### TheDataWeb & DataFerrett

DataFerrett is a highly sophisticated web-based analytical tool developed by the U.S. Census Bureau designed to equip users with the ability to analyze large amounts of data and create customized reports to support decision-making. DataFerrett draws upon the DataWeb, a distributed data dissemination system of public and private databases providing a vast amount of statistical information that is constantly updated and expanded.

DataFerrett is used for purposes such as online data integration and web-based data driven visualizations, and provides a unique and effective tool to internal and external Census customers that allows users to point and click to extract data, and create custom tabulations. The datasets in DataFerrett come from many different sources and organizations, which are the providers of the data and subsequent supporting documentation.

There are two basic types of data that DataFerrett accesses:

1. Microdata, in which the data record represents a survey response or an administrative record,

2. Aggregated data, in which a variable contains an estimate of a characteristic (e.g., the number of factories in a county.)

DataFerrett is a highly efficient research tool, in use by both internal and external customers who work for a myriad of organizations, such as agencies of the Federal government, State and Local governments, universities, non-profit organizations, and divisions within the Bureau of the Census.

DataFerrett supports the Census Bureau's mission to serve as the leading source of quality data about the nation's people and economy by providing a mechanism for external customers to analyze many varied data sets such as the American Community Survey (ACS) data, County Business Patterns (CBP), and data variables from demographic variables to business variables.

You can watch a very useful seven minute video on the basics of using DataFerrett by visiting <u>https://www.youtube.com/watch?v=STRn4XdTNo0</u> on the US Census Bureau YouTube channel. There is also a second part to the video at <u>https://www.youtube.com/watch?v=ArWHkuh0CtU</u>

The following pages contain a step-by-step introductory exercise using DataFerrett, providing a basic overview of selecting a dataset, variables and their values, and creating a table. It also illustrates several advanced and highly useful functions, including creating a recode, a table formula, and generating a thematic map.

DataFerrett Help http://dataferrett.census.gov/ 1-866-437-0171 (toll free) dsd.ferrett@census.gov

Task 1: Using the ACS PUMS data, produce a table breaking out the native born and foreign born populations by year of entry into the U.S., either before 2000 or in 2000 or later, for all states.

### START DATAFERRETT

- 1. Go to <u>http://dataferrett.census.gov/</u>, click on Get Data ~ Run: DataFerrett, and launch DataFerrett.
- 2. Enter your email address when prompted.

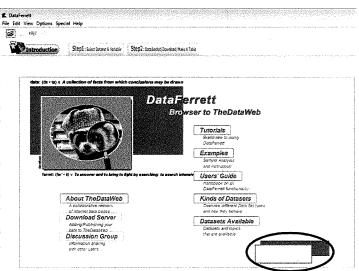
**NOTE:** The DataFerrett Team only uses the email address when returning downloaded data or responding to help messages, and on occasion to send out an announcement regarding special notices. It does not give your

email address out. Passwords are only necessary and issued to users to access their own data or private data.

The Introduction Tab is the screen you arrive at after you log in. Step 1 and Step 2 are the data application tabs.

The Introduction Screen has the Get Data Now link to go straight into the DataFerrett application.

**NOTE:** Select the links to access the user guide, tutorials, frequently asked questions, and other information about using DataFerrett.



### STEP1: SELECT DATASET & VARIABLE TAB

- 1. Select the Step 1 Select Dataset & Variables tab.
- 2. Click on the plus sign next to the American Community Survey folder.
- 3. Click on the plus sign next to the Public Use Microdata Sample folder.
- 4. Click the 2014 bullet.
- 5. A **Description** / **View Variables** fly-out will allow you to choose the Dataset Description or to View Variables.

View Variables will bring up the topics pertinent to your highlighted dataset.

- 6. Click on the Select All Topics button.
- 7. Click the Search Variables button.

You will need the following variables:

Nativity (NATIVITY), Year of entry (YOEP), and Geography(Geographic Items)=all states

- 8. Sort the NAME column (click the NAME column header).
- 9. Highlight the **NATIVITY** variable.
- 10. Scroll down to the bottom of the list, hold the Ctrl key, and highlight the YOEP variable.

Once you highlight a variable, the Browse/Select Highlighted Variables button becomes active.

11. Click the Browse / Select Highlighted Variables button.

NOTE: You can also double click on a variable to pop-up the Browse Variable window.

Selecting the **Browse/Select Highlighted Variables** button will open a new window that allows you to browse variable descriptions and values for all items that you selected in the **Select Dataset & Variables** tab. This is also called the Codebook. Additionally, this window allows you to select the variables and their values that you want to put into your DataBasket where you can recode, modify, or delete a variable.

- 12. Check the box marked Select ACS NATIVITY for the Nativity variable.
- 13. Select (highlight) the ACS YOEP.
- 14. Check the box marked Select ACS YOEP Year of Entry variable. This places boxes in front of all the values.
- 15. Uncheck the first value that says 1920) Not Eligible Born in the US (This restricts the universe to only Foreign Born).
- 16. Click OK in the upper, right corner
- 17. Confirm the addition of 2 variables to your data basket by clicking OK in the pop-up window that appears.

You have added NATIVITY and YOEP variables, with a sub-selection to YOEP, to your data basket.

-2-

Next, choose your geography variable (Step 1):

1. Double Click the → Selectable Geographies Variable → from the variable list (NOTE: the name of the variable is geography).

			Highlight ·
			377 Variables returned from search.
Торіс	Name	Availability	Variable Label
Housing	WGTP	2006 - current	Housing Weight
Population	PWGTP	2006 - current	PUMS person weight
Population	AGEP	2006 - current	Age
Population	ANC	2006 - current	Ancestry categorization
Population	DECADE	2006 - current	Decade of entry
Selectable Geographies	Geography	2006 - current	Geographic Items
Population	DRIVESP	2006 - current	Number of vehicles calculated from JWRI

- 2. Click on State in the Types of geography available: section on the left.
- 3. Highlight State in the **Hierarchies:** section and click on the Use **Hierarchy** button at the bottom of the screen.
- 4. The list of states appears in the left section. Drag the **Select All** from the left section to the far right **Selected Geographic Areas:** section.
- 5. Click the Finish button.

### STEP 2: DATABASKET/DOWNLOAD/MAKE A TABLE

We will need to create a new variable (recode) to define just 2 categories of the year of entry -1) before 2000 and 2) in 2000 or later.

- 1. Click the STEP 2: Data Basket / Download / Make A Table tab
- 2. Highlight the Year of entry variable YOEP

Selecting a variable from your list of Current Query Variables will activate the following options: Recode, Delete, and View/Modify the variables.

- 3. Click on the Recode Variable(s) button located at the top right side of the screen.
- 4. Assign a label to your recode variable in the box labeled **RECODE1** at the top left (ex. Year of Entry Recode).
- 6. Highlight all of the categories from 1921 up to and including 1999. To do this, first click on the 1919 value, then scroll down until you can see 1999 hold the Shift key and click on 1999. This will highlight all of the values between.

#### 7. Click the Recode button at the bottom left.

1921-1999 have been assigned to value 1 of the new recode variable.

E Ferrett Microdata Recodel	1	Recode/Regroup \	/ariable	es	Tell me about				Į	8
Year of Entry is label for the Variable Re	code of YOEP									
Highlight the value(s) to recode/regroup			Select		Label	Value	S	1011100/02/2000/00/200220		
Value Description			12	1	RecodeValue_1	{1921	, 1922 ,	1924,1926	,1928 ,:	1930
1997 1997				2	Not Elsewhere Classified (nec.	(2000	,2001,	2002,2003	,2004 ,2	2005
1998 1998										
1999 1999			Landa a							
2000 2000			-							
2001 2001			-							
2002 2002										
2003 2003			- Change							
2004 2004			and the second se							
2005 2005										
2006 2006		E								
2007 2007		Summer 1	in a state of the							
2007 2007		•	Table Control of							
			1.	1	1					
Set to value 2 Recode					Redefine C	urrent	Row			Alternation
		Ok	Cance	1	]	an an an an air aige an		nganang) - o yongh no dalam di sa anda		

Notice the right hand side of the screen now displays 2 categories for the new recoded variable:

- 1 RecodeValue\_1
- 2 Not Elsewhere Classified (nec.)
- 8. Type a new label by double clicking the label for the **RecodeValue\_1**. Type the words **Before 2000** and *make sure to hit the Enter key*.
- 9. Type a new label by double clicking the label for the Not Elsewhere Classified (nec). Type the words 2000 or later and *make sure to hit the Enter key*.
- 10. Finish by clicking the OK button at the bottom of the window.

### Make a Nested Table

- 1. Click on the Make A Table button.
- 2. Drag the Geog-FIPS State Code variable to the R2,C1 cell to define the rows.
- 3. Drag the Nativity variable to the R1,C2 cell to define the columns.
- 4. Nest the Year of Entry Recode variable on the columns by dropping onto any of the Nativity labels.
- 5. Click the Go Get Data button on the menu bar.

i	GO Get Data		\$ & C	I ZI AL	Ξ		<b>%</b> [%	% %	0			
												YOEP Year of entry NATIVITY Nativity
ivot(s.	) can be dropped on pive C1	นทรงมากการโรงการเหตุการ	Charles of the second state of the second	34	C5	<u>ke</u> k		C8	C9	C10 k	21	GEOG-101 FIPS State Code
4	1	<u>~-</u>	Total NATIVITY			Native	-1		Foreign born			RECODE1 Year of Entry
21	L	Total	Before 20002		Total	Before 20002	2000 or later	Total	Before 2000		-,	
22			28,215,576						24,961,487		<sup>i</sup> C	
23	Alabama	194,154	i e i el en en el en el en el en el en el el en el	88,739	41,365	en naño este manana e	11,046	o con la conserva de server	former receiver an increase of	77,693	E	
<del>R</del> 4	Alaska	74,881	41,302	33,579	19,157	in a name international	7,471		And a second second second			
25	Arizona	1,022,263	628,150	394,113	100,622	65,549	35,073	921,641	562,601	359,040	-permit	
<b>R</b> 6	Arkansas	163,446	85,460	77,986	20,054	14,640	5,414	143,392	70,820	72,572		
R7	California	10.961,993	7,358,707	3,603,286	490,426	335,340	155,086	10,471,567	7,023,367	3,448,200		
<b>R</b> 8	Colorado	615,195	367,742	247,453	82,292	61,740	20,552	532,903	306,002	226,901		
<b>R</b> 9	Connecticut	625,411	367,732	257,679	131,352	93,415	37,937	494,059	274,317	219,742		Universe: (YOEF in (1921,192
R10	Delaware	92,962	48,817	44,145	13,242	8,939	4,303	79,720	39,878	39,842		Weight used: FWGTP
R11	District of Columbia	104,358	49,371	54,987	11,538	7,762	3,776	92,820	41,609	51,211		DeceSet(s) selected: 2014
312	Florida	4,646,888	2,737,928	1,908,960	671,071	418,146	252,925	3,975,817	2,319,782	1,656,035		۱ (m) > (
<del>२</del> 13	Georgia	1,126,609	625,283	501,326	135,441	95,058	40,383	991,168	530,225	460,943		
R14	Hawaii	292,170	183,330	108,840	51,307	32,247	19,060	240,863	151,083	89,780		
R15	Idaho	118,258	72,989	45,269	15,355	11,911	3,444	102,903	61,078	41,825		
R16	Illinois	1,910,540	1,237,767	672,773	141,393	98,395	42,998	1,769,147	1,139,372	629,775		
R17	Indiana	362,546	180,782	181,764	40,274	31,873	8,401	322,272	148,909	173,363		
218	lowa	172 237	78 345	93 892	18 177	12 421	5 756	154 060	65 924	88 136	*	A A A A A A A A A A A A A A A A A A A

**NOTE:** No need to do anything about a weight variable; it is automatically applied for ACS PUMS data files. Notice the blue text on the right side of the table and you can see that this table is weighted using the PWGTP variable (person's weight). You can get unweighted counts by going to the Options menu, then select Weighting- Unweighted, and then hit the Go Get Data button.

### Create a Formula

To create a formula to calculate the percentage of the foreign born who have entered the country since 2000, do the following:

- 1. Click in the gray column header of the next empty column in this table it is column C11.
- 2. Click in the Formula bar directly below the Go Get Data button to enter your formula and type:

=comp(c10/c8\*100)

- 3. Hit the Enter key and your formula should calculate in column C11.
- 4. Format the column with a decimal by highlighting the column by clicking in the gray C11 column header. In the menu bar, choose **FORMAT > DECIMALS** and then select **One decimal place** from the dialog box and hit **OK**.
- 5. Click in the R1,C11 cell to enter a heading for our calculated column Pct Entered Since 2000
- 6. To save the table, click **FILE > SAVE AS** on the menu bar.
- 7. Save the table as acs\_foreign\_born.ftf. This is the default format and is the table shell (or layout) for re-use within DataFerrett only.

Create a Map

- To create a map, with a geographic variable in column C1, highlight a single column containing data (not the row labels). In our example, highlight the data cells in our calculated column 11 – Pct Entered Since 2000. NOTE: You cannot highlight the entire column by clicking in the header, you must select the data cells.
- 2. Click the Map Button in the toolbar (it is the button with the yellow US shape).
- 3. A separate map window will open with your map.

Close your Map window and your Table window.

In the Step 2: DataBasket/Download/Make a Table tab, click on the Empty DataBasket button (looks like a shopping cart being dumped).

### STEP 4: CHANGE TABLE DISPLAY

1. Open the DataFerrett table shell you saved in the first part of the example. Use **FILE > OPEN** and select the **acs\_foreign\_born.ftf** file. This will open the **DataFerrett Tabulation** window with your saved table layout.

The **DataFerrett Tabulation** screen is a separate window that can be left open or closed. If you wish to add variables from Step 1, they will show up in your variable list on your tabulation window.

- 2. Do not close the Tabulation window, but use your Windows task bar to navigate back to the main DataFerrett window and go to the **Step 1: Select Dataset &Variables** tab.
- 3. Select the ACS PUMS 2014 dataset, and select the View Variables option.
- 4. Choose the Selectable Geographies topic and hit the Search Variables button located at the bottom left of the screen.
- 5. Double-click on the Geography variable in the Name column.
- 6. Highlight Public Use Microdata Area in the left section.
- 7. Highlight State > Public use microdata area code (PUMA) in the Hierarchies: section and click on the Use Hierarchy button at the bottom of the screen.
- 8. The list of states will appear in the left section. Drag Maryland to the center section and click the Next Level button.
- 9. Now you will see all the PUMAs for the state listed by their PUMA number in the left section. Drag the Select All from the left section to the far right Selected Geographic Areas: section.
- 10. Click the Finish button.
- 11. Using your Windows task bar, navigate back to your Ferrett Tabulation window. You should now see the PUMA geography variable listed at the bottom of your variable list on the right side of the window.

Now, we want to replace the States in Column C1 from our saved table with the PUMAs that we just selected.

- 12. Go to the Edit menu option and select Clear > All Rows. This removes the states from the rows.
- 13. Drag the **GEOG-Public Use Microdata Area ( PUMA)** variable into the spreadsheet and drop in **C1,R1** to define the rows with the PUMAs.
- 14. Hit the Go Get Data button. The table now displays data for all PUMAs in the state of Maryland.

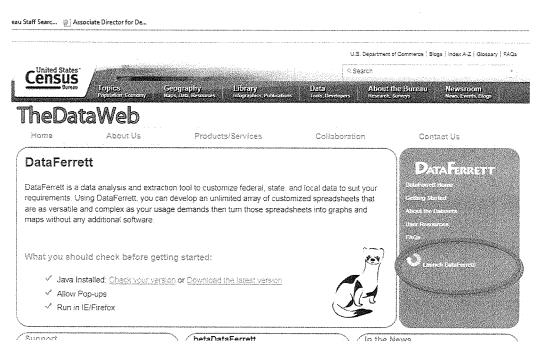
Hands-on Exercise Using DataFerrett Current Population Survey – Annual Social and Economic (ASEC) Supplement

U.S. Census Bureau

DataFerrett Help http://dataferrett.census.gov/ 1-866-437-0171 (toll free) dsd.ferrett@census.gov

### START DATAFERRETT

#### 1. CLICK LAUNCH DATAFERRETT ON THE WEBSITE:



#### 2. ENTER YOUR EMAIL ADDRESS AND SELECT OK.

NOTE: The DataFerrett Team only uses the email address when returning downloaded data or responding to help messages and on occasion to send out an announcement regarding special notices. It does not give your email address out.



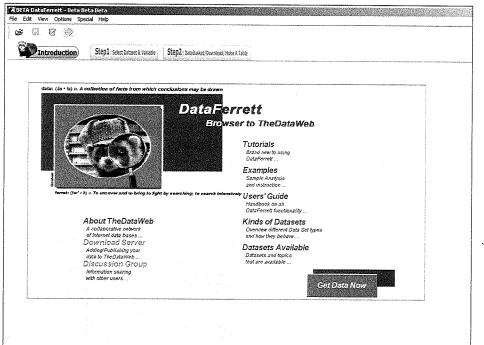
Email address: UserEmailAddressGoesHere@census.gov

The email address is used to send large extracts via email, and to inform users of new datasets available if desired. It is NOT used for any other purpose or shared with any organization.



3. ABOUT FERRETT can be found under the HELP drop-down menu at the top of the Introduction, Step 1 and Step 2 screens.

- AB	out Data	a Ferrett		x .		
File	Edit	View	Favorites	Tools	Help	
ýą				<u> </u>	💌 🔂 👻 🗇 👘 💌 Page 🖛 Safety 🕶 T	ools 🕶 🎲 🕶
			D	oli	aFerrett	
				Versio	1.3.3 (Release Notes)	
DataF	errett	(Federa	ated Elec		1 1.3.3 ( <u>Release Notes</u> ) search, Review, Extract, and Tabulation	Tool)
ls a di mining as an	ata bro tool th applica	wser fo nat acce ation on	or <u>TheDat</u> asses dat	tronic R a <u>Web</u> ( a stored ktop or u		a data e installed
Is a di mining as an differe DataF to The ability	ata bro i tool th applica nt orga errell a DataW to crea	wser for nat acce ation on inization inization allows s leb. Dat ate sum	or <u>TheDat</u> asses dat: your des ns provide earches a taFerrett a mary stat	tronic F a <u>VVeb</u> E stored ktop or u e data. across al allows ta istics. it	search, Review, Extract, and Tabulation ataFerrett is a data extraction software and a theDataWeb through the internet. It can b	a data e installed ough which published id the
Is a di mining as an differe DataF to The ability graphi This p Cente	ata bro tool th applica nt orga errett a DataW to crea cs and roject is rs for D	wser for nat access ation on mization milows s leb. Dat the sum will pro- s a joint bisease	or <u>TheDat</u> ssses dat your des ns provide earches a taFerrett i mary stat wide map t effort de	tronic R a <u>vveb</u> ( a stored ktop or u a data. across al allows ta istics. It ping and veloped vith majo	search, Review, Extract, and Tabulation ataFerrett is a data extraction software and in TheDataWeb through the internet. It can b e a java applet with an internet browser first the documentation or metadata of the data ulation of the data from different sources ar lows downloads extractions statelical hus	a data e installed bugh which published id the iness and the



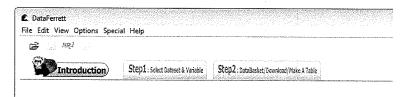
THE INTRODUCTION Tab is the screen that appears when you log in. Step 1 and Step 2 are the other two tabs.

The INTRODUCTION Screen has the Get Data Now link to go straight into the DataFerrett application.

Note: Select the links to access the users' guide, tutorials, frequently asked questions and other information about using DataFerrett.

-2-

1. SELECT THE STEP 1 SELECT DATASET & VARIABLES TAB.



- 2. CLICK ON THE PLUS SIGN NEXT TO THE CURRENT POPULATION SURVEY FOLDER. CLICK ON THE PLUS SIGN NEXT TO THE MARCH SUPPLEMENT FOLDER.
- 3. CLICK THE MAR 2014 BULLET.
- 4. A DESCRIPTION / VIEW VARIABLES FLY-OUT WILL ALLOW YOU TO CHOOSE THE DATASET DESCRIPTION OR TO VIEW VARIABLES.

View Variables will bring up the to highlighted dataset.		Search All Datasets  American Community Survey  Common Core of Data(Education)  Consumer Expenditure Survey  County Business Patterns  Current Population Survey  Common Basic  Common Basic 1989-93  Consumer Expenditure Survey  Common Core of Data(Education)  Consumer Expenditure Survey  Consu
<ol> <li>CLICK ON THE SELECT A</li> <li>CLICK THE SEARCH VA</li> </ol>		m Contingent Worker     m Disability     m Displaced Workers/Job Tenure     m Fertility
Search All Datasets American Community Survey American Community Survey Common Core of Data(Education) Consumer Expenditure Survey County Business Patterns Current Population Survey Basic Basic Basic Current Population Survey County Business Patterns Current Population Survey Basic County Business Patterns Current Population Survey County Basic County Basic	Select All Topics Person Variables Family Variables Household Variables Labor Force Variables Search Variables	

Variables will be listed on the right as seen below.

A variable list appears which includes the follow columns:

Topic - type of variable.

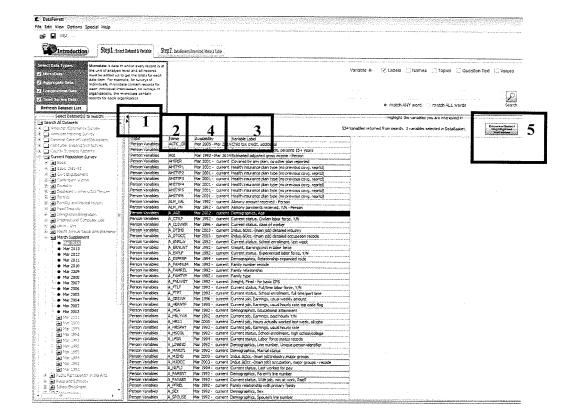
*Name* - of the variable.

Availability - the period of time for when the variable is valid.

Variable Label - short description of the variable.

#### MANAGING THE VARIABLE WINDOW

- By clicking on the **arrows (1)** indicated below, you can open up the list of Datasets pane or click the other arrow and see the full description of the variables.
- Clicking in the **gray header box (2)**, you can drag "columns" into view or any order you choose.
- Also, by clicking in the gray header box of each "column" (3) you can alphabetize by that column.
- The "Availability" column (4) shows a year and a hyphen. If there is another year after the hyphen it means that the variable was available ONLY during those years. If there is no year after the hyphen then the variable is still being asked exactly the same way today.
- Selecting the Browse/Select Highlighted Variables (5) button will open a new window that allows you to browse variable descriptions and values for all items that you selected in the Select Dataset & Variables tab.



1.

### SORT THE **NAME** COLUMN (CLICK THE NAME COLUMN HEADER) 2. HIGHLIGHT THE **A\_AGE** VARIABLE

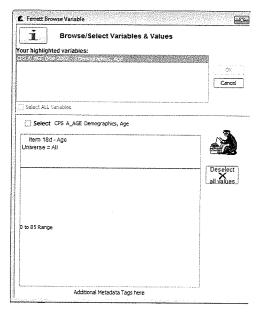
Edit View Options Special Help						
Introduction Step1 : Select Dataset &	Variable Step2: Data	lasket/Download/ Ha	ke A Table			
ect Data Types: Microdata is data in which e	very record in at §					
JecoData the unit of analysis level and must be added up to get the data item. For example, for s individuals, microdata conta	all records totals for each arveys of in records for			Variable v	Design and Sector Design an	Text 🔛 Values
engitudinal bata sech individual interviewed: organizations, the microdate trecords for each organization	contain §					
efresh Dataset List					@ match ANY word C match ALL words	Search
Search All Datasets					mighlight the variables you are interested in	
American Community Survey     American Housing Survey				534 Variab	les returned from search. O variables selected in DataBasket.	Browse/Select Righlighted Variables
Common Core of Data(Education)	Topic Person Variables	Name		ole Label		
Consumer Expenditure Survey	Person Variables	ACTC_CRD AGE1	Mar 2005 - Mar 2014 Child t Mar 1907 - Ormant Demos	graphics, Age recode, persons 15+ years		Âu.
County business Parterns	Person Variables	AGI		ated adjusted grass income - Person		
Hanse State	Person Variables	AHIPER		ed by any plan, no other plan reported		1988 B
🔆 🐜 Basic 1989-93	Person Variables	AHITYP1	Mar 2001 - Current Health	ninsurance plan type (no previous covg. reprtd)	-	10
🛞 👼 Civic Engagement	Person Variables	AHITYP2	Mar 2001 - current Health	n insurance plan type (no previous covg. reprtd)		
Contingent Worker	Person Variables	AHITYP3	Mar 2001 - current Health	n insurance plan type (no previous covg. reprtd)		
ම ලිමු Disability	Person Variables	AHITYP4	Mar 2001 - current Health	hinsurance plan type (no previous covg. reprtd)		
③ m Displaced Workers/Job Tenure	Person Variables Person Variables	AHITYP5 AHITYP6	Mar 2001 - current Health	n insurance plan type (no previous covg. reprts)		
🖲 📻 Fertőty	Person Variables	ALM VAL	Mar 2001 - current Health	n insurance plan type (no previous covg. reprtd) ny amount received - Person		
Ferbity and Marital History	Person Variables	ALM YN	Mar 1992 - Outrent Almon	ty amount received - Person ty payments received, Y/N - Person		
Fine Food Security	Person Vanables	A AGE	Mar 2002 - current Democ			
Gil Internet and Computer Use	Person Variables	A CMF		nt status, Civilian labor force, YAN		
3: (a) Charry Use	Person Variables	A CLSWKR	Mar 1995 - current Curren			
1 (a) March Annual Social and Economic Supplement	Person Variables	A_DTIND		SOcc(main job) detailed industry		
C S March Supplement	Person Variables	A_DTOCC		8Occ(main job) detailed occupation recode	1970	
Analysis      Analysis	Person Variables	A_ENRLW	Mar 1992 - ourrent Curren	nt status, School ervoliment, last week		
<ul> <li>Mar 2013</li> </ul>	Person Variables	A_ERNLSVT		t, Earnings/not in labor force		
<ul> <li>Mar 2012</li> </ul>	Person Variables	A_DPLF		nt status, Experienced labor force, Y/N		
<ul> <li>Mar 2011</li> </ul>	Person Variables	A_EXPRRP	Mar 1994 - current Demog	graphics, Relationship expanded code		
<ul> <li>Mar 2010</li> </ul>	Person Variables	A FAMANIA	Mar 1992 - current Fandy			
••• Mar 2009	Person Variables Person Variables	A_FAMREL A_FAMTYP	Mar 1992 - current Family			
<ul> <li>Mar 2008</li> <li>Mar 2008</li> </ul>	Person Variables	A_PAMITYP	Mar 1992 - current Family Mar 1992 - current (Weigh			
<ul> <li>Mar 2007</li> <li>Mar 2006</li> </ul>	Person Variables	A_FTLF		rt, Final - for basic CPS nt status, Full/time labor force, Y/N		
<ul> <li>Mar 2005</li> </ul>	Person Variables	A FIPT		nt status, School enrolment, full time/part time		
<ul> <li>Mar 2005</li> <li>Mar 2004</li> </ul>	Person Variables	A_GRSWK		nt job, Earnings, usual weekly amount	-	
• Mar 2003	Person Variables	AHERNITE		nt job, Earnings, usual hourly rate too code flag		
<ul> <li>Mar 2002</li> </ul>	Person Variables	A_HGA	Mar 1992 - current Demog	praphics, Educational attainment		
Mar 2001	Person Variables	A_HRLYV/K	Mar 1992 - current Eurren	nt job, Earnings, paid hourly Y/N	11 (1)	
<ul> <li>Har 2000</li> </ul>	Person Variables	A_HRS1		nt job, Hours actually worked last week, all jobs		
<ul> <li>Mar 1999</li> </ul>	Person Variables	A_HRSPAY		nt job, Earnings, usual hourly rate		
<ul> <li>Mar 1998</li> </ul>	Person Variables	A_HSCOL		nt status, School enroilment, high school/college		
<ul> <li>Mar 1997</li> </ul>	Person Variables	A_LFSR		nt stabus, Labor force stabus recode		
<ul> <li>Mar 1995</li> </ul>	Person Variables	A_LINENO		graphics, Line number, Unique person identifier		
<ul> <li>Mar 1995</li> </ul>	Person Variables Person Variables	A_MARITL A MJIND	Mar 1992 - current Demog			
<ul> <li>Mar 1994</li> </ul>	Person Variables Person Variables	A_MJIND A_MJOCCC	Mar 2003 - current Indus.	SOcc[main job]industry.major groups		
Mar 1993	Person Variables	A NEFLI		SOcu. (main job) occupation, major groups - recode It status, Last worked for pay		
<ul> <li>Mar 1992</li> <li>(a) Public Participation in the Arts</li> </ul>	Person Variables	A PARENT		xaphics, Last worked for pay xaphics, Parent's ine number		
<ol> <li>(a) Public Participation in the Arts</li> </ol>		A PAYABS	I'm xaze - ownen benog	proprious number of the second second		

Once you highlight a variable the Browse / Select Highlighted Variables button becomes active.

### 3. CLICK THE **BROWSE / SELECT HIGHLIGHTED VARIABLES** BUTTON AS SHOWN ABOVE.

*Note: You can also double click on a variable to pop up the Browse Variable window.* 

This is also called the Codebook. This window also allows you to select the variables (and their values) that you want to be put into your Data Basket where you can recode, modify, or delete a variable.



1	CLICK THE BOX MARKED SELECT CPS		-0-
н. А_А	AGE DEMOGRAPHICS, AGE	Image: Select Variable           Browse/Select Variables & Values	<u> </u>
		Your highlighted variables: CPS A_AGE (Mar 2002 ) Demographics, Age	I ,
			OK Cancel
		F Select ALL Variables	
		V Select CP5 A_AGE Demographics, Age	
2.	CHANGE THE AGE RANGE TO <b>0 TO 15</b>	ltem 18d - Age Universe = All	
3.	CLICK OK	0 to 15 Range	Deselect
			Unsort Ranges
		Additional Metadata Tags here	

You have added A\_AGE to your data basket and you have restricted the universe to kids 15 and younger.

Choose your remaining variables from Step 1.

- 1. WITH YOUR LIST STILL SORTED BY NAME, SCROLL DOWN AND HIGHLIGHT THE  $A\_SEX$ VARIABLE.
- 2. Scroll down until you see the  $CH_HI$  variable (child covered by health INSURANCE). HOLD DOWN THE CONTROL KEY AND HIGHLIGHT IT.
- CLICK THE BROWSE / SELECT HIGHLIGHTED VARIABLES BUTTON. 3.

	1 4			Highlight the variables you are interested in	
Search All Datasets	<b>₽</b>			All and a second se	
🕮 🔂 American Community Survey		534	Variables returned fro	m search. 1 variables selected in DataBasket.	alect ted
D American Housing Survey		***			
B- Common Core of Data(Education)	Торіс	Name	Availability	Variable Label	
E- Consumer Expenditure Survey	Person Variables	A_NLFLJ		Current status, Last worked for pay	i and the second se
County Business Patterns	Person Variables	A_PARENT		Demographics, Parent's line number	
E Current Population Survey	Person Variables	A_PAYABS		Current status, With job, not at work, Paid?	
🕀 🖻 Basic 🔤	Person Variables	A_PFREL		Family relationship with primary family	]
(i) m Basic 1989-93	Person Variables	A_SEX		Demographics, Sex	<b>(</b> ''''''''''''''''''''''''''''''''''''
😟 📠 Civic Engagement	Person Variables	A_SPOUSE	Mar 1992 - current	Demographics, Spouse's line number	
🕀 🝙 Contingent Worker	Person Variables	A_UNCOV	Mar 1992 - current	Current job, Earnings, Union contract coverage Y/N	
🕀 💼 Doablity	Person Variables	A_UNMEM	Mar 1992 - current	Current job, Earnings, Union member Y/N	1
Displaced Workers/Job Tenure	Person Variables	A_UNITYPE	Mar 1992 - current	Current status, Unemployment reason	
🕀 💼 Fertility	Person Variables	A_USLFT	Mar 1992 - current	Current status, With job, not at work, usual FT/PT	
🕀 📻 Fertäty and Marital History	Person Variables	A_USLHRS	Mar 1998 - current	Current job, Hours, usually worked at main job	4
D m Food Security	Person Variables	A_WANTJB	Mar 1992 - current	Current status, Want a regular job now	
mmgration/Emigration	Person Variables	A_WERNTF	Mar 1995 - current	Current job, Earnings, usual weekly amount top code flag	
E m Internet and Computer Use	Person Variables	A_WHENL3	Mar 1994 - current	Current status, Unemployed, duration of job search/layoff	
🔄 📻 Library Use	Person Variables	A_WHYABS	Mar 1992 - current	Current status, Absent from work last week, reason	ţ
March Annual Social and Economic Supplemen	Person Variables	A_WKSCH	Mar 1992 - current	Current status, Labor force recode by time worked or lost	
S - A March Supplement	Person Variables	A_WKSLK	Mar 2000 - current	Current status, Unemployment, duration of job search/layoff	
•	Person Variables	A_WKSTAT		Current status, Labor force full/part-time status	
<ul> <li>Mar 2013</li> </ul>	Person Variables	CAID	Mar 1996 - current	Health insurance, Covered by Medicaid-Gov't assistance that p	
Mar 2012	Person Variables	CARE	Mar 1996 - current	Health insurance, Covered by Medicare-health ins. for people	į.
<ul> <li>Mar 2011</li> </ul>	Household Variables	CARE_VAL		Annual amount paid for child care	ĺ
• Mar 2010	Person Variables	CHAMP	Mar 1992 - current	Health Insurance Champus, VA, or military - Person	
Mar 2009	Person Variables	CHELSEW YN		Child living outside the household	-
<ul> <li>Mar 2008</li> </ul>	Person Variables			Annual amt paid for child support	
<ul> <li>Mar 2007</li> </ul>	Person Variables			Person required to pay child support	
<ul> <li>Mar 2006</li> </ul>	Person Variables	CH_HI		Health insurance - child covered Y/N	i
• Mar 2005	Person Variables	CH_MC		Health insurance - child covered by medicaid Y/N	
• Mar 2003	Person Variables	CLWK		Work experience - longest job class of worker recode	
11 III III III III III III III III III	Person Variables	COV GH		Health insurance group coverage, including depend	
		*			1

Confirmation 23 You have added 1 variable for your DataBasket. OK Cancel

In the Ferrett Browse Variable window this time, you will see the two variables you have chosen in the **Your highlighted variables:** box. You can click on either of them to see its' codebook information.

### 4. CLICK THE BOX MARKED SELECT ALL VARIABLES

Note: Click **Select** <variable name> if you want to add only one variable to your Data Basket, click the **Select ALL Variables** box if you want to keep all the variables in your list.

#### 5. CLICK **OK**

You get a message: "You have added 2 variables to your data shopping basket."

Confirma	ition X
Ş	You have added 2 variables for your DataBasket,
	OK Cancel

On Step 1 you can search for a variable through the labels, names, topics, full descriptions and values.

- 1. CHECK THE BOXES OF VARIABLE SEARCH TO INCLUDE LABELS, AND NAMES.
- 2. TYPE THE WORD **RACE** IN THE BOX LOCATED ABOVE **MATCH ANY WORD**.
- 3. CLICK THE **SEARCH** BUTTON.

The list of variables is reduced to those that have the word race in the selected search options. In our example **PRDTRACE** is the only variable returned.

4. HIGHLIGHT **PRDTRACE** - CLICK THE **BROWSE / SELECT HIGHLIGHTED VARIABLES** BUTTON.

i	Browse/Select Variables & Values	
our highlighte	d variables:	
393 A. SEX (457	1992 Dummigrati finas, Sex	Contraction of the second s
CPS CH_HI (Mar	1992 - ) Health insurance - child covered Y/N	ОК
V Select ALL Va	riables	
Select	CPS A_SEX Demographics, Sex	<b>7</b>
ltem 18g - Universe = A	Sex N	
📝 1) Male		Deselect all values
🕑 2) Femal	:	
	Additional Metadata Tags here	

ta elp	
ep1 : Select Dataset & Variable Step2 : DataBa	nstet/Download/Make A Table
a is data in which every record is t of analysis level and all records isded up to get the totals for each . For example, for surveys of its, microdata contain records for widual interviewed; for surveys of hons, the microdata contain preach organization.	1. Variable & F Labels F Names F Topics F Question Text F Values 2. Face 6 match ANY word 6 match ALL words Search 3.
1	Highlight the variables you are interest
· · · · · · · · · · · · · · · · · · ·	1 Variables returned from search. 2 variables set Sunn DataBasket.
Topic Name Availability	Variable Label Mergeable
Parkson werksmitter PRD (RACE MET 2003	current Demographics-race of respondent No

6. CHECK THE BOX TO SELECT CPS PRDTRACE AND CLICK OK.

You have added this variable to your databasket.

Note: If you were searching multiple words, the system default is to search on any of the words. You also have the option to match ALL words.

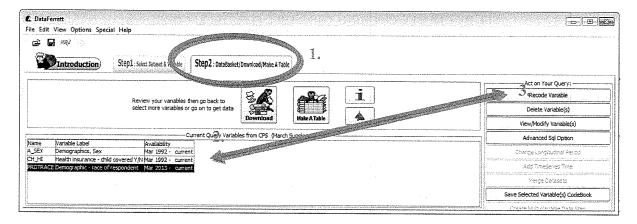
Note: Want to make a change? Simply clear your text from the search and begin another text search.

To get all of the variables back in the Dataset window re-open the Mar 2007 dataset from the dataset list on the left side.

ſ		
	Ferrett Browse Variable	
	Browse/Select Variables & Values	
	Your highlighted variables:	
100000	CPS PUBILIAGE (Mar 2013 - ) Demographic made of respondent	
The second second		ОК
and the second second		Cancel
- Statement		
-	Select A siables	
	Select CPS PRDTRACE Demographic - race of respondent	
	Universe = PMPERTYP = 1.2.0R 3	
	[√] 1) White Only	Deselect
		all values
1	2) Black Only	
	😨 3) American Indian, Alaskan Native Only	
	🕢 4) Asian Only	
	✓ 5) Hawaiian/Pacific Islander Only	
	☑ 5) White-Black	
	7) White-AI	
	Additional Metadata Tags here	
1	Automatina automatica automatica	

### 1. CLICK THE STEP 2: DATA SHOPPING BASKET / DOWNLOAD / MAKE A TABLE TAB

2. HIGHLIGHT THE RACE VARIABLE PRDTRACE



Selecting a variable from your list of Current Query Variables will activate the following options: Recoding, Deleting and Modifying the variables.

### RECODE VARIABLE(S)

3. CLICK ON THE RECODE VARIABLE(S) BUTTON LOCATED AT THE TOP RIGHT SIDE OF THE SCREEN.

			(	i	Reco	de/Reg	JLOND A	aria	bles		Tell me abou	53936 J.	2						
<u> </u>	nialius	in the second	for the Variable Re	code of PRD	TRACE			6	elect		Label	********			25	terrationer w	0.000 000 000 000 000 000 000 000 000 0		<b>27</b> - 412 - 5
Na		Description White Only	And and a second second second	A state of the second second			e de la composition de la comp	1-			White Only	491 11		{1}					
15		Black Only						μ		2	Not Elsewhere	LIASSI	iea (nec.)	ξ2,3	,4,5,	5,/,	,8,9,	10,1	1,1
1			an, Alaskan Native Only				E												
4		Asian Only	,,																
5		Hawaiian/Pacif	îc Islander Only	1				Internation											
6		White-Black		1															
17		White-AI						Contract of											
8		White-Asian					14 C												
12		White-HP					- <del>-</del> -												
		,						-	( <u>(</u> )	din 1		Karana ja jara	anailikaisinka	6.035					
II S∈	et to	value 2	Recode					1	o ann an an ann an Ann				lefine Cur			11270227	deresterter		

- 4. ASSIGN A LABEL TO THE RECODE IN THE BOX LABELED RECODE1 AT THE TOP LEFT.
- 5. HIGHLIGHT THE FIRST CATEGORY, WHITE ONLY AND CLICK THE RECODE BUTTON AT THE BOTTOM LEFT.

Notice the right hand side of the screen displays 2 categories for the new recoded variable:

- White Only
   (1)

   Not elsewhere classified
   (2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17)
- 6. HIGHLIGHT THE **BLACK ONLY** AND CLICK THE **RECODE** BUTTON AT THE BOTTOM LEFT. BLACK ONLY has been assign to value 2 in the recode.

7. CONTINUE BY ASSIGNING THE REMAINING CATEGORIES TO THE FOLLOWING VALUES:

- value <u>category</u>
- 3 American Indian, Alaskan Native Only
- 4 Asian Only 5 Hawaiian Pa
- 5 Hawaiian Pacific Islander Only
- 6 All remaining categories
- 8. TYPE A NEW LABEL BY DOUBLE CLICKING THE LABEL FOR THE RECODED VALUE CALLED **NOT ELSEWHERE CLASSIFIED (NEC.)**. TYPE THE WORD **OTHER** AND MAKE SURE TO HIT THE **ENTER** KEY.

C Ferrett Microdata Recodel			
		Recode/Regroup Variables	Tell me advort
Race Recode is label for the Variable Recode of PRDTRACE			
Highlight the value(s) to recode/regroup	Select		Values
Value Description		White Only	[1]
1 White Only		Black Only	{2}
2 Black Only		American Indian, Alaskan Native Only	
B American Indian, Alaskan Native Only			<u>{</u>
4 Asien Only	<u> </u>	Hawaiian/Pacific Islander Only	{5}
5 Hawakan/Poofic Islander Only	V 6	Other Not Elsewhere Classified (nec.)	<i>{</i> 5,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26}
5 White-Black 7 White-Al			
8 White-Asian			
B White+P			
10 Black-AI	-		
Set to value 6 Recode			
	÷		Repetitie Current Row
		Ok Cancel	······································

9. FINISH BY CLICKING THE OK BUTTON AT THE BOTTOM OF THE SCREEN.

### **Delete Variable(s)**

This function allows you to remove variable(s) from the Data Basket. Highlight the variable(s) you want to delete and click "Delete Variable(s)" button.

Modify Variable(s)

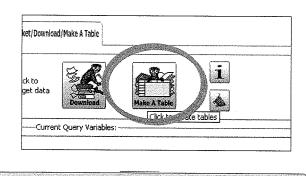
Modify will open the Browse Variables window again and you can change values that were previously selected.

**Advanced SQL Option** 

This function allows you to show the SQL routine that will be used to run your query. Advanced SQL opens the Advanced sql options to change clauses and add conditions in an SQL-like syntax.

Save Selected Variable(s) Codebook

This allows you to save the documentation for the variables in your Data Basket to an ascii text or html file. The documentation includes the variable name, label and value descriptions.



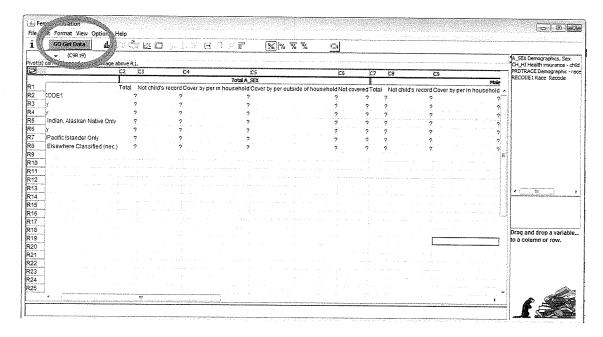
Ferrett T	ab Message
Â.	Making a Table
V	1. Click and highlight to select variable.
	<ol> <li>Hold mouse key to turn cursor into a hand, then drag selected variable to Column 1 or Row 1 on spreadsheet. You can type over any value labels in order to make them more readable.</li> </ol>
	3. Click on GO on the toolbar to get data results.
	For Formulas and other advanced spreadsheet functionality see Help,
	Contents in menu bar.
	ОК

1. CLICK ON THE MAKE A TABLE BUTTON

You will see the following pop-up window. These are directions for making a table. READ AND CLOSE by clicking the X in the upper right corner.

#### MAKE A 3-WAY TABLE.

- 2. DRAG THE RECODED RACE VARIABLE TO THE R2,C1 CELL
- 3. Drag the variable **A\_SEX** to the **R1,C2** cell
- 4. Drag the CH\_HI variable to the R1,C2 cell also



5. CLICK THE GREEN GO GET DATA BUTTON ON THE MENU BAR Note: No need to do anything about a weight variable, it is automatically applied.
6. TO SAVE THE TABLE CLICK FILE SAVE AS ON THE MENU BAR.

29 14 child's record Co 122.413.931 97.615.731 14.383.111 1.195.587 5.634.646 412.885 2.171.971
t child's record Co 122,413,931 97,615,731 14,383,111 1,195,587 6,634,646 412,885
t child's record Co 122,413,931 97,615,731 14,383,111 1,195,587 6,634,646 412,885
122,413,931 97,615,731 14,383,111 1,195,587 6,634,646 412,885
122,413,931 97,615,731 14,383,111 1,195,587 6,634,646 412,885
97,615,731 14,383,111 1,195,587 6,634,646 412,885
14,383,111 1,195,587 6,634,646 412,885
1,195,587 5,534,646 412,885
5,634,646 412,885
412.885

7. MAKE SURE TO SAVE AS (AN HTML, TEXT (TAB DELIMITED OR COMMA DELIMITED), PDF OR FERRETT TABULATION FILE). You can save the layout of your tabulation and all the variables in your databasket for later use.

Save in:	FerrettTa	bulationFile	S	and and a second se	- 2 0 11-		
	Name		Size	Item type	Date modified	Attri	
and the	acs 2014	-foreig	30.3 KB	FTF File	4/1/2016 4:32	A	
Recent Items	acs2013	esr_sex	4.46 KB	FTF File	3/30/2016 9:4	А	
	ACS_SEX	_AGE	4.31 KB	FTF File	3/15/2016 2:2	A	
Millionersen)	ACS_SEX	_AGE	219 KB	FTF File	3/11/2016 2:5	А	
	acspums	AgeL	57.3 KB	FTF File	3/9/2016 5:32	A	
Desktop	🔄 ed assist	ftf	4.39 KB	FTF File	3/8/2016 4:14	А	
	acs-forei	gn_bo	30.8 KB	FTF File	2/29/2016 3:4	А	
ander020							
Computer							
	File name:		Ferrett_Ta	ble			*
	rile name;	CPS Ferre					Save
Network	Files of type:	Sciences and sciences	abulation Fil	es			Cancel
		HTML File:	; iments - Tal				

*Please Note: The menu bar has a* "CLEAR SPREADSHEET" button which will clear all the data from the spreadsheet but will still keep the variables in your databasket (appears on the right). Click the CLEAR SPREADSHEET button to start over dragging variables into the cells of the Ferrett Tabulation.

👍 Ferrett Tabulation					en e
File Edit Format View C	)ptions Help	and the second se			and the second
i cocercas			eît	× %	©.
👻 Formula Bar					
Pivot(s) can be dropped on piv	ot image above R1.	[Lear Spreadsheet]			

The DataFerrett Tabulation screen is a separate screen that can be left open or closed. If you wish to add variables from Step 1, they will always be reflected in the list of variables on your DataFerrett Tabulation screen.

#### Accessing the DataFerrett Software

Go to <u>http://dataferrett.census.gov/</u> and click "Launch DataFerrett" on the right side of the screen to access the launch page.

### **DataFerrett Help Desk**

For technical assistance in using DataFerrett, you can call the help desk toll-free at 1-866-437-0171.

The help desk can assist you with issues concerning the use of the DataFerrett software, but not with questions regarding the subject matter contained in the datasets. The help desk can, however, usually direct you to the appropriate subject matter experts.



### **Example 1 - Joining datasets by a common variable: Creating a single table using multiple datasets**

Other features illustrated: Aggregate data multi-variable recode, computational calculation

Background: Some datasets may have variables whose value sets are defined alike (comparable). Examples include demographic variables (e.g. gender, race, marital status) and geographic variables (e.g. state code, county code). Co-tabulation allows you to tabulate information from the different datasets side by side using the comparable variable as a hinge.

Goals: 1) Estimate the number of people per gas station in each county in Maryland. 2) Add the number of households without vehicles to the table.

### Steps:

Goal 1 Estimate the number of people per gas station in each county in Maryland.

- 1. Select first dataset County Business Patterns, County Level, 2003.
- 2. Select variables Selectable Geographies (all counties in MD), Establishments (est), and industry (naics=447///).
- 3. Select second dataset Decennial Census, Summary File 3, 2000.
  - Go to Step 2: DataBasket/Download/Make a table tab.
  - Highlight the county variable.
  - Press the "Merge Datasets" button on the right-hand side.
  - The merge dialog appears with information that you will only be able to create tables and not extract data. Press the "Next" button. This will bring back a list of datasets with a comparable variable and may take a while to populate.
  - After the results return, enlarge the window by dragging the right side farther to the right.
  - Scroll down the list looking for the name of the second dataset named in step 3. Highlight this dataset and press the "Finish" button. You will get a warning that you will be taken back to step 1 where you will select variables from the second dataset.

Screenshot of merge dialog dataset selection:

You picked the option to merge data in the spreadsheet. These are other datasets that in dataset that you want to pick more variables from:	clude the match variables that you ha	ve picked. Select the
Dataset	Availability	Variable
Decennial Summary File 1 (Summary File 1)	2000 - current	COUNTY
Decennial Summary File 1 (Summary File 1)	1990 - 1990	CNTY
Decennial Summary File 3 (Summary File 3)	1990 - current	CNTY
Decennial Summary File 3 (Summary File 3)	2000 - current	COUNTY
CCD (Public Elementary/Secondary School Universe Data)	2003 - 2003	COUNTY
POPEST Census Bureau Version (Counties by Age and Sex)	1990 - 1999	COUNTY
POPEST Census Bureau Version (Counties by Age and Sex and Hispanic Origin)	1990 - current	COUNTY
POPEST Census Bureau Version (Counties by Age and Sex and Race)	1990 - current	COUNTY
POPEST Census Bureau Version (Counties by Race and Hispanic Origin)	1990 - current	COUNTY
HUD (HUD Test 1)	2000 - current	COUNTY
		•

- 4. Select the variable from Summary File 3 P1. Total Population (P001001).
- 5. Go to Step 2 tab and press the "Make a Table" button. This brings up the DataFerrett spreadsheet window.





- 6. Click and drag the first "MERGED-200 Three digit FIPS County Number" variable into C1,R1 to define the counties in the rows. This is the comparable dimension and all other variables will be dropped in the columns.
- 7. Click and drag the "est" variable into C2,R1.
- 8. Click and drag the "P001001" variable into C3,R1.
- 9. Create a formula calculating people per gas station in column 4 (C4)
  - Highlight column 4 by clicking in the gray column header.
  - In the formula bar above the column headers enter the formula =comp(c3/c2) and press the Enter key.
  - Add a description to the column header: "People per gas station."

Screenshot of co-tabulation table and a formula:

Ferrett Tabulation			_ <b>_ _ _</b> ×
File Edit Format View Options He	up		
GO Get Data			
Col C4 =COMP(C3/C2)			MERGED-202 Three-digit FIPS County N
C1	C2 C3	C4	est Total Number of Establishments
R1	Total Number of Establishments Total	population People per das st	ation Alics NAICS Code
R2 Total MERGED-202	?	?	P001001 Total population
R3 Allegany County, Maryland	?	?	?
R4 Anne Arundel County, Maryland	?	?	2
R5 Baltimore County, Maryland	?	?	?
R6 Calvert County, Maryland	?	?	2
R7 Caroline County, Maryland	?	?	2
R8 Carroll County, Maryland	?	?	2
R9 Cecil County, Maryland	?	?	2
R10 Charles County, Maryland	?	?	2
R11 Dorchester County, Maryland	?	?	2
R12 Frederick County, Maryland	2	?	2
R13 Garrett County, Maryland	2	?	21.1
R14 Harford County, Maryland	2	?	2
R15 Howard County, Maryland	?	?	
R16 Kent County, Maryland	?	?	· · · · · · · · · · · · · · · · · · ·
R17 Montgomery County, Maryland	?	?	2
R18 Prince George's County, Maryland	?	?	
R19 Queen Anne's County, Maryland	?	?	
R20 St. Mary's County, Maryland	?	?	Drag and drop a variable(s)
R21 Somerset County, Maryland	?	?	to a column or row.
R22 Talbot County, Maryland	2	2	2
R23 Washington County, Maryland	?	2	2
R24 Wicomico County, Maryland	2	2	2
R25 Worcester County, Maryland	2	?	2
R26 Baltimore city, Maryland	?	?	2
R27			
000 1			
4			

10. Press the green "Go Get Data" button in the toolbar.

Goal 2 Add the number of households without vehicles to the table.

At this point, we can explore how many households have no vehicle within each county.

- 1. DO NOT CLOSE the spreadsheet window, but go back to your main DataFerrett window using your Windows taskbar.
- 2. Go to the Step 1 tab so that you can add an additional variable from SF3.
- 3. Scroll down the dataset list under SF3 until you find table "H44. Tenure by vehicles available", and view all of the variables in this table.
- 4. Select the "D\_H044\_1 Total Housing Units: No Vehicle Available" variable. This is a pre-defined recode that sums together the H044003 and H044010 variables.
- 5. Now go back to the tabulation window from your taskbar.
- 6. The added variable is there, now drag it into column 5 (C5,R1) and press the "Go Get Data" button.
- 7. If this summed variable had not already been defined, we could have created it ourselves.





- DO NOT CLOSE the spreadsheet window, but go to your main DataFerrett window using your Windows taskbar.
- Select the H044003 and H044010 variables.
- Go to the Step 2 tab and highlight the H044003 and H044010 variables, then pre the "Create Multi-variable Data Step" button on the right-hand side. This brings up the multi-variable recode dialog window.
- Make sure you give your new variable a good label in the text field at the upper left.
- Enter the formula if  $(1 == 1) \{D_RCD1 = H044003 + H044010;\}$  in the text area as shown below.

Note: the new variable name in the formula, D\_RCD1, is the dynamic variable name shown at the top of the window and your formula must use the name shown.

The "if (1 = 1)" syntax is used to force the condition to be true in all cases.

The syntax of the formula is like programmatic code and must follow this: if (*condition*) { *action*;}

Screenshot of multi-variable recode of aggregate variables:

My no vehicle is the label for the dynamic variable D_RCD Assign labels to values	1	X
if ( 1 == 1 ) {D_RCD1 = H044003 + H044010;}	Name	Variable Label
10 P	H044003	Total: Owner occupied: No v
	H044010	Total: Renter occupied: No v
	🖉 🗖 Display Values in Va	riable List
		OK Cancel

- Press the "OK" button to save the created variable.
- 8. Now go back to the tabulation window from your taskbar.
- 9. The newly created variable is there, now drag it into column 6 (C6,R1) and press the "Go Get Data" button.

Note: The numbers in columns 6 and 5 should be identical.



### Example 2 - Multi-variable recode for microdata

Other features illustrated: Using percent buttons, sorting on a column, creating a time series graph from a table cell(s), creating a table that averages across time, creating a table that shows data over time, turning off/on Automatic Totals, cross-variable explosion, nesting vs. not nesting

Background: In microdata datasets, you can create new variables with values based on the combinations of multiple variables' values.

Goals: 1) Determine the percentage of people with and without health care coverage by age groups. 2) Examine health insurance coverage numbers over time.

### Steps:

Goal 1 Determine the percentage of people with and without health care coverage by age groups.

- 1. Open the Ferrett Session File (or saved data basket) named <u>cps\_mar\_healthins\_multi\_databasket.fsf</u>.
- 2. Go to the Step 2 tab to view the variables in the data basket. There are five health insurance variables that are required to determine if someone has some sort of health insurance coverage:
  - CHAMP Health Insurance Champus, VA, or military Person
  - COV\_HI Health insurance Group coverage Person
  - COV\_GH Health insurance group coverage, including depend
  - o MCAID Health Insurance, Medicaid coverage Y/N
  - o MCARE Health Insurance, Medicare coverage Y/N

For a person to have any coverage, the answer to any one of these must be "Yes" and for a person to NOT have coverage, the answer to ALL of these must be "No." We can create a multi-variable recode that does this for us so that we have one new variable that shows whether a person has coverage or not.

3. Highlight the "D\_RCD2 Health Insurance Coverage" variable and click on the "Modify" button on the right-hand side. This brings up the multi-variable recode window and shows us the definition of this variable. See the image below:





Screenshot of multi-variable recode for microdata variables:

		r MCAID == 1 or MCARE =	Name	Value	Label
= 1 or COV_GH == 1) {			CHAMP		Health Insurance Champus, VA, or military - I
	_	== 2 and MCAID == 2 and		1	Yes
MCARE == 2 and COV_G	H ==2) (D_	RCD2 = 2;}		2	No
			COV_HI		Health insurance - Group coverage - Person
				1	Yes
			HOUD	2	No
			MCAID		Health Insurance, Medicaid coverage Y/N Yes
				2	No
			MCARE	-	Health Insurance, Medicare coverage Y/N
			4		Fieddriffodiance, medicale corerade 174
		-	Displa	iy Value	es in Variable List
Add a new value	1	Label			
Aug a new value	1	Covered by health insurance	9	_	
	2	Not covered by health insura	nce		

- 4. Close the multi-variable recode window by pressing the "Cancel" button at the bottom right.
- 5. We will now create a new multi-variable recode defined the same way.
  - Highlight the five health insurance variables listed above (ctrl-mouse click combination), then
    press the "Create Multi-variable Data Step" button on the right-hand side. This brings up the
    multi-variable recode dialog window.
  - Give your new variable a label in the text field at the upper left, something like "Health insurance coverage Y/N".
  - Our new variable will have two values 1=Yes, covered; and 2=No, not covered and these are defined by the "if" and "else if" statements in the formulas.
  - To view the values of each variable, check the "Display Values in Variable List" box on the right side below the list of variables.
  - Enter the formula in the text area exactly as shown in the image above, except you MUST change the D\_RCD2 to the name of your new variable shown at the top where you gave it a label.
  - To assign value labels to your new variable's values, check the "Assign labels to values" box directly above the text area. This will then reveal an area at the bottom for defining value labels.
  - Click in the "label" area for value 1 and enter "Covered by health insurance".
  - Click the "Add a new value" button to get a line for value 2.
  - Click in the "label" area for value 2 and enter "NOT covered by health insurance".
  - o Click the "Ok" button. You now have a new variable in your data basket.
- 6. Press the "Make a Table" button. This brings up the DataFerrett spreadsheet window.
- 7. Drop the "RECODE3 Age groups" variable into column 1 to define the rows.
- 8. Drop our new dynamic recode into C2,R1 to define the columns and press the "Go Get Data" button.





- 9. Show the percentages of each age group covered and not covered by clicking on the "Show % of first data column" button in the toolbar.
- 10. Sort the results by column 4 not covered by health insurance to see which age groups have the highest and lowest percentages.
- 11. Turn off sorting and turn off the percentages.

Goal 2 Examine health insurance coverage numbers over time.

- 1. Create a time series graph highlight four cells in column 4 not covered, 4 age groups then press the Time series Graph button in the toolbar. This automatically creates a time series graph of the same measure going back in time as far as possible, up to 12 time periods.
- 2. Show the same table, but show the average over 5 years, and add the SEX variable to the table.
  - DO NOT CLOSE the spreadsheet window, but go to your main DataFerrett window using your Windows taskbar.
  - Go to the Step 1 tab and notice that there are several years of the March supplement available and currently just Mar 2007 is highlighted. Highlight all available years using the ctrl-click or shift-click combinations.
  - Now go back to the tabulation window from your taskbar.
  - Drag the SEX variable into column 1 to break out each age group by sex.
  - Press the "Go Get Data" button. By looking at the universe and dataset information in the spreadsheet (blue text in the middle of the right-hand side), you can see how many years are listed for the "Dataset(s) selected".
  - Go to the "Options" menu and you will see that "Average across time" is selected. You could change it to "Cumulative" if you wanted to see the cumulative totals over all years.
- 3. Create a table that shows the health insurance coverage for EACH of the available years.
  - Close the current spreadsheet window.
  - From the Step 2 tab, press the "Make a table" button.
  - Notice that there is now a variable named "Instances" at the top of your variable list in the spreadsheet. This variable appears when you have more than one instance highlighted in the dataset list and you open a spreadsheet window.
  - Drag your multi-variable recode into column 1 to define the rows.
  - Turn off the automatic totals go to "Options" and click on the "Automatic Totals" to de-select that option.
  - Drag the Instances variable into C2,R1 to show the years in the columns. There is no total for all years because we turned the automatic totals off before dropping this variable.
  - "Go Get Data".
  - Highlight the data cells in row 4, not covered by health insurance, and DO NOT include the label cell (C1,R4).
  - Press the "Graph" button in the toolbar (looks like a bar graph). The initial graph window that opens will show a bar graph by default.
  - Using the drop-down list at the bottom left, select "Line Graph" to essentially create a different time series graph.
  - Since the most recent time is on the left side, you can change it to the opposite by selecting the "Edit" menu and selecting "Reverse Axis".

### Additional Table Manipulation

- Close the spreadsheet window.
- In the Step 1 tab, select only Mar 2007.





- Step 2, "Make a Table"
- Health insurance recode in column1.
- Age recode in C2,R1 to define columns.
- Nest SEX in the rows on top of health insurance.
- Turn off "Automatic Totals".
- Drop the race variable in the rows, but not nested drop two lines below current row definitions.
- Turn on "Automatic Totals".
- Turn off "Cross variable explosion" and nest the Hispanic variable on the race wpresse only category.



### **Example 3 - Mapping address points on a thematic map**

Other features illustrated: Viewing underlying records, clearing one dimension, changing universe from the spreadsheet

Background: If a dataset has correctly-defined address point information, points can be mapped on top of a thematic map from a separate dataset.

Goals: 1) Show the locations of all the public schools in DC on top of a thematic map showing the median household income by tracts from SF3. 2) Show how you can view the underlying records from the spreadsheet to see all the information for the universe of the selected cell(s).

### Steps

*Goal 1* Show the locations of all the public schools in DC on top of a thematic map showing the median household income by tracts from SF3.

- 1. Create our underlying thematic map on which we will lay our school locations. Open the Ferrett Session File (or saved data basket) named <u>point\_map\_theme\_sf3\_median\_income.fsf</u>.
- 2. Go to the Step 2 tab to see the variables in the data basket median household income and tracts (all tracts in DC).
- 3. "Make a Table"
- 4. Drag the tract variable to column 1 to define the rows.
- 5. Drag the income variable to C2, R1 and "Go Get Data".
- 6. Highlight the data cells in column 2 and click on the "Map" button in the toolbar. The map window showing the income by tract opens. DO NOT CLOSE the map window.
- 7. Go back to the spreadsheet window and CLOSE it.
- 8. Go back to the Step 1 tab.
- 9. Clear the data basket by clicking on the "Empty Data Basket" button on the right side.
- 10. Open the Ferrett Session File (or saved data basket) named point\_ccd\_schools.fsf .
- 11. Go to the Step 2 tab to see the variables in the data basket county (DC) and some information about the public schools, including address, phone, number of students, teachers, and level of the school.
- 12. "Make a Table"
- 13. Drag the county variable to column 1 to define the rows.
- 14. Drag the "MEMBER Total students..." variable to C2, R1 and "Go Get Data".
- 15. Highlight the C2,R3 data cell.
- 16. Click on the "Point Map" button in the toolbar (US with a flag on it).
- 17. After a few seconds, a dialog box appears asking if you want to create a new map or add to an existing map. Select the "Add to existing map" option and the theme map we created is shown and highlighted. Click the "Ok" button.

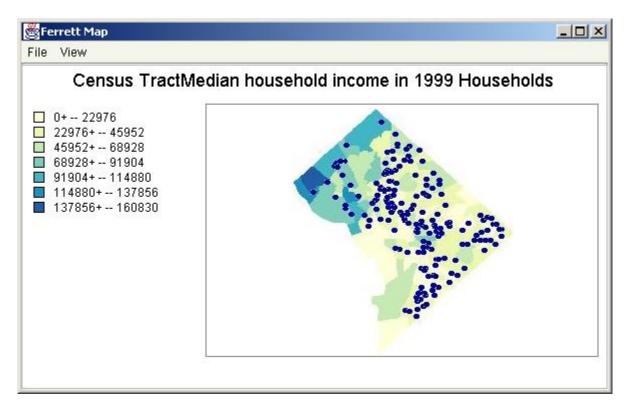


Screenshot of the dialog box for adding point to a theme map:

<u></u>			×
A DataFerrett map alre	ady exists.		
C Create a new map.			
Add data to one of the	following r	maps	
Highlight the DataFerr	ett map tha	at you wish to add this information to	5
Census TractMedian ho	usehold inc	come in 1999 Households	
	Ok	Cancel	

- 18. Find your map window in your taskbar and open it.
- 19. In the map window's "View" menu, select "Layers", and then from that select "Red Dot". The school location points will appear on your map. These dots represent the schools for those students in the selected table cell.

Screenshot of point map showing schools on median income theme in DC:



*Goal 2:* Show how you can view the underlying records from the spreadsheet to see all the information for the universe of the selected cell(s).

- 1. Back in your spreadsheet window, clear ONLY the column dimension go to the "Edit" menu, select "Clear", then select "All Columns".
- 2. Drag the LEVEL variable into C2, R1 to re-define the columns, then "Go Get Data".





- 3. Highlight the data cells in R3, C4 and C5, middle and high schools.
- 4. Click on the "View Underlying Records" button in the toolbar (looks like a spreadsheet with a magnifying glass on it). This brings up a new window showing all the variables for the records with the values highlighted. The output can be sorted by any variable column by clicking in the header.
- 5. Close the underlying record window.

### Additional Manipulations - Changing Universes in the Spreadsheet

- 1. You can change your table to show some other set of counties.
- 2. Clear your row dimension "Edit" menu "Clear" "All Rows".
- 3. Double click on the County variable in your spreadsheet variable list on the right side. This will bring up the Geography Codebook window.
- 4. Highlight "District of Columbia" shown on the right side and press the "Delete" button.
- 5. Highlight the FIPS county code on the left side and then press the "Next" button.
- 6. Highlight the FIPS state code on the left side and press the "Next" button.
- 7. Select "California" and press the "Next" button. This moves California to the right side.
- 8. Press the "Next" button again. This will show all the counties in the state.
- 9. Highlight the "Select All" at the top of the list and press the "Next" button. This moves all counties to the right side.
- 10. Press the "Finish" button. You should return to your spreadsheet window.
- 11. Drag the county variable to C1 to define the rows with all the counties in California. Press the "Go Get Data" button.



### **Example 4 - Spreadsheet functions: Ranking, weighted ranking, order of formula processing.**

Other features illustrated: Summing calculation, hiding columns, creating column spanners, viewing hidden columns, IF conditions

Background: This example shows how you can use ranking and apply weights to those ranks in order to answer questions like "What is the best county for ...?"

Goals: 1) Create a table that will rank the best county in the US in which to live, based on our three criteria - commute time, median household income, and median housing value. 2) Limit our rankings to counties that meet our population size criteria - between 50,000 and 150,000 people.

### Steps

*Goal 1* Create a table that will rank the best county in the US in which to live, based on our three criteria - commute time, median household income, and median housing value.

- 1. Open the Ferrett Session File (or saved data basket) named sf3\_bestcounty\_tolive.fsf.
- 2. Go to the Step 2 tab to see the variables in the data basket number of workers not working at home, aggregate travel time to work, median household income, median value of owner-occupied housing units, and county (all counties in US).
- 3. "Make a Table"
- 4. Turn off the "Automatic Totals" (in the "Options" menu) and drag the county variable to column 1 to define the rows.
- 5. Drag the aggregate travel time variable, P033001, to column 2 C2, R1.
- 6. Drag the workers variable, P031002, to column 3 C3, R1.
- 7. Skip two columns and drag the median household income variable, P053001, to column 6 C6, R1.
- 8. Skip 2 columns and drag the median housing unit value variable, P085001, to column 9 C9, R1.
- 9. "Go Get Data"

Create calculations and rankings.

- 1. Calculate the average commute time in column 4.
  - Click in the C4 column header to select the column in which to create the calculation.
  - In the formula bar, enter =comp(c2/c3) == aggregate travel time divided by workers not working at home equals average travel time.
- 2. Create the county ranking for commute time to work. The lower the commute time, the higher the rank number.
  - Click the C5 column header to select the column in which to create the ranking.
  - In the formula bar, enter = rank(c4)
- 3. Create the county ranking for median household income. In this case, we want the higher value to have the higher rank number. Therefore, we will need to create an inverse income column on which to create the ranking.
  - Click the C7 column header to select the column in which to create the calculation for inverting the income.
  - In the formula bar, enter =comp(0-c6)
  - Click the C8 column header to select the column in which to create the ranking.
  - In the formula bar, enter =**rank(c7)**





- 4. Create the county ranking for median housing unit value. The lower the median value, the higher the rank number. If for some reason you were to want to rank the counties higher based on HIGH housing values, then you would need to create an inverted housing value column like we did for income.
  - Click the C10 column header to select the column in which to create the ranking.
  - In the formula bar, enter =**rank(c9**)

### Hide Columns

Since several of our columns were used to create our desired columns, we can hide them. In our example, we do not need to see the columns used to calculate the commute time or the inverted income column used for ranking income.

- 1. Click the C2 column header to select the column you wish to hide, then select the "Edit" menu, then select "Hide".
- 2. Repeat for columns 3 and 7. Currently you can only highlight one column at a time.
- 3. If you ever want to see the hidden columns, select the "View" menu, then select "Hidden Rows/Columns".
- 4. If you ever want to change a hidden column back to not being hidden (unhide), first view the hidden columns as in the previous step, then highlight the column, then select the "Edit" menu and select "Unhide". The "Unhide" selection allows you to unhide only the selected column, or unhide all hidden columns.

### Create Custom Column Spanners

You can create your own custom column spanners to group columns together. In our example, we can create spanners to group the three different measures and their rankings together.

1. Select the "Edit" menu, then select "Column Spanners". This opens the "Column Spanner Editing Dialog" box. There are three fields for you to supply - 1) the text for the spanner, 2) the beginning column, and 3) the ending column. See the image.

Screenshot of column spanner creation:

👸 Column	Spanner Edit	ing Dialog	1		×
Column Sp	anner	From	To		
Commuting	2	5			
Median Hou	ne 6	8			
Median Hou	I 9	10			
		0	-		
Insert	Delete	Cance	el	Preview	ок

- 2. Press the "Insert" button at the bottom to enter the first spanner.
- 3. Click in the empty "Column Spanner" field and type in your first spanner **Commuting** then press the "Enter" key.
- 4. Click in the empty "From" column and enter the column number the spanner should start with 2 then press the "Enter" key.
- 5. Click in the empty "To" column and enter the column number where the spanner should end 5 then press the "Enter" key.
- 6. Press the "Insert" button at the bottom to enter another spanner.





- 7. Enter the information as show in the image above.
- 8. Once the three spanners have been entered, press the "OK" button.

### Create the Combined Ranking

Now we want to sum all three rankings together to find which counties rank the best based on the three criteria. The summed rankings will show us which counties have the shortest commute times, the highest median income, and the lowest median housing values.

- 1. Click in the C11 column header to select the column in which to create the summed ranking.
- 2. In the formula bar, enter =sum(c5,c8,c10) == this sums the rankings together.
- 3. A small dialog box will appear showing the computations and the last one shows a conflict. See the image.

Col         5         =RANK(C4)         rank           Col         7         =COMP(0-C6)           Col         8         =RANK(C7)         rank           Col         10         =RANK(C9)         rank		Col			commute ti
D         Col         7         =COMP(0-C6)           D         Col         8         =RANK(C7)         rank           D         Col         10         =RANK(C9)         rank			5	DALU/CO C	
Col         8         =RANK(C7)         rank           D         Col         10         =RANK(C9)         rank		Col		=RANK(C4)	rank
Col 10 =RANK(C9) rank		001	7	=COMP(0-C6)	
		Col	8	=RANK(C7)	rank
D         Conflict         Col         11         =SUM(C5,C		Col	10	=RANK(C9)	rank
	nflict	Col	11	=SUM(C5,C	0
	- 1				
		nflict	nflict Col	nflict Col 11	nflict Col 11 =SUM(C5,C

Screenshot of calculation sequence conflict dialog:

The computations must happen in order, or they cannot be completed properly. In this case, all previous ranking functions must be completed prior to our summing the ranks together, therefore we want to change the conflicting computation's sequence.

- 4. Click in the "Sequence" cell of the "Conflict" row and change the 0 to a 2 and press the "Enter" key. This will make this computation occur last.
- 5. Press the "OK" button.

To see which county ranks the highest or lowest, highlight column 11 and use the sort buttons in the toolbar.

### Adding Weights to the Ranking Criteria

We can customize our final ranking by weighting the criteria that are of more importance to us. For example, if the commute time is the most important criterion to us, we can weight it higher than the others. The weights must add up to 1. We will weight commute time as 0.5, and the other two as 0.25 each.



- 1. Unsort the table by pressting the "Return to original order" button next to the sort buttons on the toolbar.
- 2. Click in the C12 column header to select the column in which to create the weighted ranking.
- 3. In the formula bar, enter =comp((c5\*0.5)+(c8\*0.25)+(c10\*0.25)) == this weights the ranks and sums them together.
- 4. We notice that we get all 0s in our table. We have a sequence conflict again, but we were not notified this time. We can change the sequence manually.
- 5. Go to the "Options" menu and select the "Specify Sequence" option. This will bring up the dialog we saw before. Now, we want to change the last two calculations to have a sequence of 2, making sure to press the "Enter" key after each change. See the image.

Screenshot of calculation sequence specification dialog:

Sequence	Conflict		Number	Definition	Label
)		Col	4	=COMP(C2/	commute ti
)		Col	5	=RANK(C4)	rank
)		Col	7	=COMP(0-C6)	0
)		Col	8	=RANK(C7)	rank
)		Col	10	=RANK(C9)	rank
2		Col	11	=SUM(C5,C	Ranking
2		Col	12	=COMP((C5	
					1

Once the sequence is changed, the weighted rankings appear. You can easily change these weighted rankings by changing the weights in the formula for column 12. To edit a formula, simply highlight the column and the formula should appear in the formula bar. Edit the formula and press the "Enter" key and the changes will be applied.

You can sort the table to find the best and worst counties based on your weighted ranks.

*Goal 2* Limit our rankings to counties that meet our population size criteria - between 50,000 and 150,000 people.

If we only want to see the best counties of a certain size, we can include the population variable and then use an IF condition in a formula to remove counties from the top of the list.

- 1. DO NOT CLOSE your spreadsheet window.
- 2. Go back to the Step 1 tab and find the P1. Total Population table in Summary File 3.
- 3. Get the variable, and add the P001001 variable to your data basket.
- 4. Now go back to the spreadsheet window and you will see the P001001 variable in your variable list.
- 5. Drag P001001 into column 13. Press the "Go Get Data" button.
- 6. Click in the C14 column header to select the column in which to create the weighted ranking for counties of a certain size.





- 7. In the formula bar, enter =IF(((C13>50000)AND(C13<250001)),C12,3500) == this condition will use the weighted rank (C12) if the county population is between 50,001 and 250,000 and if the population is smaller or greater than that, it will set the rank to 3500 (an arbitrary number higher than the highest county rank).</p>
- 8. We notice that we get all 0s and 3500s in our table. We have a sequence conflict again, and again we were not notified. We must change the sequence manually.
- 9. Go to the "Options" menu and select the "Specify Sequence" option. This will bring up the dialog we saw before. Now, we want to change the last three calculations two to have a sequence of 2, and the last one to a sequence of 3, making sure to press the "Enter" key after each change. See the image.

Sequence	Conflict		Number	Definition	Label
0		Col	4	=COMP(C2/	commute ti
0		Col	5	=RANK(C4)	rank
0		Col	7	=COMP(0-C6)	
0		Col	8	=RANK(C7)	rank
0		Col	10	=RANK(C9)	rank
2		Col	11	=SUM(C5,C	Ranking
2		Col	12	=COMP((C5	Weighted
3		Col	14	=IF(((C13>5	

Screenshot of calculation sequence specification dialog:

Once the sequence is changed, we now can sort the list and see the best counties based on the weighted criteria we used and limiting the counties to the size we specified.





### Example 5 - List mode: show a list of only those records meeting a set criterion

Other features illustrated: Advanced SQL

Background: You can have your table only show results for the records that meet your criteria. For example, if you want to see all the counties within a select number of states that have a population within a certain size, you can get those listed in a table and include additional information from the dataset.

Goals: 1) Create a listing of all the counties in Alabama, Georgia, and Florida that have a population of less than 75,000 people. 2) Show the actual population estimate and the median household income for those counties.

### **Steps**

*Goal 1* Create a listing of all the counties in Alabama, Georgia, and Florida that have a population of less than 75,000 people.

- 1. Open the Ferrett Session File (or saved data basket) named sf3\_bestcounty\_tolive.fsf.
- 2. In the data basket, delete the FIPS County variable
- 3. Go to the step 1 tab. You will notice that there is a required variable Selectable Geographies.
- 4. Double click on this variable. This brings up the geography selection window. Select all the counties in Alabama, Florida, and Georgia.
- 5. In the step 1 tab, go to the "Geographic Traits" topic listed under Summary File 3 and view all the variables.
- 6. Sort the list by variable name and select the "NAME" and the "STUSAB" variables. These two variables will give us geographic names for our list.
- 7. Select the "P001001" variable from the P1. Total Population topic in SF3.
- 8. Go to the step 2 tab in order to set our population limit.
  - Click on the "Advanced Sql Option" button on the right side. This brings up the Advanced Sql Option window. See the image.

DATAFERRETT

Screenshot of Advanced Sql Option window:

🚰 Ferrett Advanced Sql	×
Advanced Sql Option	
You may Use the text area below to:	
- change the generated sql where clause	
- Add additional sql conditions	
***NOTE***	
Any additional conditions MUST be keyed in standard sql syntax.	
((SUMLEV = 50 AND STATE in (01,12,13)) and (P001001 between 0 and 75000))	
Ok Cancel	

- Insert text as shown in the highlighted portion of the image. This will limit the counties in our list to those with a population under 75,000.
- Press the "Ok" button.
- 9. "Make a Table" and close the green-bordered instruction window.
- 10. Go to the "Options" menu and select "List Mode". When creating a table in "List Mode", you will always leave column 1 empty.
- 11. Drag the "NAME" variable to column 2 (C2, R1).
- 12. Drag the "STUSAB" variable to column 3 (C3, R1).
- 13. "Go Get Data" and your result is a list of the county names and state abbreviations for all of the counties in the selected three states that have a population less than 75,000.

Goal 2 Show the actual population estimate and the median household income for those counties.

We can also view other variables in our list, such as the population and median household income.

- 1. Turn off the List Mode "Options" menu, select "List Mode".
- 2. Drag the P001001 (population) variable into column 4 (C4,R1).
- 3. Drag the P053001 (median income) variable into column 5 (C5,R1).
- 4. "Go Get Data"





### Example 6 - Using a single cell as part of a formula instead of the entire row or column

Other features illustrated: Column number formatting (decimals shown)

Background: The DataFerrett spreadsheet typically does row and column calculations since calculations normally apply to all cells in the row or column equally. However, sometimes you want to create a calculation that uses one specific cell as part of the formula and not the entire column or row.

Goal: Create a table that calculates the percentage of drivers within commute time ranges for a set of counties, while still showing the numbers of drivers within each commute time.

### Steps:

- 1. Open the Ferrett Tabulation File named <u>edw\_trans\_time\_sf3\_cell\_calculation.ftf</u>. Press the "OK" button on the Ferrett Chosen Instance Warning dialog.
- 2. "Go Get Data" to see the tabulation.
- 3. Column 2 was created using a formula that calculates the percent of commuters by travel time ranges. The formula in this case does not use Row 2 as the denominator, but instead just uses cell C3, R2 designated as **\$C3\$R2** see the image for the entire formula.



Screenshot of table with a formula using a cell:



		Options	Help		8		10000
i.	GO Get Data	日日		24 X 🛇 🎹 🂢	5		薔
ol C2	=COMP(C3/\$C3\$R2*1)	00)				P031002 Total: Did not v	
	C1	C2	C3	C4	C5	P031003 Total: Did not v	
₹1	1	Percent	Region Total	Cook County, Illinois	DuPage 🔺	P031004 Total: Did not v P031005 Total: Did not v	0.7000.005
२2	Did not work at home:	100.0	2,763,118	2,310,344		P031005 Total: Did not v	0.7003.005
23	Less than 5 minutes	1.7	47,797	37,573		P031007 Total: Did not v	10100100
R4	5 to 9 minutes	6.7	186,269	146,253		P031008 Total: Did not v	1.1.201.221
₹5	10 to 14 minutes	10.2	281,421	222,654		P031009 Total: Did not v	
86	15 to 19 minutes	11.8	325,414	265,222		P031010 Total: Did not v	vork at
87	20 to 24 minutes	12.5	346,375	287,461		P031011 Total: Did not v	
85	25 to 29 minutes	5.5	150,707	124,438		P031012 Total: Did not v	vork at _
9	30 to 34 minutes	16.3	449,335	388,344		D021012 Total: Did notu	unels at m
210	35 to 39 minutes	3.4	94,805	79,596		Universe: (((SUMLEV	- 50 4
211	40 to 44 minutes	5.5	152,200	130,193		Weight used: Unweight	- Contraction of the second
12	45 to 59 minutes	12.6	348,263	302,459		DataSet(s) selected:	2012 C 10 10 10 10 10 10 10 10 10 10 10 10 10
13	60 to 89 minute	10.3	284,864	242,389		A Sector Sector	
214	90 or more minutes	3.5	95,668	83,762			
15							
16							
17	2						
18							
19					*		
	4				F.		

- 4. We can create a new column and replicate this formula.
- 5. Highlight C6 by clicking in the gray column header.
- 6. In the formula bar, enter the formula as- =COMP(C3/(C3)(C3)(C3)) -then press the "Enter" key.
- 7. Highlight the column again, then select the "Decimal Format" button or the "Format" menu, then "Decimal" to bring up the decimal formatting dialog.
- 8. Select the "One decimal place" option and press the "Ok" button.
- 9. The percents in our new column (C6) should match the percents in column 2.

By creating a formula using a single cell as the denominator, we are able to create a column showing percents while also maintaining all the other data columns.





### Example 7 - Calculating medians (or any percentile) on-the-fly

Other features illustrated: Column formulas and IF conditions

Background: The DataFerrett spreadsheet allows you to calculate medians (or any percentile) on-the-fly by creating your own bins (microdata) or utilizing income distribution counts (aggregate data), and using standard formulas to calculate the linear percentile you have defined within a given distribution.

Goal: Create a table that calculates the median income for age groups (for the population 15 and over) using the March CPS supplement (Annual Social and Economic Supplement).

### Steps:

- 1. Open the Ferrett Tabulation File named <u>cps\_income\_percentile+mean.ftf</u>.
- 2. "Go Get Data" to see the tabulation.
- 3. This tabulation calculates the median (the 50th percentile) which is defined in column 2 (C2) as 0.5. This value can be changed to any percentile you want (number greater than 0 and less than 1, e.g 0.25 for first quartile). To change the value, click in the gray C2 column header. This will show the formula in the formula bar. Edit the formula.
- 4. This table uses the 2006 March file (which reports income for 2005). To calculate for a different year, leave the spreadsheet window open and go to Step 1 in the DataFerrett window to select a different year. Then go back to the spreadsheet window and "Go Get Data".
- 5. You can also calculate for different ages or other variables. To do so, you either select new variables or create new recodes (while the spreadsheet window is open). New variables or recodes will become available in the spreadsheet window. Then, go to "Edit Clear > All rows" to clear the existing age variable. Then drop your variable(s) into column 1 and "Go Get Data".
- 6. This table's universe eliminates income values of zero through the Advanced SQL Option in the Step 2 tab of DataFerrett. If you wanted to include these values, or if you wanted to determine the medians for only positive income (current table includes negative income), you can do so in the Advanced SQL Option. The existing SQL universe is

(AGE1 in (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17) and ((PTOTVAL between -999999 and -1) or (PTOTVAL between 1 and 999999)))

If you wanted to include the zero PTOTVAL values, you could simply remove everything related to PTOTVAL like:

(AGE1 in (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17))

Or if you wanted to use only positive income values, change the universe to: (AGE1 in (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17) and (PTOTVAL between 1 and 999999))

**NOTE:** These percentile calculations are not statistically accurate and will not match published numbers exactly, although they should be within the confidence interval.