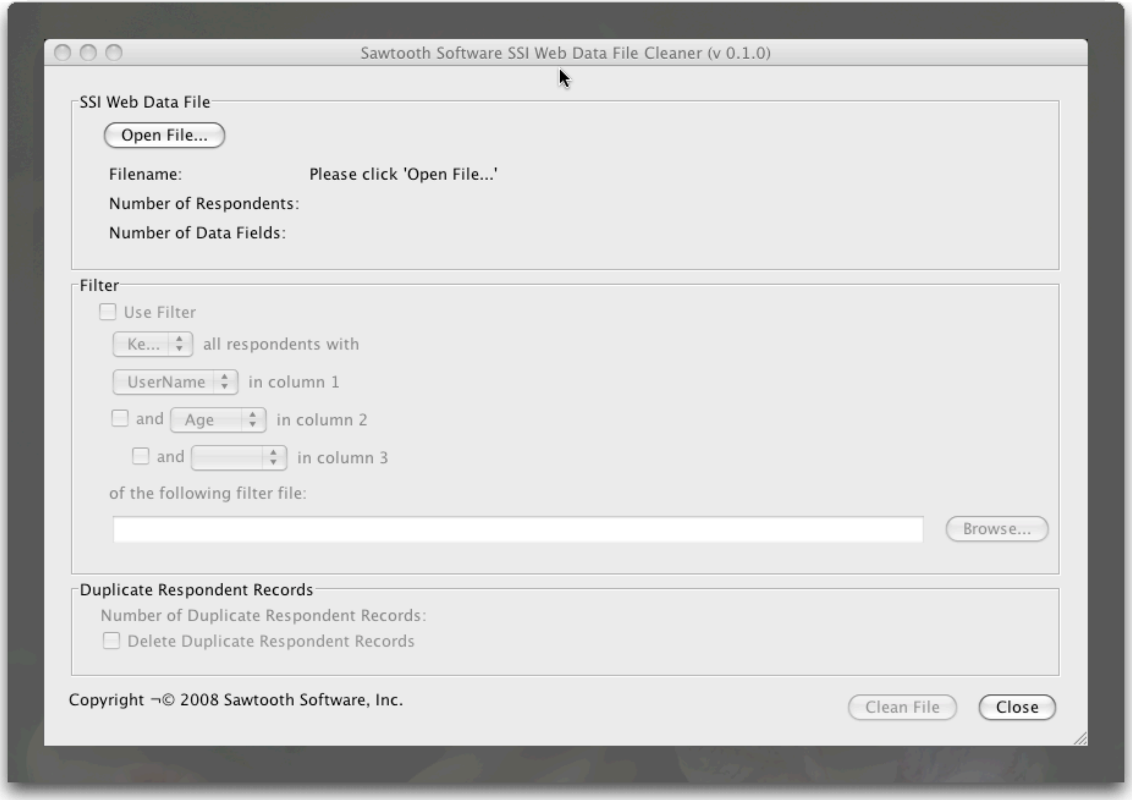
SSI Web Data Cleaner

When the SSI Web Data Cleaner is opened the main settings screen is shown:



Sawtooth Software SSI Web Data File Cleaner (v 0.1.0)

SSI Web Data File

Open File...

Filename: Please click 'Open File...'

Number of Respondents:

Number of Data Fields:

Filter

□ Use Filter

Ke... all respondents with

UserName in column 1

□ and Age in column 2

□ and in column 3

of the following filter file:

Browse...

Duplicate Respondent Records

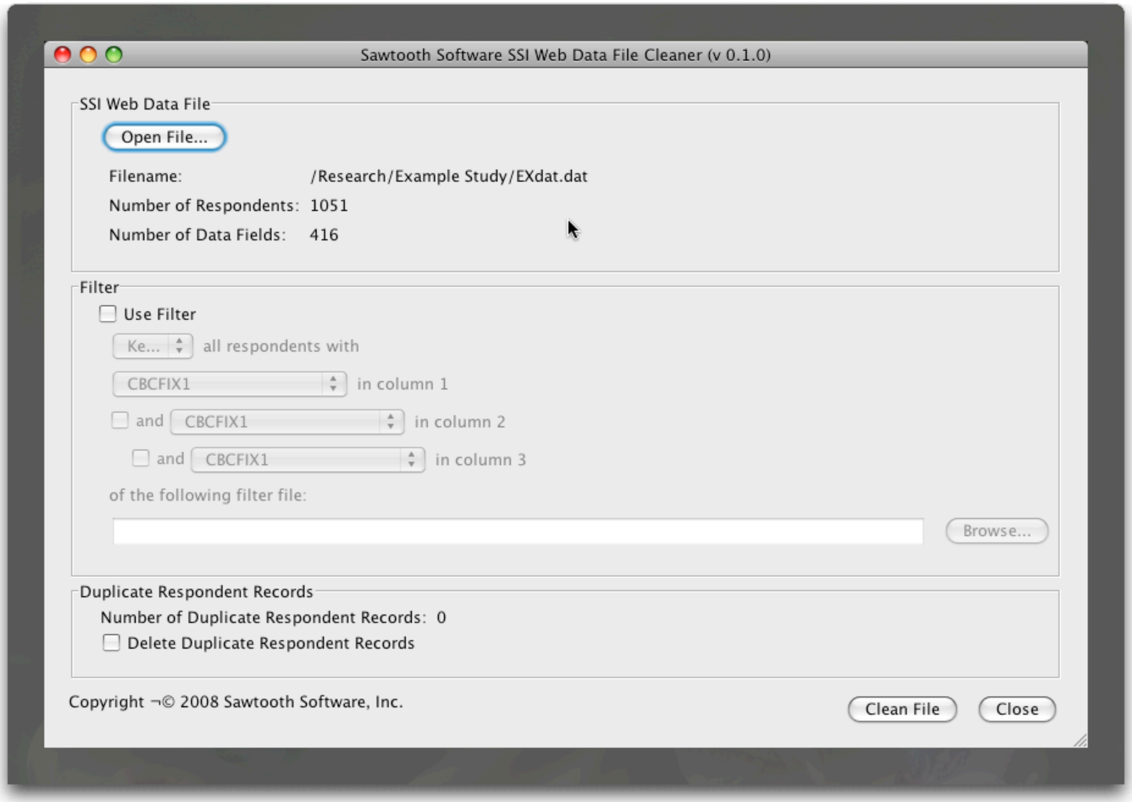
Number of Duplicate Respondent Records:

□ Delete Duplicate Respondent Records

Copyright ¬© 2008 Sawtooth Software, Inc.

Clean File Close

The first step is to select an SSI Web data file. Select ***Open File…*** to browse to a **.dat** file downloaded using the Online Administration Module or prepared by SSI Web using the ***Field + Accumulate CAPI Data ….*** or ***Field + Accumulate Paper and Pencil Files…*** When a file is selected, the Filename, Number of Respondents, and Number of Data Fields will be updated to reflect what is found in the data file:



Sawtooth Software SSI Web Data File Cleaner (v 0.1.0)

SSI Web Data File

Filename: /Research/Example Study/EXdat.dat

Number of Respondents: 1051

Number of Data Fields: 416

Filter

□ Use Filter

Ke... all respondents with

CBCFIX1 in column 1

□ and CBCFIX1 in column 2

□ and CBCFIX1 in column 3

of the following filter file:

Duplicate Respondent Records

Number of Duplicate Respondent Records: 0

□ Delete Duplicate Respondent Records

Copyright ¬© 2008 Sawtooth Software, Inc.

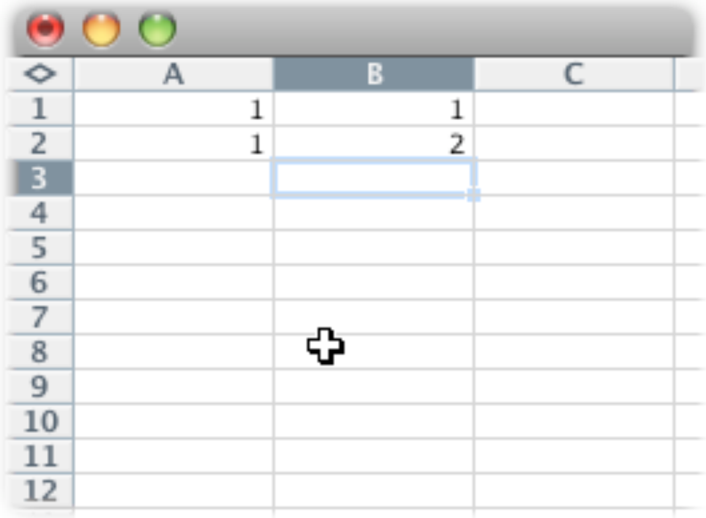
This data file is called EXdat.dat and has 1051 respondents and 416 data fields.

Each respondent record is also compared in the data file to determine whether it is a duplicate of another record. A respondent is considered to have a duplicate record if two records have the same Internal Respondent number and the same End Time. If duplicate respondents have been found and you would like them removed check ***Delete Duplicate Respondent Records*** and click ***Clean File***. (The first record will be kept and all subsequent duplicate records removed.) A backup copy of your original file will be saved with a .bak 001 extension, in this example EXdat.dat.bak 001, and the duplicate free file will be saved using the original file name, in this case EXdat.dat. If a .bak 001 file already exists, the backup will be named .bak ### where ### is the next available number.

The SSI Web Data Cleaner also allows you to filter respondents from the data file. This can be useful for removing bad respondents, test data, or simply to segment your sample prior to bringing the data into another program for estimation. To filter respondents, you will need a file containing the variables that you would like to use as filters. Generally this will be a list of respondent numbers, but it may also be an answer to one of your demographic questions.

The file format for the filter file is a .csv file. This file can be exported from SSI Web or can easily be created using Excel®, SPSS® or other data manipulation programs. The filter file does not need a header row and will contain a list of the variables that you wish to filter on. If you were filtering on respondent numbers, you would simply have a single column listing all the respondent numbers that you wish to keep or wish to remove. If you are filtering on more than one variable you would have one variable for each filtering rule.

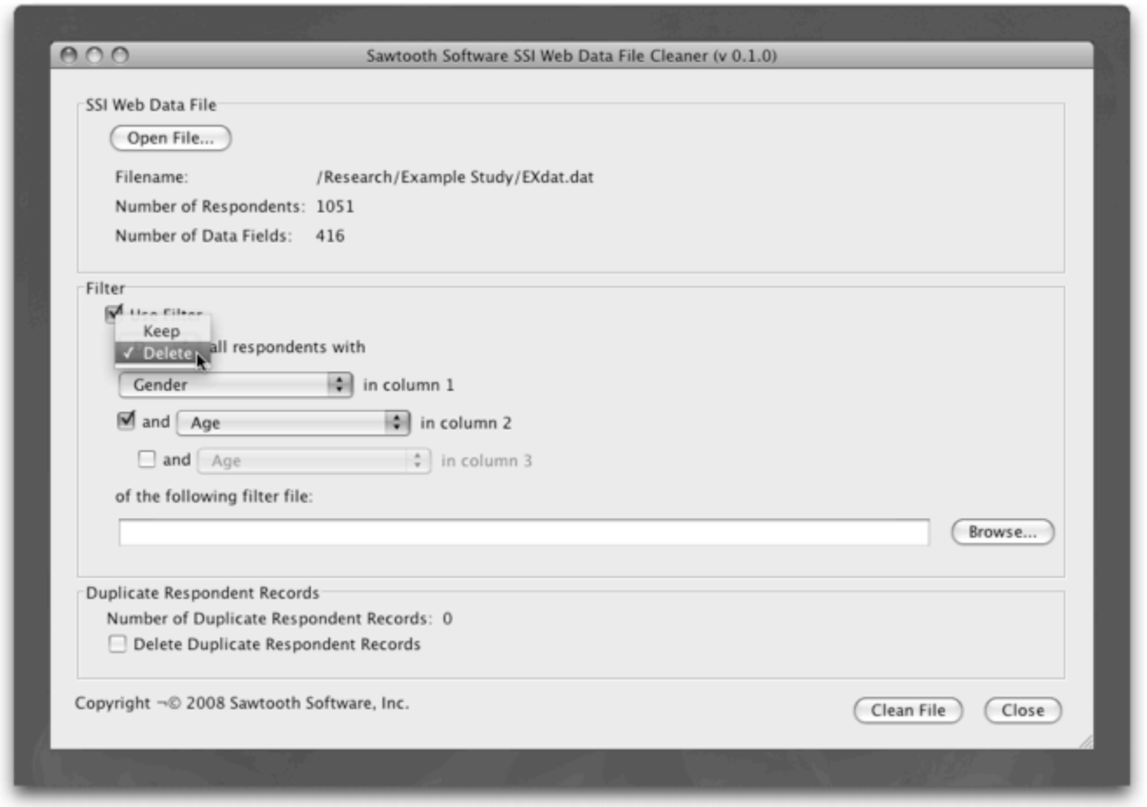
The software will support up to three filtering variables. The respondent must match all filter conditions for the filter to apply. For example, if you wanted to filter out all males under the age of 30. Assume you had two questions, Gender with response options 1 = Male and 2 = Female and Age with response options 1 = under 18, 2 = 18-29, 3 = 30-44, 4 = 45-60, and 5 = over 60. The filter file would be:



	A	B	C
1	1	1	
2	1	2	
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			

Using Excel you would then select ***Save As…*** and from the *File Type* dropdown select ***Comma Separated Value (.CSV)***. Give the file a name and then select ***Save.*** Click ***OK*** to save just the active sheet and then click ***Yes*** to keep the file as a .csv file.

In the SSI Web Data Cleaner select the checkbox labeled *Use Filter* then select whether to keep or delete respondents that match the filter. In this example we will select delete since we want to remove all males under the age of 30.



SSI Web Data File

Filename: /Research/Example Study/EXdat.dat

Number of Respondents: 1051

Number of Data Fields: 416

Filter

☑ Use Filter

Keep
✓ Delete all respondents with

Gender in column 1

☑ and Age in column 2

□ and Age in column 3

of the following filter file:

Duplicate Respondent Records

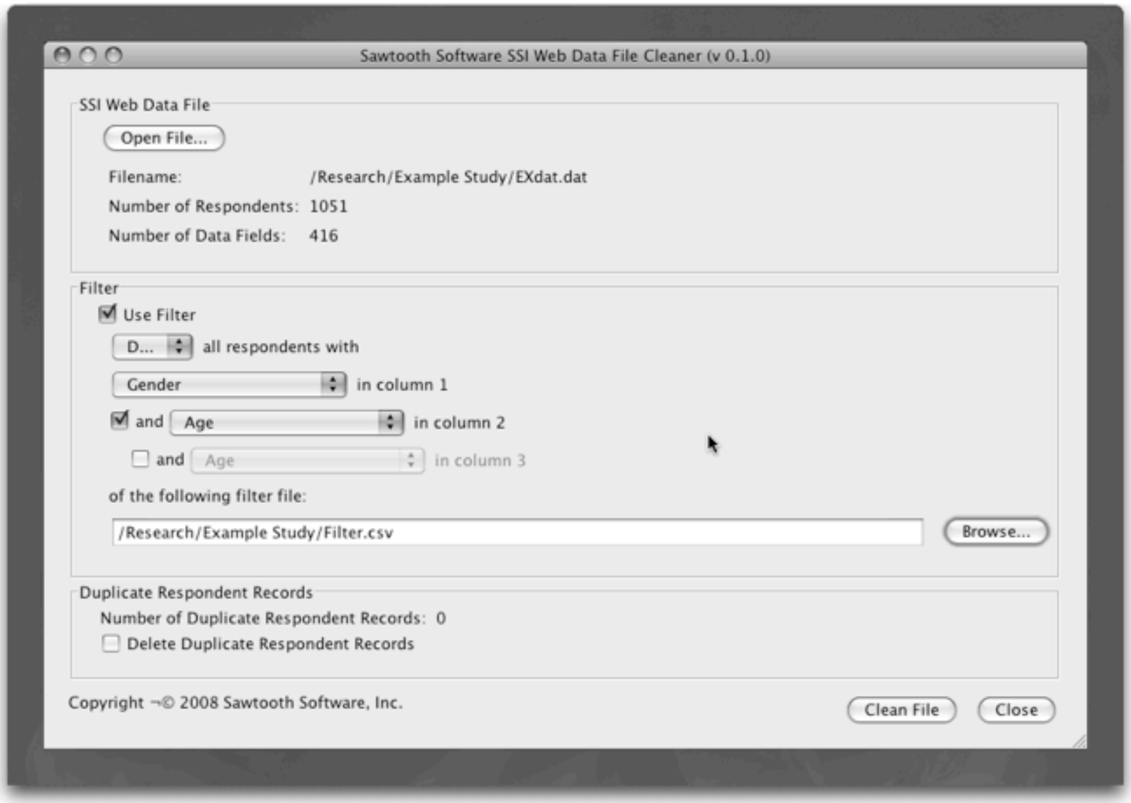
Number of Duplicate Respondent Records: 0

□ Delete Duplicate Respondent Records

Copyright ¬© 2008 Sawtooth Software, Inc.

You would then select the fields to filter by from the drop down menu. To filter based on more than one column you would need to select the checkbox to the left of the appropriate dropdown. For this example our .csv file has the variable related to Gender in the first column and the variable related to Age in the second column. We will therefore select Gender in the first dropdown box and Age in the second dropdown box. The final checkbox will be left unchecked indicating we only have two filter fields. (Most of the time you will only be filtering based on one column.)

We then need to select the .csv file we created that contains the filtering rules. Select ***Browse*** and find the .csv file you created. In our example this file is called *Filter.csv*.


Sawtooth Software SSI Web Data File Cleaner (v 0.1.0)
SSI Web Data File
Open File...

Filename: /Research/Example Study/EXdat.dat

Number of Respondents: 1051

Number of Data Fields: 416

☑ Use Filter

D... all respondents with

Gender in column 1

☑ and Age in column 2

□ and Age in column 3

of the following filter file:

/Research/Example Study/Filter.csv

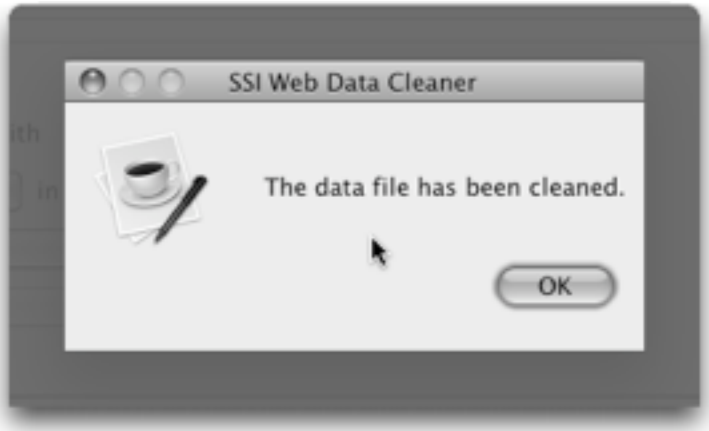
Browse...
Duplicate Respondent Records

Number of Duplicate Respondent Records: 0

□ Delete Duplicate Respondent Records

Copyright ¬© 2008 Sawtooth Software, Inc.

The data file is now ready to be cleaned. Clicking ***Clean File*** will create the backup file and save the cleaned file with the original name. If the *Clean Data* is successful, the program will display a success message:



The data file has been cleaned.

Selecting ***Close*** will close the file without changing the existing files.

FAQ

Q: What are duplicates based on?
A: Duplicates are currently based on the sys_InternalRespondentNumber and sys_EndTime. Both fields must match for a respondent to be considered a duplicate.

Q: Can I filter respondents based on Sequential Respondent Number?
A: No. SSI Web assigns a Sequential Respondent Number to respondents as they are exported. This number can change if the data file is changed. For example if I were to filter out Sequential Respondent Number 10, the next time I exported the data Respondent 11 would be come 10 and so forth. To avoid the confusion this could cause we have chosen not to implement filtering based on Sequential respondent Number.

Q: How is the number of data fields calculated?
A: The number of data fields is the total number of fields available as filters. It includes all the system fields like sys_StartTime so it will not match the number of data fields from your SSI Web study.