

GIS Workshop

Basic 2 : Census & Thematic Mapping

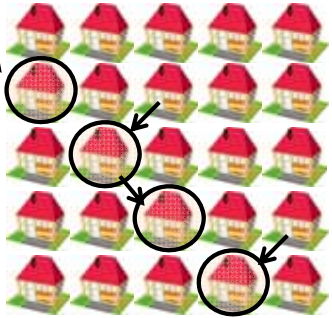
Chieko Maene
c-maene@northwestern.edu
 Maps and State Documents Librarian
 Government and Geographic Information and Data Services
 University Library (Evanston)
 Northwestern University

Historical Censuses

Census & GIS

- Why it's so important for GIS analysis?
 - Censuses provide greater precision and geographic detail than any other nation-wide surveys
- Definition:
 - Census = a count of the population of a country as of a fixed date
- In the US, generally means “decennial census”
 - Every 10 years since 1790
 - Purpose = Reapportionment! The reported total population was used to apportion seats in the House of Representatives
 - Evolved over time
 - From simple population counting to more detailed data collection

Census – How Does It Work?



- Every household receives a census questionnaire around April 1st of the census year
- Two types
 - Short form (100%)
 - Long form (sample)
 - Short form questions plus extra questions
 - Approx. one in every 6 households receives a long form

Two Questionnaire Forms

- Short Form (100 % count)
 - Population count
 - Basic demographic information
 - Population count, age, sex, race/ethnicity, household type
 - Basic housing information
 - # of units, occupancy, tenure status (owner/renter)
- Long Form (Sample, average 1 in 6 households, 1990 & 2000)
 - Introduced in 1960, to reduce cost of labor
 - Instead of asking detailed questions to every household
 - Detailed population and housing information
 - Socio-economic and housing characteristics

Long Forms (Sample) Variables

<ul style="list-style-type: none"> • Social characteristics <ul style="list-style-type: none"> – Education & enrollment – Marital status – Veterans – Disabilities – Mobility (migration) – Place of birth – Ancestry – Language 	<ul style="list-style-type: none"> • Economic characteristics <ul style="list-style-type: none"> – Employment (status, industry, occupation) – Commuting – Income – Poverty • Housing characteristics <ul style="list-style-type: none"> – Structure, conditions – Utilities & facilities – Value, rents, mortgage
---	---

2000: New Variable

- **PCT008: Grandparents as caregivers**
- Universe (subject): Population 30 years and over in households

Map on the right:
Showing ratio of grandparents responsible for own grandchildren to all population (30 years +)

Legend

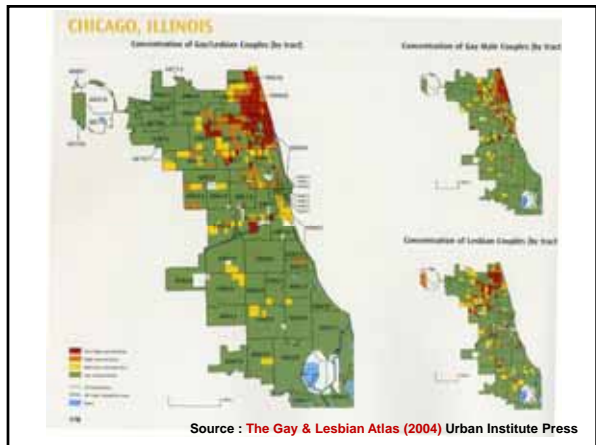
- 0%
- 0.1% - 1.5%
- 1.6% - 3%
- 3.1% - 4.5%
- 4.6% - 31.6%

2000: New Variable

- **PCT001: Unmarried partner households (same sex couples)**
- Universe (subject): Population 30 years and over in households

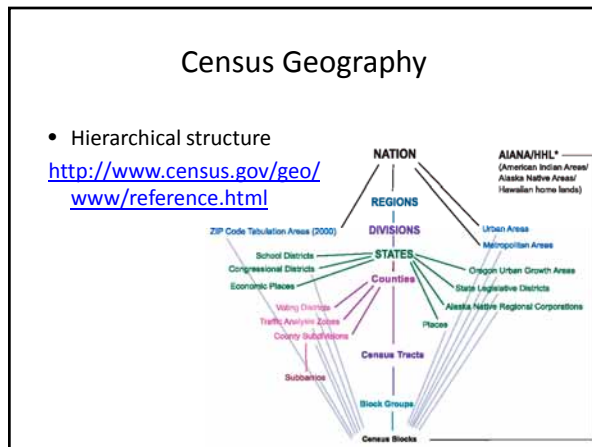
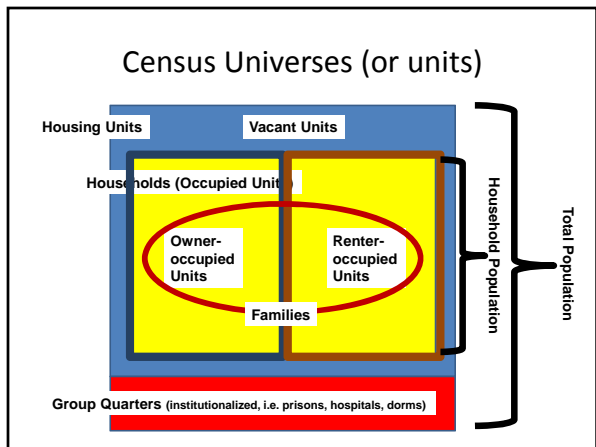
● Male-Male
● Female-Female

(1 dot = 1 couple/household)



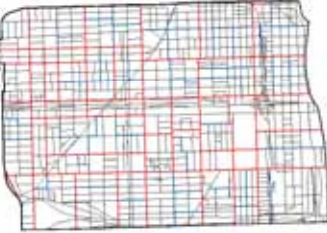
Summary File Name Changes

2000	1990	1980	1970
SF 1 (100%)	STF (Summary Tape File) 1A (down to BG), 1B (block), 1C (down to county), 1D (CD)	STF (Summary Tape File) 1A (down to BG), 1B (block), 1C (down to county), 1D (CD)	SSF (Summary Statistic File) FCF (First Count File) A (ED & BG) & B (down to places), SCF (Second Count File) A (tract) & B (others), TCF (Third Count File) (block)
SF 2 (100%) SF1 tables are repeated for various race/ethnic groups	STF 2A (down to tract) & 2B (down to place)	STF 2A (down to tract) & 2B (down to place)	
SF 3 (Sample)	STF 3A (down to BG), 3B (zipcode), 3C (down to county), 3D (CD)	STF 3A (down to BG), 3B (zipcode), 3C (down to county), 3D (CD)	SSF (Summary Statistic File) FCF (Fourth Count File) A (tract), B (places), C (others), FCF (Fifth Count File) A (3-zipcode), B (5-zipcode), C (others, including tract)
SF 4 (Sample) SF1 tables are repeated for various race/ethnic groups	STF 4	STF 4	



Census Geography

- Ex. Chicago's Near West Side Neighborhood
 - This is **Not** a census geography
 - Tract
 - Designed to contain 1,500-8,000 people (average 4,000)
 - Red outlines
 - Block Group
 - Designed to contain 600-3,000 people (average 1,000)
 - Blue outlines
 - Block
 - Bounded by physical features
 - Black outlines



Census Tracts


- GIS users ♥ tracts
 - It's a **pseudo-permanent** (stable) **geography**
 - Comparable over time – especially in older cities
 - Boundaries do not change often unlike places, districts, block groups
 - Another stable geography level is "Counties"
 - **Block groups** are smaller but boundaries change often
 - Small enough to represent **neighborhood characteristics**
- History:
 - 1910 – the first tract level data published in NY
 - Also for another 7 major cities, including **Chicago**
 - 1920 – for the same selected 8 cities
 - 1930 – for the select 13 cities
 - 1940 – nationwide tractization started (urban area only)

Census Boundary Files

- 2009 TIGER/Line Shapefiles
 - <http://www.census.gov/geo/www/tiger/tgrshp2009/tgrshp2009.html>
 - 2009 is the latest TIGER files!
 - Vastly improved in 2008 – straight lines!
 - Street lines (edges) are hard to use – stay away and use ESRI data instead
- Cartographic Boundary Files
 - <http://www.census.gov/geo/www/cob/>
 - Previous versions – older, uglier lines
 - 1990 & 2000
- ESRI Census 1990 & 2000 TIGER/Line Data
 - <http://www.esri.com/data/download/census2000-tigerline/index.html>
 - Previous versions, but still useful site

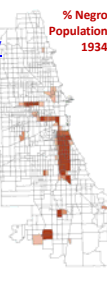
American FactFinder (Census)

- Data access tool
 - For recent censuses (1990 & 2000)
 - Limited data for 1990
- <http://www.census.gov/>
 1. On the left, click on **American FactFinder**
 2. On the left, click on **DATA SETS** and then select **Decennial Census**
 3. Select data files
 - Example: **Census 2000 Summary File 3 (SF 3) – Sample Data**
 4. Then select **Detailed Table**
 5. ... continue to select **Geography** and then **Tables** (variables)



Historical Census Data Sources

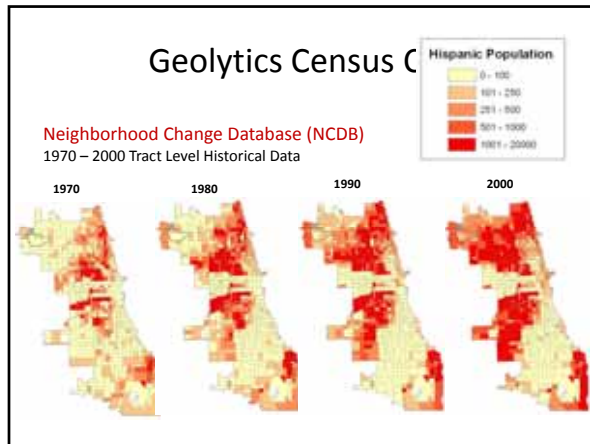
- Until recently, historical census data were available only from print sources or ICPSR archive
- Source for **tract** level historical census data:
 - **NHGIS, National Historical GIS** <http://www.nhgis.org/>
 - By Univ. of Minnesota, Minnesota Population Center
 - Download both historical census data & shapefiles
 - **NCDB, Neighborhood Change Database CD**
 - By a private census data company, Geolytics
 - 1970-2000 data & shapefiles (also normalized to 2000)
 - In-library use only, ask us (Govinfo department)
 - CD description: <http://www.geolytics.com/USCensus.Neighborhood-Change-Database-1970-2000.Products.asp>



Our Geolytics CensusCDs Holdings

<http://www.geolytics.com/>

- Easier to extract data and export as shapefiles
- Licensed data: **in-library use only**
- **Geolytics Census Products:**
 - CensusCD 1970 (down to **tracts**)
 - CensusCD 1980 (down to **block groups**)
 - CensusCD Blocks (1990) (**blocks** only)
 - CensusCD + Maps (1990) (down to **block groups**)
 - CensusCD 2000 Short Form (down to **block groups**)
 - CensusCD 2000 Short Form Blocks (**blocks** only)
 - CensusCD 2000 Long Form (down to **block groups**)
 - CensusCD 2000 Redistricting (earliest & simplest data – down to **blocks**)
 - CensusCD Neighborhood Change Database (NCDB) 1970-2000 (**tracts** only)
 - Estimates, Projections, Consumer Expenditures and Profiles (down to **block groups**)
 - Estimates & Projections – limited variables only (I don't recommend the estimates...)
 - Consumer Expenditures – based on BLS survey on "how people spend money, and on what". Broken down by small geographic levels



Historical Census Data Sources

- Another source for historical census data:
 - **Social Explorer**
<http://www.socialexplorer.com>
 - By CUNY Queens College, by a group of academic sociologists
 - Data source: NHGIS
 - Good for beginners
 - Easier to generate summary data reports and also tract-level data tables
 - No shapefiles download, just attribute data



Historical Census Reports

- If data is not available online, you may need to use “print/paper” reports instead!
 - **Online PDFs**
 - Most of them were digitized by the Census Bureau
 - <http://www.census.gov/prod/www/abs/decennial/index.htm>
 - **Library resources**
 - NU Library (us = Govinfo) has probably all the historical reports in print, including **hard to find 1920, 1930, 1934** (special census) Chicago census tract data books !

Major Change in 2010!

- **End of decennial census long form**
 - Decennial will focus on counting population
 - Short form only with 7 Questions (name, sex, age, relationship, race/Hispanic, owner/renter)
- Replacement:
 - **American Community Survey**
http://www.census.gov/acs/www/UseData/Compass/compass_series.html
 - Will replace decennial census's long form questionnaire in 2010
 - If you know you will need detailed demographic data for your research after 2010, you should look at [ACS presentations](#)

American Community Survey

- **VERY COMPLEX**: much harder to interpret than decennial
 - Unlike decennial, ACS is not a **point-in-time** survey
 - ACS are **period** estimates collecting survey information continuously nearly every day of the year and then aggregates the results over a specific period of time (**1-year, 3-year, 5-year**)
 - Data becomes available, depending on population size of the geographic areas (**more than 65K, 20K-65K, < 20K**)
 - And also smaller sampling rates:
 - 1-year: average 1 in every 40 households (2.5%)
 - 3-years: average 1 in every 13 households (7.5%)
 - 5-years: average 1 in every 8 households (12.5%)

American Community Survey

- Census long form vs. ACS
 - Pay attention to “margin of error (ME)” information
 - How to use ME to conduct significance testing
<http://www.edthefed.com/presentations/significance%20testing.ppt>
 - Not available for small geographies (i.e. tract/BG) **until 2010**
 - The first small area data will be released (hopefully) in 2010
 - (I believe it will be) 5-year estimates: 2005-2009
 - Data are hard to compare – same variable doesn't mean they represent same thing! (i.e. migration, income, employment status, rent & h. value)
<http://www.census.gov/acs/www/UseData/compACS2008.htm>
 - Pay attention to the table of variables with “comparison notes”
 - “Compare”
 - “Compare with caution”
 - “Do not compare”

Similarities & Differences

Similarities	Differences (selected)
<p>Samples The fact that both are based on samples</p>	<p>Frequencies Census – Point-in-time survey in every 10 years ACS – monthly (period) surveys in every year</p>
<p>Variables (themes) Mostly, the same variables set (sample/long form)</p>	<p>Subjects Census – residents of usual place ACS – 2-month rule</p>
	<p>Sampling Sampling method and rates, as well as weighting rate, are different</p>
	<p>Data collection ACS data are collected by census workers not temp staff – probably less operational errors, but with less publicity (thus, fewer responses)</p>

... actually, there is more differences...

New Variable in ACS (2000-)

- Fertility**
- Universe (subject): Women 15 to 50 years old in households

Map on the right:
Showing ratio of unmarried mother to all mothers (i.e. mother = who had a birth in the past 12 months)

New Variable in ACS (2008-)

- Health Insurance**
- Universe (subject): Civilian non-institutionalized population

Map on the right:
Showing ratio of people without private health insurance to all people

Other Selected Censuses

- Census Transportation Planning Package**
<http://www.fhwa.dot.gov/ctpp/>
 - Formerly known as **Journey-to-work** data
 - Special tabulation based on decennial census
 - Tabulated not just by **place of residence** but also **place of work** (and also **flows** between residence and workplace)
 - Minimum at tract or BG or TAZ (traffic analysis zone) levels (depends on metropolitan areas)

Other Selected Censuses (Cont.)

- Agricultural census**
<http://www.agcensus.usda.gov/>
 - Since 1840 (since 1997, conducted by USDA)
 - Once in every 5 years (years ending "2" and "7")
 - Demographic (i.e. information of operators) & economic information (i.e. landuse, total production of crops and livestock, values of sales, and farming practices)

Other Selected Censuses (Cont.)

- Economic census**
<http://www.census.gov/econ/>
 - Formerly known as manufactures census (since 1905)
 - Once in every 5 years (years ending "2" and "7")
 - Foundation for the nation's system of statistics about the functioning of American economy
 - Statistics for U.S., states, metro areas, counties, and cities.

Total Annual Payroll for Mining: 2002

Thematic Mapping

Graduated Colors (most common)

- Age Under 10 (no normalization)

Graduated Colors (cont.)

- Age Under 10 (normalized by total population = % Age Under 10)

Classifications

Natural Breaks (Jenks)

- Classes are based on **natural groupings** of data values
 - Breaks are set where there is a jump in values, indicated by a large step between bars.
- Pros:** good for not-evenly distributed values
- Cons:** finding (computing) good break points are difficult – doesn't always work well

Classifications

Quantile

- Each class contains an equal number of features
- Pros:** good for evenly distributed values. Also, good for comparing area that are roughly same size
- Cons:** Not appropriate for different value sets greatly different values can end up in the same class

Classifications

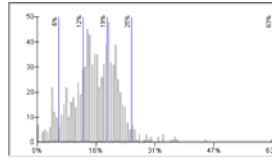
Equal (or defined) Interval

- Equal:** each class has an equal range (interval) of values
- Defined:** each class has specified interval value, the # of classes would be determined based on the interval values
- Pros:** easier to understand, works especially for evenly distributed values
- Cons:** Not appropriate for clustering values (not evenly distributed values)

Classifications

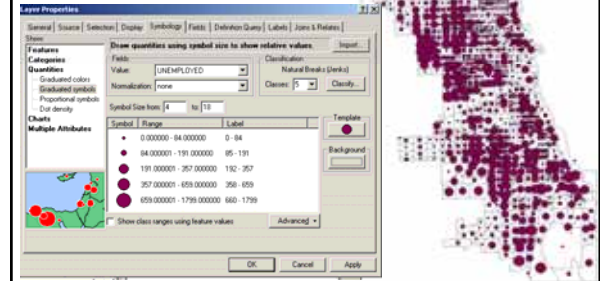
Standard Deviation

- Each class is defined by its distance from the mean value of all the features
- Pros:** good to see which features are above or below mean
- Cons:** Without basic statistics knowledge, it is hard to understand



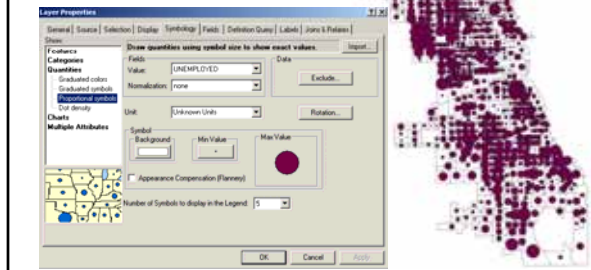
Graduated Symbols

- Unemployed Population



Proportional Symbols

- Unemployed Population
- Similar to graduated symbols but represent quantity more accurately



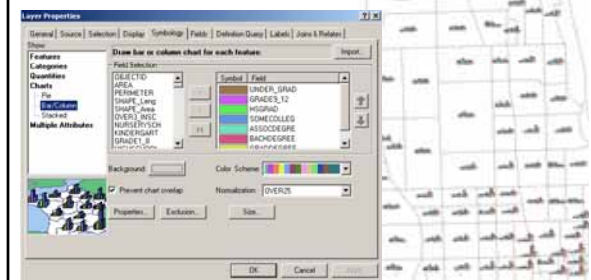
Dot Density Maps

- Housing Built Year
- Dots are distributed randomly within each area



Charts (Pie/Bar/Stacked)

- Educational Attainment (normalized by universe)



Reference

- More about symbolization
 - http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=About_symbolizing_data_to_represent_quantity
- More about classifications
 - http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Standard_classification_schemes

Exercise 6

