Reduce steps to save consumer dollars

Delivering ready-to-use ACS data for sophisticated data consumer

Jan Vink

Cornell Program on Applied Demographics

Florio Arguillas

Cornell Institute for Social and Economic Research (CISER)

CISER CORNELL INSTITUTE for Social and Economic Research

www.ciser.cornell.edu

Program on Applied Demographics

CORNELL POPULATION CENTER

pad.human.cornell.edu

Working with many variables

- Many analyses require many variables
  - To gain insight
  - For statistical relations
- Not always clear what variables are needed before hand

Would be easier to have all variables at hand at the beginning of analyses

Main alternatives for multi-variable Census Bureau data

- Summary files and Public Use Microdata
  - Download files from Census Bureau (AFF)
  - Download files from Census Bureau (FTP)
  - DataFerret
- Public Use Microdata only
  - IPUMS

Data Ferret and IPUMS

- Advantanges:
  - Easy access and selection
  - Custom tabulations

- Disadvantages:
  - Geographic filtering can be a hassle
  - Have to go back and forth for additional variables

Census FTP and AFF

- Advantages
  - Direct from the source

- Disadvantages
  - Rather complex system of files

Census files workflow – Summary files

Geographic file
One record for each geography
Geographic codes and information
Unique LOGRECNO

Data files
One record for each geography
LOGRECNO
Data from detailed ACS tables (a few tables per file)

Files need to be imported and joined to be used for analyses

Step 1: find codebooks
Step 2: Select appropriate data files
Step 3: Write import procedures
   Geographic file is fixed length
   Data files are comma separated
   Add variable names
Step 4: Join on LOGRECNO (GEOID2 if from AFF)

SAS Macro's are provided

Census files workflow – PUMS files

Household file
One record for each household
Unique SERIALNO

Person file
One record for each person
SERIALNO + SPORDER

Files need to be imported and joined to be used for analyses

Step 1: find codebooks
Step 2: Write import procedures
        Data files are comma separated
        Variable names in files
Step 3: Join on SERIALNO (many to one)

CISER approach

- Make data ready for use
  - Save steps everyone using Census data takes
  - Lower threshold to work with this kind of Census data
- Implement steps that respect need for transparencyand reproducibility
- Metadata approach (CED2AR)
  - CED2AR = Comprehensive Extensible Data Documentation and Access Repository
  - CED^{2}AR is designed to improve the discoverability of both public and restricted data

Currently available

- Download centers for:
  - Census 2010 Rural/Urban - SF1 SF2
  - ACS 2015 1 year PUMS
- Data available in multiple formats
  - SAS, SPSS, STATA, CVS
- Data in single file for each state, or single file for the nation

Future developments

- Variable labels
- Set up files
- Get codebook in CED2AR
- Register datasets at EZID to get Digital Object Indentifiers (DOI) to promote reproducibility
- Make SAS code available
- Other years depending on general feedback and specific requests by Cornell Researchers (CISER clients)

How to find the current products?

- Currently available through:
  - www.ciser.cornell.edu
    - Download centers can be found searching for 2010 "download center" or 2015 "download center"
      - http://www.ciser.cornell.edu/pub/2010SF1/census2010sf1.shtm
      - http://www.ciser.cornell.edu/pub/2010SF2/census2010sf2.shtm
      - https://cisermgmt.cornell.edu/go/pub/ACS2015.HTML

- Revised pages available this summer
  - Improved user interface and other added features

American Community Survey 2015, CISER Download Center

The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about our nation and its people. Information from the survey generates data that help determine how more than $400 billion in federal and state funds are distributed each year. Demographic variables include sex, age, relationship, households by type, race, and Hispanic origin. Social characteristics variables include school enrollment, educational attainment, marital status, fertility, grandparents caring for children, veteran status, disability status, residence one year ago, place of birth, United States citizenship status, year of entry, world region of birth of foreign born, language spoken at home, and ancestry. Variables focusing on economic characteristics include employment status, commuting to work, occupation, industry, class of worker, income and benefits, and poverty status. Variables focusing on housing characteristics include occupancy, units in structure, year structure was built, number of rooms, number of bedrooms, housing tenure, year householder moved into unit, vehicles available, house heating fuel, utility costs, occupants per room, housing value, and mortgage status. The American Community Survey is conducted under the authority of Title 13, United States Code, Sections 141 and 193, and response is mandatory.

Cornell Institute for Social and Economic Research (CISER) developed this download center to deliver ready-to-use ACS2015 1-year data sets. We made available FREE full/complete ready-to-use ACS2015 1-year datasets in multiple formats (SAS, SPSS, STATA, CSV) via downloads from the table below with no membership or account creation required. The original source CSV files for ACS 2015 1-year PUMS were downloaded from https://www2.census.gov/programs-surveys/acs/data/pums/2015/1-Year/ on October 27, 2016. This download center was built with SAS code which can update the files with newer data as needed. If you have comments/suggestions please contact Florio Arguillas (foa2@cornell.edu).

Codebook: https://www2.census.gov/programs-surveys/acs/tech_docs/pums/data_dict/PUMSDataDict15.pdf

State	PERSON						HOUSEHOLD						PERSON-HOUSEHOLD					
	Records	Variables	SAS	SPSS	STATA	CSV*	Records	Variables	SAS	SPSS	STATA	CSV*	Records	Variables	SAS	SPSS	STATA	CSV*
National (US)	3147005	284	.sas7bdat	.sav	.dta	.csv	1496678	235	.sas7bdat	.sav	.dta	.csv	3147005	514	.sas7bdat	.sav	.dta	.csv
Alabama	6619	284	.sas7bdat	.sav	.dta	.csv	3560	235	.sas7bdat	.sav	.dta	.csv	6619	514	.sas7bdat	.sav	.dta	.csv
Alaska	47476	284	.sas7bdat	.sav	.dta	.csv	24495	235	.sas7bdat	.sav	.dta	.csv	47476	514	.sas7bdat	.sav	.dta	.csv
Arizona	29605	284	.sas7bdat	.sav	.dta	.csv	15001	235	.sas7bdat	.sav	.dta	.csv	29605	514	.sas7bdat	.sav	.dta	.csv
Arkansas	67014	284	.sas7bdat	.sav	.dta	.csv	32344	235	.sas7bdat	.sav	.dta	.csv	67014	514	.sas7bdat	.sav	.dta	.csv
California	374943	284	.sas7bdat	.sav	.dta	.csv	155218	235	.sas7bdat	.sav	.dta	.csv	374943	514	.sas7bdat	.sav	.dta	.csv
Colorado	53570	284	.sas7bdat	.sav	.dta	.csv	25189	235	.sas7bdat	.sav	.dta	.csv	53570	514	.sas7bdat	.sav	.dta	.csv
Connecticut	35787	284	.sas7bdat	.sav	.dta	.csv	17170	235	.sas7bdat	.sav	.dta	.csv	35787	514	.sas7bdat	.sav	.dta	.csv
Delaware	9017	284	.sas7bdat	.sav	.dta	.csv	4547	235	.sas7bdat	.sav	.dta	.csv	9017	514	.sas7bdat	.sav	.dta	.csv
Florida	194548	284	.sas7bdat	.sav	.dta	.csv	100291	235	.sas7bdat	.sav	.dta	.csv	194548	514	.sas7bdat	.sav	.dta	.csv
Georgia	97854	284	.sas7bdat	.sav	.dta	.csv	46342	235	.sas7bdat	.sav	.dta	.csv	97854	514	.sas7bdat	.sav	.dta	.csv
Hawaii	14124	284	.sas7bdat	.sav	.dta	.csv	6095	235	.sas7bdat	.sav	.dta	.csv	14124	514	.sas7bdat	.sav	.dta	.csv

Questions?

Contact:

Florio Arguillas
foa2@cornell.edu