

CSV filter

Short script to filter csv file using the first couple of columns using a different csv file

Example usecase

I have an output csv file from cell profiler called **data.csv**.

First column is an image number and second column is an object number, the rest are the different measurements.

Based on the visual control I want to: * remove measurements single object (object 1) from an image 1 * remove measurements of all objects from image 2.

1. I create a file in excel called **filter.csv**
2. Put 1 in first column (image) and second column (object)
3. Put 2 in first column (image), **all** in second column
4. Save as csv.
5. Run the script:
 - using alias `csv-filter filter.csv data.csv` (the alias must be set in your .bashrc)
 - using the script itself `bash filter_csv.sh filter.csv data.csv`
6. Check the `filt_date-time_data.csv`

data.csv

AutoSave ☐ Off data • Saved to this PC ▾

File Home Insert Page Layout Formulas Data Review View Help Acrobat

Clipboard: Paste, Cut, Copy, Format Painter

Font: Aptos Narrow, 11, Bold, Italic, Underline, Text Color, Background Color

Alignment: Left, Center, Right, Indent, Decrease Indent, Increase Indent, Wrap Text, Merge & Center, Text Wrapping

Formula Bar: A1, fx, ImageNumber

	A	B	C	D	E	F	G	H	I	J	K	L
1	ImageNumber	ObjectNumber	FileName	FileName	FileName	FileName	PathName	PathName	PathName	PathName	AreaShape	AreaS
2	1	1	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2764	?
3	1	2	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2039	?
4	1	3	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1428	6
5	1	4	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2120	4
6	1	5	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	5297	2?
7	1	6	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2119	4
8	1	7	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	4366	20
9	1	8	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1525	?
10	1	11	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1193	?
11	1	12	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	3457	1?
12	1	14	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1262	?
13	1	15	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	6264	1?
14	1	21	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2043	?
15	1	22	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1680	?
16	1	26	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1633	?
17	1	27	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2094	?
18	1	28	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	10655	2?
19	1	30	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1688	?
20	2	2	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2404	?
21	2	3	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1761	?
22	2	5	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	4038	10

filter.csv

AutoSave Off filter

File Home Insert Page Layout Formulas Data

Clipboard: Paste, Cut, Copy, Format Painter

Font: Aptos Narrow, 11, B, I, U, [Grid], [Color]

	A	B	C	D	E
1	1	1			
2	2 all	all			
3					
4					

filt_date-time_data.csv

	A	B	C	D	E	F	G	H	I	J	K	L
1	ImageNumber	ObjectNum	FileName	FileName	FileName	FileName	PathName	PathName	PathName	PathName	AreaShap	AreaSh
2	1	2	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2039	73
3	1	3	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1428	66
4	1	4	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2120	42
5	1	5	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	5297	272
6	1	6	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2119	49
7	1	7	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	4366	202
8	1	8	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1525	31
9	1	11	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1193	18
10	1	12	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	3457	171
11	1	14	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1262	18
12	1	15	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	6264	174
13	1	21	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2043	30
14	1	22	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1680	24
15	1	26	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1633	26
16	1	27	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	2094	37
17	1	28	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	10655	271
18	1	30	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	1688	28
19	3	1	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	4947	85
20	3	2	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	668	9
21	3	3	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	FN1_dPDZ	Z:\vladimi	Z:\vladimi	Z:\vladimi	Z:\vladimi	3386	54

Files

script
data.csv
filter.csv
filt_date-time_data.csv

Script

```
#!/bin/bash
time_stamp=$(date +%d-%b_%Hh%Mm%Ss')
[ $# -lt 2 ] && echo ""
Usage:
$0 <filter> <data-to-filter>
```

Description:

This script filters out lines from <data-to-filter> that match entries in <filter>

Each line in <filter> should contain one or more comma-separated columns.

Use 'ALL' (case-insensitive) in the second, third etc. column(s) to match any value in given column.

```
"" \
    && exit 1

sed -E 's/,[Aa][Ll][Ll])$/,[^,]*/' "$1" \ # convert the all, ALL to *
| sed -E 's/^/^/; s/$/,/' \             # make sure the patterns are anchored to beginning and end with a
| grep -E -v -f - "$2" > "filt_${time_stamp}_$2" 2>/dev/null # filter the data file

[ $? -ne 0 ] && echo "Something went wrong"

echo "Filtering finished"
```