Subject: Merging files error
Posted by Adk2020 on Tue, 21 Apr 2020 20:45:32 GMT
View Forum Message <> Reply to Message

Hello All,

I both new to data warrior and this form, so forgive the naivety of the question. When I try to merge two files (both of which are csv files) I get the following error- source column not defined - see attached file. The incoming file has two more columns than the file I want to merge with. What am I missing?

While I have your attention, these two files represent a primary screen of 262 molecules and a confirmatory screen (hits) of 111 cmpds. The data file for the primary screen will eventually ave the same number of column descriptors as hit confirmation file. What I'd like to use data warrior is to compare both files across all descriptors and determine which/why molecules didn't confirm. For example, if the only 5 out of 50 aryl halides confirmed, how do the hits differ from the non hits.

I think this software is really cool and has lots of potential for visualizing analyses of this type and more. Thank you for your time and consideration.

Frank

```
File Attachments
1) merge file error.jpg, downloaded 561 times
```

Subject: Re: Merging files error
Posted by thomas on Wed, 22 Apr 2020 07:56:38 GMT
View Forum Message <> Reply to Message

Hi Frank,

I suspect that the problem is an empty column at the end of your csv file. Especially the empty column name, which you have assigned to 'new column' causes the issue. When checking, whether your merge options, DataWarrior does not expect columns to have no header. If you check your csv file, you will probably find that the first line (potentially all lines) ends with a comma. If you add a few characters to this line, you should be able to merge.

Please let me know, if I was wrong.

Thomas

Subject: Re: Merging files error
Posted by Adk2020 on Wed, 22 Apr 2020 16:36:13 GMT
View Forum Message <> Reply to Message

Thank you Thomas for the quick reply. Just so you know, I have no computational or modeling experience. I am using data warrior to learn how to perform simple correlations between molecular structures (R groups, potency and the like) to better understand the SAR.

With that said, I believe you are correct, in that the last three columns of the "primary screen" file are empty - I have yet to parse the results and fill those columns. As for your second point, please forgive my ignorance, I've opened my csv file and looked across the header row but I see no comma. I've attached a snapshot of the file "headers" common to both files as I am unsure of what you are referring to. BTW, the structure column is empty as well.

Thomas I couple of other questions that have come up as well. Should I ask them here I post to a new message. Once again thank you for your help

---

Subject: Re: Merging files error
Posted by Adk2020 on Wed, 22 Apr 2020 16:41:27 GMT
View Forum Message <> Reply to Message

sorry Thomas, here is the file header I forgot to attach in my reply

File Attachments
1) header line.jpg, downloaded 581 times

Subject: Re: Merging files error
Posted by thomas on Wed, 22 Apr 2020 18:56:45 GMT
View Forum Message <> Reply to Message

sorry, I wasn't precise. A csv file is a text file with row entries separated by commas. If you, however, double click it on Windows, it often opens in a spreadsheet application (e.g. Excel), where you cannot see the commas anymore, because they are removed in the process of splitting up a line's content into individual cells.

To see the commas and add a column header you need to open the file in a text editor. On Windows, if you don't have anything better, you should be able to use Notepad for it.

What you could do alternatively is to just copy all relevant cells in Excel and paste it into DataWarrior (Edit->Paste). That way you get a new DataWarrior window with all the information. For every column that contains Smiles you automatically get a new structure column. If you have empty columns, you can delete them within DataWarrior. You may then save the file. Do that with the other file as well. Then you can merge both DataWarrior files. The only thing to make sure is that the first line of copied cells contains column names and that none of the copied cells in the first line is empty.

Hope this works,

Thomas

PS: for other questions you may also contact me privately on the e-mail in the openmolecules.org about.

Subject: Re: Merging files error
Posted by Adk2020 on Sat, 25 Apr 2020 13:33:47 GMT
View Forum Message <> Reply to Message

Thomas no apologies necessary. Thanks for your help and the offer, but I'll post my questions publicly, this way it may help other beginners having similar issues. Have a good day.