

Split Column on Specified Delimiter

Author: Greta Linse Peterson, Golden Helix, Inc.

Overview

This script prompts the user to select a column that needs to be split on a specified delimiter and for the delimiter to use. The delimiter can be more than one character.

Possible uses include creating columns from combined family and individual IDs to two columns one for family ID and one for individual ID. Then the spreadsheet editor can be used to change the row labels to either the family ID or the individual ID.

Recommended Directory Location

Save the script to the following directory:

*..\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\Edit\

Note: The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between XP and Vista. The easiest way to locate this directory on your computer is to open SVS and go to **Tools >Open Folder > UserScripts Folder**. If saved to the proper folder, this script will be accessible from the spreadsheet **Edit** menu.

Using the Script

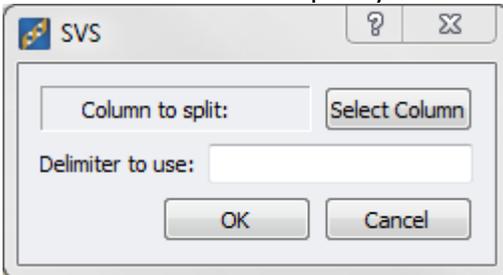
1. Open a spreadsheet containing a column with a delimiter that needs to be split into more than one column. This can be the row label column.

Note: Only active data will remain in the resulting spreadsheet.

In the sample spreadsheet below, the row labels need to be split on the “->” delimiter.

Map	Samples	R
1	fam_106->W3002	1.57
2	fam_103->W3001	2.12
3	fam_105->F5001	2.02
4	fam_109->W3003	1.67
5	fam_107->M4002	1.52
6	fam_108->F5002	1.62
7	fam_110->M4003	1.82
8	fam_101->M4000	2.17
9	fam_100->W3000	1.97
10	fam_113->M4004	2.22
11	fam_112->W3004	1.77
12	fam_111->F5003	1.87
13	fam_102->F5000	1.72
14	fam_104->M4001	1.92
15	fam_114->F5004	2.07

2. Select **Edit > Split Column on Specified Delimiter**.
3. Select the column and specify the delimiter.



4. The resultant spreadsheet contains columns that were created from the original split on the specified delimiter. If the delimiter exists more than once in some columns missings will be filled in for the columns that have fewer than the maximum number of delimiters. Any active columns in the original spreadsheet will be added to the new spreadsheet.

Additional Information:

Setting one of the new columns as the row labels.

1. From the new spreadsheet go to **Edit > Edit This Spreadsheet**. This opens the Spreadsheet Editor.

		C 1	C 2	R 3
	Samples	Samples 1	Samples2	rs123456
1	fam_106->W3002	fam_106	W3002	1.57
2	fam_103->W3001	fam_103	W3001	2.12
3	fam_105->F5001	fam_105	F5001	2.02
4	fam_109->W3003	fam_109	W3003	1.67
5	fam_107->M4002	fam_107	M4002	1.52
6	fam_108->F5002	fam_108	F5002	1.62
7	fam_110->M4003	fam_110	M4003	1.82
8	fam_101->M4000	fam_101	M4000	2.17
9	fam_100->W3000	fam_100	W3000	1.97
10	fam_113->M4004	fam_113	M4004	2.22
11	fam_112->W3004	fam_112	W3004	1.77
12	fam_111->F5003	fam_111	F5003	1.87
13	fam_102->F5000	fam_102	F5000	1.72
14	fam_104->M4001	fam_104	M4001	1.92
15	fam_114->F5004	fam_114	F5004	2.07

2. Go to **Scripts > Create Column From Row Labels** if you want to preserve the original row labels.
3. Then click on the column header for the column to use as the new row labels and select **Set Row Labels as This Column**.

		C 1	C 2
	Samples	Samples	Samples
1	fam_106->W3002	fam_106->W3002	fam_106
2	fam_103->W3001	fam_103->W3001	fam_103
3	fam_105->F5001	fam_105->F5001	fam_105
4	fam_109->W3003	fam_109->W3003	fam_109
5	fam_107->M4002	fam_107->M4002	fam_107
6	fam_108->F5002	fam_108->F5002	fam_108
7	fam_110->M4003	fam_110->M4003	fam_110
8	fam_101->M4000	fam_101->M4000	fam_101
9	fam_100->W3000	fam_100->W3000	fam_100
10	fam_113->M4004	fam_113->M4004	fam_113
11	fam_112->W3004	fam_112->W3004	fam_112
12	fam_111->F5003	fam_111->F5003	fam_111
13	fam_102->F5000	fam_102->F5000	fam_102
14	fam_104->M4001	fam_104->M4001	fam_104
15	fam_114->F5004	fam_114->F5004	fam_114

4. You can either delete the original column or keep the extra copy.
5. Go to **File > Save As Top-Level Spreadsheet**.

Joining Spreadsheets On Original Row Labels and Setting New Labels

1. Take the new spreadsheet and join on row labels to a spreadsheet that has the old row labels.
2. Follow the instructions above to select the new row labels that were created after splitting on the delimiter.