

# Census Overview: Basics, Decennial, ACS, and Estimates

An Informational Session Presented for:  
The Nebraska State Data Center 25<sup>th</sup>  
Annual Data Users Conference

August 19, 2014 – 9:00 to 10:15 a.m.  
Part of conference webcast



David Drozd, M.S.  
Center for Public Affairs Research, UNO  
402-554-2132 [ddrozd@unomaha.edu](mailto:ddrozd@unomaha.edu)  
[www.unomaha.edu/cpar](http://www.unomaha.edu/cpar) [www.facebook.com/unocpar](http://www.facebook.com/unocpar)

Please  
Like



## Terminology & Definitions

- **Decennial Census**: “the Census”; headcount taken once every 10 years; mandated by Constitution for equal representation
- **Short form: basic** decennial Census questionnaire sent to most households in 2000 and all households in 2010; records simple demographics like age, gender, race, housing tenure (own/rent)
- **Long form: detailed** decennial Census questionnaire sent to a sample of about 1 in 6 households in the 2000 Census but not used in the 2010 Census (replaced by ACS); recorded socio-economic and detailed housing data
- **American Community Survey (ACS)**: continuous sample survey (forms sent every month) compiled to provide **current annual data**; nearly identical to long form – allowed 2010 Census to be short form only (simply a headcount)
  - 2014 ACS (and those in future years) continues monthly surveying even though 2010 Census is finished



Nebraska State Data Center – CPAR @ UNO

2

## Census Datafiles

- **Summary File 1 (SF 1)**: Used for both 2000 and 2010 basic demographic data from the *short form*: age, gender, race/ethnicity, housing tenure (own vs. rent), etc.
- **Summary File 3 (SF 3)**: **Not part of 2010 Census, replaced by ACS datasets**; has detailed socio-economic and housing data from the 2000 *long form*; also has figures for SF 1 items (age) but they are based on a “weighted” sample, not the official counts

**BE CAREFUL – You must go to the right source to get correct data!**

- **ACS Datasets**: contain data for ACS variables; are based upon the timeframe over which the data was collected

- 1-year: 2012      3-year: 2010-2012      5-year: 2008-2012
- **Do NOT Compare ACS datasets to SF 1 from the decennial census**
  - Compare SF 1 from 2000 to: \_\_\_\_\_ → SF 1 from 2010
  - Compare SF 3 from 2000 to: \_\_\_\_\_ → ACS datasets



Nebraska State Data Center – CPAR @ UNO

3

## ACS Data are Released Based Upon the Population of the Geographic Unit

Timeframe of Data	Population Threshold to have Data Released	Nebraska Cities with Data
Annual Estimates (2005, 2006, 2007, etc.)	65,000+	Only Omaha and Lincoln
3-year aggregates (2005-07, 2006-08, etc.)	20,000+	Omaha, Lincoln, and regional centers like Kearney, Norfolk, etc.
5-year aggregates (2005-09; 2006-10, etc.)	No threshold (data for all areas, even census tracts, zip codes, small towns & low population counties)	All cities/towns, little suppression (Gross, NE: population 4 had 2008-12 ACS data); realize fewer people = more accuracy concerns

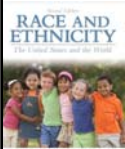


Note that even the multi-year aggregates get an annual data update (rolling timeframe as new data is released: 2007-2011 then 2008-2012, with 2009-2013 coming this fall).

More data = be more careful! Cite sources accurately and with detail (use table #s): 2012 ACS (S1701), U.S. Census Bureau

## Questionnaires & Concepts

- **Residence:** For the Census, it is based on *usual residence* – where people are most of the year (their permanent residence). It is also based upon an April 1 reference date.
  - Snowbirds spending 3 months in Texas and 9 months in Nebraska should be a resident of NE
  - College students are almost always counted in their college city (that's where they are on April 1)
  - Since the ACS forms are sent out and mailed back monthly, residence for the ACS is based upon who is living in the household for "the next 2 months"
    - Provides a more accurate picture of the populous: e.g. some snowbirds are counted in the South & college students in parent's house if home for summer
- **Race/Ethnicity:** In the census world, each person is either Hispanic/Latino or they are not, AND then they also have one or more races (option for 2+ races first utilized in 2000)
  - Hispanic/Latino question is asked first; then comes the race question
    - White non-Hispanic is "majority population"; total pop – WnH = minority population



## How does CPAR typically analyze race/ethnicity?

- We usually list totals for Hispanic/Latino, and then all races ALONE as being non-Hispanic (NH): White NH; Black NH; etc.
  - This avoids double counting and the sum of all groups when combined with 2+ races NH will equal the total population (percents sum to 100%)
  - Race ALONE means that only one race was stated on the form
  - Many groups/agencies and grants ask for simply White, Black, etc. and Hispanic – by definition many Hispanics in this case would be represented twice (their race is often listed as White or Other)
    - You heard it here first – the 2020 census will likely use a combined race/ethnicity question! It will get similar results, "other" race category will be selected less frequently, & will have areas to write in ancestry/origin (so we'd get data on African and European countries/origins in addition to Asians, Native Amer., & Hispanic)
- Also can evaluate race for those ALONE or IN COMBINATION
  - This includes those who selected that race specifically (alone) as well as those who selected that race and another one (2 or more races)
    - Double counts the people selecting 2+ races (doesn't sum to total pop. or 100%)
    - Might use more often as lessens effect of multi-race births "always" being minority

## Questionnaires & Concepts continued

- **Householder:** simply the first person on the census form
  - Is NOT necessarily the "head of the household" (that's old terminology)
  - Supposed to be the person "in whose name the residence is owned or rented"; that's usually but not always the person who fills out the form
    - Can be male or female; householder is somewhat arbitrary but makes a big difference for how families and by extension items like poverty are defined
- **Relationship:** how other persons living in the household are related or connected to the householder
  - Spouse, biological/adopted/step child, parent-in-law, roommate, unmarried partner, foster child, etc.
  - A household with 2 or more related individuals is a "family" household
    - One person households are not a family by definition, and thus are excluded from variables like median family income (but are included in median household income)
    - Unmarried partners are not a family by definition
      - If a child lives in the household of unmarried partners but is from a prior relationship, it is a function of who fills out the form (is the householder) for whether that is a "family" or not – IT GETS COMPLEX – impacts items like poverty



### Householder

Example:  
Susan



and her  
child Emily



Live with Paul



Notes: Susan and Paul are dating but not married; Paul is not the father of Emily

If Susan  
is the  
householder...



Emily is related to Susan as biological child and this is a family household; Paul is unmarried partner & not part of the 'family'

If Paul  
is the  
householder...



Emily is an "other nonrelative" and Susan is unmarried partner; this is NOT a family household

In both cases Paul's income is NOT included in Susan and Emily's poverty calculation; if they were married it would be.

# Geography

- Size continuum: large to small, or small to large
- Legal/Administrative vs. Census/Statistical
  - Nation—State—County—City—Township
  - Block—Block Group—Census Tract—Zip Code—PUMA—Metro Area (MSA)—Division—Region

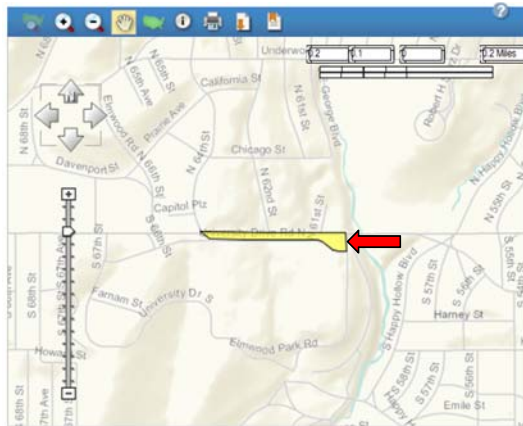


The West North Central Division of the United States

# FactFinder Address Search Page & Results

Geography Name	Geography Type	Geography Code
PUMA5 0004, Nebraska	Public Use Microdata Area (PUMA)	705
Nebraska	State	040
Douglas County, Nebraska	County	050
County Subdivision		060
Block 2022, Block Group 2, Census Tract 47, Douglas County, Nebraska	Block	100
Census Tract 47, Douglas County, Nebraska	Census Tract	140
Block Group 2, Census Tract 47, Douglas County, Nebraska	Block Group within Census Tract	150
Omaha city, Nebraska	Place within State	160
Omaha Council Bluffs, NE-IA Metro Area	Metro Statistical Area/Metro Statistical Area	310
Omaha, NE-IA MSA	MSAC/MSA	300
Omaha, NE-IA Urbanized Area	Urban Area	400
Congressional District 2 (111th Congress), Nebraska	Congressional District	500
State Senate District 9 (2010), Nebraska	State Legislative District (Upper)	410
No Lower Chamber Districts (2010), Nebraska	State Legislative District (Lower)	620
Precinct 3-12, Douglas County, Nebraska	Voting District/Remainder	700
2CTAS 68132	5-Digit ZCTA	600
ZIP 68132 (Omaha, NE)	5-Digit ZIP Code	661
Omaha Public Schools, Nebraska	School District (Unified)/Remainder	970
Omaha city, Nebraska	Economic Place	800

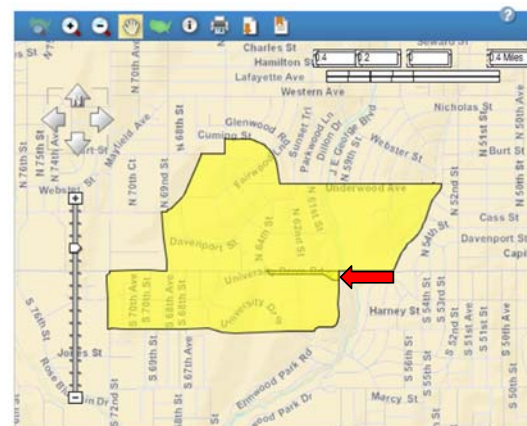
Block 2022 in Block Group 2, Census Tract 47 Douglas County, NE



This is a single block – they are “building blocks” for larger geographies.

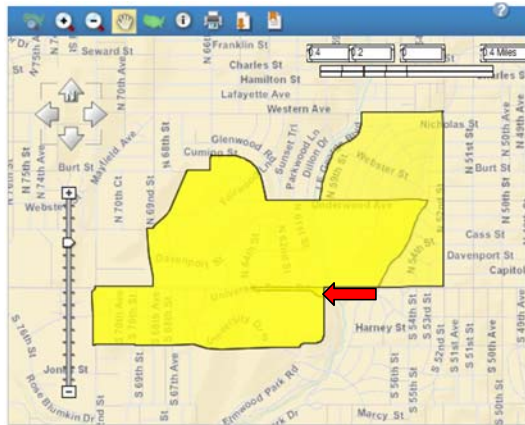
Blocks are bounded by physical features like roads or streams.

Block 2022 in Block Group 2, Census Tract 47 Douglas County, NE



A group of usually about 15-30 blocks comprise a “block group”.

Block 2022 in Block Group 2, **Census Tract 47** Douglas County, NE



Two or more block groups comprise a "census tract", a key unit for small-area analysis.

Census tracts typically have ~4,000 people and nest within counties.

13

Are you ready to rumble??

## Geography VS. Data (A Duel)

- There is usually a tradeoff between available information and the geography for which it is available
  - Can be due to confidentiality restraints
  - More specific information may require expanding to a larger geography (or a longer timeframe)
    - Even if the data is provided, you must ask whether the data is "reliable" for that geography?
    - Are the figures based on a small number of cases where the sampling error could be large?
      - Sampling error occurs if the figures would be different if different households/people were included in the sample. ACS data always lists the "margin of error".



Nebraska State Data Center – CPAR @ UNO

14

## Relationship between Data/Datafiles and Geography

- For blocks, you can only get Redistricting or SF 1 data
  - Confidentiality reasons: you'd know certain incomes on your block (also sampling/error issues)
- Block groups have most SF 3 data; only meant as building blocks for larger custom geographies in the ACS
  - No ancestry or unmarried partner data
- Census Tracts are a key geography as they are one of the smallest geographies to have most all data compiled
  - PCT (population items) and HCT (housing) tables in decennial files
  - Now have 5-yr ACS zip code data; like tracts they are a good unit of analysis for "neighborhoods" but typically larger than tracts so more accurate/less variability
- Public Use Microdata Areas (PUMAs) are combined census tracts that contain at least 100,000 people and are the smallest geography for the PUMS files; important in ACS!

Nebraska State Data Center – CPAR @ UNO

15

## PUMA geography



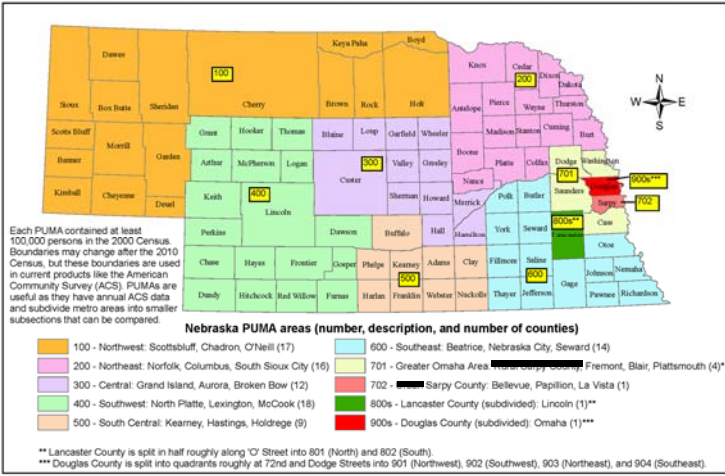
- Nebraska has 14 PUMA areas (4 in Douglas Co., 2 in Lancaster)
- ACS data used 2000 based PUMAs from 2005-2011; 2012 and future years use updated 2010 based PUMA boundaries
  - The 2010 PUMAs are nearly identical: a couple tracts changed in Douglas and Lancaster counties; Sarpy Co. is now its own PUMA
- PUMA geography has more importance in the ACS
  - Data are released annually since PUMAs exceed the annual ACS population threshold of 65,000
    - Not many Nebraska geographies get annual data; PUMAs are the smallest annual geography for making statewide comparisons
  - Breaks core metro counties into smaller areas and gives proxies for smaller geographies that are similar to the larger PUMA area
    - Northeast NE puma proxy for Pierce Co; Southeast Douglas for S. Omaha
- State Maps for PUMA boundaries can be viewed on the internet
  - <http://www.census.gov/geo/maps-data/maps/reference.html>

Nebraska State Data Center – CPAR @ UNO

16



### Geographic Boundaries of Nebraska Public Use Microdata Areas (PUMAs)



Source: 2000 Census, Geography Program, U.S. Census Bureau (a detailed map can be viewed at [http://fp2.census.gov/geo/maps/puma/puma2k/ne\\_puma5.pdf](http://fp2.census.gov/geo/maps/puma/puma2k/ne_puma5.pdf))  
 Prepared by: David Drozd, Center for Public Affairs Research, University of Nebraska at Omaha - August 27, 2009

### The American Community Survey: What is it?

- Nationwide written/mail survey conducted by the U.S. Census Bureau
  - Sent to a **sample** of households, not all households
  - Online completion option began January 2013
- Similar to the sample portion of the decennial (10 year) census but is completed every year
  - Gives us more current information: annual data rather than 10 year intervals between releases
  - Provides data on the same Census topic areas
  - Replaced “long form” of decennial census; 2010 Census primarily a population count
  - The “future” of socio-economic Census data, **IF funded into the future**

Nebraska State Data Center – CPAR @ UNO

18

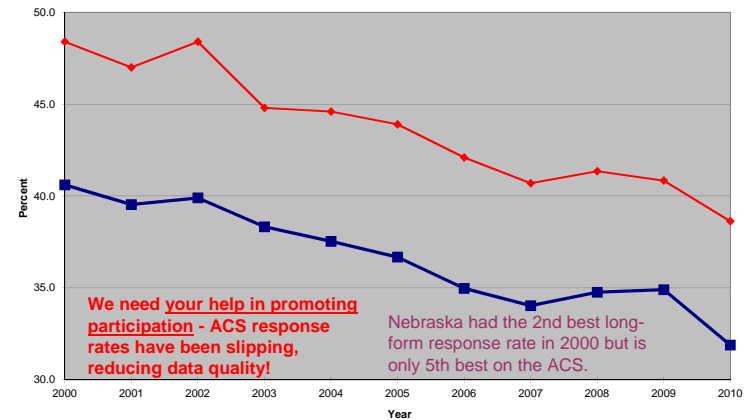
### ACS Methodology in Brief

- Surveys mailed out & received back each month
  - 3.5 million surveys annually nationwide
    - Sample about 1 in 40 housing units (1 in 8 over five years of survey collection – Census 2000 long form was 1 in 6)
      - Census assigns both household and person weights. Summing these weights produces the estimates. Base weight for households is about 40 given 1 in 40 sample.
  - Monthly surveys are combined to estimate figures for the calendar year as a whole.
  - Nebraska response rate is in the top 5 (but has been slipping—help promote participation)
    - 1 in 3 non-respondents are personally interviewed to get info. (very important—improves its data over others—but Census 2000 had full non-response follow up)

Nebraska State Data Center – CPAR @ UNO

19

Initial ACS Mail Response Rates, U.S. and Nebraska: 2000 to 2010



Source: 2010 American Community Survey Variance Memorandum Series #ACS10-S-37 and prior

United States Nebraska

Compiled and Prepared by: David Drozd, UNO Center for Public Affairs Research, July 23, 2012

## ACS New Subject Areas:

- Food stamp reciprocity of households
- Fertility (women age 15-50 giving birth in the last year)
  - Note: These items were not asked on Census 2000 but have been asked since early-on in the ACS
- Health insurance coverage, changes in marital status, and veteran service connected disabilities added in 2008; field of Bachelor's Degree added in 2009
  - Disability question was revamped in 2008: see all forms here:
    - [http://www.census.gov/acs/www/methodology/questionnaire\\_archive/](http://www.census.gov/acs/www/methodology/questionnaire_archive/)
- New questions on computer/internet access (and type like cable, dial up, etc.) started in 2013

**KEY POINT:** Remember that the goal of the ACS is not to make exact counts of the population or an item like the number of births (other sources for that), but to provide information on the characteristics of the population or those giving birth – information you can't find elsewhere!

## Point Estimates, MOE and Confidence Intervals

- The ACS data provide point estimates for various characteristics. ACS data also include a margin of error (MOE) for finding a lower and upper bound. (e.g.  $64 \pm 3$ )
  - Why?
    - The ACS is a sample and subject to sampling error.
    - Is the ACS data representative of the entire population?
    - Census 2000 long form also a sample—1 in 6 sampling rate made sampling error small and MOE was not released.
- Adding and subtracting the MOE to/from the point estimate creates a range called a confidence interval.
  - Example: 08-12 NE fertility rate:  $64 - 3$ ;  $64 + 3$ ; so the range or interval is 61 to 67
  - ACS displays the MOE for a 90% confidence interval.
    - The bounds tell us that we are 90% confident that the figure for the entire population would be in this range if all households were surveyed.

## Multi-year Aggregate Estimates: Advantages

- The aggregates provide data for more geographies
  - Annual data for NE: 3 counties (Douglas, Lancaster, Sarpy)
  - 3-yr aggregate: 17 counties; 5-yr aggregate: **all 93 counties**
- For areas that get annual data, the aggregates based on more completed surveys will be more accurate and have smaller margins of error (MOE)
  - Especially important for sub-groups—data for specific age groups or racial/ethnic groups
    - **General rule: whenever you split a stat (e.g. education by age; poverty by race, etc.) use the longest timeframe**
- Will reduce variability in year-to-year figures
- Some FactFinder tables are prepared but not released annually because of inaccuracy concerns—some of those tables will now have data released



The screenshot shows the American FactFinder interface. At the top, there are navigation tabs for 'MAIN', 'COMMUNITY FACTS', 'GUIDED SEARCH', 'ADVANCED SEARCH', and 'DOWNLOAD CENTER'. The 'Advanced Search' tab is active, showing a search result for 'RECEIPT OF FOOD STAMPS/SNAP IN THE PAST 12 MONTHS BY RACE OF HOUSEHOLDER (AMERICAN INDIAN AND ALASKA NATIVE ALONE)'. Below the search result, there are options for 'Table View', 'Modify Table', 'Bookmark', 'Print', 'Download', and 'Create a Map'. A message at the bottom of the page states: 'The data you requested are not available. Possible reasons include: No data are available for the combination of geographic area(s) and other items in "Your Selections"'. A red box highlights this message with the text: 'Example of Standard Table without Data Released; this was for Sarpy County – the 3-yr dataset did list figures'.

## ACS Sample Size Increase Began in 2011 and Continues!!

- Higher sampling rate means more completed surveys and more data items available with better accuracy!!
- National sample expanded from 2.9 million to 3.54 million addresses per year (+ 22%)
  - Nebraska housing unit sample increased by about 5,700 (3,450 more completes or + 21%)
    - >21,500 NE ACS completes in 2012 vs. U.S. polls of 1,000
    - Statistical quality improvements are seen with 2011 & 2012 data vs. prior years (smaller error factors with bigger sample)
    - Full impact will be seen later as we move forward in time
    - This may prove to be the "golden time" for ACS data
      - Larger sample with solid weighting based on 2010 census



## 2013 ACS Data to be released later in 2014

- Basically all the annual (one-year) data for 2013 will be released at one time (for areas with 65,000+ persons)
  - **Thursday Sept 18, 2014** (media embargo on Tues. 9-16-14)
    - Annual CPS data on income/poverty and health insurance also released Sept 16 along with state ACS health insur. (no embargo)
  - Will include the first computer availability/internet access data
  - 2013 ACS response rates and data could be impacted by the new internet response option
- The 3-year aggregate estimates for 2011-13 will be released on **October 23, 2014** (20,000+ person areas; no embargo)
  - Will have two 3-year periods that don't overlap to compare (2008-10 vs. 2011-13; recession vs. recovery)
- New 5-year aggregate estimates for 2009-13 will be released on **December 4, 2014** (all geographies; embargo on 12-2-14)
  - Increased sampling rate is important, especially for small areas
    - We must keep updating our data files to the most current timeframe as it will also be the most accurate given sample size increase
  - Field of Bachelor's Degree data for all places for the 1<sup>st</sup> time

## The 2009 ACS Nebraska "Double Whammy"



- Two separate methods adjustments affected the 2009 ACS sample size and weighting approach
  - Implementation of city weighting controls
    - ACS had always controlled to county estimates, but in 2009 they also started controlling to city estimates
      - Omaha 2000 pop = 390,007; ACS pop 388,894 in '08; 454,714 in '09; (2010 Census: 408,958)
        - Different characteristics—more accurate but may have to consider it a 'break in series'
    - NE dissolution of elementary school districts affects ACS
      - Now are "unified" school districts – larger territory and population means a smaller sampling rate in the ACS
      - Distribution of sample changes: a "different Nebraska"
        - 500 less in micro counties; 70 less in Big 3; 85 less in Cherry Co
- Realize Omaha change impacts all NE data; try not to utilize 2009 products – rely on newer data (2010, 2011, 2012, etc.)

## Population Estimates Program: Basics

- The population estimates program provides the official head and housing unit counts as well as counts by age, sex, and race in non-census years
  - Shows how the population has changed since the Census
- As of July 1 of the specific year
- Staggered releases throughout the year
  - Large geographies first, most detailed data last
- The current release/vintage always supersedes prior
  - Can create confusion for why one 7-1-11 estimate will differ from another—cite the source and release date
  - You always have to be going and getting/using the most currently released data
- Estimates differ from projections
  - Estimates are the newest look at our current population
  - Projections predict the population structure in the future



## Population Estimates Program: Census Methodology in Brief

- Start with the Census 2010 headcount (4-1-10)
- Adjust for headcount revisions, boundary changes (annexations) so that everything is consistent over time
  - This is called the Estimates Base (refers to 4-1-10)
- Add births, subtract deaths from vital records
- Make estimates of domestic and international migration
  - Partner with IRS to show changes in where people file returns
    - Limited to people/families who file their taxes, doesn't account for people leaving the U.S. very well
  - Use Medicare records to be more precise in 65+ population
- Various methods estimate changes in group quarters and military populations (deployment affects)
- Use building permit data and recorded demolitions
  - Tornados don't file demolition permits!! (Hallam, Pilger)
  - Permit allocation has some issues (Omaha ETJ vs. city limits – Omaha city estimates have been/continue to be overstated)

## Population Estimates Program: Its Importance

- The estimates give us our best look and official figures for how the population has changed since the Census
  - Some grants and \$ allocations are calculated based on the estimates
  - On SF 1 style demographics, it is the estimates and not the ACS that are the official population source – use it for age, gender, race
- Is Nebraska growing more/less than regional/neighboring states and the U.S.?
  - **Will be the basis of predicting whether we'll keep all 3 Congressional seats in 2020!**
- What population components are changing
  - Births/Deaths; Foreign and Domestic net migration
- How is the state's population distribution (more growth in metro counties) and age/race structure changing?
  - Migration of baby boomers and those age 20-34 is key
  - Will Nebraska's growth continue to be dominated by minority populations, especially Hispanic/Latino?



## Estimates Program Data



- Our office has a lot of the data compiled and will make an annual Nebraska Population Report
  - Is a good reference document as data changes year to year
  - If you need something specific, contact us to see if we have compiled what you're looking for
- Realize that you may just want to use 2010 data rather than the estimates—the estimates have more value as we move further away from 2010
- Pop. estimates website is: [www.census.gov/popest/](http://www.census.gov/popest/)
- Next big release will be for data as of July 1, 2014
  - Data for states coming in December 2014
  - Data for counties/metro areas released March 2015

## Bottomline: Ask yourself these questions

- What would I like to have ideally?
  - Exact data items (variables)
  - Geography
  - Timeframe
- What source is best to use for that – Decennial Census, ACS, Estimates Program, other source?
- Is that data available? Is it accurate or have large error factors?
  - If not, what are my best tradeoffs?
    - Is it more important to have that exact geography or can I use a larger one as a proxy? (Douglas County census tract 11 versus PUMA 903)
    - Do I want the most current data or can I extend to a longer timeframe (to increase accuracy or availability)? (2012 versus 2010-12)
      - Any item split by age or race should use longest timeframe unless geography is sufficiently large (State level might not be large enough)

