Census 2000 Public Use Microdata Sample (PUMS) Files

Jerry Wong
Information Services Specialist
U.S. Bureau of the Census
Los Angeles Regional Office

USCENSUSBUREAU ----

What are Microdata?

- Individual records which contain information collected about each person and housing unit
- They are used to produce the summary data that go into various reports, summary files, and special tabulations
- The PUMS files are extracts from the long form confidential microdata taken in a manner that avoids disclosure of information about households or individuals

Public Use Microdata File Types

	1% Sample File	5% Sample File
Data Detail	Most detail in PUMS files	Some detail filtered out for confidentiality reasons
Geography	400,000 population threshold Super- Public Use Microdata Areas (PUMAs)	100,000 population threshold Public Use Microdata Areas (PUMAs)
Records	2.8 million people Over 1 million housing units	14 million people over 5 million housing units

Confidentiality of the Data

Confidentiality is protected, in part, by the use of the following processes:

- Data swapping (exchanging selected characteristics for a sample of cases)
- Top-coding (all cases at or above a certain percentage of the distribution are placed in a single category)

Confidentiality of the Data

- Geographic population thresholds (no disclosure of data for geographic units with a population below a specified level)
- Age perturbation (age of household members is modified for households containing 10 people or more)
- Collapsing of categorical variables (detail is collapsed if categories do not meet a specified minimum threshold)

USCENSUSBUREAU----

_

PUMS Files vs. Standard Data Products

- PUMS files allow the user to create tabulations that are not available in standard products
 - For example, tables that cross single years of age or user-defined age groups with other characteristics

Limitations of the PUMS files

- Two independently drawn samples of the full census sample (1 percent sample and 5 percent sample)
- Limited geography
 - The smallest unit for the 1 percent files is the Super-PUMA which contains a population of 400,000 or more
 - The smallest unit for the 5 percent files is the PUMA which contains a population of 100,000 or more

USCENSUSBUREAU ----

7

Limitations of the PUMS Files

- Continuous variables (age, income, etc.) are topcoded in order to protect the confidentiality of the data
 - For example, age is topcoded at 90. All individuals with ages at or above the topcode receive the state mean of topcoded age. In the case of Alabama for the 5 percent file, everyone age 90 and above is shown as being age 93.

Limitations of the PUMS Files

- 1 percent files all values for the variables are shown (Except race and Hispanic Origin categories meet a national population threshold of 8,000)
- 5 percent files groups within categorical variables meet a minimum national population threshold of 10,000
- Information on how to combine the two samples is found in the technical documentation

USCENSUSBUREAU ----

_

File Structure

These are state files (includes Puerto Rico)

- Beyond 20/20 software aggregates the data for the 1 percent files and presents a "total" for the U.S. The 5 percent files are state files only and do not include a "total" for the U.S.
- Users of the ASCII version of the 1 percent and 5 percent files must use their own software to aggregate the state data and produce a "national" number
- Geographic equivalency files and maps for each state are online and on the DVD

2 Record Types

Housing unit record

- Includes all of the housing variables such as acreage, annual cost of electricity, property value, and gross rent, as well as many others
- It also includes household variables such as household type, number of people 65 years and over in the household, number of related children under 18 years in the household, and household total income in 1999

USCENSUSBUREAU

..

Record Type

Person record

- Includes all of the person record variables such as sex, age, race, citizenship, veteran status, place of work, and means of transportation to work, as well as many others
- Vacant housing units have no person data
- The housing unit record and the person record are linked using the variables STATE and SERIALNO

Weighting

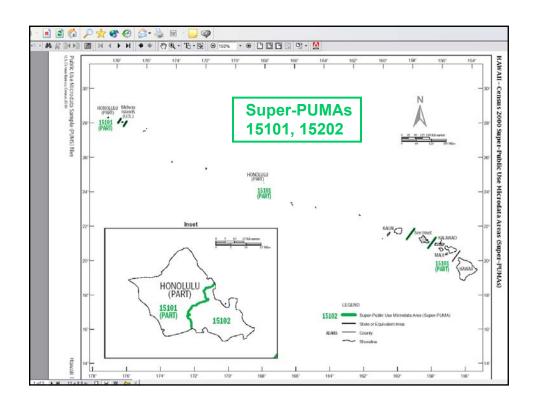
- There is a housing unit weight on the housing unit record and a person weight on the person record
- Information on when to use which weight is available as Data Note 5 in Chapter 8 of the technical documentation

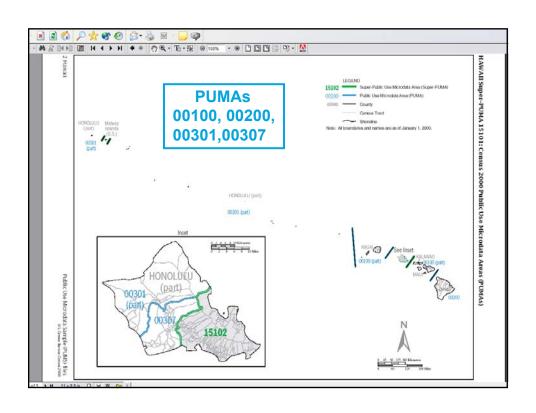
USCENSUSBUREAU ---

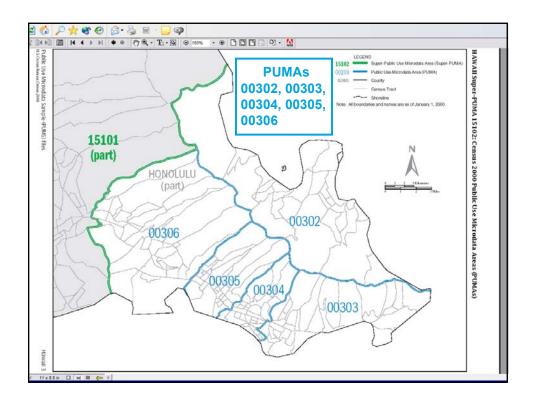
13

Geography

1 Percent Sample	5 Percent Sample
h_msacmsa1 (for super PUMA)	h_msacmsa5 (for PUMA)
h_msapmsa1 (for super PUMA)	h_msapmsa5 (for PUMA)
h_puma1 (Hawaii Super PUMAs: 15101, 15102)	h_puma5 (PUMAs: 00100, 00200, 00301, 00302, 00303, 00304, 00305, 00306, 00307
h_region	h_region
h_state (Hawaii = state code 15)	h_state







PUMS Files

Availability: Available on a single DVD containing Beyond 20/20 software. States also available as downloadable ASCII files via FTP (5-percent files) and FTP (1-percent files). 1-percent files also available separately on CD-ROM.

Subject Content: DVD containing both the 1-percent and 5-percent PUMS files providing individual records of responses to questionnaires with unique identifiers (names, address, etc.) removed so the confidentiality of respondents is protected. These files enable users to produce their own tabulations within the limits of the data provided.

Product ID and Pricing: V1-D00-PUMS-08-US1 \$70.00 Released December 31, 2003. Customer Services Center (orders) 301-763-INFO (4636)

Census Contact: Population Division (content), 301-763-2422

Helpful Websites

PUMS 1% data and related files

http://www.census.gov/PressRelease/www/2003/PUMS.html

PUMS 5% data and related files

http://www.census.gov/PressRelease/www/2003/ PUMS5.html

Technical Documentation

www.census.gov/prod/cen2000/doc/pums.pdf

USCENSUSBUREAU ----

19

Helpful Websites

Notes and Errata

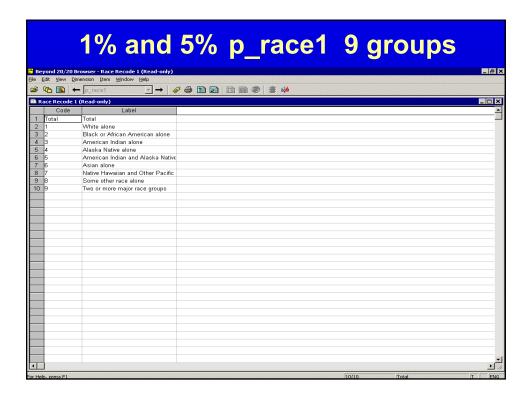
http://www.census.gov/prod/cen2000/notes/errata.pdf

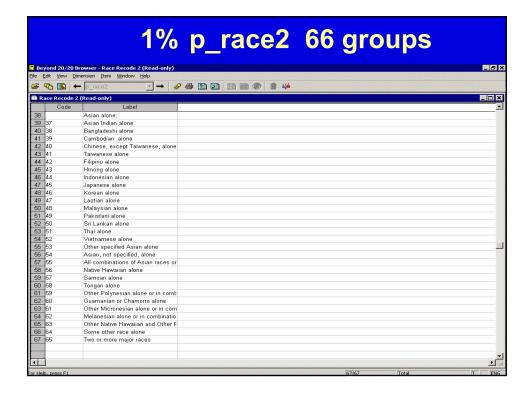
PUMS support

Censtats.census.gov/techdoc/pums.htm

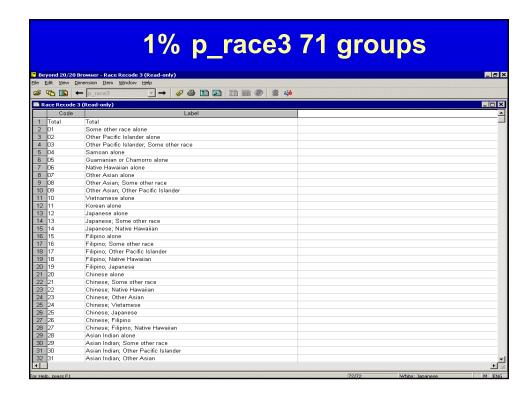
PUMS support for the DVD

www.census.gov/support/PUMSdata.html

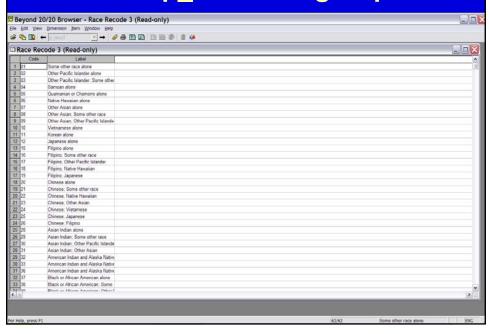


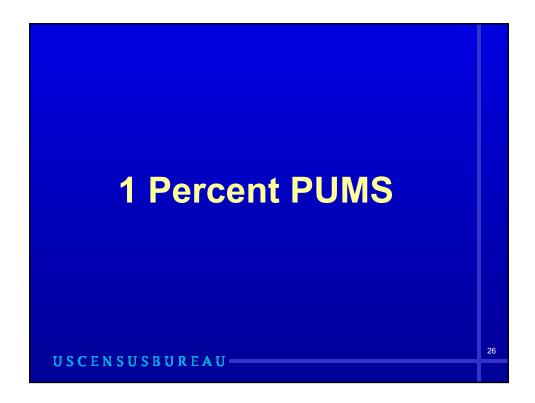


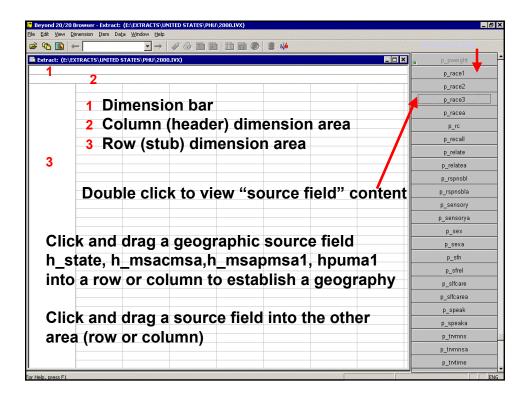
5 Beyond 20/20 Browser - Race Recode 2 (Read-only) See Set your generous pare yourselves The Code of the Code of

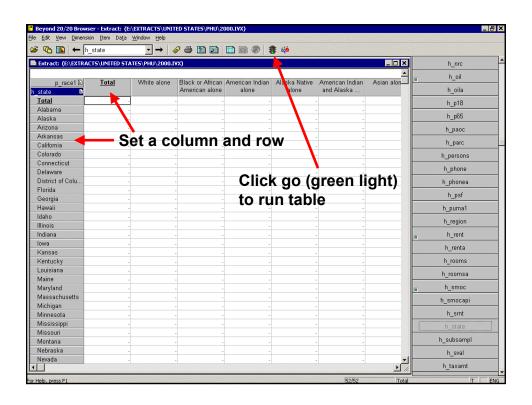


5% p_race3 62 groups

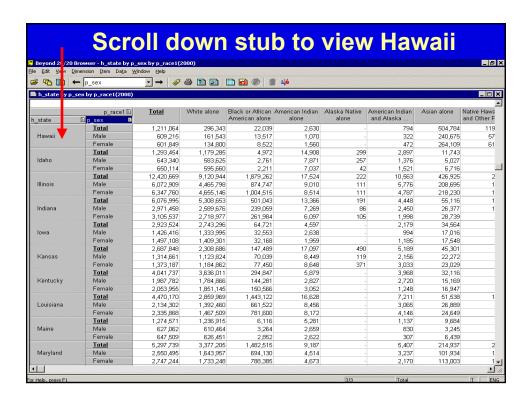


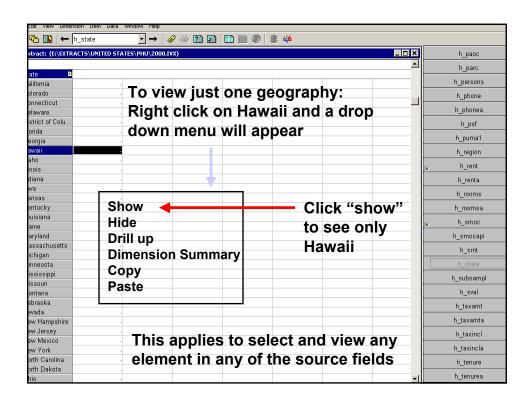


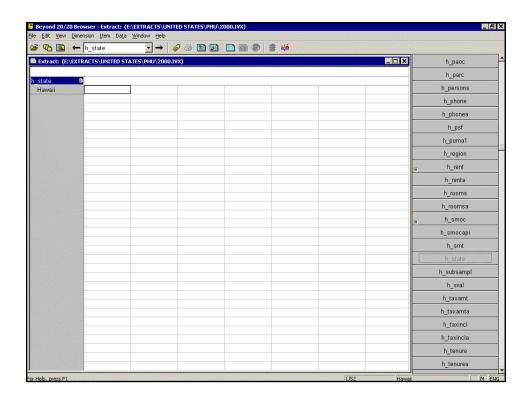


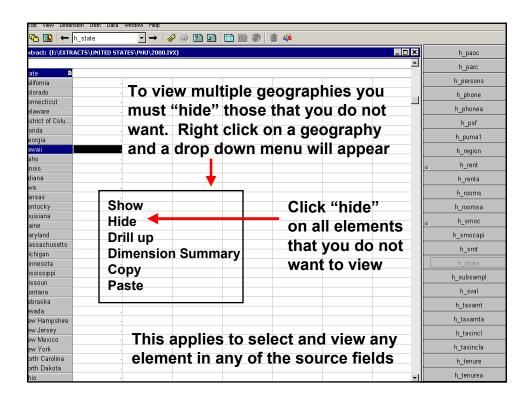


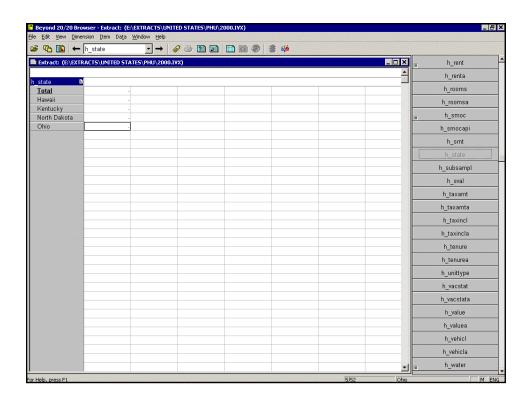
]→ (111 🗗	p sex	- → ∅	a m m		le ajé				
h_state by p_sex									
state 🗈	p_race1 🗈 p_sex 🖹	<u>Total</u>	White alone	American alone	American Indian alone	Alaska Native alone	American Indian and Alaska	Asian alone	Native Haw and Other I
Total	<u>Total</u>	281,421,906	211,264,815	34,404,747	1,852,290	97,589	490,290	10,149,943	40
	Male	138,031,205	103,742,815	16,306,775	912,786	50,011	253,787	4,899,727	20
	Female	143,390,701	107,522,000	18,097,972	939,504	47,578	236,503	5,250,216	20
Alabama Alaska Arizona Arkansas	<u>Total</u>	4,445,562	3,149,954	1,157,451	19,677	-	4,038	29,092	
	Male	2,145,646	212, 537, 1	538,427	10,032	-	1,797	13,322	
	Female	2,299,916	1,612,742	619,024	9,645		2,241	15,770	
Alaska	<u>Total</u>	628,493	431,236	19,830	4,120	79,630	13,218	25,749	
	Male	332,145	230,056	10,850	1,424	40,609	5,964	12,380	
	Female	296,348	201,180	8,980	2,696	39,021	7,254	13,369	
Arizona	<u>Total</u>	5,129,713	3,871,901	164,581	227,646	385	25,711	84,631	
	Male	2,556,941	1,914,117	88,225	110,409	251	12,619	40,459	
	Female	2,572,772	1,957,784	76,356	117,237	134	13,092	44,172	
Arkansas	<u>Total</u>	2,675,687	2,123,431	420,394	15,372	201	5,340	21,344	
	Male	1,317,560	1,048,469	201,126	7,080	53	3,338	9,960	
	Female	1,358,127	1,074,962		8,292	148	2,002	11,384	
	<u>Total</u>	33,879,320	20,105,863		207,514	1,917	96,338	3,655,699	12
California	Male	16,878,673	10,006,434	1,096,222	104,349	1,167	48,415	1,760,798	6
	Female	17,000,647	10,099,429	1,127,451	103.165	750	47.923	a 10,149,943 77 4,899,727 4,899,727 3 5,250,216 3 29,092 44 12,380 44 12,380 44 12,380 9,960 11,760,799 3 1,894,901 86,599 5 1,760,799 3 1,894,901 86,599 16,421 9,960 13,956,699 16,421 9,960 13,956,699 16,421 9,960 13,956,699 16,421 9,960 13,956,699 16,421 9,960 13,956,699 16,421 9,960 13,956,699 16,421 9,960 13,956,69 16,421 9,960 13,956,69 16,421 9,960 13,956,69 16,421 9,960 13,956,69 16,421 9,960 13,956,69 16,421 9,960 13,956,69 16,421 9,960 13,956,69 16,421 9,960 14,725 14,725 14,725 14,725 15,725	6
	Total	4,301,983	3,555,781	155,984	31.577	-	13,438		
Colorado	Male	2,176,042	1,785,376	83,329	15,935		7,403	41.725	
Colorado	Female	2,125,941	1,770,405		15,642		6,035		
	<u>Total</u>	3,406,431	2,779,948		5,757	123			
Connecticut	Male	1,639,705	1.336.329		2.434	-	1,515		
Connecticut	Female	1.766.726	1.443.619		3.323	123			
	Total	783,683	584,292		1,702	-	619		
Delaware	Male	381,416	282,532		921	-	533		
	Female	402,267	301,760			_	86		
	Total	572,781	176.231	343,770	208	-	1.047		
District of Colu	Male	269,383	87,405		104	-	524		
	Female	303,398	88.826		104		523	6.703	
		222,200	1020	122,101	, ,,,,,		520	2,100	Th.

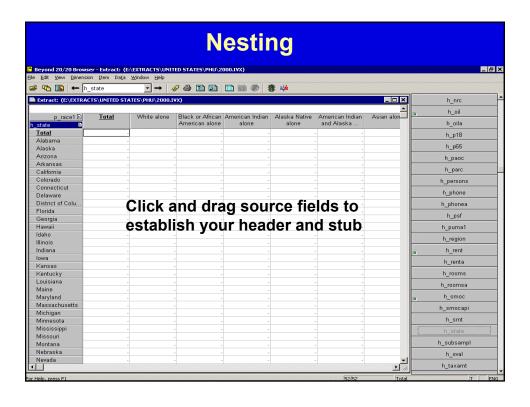


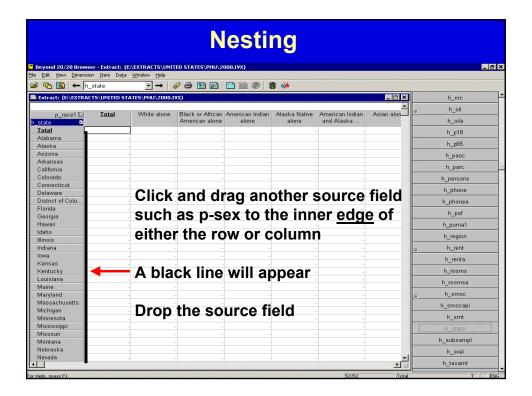


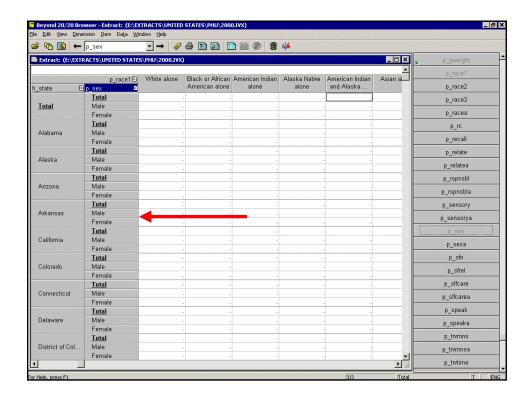


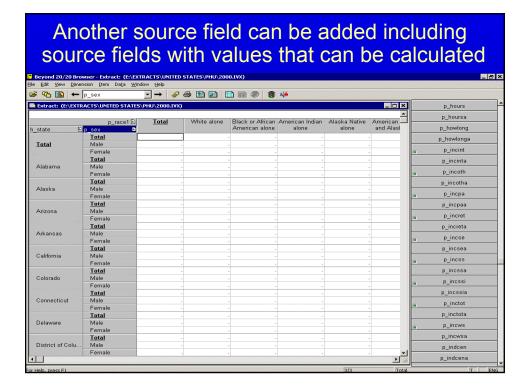


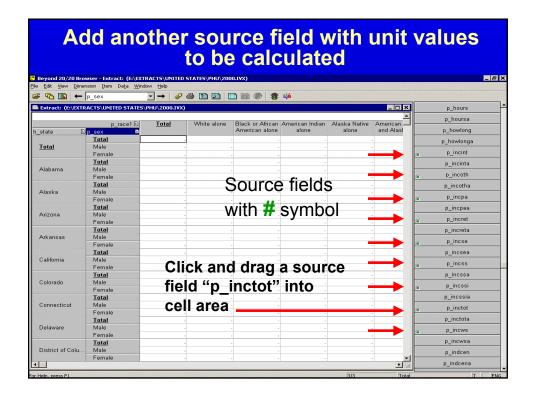


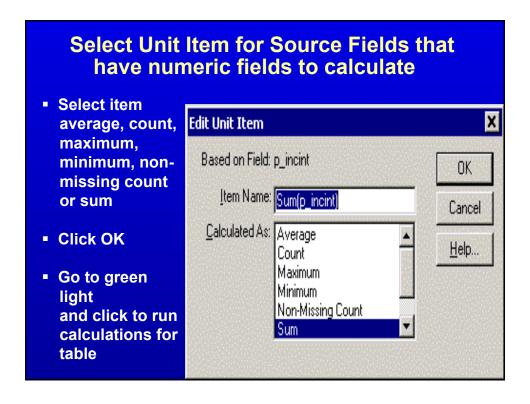


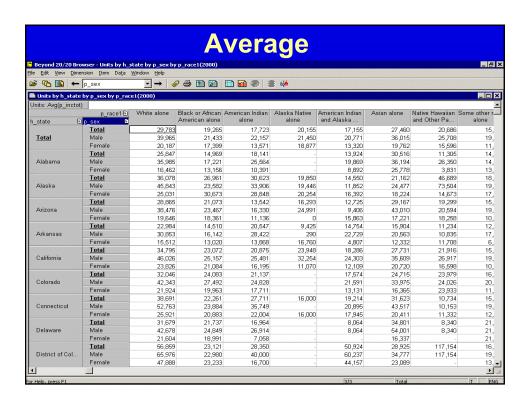


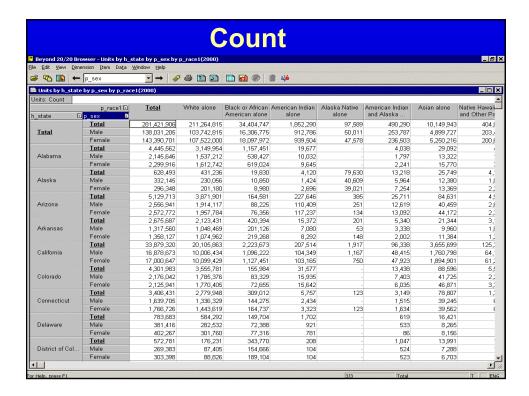


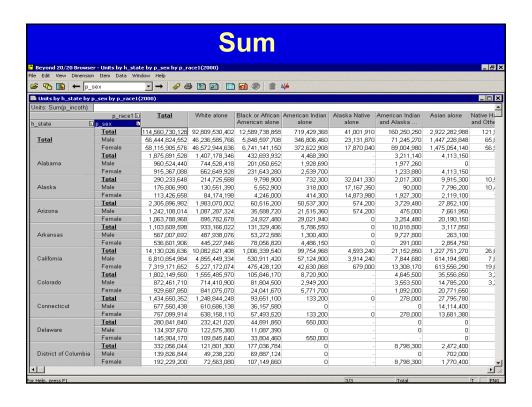


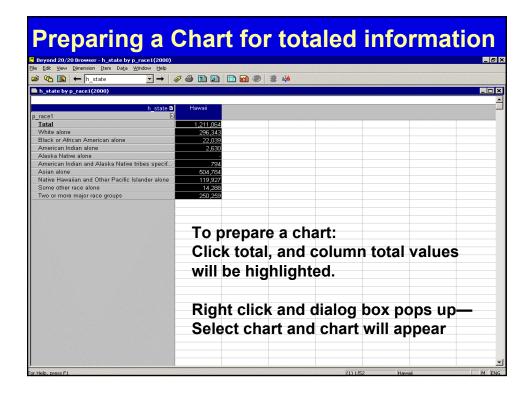


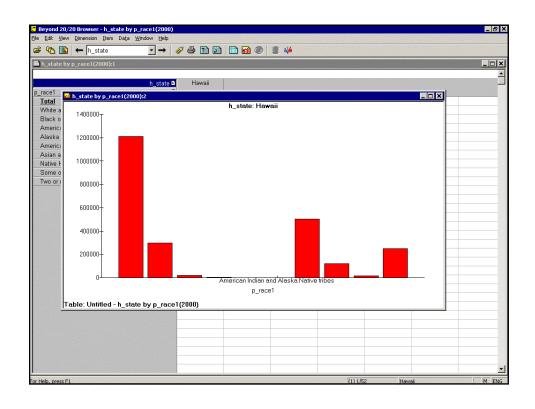


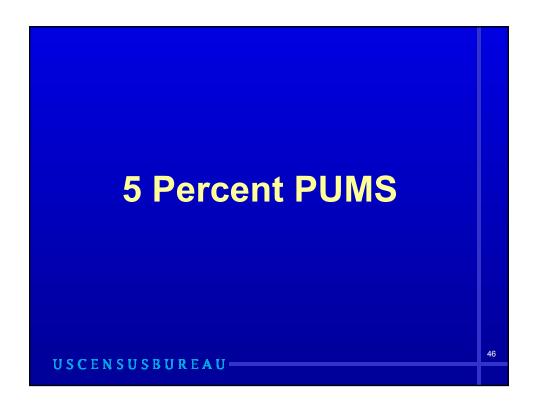


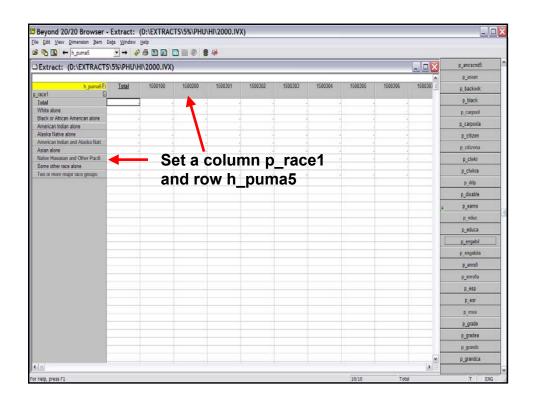


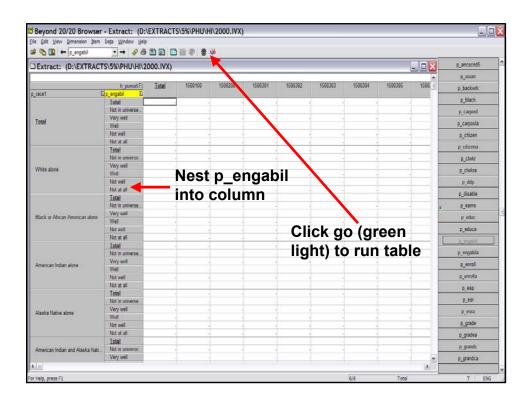












Edit Ylew Dimension Item ↑ Dimension Item	Daţa Window Help → → Ø ⊕ D D □ ■ ●	書 棒								
p_race1 by p_engabil	l by h_puma5(2000)									
	h puma6 B	Total	1500100	1500200	1500301	1500302	1500303	1500304	1500305	1500306
race1 E	p_engabil D									
	Total	1,211,717	186,193	149,458	113,717	118,050	101,923	110,264	111,192	142,46
	Not in universe (Under 5 years or SPEAK = 2)	908,485	145,659	124,301	90,612	102,284	76,522	72,115	61,783	107,32
	Very well	158,782	20.033	13.901	14,430	11,540	13.728	18.078	19.872	19.28
Total	Well	87.641	13,562	7.296	5,149	2,754	6,503	10,979	16,539	10,70
	Not well	51,051	6.411	3,751	3.246	1,402	4.398	7.912	11,176	4.81
	Not at all	5,758	528	209	280	70	772	1,180	1,822	32
White alone	Total	292.086	60.675	46.296	28.010	38.548	23,197	26.626	10,460	33.22
	Not in universe (Under 5 years or SPEAK = 2)	269.653	56.356	43.177	26.067	35.981	21,267	23.756	9.442	30.12
	Very well	17.374	3 100	2.358	1 648	2 136	1.692	2 283	652	2.19
	Well	3.256	785	569	187	301	134	395	255	58
	Not well	1.746	390	192	108	130	91	192	111	31
	Not at all	57	44			- 1	13			
	Total	21 940	636	1 025	5.416	1 993	320	1 129	1.062	6.76
	Not in universe (Under 5 years or SPEAK = 2)	20.330	553	966	5.130	1.832	320	1.062	970	6.29
	Very well	1.295	83	59	220	126		67	92	34
Black or African American alone	Well	258			59	36		-		10
	Not well	57			7					- 1
	Not at all		-		- 1	- 1				-
	Total	2.570	755	408	129	251	127	307	45	41
	Not in universe (Under 5 years or SPEAK = 2)	2.072	746	373	114	179	127	215	15	20
	Very well	300	9	28	- 11-	52	NE1	92		- 1
American Indian alone	Well	103	- 1	7	15	20		54	14	
	Not well	95	- 1	- 1	- 3	- 20		-	16	- 1
	Not at all	33							10	
Alaska Native alone	Total	335	-	80	26	-	229			
	Not in universe (Under 5 years or SPEAK = 2)	208		80	26		102			
	Very well	127		- 00	20		127	-	- 1	
	Well	121	- 1	- 1		- 1	141			
	Not well							1		
	Not at all	- 1	- 1	- 1	- 1	1		1	- 1	
American Indian and Alaska N	Total	895	201	346	18	104	-	- 1	165	2
	Not in universe (Under 5 years or SPEAK = 2)	660	183	290	18	82			62	
	Very well	168	18	56	10	22	- 1	1	72	
lal I	FOI HER	100	10	.00		- 11			16	- 1

