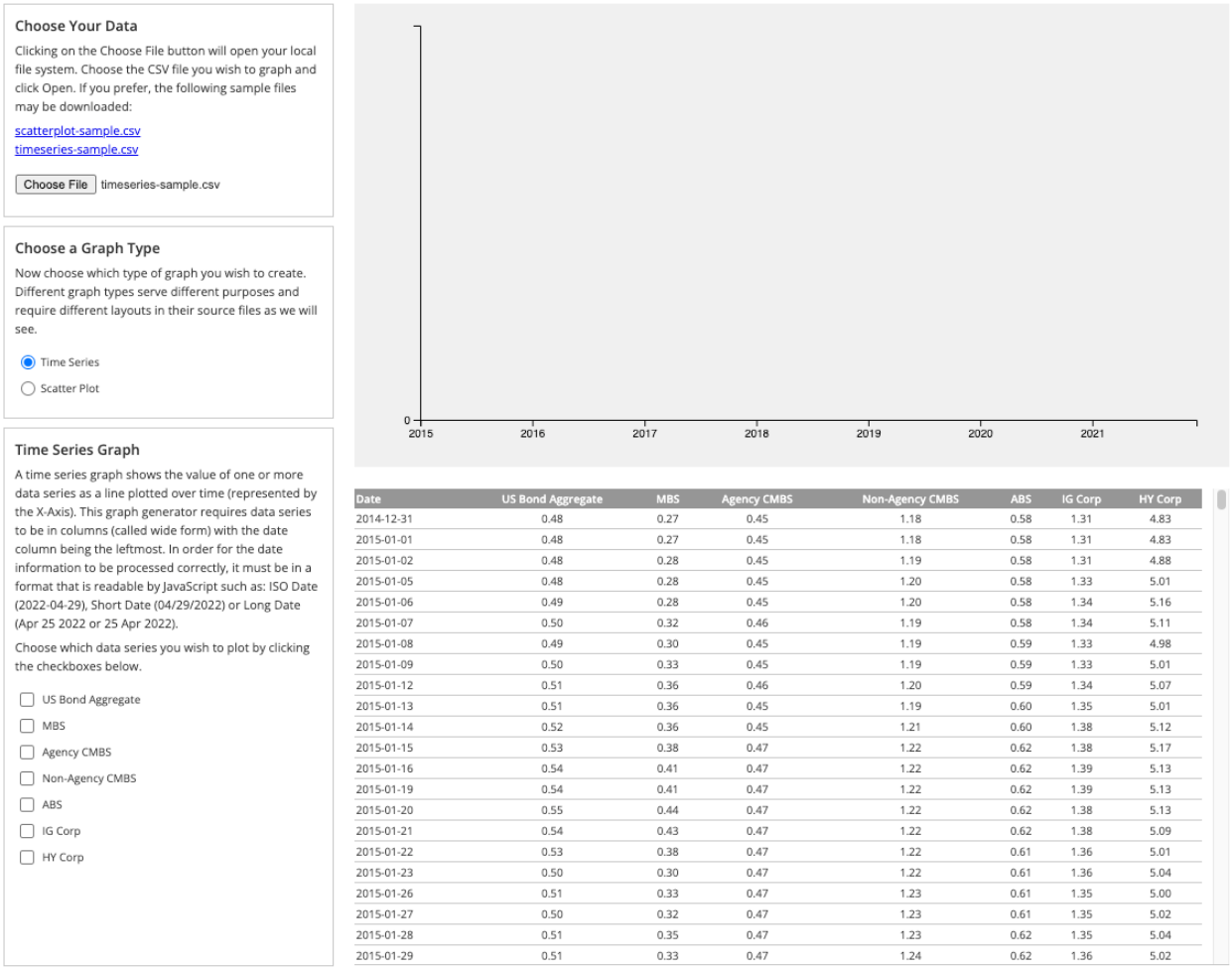
  The Graph section is unchanged in State 2, but the data from the previously selected file now appears as a scrollable table beneath it.

State 3 – Interactive Graph

Upon selection of a graph type, a third panel appears. As with the previous two panels, the first element is a set of instructions. In this case, however, the instructions are more crucial to the graphing function as they explain what form the data must take in order to be processed correctly.

Beneath the instructions are HTML input elements, the type of which depends on the user's graph selection.

State 3 – Time Series



- **State 3 – Time Series: Control Area**

  As mentioned above, the instructions in the Time Series panel lay out the specific requirements for this type of graph, that is, a JavaScript readable date column in the leftmost position and data series laid out in columns (wide form).

  At this point, the user may examine the data table in the Display area to determine whether the chosen data source conforms to the requirements. If it does not, the user has the option to change the selection in the Graph Type Control, and continue with the current data source, or choose a new data file in the File Selection Control and essentially reset the process.

Below the instructions is a checklist with an entry for each column appearing to the right of the date in the Table. Selecting any of these entries will plot its column's data in the Graph section of the Display area.

- **State 3 – Time Series: Display Area**

  Assuming that the data file is in the proper format, upon entering State 3, the X axis in the Graph section will show the date range of the data. As at this point no data series has been selected, the Y axis remains unmarked.

  Upon selection of a data series, the Y axis will display the data range. The scale is automatically adjusted to the extent of the data series and will automatically re-adjust as data series are added to and deleted from the graph. Each data series has a unique color which can be seen in three places: the line plotted on the graph itself as well as the corresponding table header in the Table section and checkbox in the Control area, either of which may serve as a legend for the graph.

Choose Your Data

Clicking on the Choose File button will open your local file system. Choose the CSV file you wish to graph and click Open. If you prefer, the following sample files may be downloaded:

scatterplot-sample.csv
timeseries-sample.csv

Choose File timeseries-sample.csv

Choose a Graph Type

Now choose which type of graph you wish to create. Different graph types serve different purposes and require different layouts in their source files as we will see.

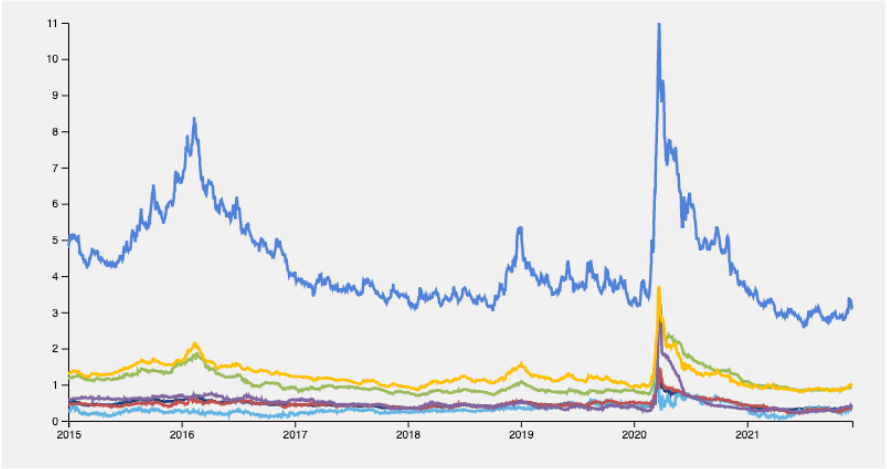
- (•) Time Series
- ( ) Scatter Plot

Time Series Graph

A time series graph shows the value of one or more data series as a line plotted over time (represented by the X-Axis). This graph generator requires data series to be in columns (called wide form) with the date column being the leftmost. In order for the date information to be processed correctly, it must be in a format that is readable by JavaScript such as: ISO Date (2022-04-29), Short Date (04/29/2022) or Long Date (Apr 25 2022 or 25 Apr 2022).

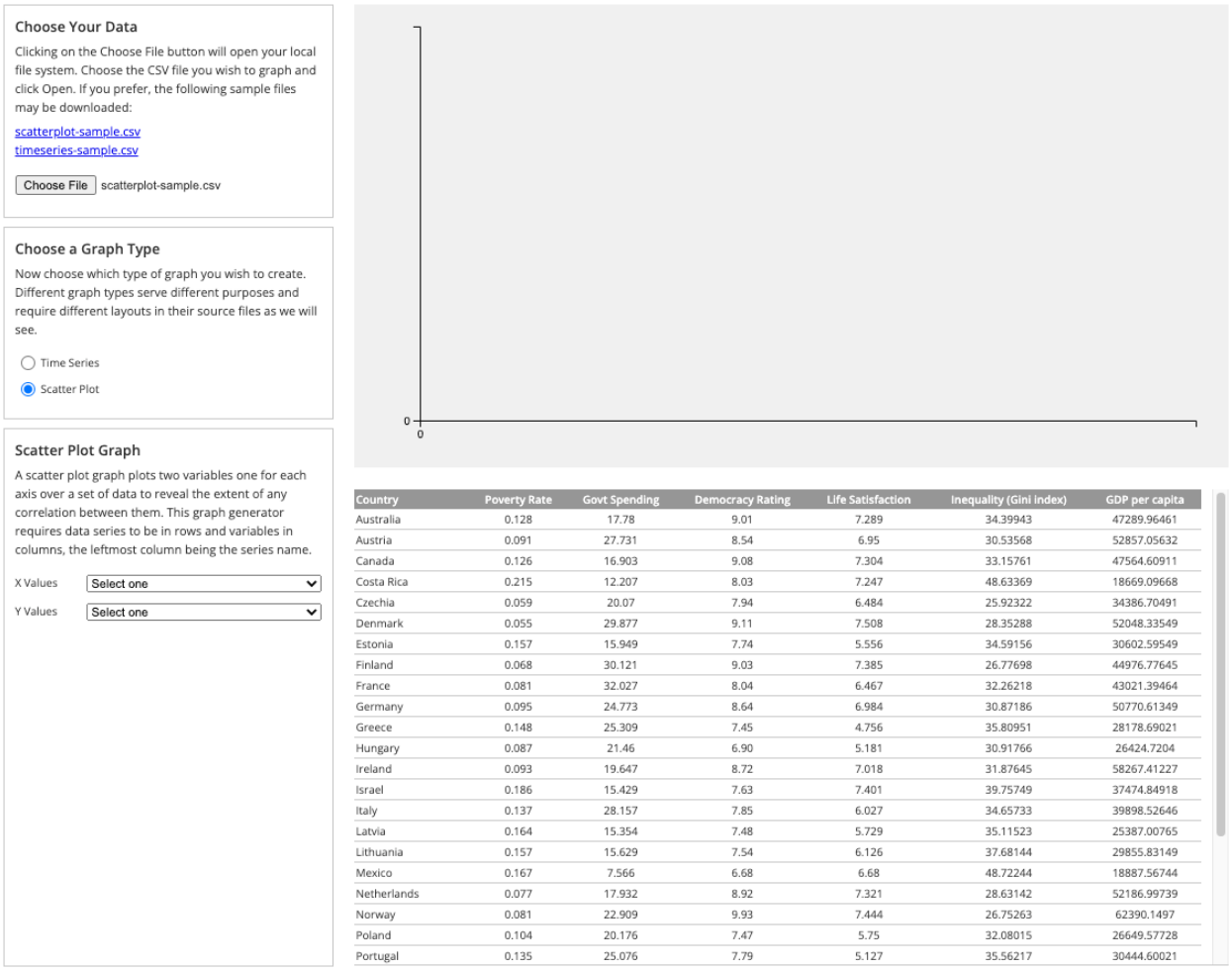
Choose which data series you wish to plot by clicking the checkboxes below.

- [x] US Bond Aggregate
- [x] MBS
- [x] Agency CMBS
- [x] Non-Agency CMBS
- [x] ABS
- [x] IG Corp
- [x] HY Corp



Date	US Bond Aggregate	MBS	Agency CMBS	Non-Agency CMBS	ABS	IG Corp	HY Corp
2014-12-31	0.48	0.27	0.45	1.18	0.58	1.31	4.83
2015-01-01	0.48	0.27	0.45	1.18	0.58	1.31	4.83
2015-01-02	0.48	0.28	0.45	1.19	0.58	1.31	4.88
2015-01-05	0.48	0.28	0.45	1.20	0.58	1.33	5.01
2015-01-06	0.49	0.28	0.45	1.20	0.58	1.34	5.16
2015-01-07	0.50	0.32	0.46	1.19	0.58	1.34	5.11
2015-01-08	0.49	0.30	0.45	1.19	0.59	1.33	4.98
2015-01-09	0.50	0.33	0.45	1.19	0.59	1.33	5.01
2015-01-12	0.51	0.36	0.46	1.20	0.59	1.34	5.07
2015-01-13	0.51	0.36	0.45	1.19	0.60	1.35	5.01
2015-01-14	0.52	0.36	0.45	1.21	0.60	1.38	5.12
2015-01-15	0.53	0.38	0.47	1.22	0.62	1.38	5.17
2015-01-16	0.54	0.41	0.47	1.22	0.62	1.39	5.13
2015-01-19	0.54	0.41	0.47	1.22	0.62	1.39	5.13
2015-01-20	0.55	0.44	0.47	1.22	0.62	1.38	5.13
2015-01-21	0.54	0.43	0.47	1.22	0.62	1.38	5.09
2015-01-22	0.53	0.38	0.47	1.22	0.61	1.36	5.01
2015-01-23	0.50	0.30	0.47	1.22	0.61	1.36	5.04
2015-01-26	0.51	0.33	0.47	1.23	0.61	1.35	5.00
2015-01-27	0.50	0.32	0.47	1.23	0.61	1.35	5.02
2015-01-28	0.51	0.35	0.47	1.23	0.62	1.35	5.04
2015-01-29	0.51	0.33	0.47	1.24	0.62	1.36	5.02

State 3 – Scatter Plot



Choose Your Data

Clicking on the Choose File button will open your local file system. Choose the CSV file you wish to graph and click Open. If you prefer, the following sample files may be downloaded:

scatterplot-sample.csv
timeseries-sample.csv

Choose File scatterplot-sample.csv

Choose a Graph Type

Now choose which type of graph you wish to create. Different graph types serve different purposes and require different layouts in their source files as we will see.

Time Series
Scatter Plot

Scatter Plot Graph

A scatter plot graph plots two variables one for each axis over a set of data to reveal the extent of any correlation between them. This graph generator requires data series to be in rows and variables in columns, the leftmost column being the series name.

X Values Select one
Y Values Select one

0
0

Country	Poverty Rate	Govt Spending	Democracy Rating	Life Satisfaction	Inequality (Gini index)	GDP per capita
Australia	0.128	17.78	9.01	7.289	34.39943	47289.96461
Austria	0.091	27.731	8.54	6.95	30.53568	52857.05632
Canada	0.126	16.903	9.08	7.304	33.15761	47564.60911
Costa Rica	0.215	12.207	8.03	7.247	48.63369	18669.09668
Czechia	0.059	20.07	7.94	6.484	25.92322	34386.70491
Denmark	0.055	29.877	9.11	7.508	28.35288	52048.33549
Estonia	0.157	15.949	7.74	5.556	34.59156	30602.59549
Finland	0.068	30.121	9.03	7.385	26.77698	44976.77645
France	0.081	32.027	8.04	6.467	32.26218	43021.39464
Germany	0.095	24.773	8.64	6.984	30.87186	50770.61349
Greece	0.148	25.309	7.45	4.756	35.80951	28178.69021
Hungary	0.087	21.46	6.90	5.181	30.91766	26424.7204
Ireland	0.093	19.647	8.72	7.018	31.87645	58267.41227
Israel	0.186	15.429	7.63	7.401	39.75749	37474.84918
Italy	0.137	28.157	7.85	6.027	34.65733	39898.52646
Latvia	0.164	15.354	7.48	5.729	35.11523	25387.00765
Lithuania	0.157	15.629	7.54	6.126	37.68144	29855.83149
Mexico	0.167	7.566	6.68	6.68	48.72244	18887.56744
Netherlands	0.077	17.932	8.92	7.321	28.63142	52186.99739
Norway	0.081	22.909	9.93	7.444	26.75263	62390.1497
Poland	0.104	20.176	7.47	5.75	32.08015	26649.57728
Portugal	0.135	25.076	7.79	5.127	35.56217	30444.60021

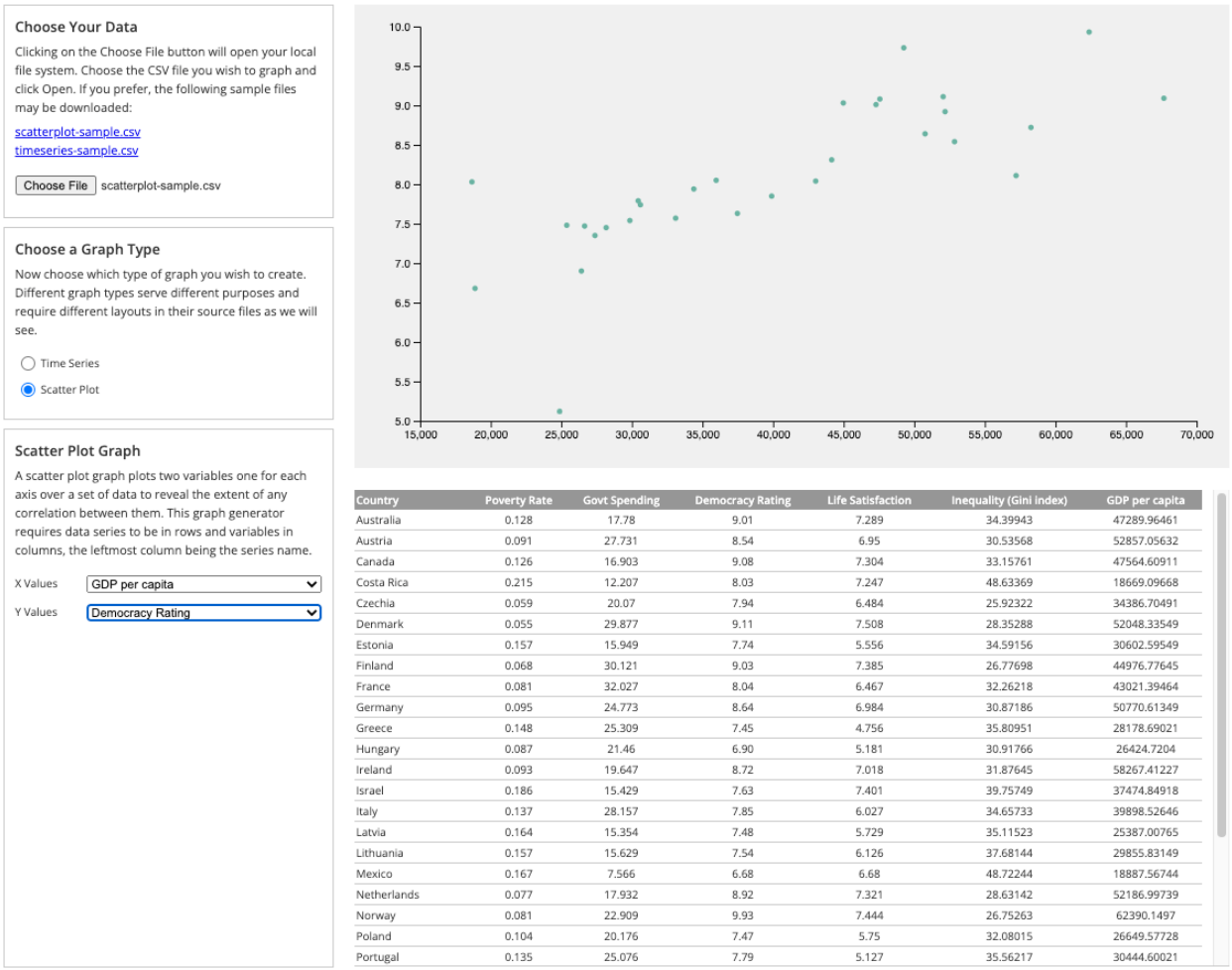
- **State 3 – Scatter Plot: Control Area**

  Instructions for a Scatter Plot graph, like those for the Time Series above detail data format required for processing. Data series for this graph type must be in rows with variables in columns not including the leftmost, which is reserved for a series name. The Scatter Plot graph is a bit more forgiving than the Time Series as it will accept and plot any column containing numeric data (the data in the timeseries-sample.csv file for instance, will be drawn correctly in a Scatter Plot, though the relationship between the variables may well be meaningless.

In place of the checklist shown with the Time Series, the Scatter Plot Control contains two drop-down Select inputs, one for each axis and allowing only one selection each.

- **State 3 – Scatter Plot: Display Area**

  When a data series is selected for either axis, that axis will display the range for that series. After both axes have been selected, the data will be plotted on the Graph. Changing the data series for either axis will automatically re-plot the data and adjust its axis accordingly.



Country	Poverty Rate	Govt Spending	Democracy Rating	Life Satisfaction	Inequality (Gini index)	GDP per capita
Australia	0.128	17.78	9.01	7.289	34.39943	47289.96461
Austria	0.091	27.731	8.54	6.95	30.53568	52857.05632
Canada	0.126	16.903	9.08	7.304	33.15761	47564.60911
Costa Rica	0.215	12.207	8.03	7.247	48.63369	18669.09668
Czechia	0.059	20.07	7.94	6.484	25.92322	34386.70491
Denmark	0.055	29.877	9.11	7.508	28.35288	52048.33549
Estonia	0.157	15.949	7.74	5.556	34.59156	30602.59549
Finland	0.068	30.121	9.03	7.385	26.77698	44976.77645
France	0.081	32.027	8.04	6.467	32.26218	43021.39464
Germany	0.095	24.773	8.64	6.984	30.87186	50770.61349
Greece	0.148	25.309	7.45	4.756	35.80951	28178.69021
Hungary	0.087	21.46	6.90	5.181	30.91766	26424.7204
Ireland	0.093	19.647	8.72	7.018	31.87645	58267.41227
Israel	0.186	15.429	7.63	7.401	39.75749	37474.84918
Italy	0.137	28.157	7.85	6.027	34.65733	39898.52646
Latvia	0.164	15.354	7.48	5.729	35.11523	25387.00765
Lithuania	0.157	15.629	7.54	6.126	37.68144	29855.83149
Mexico	0.167	7.566	6.68	6.68	48.72244	18887.56744
Netherlands	0.077	17.932	8.92	7.321	28.63142	52186.99739
Norway	0.081	22.909	9.93	7.444	26.75263	62390.1497
Poland	0.104	20.176	7.47	5.75	32.08015	26649.57728
Portugal	0.135	25.076	7.79	5.127	35.56217	30444.60021

Domain Expert Evaluation and Feedback

For expert evaluation I enlisted Mary Beth Fisher, PhD, a senior financial analyst with Amherst Pierpont Securities in New York. Her evaluation of the prototype overall was positive for the most part, particularly concerning the simplicity of design and ease of use, according to her comments:

- **Ease of use**

  Constructing graphs was effortless and "…so much easier than doing it myself in Excel."

- **Simplicity of Design**

  "I like that the tool is simple, I've tried to use the super fancy visualization tools and I usually give up because trying to remember how to use them and do plots is too much trouble."

  This comment calls to mind Jakob Nielsen's 6th Usability Heuristic, "Recognition Rather than Recall":

  > *The user should not have to remember information from one part of the interface to another. Information required to use the design (e.g., field labels or menu items) should be visible or easily retrievable when needed.*[1]

Certain other aspects of the design, however, were found to be either unintuitive or lacked sufficient feedback when performing tasks:

- **Instructions on File Loading are Unclear**

  The initial file loading control (State 1) was found to be confusing in that it is not made clear in the instructions that the tool expects the user to upload their own data and that the sample links provided were for download and would not open directly within the app.

[1] Jakob Nielsen, 10 Usability Heuristics for User Interface Design

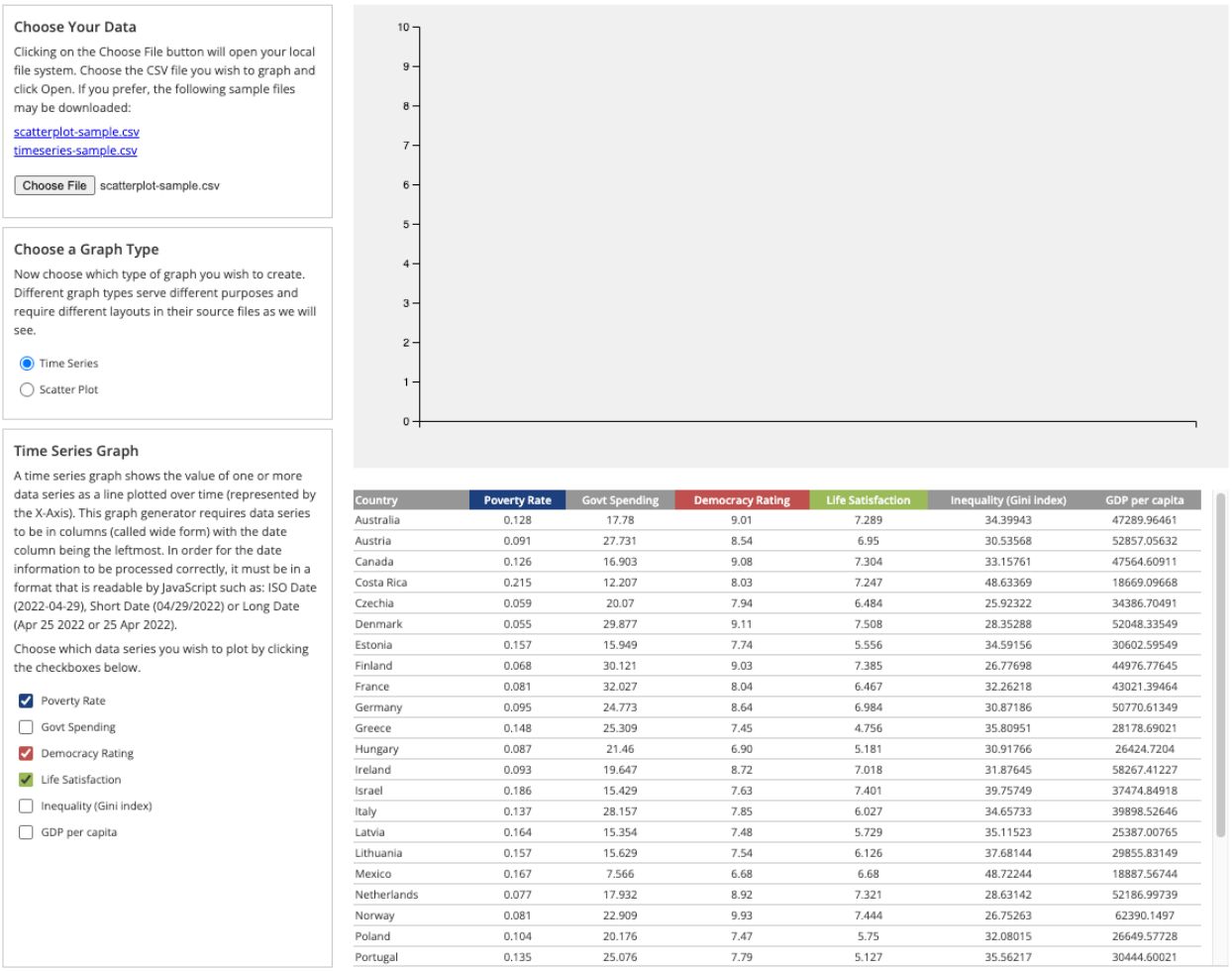
Two possible approaches to this problem would be:

- Clearer instructions at the beginning of State 1 and more separation between the downloadable samples and the File Selection Control, i.e., removing the sample links to their own panel, or
- Re-coding the links to perform as she said she expected them to, i.e., to load the sample files into to tool rather than download to the user's file system.

- **Tool continues to operate when data file and graph type are incompatible**

  Mary Beth noted that at one point she "…accidentally tried to plot the scatter data as a timeseries. It looked like it was going to do it. Then I realized I had to choose the timeseries file." This comment highlights two issues in the design:

  - **Shifting responsibility to the user that should belong to the program** – It should not be up to the user to ensure the data format of their source file is compatible with their selected graph type. The tool needs to perform its own compatibility test.
  - **Continuing to operate when incompatible data is used** – The tool should halt when it detects incompatibility between data and chart type. I reproduced this error (see below). Rather than continuing in a non-sensical state as the screenshot shows, the tool should give the user feedback on the error and how to correct it (see Nielsen's Usability Heuristics #9).



This screenshot shows the results of loading a scatter plot formatted data file and selecting a Time Series graph type. While the user can still interact with the Main Control, the prototype is unable to process the data and the Graph section of the Display area is blank, save for the meaningless adjustment of the Y axis.

Her suggestions for "nice to have" improvements to the prototype are as follows:

- **More textual information in or around the Graph**

  More information (including non-numerical) about the data series could be shown automatically or through some type of interaction with the Graph e.g., on rollover.

- **Improved statistical functionality**

  The Scatter Plot graph could use functionality such as a linear regression line and display of R-squared function for instance.

Wish List for a Full Version of this Web Application

To conclude this paper, I'd like to speculate of what a full, production version of this web application might look like. As a prototype the app is by necessity bare-bones simple, but in the course of development ideas were continually occurring to me of what features might be added if I were to continue developing this tool. These, plus some suggested by my domain expert reviewer are enumerated below:

- **More graph types**

  The most obvious improvement would be to give users a wider selection of graph types to choose from

- **More comprehensive instructions**

  If the app is to serve as a teaching tool, it would need to have more than the brief explanations of various graph types and data formats of the prototype. My models for an interactive teaching tool would be Josh Comeau's excellent "Let's Learn about Waveforms" and Marijn Haverbeke's interactive online textbook, Eloquent JavaScript.

- **Automated Compatibility Detection**

  Suggested by my domain expert, the app should be able to detect whether data sources are compatible with chart types and throw an error if not.

- **Ability to process "long-form" data files**

  Being able to read this form of data, as programs like Tableau do, would enable more complex interactive visualizations such as Gapminder-style motion charts.

- **Additional data series information**

The graph could show additional, not necessarily quantitative information as tooltips or other similar sidebar views, on rollover or other interaction.

References and Acknowledgements

Scott Murray, *Interactive Data Visualization for the Web: An Introduction to Designing with D3*, ©2017 O'Reilly Media Inc.

The D3.js Graph Gallery, the basics of graph construction in D3, maintained by Yan Holtz

D3 Data Driven Documents, the home of the D3 library maintained by its developer Mike Bostock

Stack Overflow Questions Tagged D3.js, an invaluable resource

Special thanks to Mary Beth Fisher for her review and feedback.