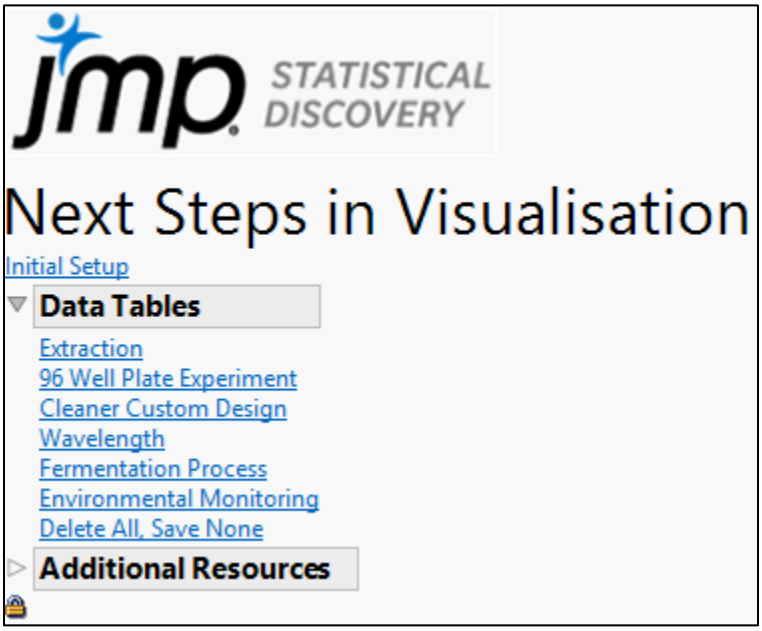
Visualising Data for Scientific Discovery: Next Steps in Visualisation

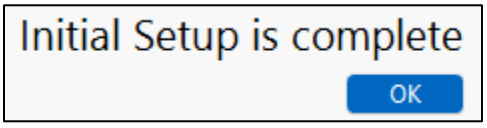
The datasets used in this workshop are all contained within a JMP journal file for your convenience.

1. Open **Data Visualisation.jrn**.



   A JMP journal is a convenient way to organize presentation materials or project data and analysis.

2. Select **Initial Setup**.



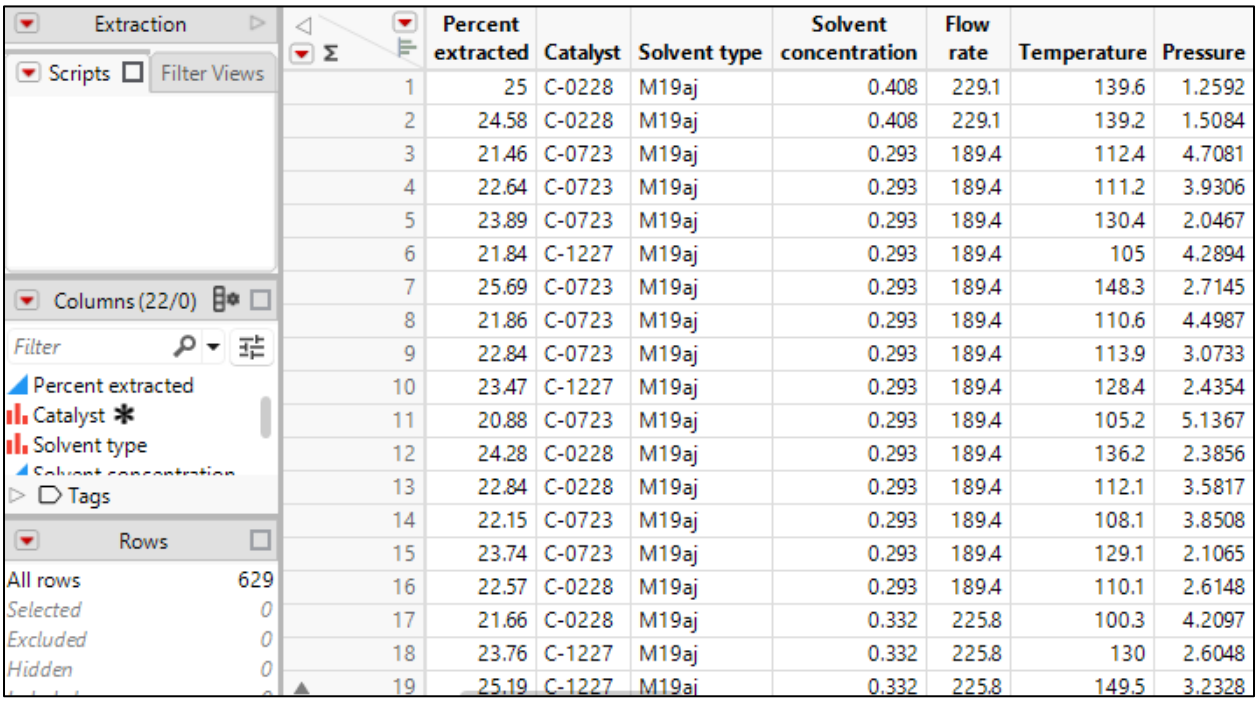
   The initial setup procedure writes some files in order to making graphing easier in the workshop.

Graphs Useful in Modeling Data

We have historical data on a chemical extraction process. The goal of our study is to find settings which maximize **Percent extracted**, the percent of desired substance extracted from a substrate. First, we'll find the important predictors, then search for correlations and ways to model our process and find the best model.

Note: we're demonstrating just one method for analysis here. There are many statistical models you could (and should!) fit to the data while trying to achieve the goal.

1. Select **Extraction**.

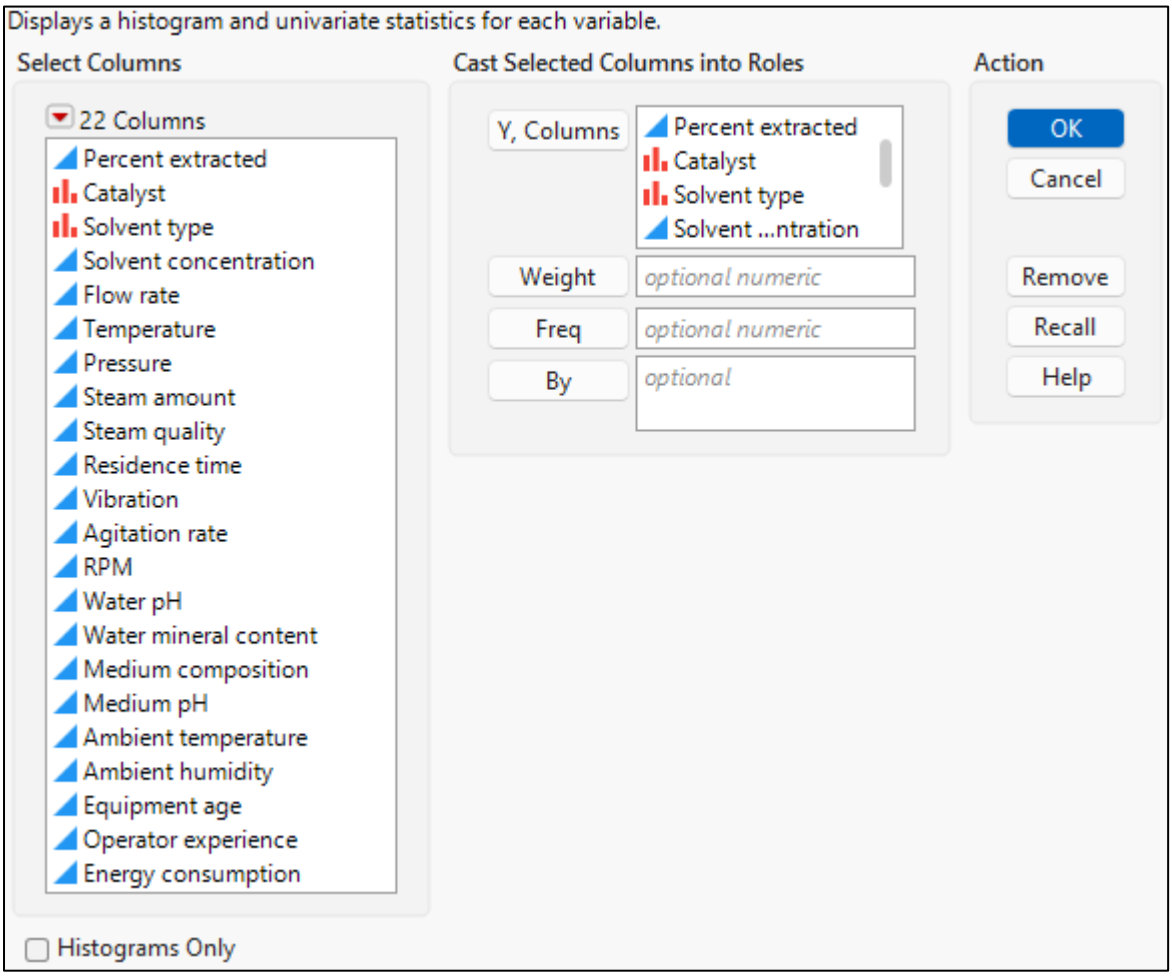


	Percent extracted	Catalyst	Solvent type	Solvent concentration	Flow rate	Temperature	Pressure
1	25	C-0228	M19aj	0.408	229.1	139.6	1.2592
2	24.58	C-0228	M19aj	0.408	229.1	139.2	1.5084
3	21.46	C-0723	M19aj	0.293	189.4	112.4	4.7081
4	22.64	C-0723	M19aj	0.293	189.4	111.2	3.9306
5	23.89	C-0723	M19aj	0.293	189.4	130.4	2.0467
6	21.84	C-1227	M19aj	0.293	189.4	105	4.2894
7	25.69	C-0723	M19aj	0.293	189.4	148.3	2.7145
8	21.86	C-0723	M19aj	0.293	189.4	110.6	4.4987
9	22.84	C-0723	M19aj	0.293	189.4	113.9	3.0733
10	23.47	C-1227	M19aj	0.293	189.4	128.4	2.4354
11	20.88	C-0723	M19aj	0.293	189.4	105.2	5.1367
12	24.28	C-0228	M19aj	0.293	189.4	136.2	2.3856
13	22.84	C-0228	M19aj	0.293	189.4	112.1	3.5817
14	22.15	C-0723	M19aj	0.293	189.4	108.1	3.8508
15	23.74	C-0723	M19aj	0.293	189.4	129.1	2.1065
16	22.57	C-0228	M19aj	0.293	189.4	110.1	2.6148
17	21.66	C-0228	M19aj	0.332	225.8	100.3	4.2097
18	23.76	C-1227	M19aj	0.332	225.8	130	2.6048
19	25.19	C-1227	M19aj	0.332	225.8	149.5	3.2328

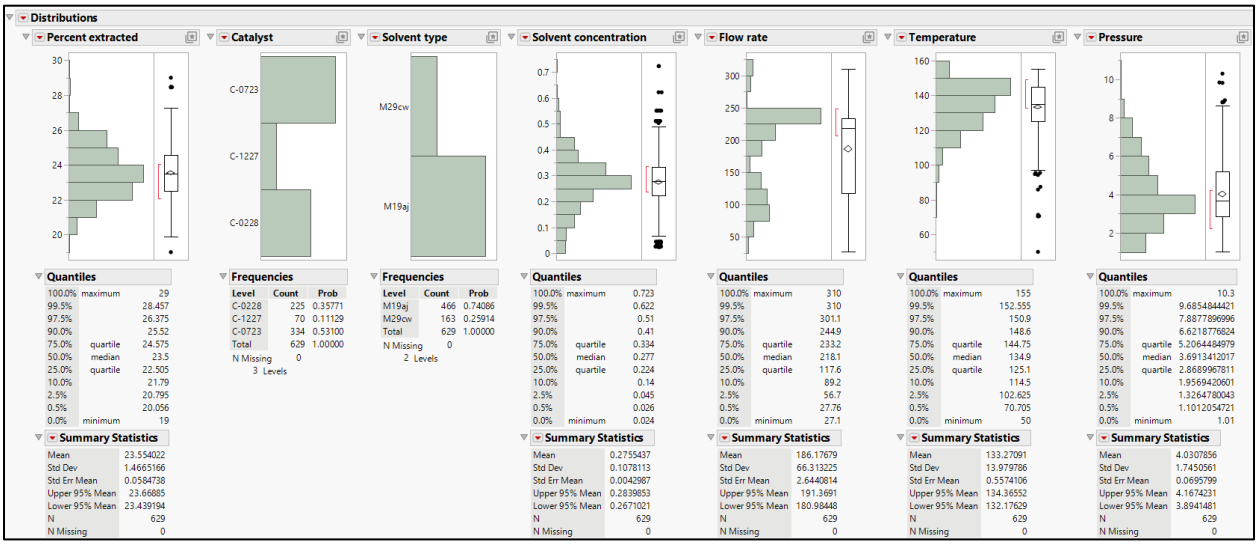
   There are 22 columns (variables) and 629 rows (observations) in the data table. The first column, **Percent extracted**, is the response variable; the others are predictor variables. Examine the distributions of the variables.

2. Select **Analyze** > **Distribution**.

3. Select **Percent extracted** through **Energy Consumption**, then click **Y, Columns**.



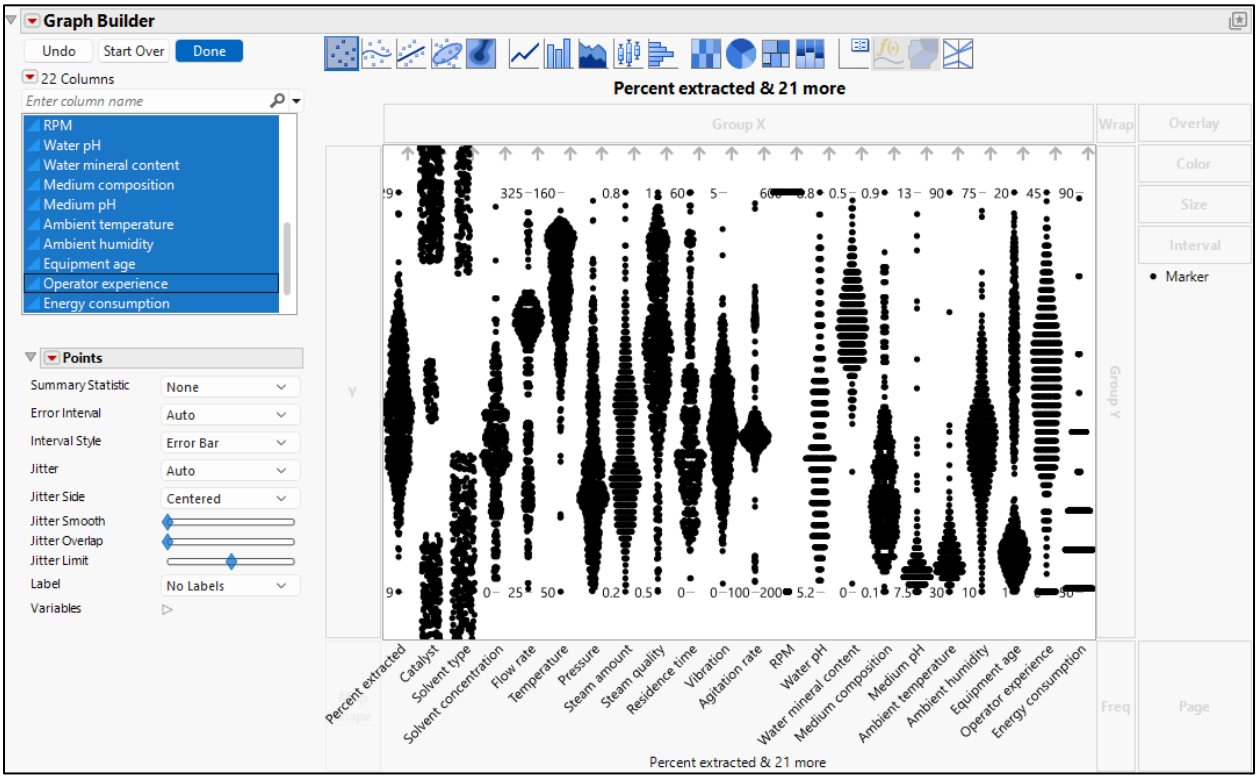
4. Click **OK**.



The data have already been cleaned. There are no outliers (points that don't belong to the process we are studying), all variables have the correct modeling type (nominal or continuous), and no variables need a transformation.

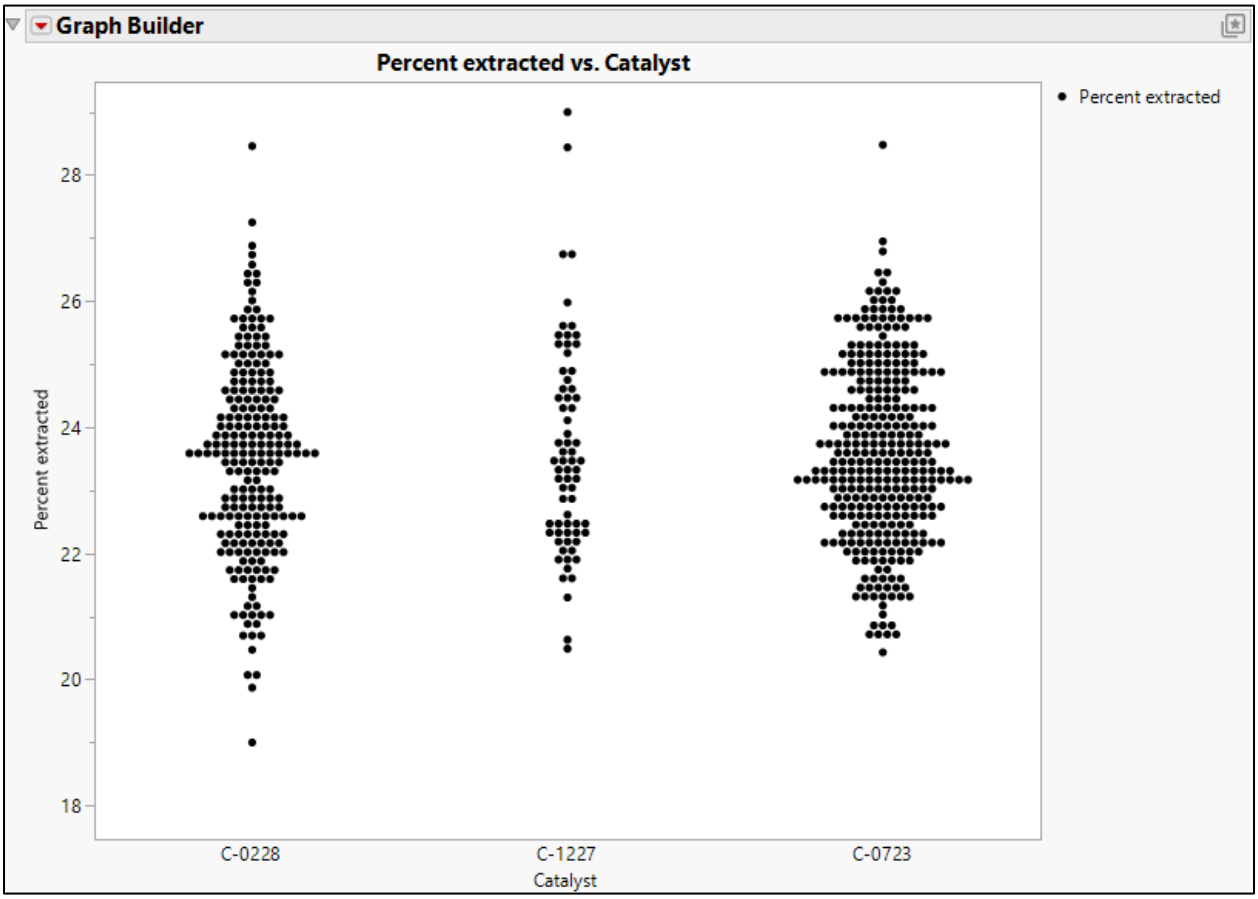
The Distribution platform is interactive – selecting bars in one histogram will select the corresponding rows in the data table and highlight those observations on other open graphs.

5. Select **Graph** > **Graph Builder**.
6. Select all variables and drag them to the X drop zone.



7. Drag **Percent extracted** to the Y drop zone.
8. Drag **Catalyst** to the X drop zone.

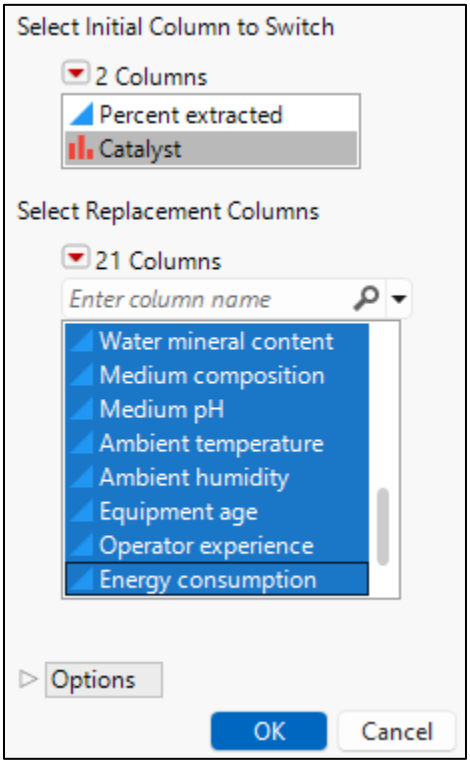
9. Click **Done**.



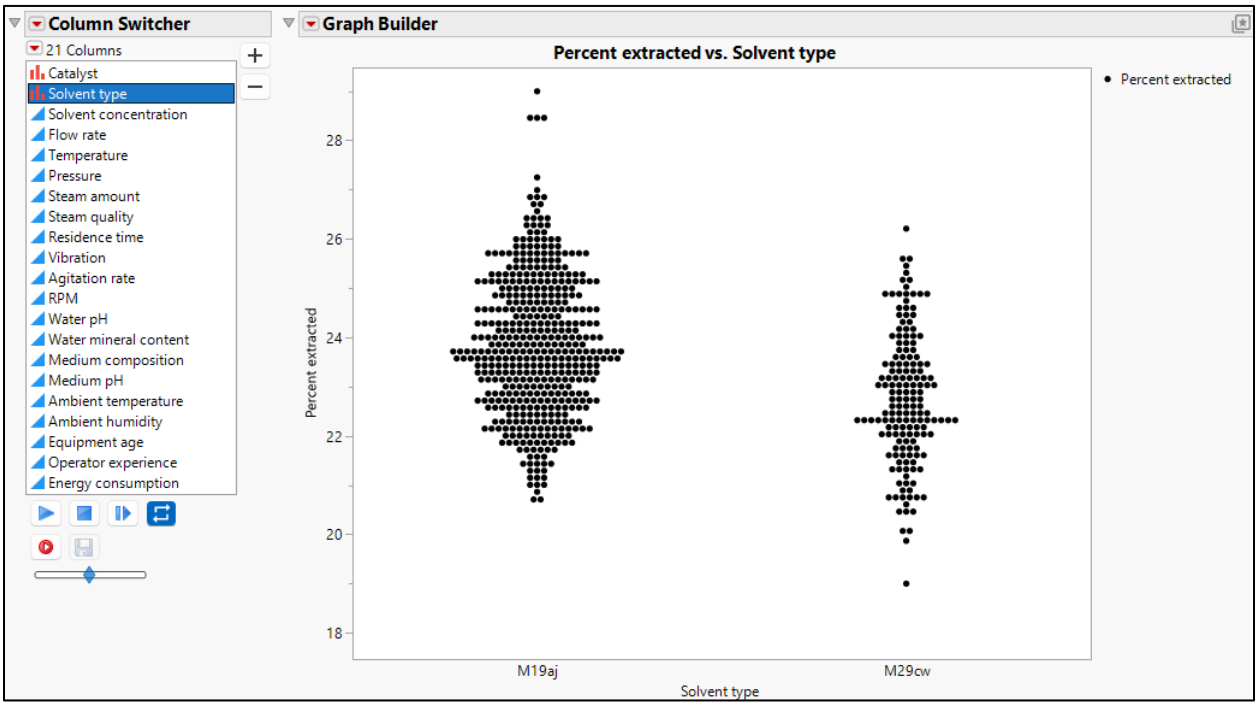
There is not a strong relationship between the variables.

10. Click the red triangle next to **Graph Builder** and select **Redo** > **Column Switcher**.
11. In the **Select Initial Column to Switch** box, select **Catalyst**.

12. In the **Select Replacement Columns** box, select **Catalyst** through **Energy Consumption**.



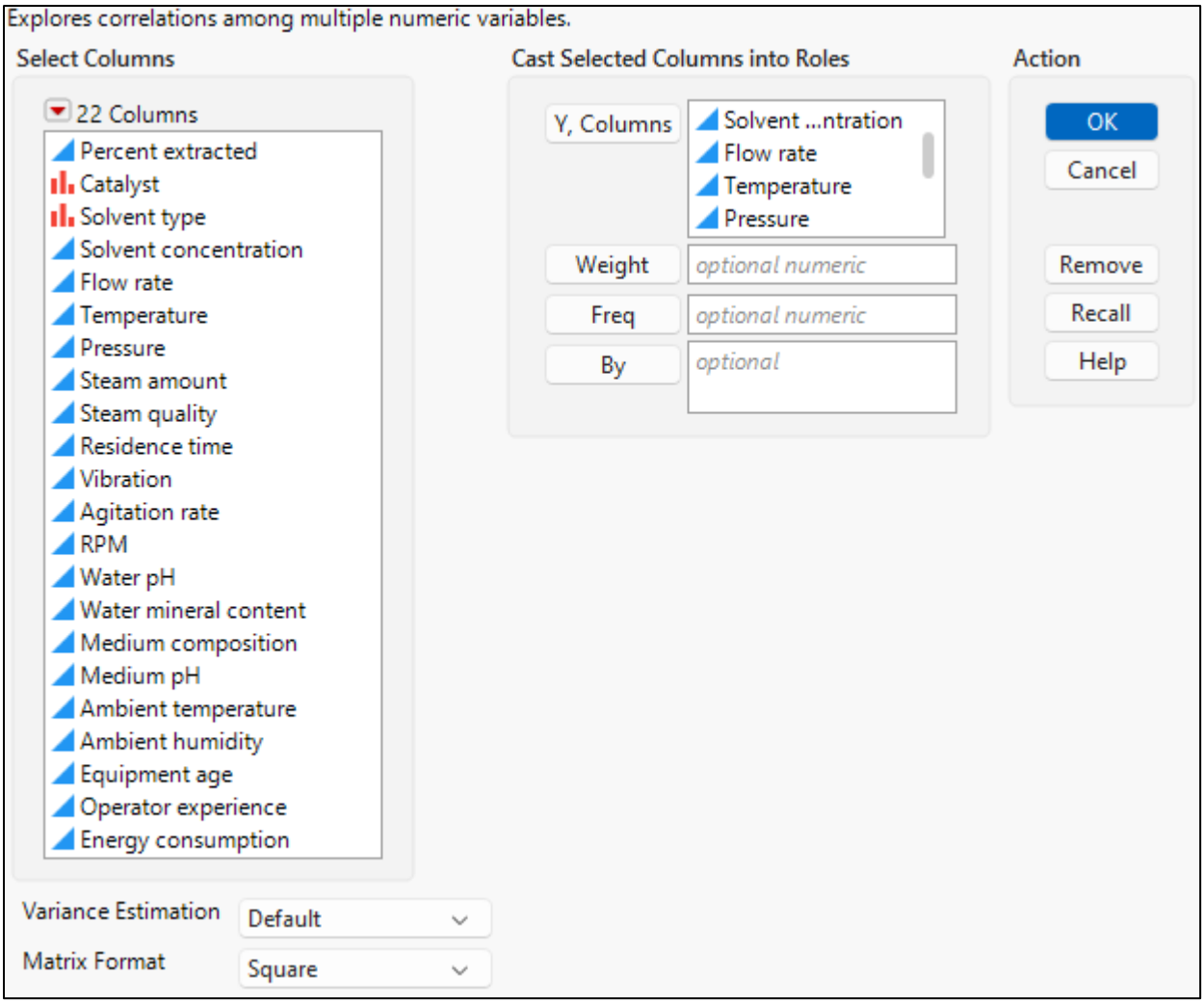
13. Click **OK**.



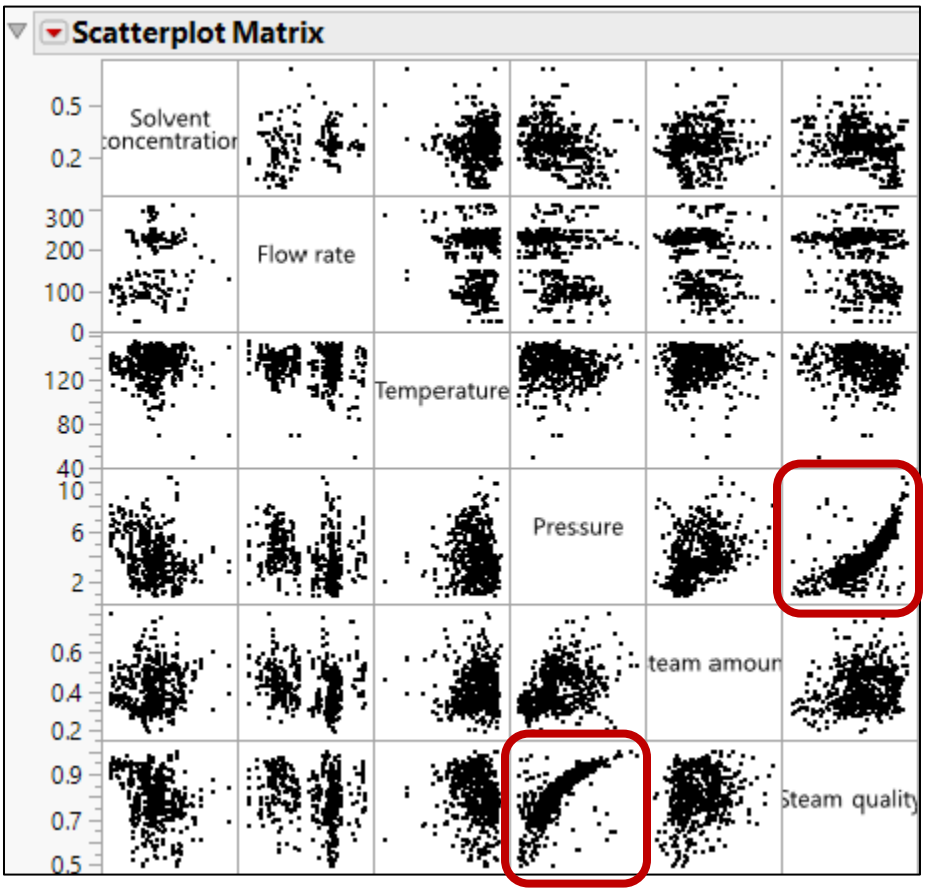
You can click through the columns on the left to visualize their main effect on **Percent extracted**.

It can also be useful to visualize the relationships among continuous predictors. Linear regression models are unstable when predictors are highly correlated.

14. Select **Analyze** > **Multivariate Methods** > **Multivariate**.
15. Select **Solvent Concentration** through **Energy Consumption**, then click **Y, Columns**.

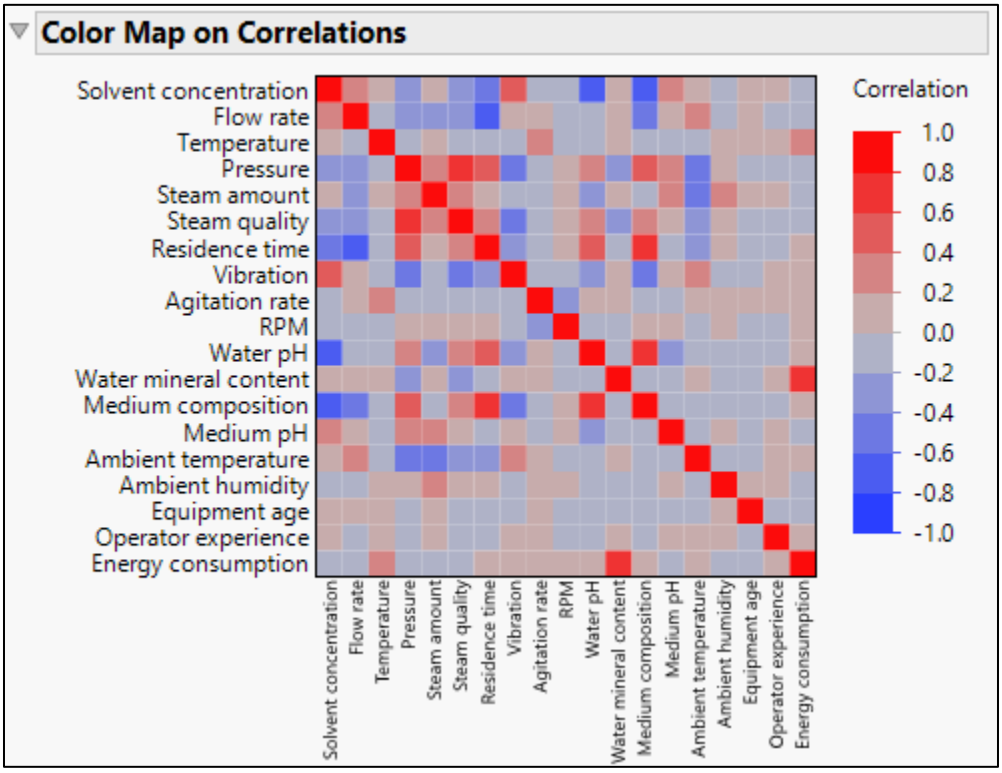


16. Click **OK**.



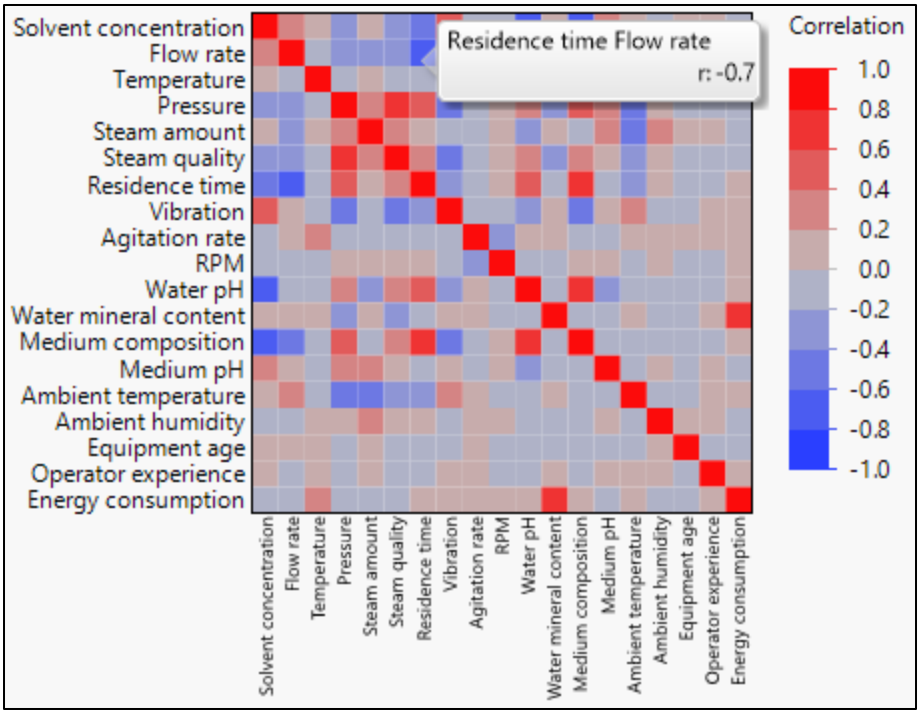
The screen capture of the entire scatterplot matrix is too large to fit here, but there are some interesting relationships between predictors, for example **Pressure** and **Steam quality** highlighted in the screen capture above.

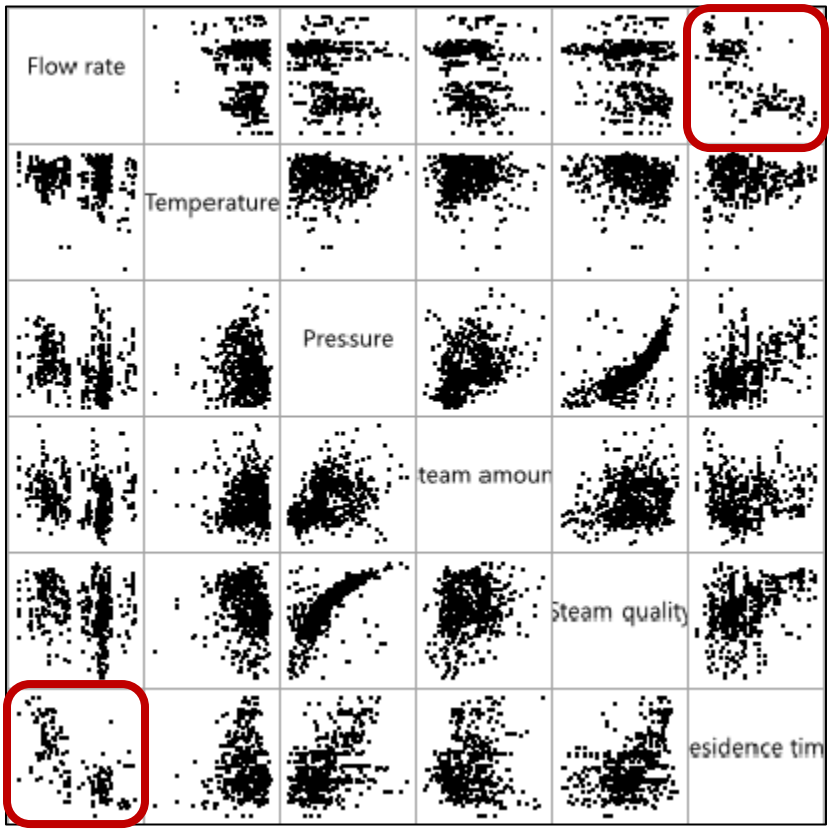
17. Click the red triangle next to **Multivariate** and select **Color Maps** > **Color Maps on Correlations**.



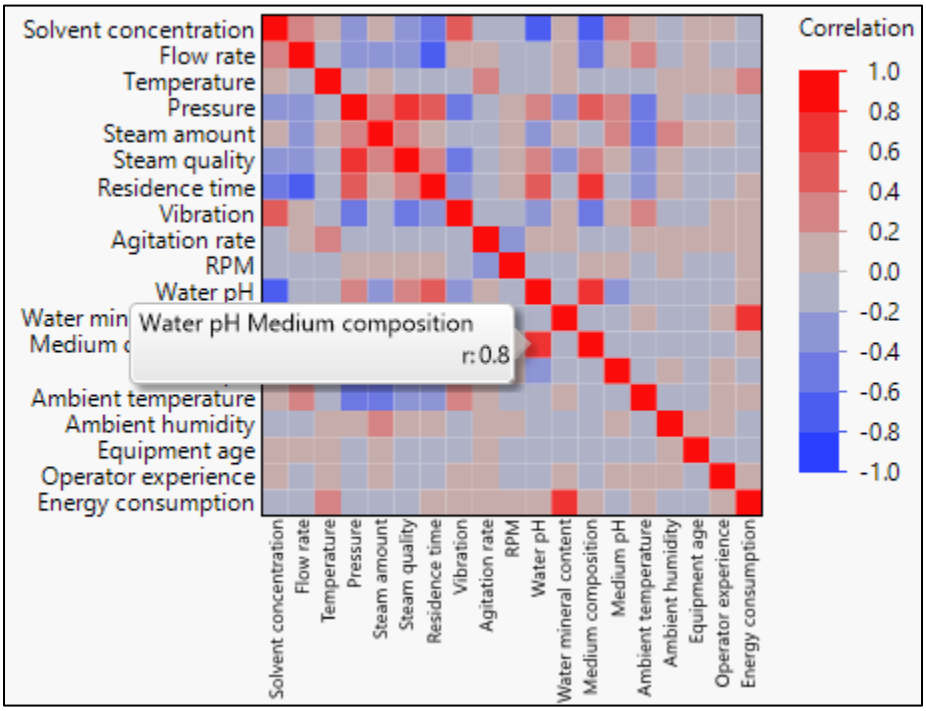
The correlation matrix of numbers that represent the strength of the linear relationship between two predicators has been simplified to a grid of colors. You can change the color theme by right-clicking the legend and selecting **Color Theme**.

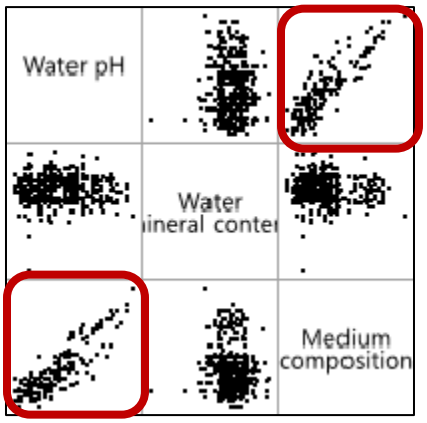
Correlations near +1 are denoted with red, correlations near –1 are denoted with blue, correlations near 0 are denoted with grey. The main diagonal of this grid is read because each predictor is perfectly correlated with itself. Compare some correlations with their graphs. You can hover over the colored squares to see the variables and correlation.



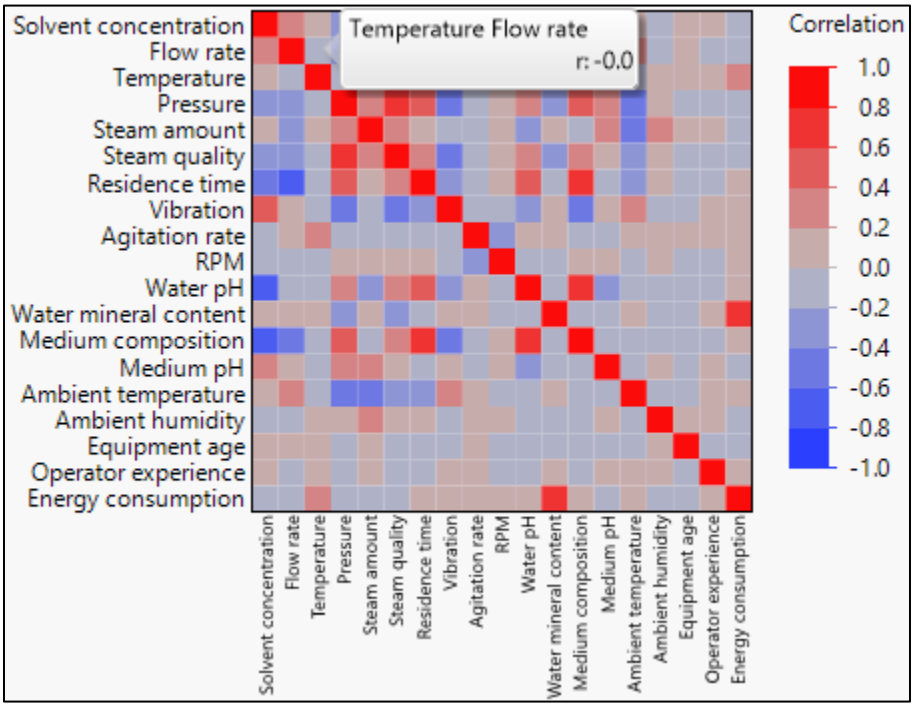


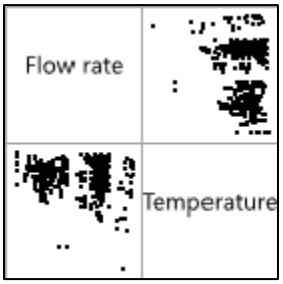
Examine a positive correlation.





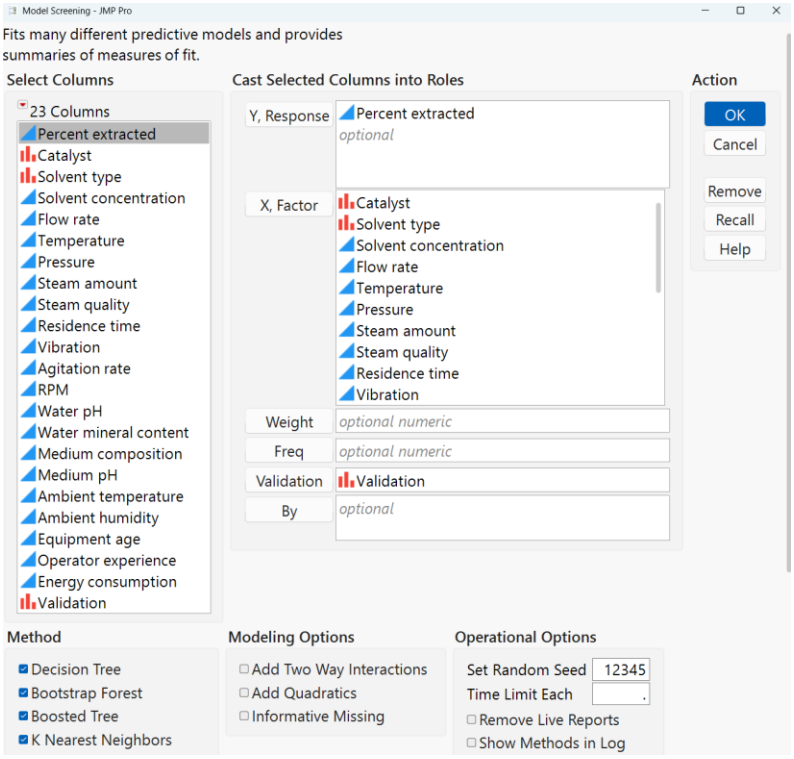
Examine a correlation near zero.



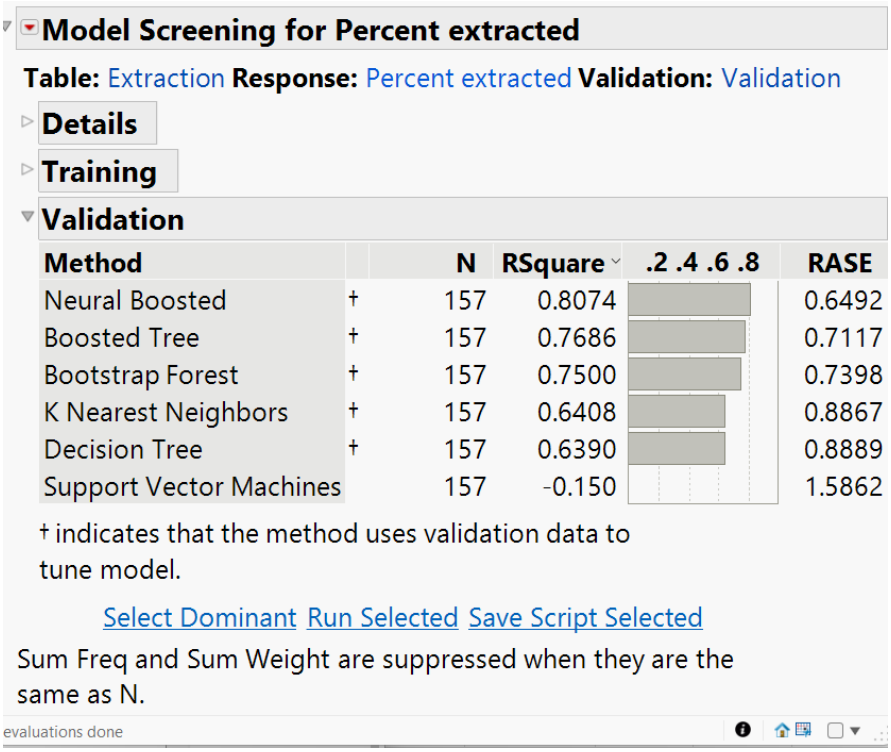


18. Linear regression models are unstable when predictors are highly correlated. So fitting a simple regression model will not describe the process properly. An advanced modeling capability is required. On your own, you can fit a variety of models for **Percent extracted**, including linear models, penalized regression (Generalized Regression) models, decision tree (Partition) models, and others. How do you know which model is the best? See the Model Screening platform in JMP Pro to fit and evaluate many models at once.
19. Select **Analyze** > **Predictive Modeling**> **Model Screening**.
20. Select **Percent extracted**, then click **Y, Response**.
21. Select **Catalyst** through **Energy Consumption**, then click **X**.
22. Select **Validation** then click **Validation**

23. For **Set Random Seed**, enter 12345.



24. Here, we can see that the best performing model is the neural boosted.



We can select the Dominant model or dig deeper by exploring the models under the **Details** menu.

Visualising a well plate experiment using map shapes

An experiment was performed to understand the relationship between temperature, pH, and amounts of an acid and two additives on the relative potency of a drug substance. Six 96 well plates were used in the experiment. Each plate was processed at one of three temperatures. pH and acid amount were randomized over the wells. One additive was randomized over the rows, the other was randomized over the columns.

You can use Graph Builder to visualise the factors and responses over the plates.

1. Return to the workshop journal and select **96 Well Plate Experiment**.

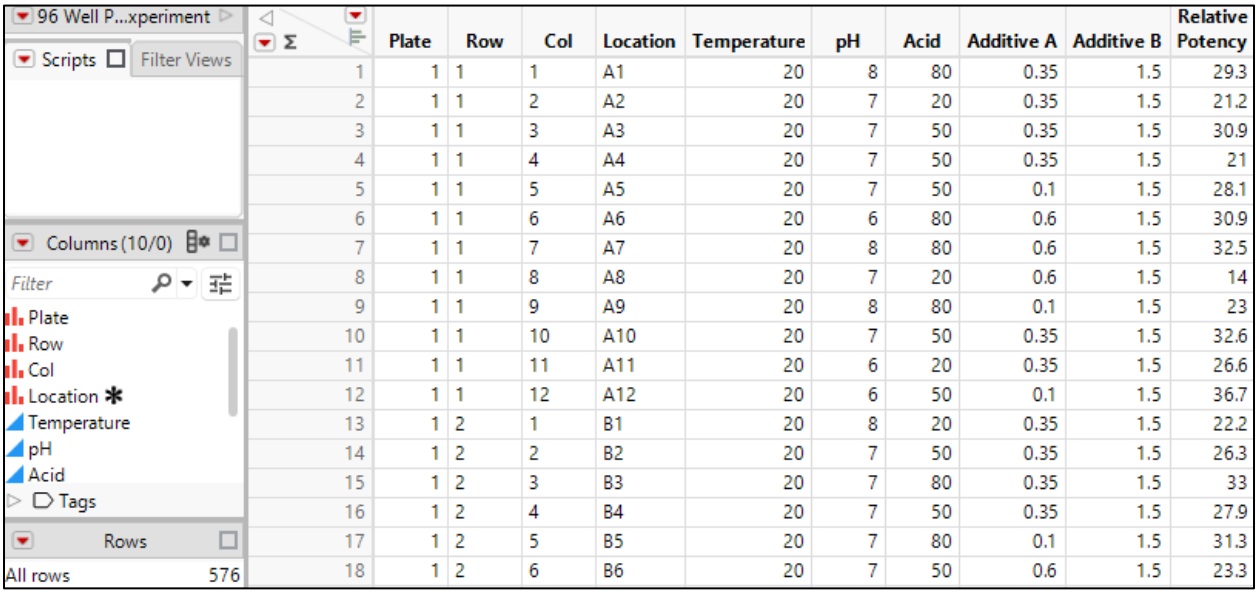
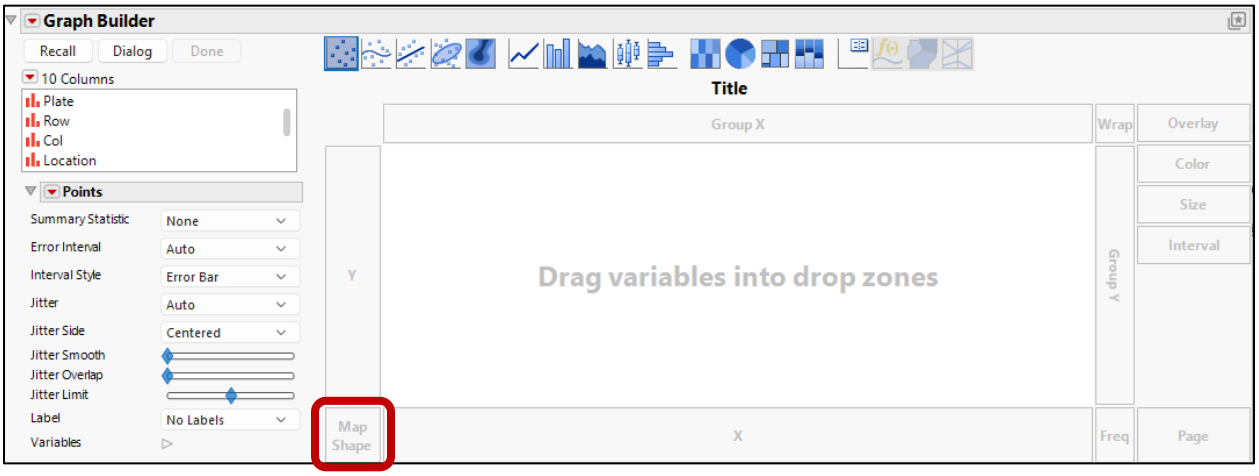
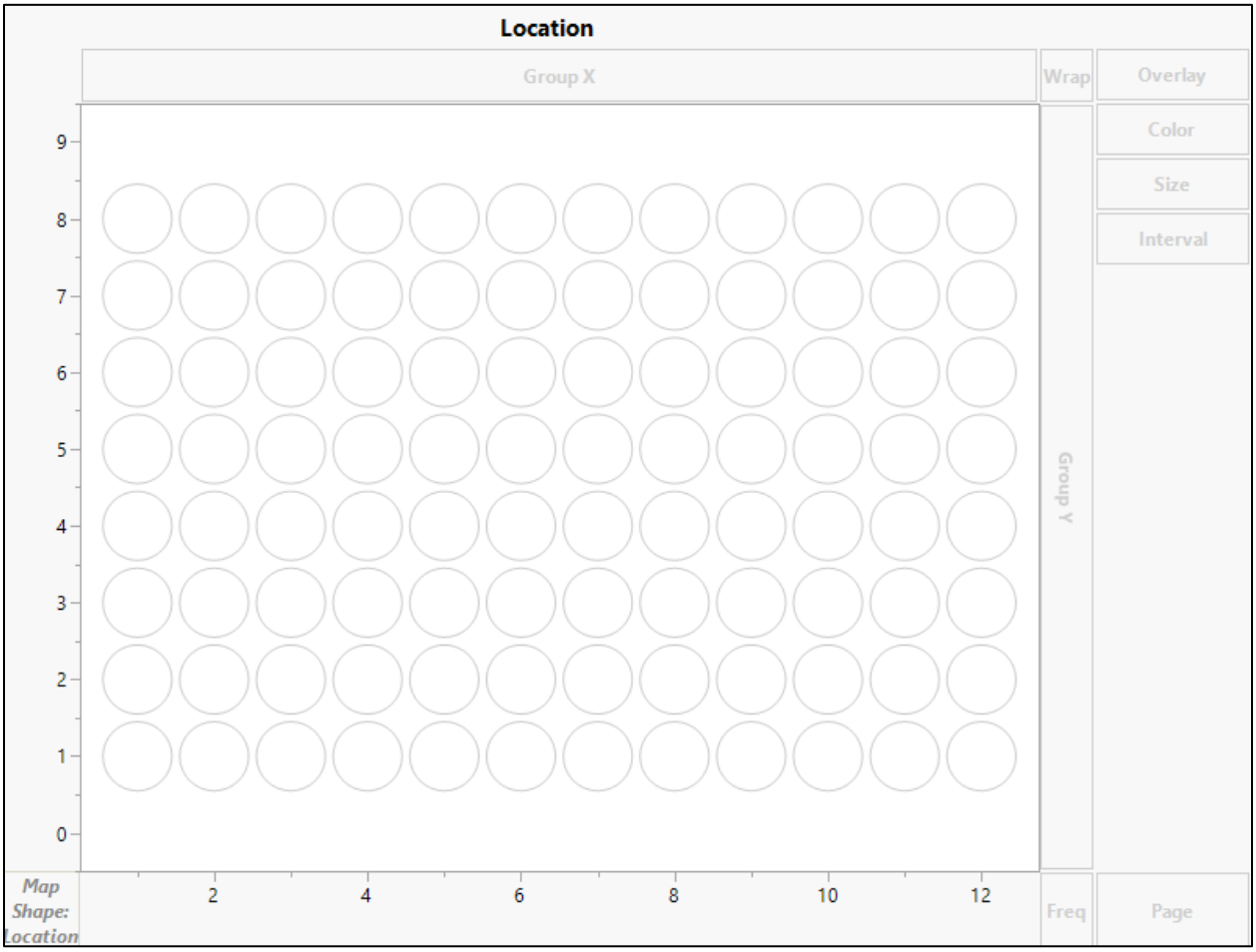


	Plate	Row	Col	Location	Temperature	pH	Acid	Additive A	Additive B	Relative Potency
1	1	1	1	A1	20	8	80	0.35	1.5	29.3
2	1	1	2	A2	20	7	20	0.35	1.5	21.2
3	1	1	3	A3	20	7	50	0.35	1.5	30.9
4	1	1	4	A4	20	7	50	0.35	1.5	21
5	1	1	5	A5	20	7	50	0.1	1.5	28.1
6	1	1	6	A6	20	6	80	0.6	1.5	30.9
7	1	1	7	A7	20	8	80	0.6	1.5	32.5
8	1	1	8	A8	20	7	20	0.6	1.5	14
9	1	1	9	A9	20	8	80	0.1	1.5	23
10	1	1	10	A10	20	7	50	0.35	1.5	32.6
11	1	1	11	A11	20	6	20	0.35	1.5	26.6
12	1	1	12	A12	20	6	50	0.1	1.5	36.7
13	1	2	1	B1	20	8	20	0.35	1.5	22.2
14	1	2	2	B2	20	7	50	0.35	1.5	26.3
15	1	2	3	B3	20	7	80	0.35	1.5	33
16	1	2	4	B4	20	7	50	0.35	1.5	27.9
17	1	2	5	B5	20	7	80	0.1	1.5	31.3
18	1	2	6	B6	20	7	50	0.6	1.5	23.3

2. Select **Graph** > **Graph Builder**.



3. Drag **Location** to the **Map Shape** zone in the bottom left of the graphing region, see the screen capture above for the location.

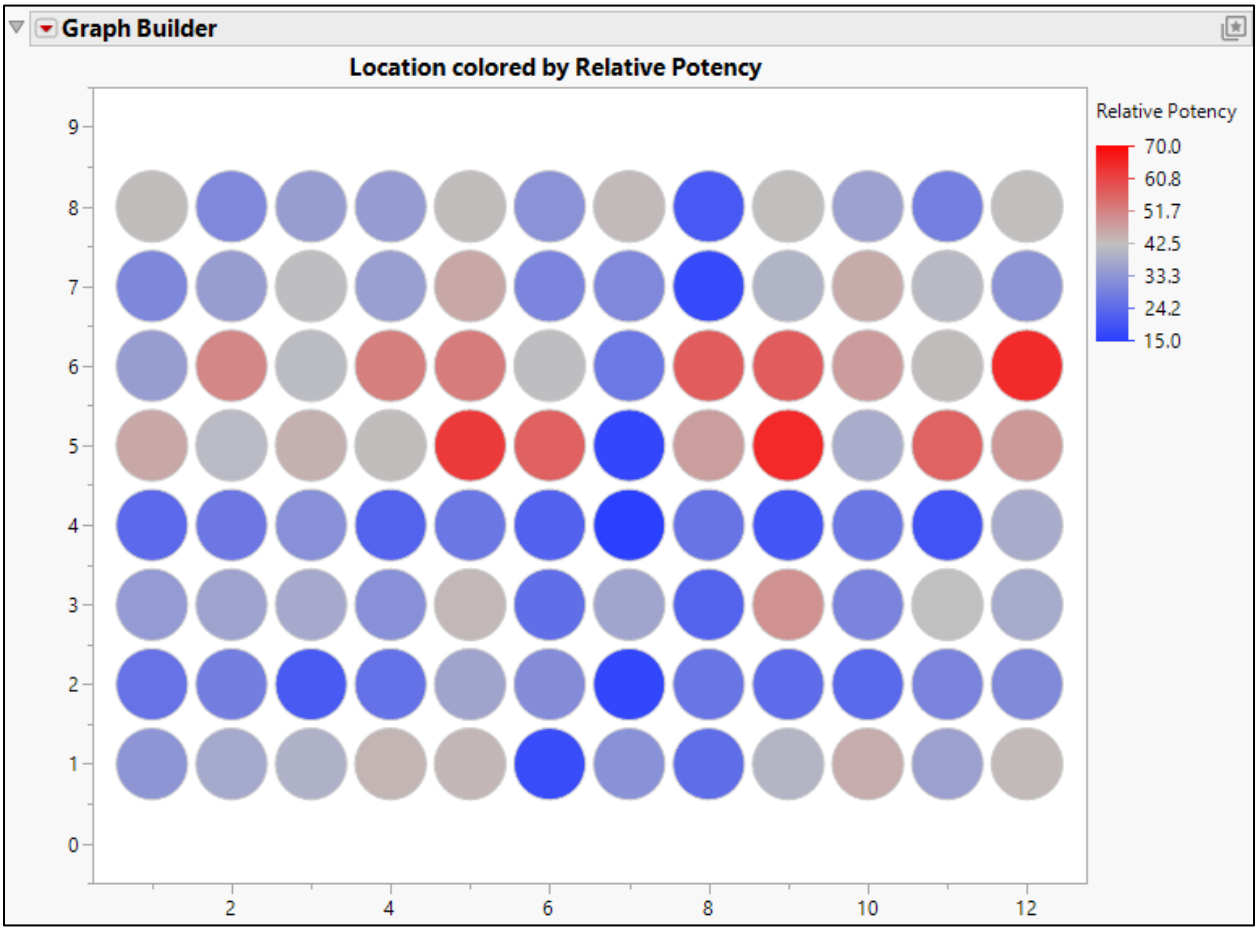


Why are the wells appearing in the graph? The **Location** column has the Map Role column property. That property is related to a shape name file which is related to a shape coordinates file. Those files were added to your workshop folder when you clicked **Initial Setup** at the beginning of these notes.

Country shape files are automatically included with JMP. You can create your own shape files to correspond to floor plans, the human body, or any other possibility. Information on graphing using maps and creating custom shape files can be found in the **Additional Resources** section of your workshop journal.

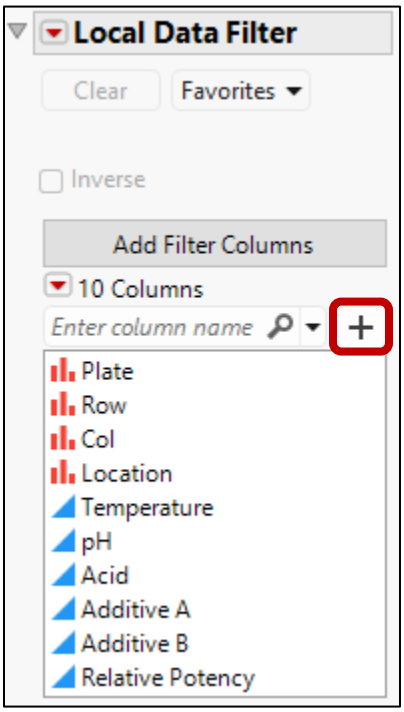
4. Drag **Relative Potency** to the Color drop zone.

5. Click **Done**.



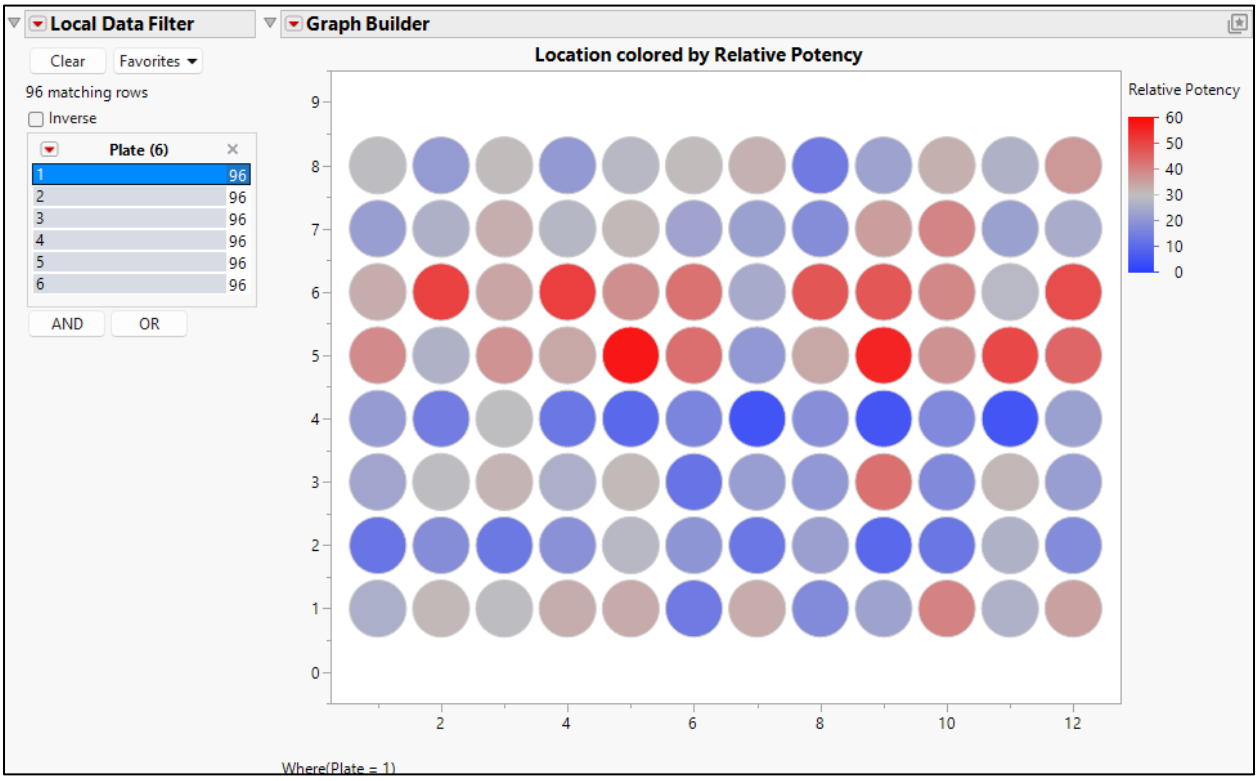
Each well has been colored by the average value of all response values for the well. You can view individual plates by adding a data filter.

6. Click the red triangle next to **Graph Builder** and select **Local Data Filter**.



7. In the **Add Filter Columns** box, select **Plate**, then click the plus sign icon, see the screen capture above for its location.

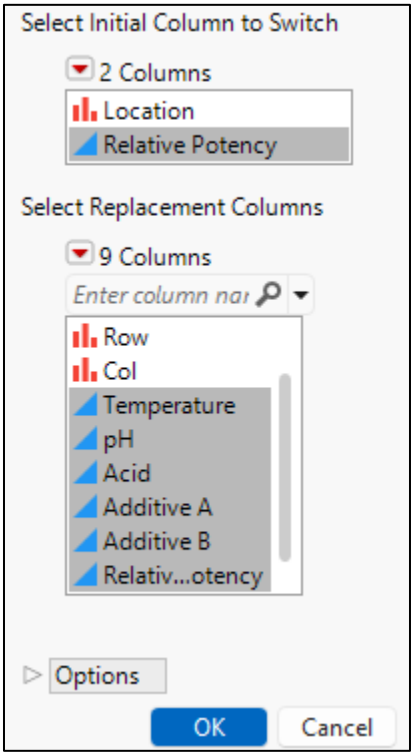
8. Select 1 to restrict the graph to the first plate.



Examine other variables in the Color zone by using the Column Switcher.

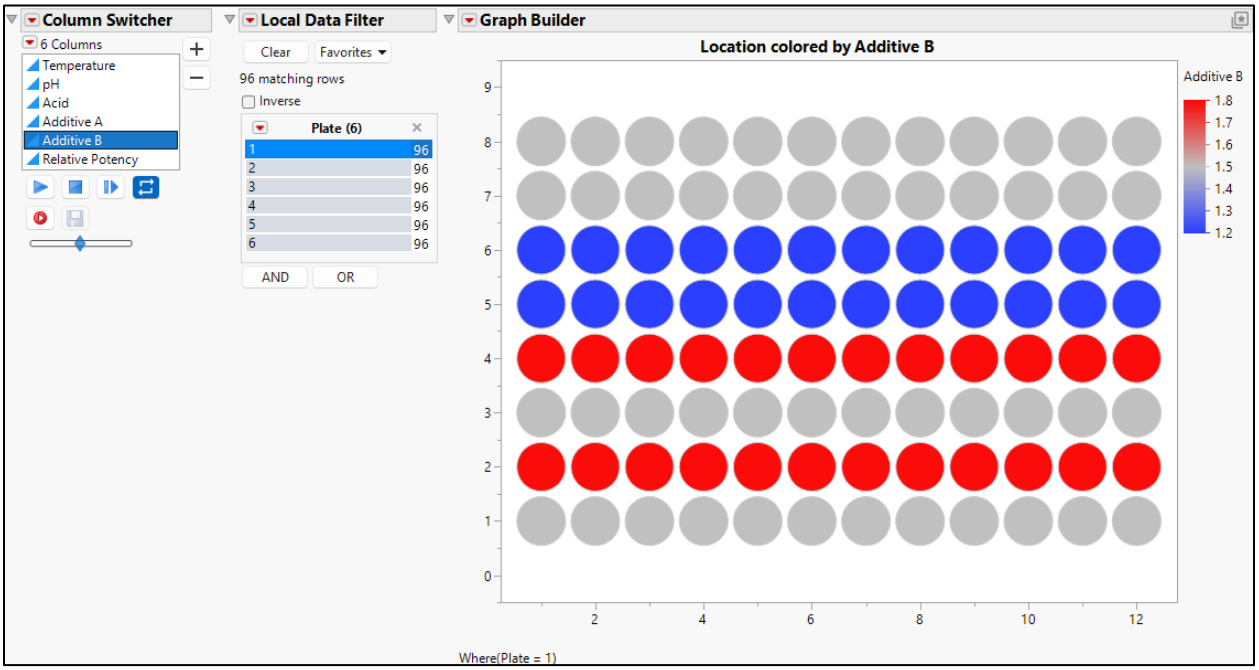
9. Click the red triangle next to **Graph Builder** and select **Redo** > **Column Switcher**.
10. In the **Select Initial Column to Switch** box, select **Relative Potency**.

11. In the **Select Replacement Columns** box, select **Temperature** through **Relative Potency**.



12. Click **OK**.

13. In the Columns Switcher, select **Additive B**.



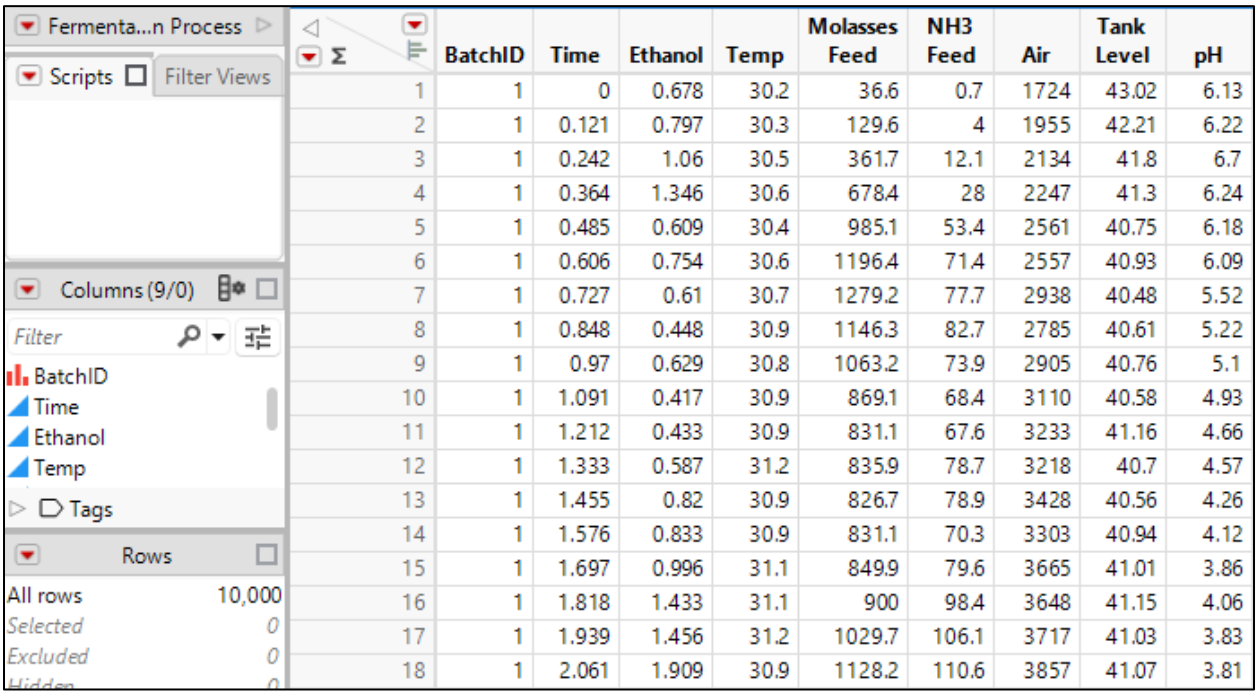
It is easy to visualise the experimental conditions and the response variable over those conditions with this graph.

14. Close the **96 Well Plate Experiment** data table and all associated report windows without saving.

Using multiple Y axes in Graph Builder

In a fermentation process, characteristics of the environment such as temperature and pH and characteristics of the product, such as amount of molasses and ammonium feeds and amount of ethanol produced, are monitored over time for each batch. Graph Builder makes it easy to build separate graphs each variable over time, or to build one graph with separate Y axes.

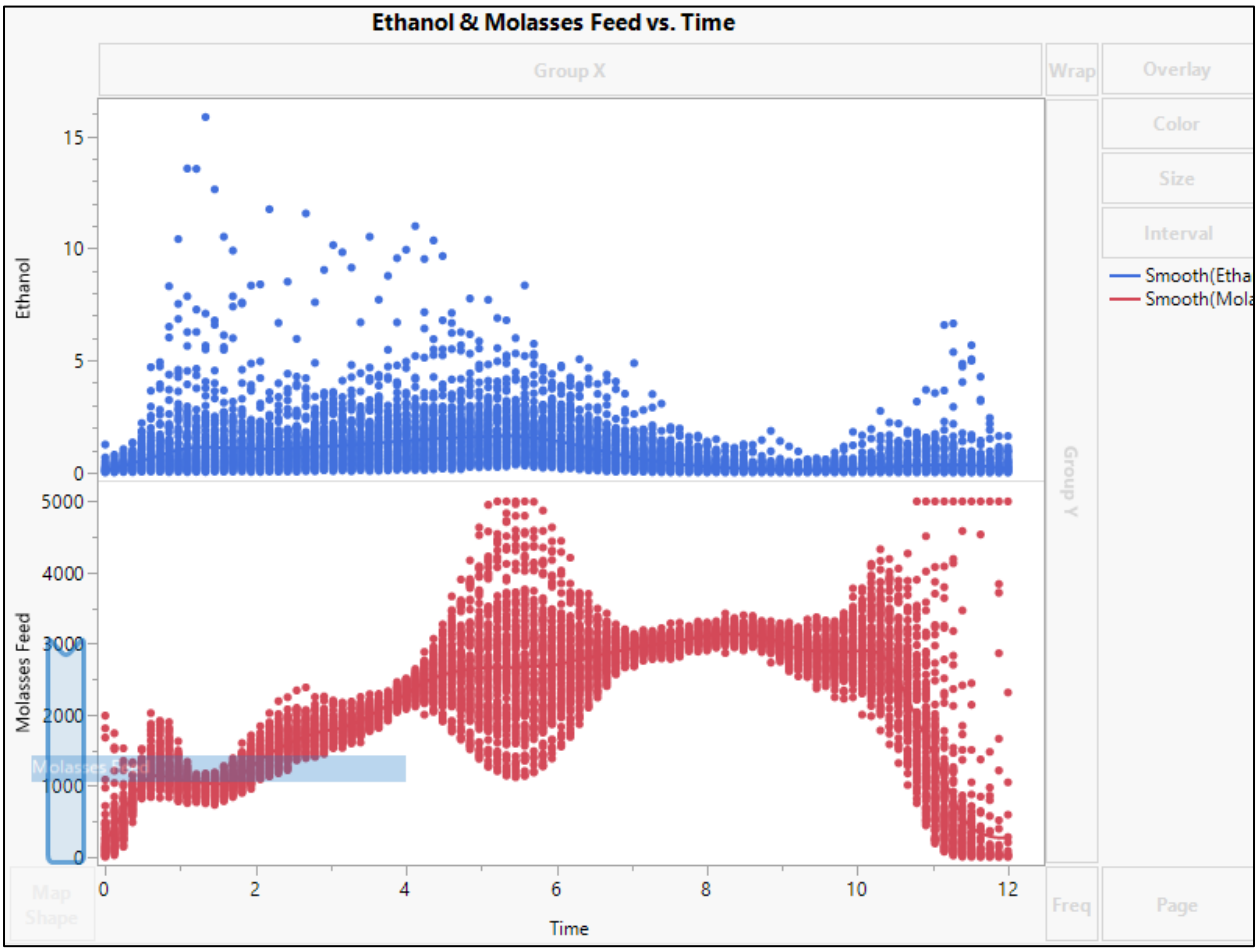
1. Return to the workshop journal and select **Fermentation Process**.



	BatchID	Time	Ethanol	Temp	Molasses Feed	NH3 Feed	Air	Tank Level	pH
1	1	0	0.678	30.2	36.6	0.7	1724	43.02	6.13
2	1	0.121	0.797	30.3	129.6	4	1955	42.21	6.22
3	1	0.242	1.06	30.5	361.7	12.1	2134	41.8	6.7
4	1	0.364	1.346	30.6	678.4	28	2247	41.3	6.24
5	1	0.485	0.609	30.4	985.1	53.4	2561	40.75	6.18
6	1	0.606	0.754	30.6	1196.4	71.4	2557	40.93	6.09
7	1	0.727	0.61	30.7	1279.2	77.7	2938	40.48	5.52
8	1	0.848	0.448	30.9	1146.3	82.7	2785	40.61	5.22
9	1	0.97	0.629	30.8	1063.2	73.9	2905	40.76	5.1
10	1	1.091	0.417	30.9	869.1	68.4	3110	40.58	4.93
11	1	1.212	0.433	30.9	831.1	67.6	3233	41.16	4.66
12	1	1.333	0.587	31.2	835.9	78.7	3218	40.7	4.57
13	1	1.455	0.82	30.9	826.7	78.9	3428	40.56	4.26
14	1	1.576	0.833	30.9	831.1	70.3	3303	40.94	4.12
15	1	1.697	0.996	31.1	849.9	79.6	3665	41.01	3.86
16	1	1.818	1.433	31.1	900	98.4	3648	41.15	4.06
17	1	1.939	1.456	31.2	1029.7	106.1	3717	41.03	3.83
18	1	2.061	1.909	30.9	1128.2	110.6	3857	41.07	3.81

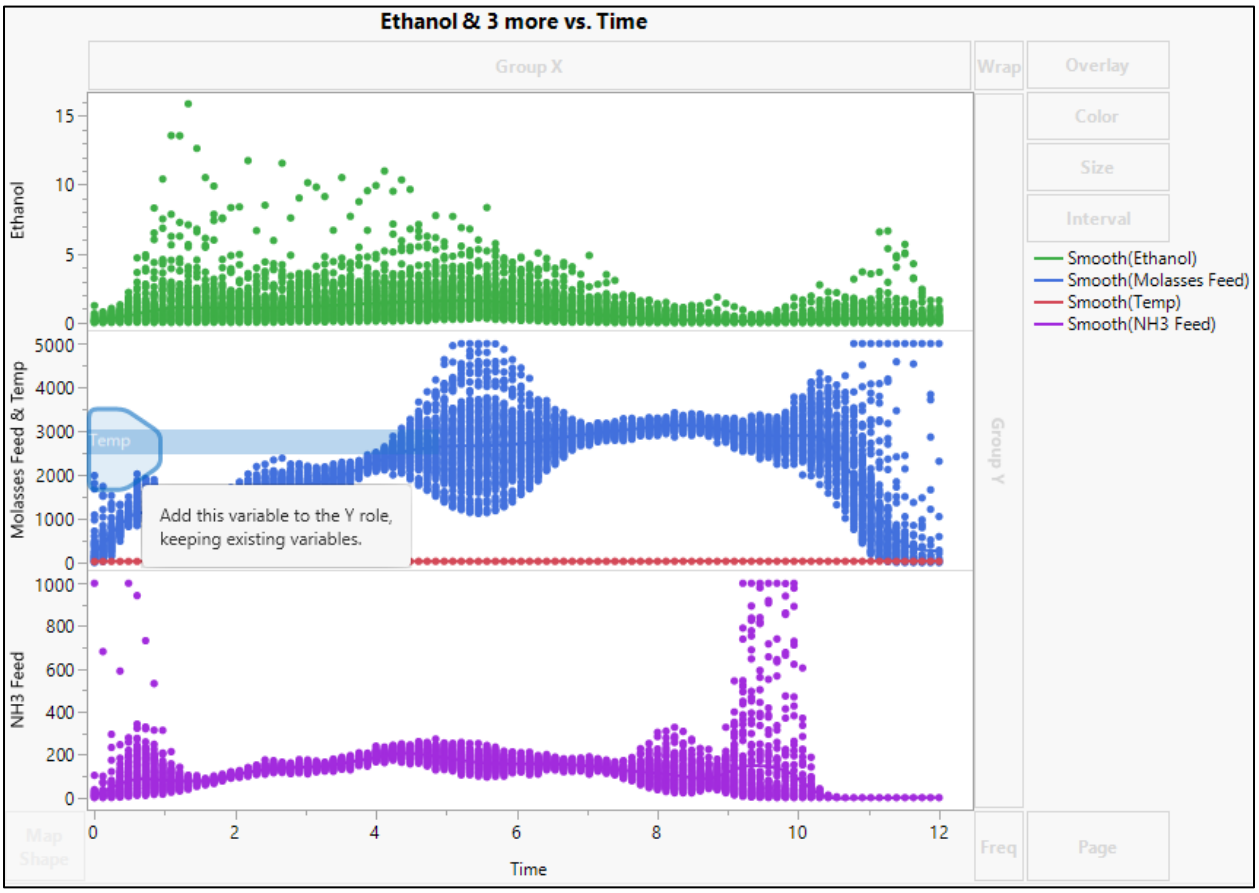
2. Select **Graph** > **Graph Builder**.
3. Drag **Time** to the X drop zone.
4. Drag **Ethanol** to the Y drop zone.

5. Drag **Molasses Feed** to the Y drop zone under **Ethanol**.

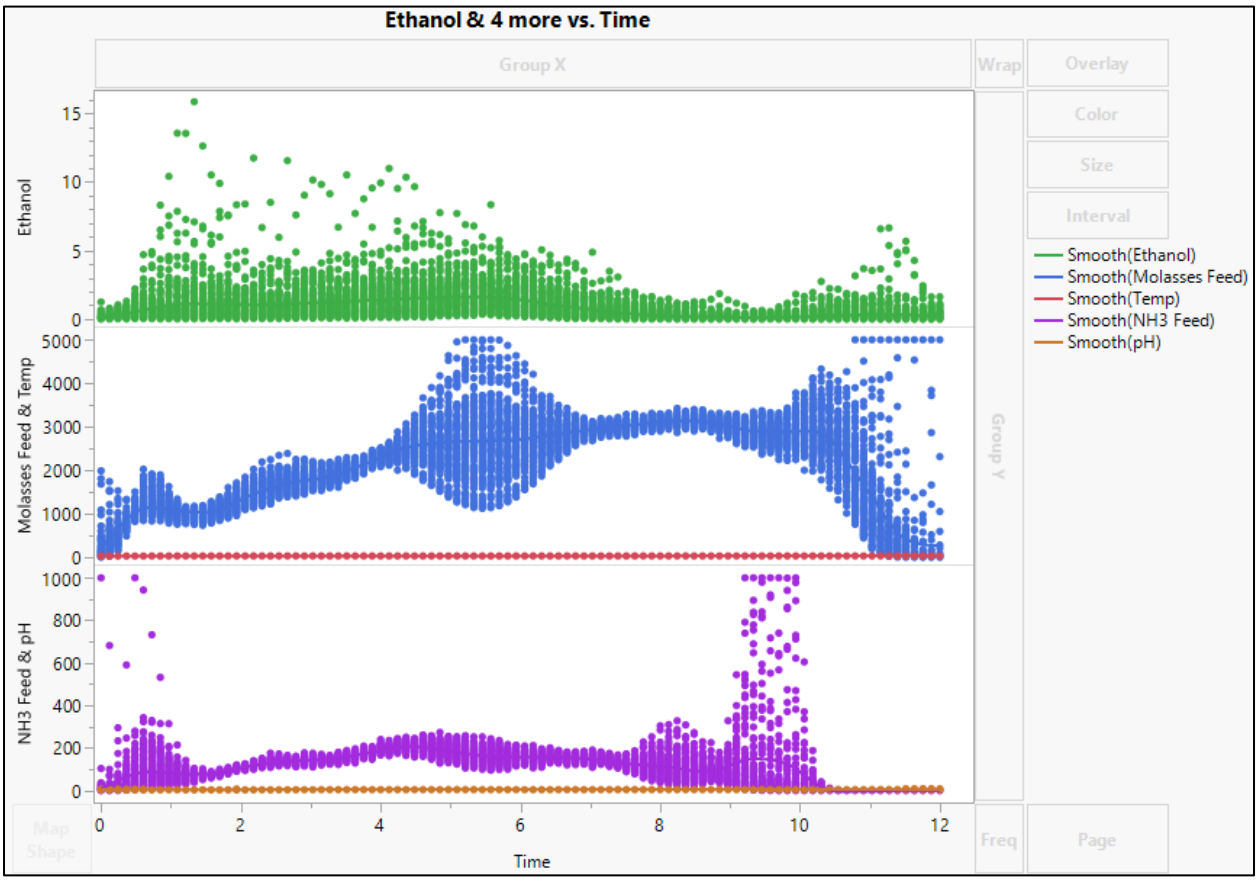


6. Drag **NH3 Feed** to the Y drop zone under **Molasses Feed**.

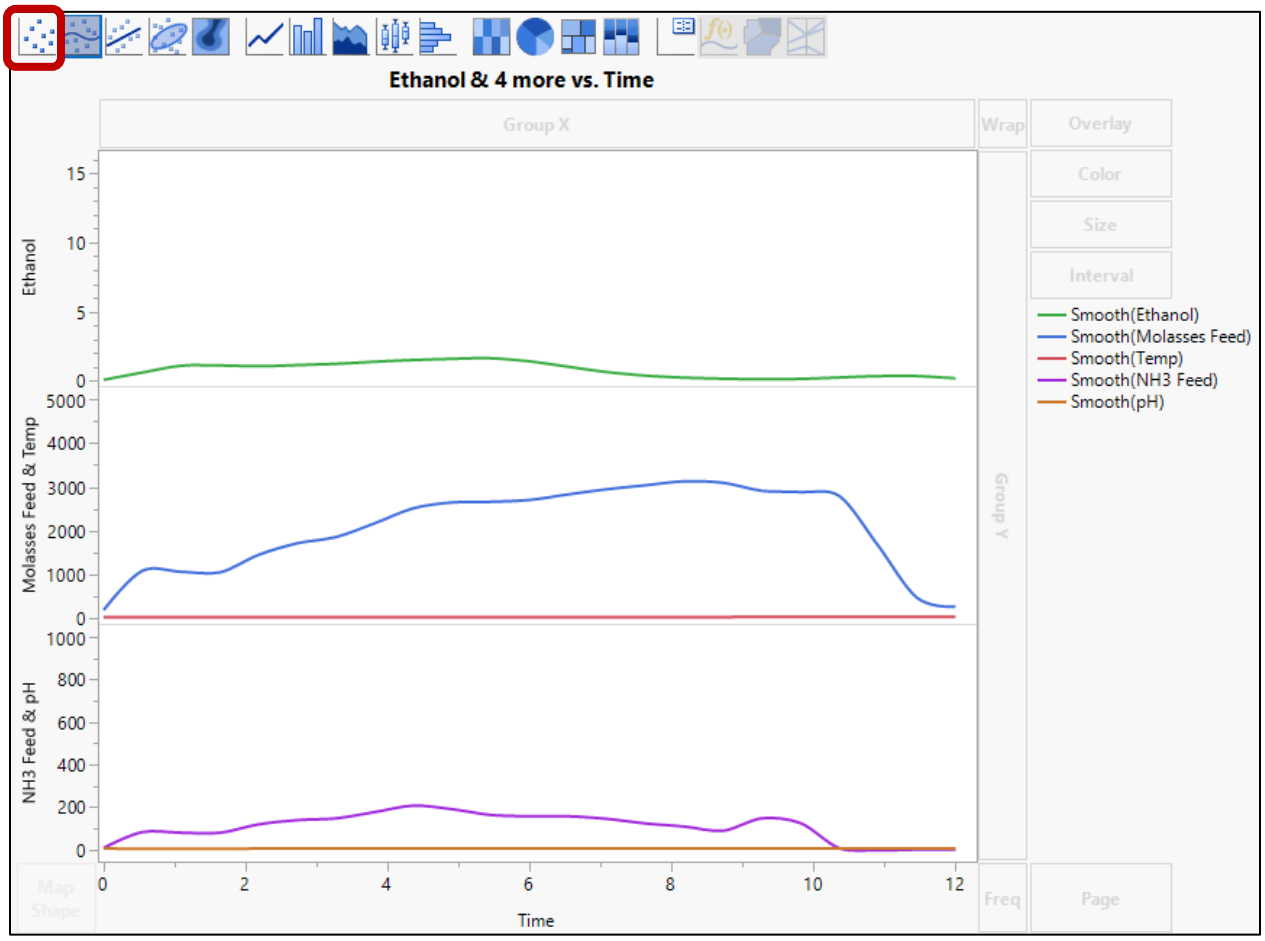
7. Drag **Temp** to the Y drop zone of the **Molasses Feed** panel, just inside the Y axis.



8. Drag **pH** to the Y drop zone of the **NH3 Feed** panel, just inside the Y axis.

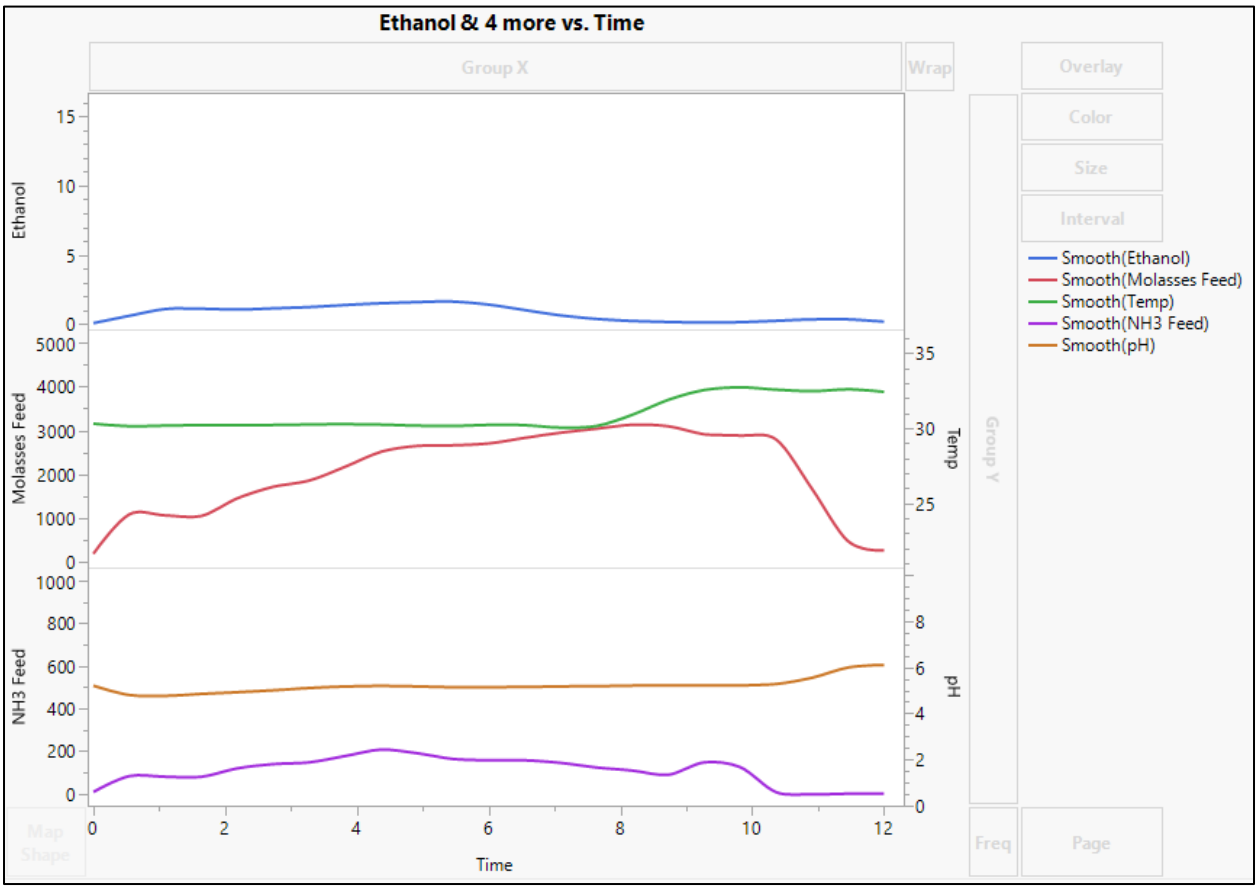


9. Click the Points icon in the graphing elements bar above the graph to remove the points element but keep the smoother element.



10. Right-click the **Molasses Feed & Temp** Y axis and select **Move Right** > **Temp**.

11. Right-click the **NH2 Feed & pH** Y axis and select **Move Right** > **pH**.



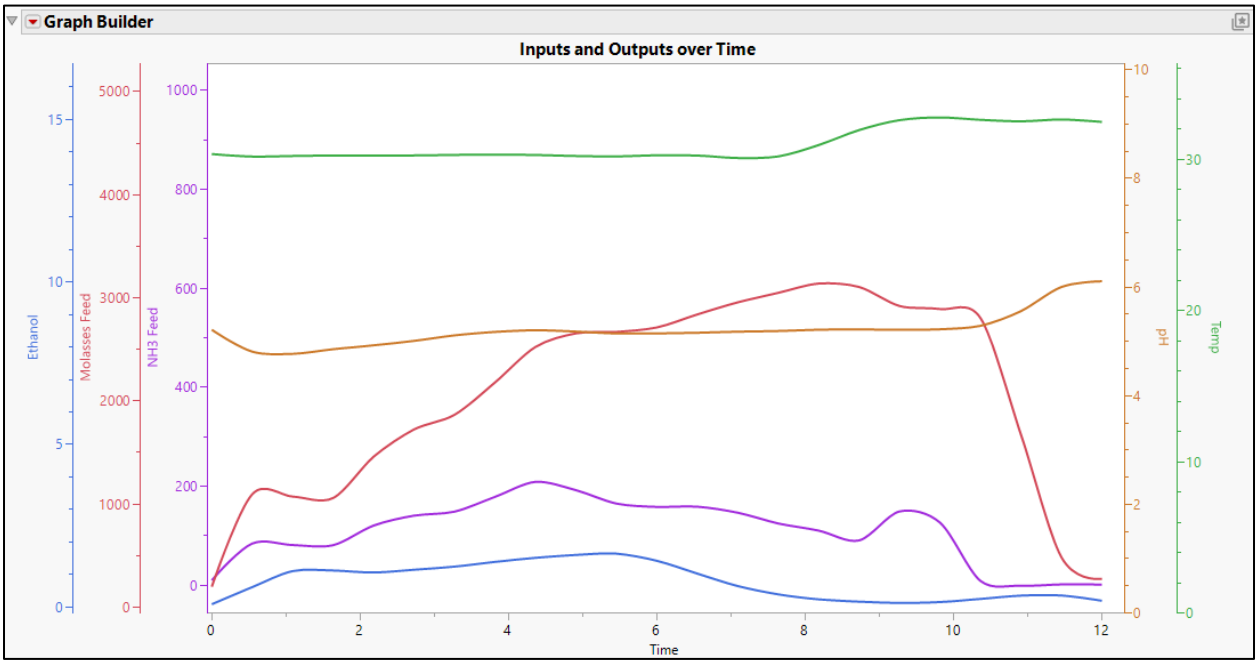
12. Click the red triangle next to **Graph Builder** and select **Parallel Y Axes**.

13. Click **Done**.

14. Click the red triangle next to **Graph Builder** and select **Show** > **Legend**.

15. Click the graph title and enter **Inputs and Outputs over Time**.

16. Resize the window to be wider than it is tall.



Each variable on the graph has its own scale.

17. Close the **Fermentation Process** data table and all associated report windows without saving.

Summary

We hope these examples of graphs and graphing techniques are useful to you in visualizing your own data and models. You can find much more information, including short videos, short articles, full courses of study, and hands-on activities in the Learn JMP section of the JMP user community: https://community.jmp.com/t5/Learn-JMP/ct-p/learn-jmp.