

How do I count the number of people in a household by gender?

```
data gender (keep =ctmale ctfemale hhid);
set pums;
by hhid;
retain ctmale ctfemale;
if first.hhid then do ;
ctfemale= 0 ; ctmale= 0 ; end ;
if gender= 1 then ctmale=ctmale+ 1 ;
else if gender= 2 then ctfemale=ctfemale+ 1 ;
if last.hhid then output ;
run ;
```

So, the keep statement tells SAS to create a new data set with only those three variables (ctmale, ctfemale and hhid).

- Retain holds the value over through the course of the iterations performed in the datastep.
- If first.hhid tells SAS to perform the following command on the first household id. ctfemale=0 initializes the variable to zero, then with the end statement, SAS moves on to the other conditions.
- Ctmale+1 is creating a running count for the whole household.
- If last.hhid tells SAS to output to the dataset once the last id is reached.

Now you have a dataset called 'gender' that looks like this:

<u>hhid</u>	<u>ctmale</u>	<u>ctfemale</u>
123	1	1
456	3	1
789	0	3

The next move is to merge this file with only three variables back on to the original PUMS file.

```
data newpums;
merge pums gender;
by hhid;
run ;
```

- Data newpums is the name of the new dataset with the variables we created (ctmale and ctfemale).
- Merge is a simple joining of the original 'PUMS' data with the new file we just created called 'gender'.
- by hhid is the common merge variable.

Once you finish the merge you are left with a new dataset called 'newpums' that looks like this

<u>Household Id</u>	<u>Person Id</u>	<u>Age</u>	<u>Gender</u>	<u>Marstat</u>	<u>Pob</u>	<u>Educ at</u>	<u>Hhinc</u>	<u>Emp loy</u>	<u>Ctmale</u>	<u>Ctfemale</u>
123	1	67	1	1	10	12	45000	3	1	1
123	2	66	2	1	70	12	45000	3	1	1
1	1	1	1	1	1	1	1	1	1	1
789	3	8	2		32		24000		0	3