Public Use Microdata Samples

Using PDQ Explore Software

Grace York

University of Michigan Library

December 2003

Public Use Microdata Samples

- Copies of the original questionnaires with identifying information edited out

- Create your own cross tabulations of census data

  (e.g. Census Bureau does not tabulate teachers by single years of age)

Public Use Microdata Samples

Two files – 1% and 5% sample

- 1% has data for the nation, states, MSAs and super-Pumas (areas of 400,000)

- 5% has data for the nation, states, MSAs and Pumas (areas of 100,000)

PUMS Software Programs

- **FTP data from Census Bureau (and manipulate with SAS or SPSS**

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

- **Census Bureau CD-ROMS (Beyond 20/20 software)**

  http://www.census.gov/mp/www/Tempcat/PUMS.html

- **SDA Software (coming in Winter 2004 for California and Michigan**

  http://nds.umdl.umich.edu/n/nds/

- **PDQ Explore**

  http://www.pdq.com

PDQ Explore Software

- **Easy interface to**
  - Public Use Microdata Samples, 1 and 5%, 1980-2000
  - IPUMS, edited PUMS, 1850-1880, 1900-1920, 1940-1990
  - Current Population Survey, 1991+
  - Mortality Schedules
- **Permits users to tabulate their own variables**

Access to PDQ

- **UMich Documents Center:**

  Reference Desk and Scholar's Workstation
  (request login help from Denise or Grace)

- **Academics and librarians may request free IDs and passwords from PDQ. See web site at:**

  - **http://www.pdq.com**

Before Beginning...

Define the data you want in terms of a spreadsheet

I want the single years of age by sex for all Vietnam veterans in the United States

Universe = Vietnam veterans in the U.S.

Column=sex (not very wide)

Row=single years of age (could be long)

Before Beginning…

Consult Chapter 6 of the PUMS codebook if you want to check the possibilities
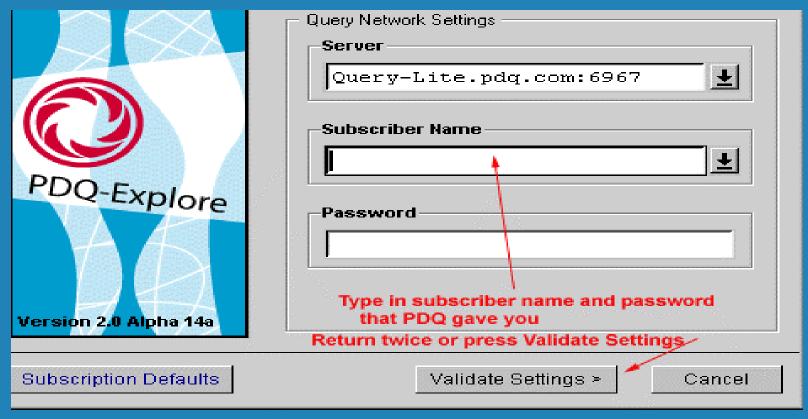
http://www.census.gov/prod/cen2000/doc/pums.pdf

Software

- Download the software per instructions to your hard drive
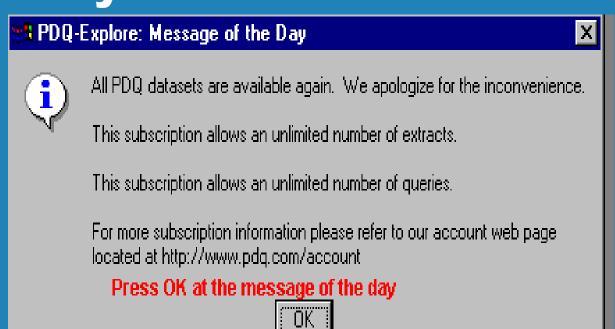- To begin searching, open the icon on your desktop

explore.exe

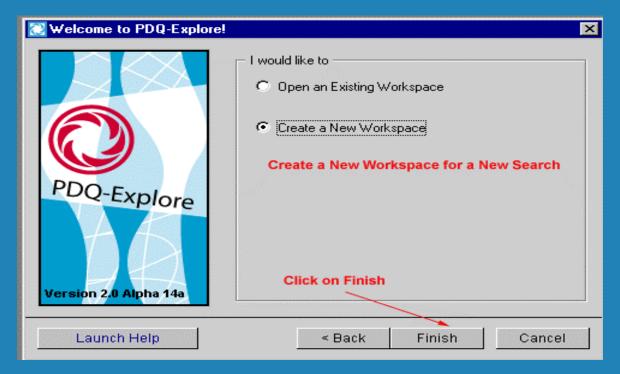# Logging On

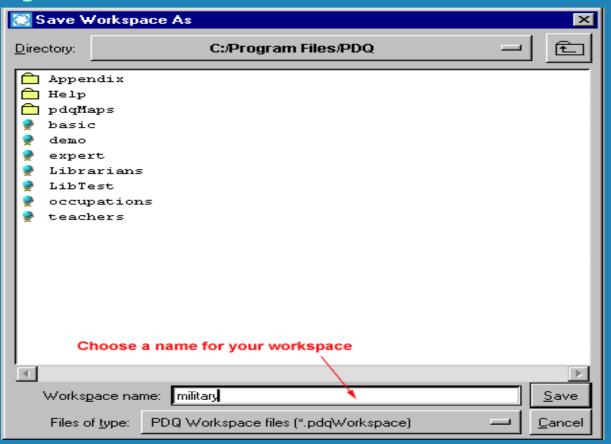Enter the subscriber name and password that you were given

PDQ-Explore

Version 2.0 Alpha 14a

Query Network Settings

Server

Query-Lite.pdq.com:6967

Subscriber Name

Password

Type in subscriber name and password that PDQ gave you

Return twice or press Validate Settings

Subscription Defaults

Validate Settings >

Cancel

Logging On

Press OK to close the message of the day

PDQ-Explore: Message of the Day

All PDQ datasets are available again. We apologize for the inconvenience.

This subscription allows an unlimited number of extracts.

This subscription allows an unlimited number of queries.

For more subscription information please refer to our account web page located at http://www.pdq.com/account

Press OK at the message of the day

[OK]

Defining Workspace

- To conduct a new search, create a new workspace

- Press Finish or return twice

Defining Workspace

Name your file on the hard disk and save.

Save Workspace As

Directory: **C:/Program Files/PDQ**

- 📁 Appendix
- 📁 Help
- 📁 pdqMaps
- basic
- demo
- expert
- Librarians
- LibTest
- occupations
- teachers

Choose a name for your workspace

Workspace name: military

Files of type: PDQ Workspace files (*.pdqWorkspace)

Save

Cancel

Defining Workspace

At the next screen, use the top menu to choose Workspace; then Add a Data Set

PDQ-Explore Workspace: military

File | View | Workspace | Query | Help

Add Data Set
Remove Data Set

✔ Show Names Hierarchically
✔ Show Names Alphabetically

✔ Show All Names
Show Subset of Names ▶
Manage Symbol Subsets

Data

1. Click on Workspace

2. Add a Data Set

Defining Workspace

Browse data sets; highlight ipums, pums, cps, or mortality file; Open

Defining Variables

- Once you choose a data set, its codebook will open up
- Click on the plus button to get a list of variables, their alphabetic symbols, and any numeric values

Defining Variables

With the PUMS files, geographic names (state, metropolitan area puma, etc. appear in housing file)

| File | View | Workspace | Query | Help |

Data	Short
⊞ 📁 UNASSOCIATED	Items not associated with a particlar codebo
⊟ 📂 pums_2000_5pct	2000 5% Public Use Microsample
⊟–housing	
⊞–person	
⊞–PDQ UNIVERSES	
⊞–PDQ CUSTOM ITEMS	
⊟–USER UNIVERSES	
⊞–USER CUSTOM ITEMS	
⊟–USER QUERIES	

HIERARCHY: pums_2000_5pct person
Description:

Click on the plus to open the codebook

Geographic terms appear in housing
Most data items are found by expanding person record

Defining Variables

- Determine the first alphanumeric variable you want (e.g. Vietnam veteran: yes is VPS5=1)
- Use Top Menu to Choose Query/Setup New Expert Query

Expert Query Form

1. Make sure you have the correct data set
2. Determine if you want a tabulation (counts or numbers)
3. Name your file

Expert Query Form

Enter the code for UNIVERSE (what you're counting)
in the Universe box (e.g. vps5=1 are Vietnam veterans
for the entire U.S.)

Expert Query Form

- Enter the code for the variables in the ROW box (age = single years of age) and the code for the variables in the COLUMN box (e.g. sex)
- Press RESULTS to run the query

Search Results

Search results appear in spreadsheet format

Saving Results

- **Click on File/Export Query Results**
- **You can save as CSV , tab deliminted and several other formats**
- **CSV recommended for use with Excel**

Narrowing the Universe

Narrow the universe by using & newcode (e.g. vps5=1 & state=26)

Altering the Spreadsheet Tabulations

Once you have a spreadsheet, click on Options to create totals or percentages for tables or columns

Adding More Parameters

Expand the table detail by repeating the row and column data for another parameter (e.g. race) as shown in Dimension 3

Altering Spreadsheet Appearance

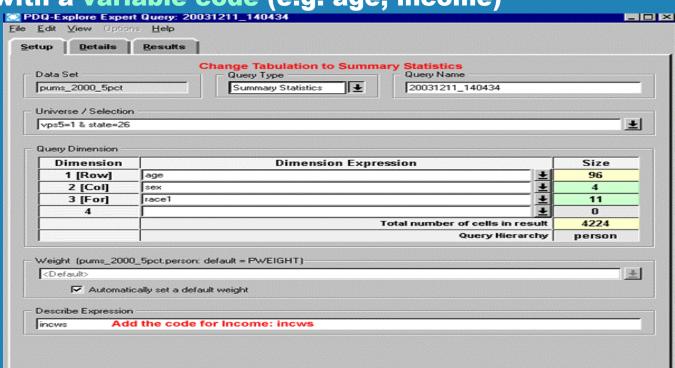- The default shows separate tables for each of the values in the third dimension (e.g. separate spreadsheets for white and black)
- Change Axis3 tab to FOREACH everything on same spreadsheet

Calculating Means or Averages

- Calculate averages by changing the query type to summary statistics at the top
- Fill in the new Describe Expression box at the bottom with a variable code (e.g. age, income)

Complex Table

Mean income of white male Vietnam veterans in Michigan by age

Setup | Details | Results

Data Set: pums_2000_5pct
Query Type: Summary Statistics
Universe / Selection: vps5=1 & state=26
Axis 1: age
Axis 2: sex
Axis 3: race1
Weight: <Default>

Mean income of white male Vietnam vets in Michigan by single year of age

Dimension	Swap Dimension	
Axis-01: "age"	<<ROW>>	±
Axis-02: "sex"	Male	±
Axis-03: "race1"	White alone	±

age	Frequency	Mean	Std. Dev.	Low	High
40	154	31,161.04	23,294.31	0	100,000
41	325	47,887.38	54,263.79	0	316,000
42	1,269	41,740.26	37,080.38	0	316,000
43	3,784	40,493.53	33,529.92	0	316,000
44	5,307	35,432.27	32,354.54	0	316,000
45	6,330	36,691.42	26,275.84	0	108,000
46	7,076	43,821.46	41,562.97	0	316,000
47	11,010	44,216.63	44,644.74	0	316,000
48	11,794	45,650.45	37,056.92	0	316,000
49	14,467	49,157.72	51,423.00	0	316,000
50	21,180	43,434.30	33,137.83	0	316,000
51	23,082	47,201.78	45,242.20	0	316,000
52	25,684	50,317.90	49,733.44	0	316,000
53	27,434	48,112.29	47,735.66	0	316,000
54	18,106	52,032.12	57,261.70	0	316,000
55	16,837	46,741.23	48,387.48	0	316,000
56	14,213	47,184.23	51,122.07	0	316,000
57	11,313	47,211.51	57,738.53	0	316,000
58	7,610	47,834.01	66,150.78	0	316,000
59	4,901	45,236.24	57,505.73	0	316,000
60	3,424	50,184.67	72,612.94	0	316,000

Small Area Geography

- Data from the PUMS 5% file is available for states, metropolitan areas, and Public Use Microdata Areas (PUMAS) of 100,000

- You can identify a PUMA or group of PUMAs using
  - Maps in American Factfinder (http://factfinder.census.gov/)
  - PDF maps on the Census Bureau web site (http://www.census.gov/geo/www/maps/puma5pct.htm)
  - Mable/Geocorr Search Engine (http://mcdc2.missouri.edu/websas/geocorr2k.html)

Small Area Geography

This map shows Detroit and its closest suburbs as PUMA 3701-3801

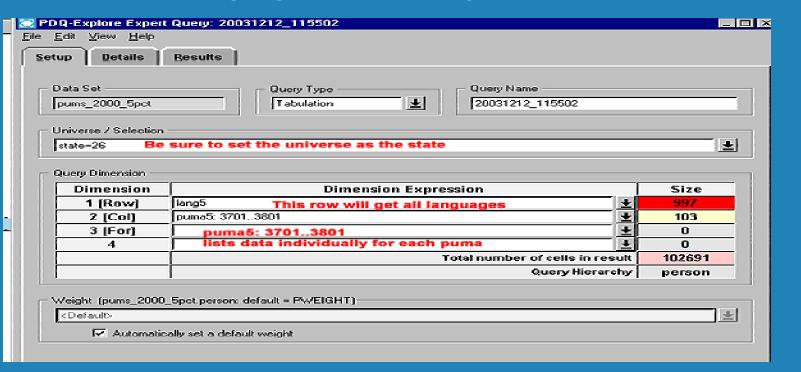Determine the PUMA numbers using American Factfinder or Census Bureau pdf maps

Codebook and PUMAS

The Explore Codebook shows PUMA5 as term for 5% PUMA boundaries

⊞ MRT2AMTA	Second Mortgage Payment (monthly amo...
⊞ MSACMSA1	Metropolitan Area: MSA/CMSA for SuperP...
⊞ MSACMSA5	Metropolitan Area: MSA/CMSA for PUMA
⊞ MSAPMSA1	Metropolitan Area: MSA/PMSA for SuperP...
⊞ MSAPMSA5	Metropolitan Area: MSA/PMSA for PUMA
⊞ NOC	Number of own children under 18 years i...
⊞ NPF	Number of people in family
⊞ NRC	Number of related children under 18 years...
⊞ OIL	Cost of Oil, Kerosene, or Wood (annual)
⊞ OILA	Cost of Oil, Kerosene, or Wood (annual) Al...
⊞ P18	Number of people under 18 years in hous...
⊞ P65	Number of people 65 years and over in ho...
⊞ PAOC	Presence and Age of Own Children under...
⊞ PARC	Presence and Age of Related Children un...
⊞ PERSONS	Number of person records following this...
⊞ PHONE	Telephone Availability
⊞ PHONEA	Telephone Availability Allocation Flag
⊞ PSF	Presence of Subfamily in Household
─ PUMA1	Super Public Use Microdata Area Code (Su...
─ PUMA5	Public Use Microdata Area Code (PUMA)

**Determine the geographic code
for pumas in the housing unit record
(pumas are "puma5")**

Small Area Geography and Ranges

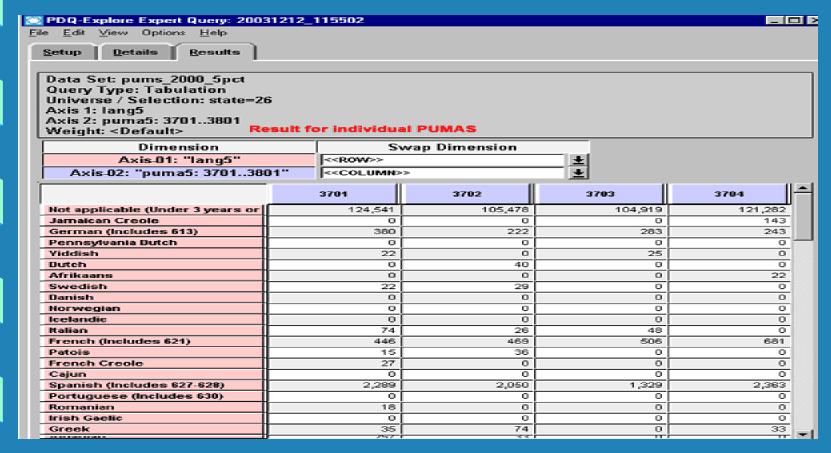When creating data sets for PUMAS, be sure to include the correct state as the universe (e.g. state=26)

Small Area Geography and Ranges

- **Puma5: 3701..3801 will list the data for each individual area**

Small Area Geography and Ranges

Search result for each individual PUMA

PDQ-Explore Expert Query: 20031212_115502

File   Edit   View   Options   Help

Setup   **Details**   **Results**

Data Set: pums_2000_5pct
Query Type: Tabulation
Universe / Selection: state=26
Axis 1: lang5
Axis 2: puma5: 3701..3801
Weight: <Default>   **Result for individual PUMAS**

Dimension	Swap Dimension		
Axis-01: "lang5"	<<ROW>>		
Axis-02: "puma5: 3701..3801"	<<COLUMN>>		

	3701	3702	3703	3704
Not applicable (Under 3 years or	124,541	105,478	104,919	121,282
Jamaican Creole	0	0	0	143
German (Includes 613)	360	222	283	243
Pennsylvania Dutch	0	0	0	0
Yiddish	22	0	25	0
Dutch	0	40	0	0
Afrikaans	0	0	0	22
Swedish	22	29	0	0
Danish	0	0	0	0
Norwegian	0	0	0	0
Icelandic	0	0	0	0
Italian	74	26	48	0
French (Includes 621)	446	469	506	681
Patois	15	36	0	0
French Creole	27	0	0	0
Cajun	0	0	0	0
Spanish (Includes 627-628)	2,289	2,050	1,329	2,363
Portuguese (Includes 630)	0	0	0	0
Romanian	18	0	0	0
Irish Gaelic	0	0	0	0
Greek	35	74	0	33

Small Area Geography for Ranges

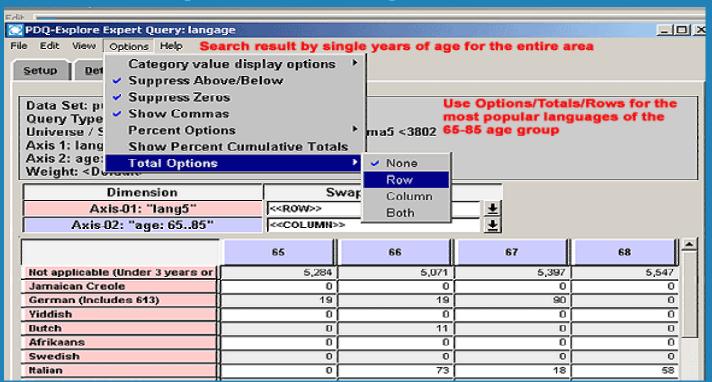To get the total for the area, list it in the universe as puma5 >3700 & puma5 >3802 & state=26

Small Area Geography for Ranges

To get a listing of single years of age between 65 and 85, list column as age: 65..855

Calculating Totals

- To calculate the most spoken languages by 65-85 year olds as a group
- Click on Options/Total Options/Row

Complex Result

Polish and Spanish are two most popular languages spoken by seniors 65-85 in Detroit

PDQ-Explore Expert Query: langage

File   Edit   View   Options   Help

Setup | Details | Results

Data Set: pums_2000_5pct
Query Type: Tabulation
Universe / Selection: state=26 & puma5 >3700 & puma5 <3802
Axis 1: lang5   **Search results show Polish and Spanish**
Axis 2: age: 65..85   **Other interesting results**
Weight: <Default>

Dimension	Swap Dimension		
Axis-01: "lang5"	<<ROW>>		
Axis-02: "age: 65..85"	<<COLUMN>>		

	84	85	Total
Patois	0	0	51
French Creole	0	0	18
Spanish (Includes 627-628)	63	36	**2.** 2,120
Portuguese (Includes 630)	0	0	0
Romanian	0	0	100
Greek	0	0	396
Albanian	0	0	26
Russian	0	0	71
Ukrainian	0	18	257
Czech	0	0	15
Polish	81	95	**1.** 2,421
Slovak	0	0	56
Bulgarian	0	0	0
Macedonian	0	0	41
Serbocroatian	0	0	37
Croatian	0	0	14
Serbian	0	0	30
Lithuanian	0	0	0
Armenian	0	0	0
Persian	0	0	16