

\*

Connecting the Dots

- or -

*Using Plotly for Straight Line and Broken Line Plots*

ISC1057

Janet Peterson and John Burkardt

Computational Thinking

Fall Semester 2016

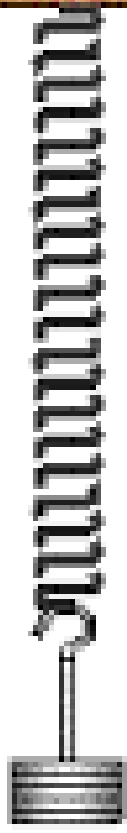
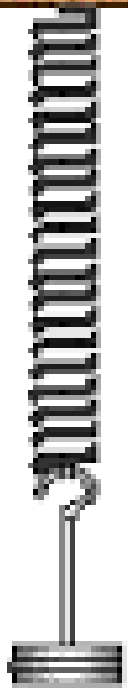
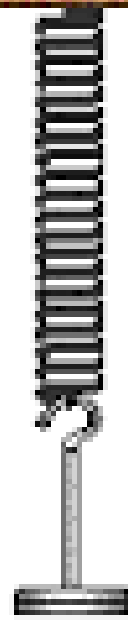
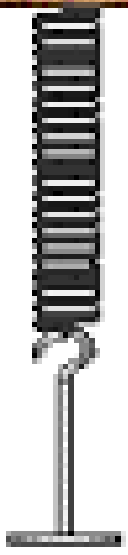
## Example 1: Stretching a Spring

A rubber band is useful because it has a property known as **elasticity**.

We know a few things about a rubber band:

- it has a natural length of one or two inches;
- as we pull on it harder and harder, it stretches to six or seven inches;
- if we let go, it snaps back;
- if we pull too hard, it breaks.

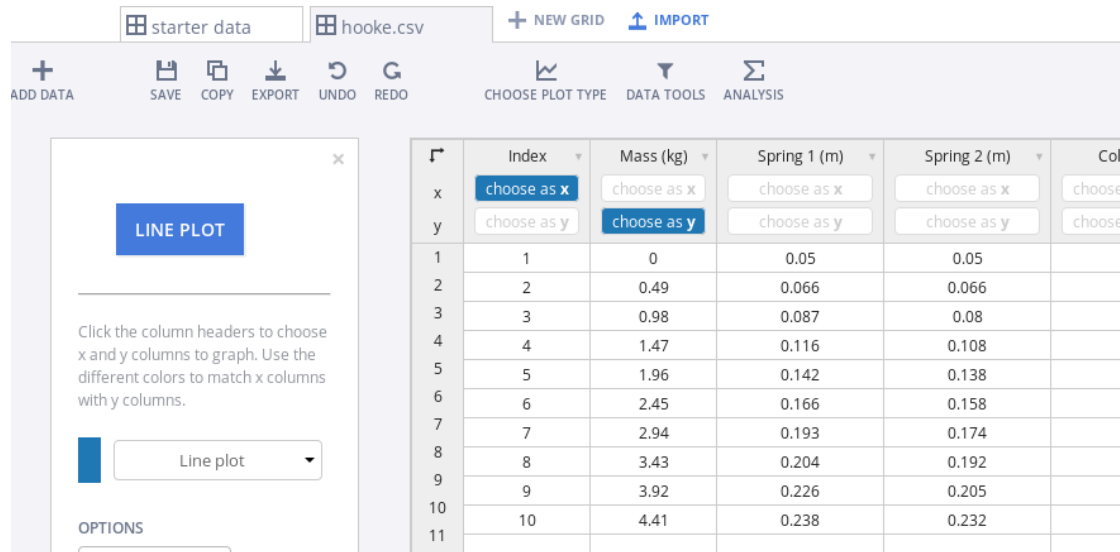
Robert Hooke searched for a pattern to describe the stretching of a spring.



Here is sample data stored in the file *hooke.csv*, in which two springs were stretched by adding one more mass at a time to the end of the spring and measuring the position of the end of the spring.

| "Index", | "Mass (kg)",     | "Spring 1 (m)", | "Spring 2 (m)" |
|----------|------------------|-----------------|----------------|
| 1,       | 0.00,            | 0.050,          | 0.050          |
| 2,       | 0.49,            | 0.066,          | 0.066          |
| 3,       | 0.98,            | 0.087,          | 0.080          |
| —        | <i>more data</i> | —               | —              |
| 8,       | 3.43,            | 0.204,          | 0.192          |
| 9,       | 3.92,            | 0.226,          | 0.205          |
| 10,      | 4.41,            | 0.238,          | 0.232          |

To examine this data, we can log into Plotly, then use the “import” menu to browse for the file *hooke.csv* and bring it into the Plotly grid, so that we see something like this:

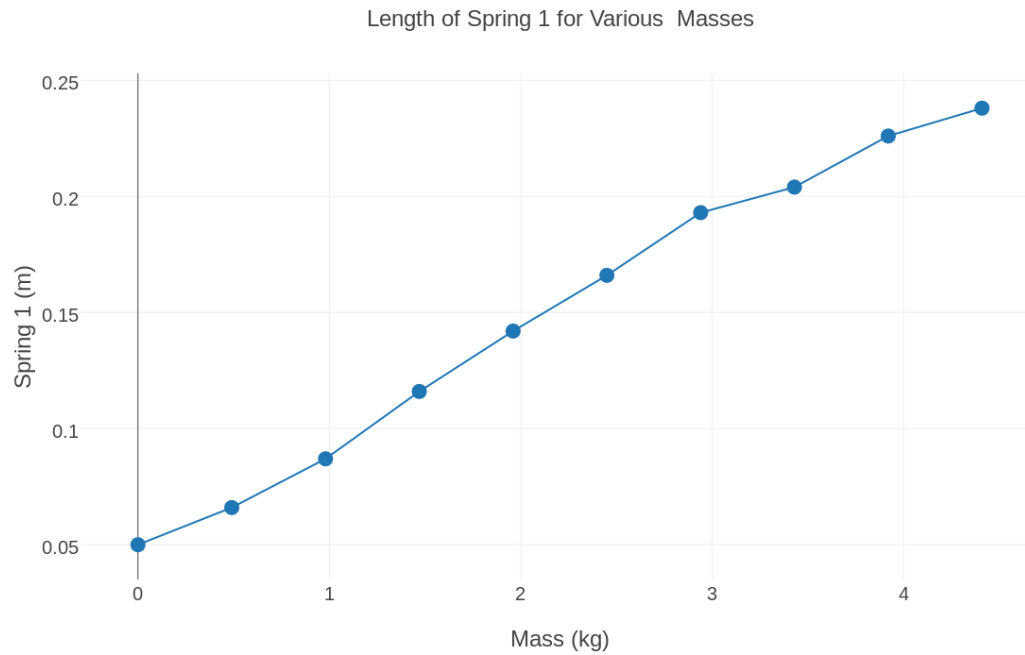


We need to change “Choose as x” and “Choose as y” so that we plot Mass versus Spring 1 Length,

The screenshot shows a software interface for data analysis. At the top, there are tabs for 'starter data' and 'hooke.csv', along with buttons for '+ NEW GRID' and 'IMPORT'. Below the tabs is a toolbar with icons for 'DATA', 'SAVE', 'COPY', 'EXPORT', 'UNDO', 'REDO', 'CHOOSE PLOT TYPE', 'DATA TOOLS', and 'ANALYSIS'. On the left, a 'LINE PLOT' configuration panel is open, showing a dropdown menu set to 'Line plot' and an 'OPTIONS' section. The main area displays a table with the following data:

|    | Index       | Mass (kg)   | Spring 1 (m) | Spring 2 (m) | Col    |
|----|-------------|-------------|--------------|--------------|--------|
| x  | choose as x | choose as x | choose as x  | choose as x  | choose |
| y  | choose as y | choose as y | choose as y  | choose as y  | choose |
| 1  | 1           | 0           | 0.05         | 0.05         |        |
| 2  | 2           | 0.49        | 0.066        | 0.066        |        |
| 3  | 3           | 0.98        | 0.087        | 0.08         |        |
| 4  | 4           | 1.47        | 0.116        | 0.108        |        |
| 5  | 5           | 1.96        | 0.142        | 0.138        |        |
| 6  | 6           | 2.45        | 0.166        | 0.158        |        |
| 7  | 7           | 2.94        | 0.193        | 0.174        |        |
| 8  | 8           | 3.43        | 0.204        | 0.192        |        |
| 9  | 9           | 3.92        | 0.226        | 0.205        |        |
| 10 | 10          | 4.41        | 0.238        | 0.232        |        |
| 11 |             |             |              |              |        |

Now we have a “broken line” plot of the data for spring 1.



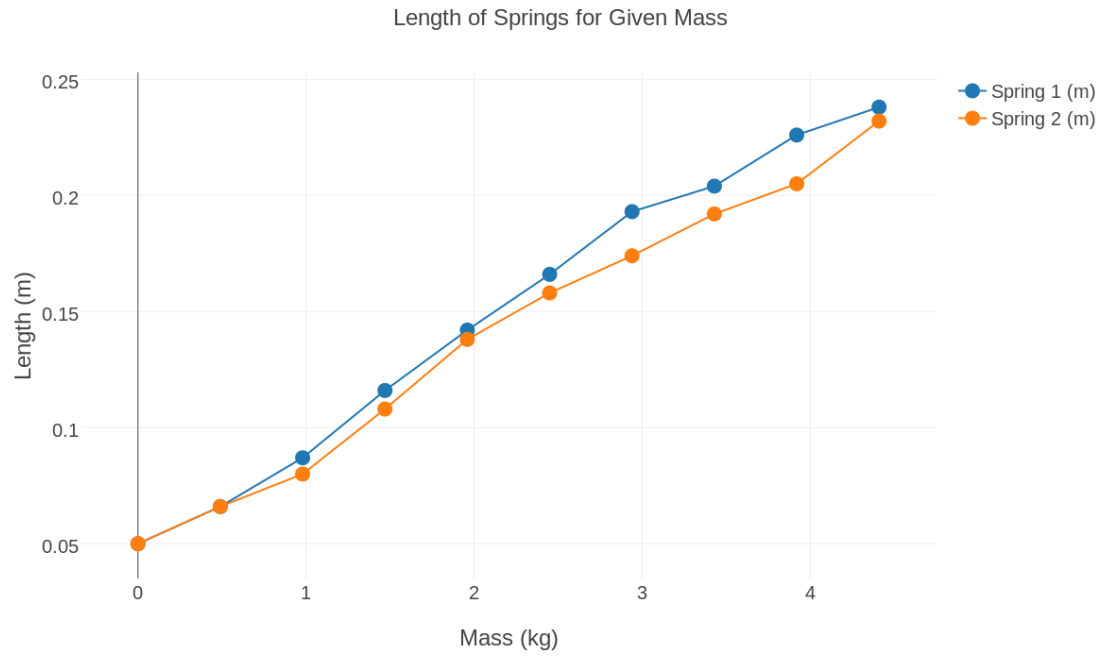
We can go back to the data grid and add the Spring 2 data using “choose as y”

The screenshot shows a software interface for data analysis. At the top, there are tabs for 'starter data', 'hooke.csv', and 'Plot'. Below the tabs is a toolbar with icons for '+ ADD DATA', 'SAVE', 'COPY', 'EXPORT', 'UNDO', 'REDO', 'CHOOSE PLOT TYPE', 'DATA TOOLS', and 'ANALYSIS'. On the left, a 'LINE PLOT' configuration panel is open, showing a 'Line plot' dropdown menu. The main area is a data grid with the following columns: Index, Mass (kg), Spring 1 (m), Spring 2 (m), and Co. The grid contains 10 rows of data. The 'x' and 'y' rows at the top of the grid have dropdown menus for selecting columns. The 'Mass (kg)' column is selected for 'x' and the 'Spring 2 (m)' column is selected for 'y'.

|    | Index       | Mass (kg)   | Spring 1 (m) | Spring 2 (m) | Co     |
|----|-------------|-------------|--------------|--------------|--------|
| x  | choose as x | choose as x | choose as x  | choose as x  | choose |
| y  | choose as y | choose as y | choose as y  | choose as y  | choose |
| 1  | 1           | 0           | 0.05         | 0.05         |        |
| 2  | 2           | 0.49        | 0.066        | 0.066        |        |
| 3  | 3           | 0.98        | 0.087        | 0.08         |        |
| 4  | 4           | 1.47        | 0.116        | 0.108        |        |
| 5  | 5           | 1.96        | 0.142        | 0.138        |        |
| 6  | 6           | 2.45        | 0.166        | 0.158        |        |
| 7  | 7           | 2.94        | 0.193        | 0.174        |        |
| 8  | 8           | 3.43        | 0.204        | 0.192        |        |
| 9  | 9           | 3.92        | 0.226        | 0.205        |        |
| 10 | 10          | 4.41        | 0.238        | 0.232        |        |
| 11 |             |             |              |              |        |



The blue and orange lines are for springs 1 and 2.

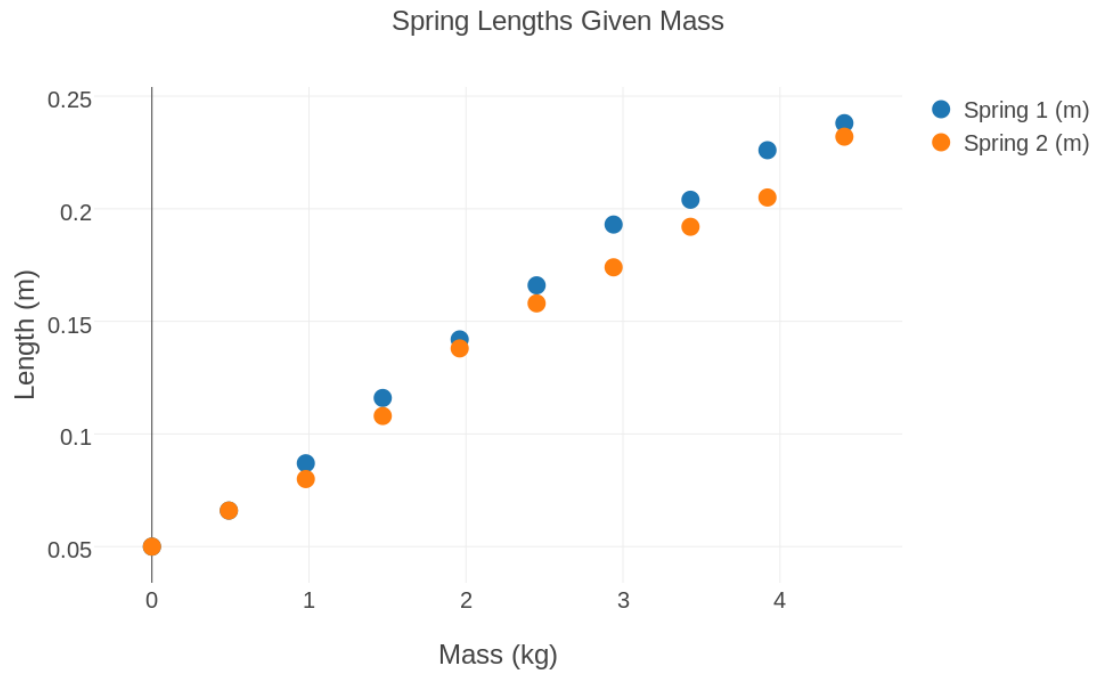


Our data doesn't lie on straight lines. Let's redraw the points as scatterplots.

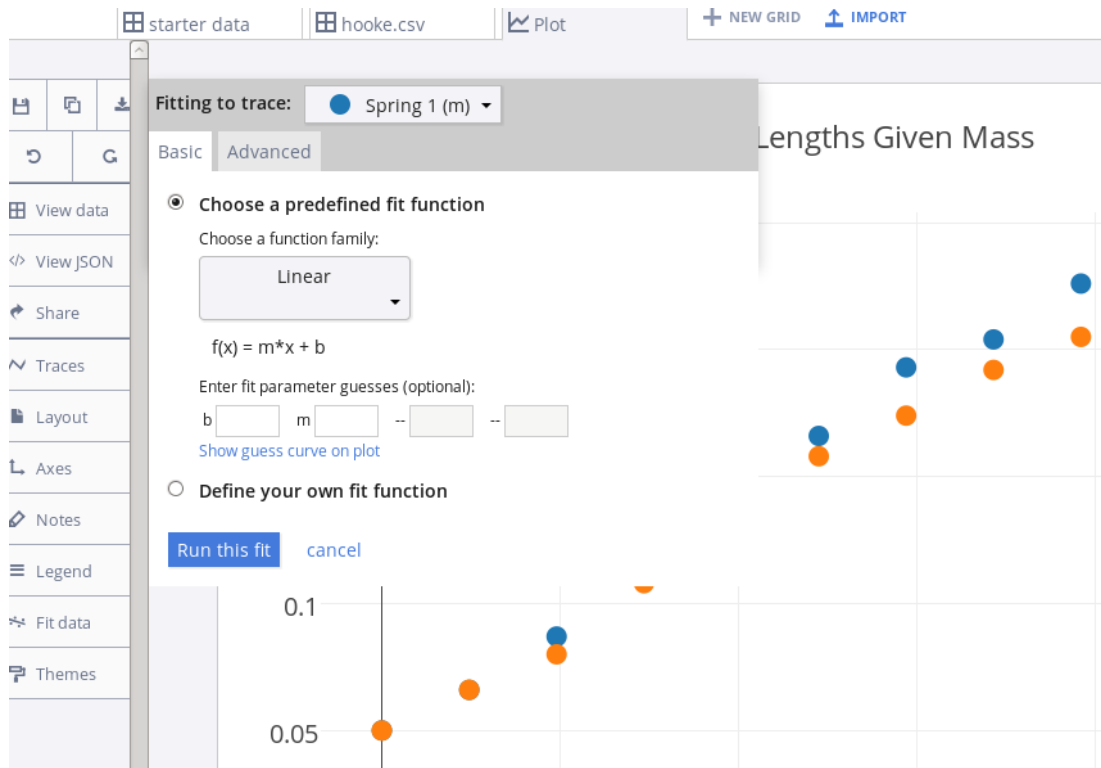
The screenshot shows the Plotly web application interface in a Mozilla Firefox browser. The browser address bar shows the URL <https://plot.ly/210/~jburkardt/#>. The Plotly logo and tagline "make charts and dashboards online" are visible at the top. The main workspace contains a data table with the following columns: Index, Mass (kg), Spring 1 (m), Spring 2 (m), Col5, Col6, and Col7. The data rows are numbered 1 through 10. A configuration panel on the left is open, showing a dropdown menu for "LINE PLOT" with options: Line plot, Scatter plot (selected), Bar chart, Histogram, Area plot, and Text. Below the dropdown is an "INSERT INTO" section with a "Make a new plot" button. A notification at the bottom of the browser window reads: "It looks like you haven't started Firefox in a while. Do you want to clean it up for a fresh, like-new experience? And by the way, welcome back! Refresh Firefox... x".

|    | Index       | Mass (kg)   | Spring 1 (m) | Spring 2 (m) | Col5        | Col6        | Col7        |
|----|-------------|-------------|--------------|--------------|-------------|-------------|-------------|
| x  | choose as x | choose as x | choose as x  | choose as x  | choose as x | choose as x | choose as x |
| y  | choose as y | choose as y | choose as y  | choose as y  | choose as y | choose as y | choose as y |
| 1  | 1           | 0           | 0.05         | 0.05         |             |             |             |
| 2  | 2           | 0.49        | 0.066        | 0.066        |             |             |             |
| 3  | 3           | 0.98        | 0.087        | 0.08         |             |             |             |
| 4  | 4           | 1.47        | 0.116        | 0.108        |             |             |             |
| 5  | 5           | 1.96        | 0.142        | 0.138        |             |             |             |
| 6  | 6           | 2.45        | 0.166        | 0.158        |             |             |             |
| 7  | 7           | 2.94        | 0.193        | 0.174        |             |             |             |
| 8  | 8           | 3.43        | 0.204        | 0.192        |             |             |             |
| 9  | 9           | 3.92        | 0.226        | 0.205        |             |             |             |
| 10 | 10          | 4.41        | 0.238        | 0.232        |             |             |             |

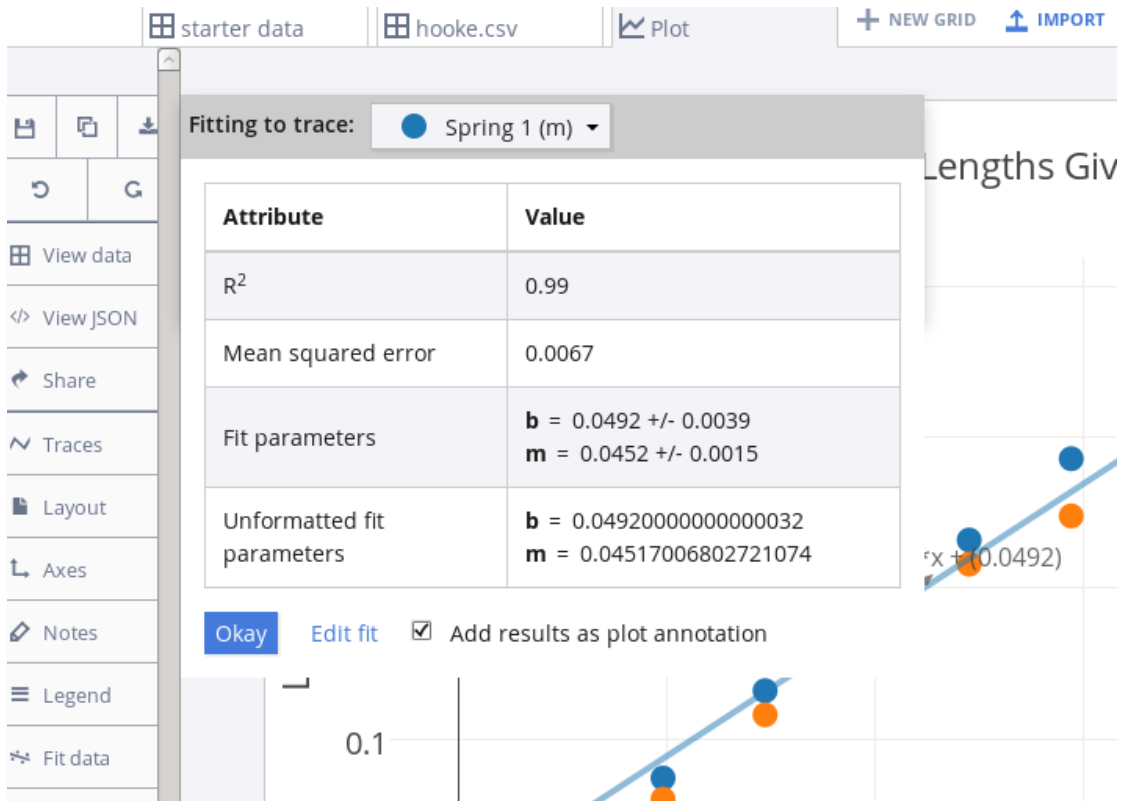
Here is just the data.



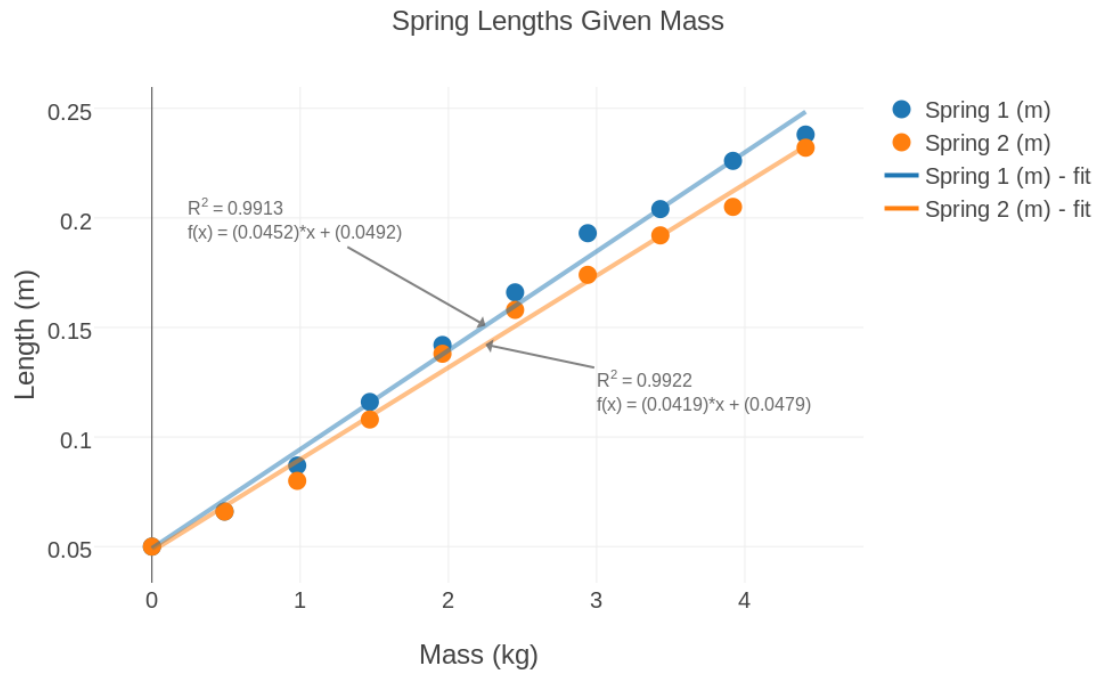
By using the **Fit data** menu, we can request a line like the (blue) data.



Plotly finds a formula  $y = 0.0452x + 0.0492$  that best fits the blue data.



We can also get a fitting line for the orange data.



## Example 2: Housing Data from Zillow

A realtor considers many factors when suggesting a listing price for a house:

- the square footage;
- number of bedrooms and bathrooms;
- the zip code;
- the year the house was built.

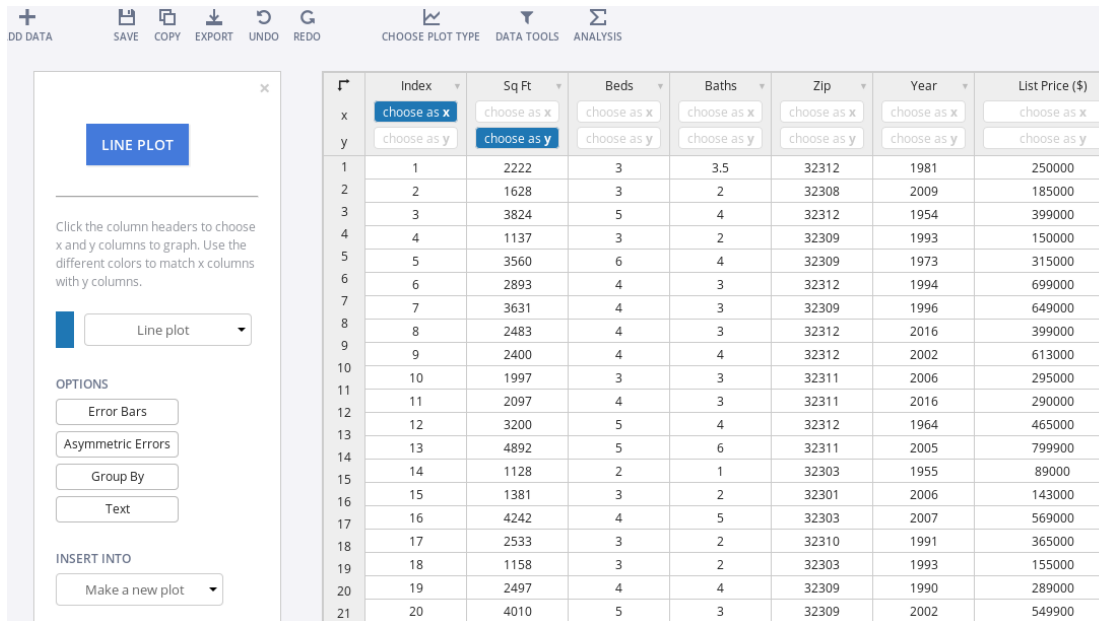
Experience suggests that the square footage is one of the most important.

Here is sample data stored in the file *zillow.csv*

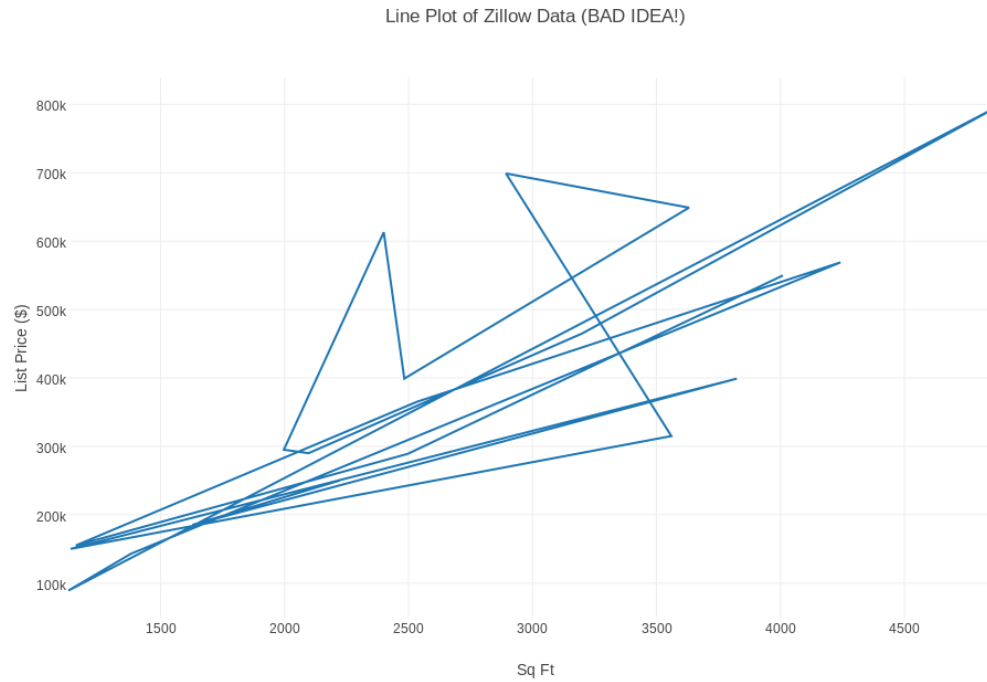
| "Index", | "(Sq Ft)", | "Beds",          | "Baths", | "Zip", | "Year", | "List Price (\$)" |
|----------|------------|------------------|----------|--------|---------|-------------------|
| 1,       | 2222,      | 3,               | 3.5,     | 32312, | 1981,   | 250000            |
| 2,       | 1628,      | 3,               | 2,       | 32308, | 2009,   | 185000            |
| 3,       | 3824,      | 5,               | 4,       | 32312, | 1954,   | 399000            |
| —        | —          | <i>more data</i> | —        | —      | —       | —                 |
| 18,      | 1158,      | 3,               | 2,       | 32303, | 1993,   | 155000            |
| 19,      | 2497,      | 4,               | 4,       | 32309, | 1990,   | 289000            |
| 20,      | 4010,      | 5,               | 3,       | 32309, | 2002,   | 549900            |



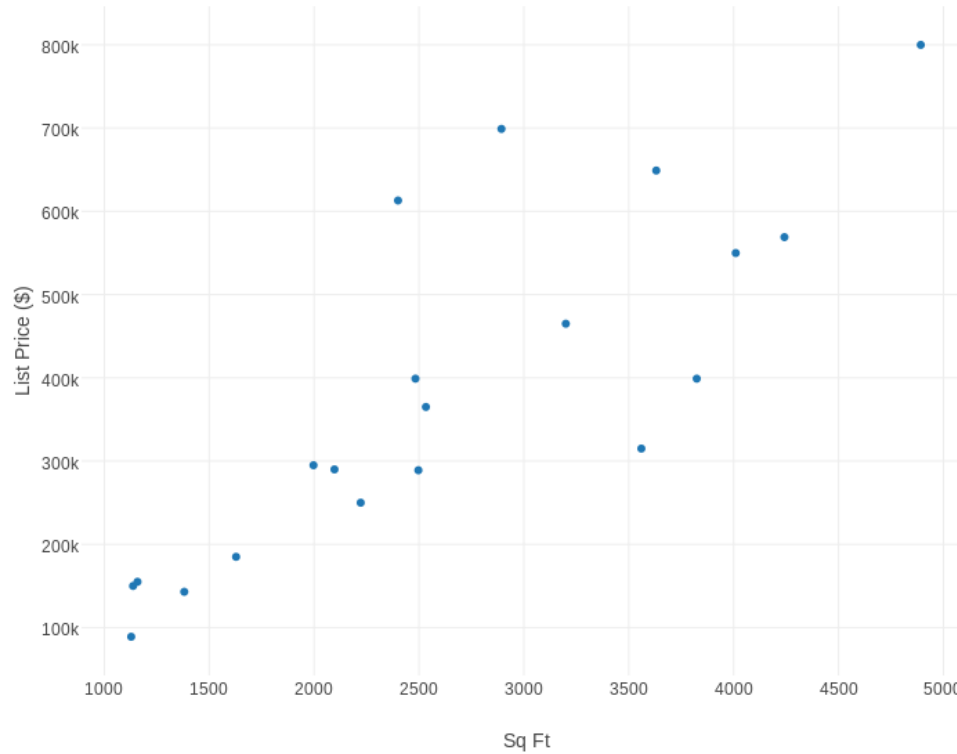
To examine this data, we can log into Plotly, then use the “import” menu to browse for the file *zillow.csv* and bring it into the Plotly grid, so that we see something like this:



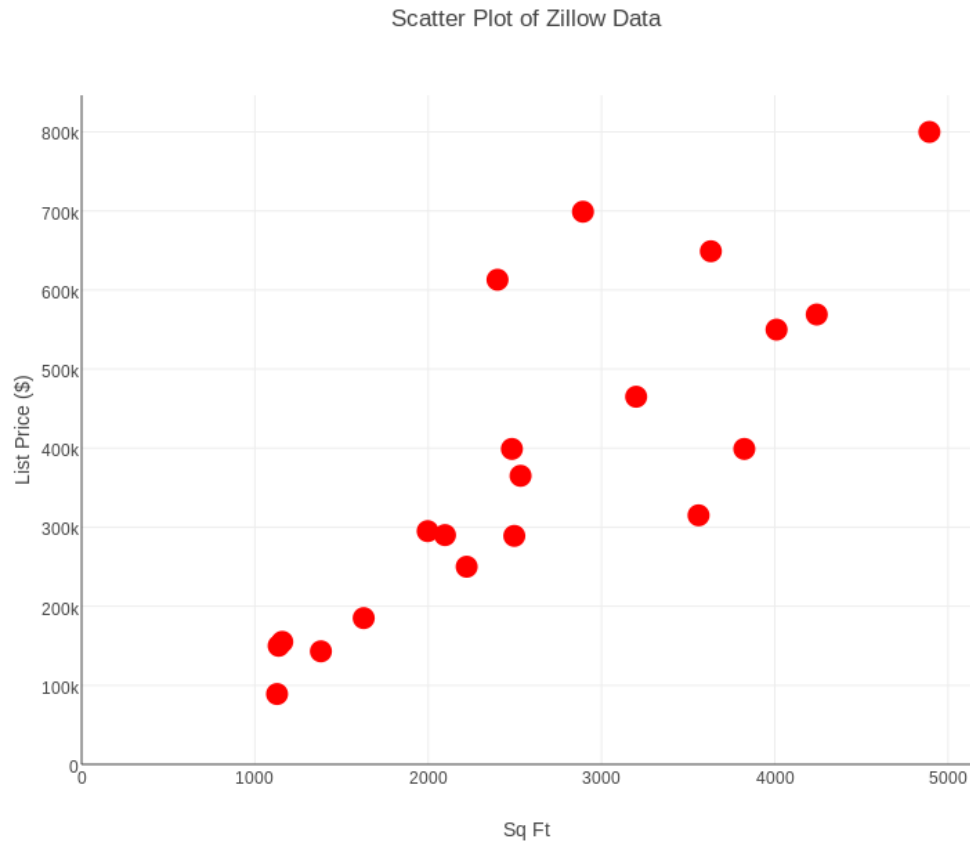
We can choose as  $x$  square footage, and choose as  $y$  the price and hit Line Plot, but here's what we get:



Our data is not suitable for line plots; we can make a scatter plot.



The **Traces** menu changes the size and color of dots. The **Axes** menu shifts the graph so we see the zero values for Footage and Price.



Now it seems like there is a general trend of the data, that the list price increases with the square footage.

If we think about it carefully, we can even find a formula that exact predicts the relationship between footage and price, if we just consider the first two pairs of data:

footage  $x = 2,222$ , price  $y = \$250,000$

footage  $x = 1,628$ , price  $y = \$185,000$

The formula is  $y = 6851.85 + 109.428x$ . We'd like to compare this formula to all 20 of our data points.

If we select the **Data Tools** item, and then **formula f(x)**, we see a box to enter our formula:

The screenshot shows a data analysis tool interface. On the left, a dialog box titled "RUN" is open, containing the following text: "Choose x and/or y columns as function variables, write an expression in the box below, then click 'Run'." Below this text is a text input field with the formula  $6851.85 + 109.428 * x$  and a dropdown menu showing "abs". A blue "RUN" button is positioned above the text. The main area of the interface displays a data table with the following columns: Index, Sq Ft, Beds, Baths, Zip, Year, and List Price (\$). The table contains 16 rows of data.

|    | Index       | Sq Ft       | Beds        | Baths       | Zip         | Year        | List Price (\$) |
|----|-------------|-------------|-------------|-------------|-------------|-------------|-----------------|
| x  | choose as x | choose as x | choose as x | choose as x | choose as x | choose as x | choose as x     |
| y  | choose as y | choose as y | choose as y | choose as y | choose as y | choose as y | choose as y     |
| 1  | 1           | 2222        | 3           | 3.5         | 32312       | 1981        | 250000          |
| 2  | 2           | 1628        | 3           | 2           | 32308       | 2009        | 185000          |
| 3  | 3           | 3824        | 5           | 4           | 32312       | 1954        | 399000          |
| 4  | 4           | 1137        | 3           | 2           | 32309       | 1993        | 150000          |
| 5  | 5           | 3560        | 6           | 4           | 32309       | 1973        | 315000          |
| 6  | 6           | 2893        | 4           | 3           | 32312       | 1994        | 699000          |
| 7  | 7           | 3631        | 4           | 3           | 32309       | 1996        | 649000          |
| 8  | 8           | 2483        | 4           | 3           | 32312       | 2016        | 399000          |
| 9  | 9           | 2400        | 4           | 4           | 32312       | 2002        | 613000          |
| 10 | 10          | 1997        | 3           | 3           | 32311       | 2006        | 295000          |
| 11 | 11          | 2097        | 4           | 3           | 32311       | 2016        | 290000          |
| 12 | 12          | 3200        | 5           | 4           | 32312       | 1964        | 465000          |
| 13 | 13          | 4892        | 5           | 6           | 32311       | 2005        | 799900          |
| 14 | 14          | 1128        | 2           | 1           | 32303       | 1955        | 89000           |
| 15 | 15          | 1381        | 3           | 2           | 32301       | 2006        | 143000          |
| 16 |             |             |             |             |             |             |                 |

Using square footage as  $x$ , Plotly creates a new column of formula values.

|    | Index       | Sq Ft       | Beds        | Baths       | Zip         | Year        | List Price (\$) | 6851.85+109.42*x   |
|----|-------------|-------------|-------------|-------------|-------------|-------------|-----------------|--------------------|
| x  | choose as x | choose as x | choose as x | choose as x | choose as x | choose as x | choose as x     | choose as x        |
| y  | choose as y | choose as y | choose as y | choose as y | choose as y | choose as y | choose as y     | choose as y        |
| 1  | 1           | 2222        | 3           | 3.5         | 32312       | 1981        | 250000          | 249983.09          |
| 2  | 2           | 1628        | 3           | 2           | 32308       | 2009        | 185000          | 184987.61000000002 |
| 3  | 3           | 3824        | 5           | 4           | 32312       | 1954        | 399000          | 425273.93          |
| 4  | 4           | 1137        | 3           | 2           | 32309       | 1993        | 150000          | 131262.39          |
| 5  | 5           | 3560        | 6           | 4           | 32309       | 1973        | 315000          | 396387.05          |
| 6  | 6           | 2893        | 4           | 3           | 32312       | 1994        | 699000          | 323403.91          |
| 7  | 7           | 3631        | 4           | 3           | 32309       | 1996        | 649000          | 404155.87          |
| 8  | 8           | 2483        | 4           | 3           | 32312       | 2016        | 399000          | 278541.70999999996 |
| 9  | 9           | 2400        | 4           | 4           | 32312       | 2002        | 613000          | 269459.85          |
| 10 | 10          | 1997        | 3           | 3           | 32311       | 2006        | 295000          | 225363.59          |
| 11 | 11          | 2097        | 4           | 3           | 32311       | 2016        | 290000          | 236305.59          |
| 12 | 12          | 3200        | 5           | 4           | 32312       | 1964        | 465000          | 356995.85          |
| 13 | 13          | 4892        | 5           | 6           | 32311       | 2005        | 799900          | 542134.49          |
| 14 | 14          | 1128        | 2           | 1           | 32303       | 1955        | 89000           | 130277.61          |
| 15 | 15          | 1381        | 3           | 2           | 32301       | 2006        | 143000          | 157960.87          |
| 16 | 16          | 4242        | 4           | 5           | 32303       | 2007        | 569000          | 471011.49          |

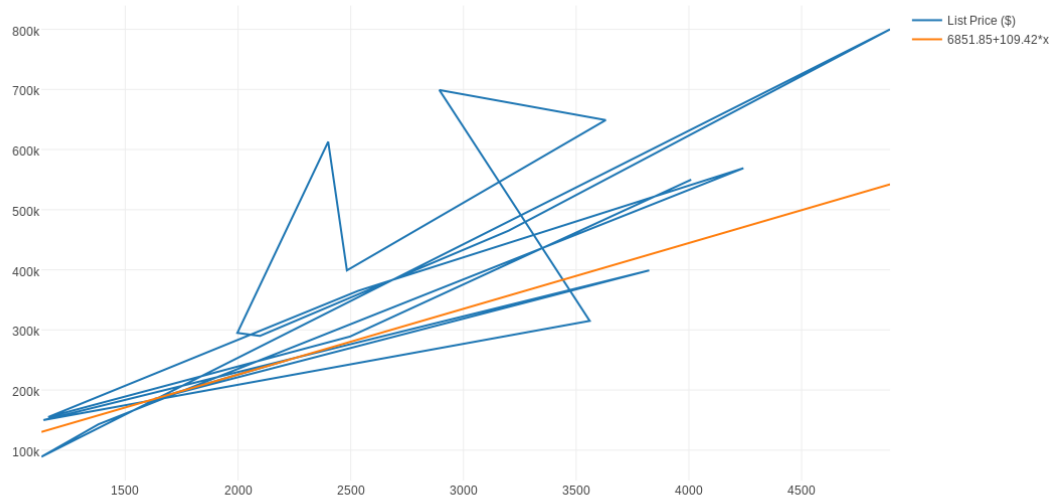
Now we can plan to plot square footage against our data and our formula values.

The screenshot shows a software interface for data analysis. On the left, a 'LINE PLOT' configuration window is open, with a dropdown menu set to 'Line plot'. Below it are 'OPTIONS' for 'Error Bars', 'Asymmetric Errors', and 'Group By'. The main area displays a data table with the following columns: Index, Sq Ft, Beds, Baths, Zip, Year, List Price (\$), and a formula  $6851.85 + 109.42 * x$ . The 'Sq Ft' column is highlighted in blue, and the formula column is also highlighted in blue. The data table contains 15 rows of data.

|    | Index | Sq Ft | Beds | Baths | Zip   | Year | List Price (\$) | $6851.85 + 109.42 * x$ |
|----|-------|-------|------|-------|-------|------|-----------------|------------------------|
| 1  | 1     | 2222  | 3    | 3.5   | 32312 | 1981 | 250000          | 249983.09              |
| 2  | 2     | 1628  | 3    | 2     | 32308 | 2009 | 185000          | 184987.61000000002     |
| 3  | 3     | 3824  | 5    | 4     | 32312 | 1954 | 399000          | 425273.93              |
| 4  | 4     | 1137  | 3    | 2     | 32309 | 1993 | 150000          | 131262.39              |
| 5  | 5     | 3560  | 6    | 4     | 32309 | 1973 | 315000          | 396387.05              |
| 6  | 6     | 2893  | 4    | 3     | 32312 | 1994 | 699000          | 323403.91              |
| 7  | 7     | 3631  | 4    | 3     | 32309 | 1996 | 649000          | 404155.87              |
| 8  | 8     | 2483  | 4    | 3     | 32312 | 2016 | 399000          | 278541.70999999996     |
| 9  | 9     | 2400  | 4    | 4     | 32312 | 2002 | 613000          | 269459.85              |
| 10 | 10    | 1997  | 3    | 3     | 32311 | 2006 | 295000          | 225363.59              |
| 11 | 11    | 2097  | 4    | 3     | 32311 | 2016 | 290000          | 236305.59              |
| 12 | 12    | 3200  | 5    | 4     | 32312 | 1964 | 465000          | 356995.85              |
| 13 | 13    | 4892  | 5    | 6     | 32311 | 2005 | 799900          | 542134.49              |
| 14 | 14    | 1128  | 2    | 1     | 32303 | 1955 | 89000           | 130277.61              |
| 15 | 15    | 1381  | 3    | 2     | 32301 | 2006 | 143000          | 157960.87              |



Choosing **Line Plot**, Plotly will make a mess of our data, but the line shows up just fine.



We can clean things up.

- Use **Traces/Mode/Lines/Markers** to change the marker for the list price data from a line to a dot;
- Use **Traces/Style** to change the size of dots from 6 to 16;
- Use **Axes** to shift the  $x$ -axis and  $y$ -axis to start at 0;

? Price = 6851.85 + 109.42 Square Footage ?

