

03 – Social Demography Concepts and Databases

Ness Sandoval
Sociology
Saint Louis University

Outline

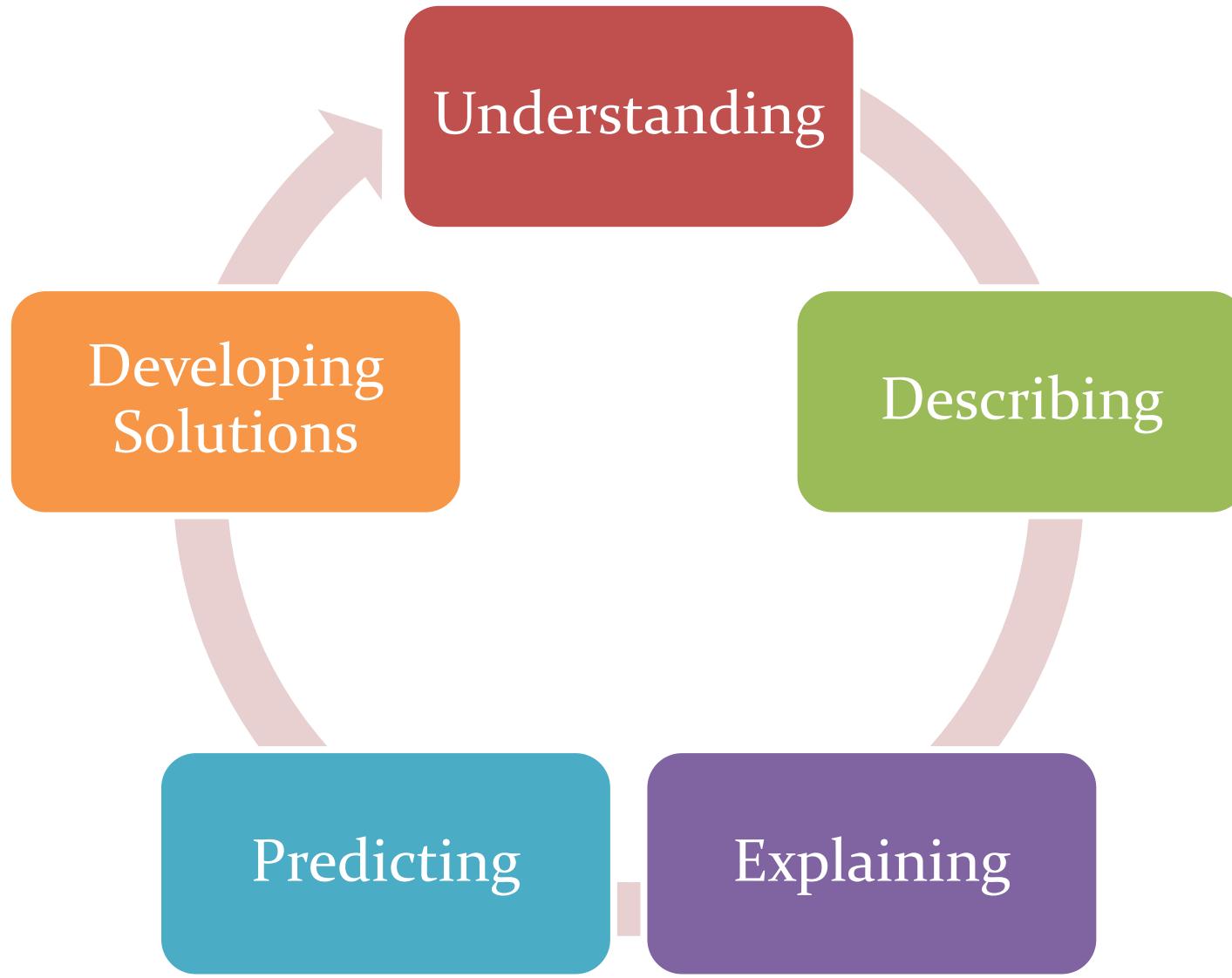
- Thinking Spatially in the Social Sciences
- Basic Social Demographic Measures
- Translation of Social Demographic Measures

Thinking Spatially in the Social Sciences

"We are drowning in information, while starving for wisdom. The world henceforth will be run by synthesizers, people able to put together the right information at the right time, think critically about it, and make important choices wisely."

-*Edward O. Wilson in "Consilience: The Unity of Knowledge" (1998)*

The Importance of Space



Center for Spatially Integrated Social Science

(<http://www.csiss.org/>)

- Founded in 1999 by an NSF
- Multidisciplinary
 - Computational obstacles have gone away
 - Accepted as axiomatic that space matters
 - Every year new data geo-coded data is becoming available
 - Science should be user-oriented
 - Creative play and experimentation with ideas



NESCent

National Evolutionary Synthesis Center

search NESCent site

GO

[about](#)[people](#)[news](#)[contact](#)[donate](#)[SCIENCE](#)[INFORMATICS](#)[EDUCATION & OUTREACH](#)

[Home](#) > [About The Center](#)

[About the Center](#)[Partners](#)[Employment](#)[Information for Visitors](#)[People](#)[Contact Us](#)[Driving Directions](#)[FAQ](#)

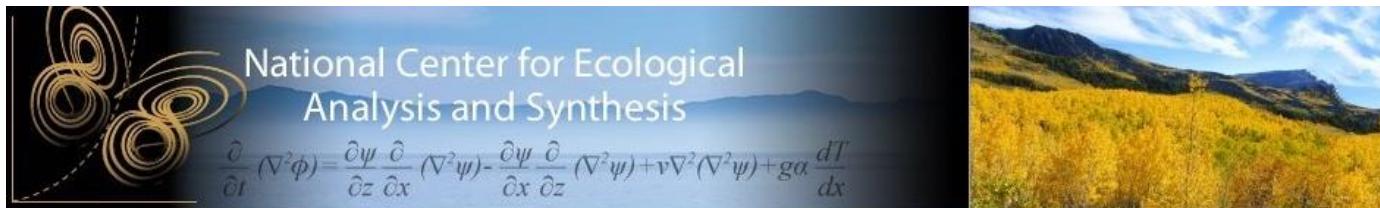
About the Center

The National Evolutionary Synthesis Center (NESCent) is a nonprofit science center dedicated to cross-disciplinary research in evolution. NESCent is jointly operated by Duke University, The University of North Carolina at Chapel Hill, and North Carolina State University, and is sponsored by the National Science Foundation.



Our Mission

NESCent promotes the synthesis of information, concepts and knowledge to address significant, emerging, or novel questions in evolutionary science and its applications. NESCent achieves this by supporting research and education across disciplinary, institutional, geographic, and demographic boundaries.



- Overview
- The Impact of NCEAS
- People
- Collaborators
- Computing at NCEAS
- Support NCEAS

Overview

The National Center for Ecological Analysis and Synthesis (NCEAS), located in downtown Santa Barbara, is a research center of the University of California, Santa Barbara. NCEAS supports cross-disciplinary research that uses existing data to address major fundamental issues in ecology and allied fields, and encourages the application of science to management and policy. NCEAS is a unique institution with an explicit mission to foster synthesis and analysis, turn information into understanding and, through effective collaboration, alter how science is conducted. NCEAS' success is [evident in the broad impact](#) of its research and programs.

The NCEAS mission is to:

- Advance the state of ecological knowledge through the search for general patterns and principles in existing data
- Organize and synthesize ecological information in a manner useful to researchers, resource managers, and policy makers addressing important environmental issues
- Influence the way ecological research is conducted by promoting a culture of synthesis, collaboration, and data sharing

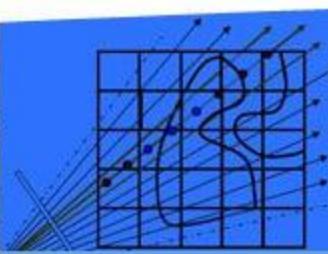


© Jeff Jones

NCEAS **facilitates** integrative research aimed at synthesizing existing data and information, and subsequently [making these data and inferences widely available](#). NCEAS fosters new techniques in mathematical and geospatial modeling, dynamic simulation, and visualization of ecological systems.



National Socio-Environmental Synthesis Center



The National Socio-Environmental Synthesis Center is a research center dedicated to creating synthetic, actionable science related to the structure, functioning, and sustainability of socio-environmental systems.



WHAT IS ACTIONABLE SCIENCE?

[Read why it's important](#)

RESEARCH THEMES

SESYNC uses a [thematic structure](#) to organize synthesis efforts



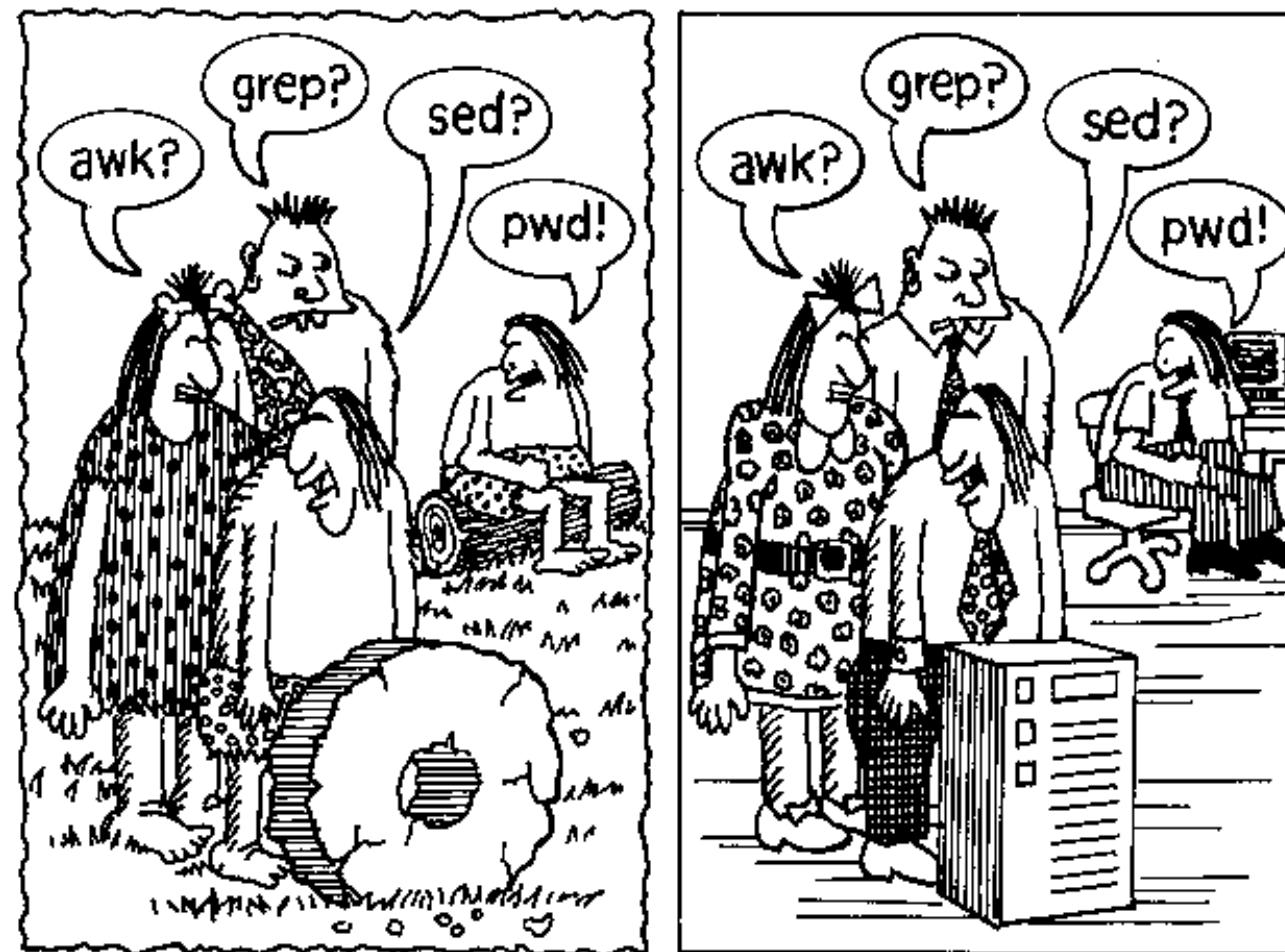
ALWAYS ACCEPTING

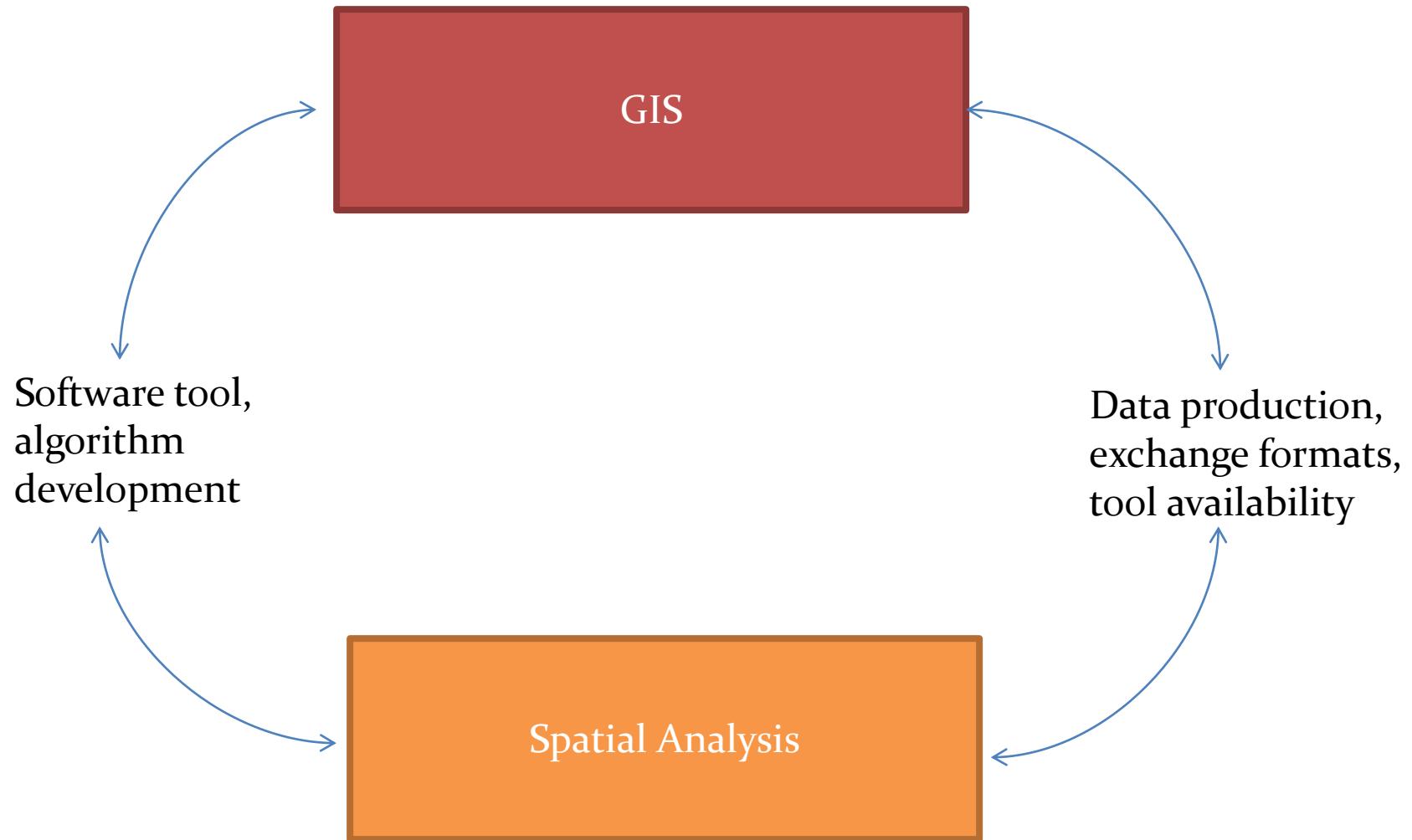
Ventures
Workshops
Short Courses
Research Fellows



SESYNC [programs](#) support the [synthesis](#) of natural and social science in addressing problems of the environment. Fostering fundamental, discovery-driven synthesis research that contributes to [actionable](#) science and policy-relevant applications.

What Changed?





Spatial Perspectives

Functional
Distance

Spatial
Position

Spatial
Order

Basic Social Demographic Measures

Statistics	St. Louis	
SE:T4. Sex		
Total Population:	2,801,914	
Male	1,357,051	48.4%
Female	1,444,863	51.6%
SE:T8. Age (Detailed version)		
Total:	2,801,914	
Under 5 Years	170,736	6.1%
5 to 9 Years	177,636	6.3%
10 to 14 Years	183,488	6.6%
15 to 17 Years	112,825	4.0%
18 and 19 Years	70,044	2.5%
20 Years	34,270	1.2%
21 Years	38,274	1.4%
22 to 24 Years	109,084	3.9%
25 to 29 Years	190,123	6.8%
30 to 34 Years	187,772	6.7%
35 to 39 Years	171,349	6.1%
40 to 44 Years	176,327	6.3%
45 to 49 Years	191,525	6.8%
50 to 54 Years	214,794	7.7%
55 to 59 Years	201,243	7.2%
60 and 61 Years	72,897	2.6%
62 to 64 Years	95,919	3.4%
65 and 66 Years	55,411	2.0%
67 to 69 Years	70,825	2.5%
70 to 74 Years	94,955	3.4%
75 to 79 Years	71,141	2.5%
80 to 84 Years	55,188	2.0%
85 Years and Over	56,088	2.0%

Statistics	St. Louis	
SE:T14. Hispanic or Latino by Race		
Total Population	2,801,914	
Not Hispanic or Latino:	2,723,282	97.2%
White Alone	2,085,319	74.4%
Black or African American Alone	509,235	18.2%
American Indian and Alaska Native Alone	4,490	0.2%
Asian Alone	64,212	2.3%
Native Hawaiian and Other Pacific Islander Alone	721	0.0%
Some Other Race Alone	3,608	0.1%
Two or More Races	55,697	2.0%
Hispanic or Latino:	78,632	2.8%
White Alone	55,789	2.0%
Black or African American Alone	2,219	0.1%
American Indian and Alaska Native Alone	543	0.0%
Asian Alone	540	0.0%
Native Hawaiian and Other Pacific Islander Alone	216	0.0%
Some Other Race Alone	13,035	0.5%
Two or More Races	6,290	0.2%

Sex Ratio

The number of males per hundred females

$$\frac{\text{number of males}}{\text{number of females}} * 100 \text{ or } \frac{M}{F} * 100$$

Males = 1,357,051

Females = 1,444,863

$$\text{Sex ratio} = \frac{1,357,051}{1,444,863} * 100 = 94 \text{ males per hundred females}$$

Dependency Ratio

The number of children and aged person per hundred people of working age

$$dr = \frac{P_{0-14} + P_{65+}}{P_{15-64}} * 100$$

Numbers aged 0-14 = 531,860

Numbers aged 15-64 = 1,866,446

Numbers aged 65+ = 403,608

$$dr = \frac{531,860 + 1,866,446}{403,608} * 100$$

dr=50 dependent individuals per hundred working age individuals

Child Dependency Ratio

The number of children per hundred people of working age

$$cdp = \frac{P_{0-14}}{P_{15-64}} * 100$$

Numbers aged 0-14 = 531,860

Numbers aged 15-64 = 1,866,446

Numbers aged 65+ = 403,608

$$cdp = \frac{531,860}{1,866,446} * 100$$

cdp=28 dependent children per hundred working age individuals

Senior Dependency Ratio

The number of aged people per hundred people of working age

$$adr = \frac{P_{65+}}{P_{15-64}} * 100$$

Numbers aged 0-14 = 531,860

Numbers aged 15-64 = 1,866,446

Numbers aged 65+ = 403,608

$$adr = \frac{403,608}{1,866,446} * 100$$

adr=22 dependent seniors per hundred working age individuals

Ageing Index

The number of people aged per hundred children

$$ai = \frac{P_{65+}}{P_{0-14}} * 100$$

Numbers aged 0-14 = 531,860

Numbers aged 65+ = 403,608

$$ai = \frac{403,608}{531,860} * 100$$

ai=76 seniors per hundred youth

Diversity Score

$$E_i = \left| \sum_{r=1}^n \frac{(\ln(P_{(ik)}) * P_{(ik)})}{(\ln(n))} \right|$$

where:

E_i = Diversity score for tract j

$P_{(ik)}$ =Proportion of the tract population in race/ethnic group

n=the total number of racial/ethnic categories

Diversity score will range from 0=(No Diversity) to 1=(Very Diverse)

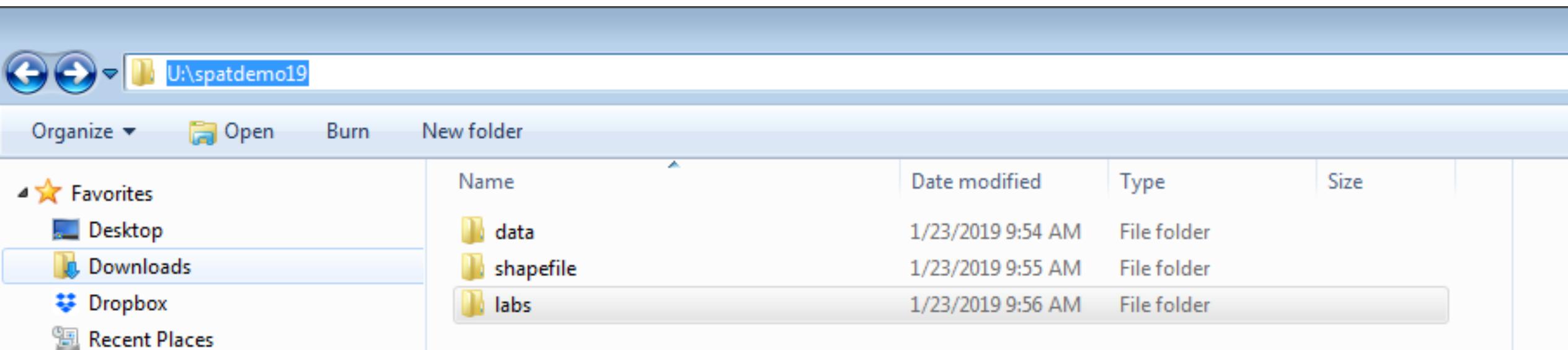
The racial diversity score for the U.S. based on five racial categories is .676

	White	Black	Asian	Latino	Other	Total
Population	2,085,319	509,235	64,933	78,632	63,795	2,801,914
Percent	74%	18%	2%	3%	2%	100%
$\ln(P_{ik})^*P_{ik}$	-0.220	-0.310	-0.087	-0.100	-0.086	0.499

Lab Exercise

Part One – Getting Organized and Creating your Social Data Variables

Step 1. – Let's get organize – Set up a folder and sub-folders



Step 2. – Let's get our data – www.socialexplorer.com

We will use the ACS 2017 5-Year Estimates

Make sure you download level 140 – census tracts

ACS 2017 (5-Year Estimates)

Choose survey/year ▶ Geographies ▶ Tables ▶ Results

List Geographies FIPS Codes

Select a geographic type: ([Show all geographies](#))

140 Census Tract

Select a State:

Missouri

Select a County:

Select a County

Select one or more geographic areas and click 'add':

All census tracts
All census tracts in Missouri

Add ▾ Remove ▲

Current Geography Selections:

===== Census Tract - 140 =====
All census tracts in Illinois
All census tracts in Missouri

Proceed to Tables ▶

Step 3 – We want to focus on three variables

ACS 2017 (5-Year Estimates)

Choose survey/year ▶ Geographies ▶ Tables ▶ Results

List Tables Search by Keyword Premade Reports

Select a dataset:

Social Explorer Tables: ACS 2017 (5-Year Estimates) ▾

Select one or more tables and click 'add':

- T8. Age (Detailed Version)
- T8A. Age (Detailed Version) - Cumulative (Less)
- T8B. Age (Detailed Version) - Cumulative (More)
- T9. Age (Short Version)
- T10. Age for Male Population (Short Version)
- T11. Age for Female Population (Short Version)
- T12. Median Age by Sex
- T13. Race
- T14. Hispanic or Latino by Race
- T15. Hispanic or Latino by Specific Origin
- T16. Asian by Specific Origin (Asian with One Asian Category for Selected Groups)
- T17. Households by Household Type
- T165. Housing Units by Household Type (Including Living Alone) and Age of Householder

Abbreviations: AIAN - American Indian and Alaska Native; NHPI - Native Hawaiian and Other Pacific Islander

Add ▾ Remove ▲

Current Table Selections:

===== Social Explorer Tables: ACS 2017 (5-Year Estimates) =====

T4. Sex
T8. Age (Detailed Version)
T14. Hispanic or Latino by Race

Show results ▶

Step 4 – We now want to download the data. Make sure you select Tab and download, data, dct and do files and Data Dictionary

ACS 2017 (5-Year Estimates)

Choose survey/year ► Geographies ► Tables ► Results

Report Excel Data Download

Download data in CSV format. CSV stands for Comma Separated Values, and it is a highly portable format that can be used in Microsoft Access. Encoding: Western Latin-1 (ISO-8859-1).

Please set options before you download any data or import programs (these options affect both).

Output options:

TAB delimited files (STATA users should use this option to make import more efficient & accurate!)

Output column labels in the first row

Output ALL geographic identifiers

Output percents (first variable in each table is the base)

Output DBF friendly column names

Download data by geography type:

 [Census Tract data \(TAB\)](#)

Summary Level: 140

Geography Nesting: State-County-Census Tract

Selected: 4516

Download programs to import and label data:

 [SAS](#)

 [SPSS](#)

 [STATA: .dct file .do file](#)

 [Data dictionary \(text file\)](#)

Step 5 - Open STATA and we can start to write our do file and run the program. Line 13 is the most important line.

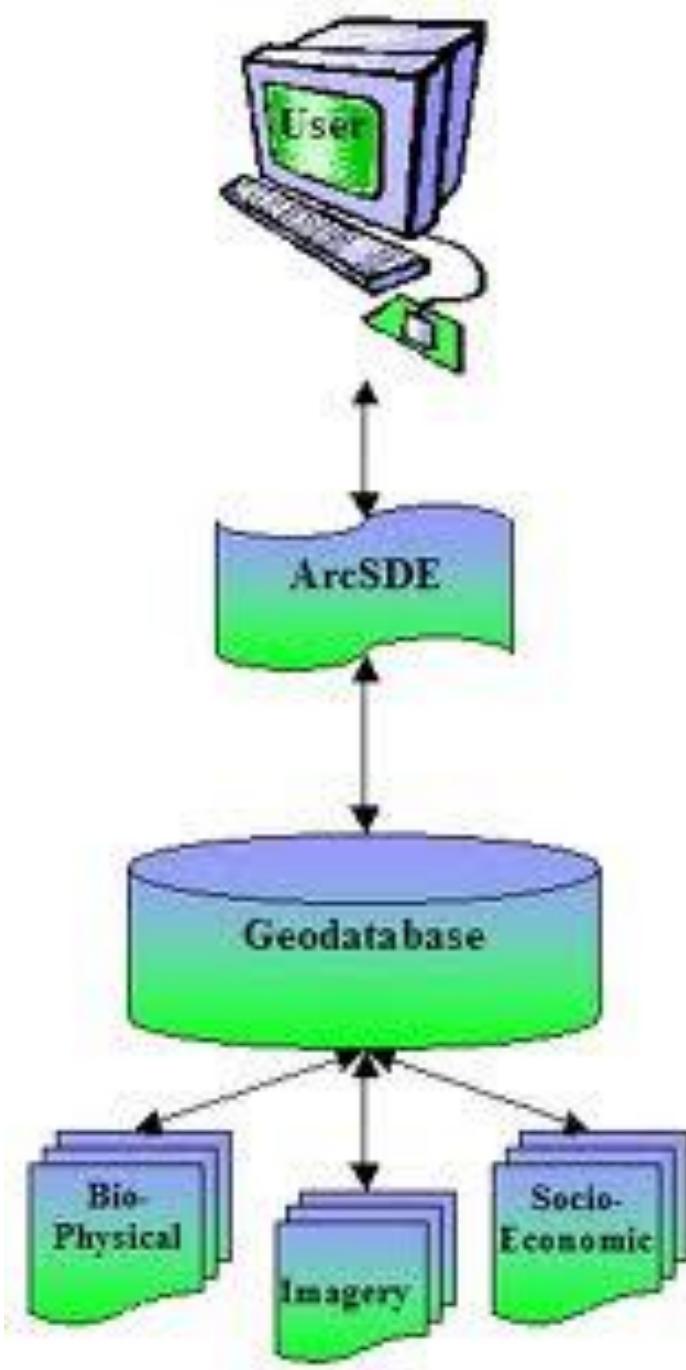
```
10 clear
11 capture log close
12 set more off
13 infile using "R11998837.dct", using("U:\spatdemo19\data\R11998837_SL140.txt")
14 log using analysis.log, replace
15
16 *Sex Ratio
17 gen sr17=T004_002/T004_003
18
19 *Dependency Ratio
20 gen depc17=T008_002+T008_003+T008_004
21 gen age17=T008_005+T008_006+T008_007+T008_008+T008_009+T008_010+T008_011+T008_012+T008_013+T008_014+T008_015+T008_016+T008_017+T008_018
22 gen deps17=T008_019+T008_020+T008_021+T008_022+T008_023+T008_024
23 gen check=(depc17+age17+deps17)-T008_001
24
25 gen dp17=(depc17+deps17)/age17
26
27 *Child Dependency Ratio
28 gen cdp17=depc17/age17
29
30 *Senior Dependency Ratio
31 gen sdp17=deps17/age17
32
33 *Ageing Index
34 gen ai17=deps17/depc17
35
36 *Racial Diversity Score
37 gen oth17=T014_005+T014_006+T014_007+T014_008+T014_009
38
39 gen pwht17=T014_003/T014_001
40 gen pblk17=T014_004/T014_001
41 gen phis17=T014_010/T014_001
42 gen poth17=oth17/T014_001
43 gen ptot17=pwht17+pblk17+phis17+poth17
44
45 gen ewht17=pwht17*ln(pwht17)
46 gen eblk17=pblk17*ln(pblk17)
47 gen ehis17=phis17*ln(phis17)
48 gen eoth17=poth17*ln(pothe17)
49
50 recode ewht17(mis = 0)
51 recode eblk17(mis = 0)
52 recode ehis17(mis = 0)|
```

53 recode eoth17(mis = 0)

```

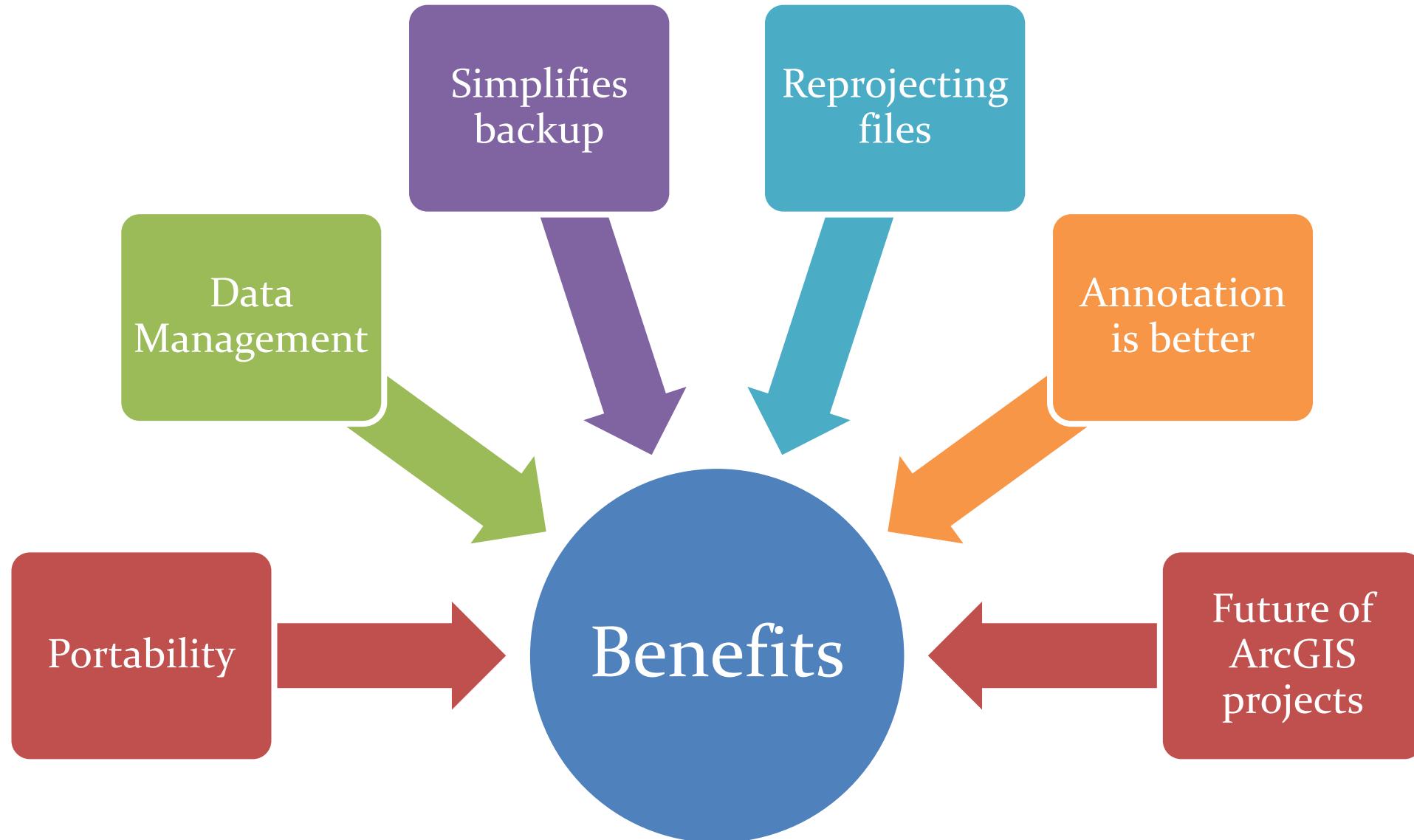
54
55 gen e17=abs(ewht+eblk+ehis+eoth)/ln(4)
56
57 *recode some variables
58 gen wht17=T014_003
59 gen blk17=T014_004
60 gen lat17=T014_010
61 gen tot17=T014_001
62
63 export excel FIPS dp17 cdp17 sdp17 ai17 e17 wht17 blk17 lat17 oth17 tot17 using "U:\spatdemo19\data\part01.xls", firstrow(variables) nolabel replace,
64
65 save "U:\spatdemo19\data\st1_part01.dta", replace
66
67 log close
68 exit
69
```

Part Two – Creating you Geodatabase

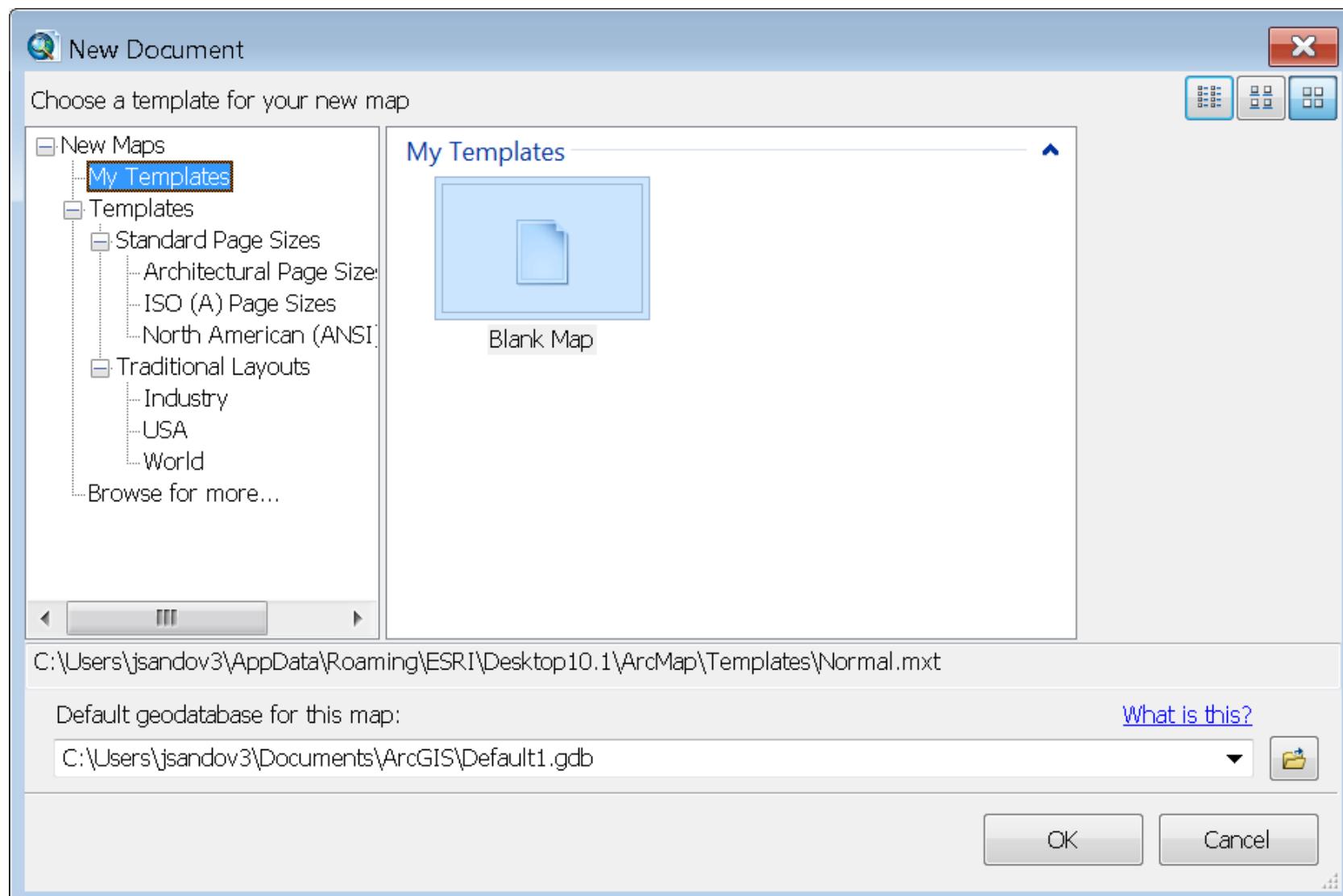


Geodatabases

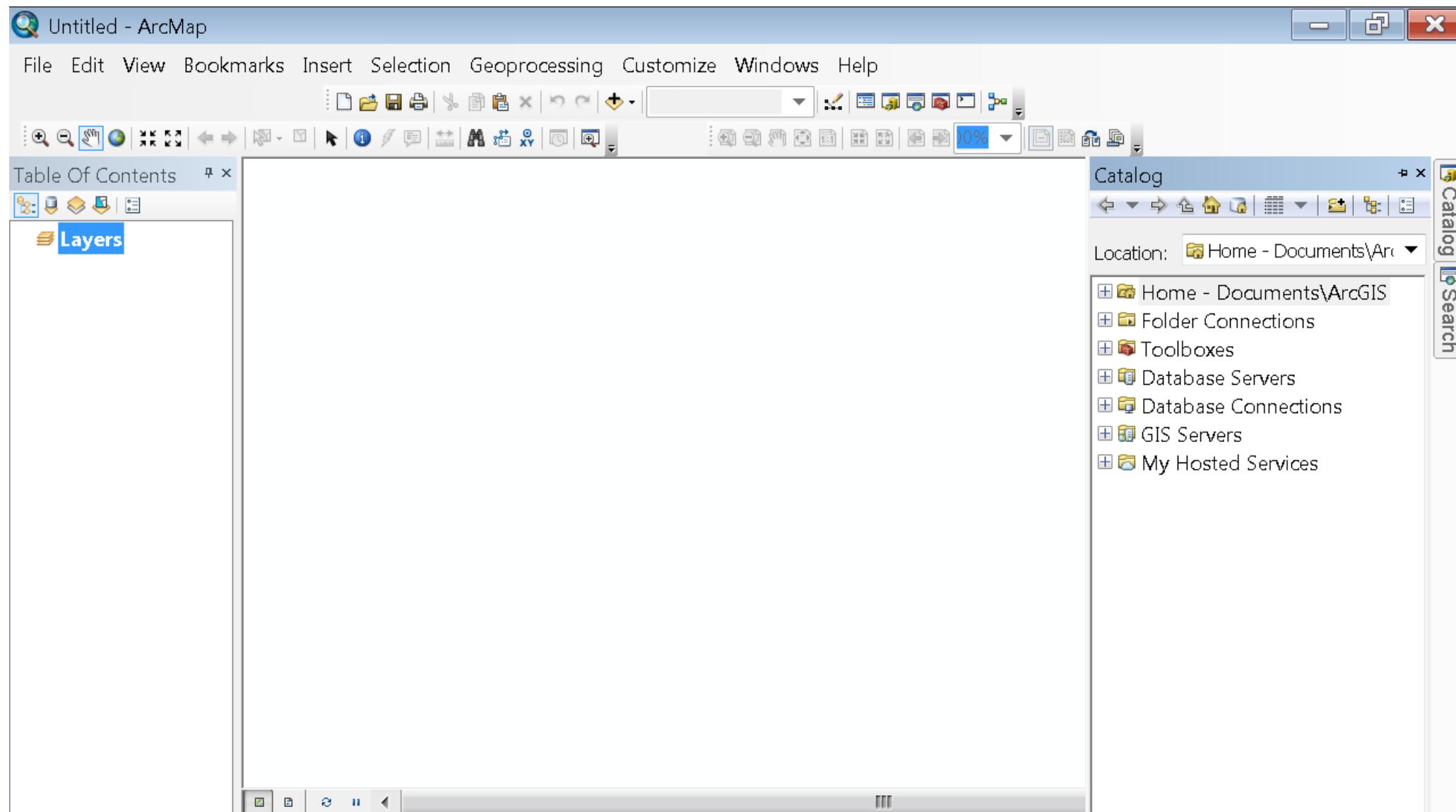
- Solutions to many problems
 - Point and Polygons can be in the same folder
 - We don't lose topology
- Personal
 - Single File
 - Easily portable
 - Data storage limit (2gb of data)
 - Single user at a time
- File Geodatabase
 - Not dependent MS product
 - Portable is problematic (1tb of data)
 - Raster is better
 - Faster processing time



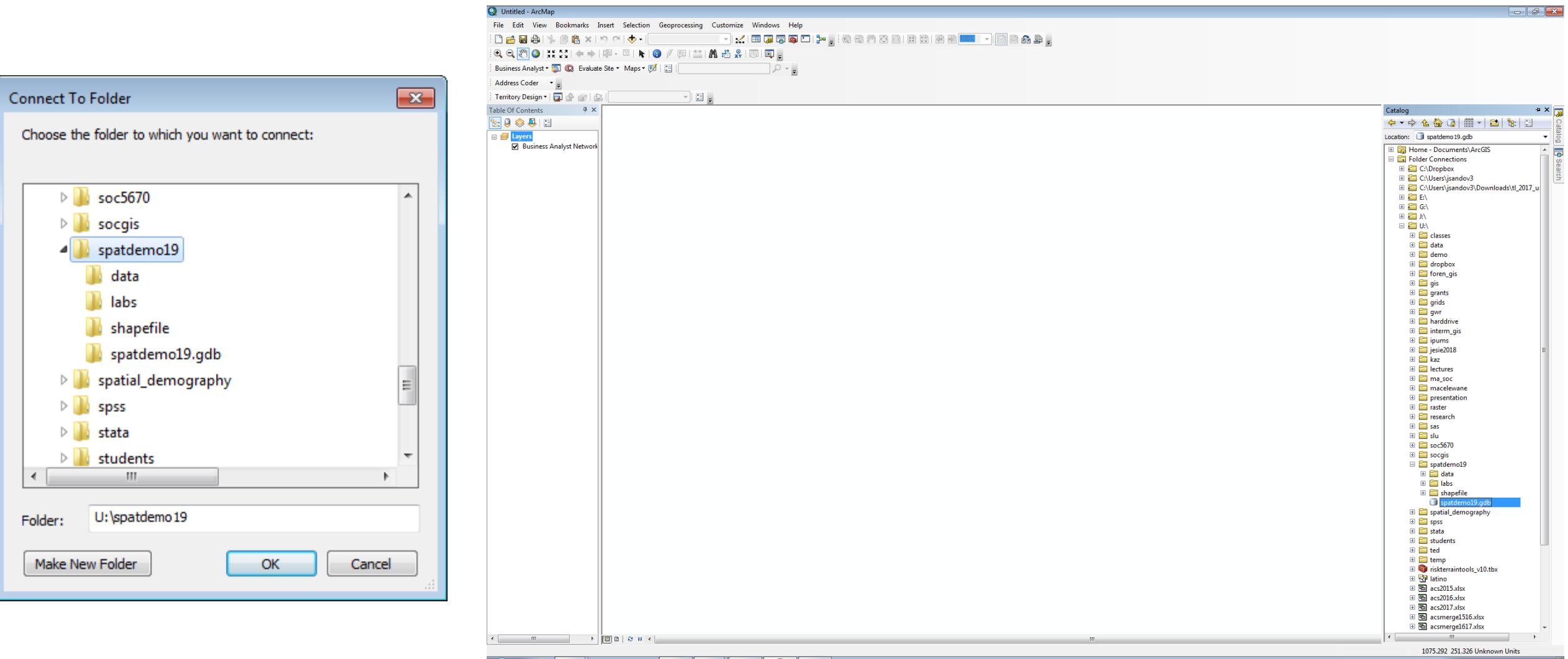
Working with Geodatabase



Creating Geodatabase



- Make sure you have folder connection
- Click New-Right Click-select new file Geodatabase



Part Three – Download your shapefiles from the Census

Step 1 – We now want to download the shapefiles. Go to www.census.gov

NOTICE: Due to a lapse in federal funding this website is not being updated. [Learn More](#)

U.S. Department of Commerce | Blogs | Index A-Z | Glossary | FAQs

United States®
Census
Bureau

TOPICS Population, Economy GEOGRAPHY Maps, Products LIBRARY Infographics, Publications DATA Tools, Developers SURVEYS/PROGRAMS Respond, Survey Data NEWSROOM News, Blogs ABOUT US Our Research

Search 

You are here: [Census.gov](#) > [Geography](#) > [Maps & Data](#) > [TIGER Products](#) > TIGER/Line® Shapefiles and TIGER/Line® Files

Geography

Main About Maps & Data Reference Partnerships Education Research GSS-I Contact Us

Maps & Data

Maps & Data Main Page

Maps

- Census Data Mapper
- Reference
- Thematic
- Maps Available for Purchase

Data

- TIGER Products
- Census Geocoder
- Partnership Shapefiles
- Relationship Files
- Gazetteer Files
- Block Assignment Files
- Name Lookup Tables
- Tallies
- LandView

TIGER/Line® Shapefiles and TIGER/Line® Files

Format:

- Shapefile - 2007 to Present
- TIGER/Line ASCII format - 2006 and earlier
- Census 2000 available in both formats

The core TIGER/Line Files and Shapefiles do not include demographic data, but they do contain geographic entity codes (GEOIDs) that can be linked to the Census Bureau's demographic data, available on [American FactFinder](#).

- [How Do I Choose Which Vintage to Use? \[PDF\]](#)
- [Note on Special Characters Not Displaying Correctly](#)
- [Working with TIGER/Line Shapefiles How-To Guides](#)
- Geography Change & Errata
 - [Geographic Boundary Change Notes](#)
 - [Geography Notes and Errata from the 2010 Census \[PDF\]](#)
 - [Substantial County Changes](#)



2018 2017 2016 2015 2014 2013 113th CD 2012 2011 2010 2009 2008 2007 2006SE Census 2000 1992

2018 TIGER/Line Shapefiles

All legal boundaries and names are as of January 1, 2018. Released September 28, 2018.

Download
Technical Documentation
File Availability
User Notes

The boundaries shown are for Census Bureau statistical data collection and tabulation purposes only; their depiction and designation for statistical purposes does not constitute a determination of jurisdictional boundaries.

- Download Census Tracts for:
 - Illinois
 - Missouri
- Download Core Based Statistical Areas
 - Metropolitan/ Micropolitan Statistical Area (current)

TIGER/Line® Shapefiles

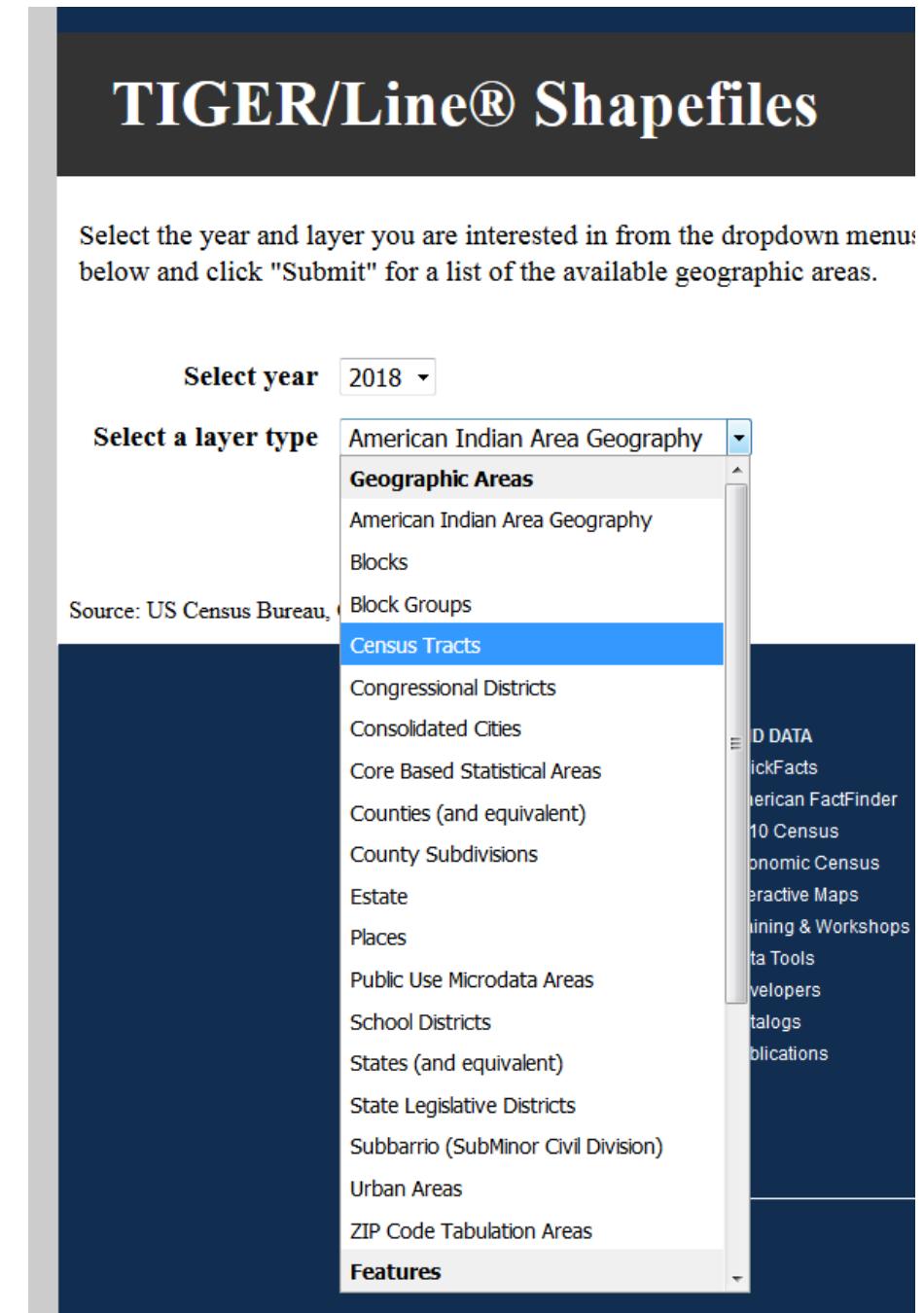
Select the year and layer type you are interested in from the dropdown menus below and click "Submit" for a list of the available geographic areas.

Select year 2018

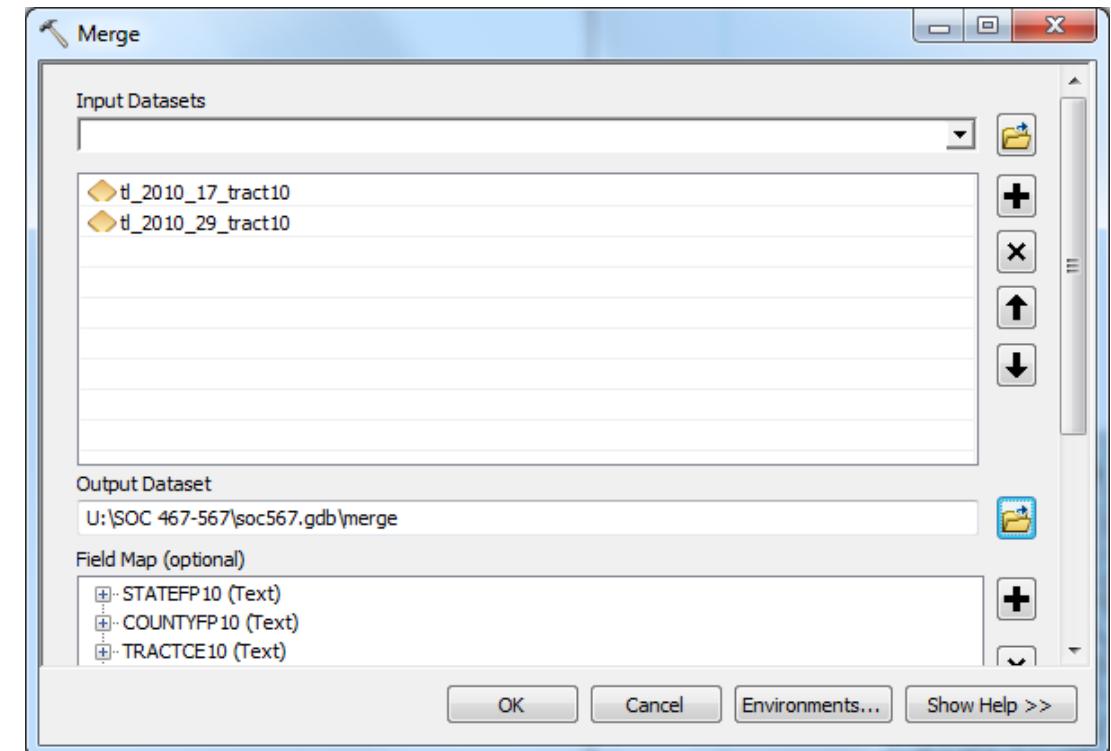
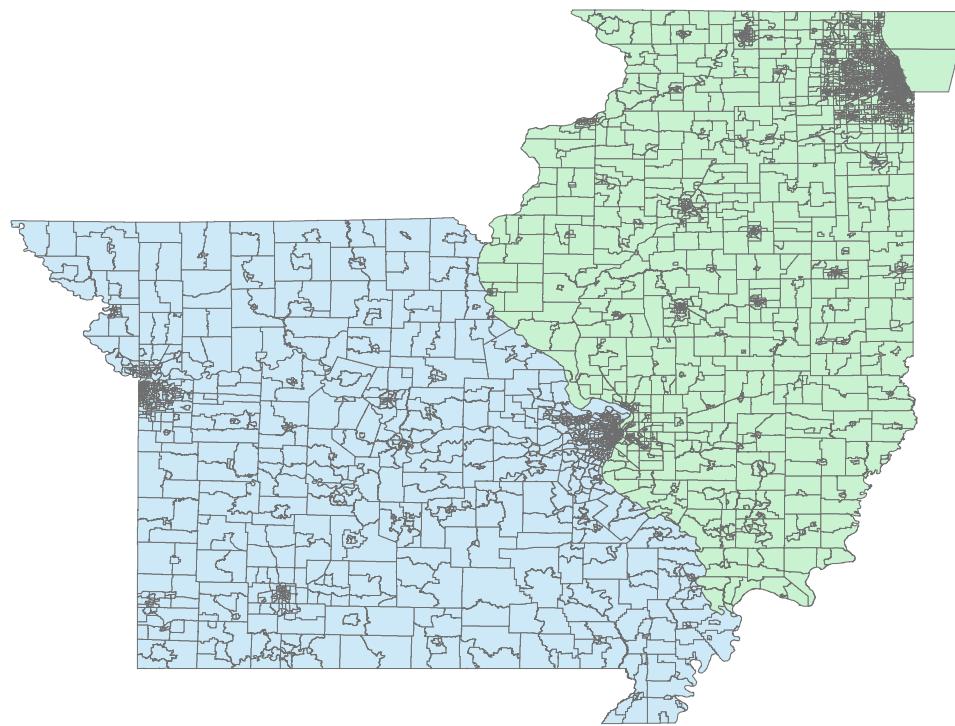
Select a layer type

American Indian Area Geography
Census Tracts
 Geographic Areas
 American Indian Area Geography
 Blocks
 Block Groups
 Congressional Districts
 Consolidated Cities
 Core Based Statistical Areas
 Counties (and equivalent)
 County Subdivisions
 Estate
 Places
 Public Use Microdata Areas
 School Districts
 States (and equivalent)
 State Legislative Districts
 Subbarrio (SubMinor Civil Division)
 Urban Areas
 ZIP Code Tabulation Areas
 Features

Source: US Census Bureau

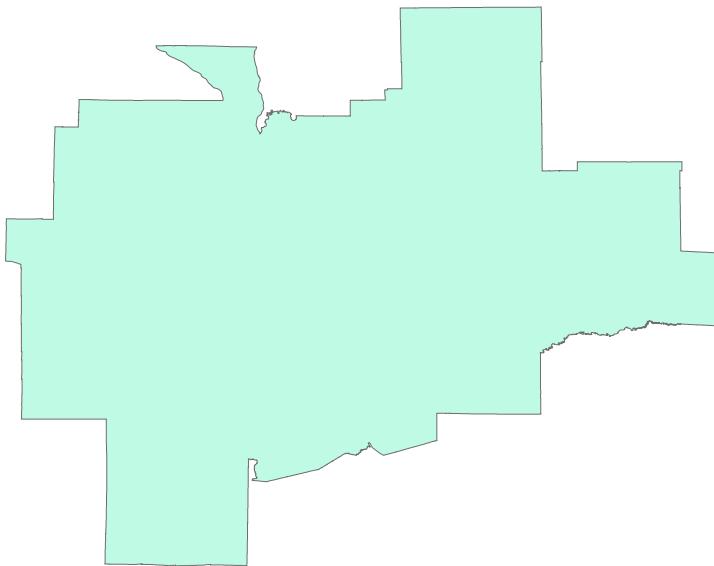


- In ArcGIS use the merge function to create a combined shapefile.

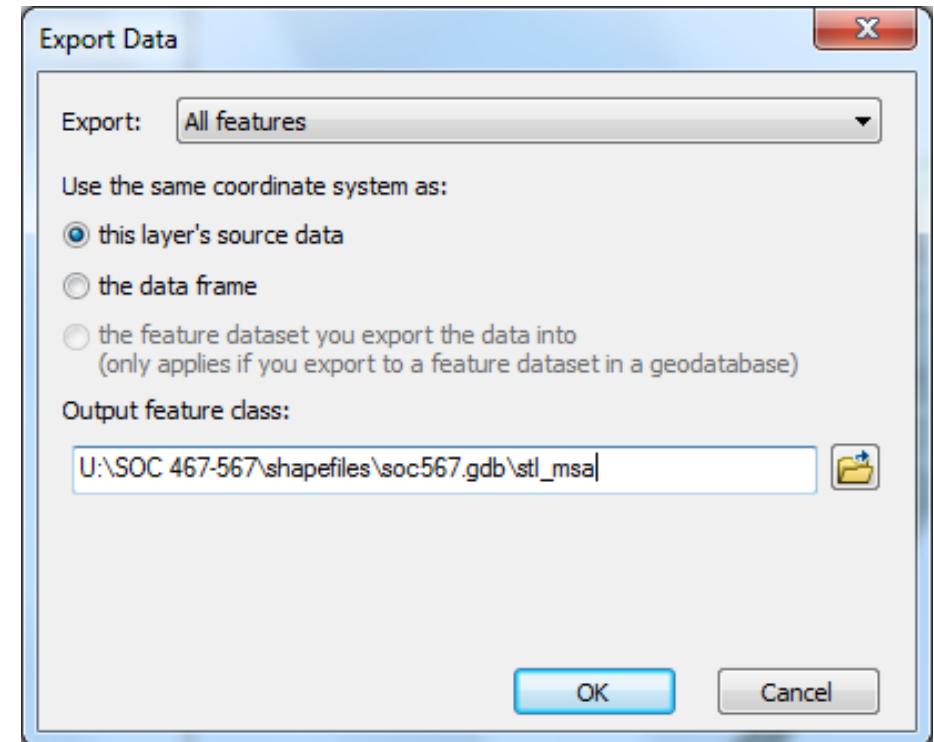


- Download CBSA Boundary Files for US

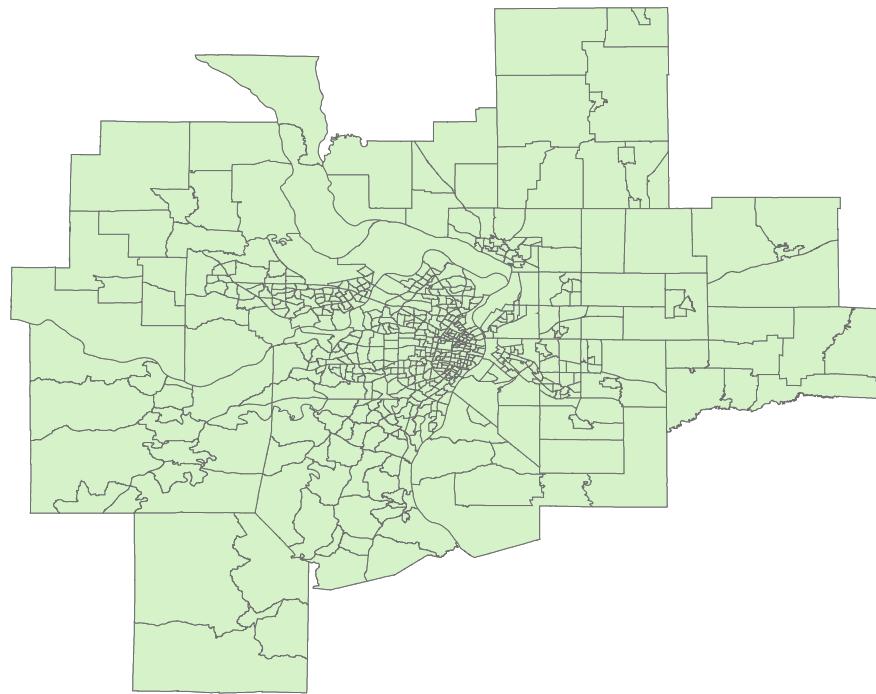
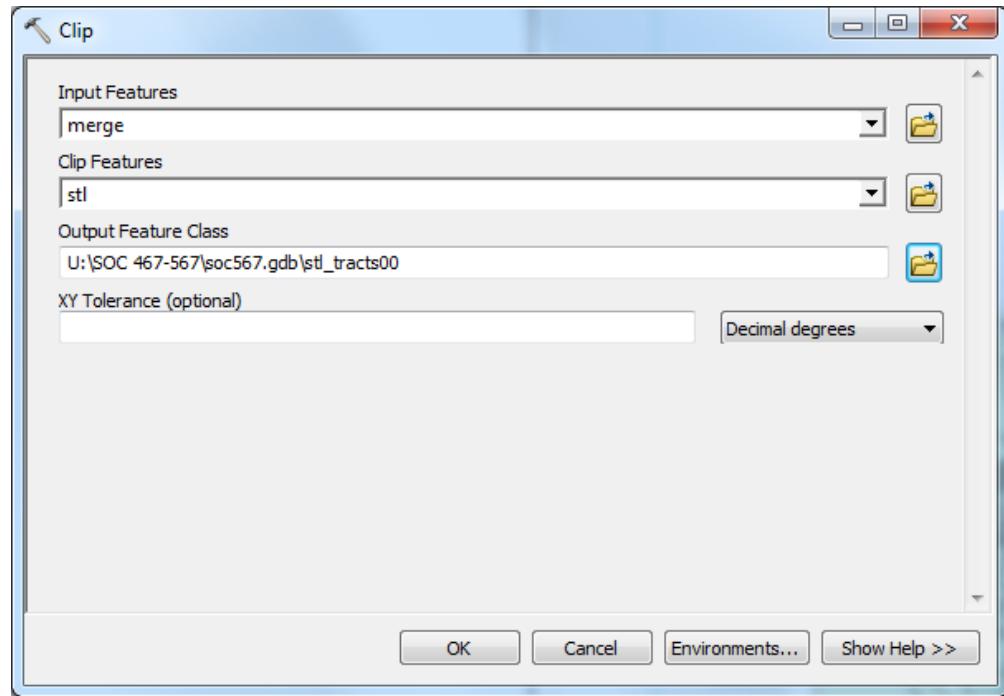
- Query for CBSA Saint Louis (Several Methods)



- Extract Query to new file.

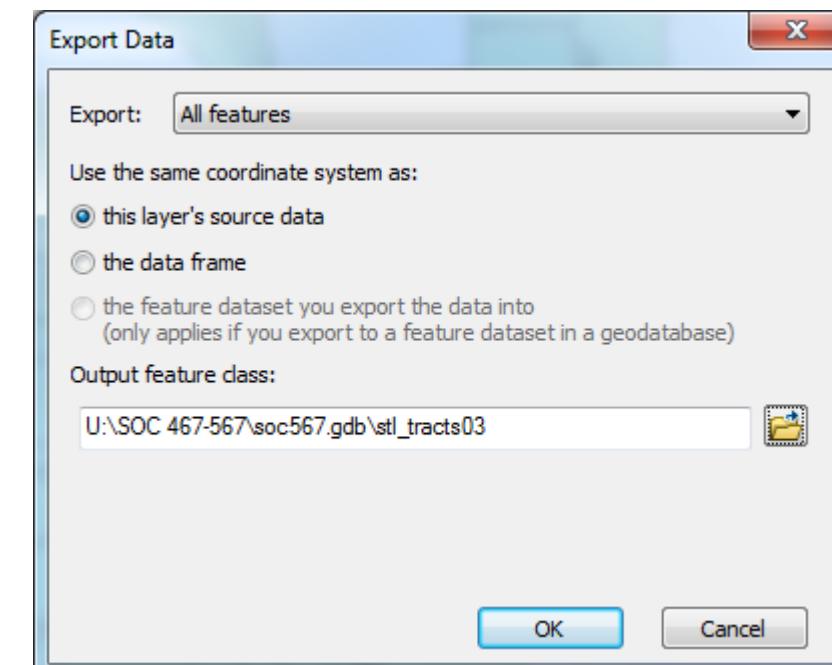
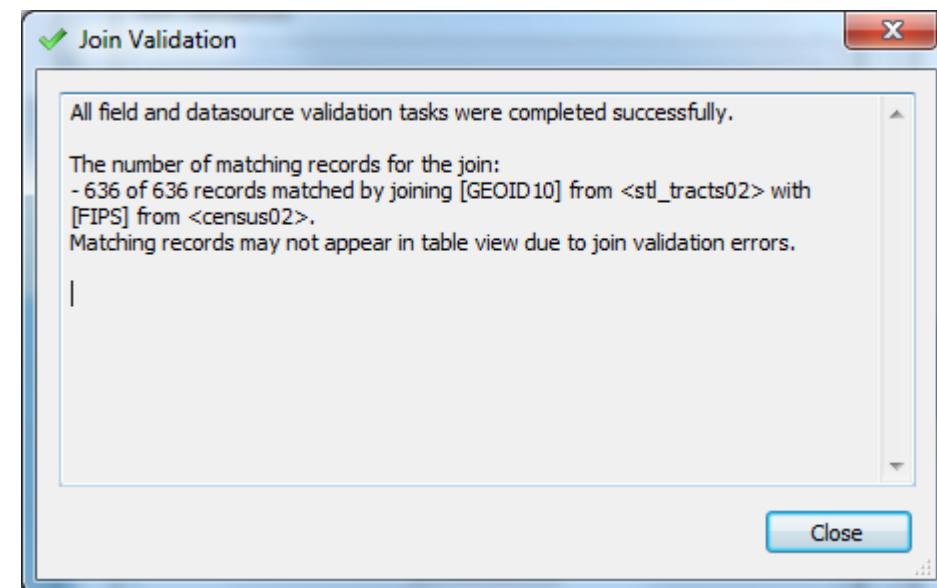
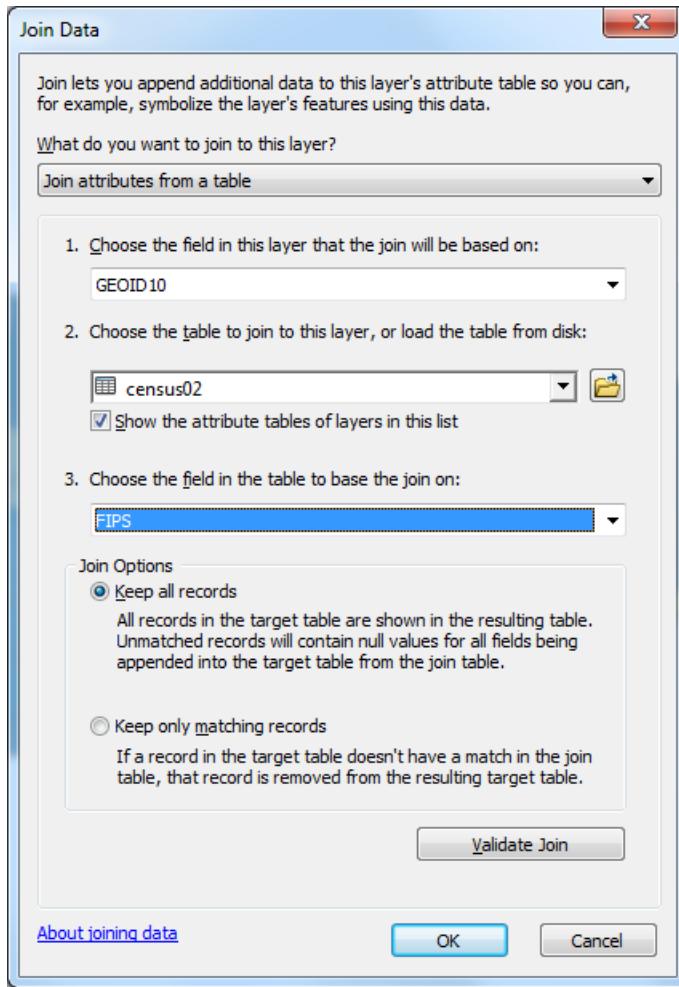


- Clip census tract file to CBSA file



Part Four – Merge Data with Shapefile

- Merge the data file with the shapefile – The merge field should be “GEOID10” and “FIPS”

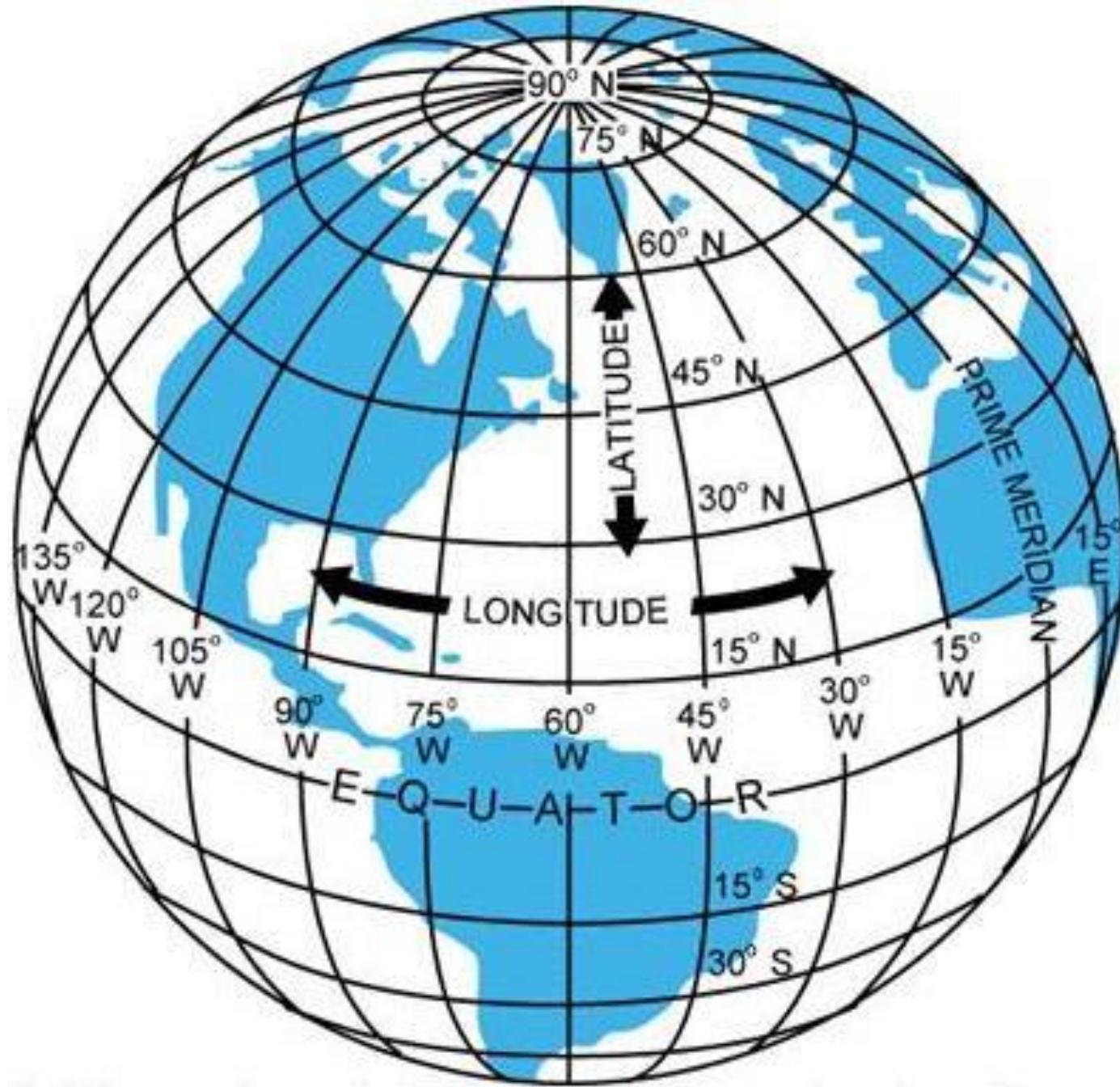


- Make a permanent shapefile

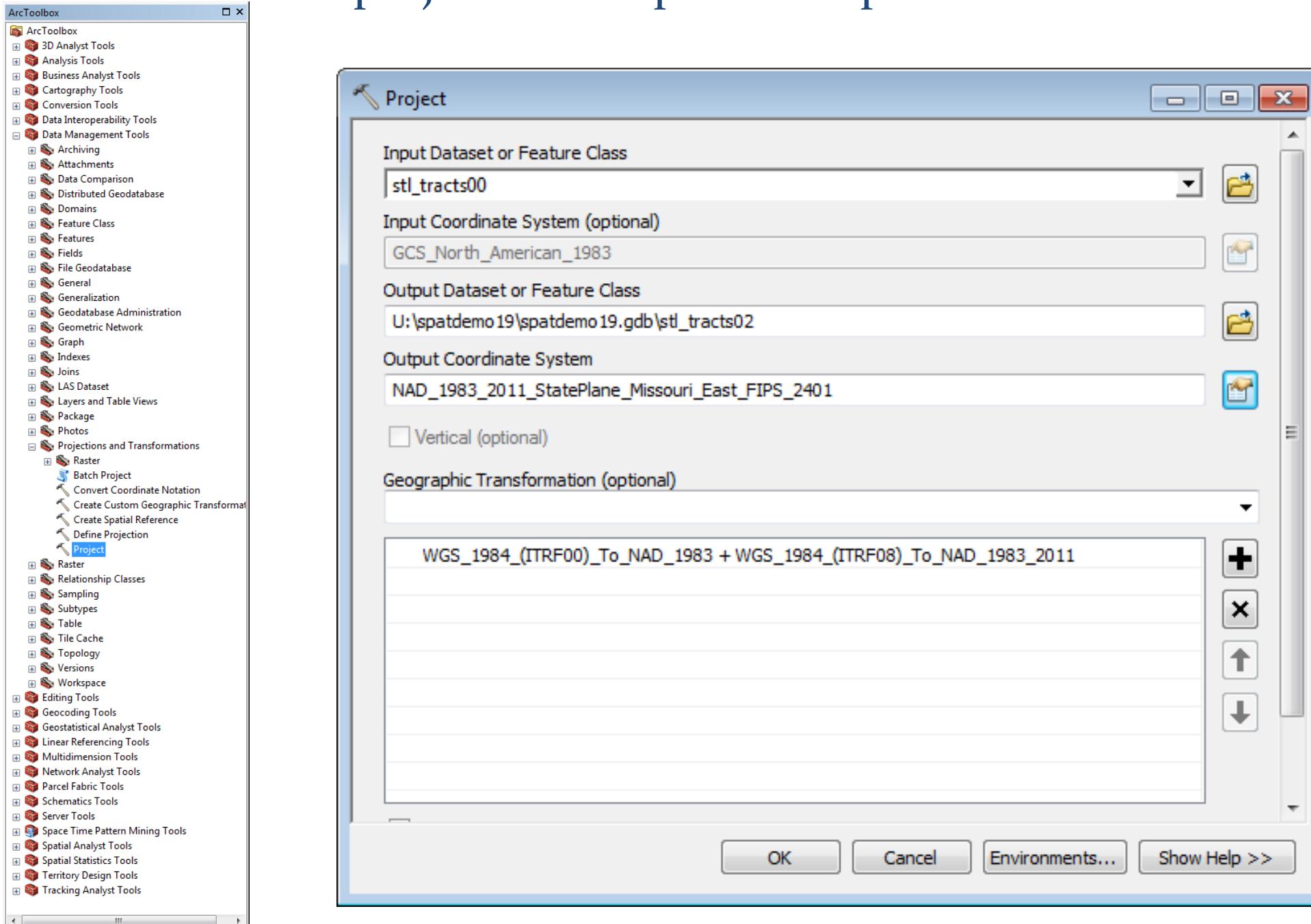
Part Five - Projection



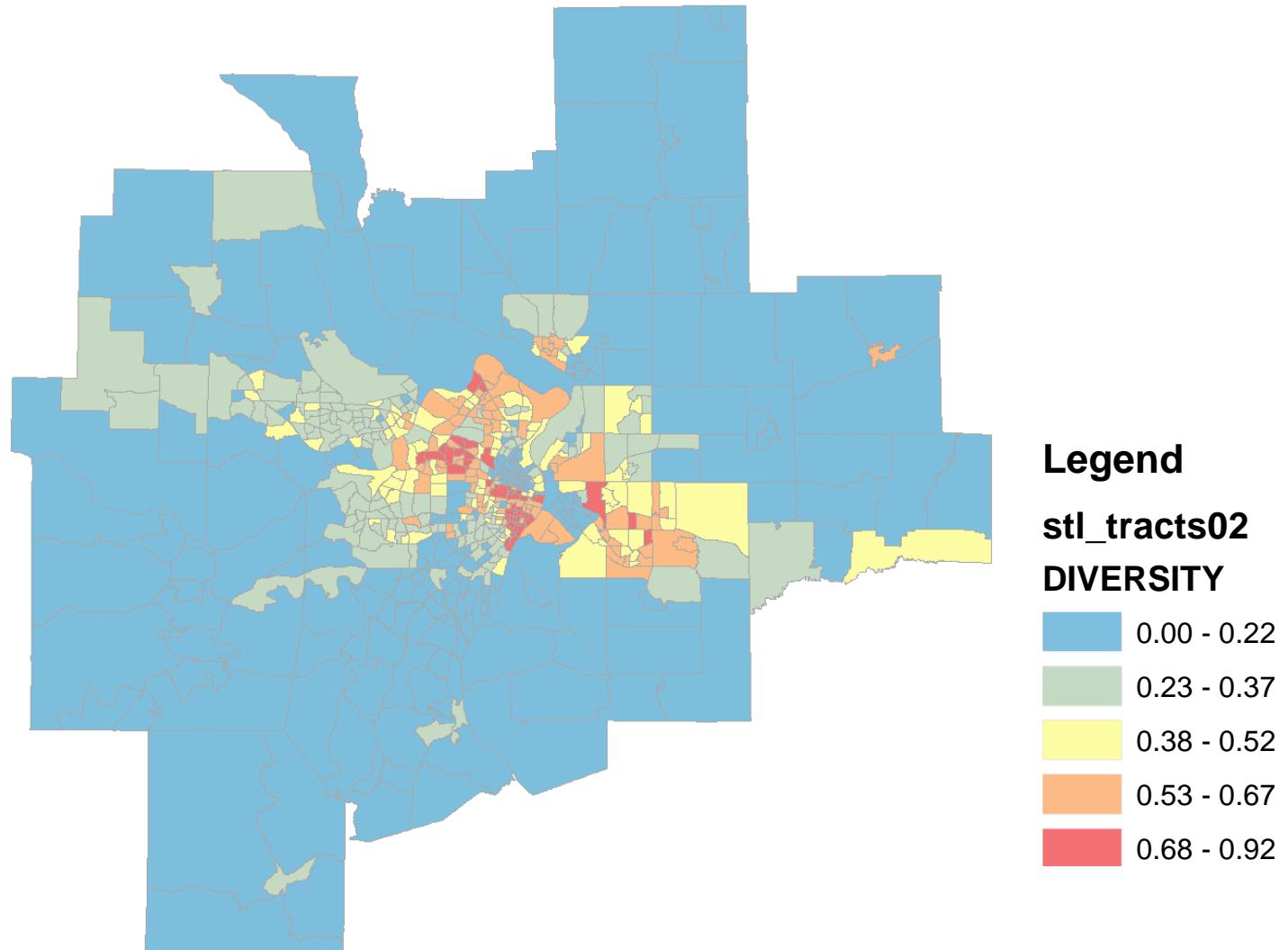
The
Earth is
Round



- Need to reproject the shapefile ->Open ArcToolbox



Make Thematic Maps



Part Six - Data Management

- Rules of Thumb
 - Always store your work in a geodatabase and create a data dictionary
 - Keep your path simple
 - U:\gisc
 - U:\gisc\data
 - U:\gisc\shapefile
 - Keep your variable names simple
 - 10-12 characters
 - Use lowercase
 - Never use spaces for variables names or path
 - Wrong
 - Percent black
 - Right
 - per_black
 - pblk17
 - ro2_17 (percent black for 2017)
 - No special characters
 - Always keep a data dictionary of your work