

# RDC Demographic Data

Jim Davis

Boston Census Research Data Center

US Bureau of the Census

*Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.*

# Demographic data available

- Individual and household data
- Restricted use RDC data
  - more detailed geography
  - less top-coding/censoring
  - additional variables

# More Detailed Geography

Data Set	RDC Years	Geography
Decennial Census	1970-2000	Block
American Community Survey (ACS)	1996-2009	Block
Survey of Income and Program Participation (SIPP)	1984-2008	Tract
Current Population Survey (CPS) – March Supplement	1967-2008	Tract
American Housing Survey (AHS)	1984-2009	Tract
National Longitudinal Survey (NLS) – Young/Mature Women	1966-2003	Lat/Lon Block Group
National Longitudinal Survey (NLS) – Young/Old Men	1966-1990	County
National Longitudinal Mortality Study (NLMS)	1973-2002	County

# American Community Survey (ACS) (1996-2009)

Replaced the decennial long form going forward.

The questions asked include age, race, sex, educational attainment, income, place of work, occupation, household relationships, housing unit characteristics, etc.

RDC data include tract, school and congressional district, birthday, migration place code, place of work tract code..

Questionnaires

[http://www.census.gov/acs/www/methodology/questionnaire\\_archive/](http://www.census.gov/acs/www/methodology/questionnaire_archive/)

# Top-coded 2006 Public Use ACS

State	Obs (000)	Age Top Code	Wage Top Code	Mortgage Top Code
AL	47	93	307,000	4300
FL	185	94	380,000	6900
GA	92	91	351,000	5500
SC	44	92	332,000	5400
TN	61	92	330,000	5000

RDC 2006 ACS age and mortgage expenditures are not top-coded, wages are top-coded at one million dollars.

# 2006 ACS PUMS Top-coded Variables

Housing Variables		Person Variables
Electricity	Property tax	Wages, salary income
Gas	Insurance	Self-employment income
Water	Mortgage	Interest, dividends
Heating fuel	2 <sup>nd</sup> mortgage	Retirement income
Condo fee		All other income
Rent		Age
Mobile home costs		Travel time to work

# Decennial Census – Long Form (1970-2000)

It is a 1-in-6 sample of all households.

The questions asked include age, race, sex, educational attainment, income, place of work, occupation, household relationships, housing unit characteristics, etc.

<http://www.census.gov/dmd/www/2000quest.html>

<http://www.census.gov/prod/1/90dec/cph4/appdxe.pdf>

# RDC version of Decennial – part 1

Detailed geography on residence and place of work

Bayer, Ross, and Topa (JPE, 2008) “Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes”

Use census block of residence and census block of work to look for social hiring networks.

They find a significant effect of social networks on hiring, especially among those with similar socio-demographic characteristics.

# RDC version of Decennial – part 2

Additional Variable on birth date

Angrist and Chen (AEJ: Applied Economics, 2011)  
“Schooling and Vietnam-Era GI Bill: Evidence  
from the Draft Lottery”

Use census birth date to link to Vietnam draft lottery  
dates to use as an instrument for military service.

Long run consequence of Vietnam service were  
found to be schooling gains and no earnings  
losses.

# RDC version of Decennial – part 3

Detailed geography – merging in outside data

Lucas Davis “The Effect of Power Plants on Local Housing Prices and Rents,” *Review of Economics and Statistics* forthcoming

Using census block, he merged in data on the location of waste incinerators, coal-burning plants, and nuclear power plants.

Neighborhoods within two miles of plants experienced a 3-7 percent decrease in housing values and rents.

# RDC version of Decennial – part 4

## Additional variables

McKinnish, Walsh and White (JUE 2010) “Who Gentrifies Low-Income Neighborhoods?”

Used whether the household moved into the housing unit in the past 10 years, and they used whether the household moved within the same county within the last five years.

They find that the gentrification of urban neighborhoods during the 1990s was not consistent with displacement and harm to minority households.

# Current Population Survey (CPS) (1967-2008)

General description – We have the March Supplement  
(aka - Annual Social and Economic Supplement)

Collects data concerning work experience, several sources of income, migration, household composition, health insurance coverage, and receipt of non-cash benefits.

[http://www.nber.org/data/mare\\_winship.html](http://www.nber.org/data/mare_winship.html)

[http://www.census.gov/aprd/techdoc/cps/CPS\\_Interviewing\\_Manual\\_July2008rv.pdf](http://www.census.gov/aprd/techdoc/cps/CPS_Interviewing_Manual_July2008rv.pdf)

# RDC version of the CPS – part 1

## Less top-coding

Burkhauser, Feng, and Jenkins "Using a P90/P10 Ratio to Measure Inequality Trends with the Public Use Current Population Survey: A View from Inside the Census Bureau Vaults." *The Review of Income and Wealth* (March 2009).

Changes in the top-codes for earnings led previous research to use the 90/10 ratio to measure inequality, but they find that this does not obviate the effects of the top-code changes.

# Survey of Income and Program Participation (SIPP) (1984-2008)

The main objective of SIPP is to provide accurate and comprehensive information about income and program participation, and about the principal determinants of income and program participation.

SIPP offers detailed information on cash and non-cash income on a sub-annual basis. The survey also collects data on taxes, assets, liabilities, and participation in government transfer programs.

# RDC version of the SIPP

Additional variables – the SIPP has topical modules that include detailed information on special topics. The topical modules are too varied to name them all.

- Retirement and pension plan coverage
- Child care
- Medical expenses

[http://www.census.gov/sipp/top\\_mod/top\\_mods\\_chart.html](http://www.census.gov/sipp/top_mod/top_mods_chart.html)

<http://www.census.gov/sipp/questionnaires.html>

## RDC version of the SIPP – part 2

Detailed geography merging in context variables

Fitzgerald and Ribar (Demography, 2004) “Welfare Reform and Female Headship.”

Used county geography to link in data on local labor market and marriage market conditions

Found no consistent effect of welfare reform waivers and TANF on women’s propensity to become unmarried heads of families.

# RDC version of the SIPP – part 3

Additional variables link to SSA earnings records

Feldstein and Liebman (NBER, 2002) “The Distributional Effects of an Investment-Based Social Security System.”

Used SSA lifetime earnings histories for a cohort born 1925-1929 linked to SIPP to simulate redistribution aspects of changes to Social Security

Show that an investment based system doesn't necessarily threaten the progressivity of SS.

# Survey of Program Dynamics

It was created in response to the 1996 welfare reform law (PRWORA).

Households from the 1992-3 SIPP panel were followed from 1997-2002.

Core data are collected on employment, income, program participation, health insurance and utilization, child well-being, marital relationships, and parents' depression. The SPD also has topical modules that vary by year.

# American Housing Survey (AHS) (1984-2009)

Collects data on housing (characteristics of homes, apartments, mobile homes, vacant units, etc.) and households (e.g. income, mobility, housing size, quality and costs).

National data in odd years, and 47 metropolitan areas each every 4 years (12 / year)

<http://www.huduser.org/portal/datasets/ahs.html>

# RDC version of the AHS

## Detailed geography

Ioannides and Zabel (JUE 2008) " Interactions, neighborhood selection and housing demand."

Use a tract level neighborhood choice model to control for sorting, and the AHS neighborhood clusters sub-sample for data on neighbor houses.

Find evidence of an influence of neighbor demand on own housing structure demand

# National Longitudinal Mortality Study (NLMS)

The NLMS is a national mortality study for the purpose of studying the effects of differentials in demographic and socio-economic characteristics on mortality.

It uses cohorts from the CPS and matches in subsequent data on a mortality outcome, the underlying cause of death, and follow-up time until death. The sample is designed to be representative of the April 1, 1983 U.S. non-institutionalized population.

# RDC version of the NLMS

These data are just becoming available through the RDC network.

The RDC data will include more detailed geographical information, actual month and year of birth, and additional information not available in the public use data.

# National Longitudinal Surveys (NLS)

Consists of data sets for four original cohorts of National Longitudinal Surveys (NLS).

- The Young Women's survey includes women who were ages 14-24 when first interviewed in 1968. Stopped in 2003.
- The Mature Women's survey includes women who were ages 30-44 when first interviewed in 1967. Stopped in 2003.
- The Young Men's survey includes men who were ages 14-24 when first interviewed in 1966. Stopped in 1981
- The Older Men's survey includes men who were ages 45-59 when first interviewed in 1966. Stopped in 1990.

## NLS part 2

- These files add detailed geographic information to the public use files.
- Prior approval by BLS is required before Census will entertain proposals requesting access to the NLS. Interested researchers should consult:  
<http://www.bls.gov/nls/origcohortgeo.htm>

# More Information

- Links
  - Center for Economic Studies  
<http://www.ces.census.gov/>
  - Atlanta Census Research Data Center  
<http://aysps.gsu.edu/acrdc/>