To start, let's circle back to the task of merging 3 CSV files into 1 file. We can use [csvstack tool](http://csvkit.readthedocs.io/en/0.9.1/scripts/csvstack.html" \l "description" \t "_blank) to consolidate the rows from multiple CSV files and redirect the stdout to a new file:



csvstack file1.csv file2.csv file3.csv > final.csv

As long as the header row for each file in the stdin to csvstack is the same, the first row in the resulting file will match this header row. After the header row, final.csv will contain all of the non-header rows from file1.csv, then all of the non-header rows from file2.csv, then finally the non-header rows from file3.csv. If you don't redirect the stdout of csvstack to a file or a tool like head, the full output will be rendered in the terminal. This can cause your terminal to grind to a halt as it tries to process and display all of the output and you want to be extra careful to avoid doing so.

If you peeked at the [documentation](http://csvkit.readthedocs.io/en/0.9.1/scripts/csvstack.html#description), you may have noticed that the behavior of csvstack can be modified using a few different flags. For example,

if you want to be able to trace the file where each row originated from in the merged file, you can use the -g flag to specify a grouping value for each filename. When stacking the rows from a file, csvstack will add the corresponding value in a new column. Lastly, you can use the -n flag to specify the name of this new column. The following code will create a new column named origin, containing the values 1, 2, or 3 depending on which file that row originated from:



csvstack -n origin -g 1,2,3 file1.csv file2.csv file3.csv > final.csv

The rows in final.csv that originated from file1.csv will contain the value 1 in the origincolumn and those from file2.csv will contain the value 2 in the origin column. Let's now use csvstack to combine the 3 datasets on U.S. housing affordability from the last challenge.

Instructions

* Merge Hud\_2005.csv, Hud\_2007.csv, and Hud\_2013.csv in that order into one file:
  + Name the resulting file Combined\_hud.csv.
  + Add an extra column named year which contains the year value from the file name for each row. E.g. the rows that originated from Hud\_2005.csv should have 2005 as the value in the yearcolumn.
* Use head to preview the first few rows of Combined\_hud.csv.

/home/dq$ csvstack -n year -g 2005,2007,2013 Hud\_2005.csv Hud\_2007.csv Hud\_2013.

csv > Combined\_hud.csv

/home/dq$ head -5 Combined\_hud.csv

year,AGE1,BURDEN,FMR,FMTBEDRMS,FMTBUILT,TOTSAL

2005,43,0.513,680,'3 3BR','1980-1989',20000

2005,44,0.223,760,'4 4BR+','1980-1989',71000

2005,58,0.218,680,'3 3BR','1980-1989',63000

2005,22,0.217,519,'1 1BR','1980-1989',27040