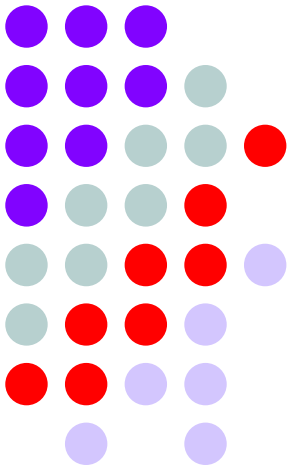
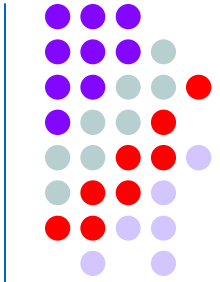
Microdata - PUMS

Pam Perlich
URBPL 5/6010: Urban Research
University of Utah
Rev. Nov. 27, 2006



Public Use Microdata Sample



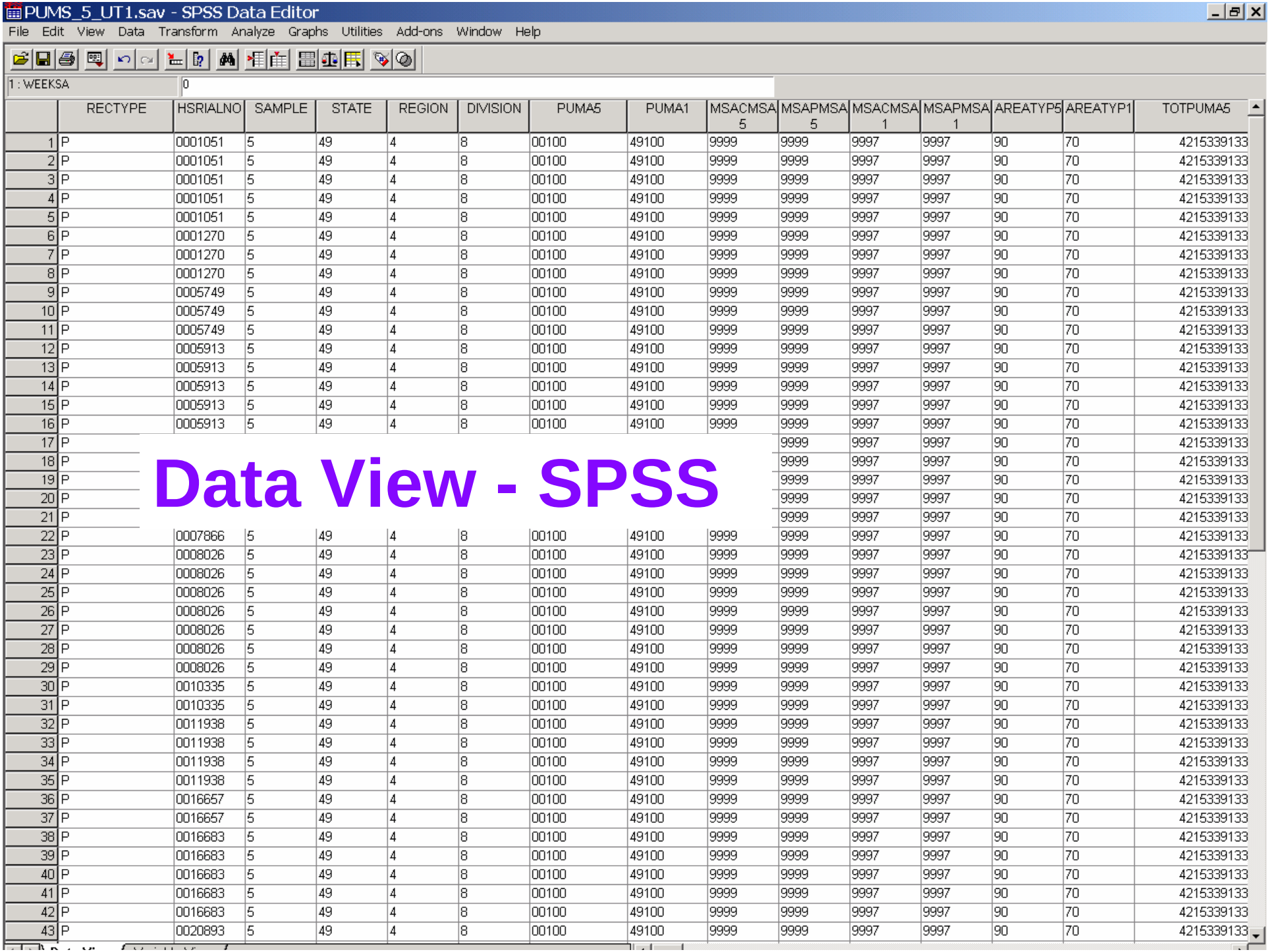
- Many surveys make microdata available
- Decennial Census data
- Allows customize estimates
- Long form is 1 in 6 household
- 5% PUMS and 1% PUM
- Household ID links
  - Household records
  - Person records

Record Layouts

- Rectangular files
- Fixed lengths for variables
- Write code to read files

PERSON RECORD

DATA	SIZE	BEGIN	END
D RECTYPE	1	1	1
T Record Type			
V		P . Person record	
D SERIALNO	7	2	8
T Housing/Group Quarters (GQ) Unit Serial No.			
R	0000001..9999999	. Unique identifier assigned within state	
D PNUM	2	9	10
T Person Sequence Number			
R	01..97	. Person Number	



	RECTYPE	HSRIALNO	SAMPLE	STATE	REGION	DIVISION	PUMA5	PUMA1	MSACMSA5	MSAPMSA5	MSACMSA1	MSAPMSA1	AREATYP5	AREATYP1	TOTPUMA5
1	P	0001051	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
2	P	0001051	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
3	P	0001051	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
4	P	0001051	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
5	P	0001051	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
6	P	0001270	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
7	P	0001270	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
8	P	0001270	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
9	P	0005749	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
10	P	0005749	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
11	P	0005749	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
12	P	0005913	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
13	P	0005913	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
14	P	0005913	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
15	P	0005913	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
16	P	0005913	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
17	P									9999	9997	9997	90	70	4215339133
18	P									9999	9997	9997	90	70	4215339133
19	P									9999	9997	9997	90	70	4215339133
20	P									9999	9997	9997	90	70	4215339133
21	P									9999	9997	9997	90	70	4215339133
22	P	0007866	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
23	P	0008026	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
24	P	0008026	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
25	P	0008026	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
26	P	0008026	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
27	P	0008026	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
28	P	0008026	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
29	P	0008026	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
30	P	0010335	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
31	P	0010335	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
32	P	0011938	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
33	P	0011938	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
34	P	0011938	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
35	P	0011938	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
36	P	0016657	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
37	P	0016657	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
38	P	0016683	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
39	P	0016683	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
40	P	0016683	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
41	P	0016683	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
42	P	0016683	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133
43	P	0020893	5	49	4	8	00100	49100	9999	9999	9997	9997	90	70	4215339133

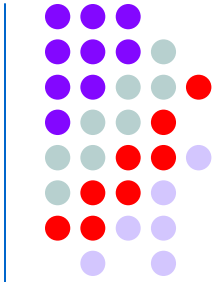
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Variable View - SPSS

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	RECTYPE	String	1	0	Record Type	{H, Household record}...	None	13	Left	Nominal
2	HSRIALNO	String	7	0	HU Record Housing/GQ Unit Serial Number	None	None	8	Left	Nominal
3	SAMPLE	String	1	0	Sample Identifier	{1, 1% sample}...	None	8	Left	Nominal
4	STATE	String	2	0	State (FIPS)	{10, Delaware}...	None	8	Left	Nominal
5	REGION	String	1	0	Region	{0, Region not identified}...	None	8	Left	Nominal
6	DIVISION	String	1	0	Division	{0, Division not identified}...	None	8	Left	Nominal
7	PUMA5	String	5	0	Public Use Microdata Area Code (PUMA)	None	None	11	Left	Nominal
8	PUMA1	String	5	0	Super Public Use Microdata Area Code (SuperPUMA)	None	None	9	Left	Nominal
9	MSACMSA	String	4	0	Metropolitan Area(MA):MSA/CMSA for PUMA	{9997, Mixed MSA/CMSA:nonmetro t	None	8	Left	Nominal
10	MSAPMSA	String	4	0	Metropolitan Area:MSA/PMSA for PUMA	{9997, Mixed MSA/CMSA:nonmetro t	None	8	Left	Nominal
11	MSACMSA	String	4	0	Metropolitan Area:MSA/CMSA for SuperPUMA	{9997, Mixed MSA/CMSA:nonmetro t	None	8	Left	Nominal
12	MSAPMSA	String	4	0	Metropolitan Area:MSA/PMSA for SuperPUMA	{9997, Mixed MSA/CMSA:nonmetro t	None	8	Left	Nominal
13	AREATYP5	String	2	0	Metropolitan Area:PUMA relationship to MA	{70, Both metro and nonmetro territor	None	8	Left	Nominal
14	AREATYP1	String	2	0	Metropolitan Area:SuperPUMA Relationship to MA	{70, Both metro and nonmetro territor	None	8	Left	Nominal
15	TOTPUMA5	Numeric	14	0	Total Area of PUMA	None	None	14	Right	Scale
16	LNDPUMA	Numeric	14	0	Land Area of PUMA	None	None	14	Right	Scale
17	TOTPUMA1	Numeric	14	0	Total Area of SuperPUMA	None	None	14	Right	Scale
18	LNDPUMA	Numeric	14	0	Land Area of SuperPUMA	None	None	14	Right	Scale
19	SUBSAMP	String	2	0	Subsample Number	None	None	8	Left	Nominal
20	HWEIGHT	N					None	7	Right	Scale
21	PERSONS	S					None	7	Left	Nominal
22	UNITTYPE	S					None	8	Left	Nominal
23	HSUB	S					None	4	Left	Nominal
24	HAUG	S					None	4	Left	Nominal
25	VACSTAT	String	1	0	Vacancy Status	{0, Not in universe}...	0	7	Left	Nominal
26	VACSTATA	String	1	0	Vacancy Status Allocation Flag	{0, Not allocated or GQ}...	None	8	Left	Nominal
27	TENURE	String	1	0	Home Ownership	{0, Not in universe}...	0	6	Left	Nominal
28	TENUREA	String	1	0	Home Ownership Allocation Flag	{0, Not allocated or GQ}...	None	7	Left	Nominal
29	BLDGSZ	String	2	0	Size of Building	{, Not in universe}...	00	6	Left	Nominal
30	BLDGSZA	String	1	0	Size of Building Allocation Flag	{0, Not allocated or GQ}...	None	7	Left	Nominal
31	YRBUILT	String	1	0	Year Building Built	{, Not in universe}...		7	Left	Nominal
32	YRBUILTA	String	1	0	Year Building Built Allocation Flag	{0, Not allocated or GQ}...	None	8	Left	Nominal
33	YRMOVED	String	1	0	Year Moved In	{, Not in universe}...		7	Left	Nominal
34	YRMOVED	String	1	0	Year Moved In Allocation Flag	{0, Not allocated or GQ}...	None	8	Left	Nominal
35	ROOMS	String	1	0	Number of Rooms	{, Not in universe}...		5	Left	Nominal
36	ROOMSA	String	1	0	Number of Rooms Allocation Flag	{0, Not allocated or GQ}...	None	6	Left	Nominal
37	BEDRMS	String	1	0	Number of Bedrooms	{, Not in universe}...		6	Left	Nominal
38	BEDRMSA	String	1	0	Number of Bedrooms Allocation Flag	{0, Not allocated or GQ}...	None	7	Left	Nominal
39	CPLUMB	String	1	0	Complete Plumbing Facilities	{, Not in universe}...		6	Left	Nominal
40	CPLUMBA	String	1	0	Complete Plumbing Facilities Allocation Flag	{0, Not allocated or GQ}...	None	7	Left	Nominal
41	CKITCH	String	1	0	Complete Kitchen Facilities	{, Not in universe}...		6	Left	Nominal
42	CKITCHA	String	1	0	Complete Kitchen Facilities Allocation Flag	{0, Not allocated or GQ}...	None	7	Left	Nominal
43	PHONE	String	1	0	Telephone Availability	{, Not in universe}...		5	Left	Nominal
44	PHONEA	String	1	0	Telephone Availability Allocation Flag	{0, Not allocated or GQ}...	None	6	Left	Nominal
45	FUEL	String	1	0	Heating Fuel	{, Not in universe}...		4	Left	Nominal

Data View | Variable View

Estimates from PUMS



- Extract data – may require linking person and household files
- Apply person or household weights
- Calculate estimate
- Calculate standard error
  - Unadjusted SE
  - Design factors

Table A. Unadjusted Standard Errors for Estimated Totals from Census 2000 5-Percent PUMS

Estimated total[1]	Size of geographic area tabulated[2]							
	100,000	250,000	500,000	750,000	1,000,000	5,000,000	10,000,000	25,000,000
1,000	137	138	138	138	138	138	138	138
2,500	215	217	217	218	218	218	218	218
5,000	300	305	307	307	307	308	308	308
10,000	414	427	432	433	434	435	436	436
15,000	492	518	526	528	530	533	533	534
25,000	597	654	672	678	681	687	688	689
75,000	597	999	1,101	1,132	1,148	1,185	1,189	1,192
100,000		1,068	1,233	1,283	1,308	1,365	1,371	1,376
250,000			1,541	1,780	1,887	2,124	2,152	2,169
500,000				1,780	2,179	2,924	3,004	3,051
750,000					1,887	3,480	3,631	3,718
1,000,000						3,899	4,135	4,271
5,000,000							6,892	8,718
10,000,000								10,677

[1]For estimated totals larger than 10,000,000, the standard error is somewhat larger than the table values. The formula (1) given below should be used to calculate the standard error.

[2]Total count of people, housing units, households, or families in the area if the estimated total is a person, housing unit, household, or family characteristic, respectively.

$$SE(\hat{Y}) = \sqrt{19(\hat{Y})\left(1 - \frac{\hat{Y}}{N}\right)}$$

N = Size of geographic area

$\hat{Y}$ = Estimate of characteristic total

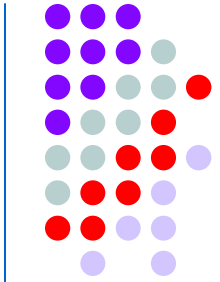
Unadjusted Standard Errors

Design Factors

Table E. Census 2000 PUMS Standard Error Design Factors—United States

Characteristic	Design factor
POPULATION	
Age	1.3
Sex	1.2
Race	2.2
Hispanic or Latino	2.1
Marital status	1.1
Household type and relationship	1.1
Disabled and employment disability	1.4
Ancestry	1.8
Place of birth	1.5
Citizenship status	1.6
Residence in 1995	2.0
Year of entry	1.7
Language spoken at home and ability to speak English	1.5
Educational attainment	1.2
School enrollment	1.5
Type of residence (urban/rural)	1.6
Household type	1.1
Family type	2.2
Subfamily type and presence of children	1.3
Grandparent status and responsibility for grandchild	1.4
Employment status	1.2

Reporting Confidence Levels



- Example: Report the 95% confidence interval (+ and - 2 SEs)
- Example: Estimate = 24,104

  Computed Unadjusted SE: 658

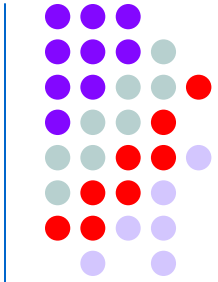
  Design Factor: 1.1

  Adjusted SE = 742

  95% CI = 2 x 742 = 1,447

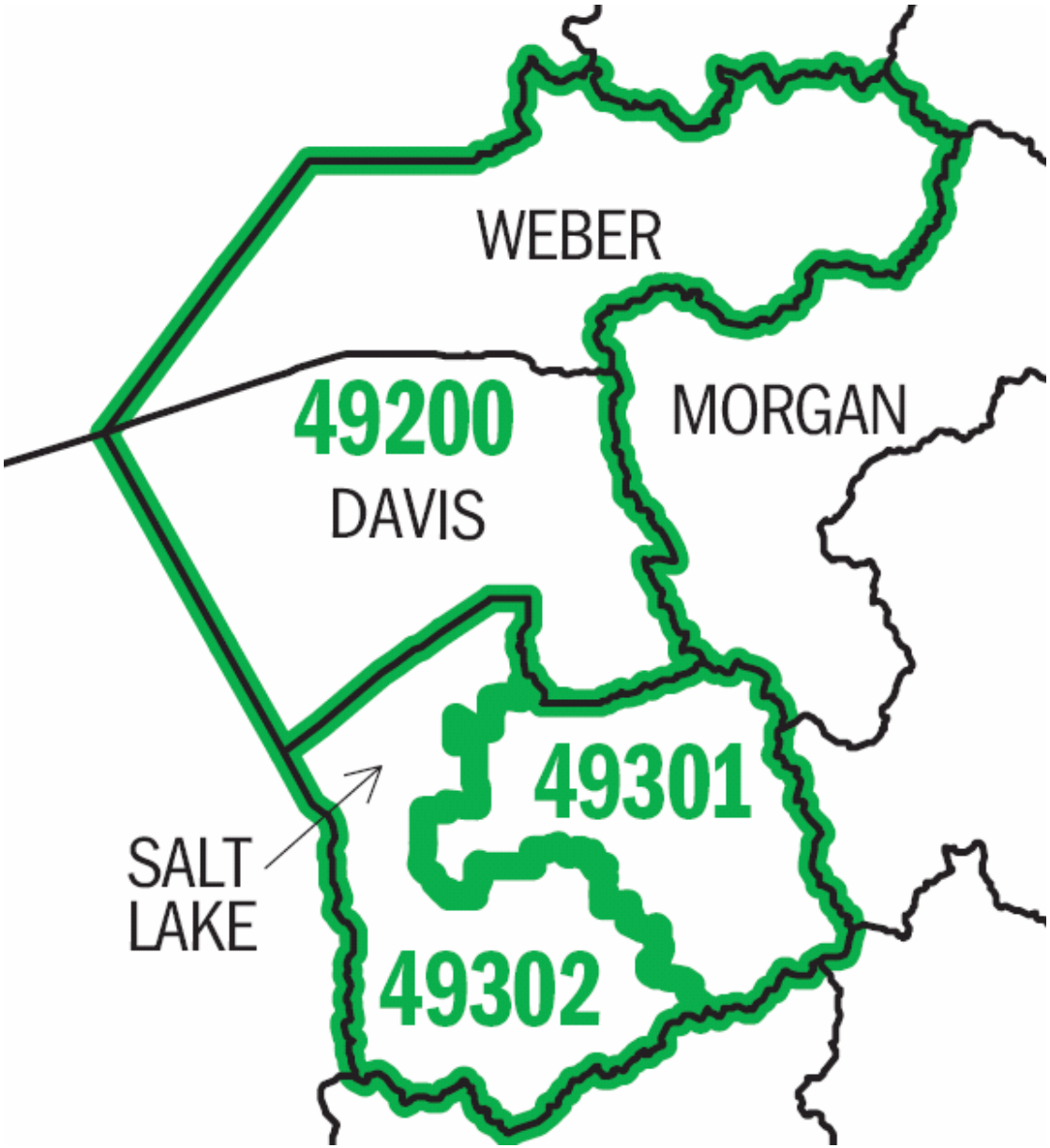
  LB: 22,657 UB: 25,551

PUMAs: Geography



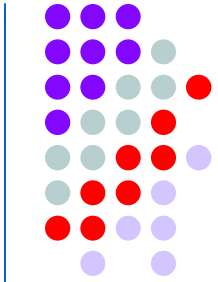
- PUMA = Public Use Microdata Area
- Super PUMAs
  - 1% PUMS
  - 400,000+ population
  - Utah has 4 Super PUMAs
- PUMAs
  - 5% PUMS
  - Nested within Super PUMAs

Utah Super PUMAs



- 49301 – Salt Lake County (pt.)
- 49302 – Salt Lake County (pt.)
- 49200 – Davis & Weber Counties
- 49100- Rest of State

2000 PUMS Resources



- PUMS data

http://www.census.gov/Press-Release/www/2003/PUMS5.html

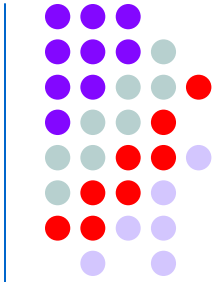
- Maps for Utah

http://ftp2.census.gov/geo/maps/puma/puma2k/ut_puma5.pdf

- Example of PUMS extractions – foreign born characteristics

http://www.business.utah.edu/bebr/bebrFiles/bebr_May_Jun2004.pdf

2000 – 2005 ACS PUMS



- PUMS data

http://factfinder.census.gov/home/en/acs_pums.html

- Currently available at state level
- Not always comparable to 2000 census results
- Only household population
- VERY LARGE confidence intervals

http://factfinder.census.gov/home/en/acs_pums_2005.html

uguide.pdf | American Communit... | Bureau of the Census

ACS PUMS 2005

U.S. Census Bureau

American FactFinder

Main | Search | Feedback | FAQs | Glossary | Site Map | Help

ACS Public Use Microdata Sample (PUMS) 2005

You are here: Main ▸ Data Sets ▸ ACS PUMS 2005

PUMS Data and Documentation:

- Overview
- **2005**
- 2004
- 2003
- 2002
- 2001
- 2000

Note: The population estimates of the American Community Survey (ACS) data are currently limited to the household population and exclude the population living in institutions, college dormitories, and other group quarters.

Download 2005 PUMS Data

Select a data type, data format, and state. Click 'GO'.

Data Type

- (•) Population Records
- ( ) Housing Records

Data Format

- (•) CSV (comma separated values)
- ( ) PC SAS Data Set
- ( ) UNIX SAS Data Set

State

Utah GO

Documentation

- Subjects available in PUMS files
- 2005 PUMS top coded values
- 2005 PUMS Code Lists

  Ancestry
  CMSA (Note: CMSAs are not available on the ACS PUMS)
  Hispanic Origin
  Industry
  Language
  Migration
  Occupation
  Place of Birth
  Place of Work
  Race 1 (9 categories)
  Race 2 (67 categories)
  Race 3 (72 categories)
  State

- 2005 PUMS Accuracy (PDF - 411 KB)
- 2005 Data Dictionary (PDF - 114 KB)

Source: U.S. Census Bureau. Last Revised: October 17, 2006

The letters PDF or symbol indicate a document is in the Portable Document Format (PDF). To view the file you will need the Adobe® Acrobat® Reader, which is available for **free** from the Adobe web site.

IPUMS

(ĭp'-ums)

- Integrated Public Use Microdata Series
- US Files:
  - 38 samples of demographic data sets
  - 15 decennial censuses
  - 2000 – 2004 American Community Surveys
- Single data set – uniform codes across time
  - Study change over time
- Major funding:

Funded by

NSF Sun microsystems University of Minnesota National Institutes of Health

- http://www.ipums.org/

MPC

Minnesota Population Center
University of Minnesota

Integrated Public Use Microdata Series

census microdata for social and economic research

IPUMS-USA is a coherent national census database spanning 1850 to 2005.

- High precision samples
- Integrated microdata
- Comprehensive documentation
- Free access and use!

IPUMS-International is a new project dedicated to collecting and distributing census data from around the world. Its goals are to:

- Collect and preserve data and documentation
- Harmonize data
- Disseminate the data absolutely free!

"An absolutely extraordinary project--one of the best 20 sites on the web"

Liens-Socio (French portal for the social sciences)