

Q&A

 Please submit all questions concerning webinar content through the Q&A panel.

Reminder:

- If you have participants watching this webinar at your site, please collect their names and emails.
 - We will be distributing a Q&A document in about one week. This document will fully answer questions asked during the webinar and will contain any corrections that we may discover after the webinar.



NAACCR Webinar
Topics in GIS – July 11, 2013
Using Census 2010/American
Community Survey Data for Cancer
Surveillance

Overview

Faculty

- Dr. Frank Boscoe
 - New York State Cancer Registry
- Dr. Dan Goldberg
 - Texas A & M University
- Recinda Sherman
 - Florida Cancer Data System
- David Stinchcomb
 - Westat

5

NAACCR

Agenda

- Census 2010 David
- ACS Recinda

15 minute break

- Denominator Data and Cancer Rates Dan, All
- Accessing Census Data David
 - NAACCR Tools/Datasets

6

Questions

During Webinar:

- Please send to Q&A as they come up
- Will be answered at end of each section or end of session

Post Webinar:

Dan: <u>daniel.goldberg@tamu.edu</u>
David: <u>DavidStinchcomb@westat.com</u>

Frank: <u>fpb01@health.state.ny.us</u>
Recinda: <u>rsherman@med.miami.edu</u>

7

NAACCR

GIS Committee



Census Work Group user forum, resource provider

8

NAACCR 2013 Topics in GIS Webinar

Census 2010

Dave Stinchcomb, Westat, Inc.



Census - Background

- Every 10 years for congressional apportionment
- Prior to 2010:
 - Short form "SF1" & "SF2" (Summary Files 1 & 2)
 - 100% of population, complete enumeration
 - · Basic demographics: age, sex, race/ethnicity
 - By household: relationships, tenure (rent/own)
 - Long form "SF3" & "SF4" (Summary Files 3 & 4)
 - 1 in 6 households (about a 16% sample) in 2000
 - Income, education, occupation, household characteristics, languages spoken, residential history, commuting, etc.
- Data released:
 - Aggregated counts by census geographic area
 - A Public-Use Microdata Sample (PUMS) of long-form answers

10

Census 2010

- Census 2010 short-form only
 - 100% of population, complete enumeration
 - Basic demographics: age, sex, race/ethnicity
 - By household: relationships, tenure (rent/own)
 - As of April 1, 2010
- No more long-form transition to annual American Community Survey (ACS)
 - Began in 1995, fully implemented in 2005
 - Recinda will have more on the ACS in the next section

11

NAACCR

Data Products and Release Schedule

- Redistricting Data
 - Basic counts for redistricting (no age breakdowns)
 - Released by state February-March 2011
- Demographic Profile
 - Tables of selected population & housing characteristics
 - Released by state May-August 2011
- Summary File 1 (SF 1)
 - Detailed data with 63 race/ethnicity categories
 - Released by state June-August 2011
- Summary File 2 (SF 2)
 - 331 detailed race and Hispanic or Latino subgroups. For example:
 - 114 American Indian and Alaska Native categories
 - 47 Asian categories; 43 Native Hawaiian and Other Pacific Islander categories
 - Released by state December 2011 April 2012
- Etc.

12

http://www.census.gov/population/www/cen2010/glance/

Census 2010 – SF1 Data Tables

• There are 771 SF1 data tables

	Population Tables Available Down to the Block Level			Housing Tables Available Down to Block Level
P1.	TOTAL POPULATION	H	11.	HOUSING UNITS
P2.	URBAN AND RURAL		12.	URBAN AND RURAL
P3.	RACE		13.	OCCUPANCY STATUS
P4.	HISPANIC OR LATINO ORIGIN		14.	TENURE
P5.	HISPANIC OR LATINO ORIGIN BY RACE		15.	VACANCY STATUS
P6.	RACE (TOTAL RACES TALLIED)		16.	RACE OF HOUSEHOLDER
P7.	HISPANIC OR LATINO ORIGIN BY RACE (TOTAL RACES TALLIED)		17.	HISPANIC OR LATING ORIGIN OF HOUSEHOLDE
P8.	RACE		11.	HISTANIC ON EXTINO ONIOIN OF HOUSEHOLDS
P9.	HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE			
P10.	RACE FOR THE POPULATION 18 YEARS AND OVER			
P11	HISPANIC OR LATINO AND NOT HISPANIC OR LATINO BY RACE F	n		
	Population Tables Available Down to the Census Tract Level			
PCT1.	AMERICAN INDIAN AND ALASKA NATIVE ALONE WITH ONE TRIBE	REPORTE	D FOR S	BELECTED TRIBES
PCT2.	AMERICAN INDIAN AND ALASKA NATIVE ALONE WITH ONE OR MC	RE TRIBE	S REPO	RTED FOR SELECTED TF
PCT3.	AMERICAN INDIAN AND ALASKA NATIVE ALONE OR IN COMBINATION WITH ONE OR MORE OTHER RACES AN			
PCT4.	AMERICAN INDIAN AND ALASKA NATIVE ALONE OR IN COMBINATION WITH ONE OR MORE OTHER RACES			
PCT5.	ASIAN ALONE WITH ONE ASIAN CATEGORY FOR SELECTED GRO	UPS		
PCT6.	ASIAN ALONE WITH ONE OR MORE ASIAN CATEGORIES FOR SELECTED GROUPS			
PCT7	ASIAN ALONE OR IN COMBINATION WITH ONE OR MORE OTHER	RACES AN	ID WITH	LONE OR MORE ASIAN CA
			Hous	ing Tables Available Down to the Census Tract Level
	Population Tables Available Down to the County Level	HCT1.	TENU	JRE BY HISPANIC OR LATINO ORIGIN OF HOUSEHOLDER I
			TENI	JRE BY PRESENCE AND AGE OF OWN CHILDREN
	GROUP QUARTERS POPULATION BY SEX BY AGE	HCT2.		
PC02.	GROUP QUARTERS POPULATION BY SEX BY AGE GROUP QUARTERS POPULATION IN INSTITUTIONAL FACILITIE	HCT2. HCT3.		JRE BY PRESENCE AND AGE OF RELATED CHILDREN
PC02. PC03.	GROUP QUARTERS POPULATION BY SEX BY AGE GROUP QUARTERS POPULATION IN INSTITUTIONAL FACILITIE GROUP QUARTERS POPULATION IN CORRECTIONAL FACILITI		TENU	JRE BY PRESENCE AND AGE OF RELATED CHILDREN JRE BY PRESENCE AND AGE OF PEOPLE UNDER 18 YEAR
PC02. PC03. PC04.	GROUP QUARTERS POPULATION BY SEX BY AGE GROUP QUARTERS POPULATION IN INSTITUTIONAL FACILITIE GROUP QUARTERS POPULATION IN CORRECTIONAL FACILITI GROUP QUARTERS POPULATION IN JUVENILE FACILITIES BY	нстз.	TENU	
PC02. PC03. PC04. PC05.	GROUP QUARTERS POPULATION BY SEX BY AGE GROUP QUARTERS POPULATION IN INSTITUTIONAL FACILITIE GROUP QUARTERS POPULATION IN CORRECTIONAL FACILITI GROUP QUARTERS POPULATION IN JUVENILE FACILITIES BY GROUP QUARTERS POPULATION IN NURSING FACILITIES ISKI	нстз.	TENU	
PC02. PC03. PC04. PC05. PC06.	GROUP QUARTERS POPULATION BY SEX BY AGE GROUP QUARTERS POPULATION IN INSTITUTIONAL FACILITIE GROUP QUARTERS POPULATION IN CORRECTIONAL FACILITI GROUP QUARTERS POPULATION IN JUVENILE FACILITIES BY GROUP QUARTERS POPULATION IN NURSING FACILITIES/SKI GROUP QUARTERS POPULATION IN OTHER INSTITUTIONAL F.	нстз.	TENU	
PC01. PC02. PC03. PC04. PC05. PC06.	GROUP QUARTERS POPULATION BY SEX BY AGE GROUP QUARTERS POPULATION IN INSTITUTIONAL FACILITIE GROUP QUARTERS POPULATION IN CORRECTIONAL FACILITI GROUP QUARTERS POPULATION IN JUVENILE FACILITIES BY GROUP QUARTERS POPULATION IN NURSING FACILITIES ISKI	нстз.	TENU	

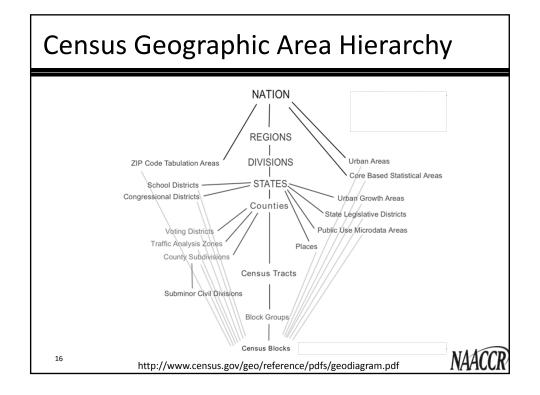
Census Boundaries

14

Census Geographic Area Definitions

- Area boundaries are updated every 10 years
 - "Using Census 2000 geographies"
- First step in the enumeration process
- Design goals for census tracts:
 - Ideally, about 4,000 people
 - Required thresholds:
 - Population: 1,200 minimum; 8,000 maximum
 Housing units: 480 minimum; 1,600 maximum
 - Relatively homogeneous neighborhood
- Done in consultation with local officials
- Effort to limit changes from previous boundaries

15



Summary Levels

- A three digit code to indicate the geographic level of detail
- For example:
 - 040: State
 - 050: State-County
 - 140: State-County-Census Tract
 - » 150: State-County-Census Tract-Block Group
- Useful in American FactFinder as well as for custom processing of Census data files

17

NAACCR

Census Area Codes (FIPS/ANSI codes)

- Each census area is identified with a numeric code
 - States, counties, and Census places use national standards
 - Formerly the Federal Information Processing Standards (FIPS) from National Institute of Standards and Technology (NIST)
 - Converting to American National Standards Institute (ANSI) for future maintenance of place codes
- Number of digits for the main census hierarchy:
 - State (2), county (3), tract (6), block group (1), block (3)
- Census tract code has an implied 2 decimals
 - 001234 = "Tract 12.34", 123400 = "Tract 1234"
 - Source of coding errors in NAACCR records
- Fully-qualified census tract: 11 digits

18

Urban and Rural Areas

- Census defines urban areas:
 - Densely developed; includes residential, commercial, and other non-residential urban land uses
 - Two sizes:
 - Urbanized Areas (UAs): 50,000 or more people;
 - Urban Clusters (UCs): at least 2,500 and less than 50,000
 - A collection of contiguous tracts and blocks
- Updated every 10 years, after the decennial census
 - Census 2010 totals announced in March 2012
 - The Census 2010 SF1 Urban/Rural Update released in September 2012

 P2. URBAN AND RURAL [6] (Urban/Rural Update only)

URBAN AND RURAL [6] (Urban/Rural Update only)
Universe: Total population
Total:
Urban:
Inside urbanized areas
Inside urban clusters

19

http://www.census.gov/geo/reference/ua/urban-rural-2010.html



Census 2010 Urban Areas Urbanized Areas and Urban Clusters: 2010 Urbanized Areas and Urban Clusters: 2010

Census Tract Relationship Files

- Files showing how 2000 census tracts relate to 2010 census tracts
- One record for each part of a 2000 census tract/2010 census tract
 - The area that is uniquely shared between a 2000 census tract and a 2010 census tract
 - A part of a census tract (a piece of land)
 - Intersection of the two polygon layers
 - Unique identifier: the 2000 CT and the 2010 CT
- Relationship files have:
 - Population/housing units (HUs) of the piece of land (the record)
 - Population/HUs of the whole 2000 and 2010 CTs
 - All population/HUs are from 2010 census
- Useful for converting data between Census geographies

21

http://www.census.gov/geo/maps-data/data/tract_rel_layout.html

NAACCR

Data Conversions

- Count data (e.g. number of cancer cases)
 - The New CT value: the sum of the values in the Old CT pieces multiplied by the proportion of the Old CT

$$V_{\text{NewCT}} = \sum_{i=1}^{n} \left(V_{\text{OldCT}_i} \times \text{Pct}_{\text{OldCT}_i} / 100 \right)$$

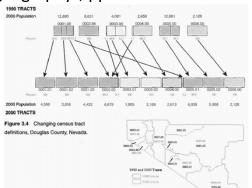
- Rate data (e.g. percent unemployed)
 - The New CT value: the average of the values in the Old CT pieces weighted by the proportion of the New CT

$$V_{\text{NewCT}} = \sum_{i=1}^{n} \left(V_{\text{OldCT}_i} \times \text{Pct}_{\text{NewCT}_i} / 100 \right)$$

22

Additional Discussion and Examples

- Unlocking the Census with GIS by Alan Peters and Heather MacDonald (ESRI Press, 2004)
 - "Example one: Analyzing trends despite changing census geography", pp 115-120



NAACCR

American FactFinder Accessing Census 2010 Data

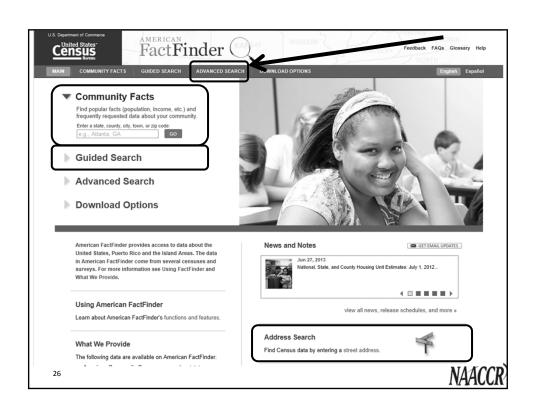
24

American FactFinder – Introduction

- Interactive Web interface to most Census Bureau data products
- Original "legacy" version has been retired
- The New American FactFinder FactFinder2
 - http://www.census.gov/
 - Data tab -> American FactFinder
 - http://factfinder2.census.gov/
 - Google "american factfinder" or "factfinder2"
- New features are being added quarterly

25









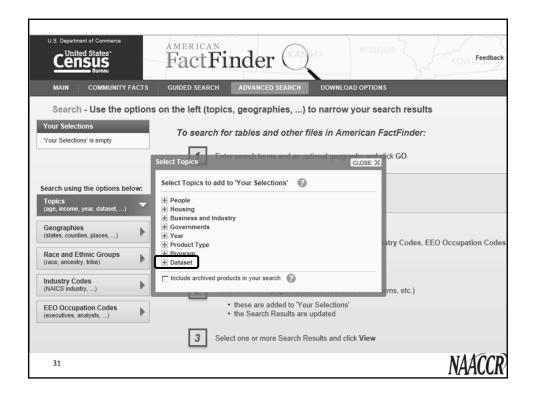
AFF Example – Census 2010 Data

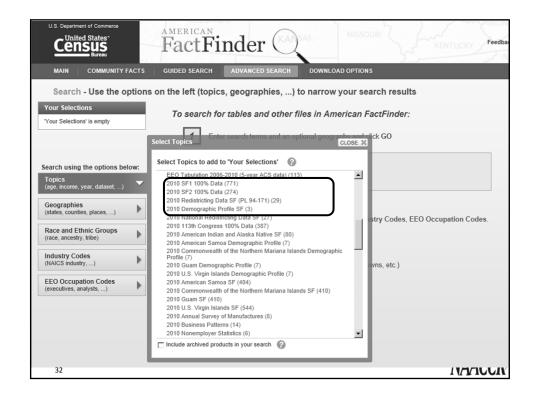
- Hispanic populations
- For Oregon census tracts
- By age and sex
- Download a table for analysis

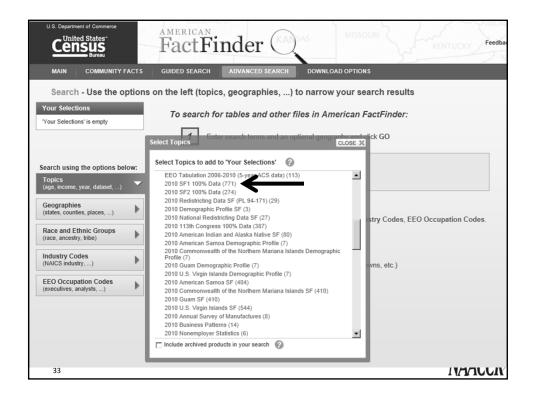
29

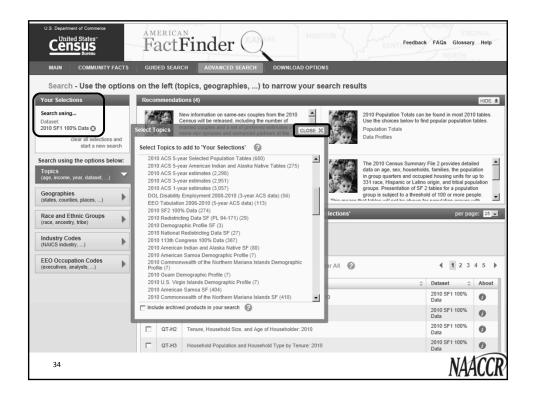






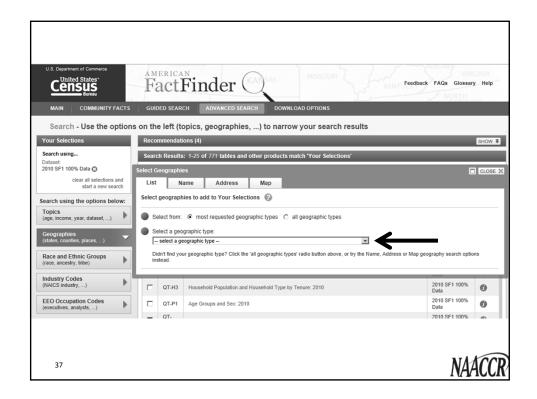


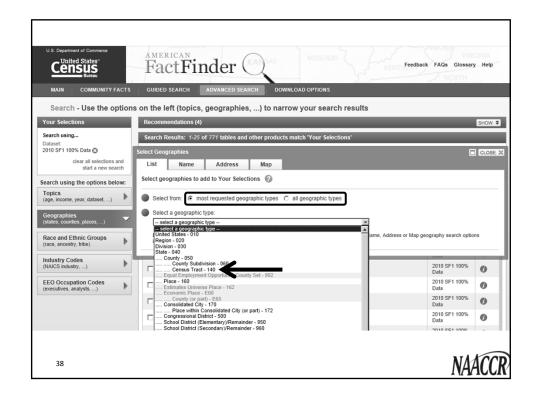


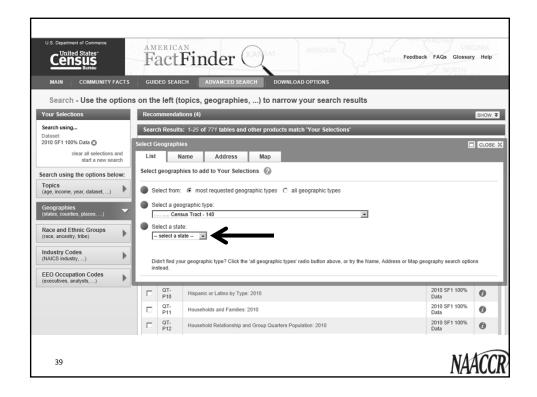


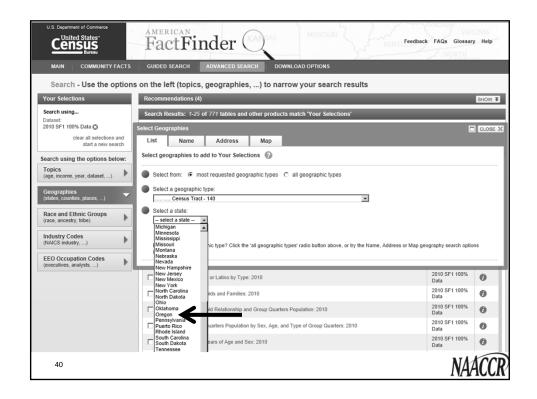


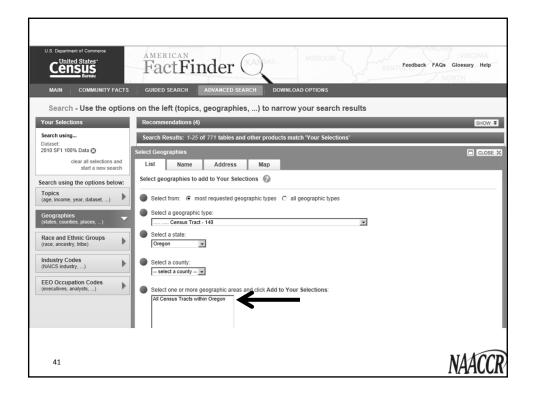


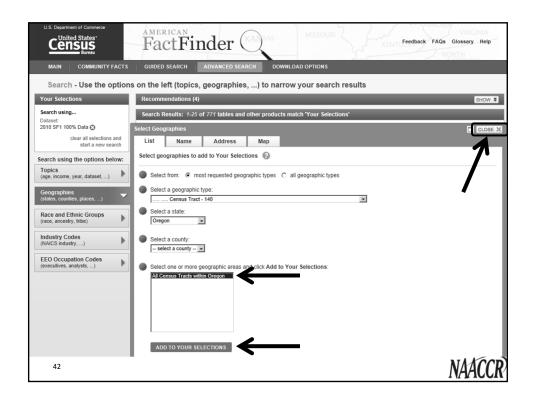


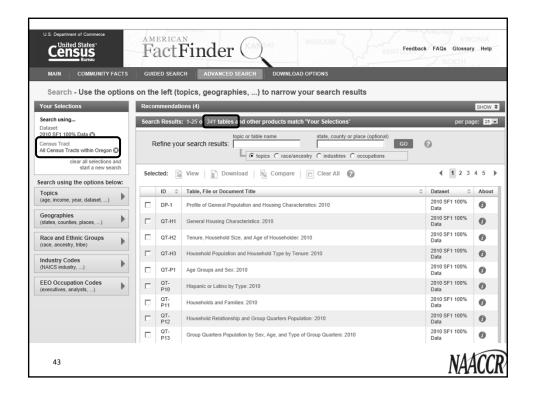


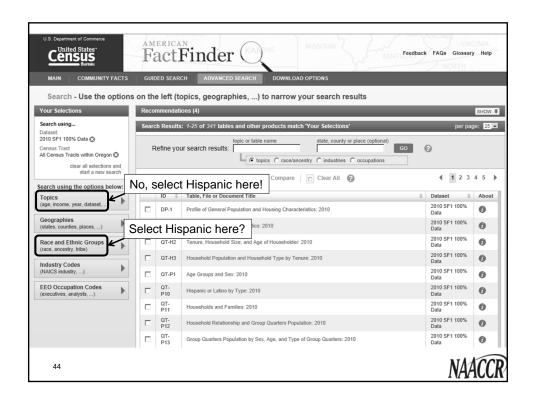


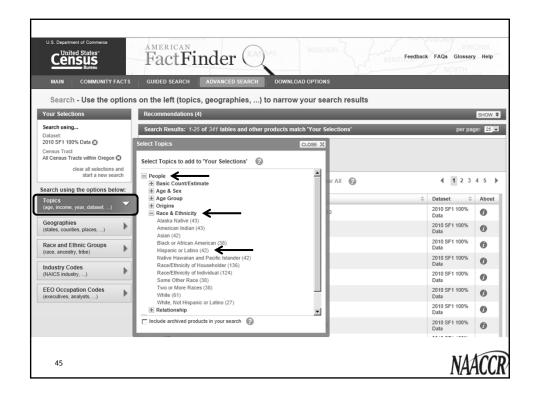


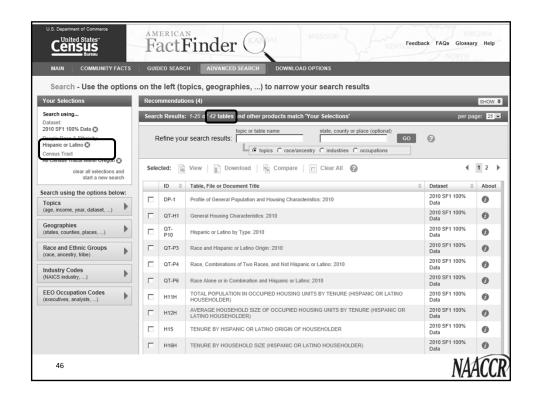


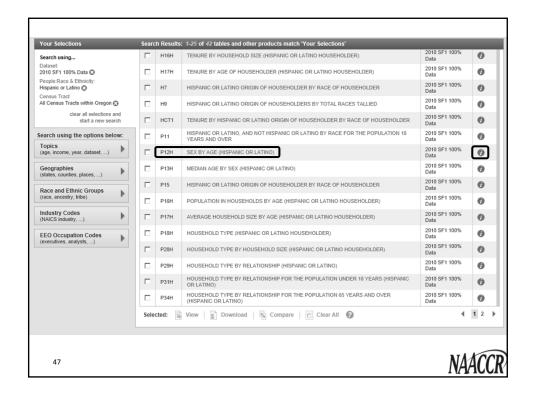


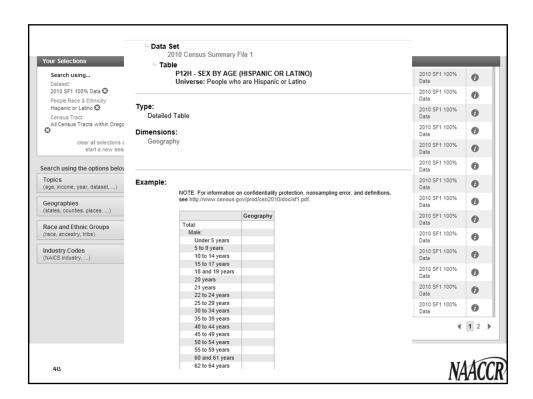


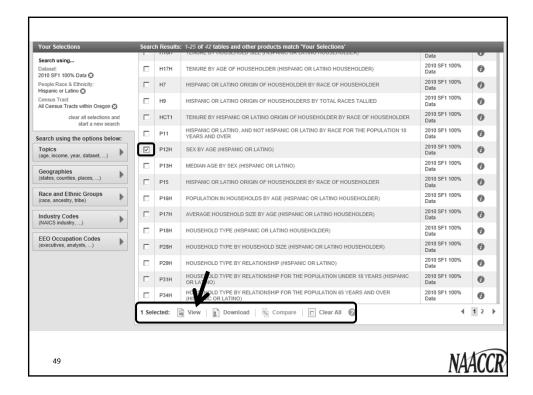


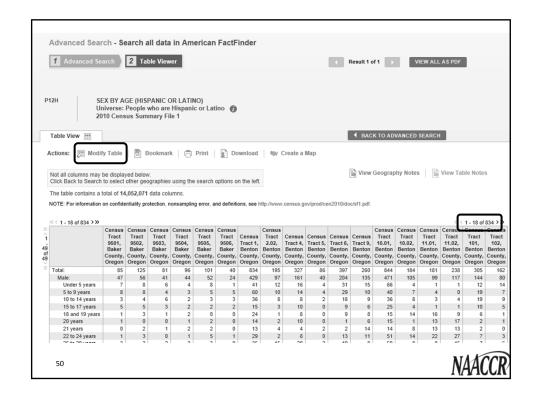


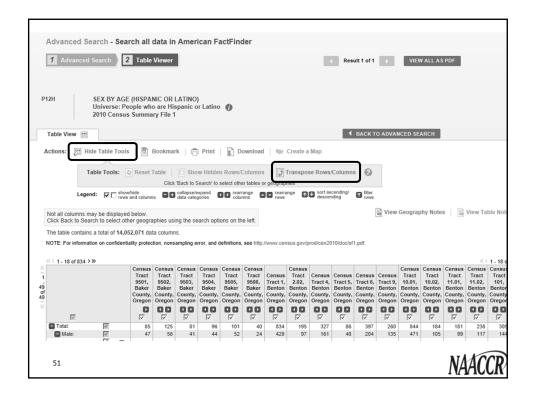


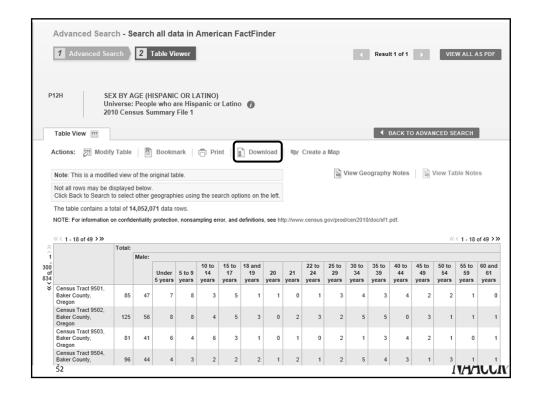


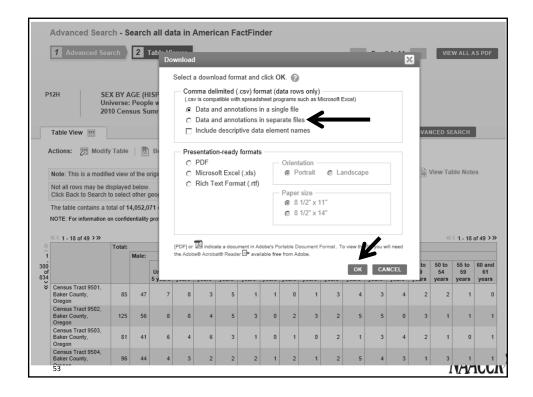


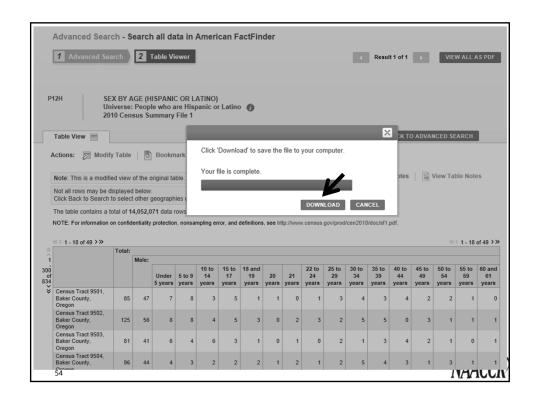


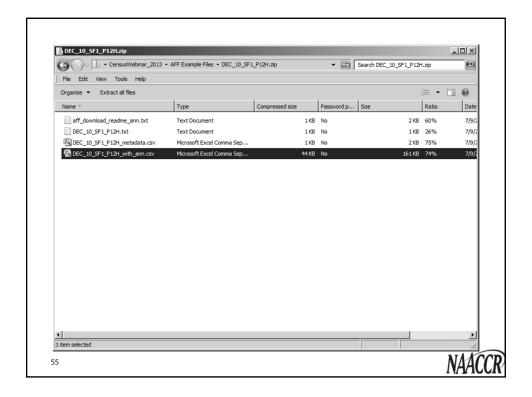


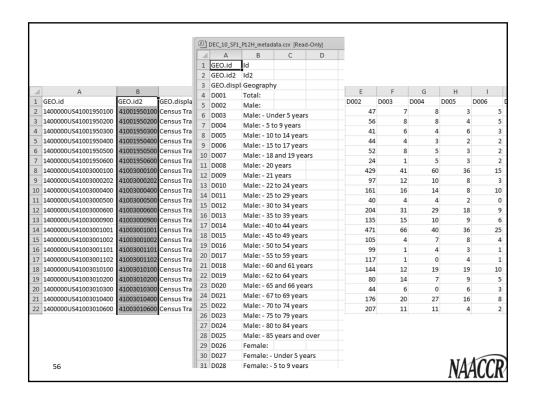












AFF Census 2010 - Exercises

- Your turn
 - Urban/rural populations for your county's census tracts
 - Other data of your choice (households, group quarters populations, ...)
 - Total population by age and sex for all Oregon block groups (challenging)
 - Hint: Geographies, Name tab, Summary Level, Groups of geographies

NAACCR

QUESTIONS?

58

NAACCR Webinar

Topics in GIS – July 11, 2013

Using Census 2010/American

Community Survey Data for Cancer

Surveillance

American Community Survey

Recinda Sherman, Florida Cancer Data System



Purpose

- GIS/Geospatial analysis
 - "Novel" ways to communicate & interpret surveillance data
 - Spatial link to external datasets
- SES data are integral to cancer surveillance
 - Describe disparities for cancer control
 - Hypothesize causality
 - Prioritize resources, inform policy, target intervention
 - Evaluation
- Is it spurious?



Sources: US Census

- Website: http://www.census.gov/acs/www/
 - Tutorials, Technical Documentation
 - What Researchers Need to Know
 - http://www.census.gov/acs/www/Downloads/handbooks/ACSResearch.pdf
 - Design and Methodology
 - <a href="http://www.census.gov/acs/www/methodology/m
- Personal communication
 - General 1-800-923-8282
 - Questionnaire 1-800-354-7271





NAACCE

61

Overview

- Overview of ACS
- Difference from SF3
 - Variables
 - Period versus Fixed Estimates
 - Sample Design
 - Population Counted
 - Sampling Error
- Access/Availability
 - AFF2 Features, DataFerrett
- Recommendations

62

Overview: ACS

- American Community Survey
- Estimates not counts
 - Description of population
 - Social, demographic, housing, economic characteristics
- Replaces Summary File 3
 - Rolling survey
 - Puerto Rico Community Survey
 - "essentially" same information
 - 55 individual q's, 30 household
- Annual estimates
 - Single year (65,000+), 3 year (20,000+), 5 year(all geographies)

63

NAACCR

ACS History

- Proposed in 1986
- 1994 2004 demonstration period
- Fully implemented in 2005
- Replaced SF3 in 2010
- Goal
 - More timely & frequent data
 - Streamline operations & improve data quality

64

Variables

- Demographic characteristics
 - Same as Census 2010
 - Sex, race, age, Hispanic origin
- Social characteristics
 - Education, marital status, fertility, grandparent caregivers, veterans, disability status, disability status, place of birth, citizenship, year of entry, language spoken at home, ancestry/tribal affiliation
- Economic characteristics
 - Income, benefits, employment status, occupation, industry, commuting to work, place of work
 - Income not compatible with SF3
- Housing characteristics
 - Tenure, occupancy & structure, housing value, taxes & insurance, utilities, mortgage/rent

65



Period vs Fixed Estimates

- ACS Period Estimate
 - Average over specific time period
 - 1, 3 & 5 year averages
 - Updated annually
 - Changes in questions and circumstances
 - Economic indicators
- SF3 Fixed Estimate
 - April 1, XXX9
 - Although collected over several months

66



Continuous versus point-in-time

- No longer "snapshot" now an abstract picture
 - Averaged data
- Median income
 - Averaged throughout year & adjusted for inflation
- Permanent, professional staff
 - Smaller, cost efficient
 - Highly trained
- Same number of households sampled annually
 - Smaller percentage each year



Continuous versus point-in-time

- Reference period changes throughout year
 - Seasonal or rapid trends
- Multiyear estimates are *not* average of single year
 - Adjusted for age, sex, race, & Hispanic origin
 - More complex weighting scheme
 - More precise, more reliable

Sample Design

- Continuous sampling
 - 250,000 addresses monthly; every county
- 3 million addresses annually
 - 2.3% of 2005 US population
 - Roughly same at state level
 - However, sample size not met
 - Closer to 2 million
- Mail, phone, personal visit
- Unique address once per 5 years
- Numerous geographic levels
 - Single year
 - States, congressional districts, 700 counties
 - 5 year aggregates: all geographies

69

NAACCR

Sample design

- Sent to address
- Phone f/u for non-responders
- Interviewer sent 3 months later
 - subsample
- Oversample: small areas (county, towns, school districts), AI/AN, and tracts expected to have low response rate (-8% in tracts expected to have high response rate)
- Highly complex design

Processing

- Monthly coding for write-ins
 - i.e. occupation, ancestry
- Annual completeness assessment
 - Accept/reject
- Annual automated editing
 - Hot-deck imputation for missing values
 - 100% "complete"
 - Dollar amounts adjusted for inflation (adjusted to single year or most current)
 - 2005-2009 adjusted to 2009



Weighting

- Multiple step
- Base weight(s)
 - based on % of total pop and % of f/u for in-person interviews
- Adjusted for nonresponse
 - Based on % of occupied housing units
- Large or group county weighting
 - Adjusted for missing addresses
 - Adjusted for missed persons in household
- 3 year, 5 year additional adjustments
 - On pooled data

Sampling error

- Sample to lower cost
- 2000 SF3 also sample
 - 1 in 6 households
 - Results from 1 household represent 5 additional
- 2010 ACS
 - 1 in 8 (5 year estimate); 1 in 40 (single-year estimate)
 - Less precise
- SF3 errors significantly smaller, published for table not data item



Sampling Error

- Uncertainty of data based on sample versus full population
- Used to show reliability and usability, compare estimates, conduct statistical tests, inform conclusions
- Margin of Error is reported for each data point
 - MOE measures the precision of the estimate at given level of confidence
 - 90% CI (90% likelihood that difference between population and sample is less than or equal to MOE)
 - 95% etc can be calculated (Compass Handbook)
 - SE = MOE/1.645; CI = Estimate ±MOE
- How do we account for the MOE in analysis, presentation, interpretation?

Consistency of response

- 800,000 address test 2000
 - Results compared to 2000 SF3
- Race/Ethnicity—more accurate
- Ancestry more accurate
- Disability more accurate
- Employment more accurate
- Renter possibly more accurate
- Vacancy unknown
- Median Income unknown
- Poverty mixed



Population "Counted"

- SF3
 - EVERYONE in US
 - plus overseas federal employees & family
 - Usually living at address
- ACS
 - 2005 "everyone" in US
 - EXCEPT homeless, shelters, transitory locations
 - · Young & elderly
 - 2006 EVERYONE
 - Are living at address for 2+ months
 - Legal/citizenship status irrelevant

76

Content changes

- 2006
 - Include group quarters
- 2008
 - Additions
 - health insurance coverage, veteran's service-connected disability, marital history
 - Deletions
 - Time & reason for staying at address
- Wording/Format
 - Changes from SF3, changes over time

77

NAACCR

Strengths

- Timeliness
 - Released 8-10 months after collection
 - 2 years SF3; bg level longer
- Frequency
 - Annual versus decennial
- Higher data quality
 - Higher training level for interviewers
 - Higher response rate
 - Computer-assisted follow-up and personal interview for nonresponse by mail
 - More complete
 - f/u on missing data via phone or visit

Weakness

- Less precise
 - 12.5% versus 17%
 - Census/bg level highly impacted
- Significantly larger MOE
 - Smaller sample size
 - 3 million addresses a year
 - 16 million SF3
 - Weighting issue
 - In-person interview on subsample of households that do not respond by mail or phone
 - 15 million housing unit sample -> 10-11



Type of Product Outputs Available

- Similar to Decennial Census
- Profiles
 - Novice users, fact sheets
 - Tabular, textual, comparisons
 - Limited geography (65,000/500,000)
- Tables
 - Detailed, most relevant
 - 1,200 (11 race/ethnic groups)
 - Sex by age (for rates)
 - Poverty Status (for SES)
- Thematic Maps



Public Use Microdata Sample (PUMS)

- Subsample of individual records
 - Anonymized
- More subject detail
 - Detailed population group; Custom variables
- Less geographic detail
 - 100,000 threshold
- Documented data errors
 - Compared to tables
 - Age & Sex
 - 65+ 15% difference in counts
 - 65+ not representative
 - Byproduct of "misapplication of disclosure avoidance"
 - Data swapping, imputation

81



How to Access Data Products

- American FactFinder
- DataFerrett
 - Block Group Level for limited tables
- FTP
 - http://www2.census.gov/acs2009 5yr/summaryfile/
- D\/D
 - acso.dvd.order@census.gov
 - Comma delimited
- Summary File Retrieval Tool
 - Excel macro



Availability

- Release thresholds
 - 3 households
 - 50 respondents
- Population thresholds
 - 1-year 65,000+
 - 3-year 20,000+
 - 5-year all levels of geography
- ZCTA, census tracts, block groups
 - Only 5-year

83

NAACCR

Available Geographies

- Block group level
- Many, many tables suppressed
 - Reliability/Privacy concerns
- Appendix E
- The 2006-2010 5-Year Summary File Technical Documentation

84

Recommendations

- Use ACS to describe population characteristics
 - Not as denominator
- Don't use overlapping period estimates
 - If analyzing trend data, use 1-year estimates if possible
- Use 2006+
 - Elderly more likely to live in group quarters
- Examine MOE before analysis
 - Use multiyear data if available

85

NAACCR

Recommendations

- When using multiple years or trends
 - Review for any geographic boundary changes
 - Examine history of question
- Review item specific guides if comparing ACS to SF3
 - Compare, compare with caution, do not compare
 - "Guidance for Data Users" Tab on ACS website
- Review non-sampling error information to guide interpretation
 - Response Rates, Coverage, Imputation
 - Response and processing error

Detailed Tables for Quality Measures

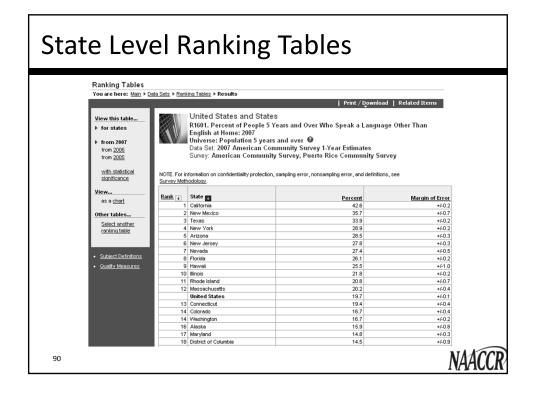
 Methodology; "Sample Size & Data Quality" Sample size, coverage rates, unit and item non-Choose a table selection method by subject by keyword show all tables Select one or more tables and click 'Add' numbers (explain table numbers) B98001. Unwieghted Housing Unit Sample B98002. Unweighted Group Quarters Population Sample)S B98011. Housing Unit Coverage Rate B98012. Total Population Coverage Rate by Sex B98013. Total Population Coverage Rate by Weighting Race and Hispanic or Latino Groups B98014. Group Quarters Population Coverage Rate B98021. Housing Unit Response and Nonresponse Rates with Reasons for Noninterviews B98022. Group Quarters Population Response and Nonresponse Rates with Reasons for Noninterviews B98031. Overall Person Characteristic Imputation Rate B98032. Overall Housing Unit Characteristic Imputation Rate Add ▼

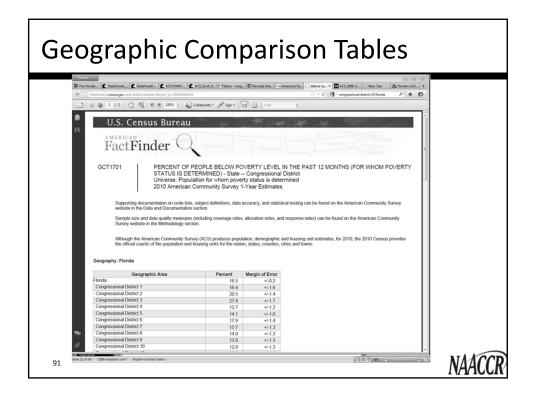
On-going Issues

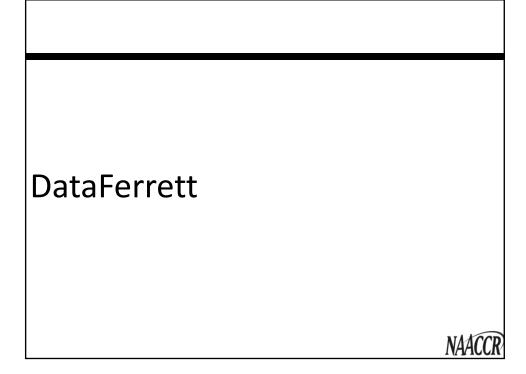
- Sample Size
- Cost & Politics
 - Census Survey Costs
 - 1950 \$2.50 per capita; 2000 \$16 per capita
 - Political target
 - Budget cutting, eliminate ACS & Economic Survey
- Continuous collection and annual changes

American Fact Finder: Neat Features









DataFerrett

- Data mining and extraction tool
 - Select data and recode variables as needed
 - Customize and export tables (ascii, SAS, SPSS, Excel/Access)
 - Create charts, graphs, maps
- Variety of datasets available
 - The ACS 5-year summary file is aggregated
 - Can't modify or recode
 - PUMS
- Can't download entire summary file
 - Limit on # of variables and size of extract

93

NAACCR

Configuration

- Must run Java
 - 1.6.0 3X or higher
 - http://javatester.org/
 - Increase JAVA applet memory size
 - Start > Control panel > JAVA (double click) > JAVA tab > Runtime Settings > in cell "Runtime Parameters" type – Xmx512m for each version
- Browsers
 - − IE, Mozilla, Google Chrome (PC), Safari (MAC)
 - Allow pop-ups

94

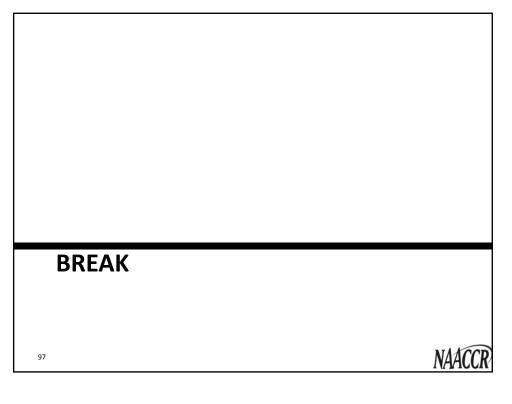
DataFerrett

- http://dataferrett.census.gov/run.html
- Supplement for ACS 5-year
 - http://dataferrett.census.gov/acs/ACS_5yrsf_UserGuide.pdf

95

NAACCI

Questions?	
96	NAACCR



NAACCR Webinar
Topics in GIS – July 11, 2013
Using Census 2010/American
Community Survey Data for Cancer
Surveillance:

Denominator Data and Cancer Rates: Applications and Complications

Frank Boscoe, New York State Cancer Registry Dan Goldberg, Texas A&M University

Outline

- 1. Available denominators
- 2. End of decade issues
- 3. Race bridging
- 4. Relating cancer and SES using subcounty pops
- 5. ACS margins of error

99

NAACCR

1. SEER*Stat populations

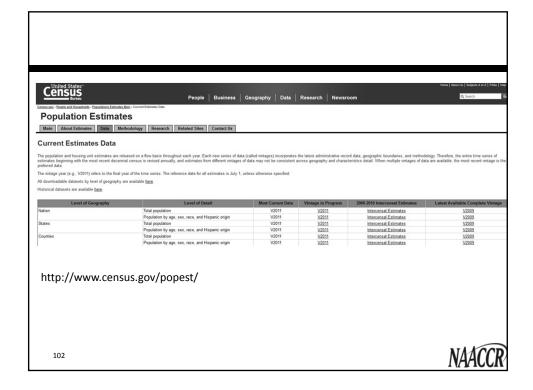
- Developed through a special arrangement with the Census Bureau, they are different than what is publicly available from the Census.
 - Shift reference date from April 15 to July 1
 - Single year of age
 - Bridged race groups (eliminates other & multiple race)
 - Special rules for Hawaii
 - Only to the county level

100

Census populations

- Years ending in zero (1990, 2000, 2010)
 - Much data to block/block group level
 - 7 race groups (including other & multiple)
 - Modified Race Data Summary file resolves "other" (2010 not yet released)
- For other years projections & estimates, to county level
 - 5 year age groups
 - 6 race groups ("other" removed)
- American Community Survey 5-year estimates, to block group level
 - 5 year age groups
 - 7 race groups





2. End of decade issues

- Affects all population sources (except arguably ACS)
- Population estimates are based on births (highly accurate), deaths (highly accurate), and migration (not very accurate)
 - Gets worse the further removed from a decadal census

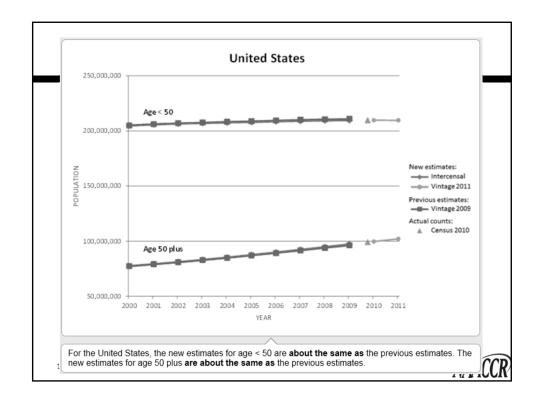
103

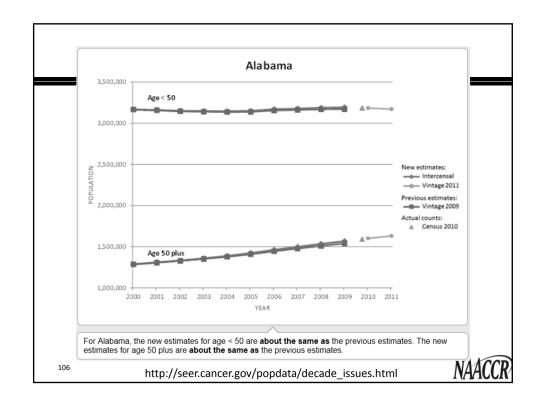


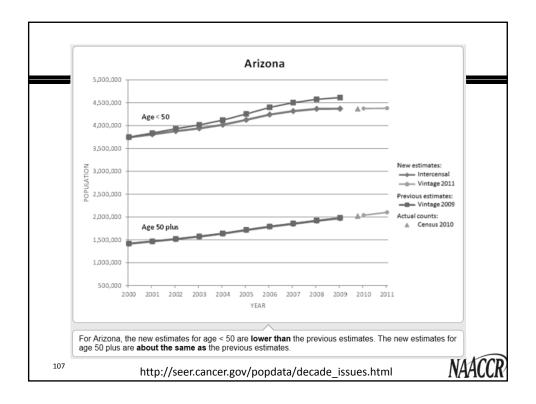
Table 1 Important Potential Sources of Bias in Administrative Records Used for Estimating Populations (after Judson, Popoff, and Batutis 2001, except for *)

County Characteristic	Direction of Bias	Reason for Bias
High growth rate	Underestimate	Some in-migrants are not captured by tax return matching process.
Negative growth rate	Overestimate	Some out-migrants are not captured by tax return matching process.
Lack of hospital in county	Underestimate	Births may be misassigned to the county in which the hospital is located
High % living on military bases & in college dorms	Underestimate	Lag in reporting of new quarters (but if quarters are being eliminated, then bias is opposite).
High % prisoners	Overestimate	Prisons have tendency to overreport.
High % over age 65 not enrolled in Medicare	Underestimate	These people are not captured in the Medicare database.
High % foreign born	Underestimate	Immigrants are less likely to be captured in administrative records.
High % poverty	Overestimate	Such counties tend to generate out-migrants less likely to have filed tax returns.
High % Native American	Underestimate	Native Americans are less likely to be captured in administrative records
Retirement destination*	Underestimate	Some 65 year old in-migrants are not captured.









Population Estimation Error and Its Impact on 1991–1999 Cancer Rates*

Francis P. Boscoe

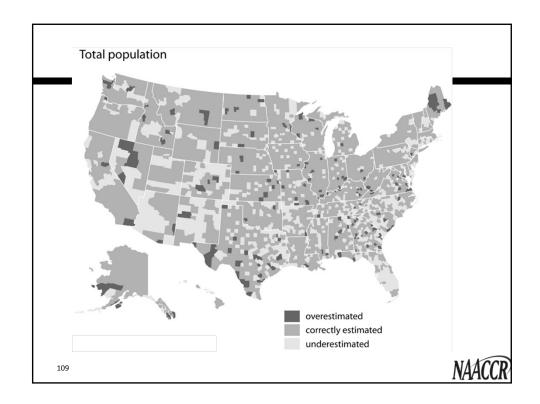
New York State Department of Health

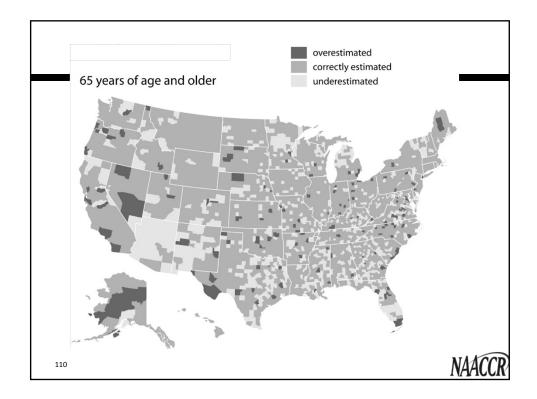
Barry A. Miller

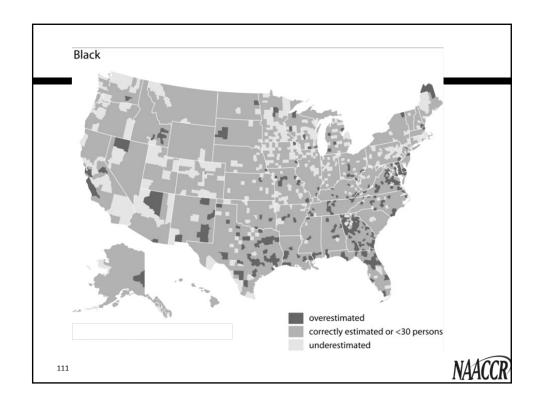
National Cancer Institute

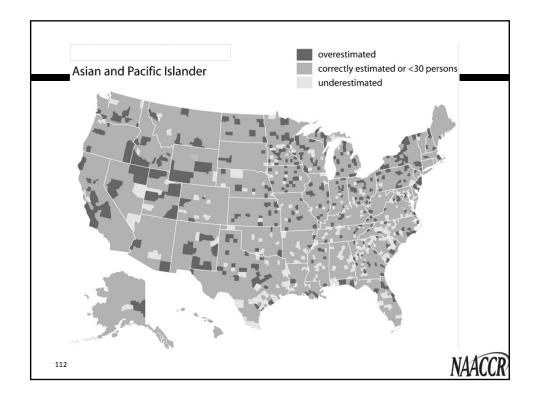
The Professional Geographer, 56(4) 2004, pages 516–529 © Copyright 2004 by Association of American Geographers. Initial submission, January 2003; revised submission, July 2003; final acceptance, September 2003. Published by Blackwell Publishing, 350 Main Street, Malden, MA 02148, and 9600 Garsington Road, Oxford OX4 2DQ, U.K.

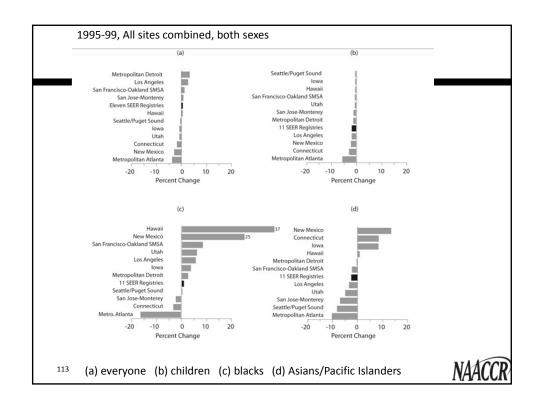


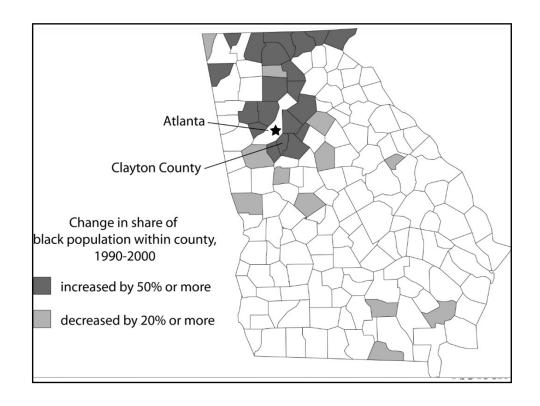


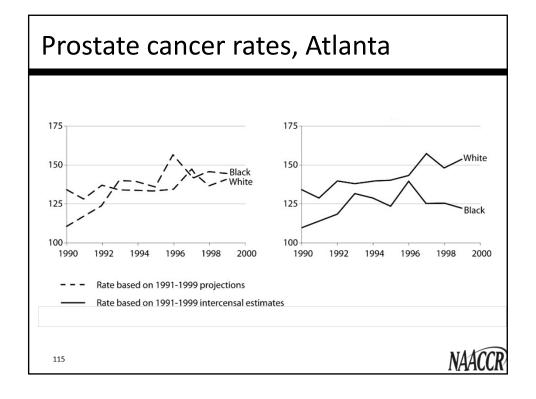


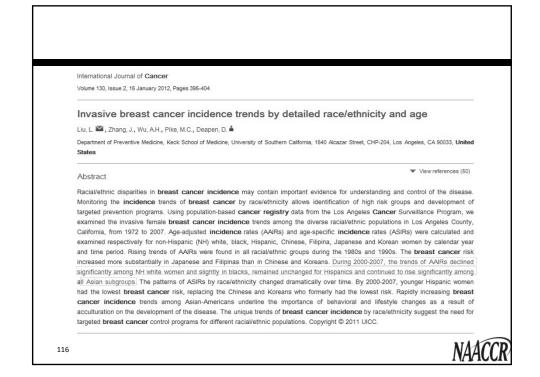


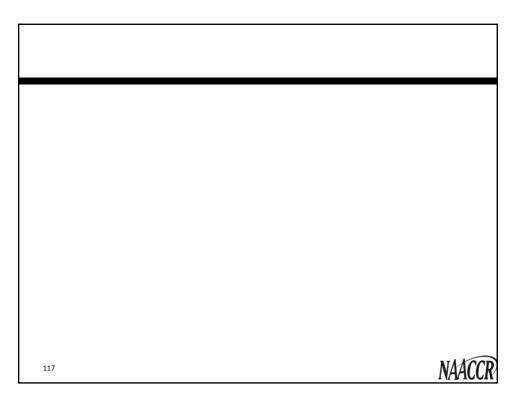












3. Race bridging

- Multiple Race in California (2010 census)
 - Imperial County 1.7%
 - Los Angeles County 2.7%
 - Sacramento County 5.6%
 - Solano County 6.5%

118

Building your own bridge

- Necessary for subcounty analysis
- "Other" race
 - Hispanic = white
 - Non-Hispanic = fractional allocation
- Multiple race
 - Fractional allocation
 - NCHS logit model

119

NAACCR

Fractional allocation

- Suppose in some census tract there are the following population counts by race:
 - White 500
 - Black 300
 - API 150
 - -AI/AN 50
 - Other 30
 - Multiple race 20

120

Fractional allocation

- White 500 (50%)
- Black 300 (30%)
- API 150 (15%)
- AI/AN 50 (5%)
- Other 30 (Hispanic 26, Non-Hispanic 4)
- Multiple 20

With this method, first assign the 26 Hispanic Other to white, of the remaining 24 assign 12 to white, 7 to black, 4 to API, 1 to AI/AN (in actuality, apply this separately to each age/sex strata)

121

NAACCR

Equal allocation

- White 500 (50%)
- Black 300 (30%)
- API 150 (15%)
- AI/AN 50 (5%)
- Other 30 (Hispanic 26, Non-Hispanic 4)
- Multiple 20

With this method, first assign the 26 Hispanic Other to white, of the remaining 24 assign 6 to white, 6 to black, 6 to API, 6 to AI/AN (in actuality, apply this separately to each age/sex strata)

122

NCHS logit model

- Based on the National Health Interview Survey, which asks multiple-race respondents to choose the single race with which they most closely identify
- A separate model was developed for each multirace combination (e.g. white and black, Asian and American Indian)

123

NAACCR

NCHS logit model

- Other variables used in the model:
 - Age
 - Hispanic origin
 - Sex
 - Region (NE, Midwest, South, West)
 - Urbanization
 - Percent of each race group in the county
 - Percent multiple race in the county

124

NCHS logit model

Health Interview Survey 1997–2000

	Multiple-race group and predicted primary race						
	AIAN¹/Black	AIAN/White	API ² /Black ³	API/White	Black/White	AIAN/Bla	ck/White
Covariates	Black	AIAN	Black	API	Black	AIAN	Black
Age (in years) per 10 years	-0.05461	*-0.08968	0.05669	0.09568	0.05532	0.26212	*0.36140
Hispanic origin (yes)	*-1.92602	*0.88834	-0.10458	0.19303	-0.52253	0.35986	-0.83526
Sex (male)	-0.12359	0.00972	0.33642	0.01393	0.11948	-0.43898	0.50777
Region (West reference)							
Northeast	-0.88349	0.21233	-0.45997	-0.05520	-0.25363	*-4.53976	#-3.45593
Midwest	*-1.70126	0.09144	*-3.92403	-0.06453	0.17140	#-3.82328	#-3.79144
South	-0.97935	-0.28494	-1.48264	0.12694	-0.64386	*-5.73385	-2.27313
Urbanization level (Large urban reference)							
Large suburban	-0.44211	-0.22069	1.46590	0.50556	-0.07649	#2.78910	*2.31011
Medium/small metropolitan	0.88281	#-0.44238	1.67953	0.07443	0.28938	*2.27176	0.75477
Nonmetropolitan	-0.38427	-0.13978	0.13301	-0.62956	0.57636	*4.17804	1.64725
Percent AIAN in county ⁴	*-0.43045	*0.51235				0.54579	0.39101
Percent API in county			-0.13245	0.00735			
Percent Black in county ⁵	0.0000258		0.02078		*0.00079	*0.11100	#0.04985
Percent multiple-race in county	*-0.16934	#-0.07906	0.31250	*0.09791	#0.31679	*-0.23972	-0.02919
Constant	3.08086	-0.70527	0.45883	-1.18887	-0.17533	-0.64594	0.77004

perficient differs from zero, p < .05. coefficient differs from zero, p < .10. Variable not in model. merican Indian or Alaska Native. sian or Pacific Islander.

www.cdc.gov/nchs/data/series/sr 02/sr02 135.pdf



NCHS logit model

- Los Angeles resident, choosing between white/black
 - Male, age 18, Hispanic 64% black
 - Female, age 18, Hispanic 61%
 - Male, age 65, Hispanic 69%
 - Male, age 18, non-Hispanic 79%



NCHS logit model

- Limitations
 - Some data sets only provide the total number of multirace, not the actual combinations
 - Model is probably seriously overfitted
 - All methods give us strange results in Guyanan and Trinidadian neighborhoods in Queens

127

NAACCR

4. Subcounty SES and cancer

- County is not a useful unit of measure for ecologic analysis
 - Small number of highly heterogeneous urban counties with a large share of the population
 - Directions of associations are often reversed!



NAACCR Census Tract Poverty Indicator

- Helpful for interstate analyses
 - Poverty rate has been found to be a "singularly effective" measure of SES
 - Cut points are <5%, 5-10%, 10-20%, and >20%
 - Within your own state, higher precision is possible
 - Finer cut points (or even continuous)
 - Block group
 - Other measures than poverty

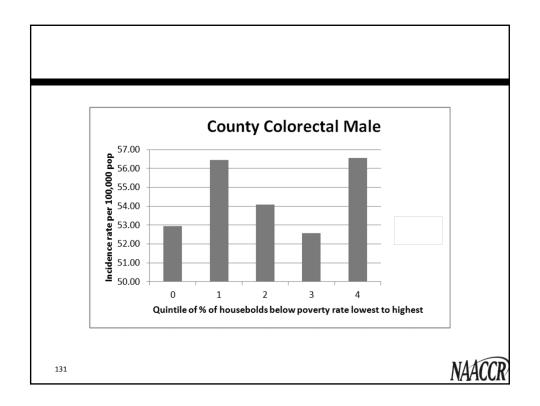
129

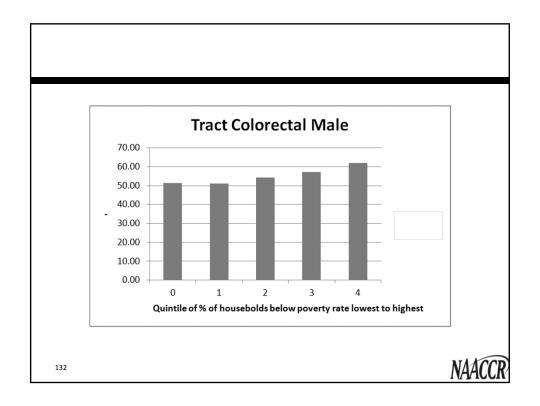
NAACCR

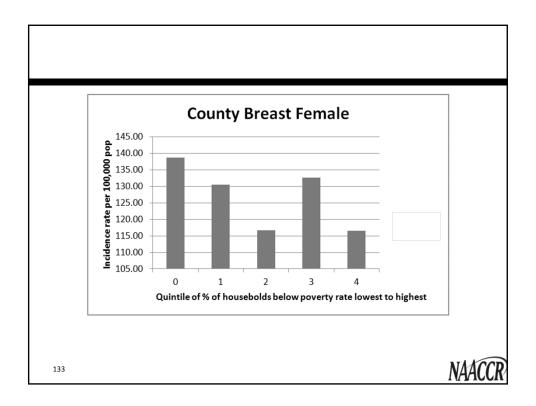
New York State analysis

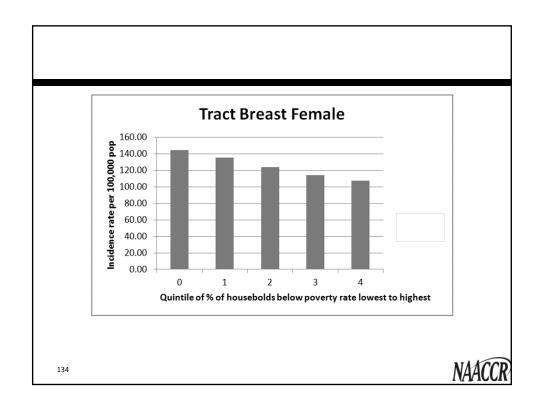
- Divided the state into quintiles at the block group, tract, and county level using different SES measures
 - 3,000 block groups per quintile
 - 1,000 tracts per quintile
 - 12-13 counties per quintile
- Calculated cancer rates for each quintile for 23 common cancer sites
- SES measures used were median income, poverty rate, % foreign born, and % without HS education (only results for poverty reported here)

130









Summary

- Cancers associated with poverty at the county level:
 - none

135

NAACCR

Summary

- Cancers associated with poverty at the tract level:
 - oral (M)
 - esophagus (M)
 - stomach (M,F)
 - colorectal (M,F)
 - liver (M,F)
 - larynx (M,F)
 - lung (M)
 - cervix (F)
 - myeloma (F)

136

Summary

- Cancer associated with affluence at the tract level:
 - all sites (F)
 - Breast (F)
 - Uterus (F)
 - Ovary (F)
 - Bladder (M,F)
 - Kidney (M)
 - Melanoma (M,F)
 - Testis (M)
 - Thyroid (M,F)
 - Brain (M,F)
 - NHL (M,F)
 - Leukemia (M,F)

NAACCR

But aren't the errors on these ACS subcounty estimates enormous? Doesn't that invalidate this sort of analysis?

139

NAACCI

5. Margins of error in ACS data

- We now have margins of error, which we never did before.
 - How should we deal with them?
 - How bad are they?
- At the tract level, errors appear to be large but the data are still usable
- Errors are only 20-50% higher than they were 10 years ago
- Block group data are very poor

140

Albany County

Geography	Total Pop	Below Poverty	Above Poverty
Tract 1, BG 1	472 ± 104	75 ± 62	397 ± 105
Tract 1, BG 2	322 ± 93	54 ± 56	268 ± 86
Tract 2, BG 1	1030 ± 177	373 ± 152	659 ± 152
Tract 2, BG 2	307 ± 113	0 ± 123	307 ± 113
Tract 2, BG 3	607 ± 162	316 ± 148	291 ± 119

Problems with M.O.E.

- The errors are not normally distributed if the confidence intervals include negative numbers
- Measures are treated as independent when they are not
 - Under 65 plus 65 & over must equal total!
- Geographic units are treated as independent when they are not
 - Adjacent block groups are likely to be similar!



Problems with M.O.E.

- When considering the relationship between cancer and some demographic variable, we are pooling across many hundreds or thousands of observations
 - New York State poverty quintiles:
 - Tract 1000 in each quintile
 - Block group 3000 in each quintile

143

NAACCR

M.O.E. - conclusion

- There are ways to incorporate this error into analysis – not easy, but doable
- Continuing to ignore the error seems to be a viable option if we pool observations

144

NAACCR 2013 Topics in GIS Webinar

Accessing Census Data Using NAACCR Tools/Datasets

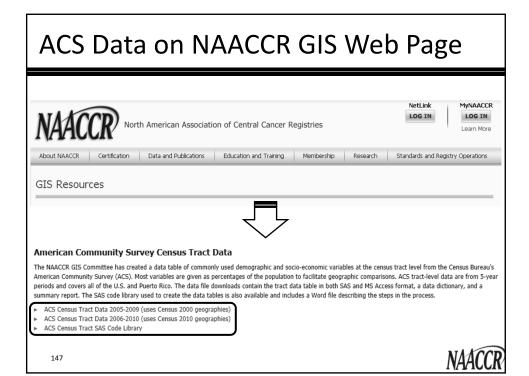
Dave Stinchcomb, Westat, Inc.

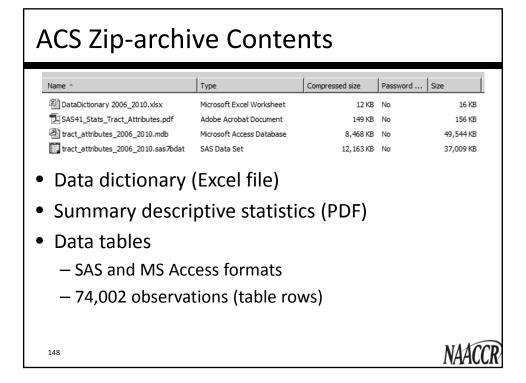
NAACCE

ACS Data on NAACCR GIS Web Page

- Commonly used demographic and socio-economic variables at the census tract level
 - Percentages for comparison across areas
 - All of the U.S. and Puerto Rico
 - Currently ACS 2005-2009 and 2006-2010 (2007-2011 coming soon)
- Finding the NAACCR GIS Resources Web page:
 - NAACCR Home Page / Research / GIS Resources
 - Google "NAACCR GIS"
 - http://www.naaccr.org/Research/GISResources.aspx

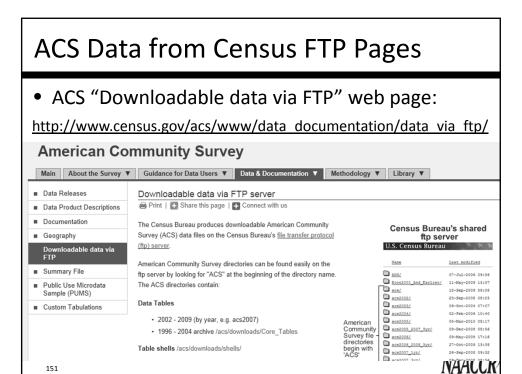
146

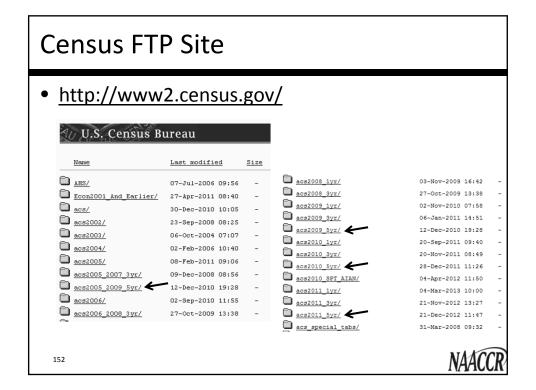


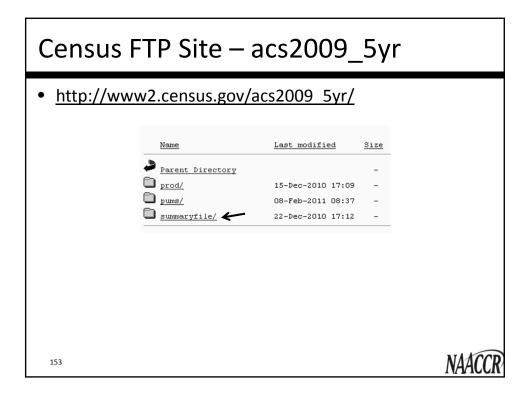


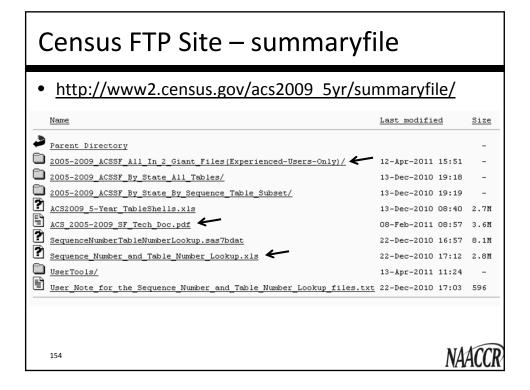
Census ACS 5-year Dat	a 2006-20	10 by C	ensus T	ract	
(MOE: the Margin of Er	ror for th	e value	.)		
Variable		Len F		ACS Source Table	Label
STUSAB	Char	2	\$2	GEO	State/U.SAbbreviation (USPS)
STATE	Char	2	ΨŁ	GEO	State (FIPS)
COUNTY	Char	3		GEO	County (FIPS)
TRACT	Char	6		GEO	Census Tract
NAME	Char	200		GEO	Area Name
StCntyTract	Char	11		GEO	Census Tract 11-digit FIPS Code
Popsamps	Num	8		B00001	Unweighted population sample count
Totpop	Num	8		B01003	Total population
Totpop_MOE	Num	8		B01003	Total population MOE
Pct_sampled	Num	8	8.3	B00001, B01003	Percent sampled
Pct nonwhite	Num	8	8.3	B02001	Percent race not white alone
Pct_nonwhite_MOE	Num	8	8.3	B02001	Percent race not white alone MOE
Pct_Hispanic	Num	8	8.3	B03002	Percent Hispanic
Pct_Hispanic_MOE	Num	8	8.3	B03002	Percent Hispanic MOE
Pct_minority	Num	8	8.3	B03002	Percent minority (other than non-Hisp white)
Pct_minority_MOE	Num	8	8.3	B03002	Percent minority (other than non-Hisp white) MOE
Pct_forborn	Num	8	8.3	B05012	Percent foreign born

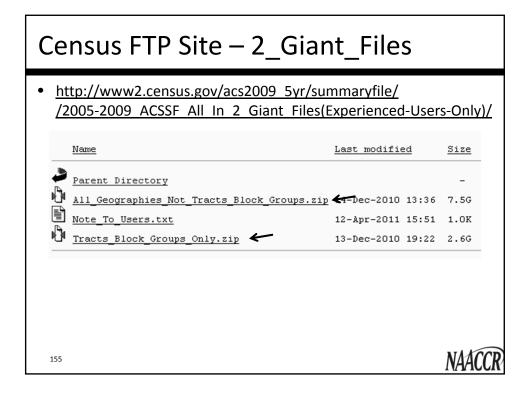
Variable	Туре	Len	Format	ACS Source Table	Label
Totpop_educ	Num	8		S1501	Total pop for education (25+)
Totpop_educ_MOE	Num	8		S1501	Total pop for education (25+) MOE
Pctsubj_highsch	Num	8	5.1	S1501	Percent high school (from subj table)
Pctsubj_highsch_MOE	Num	8	5.1	S1501	Percent high school MOE (from subj table)
Pctsubj_notHS	Num	8	5.1	S1501	Percent less than HS (from subj table)
Pctsubj_notHS_MOE	Num	8	5.1	S1501	Percent less than HS MOE (from subj table)
Pctsubj_college	Num	8	5.1	S1501	Percent college (from subj table)
Pctsubj_college_MOE	Num	8	5.1	S1501	Percent college MOE (from subj table)
Tothh_lingiso	Num	8		B16002	Total households for linguistic isolation
Tothh_lingiso_MOE	Num	8		B16002	Total households for linguistic isolation MOE
Pctsubj_lingiso	Num	8	5.1	S1602	Percent linguistically isolated households (from subj table)
Pctsubj_lingiso_MOE	Num	8	5.1	S1602	Percent linguistically isolated households MOE (from subj table)
Totpop_engl	Num	8		B16005	Total pop for English language (5+)
Totpop_engl_MOE	Num	8		B16005	Total pop for English language (5+) MOE
Pct_noEngl	Num	8	8.3	B16005	Percent non-English speakers
Pct_noEngl_MOE	Num	8	8.3	B16005	Percent non-English speakers MOE
Totpop_pov	Num	8		S1701	Total pop for poverty (non-group)
Totpop_pov_MOE	Num	8		S1701	Total pop for poverty (non-group) MOE
Pctsubj_belowpov	Num	8	5.1	S1701	Percent below poverty (from subj table)
Pctsubj_belowpov_MOE	Num	8	5.1	S1701	Percent below poverty MOE (from subj table)
Pct_be150pov	Num	8	8.3	C17002	Percent below 150% poverty
Pct_be150pov_MOE	Num	8	8.3	C17002	Percent below 150% poverty MOE

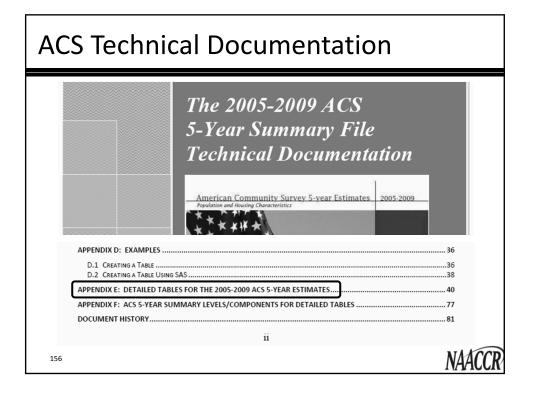




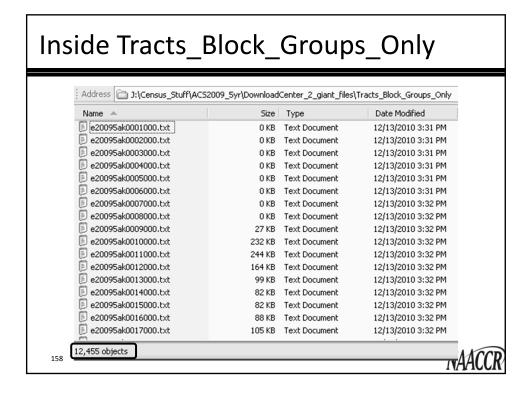


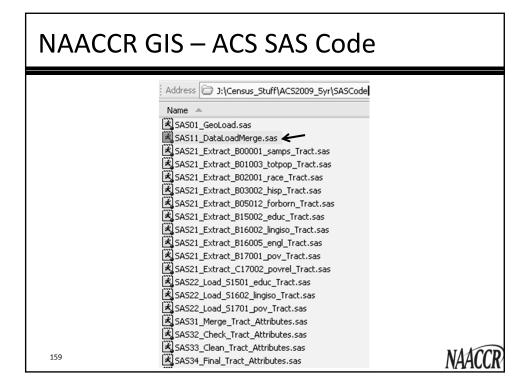






Appendix E						
	Appendix E: Detailed Tables for the 20	005-2009	ACS 5-Y	Year Estimate	s	
						•
		for geo other th	ity of data graphies an Block oups	Availability of data for Block Groups	Summary File Information	
Table Number	Table Title	In Summary File and American FactFinder	Only in Summary File	Only In Summary File	Sequence Number (If there are two numbers: first is US, second is PR)	Starting And Ending Positions
B00001	Unweighted Sample Count Of Population	See Note 1		See Note 1	9	7-7
B00002	Unweighted Sample Housing Units	See Note 1		See Note 1	9	8-8
B01001	Sex By Age	х		x	10	7-55
B01001A	Sex By Age (White Alone)	х			10	56-86
B01001B	Sex By Age (Black Or African American Alone)	х			10	87-117
B01001C	Sex By Age (American Indian And Alaska Native Alone)	х			10	118-148
B01001D	Sex By Age (Asian Alone)	х			10	149-179
B01001E	Sex By Age (Native Hawaiian And Other Pacific Islander Alone)	х			11	7-37
D04004E	Cau Du Ana / Cama Other Dana Alama)				11 - 4	_ 2000





NAACCR GIS – Data Load & Merge

```
/* Load the two Census ACS sequence data files: Estimates (Est) and Margin of Error(MOE) */
/* for a given sequence that contains one or more desired tables and combine them. */
/* See "Sequence_Number_and_Table_Number_Lookup.xls" to get sequence number for the desired table */
options sysprintfont=("Courier New" 8) leftmargin=0.75in nocenter;

/* Specify ACS sequence number and number of variables: */
%let seqnum=011; * The sequence number of files that contain the table you want (3 digits);
%let varnum=192; * The number of variables in the sequence that you need;
/* Example: seqnum 046 and varnum 226 for table C17002 (Ratio of Income to Poverty Level) */

NAACCR
```

NAACCR GIS – Data Load & Merge

```
/* Create a dataset with the state abbreviations and Fips codes */
        proc sort data=sashelp.zipcode nodupkey out=statetbl;
          by state statecode;
        data statetbl;
          set statetbl;
           keep state statecode;
          if state <= 56 or state = 72; /* Just the 50 states, DC, and Puerto Rico */
        %macro readEst;
        /* Read in state ACS Est File */
        options NOSOURCE; /* Suppress generated source from IMPORT */
        PROC IMPORT DATAFILE="&pathin.&filegrplong.\e20095&statem.O&seqnum.000.txt"
           OUT=tempEst DBMS=CSV REPLACE ;
           getnames=no;
           DATAROW=1;
        options SOURCE;
        proc datasets nolist force;
          append out=EstSeq&seqnum data=tempEst;
           quit:
        run;
161
       %mend readEst;
```

MOEs for Derived Variables

- Census Bureau publication:
 - American Community Survey: Multiyear Accuracy of the Data
 - http://www.census.gov/acs/www/Downloads/data_documentat_ ion/Accuracy/MultiyearACSAccuracyofData2011.pdf

10/2/2012 American Community Survey Office

American Community Survey Multiyear Accuracy of the Data (2009-2011 ACS 3-year and 2007-2011 ACS 5-year)

INTRODUCTION

This multiyear ACS Accuracy of the Data document pertains to both the 2009-2011 ACS 3-year data products and the 2007-2011 ACS 5-year data products. Differences will be noted where applicable.

162

MOEs for Derived Variables

Sums and Differences of Direct Standard Errors

The standard errors estimated from these tables are for individual estimates. Additional calculations are required to estimate the standard errors for sums of or the differences between two or more sample estimates.

The standard error of the sum of two sample estimates is the square root of the sum of the two individual standard errors squared plus a covariance term. That is, for standard errors $SE(\hat{X}_1)$ and $SE(\hat{X}_2)$ of estimates \hat{X}_1 and \hat{X}_2 :

$$SE(\hat{X}_1 \pm \hat{X}_2) = \sqrt{(SE(\hat{X}_1))^2 + (SE(\hat{X}_2))^2 \pm covariance}$$
 (1)

The covariance measures the interactions between two estimates. Currently the covariance terms are not available. Data users should use the approximation:

$$SE(\hat{X}_1 \pm \hat{X}_2) \approx \sqrt{(SE(\hat{X}_1))^2 + (SE(\hat{X}_2))^2}$$
 (2)

However, this method will underestimate or overestimate the standard error if the two estimates interact in either a positive or negative way.

I W K IUU

MOEs for Derived Variables - in SAS

```
/* Macro to calculate the MOE of a sum or difference */
/* temporary variables: sumdiff1 sumdiff2 sumdiff */
%macro moe_sumdiff (moe1=, moe2=, moesumdiff=);
    sumdiff1_se = &moe1 / 1.645; /* convert MOE1 to std error */
    sumdiff2_se = &moe2 / 1.645; /* convert MOE2 to std error */
    sumdiff_se = SQRT((sumdiff1_se**2)+(sumdiff2_se**2));
    &moesumdiff = sumdiff_se * 1.645; /* convert std error to MOE */
%mend moe_sumdiff;
```

164

MOEs for Derived Variables – Example

Proportion with income below 150% of poverty

C17002 RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS
Universe: Population for whom poverty status is determined
2005-2009 American Community Survey 5-Year Estimates

	Census Tract 9501, Baker County, Oregon				
	Estimate	Margin of Error			
Total:	2,714	+/-213			
Under .50	79	+/-42			
.50 to .99	194	+/-108			
1.00 to 1.24	109	+/-89			
1.25 to 1.49	88	+/-59			
1.50 to 1.84	105	+/-63			
1.85 to 1.99	54	+/-39			
2.00 and over	2,085	+/-220			

Proportion below 150% of poverty = (79 + 194 + 109 + 88) / 2714

NAACCR

An ACS GIS Tool from the CDC

- Esri ArcGIS ACS Toolbox
 - Calculates MOEs for user-derived data
 - Determines statistically significant differences
 - Helps find a suitable mapping scheme considering data uncertainty
- Availability:
 - Current version running on ArcGIS 10
 - Contact Elaine Hallisey, CDC/ATSDR (ehallisey@cdc.gov)

166

Future Plans

- Census tract data for ACS 2007-2011
- Expanded variables
- Suggestions from user feedback

167

NAACCR

Questions?

168

Coming up!

- 8/1/13
 - Cancer Registry Quality Control
- 9/5/13
 - Coding Pitfalls

Certificate phrase: Census 2010

http://www.surveygizmo.com/s3/1303184/GIS-Webinar



