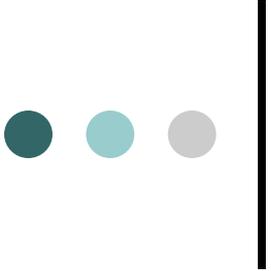




# **GIS Summer Institute**

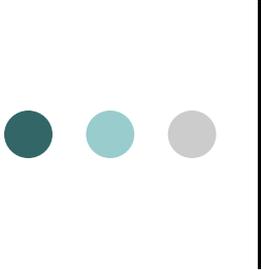
## **Center for Geographic Analysis**

### **Databases & Queries**



# Outline

- Databases
- Joining Tables
  - types of joins
  - spatial or non spatial
- Queries
  - by attribute
  - by location
- US Census data

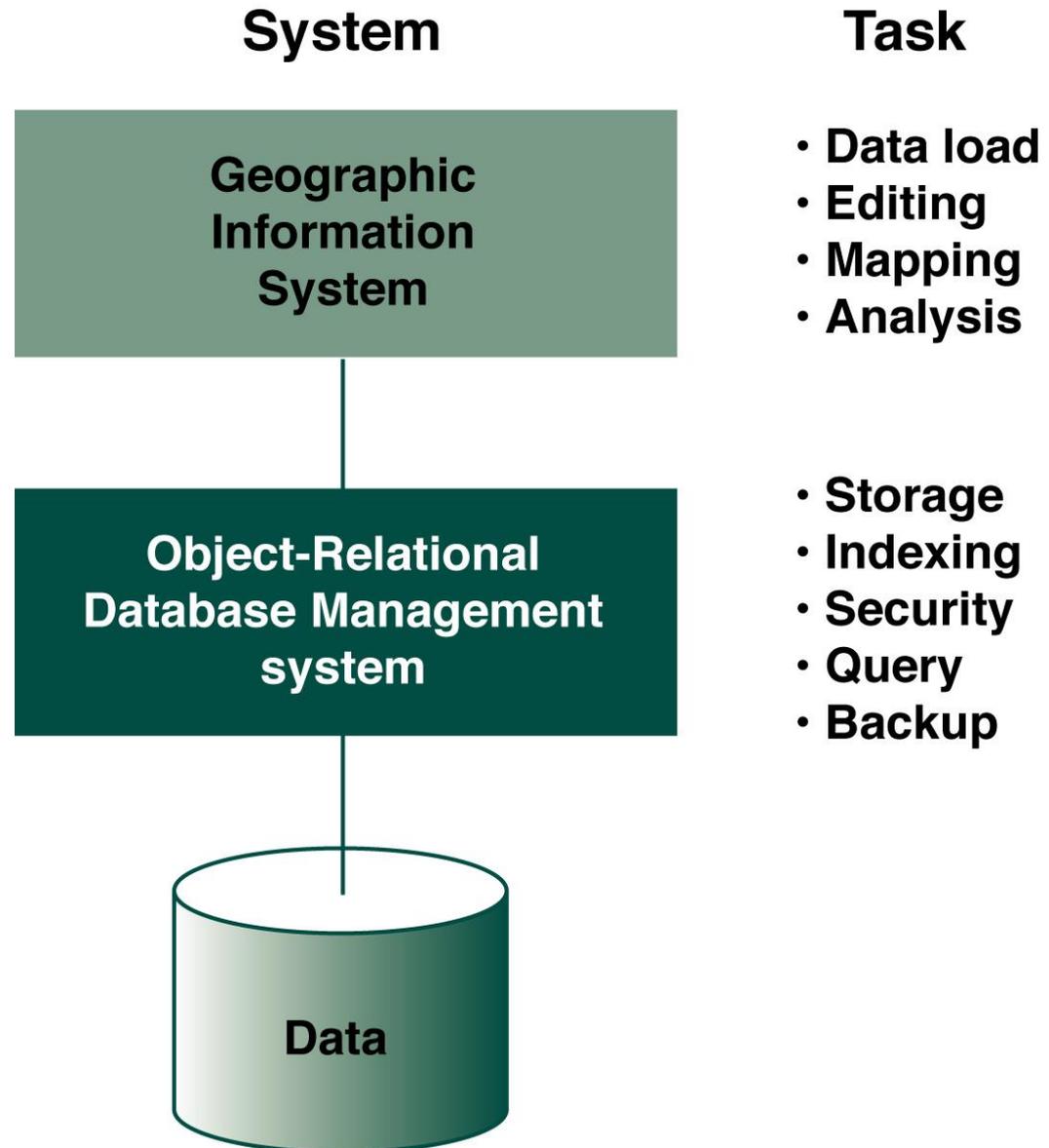


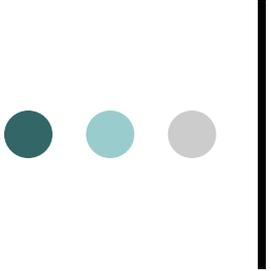
# GIS and Database Management System (DBMS)

1. Non spatial or attribute data managed by DBMS cannot be related to **locations** without GIS
2. GIS has embedded **DBMS**, or links to a DBMS for data management
3. GIS in combination with DBMS allows for **geographic search** which is key to GIS data retrieval



# Role of DBMS





# Definitions

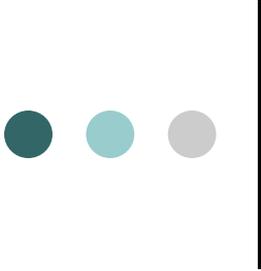
A **database** is a collection of files, an integrated set of data on a particular subject

A **file** is a collection of records

A **record** is a set of items for an object or event

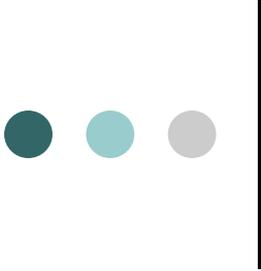
An **item** is a description of an object or an event

Items, records, and files are usually linked according to some well defined structure



# Geodatabase

- Represents geographic features and attributes as objects and is hosted inside a relational database management system
- Manages coverages, grids, and shapefiles *inside* a database management system, or DBMS



# Joining Tables

Tables can be linked by a **relational join**

A **primary key** of a table is linked to another key on another table

Joins can be:

- one-to-one
- one-to-many
- many-to-one
- many-to-many

# Keys

Primary Key

Foreign Key

Attributes of lu99				
	FID	Shape	ID	GRIDCODE
▶	0	Polygon	1	18
	1	Polygon	2	6
	2	Polygon	3	18
	3	Polygon	4	18
	4	Polygon	5	8
	5	Polynn	6	18

Primary Key

A1 = GRIDCODE			
	A	B	C
1	GRIDCODE	DEFINITION	
2	1	Cropland	
3	2	Pasture	
4	3	Forest	
5	4	Wetland	
6	5	Mining	
7	6	Open Land	
8	7	Recreation	
9	8	Spectator Recreation	
10	9	Water Based Recreation	
11	10	R0	
12	11	R1	
13	12	R2	
14	13	R3	
15	14	Salt Wetland	
16	15	Commercial	
17	16	Industrial	
18	17	Urban Open	
19	18	Transportation	
20	19	Waste Disposal	
21	20	Water	

# Types of Joins: Example 1

One to one

- Census tracts
- Income by household for each census tract

census\_tract\_2010\_esri

FID	Shape	ObjectID	STATE_FIPS	CNTY_FIPS	STCOFIPS	TRACT	FIPS
32418	Polygon	32418	25	027	25027	710400	25027710400
32419	Polygon	32419	25	027	25027	710500	25027710500
32420	Polygon	32420	25	027	25027	710600	25027710600
32421	Polygon	32421	25	027	25027	710700	25027710700
32422	Polygon	32422	25	027	25027	710800	25027710800

ACS\_11\_5YR\_S1701\_with\_ann.csv

GEOid	GEOid2	HC01_EST_VC01	HC01_MOE_VC01
1400000US25017300100	25017300100	2993	62
1400000US25017301101	25017301101	3874	205
1400000US25017301102	25017301102	4913	228
1400000US25017310100	25017310100	4212	388
1400000US25017310200	25017310200	6201	627

# Types of Joins: Example 2

One to Many

- One Town has
- Many homes sold

	FID	Shape *	OBJECTID	TOWNID	TOWN	POP1980	POP1990	POP
▶	0	Polygon	1	1	BRINGTON	13579	13817	
	1	Polygon	2	2	ACTON	17672	17872	
	2	Polygon	3	3	ACUSHNET	8808	9554	
	3	Polygon	4	4	ADAMS	10181	9445	

	FID	Shape	POLYID	AREA	PERIMETER	ID	TOWN	TOWNNO	TRAC
	3	Polygon	4	3.15512	12.6604	4	0	2	20
	4	Polygon	5	14.9663	26.161	5	0	2	20
	5	Polygon	6	28.262	26.6096	6	0	2	20
	6	Polygon	7	1.63288	4.94628	7	0	3	20
	7	Polygon	8	0.948014	3.94009	8	0	3	20

# Types of Joins: Example 3

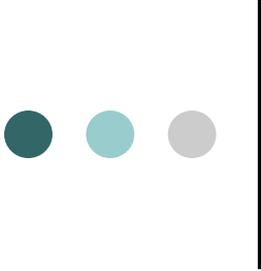
Many to Many

○ Parcels and Owners

Parcels		Parcel Owners	
Shape	PARCELID	PARCELID	OWNERID
Polygon	1-1-1	1-1-1	1
Polygon	1-1-2	1-1-1	2
Polygon	1-1-3	1-1-2	2
		1-1-2	3

**Owners**

OWNERID	NAME
1	Donald Trump
2	Bill Gates
3	Rupert Murdoch



# Spatial data and non spatial data joins

# Add Data Table in ArcMap

Table Of Contents

- Layers
  - D:\users\Sumeeta\courses\Harvard\Gov10
    - housing\_value

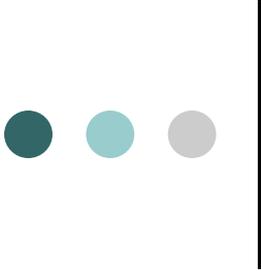
Table

housing\_value

OBJECTID *	LOGRECNO *	TOT_OWN_OC	VAL_LT10K	VAL10_15K	VAL15_20K	VAL20_25K	VAL25_30K	VAL30_35K	VAL35_40K	VAL40_50K	VAL50_60K
1	0000855	0	0	0	0	0	0	0	0	0	0
2	0000857	264	0	0	0	0	0	0	0	0	0
3	0000858	171	0	0	0	0	0	0	0	0	0
4	0000859	175	0	0	0	0	0	0	0	0	0
5	0000860	146	0	0	0	0	0	0	0	0	0
6	0000861	230	0	0	0	0	0	0	0	0	0
7	0000863	297	0	0	0	0	0	0	0	0	0
8	0000864	300	0	0	0	0	0	0	0	0	0
9	0000865	281	0	0	0	0	0	0	0	0	0
10	0000866	247	0	0	0	0	0	0	3	0	0
11	0000867	331	0	0	0	0	0	0	0	0	0
12	0000868	245	0	4	0	3	0	8	7	0	0
13	0000870	281	0	0	0	0	0	0	0	0	0
14	0000871	305	0	0	0	0	0	0	0	0	0
15	0000872	260	0	0	0	0	0	0	0	0	0
16	0000873	249	0	0	0	0	0	0	0	0	0
17	0000874	478	0	0	0	0	0	0	0	0	0
18	0000875	398	0	0	0	0	0	0	0	0	0
19	0000877	262	0	0	0	0	0	0	0	0	0
20	0000878	235	0	0	0	0	0	0	0	0	0
21	0000879	203	0	0	0	0	0	0	0	0	0

(0 out of 5053 Selected)

housing\_value



# Download Shapefile

## Census 2000: Census Tracts in ArcView Shapefile (.shp) format

Alabama - [tr01\\_d00\\_shp.zip](#) (878,228 bytes)  
Alaska - [tr02\\_d00\\_shp.zip](#) (538,034 bytes)  
Arizona - [tr04\\_d00\\_shp.zip](#) (609,863 bytes)  
Arkansas - [tr05\\_d00\\_shp.zip](#) (650,769 bytes)  
California - [tr06\\_d00\\_shp.zip](#) (2,303,684 bytes)  
Colorado - [tr08\\_d00\\_shp.zip](#) (593,672 bytes)  
Connecticut - [tr09\\_d00\\_shp.zip](#) (220,667 bytes)  
Delaware - [tr10\\_d00\\_shp.zip](#) (66,456 bytes)  
District of Columbia - [tr11\\_d00\\_shp.zip](#) (26,099 bytes)  
Florida - [tr12\\_d00\\_shp.zip](#) (967,180 bytes)  
Georgia - [tr13\\_d00\\_shp.zip](#) (1,079,095 bytes)  
Hawaii - [tr15\\_d00\\_shp.zip](#) (95,658 bytes)  
Idaho - [tr16\\_d00\\_shp.zip](#) (383,992 bytes)  
Illinois - [tr17\\_d00\\_shp.zip](#) (1,037,888 bytes)  
Indiana - [tr18\\_d00\\_shp.zip](#) (542,204 bytes)  
Iowa - [tr19\\_d00\\_shp.zip](#) (426,715 bytes)  
Kansas - [tr20\\_d00\\_shp.zip](#) (400,984 bytes)  
Kentucky - [tr21\\_d00\\_shp.zip](#) (959,474 bytes)  
Louisiana - [tr22\\_d00\\_shp.zip](#) (730,734 bytes)  
Maine - [tr23\\_d00\\_shp.zip](#) (179,306 bytes)  
Maryland - [tr24\\_d00\\_shp.zip](#) (447,335 bytes)  
Massachusetts - [tr25\\_d00\\_shp.zip](#) (399,790 bytes)  
Michigan - [tr26\\_d00\\_shp.zip](#) (947,308 bytes)  
Minnesota - [tr27\\_d00\\_shp.zip](#) (640,203 bytes)  
Mississippi - [tr28\\_d00\\_shp.zip](#) (482,941 bytes)  
Missouri - [tr29\\_d00\\_shp.zip](#) (1,013,417 bytes)  
Montana - [tr30\\_d00\\_shp.zip](#) (586,050 bytes)  
Nebraska - [tr31\\_d00\\_shp.zip](#) (254,423 bytes)  
Nevada - [tr32\\_d00\\_shp.zip](#) (205,679 bytes)

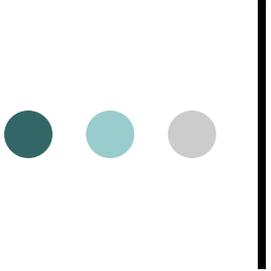
# Check Fields to Join

The image shows two attribute tables from ArcGIS. The top table, 'Attributes of Tract003', has columns: Shape, AREA, PERIMETER, TR42\_D00, TR42\_D00\_I, STATE, COUNTY, TRACT, and TRACT. The bottom table, 'Attributes of gct\_dec\_2000\_sf3\_u\_data0', has columns: GEOCOMP, GEO\_NAME, STATE, COUNTY, TRACT, and GCTP14\_C01. A red box highlights the TRACT field in the top table, with an arrow pointing to the TRACT field in the bottom table. The text 'Problem! Fields are not the same' is written in red inside the box.

Shape*	AREA	PERIMETER	TR42_D00	TR42_D00_I	STATE	COUNTY	TRACT	TRACT
Polygon			1335	42	003	0103	0103	
Polygon			1303	42	003	0201	0201	
Polygon			1234	42	003	0203	0203	
Polygon			1301	42	003	0305	0305	
Polygon			1310	42	003	0402	0402	
Polygon	0.00005	0.03509	1291	42	003	0403	0403	
Polygon	0.00005	0.02887	1295	42	003	0404	0404	
Polygon	0.00005	0.03237	1321	42	003	0405	0405	
Polygon	0.00003	0.02344	1327	42	003	0406	0406	
Polygon	0.00017	0.06193	1337	42	003	0409	0409	

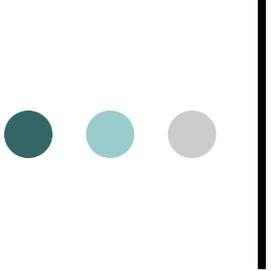
  

GEOCOMP	GEO_NAME	STATE	COUNTY	TRACT	GCTP14_C01
00	Allegheny County	42	003		38329
00	Tract 103	42	003	010300	11519
00	Tract 201	42	003	020100	20500
00	Tract 203	42	003	020300	41719
00	Tract 305	42	003	030500	13449
00	Tract 402	42	003	040200	23517
00	Tract 403	42	003	040300	90619
00	Tract 404	42	003	040400	17022
00	Tract 405	42	003	040500	14178
00	Tract 406	42	003	040600	15597



# SQL Query

```
SELECT tracts  
FROM census_tracts_us  
WHERE pop2010 > pop2000  
ORDER BY tracts_ID;
```



# Spatial SQL

**Query:** List the ID, population, and area of each tract listed in a table with data about census block groups (B)

```
SELECT B.ID,B.Pop, Area(B.Shape)AS "Area"  
FROM B
```

**Query:** Find the names of all tracts which are neighbors of the tract which includes 1730 Cambridge St in the table

```
SELECT B1.Name AS "Neighbors of 1730"  
FROM B1, B2  
WHERE Touch(B1.Shape,B2.Shape)=1  
AND B2.Name = '1730 cambridge st'
```

# Attribute Query Example

- Select tracts that have had population increase since 2000

Select By Attributes

Layer: census\_tract\_2010\_esri  
 Only show selectable layers in this list

Method: Create a new selection

"FIPS"  
"POP2000"  
"POP00\_SQMI"  
"POP2010"  
"POP10\_SQMI"  
"POP10\_SQMI"

= <> Like  
> >= And  
< <= Or  
\_ % ( ) Not  
Is

Get Unique Values Go To:

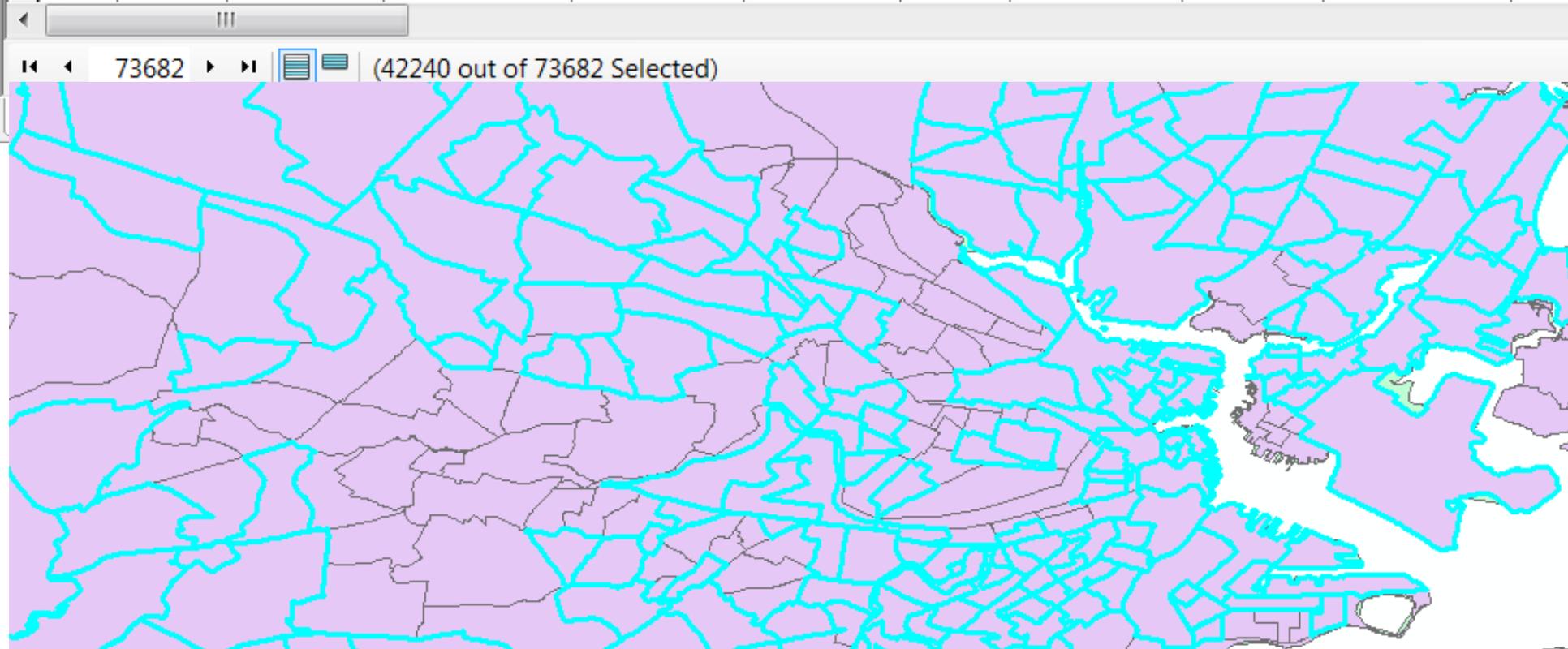
SELECT \* FROM census\_tract\_2010\_esri WHERE:  
"POP2010" > "POP2000"

Clear Verify Help Load... Save...  
OK Apply Close

# Attribute Query Example

census\_tract\_2010\_esri

FID	Shape	ObjectID	STATE_FIPS	CNTY_FIPS	STCOFIPS	TRACT	FIPS	POP2000	POP00_SQMI	PO
32418	Polygon	32418	25	027	25027	710400	25027710400	2219	3761	
32419	Polygon	32419	25	027	25027	710500	25027710500	3537	9307.9	
32420	Polygon	32420	25	027	25027	710600	25027710600	6603	5368.3	
32421	Polygon	32421	25	027	25027	710700	25027710700	1699	6796	
32422	Polygon	32422	25	027	25027	710800	25027710800	4506	4951.6	



● ● ●

# Select by location: select all tracts within 0.5 miles of Charles river

Select By Location

Select features from one or more target layers based on their location in relation to the features in the source layer.

Selection method:  
select features from

Target layer(s):

- charles
- census\_tract\_2010\_esri

Only show selectable layers in this list

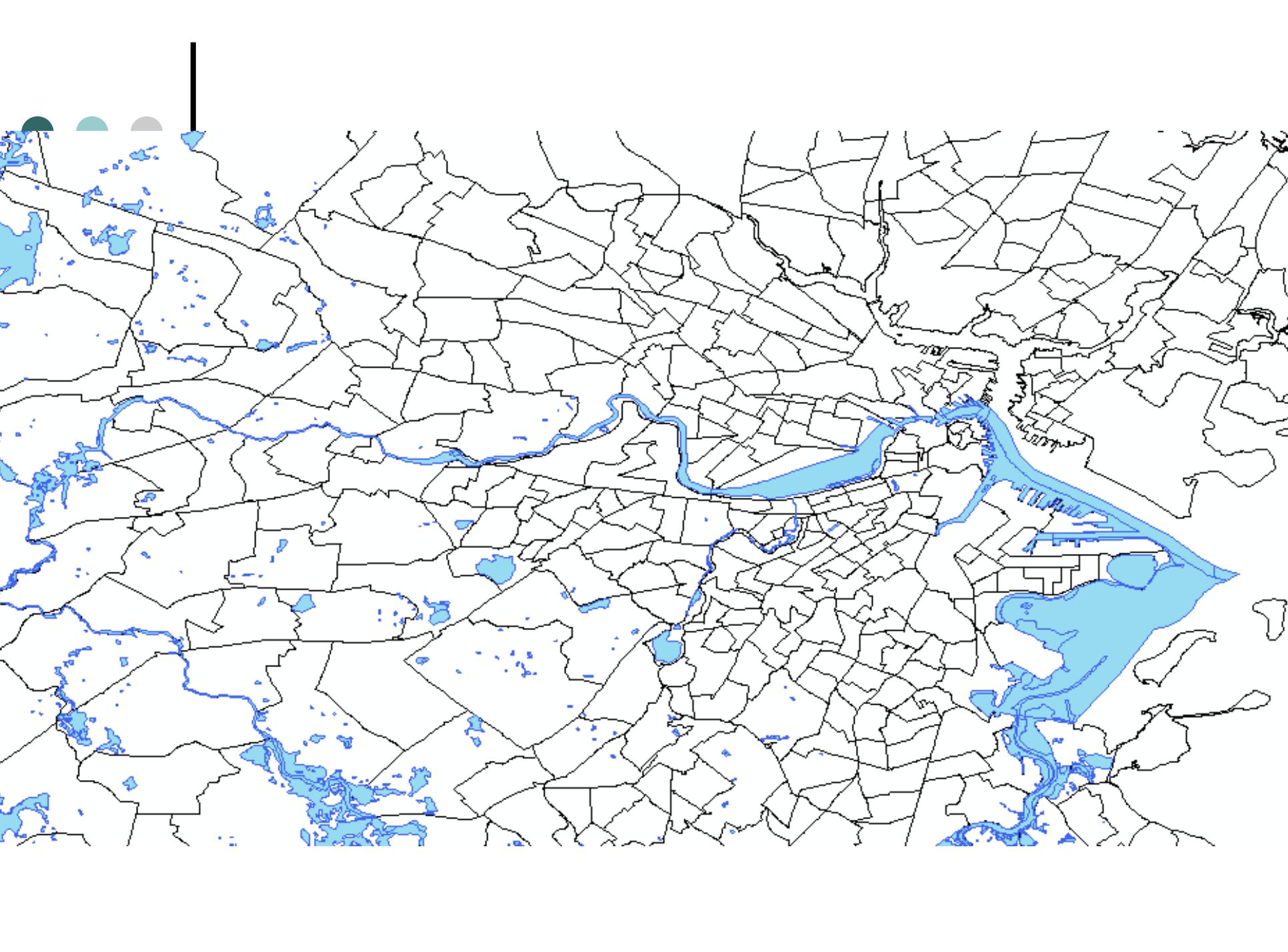
Source layer:  
charles

Use selected features (0 features selected)

Spatial selection method for target layer feature(s):  
are within a distance of the source layer feature

Apply a search distance  
0.5 Miles

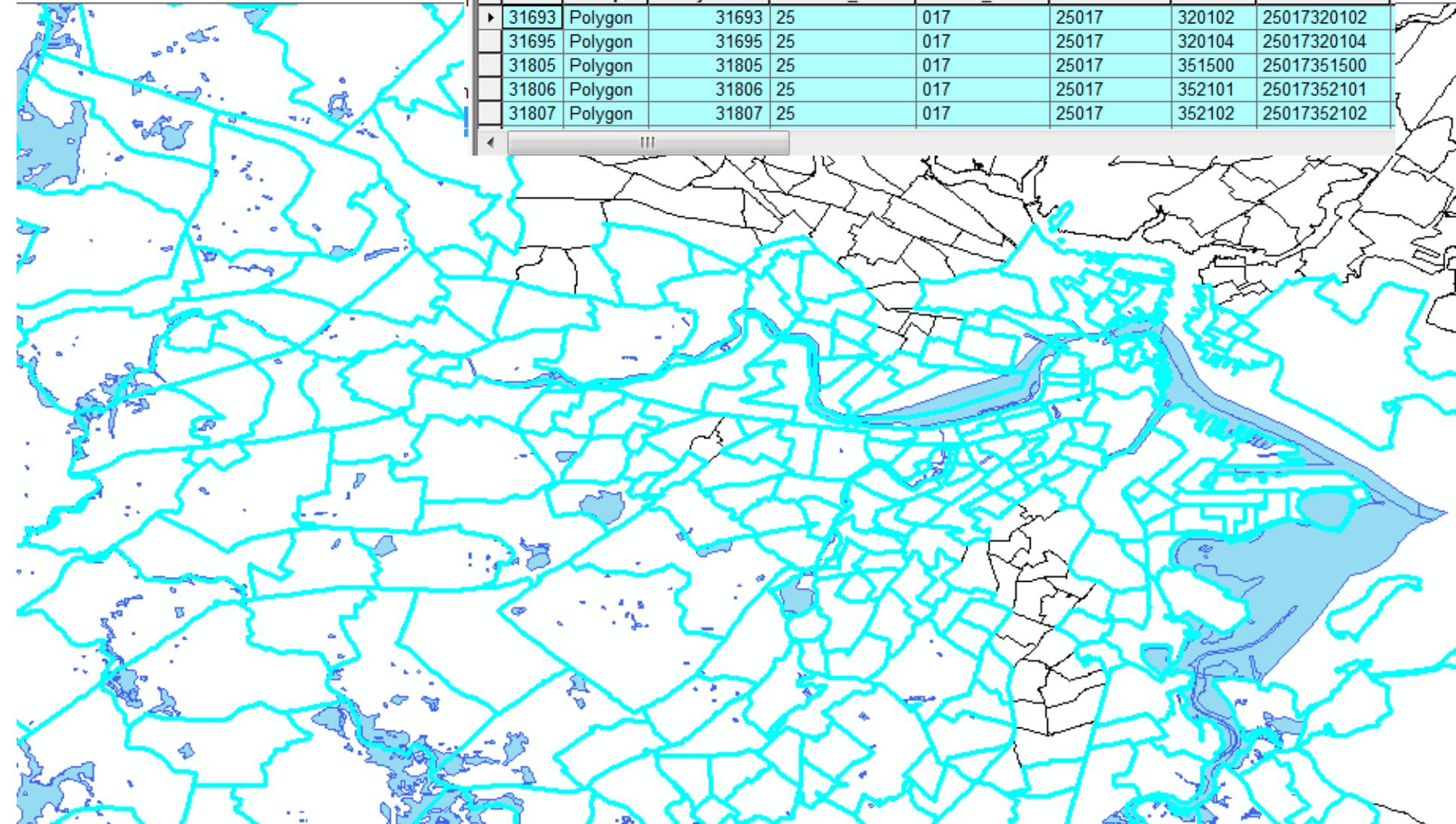
[About select by location](#) OK Apply Close



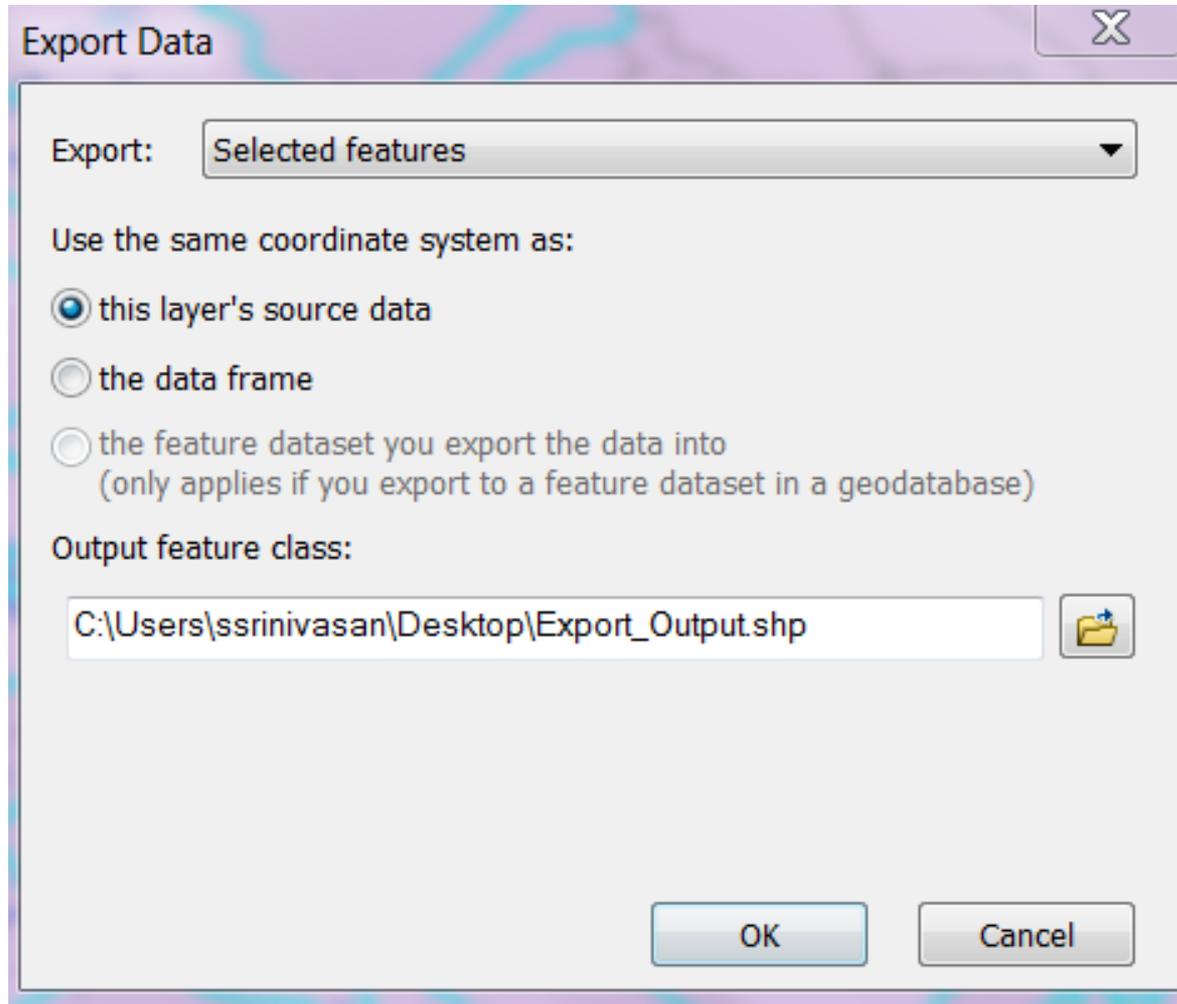
# Select by Location Query

census\_tract\_2010\_esri

	FID	Shape	ObjectID	STATE_FIPS	CNTY_FIPS	STCOFIPS	TRACT	FIPS
▶	31693	Polygon	31693	25	017	25017	320102	25017320102
	31695	Polygon	31695	25	017	25017	320104	25017320104
	31805	Polygon	31805	25	017	25017	351500	25017351500
	31806	Polygon	31806	25	017	25017	352101	25017352101
	31807	Polygon	31807	25	017	25017	352102	25017352102

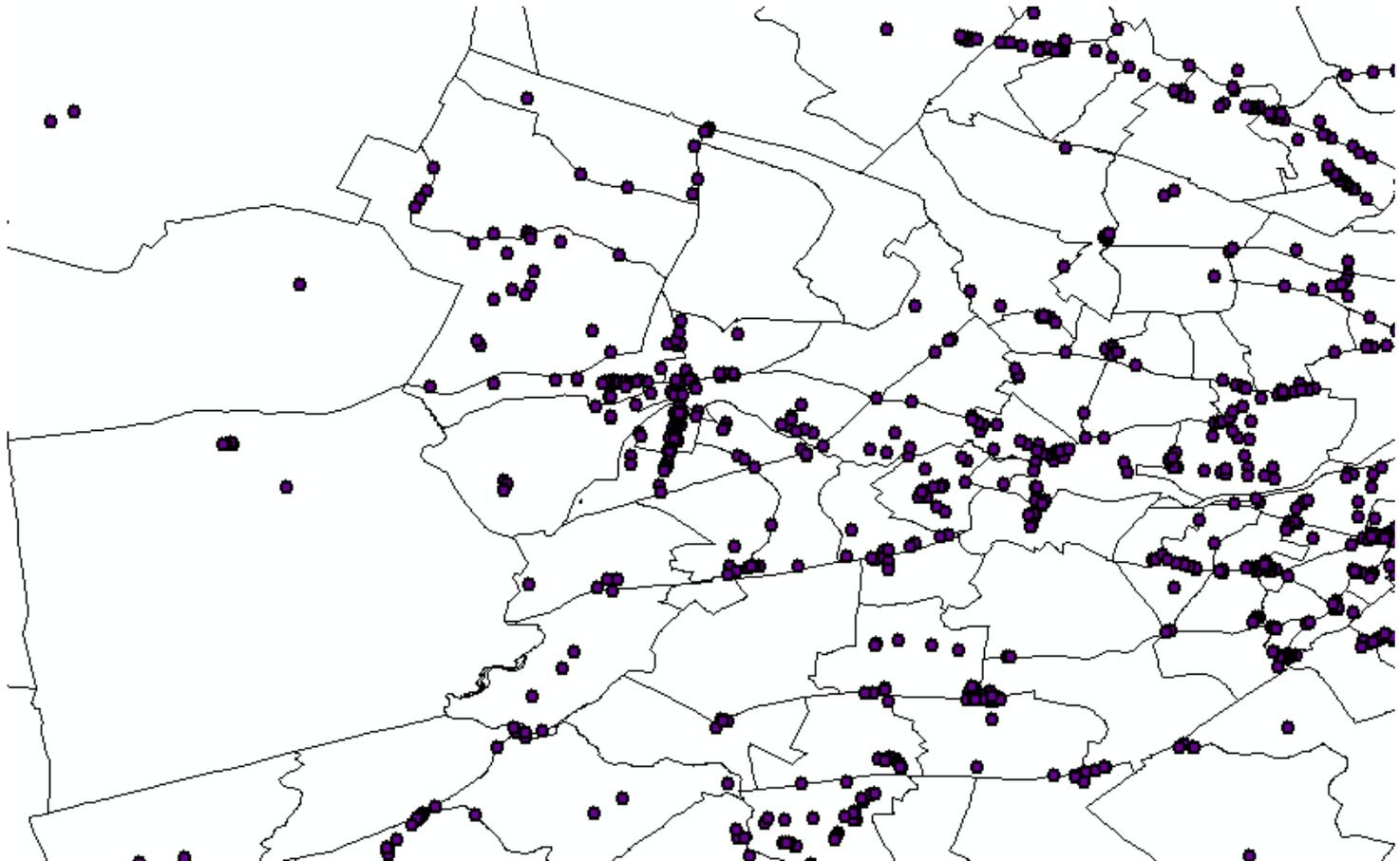


# Export Selection and Add Layer



# Spatial Joins

- Join points, lines, areas to polygons or vice versa



# Spatial Joins

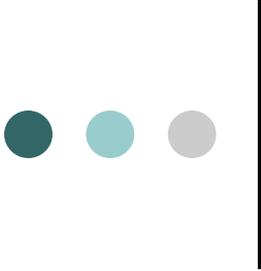
- No common attributes

census\_tract\_2010\_esri

	FID	Shape	ObjectID	STATE_FIPS	CNTY_FIPS	STCOFIPS	TRACT	FIPS	POP2000
▶	0	Polygon	0	01	001	01001	020100	01001020100	192
	1	Polygon	1	01	001	01001	020200	01001020200	189
	2	Polygon	2	01	001	01001	020300	01001020300	333
	3	Polygon	3	01	001	01001	020400	01001020400	455
	4	Polygon	4	01	001	01001	020500	01001020500	604

MA\_restaurant\_grocery

	FID	Shape *	OBJECTID	CONAME	CITY16	STATE	STAT
▶	0	Point	3688870	MOLLY'S AT THE MEADOWS	BRIMFIELD	MA	Massac
	1	Point	3688897	JENNY'S DINER	WARREN	MA	Massac
	2	Point	3688904	TOWN PIZZA	WARREN	MA	Massac
	3	Point	3688906	GIRLY'S GRILL	WARREN	MA	Massac
	4	Point	3688959	WARREN CASH MARKET	WARREN	MA	Massac
	5	Point	3688965	MAD DOG CAFE	WARREN	MA	Massac
	6	Point	3688967	TEBO'S MINI MARKET	WARREN	MA	Massac
	7	Point	3688980	DES'S PIZZA	WEST WARREN	MA	Massac
	8	Point	3688984	TRASKA VILLAGE MARKET	WEST WARREN	MA	Massac



# Spatial Joins: how many restaurants and grocery stores in each tract?

Join Data

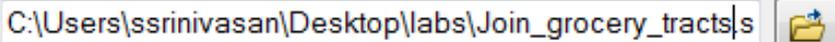
Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join data from another layer based on spatial location

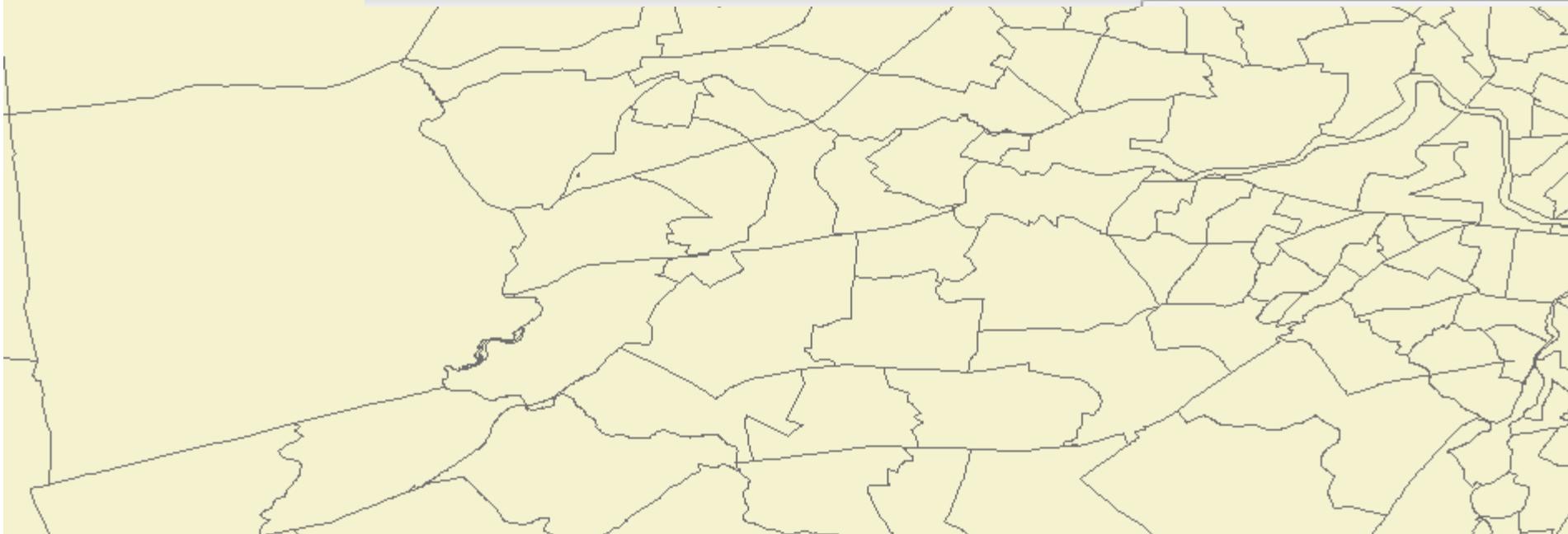
1. Choose the layer to join to this layer, or load spatial data from disk:  

2. You are joining: Points to Polygons  
Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.
  - Each polygon will be given a summary of the numeric attributes of points that fall inside it, and a count field showing how many points inside it.  
How do you want the attributes to be summarized?

<input type="checkbox"/> Average	<input type="checkbox"/> Minimum	<input type="checkbox"/> Standard Deviation
<input checked="" type="checkbox"/> Sum	<input type="checkbox"/> Maximum	<input type="checkbox"/> Variance
  - Each polygon will be given all the attributes of the point that is closest to its boundary, and a distance field showing how close the point is (in units of the target layer).  
  
Note: A point falling inside a polygon is treated as being closest to the polygon, (i.e. a distance of 0).
3. The result of the join will be saved into a new layer.  
Specify output shapefile or feature class for this new layer:  


# Spatial Joins

FIPS2	Count_	Sum_OBJECT	Sum_SALES_	Sum_NUMBER
25023523201	8	30158515	3760	54
25023523202	11	41473596	8832	143
25023524101	12	45190031	7592	662
25023524102	22	82888653	74542	424
25023525101	16	60263152	53370	387
25023525104	16	60266951	15868	154
25023525203	12	45198761	31294	208
25023525204	1	3766650	836	0
25023525200	0	0	0	0



# US Census

U.S. Census Bureau

AMERICAN FactFinder

Feedback FAQs Glossary Help

English Español

MAIN SEARCH WHAT WE PROVIDE USING FACTFINDER

Your Selections  
Your Selections is empty

Search using the options below:

Topics (age, income, year, dataset, ...)

Geographies (states, counties, places, ...)

Race and Ethnic Groups (race, ancestry, tribe)

Industry Codes (NAICS industry, ...)

EEO Occupation Codes (executives, analysts, ...)

Quick Facts  
Quick, easy access to facts about people, business, and geography.  
Select the U.S. or a state to begin

Your source for population, housing, economic, and geographic information

Updates to American FactFinder are coming in January 2013!  
Watch a [Virtual Tour](#) of the new American FactFinder.

Quick Start

Enter search term(s) and click 'GO' ?

topic or table name for state, county or place (optional)

GO

topics race/ancestry industries occupations

Or use the options on the left to begin your search

News and Notes [GET EMAIL UPDATES](#)

Dec 13, 2012

The Census Bureau has released the 2010 American Indian and Alaska Native Summary File (AIANSF). This file contains population and housing characteristics iterated for the total population, the total American Indian and Alaska Native population alone, the total American Indian and Alaska Native population alone or in combination, and 1,567 detailed tribes and tribal groupings. The 2010 AIANSF presents data for the nation, the 50 states, the District of Columbia, and Puerto Rico. [more »](#)

Address Search  
Find Census data by entering a [street address](#).

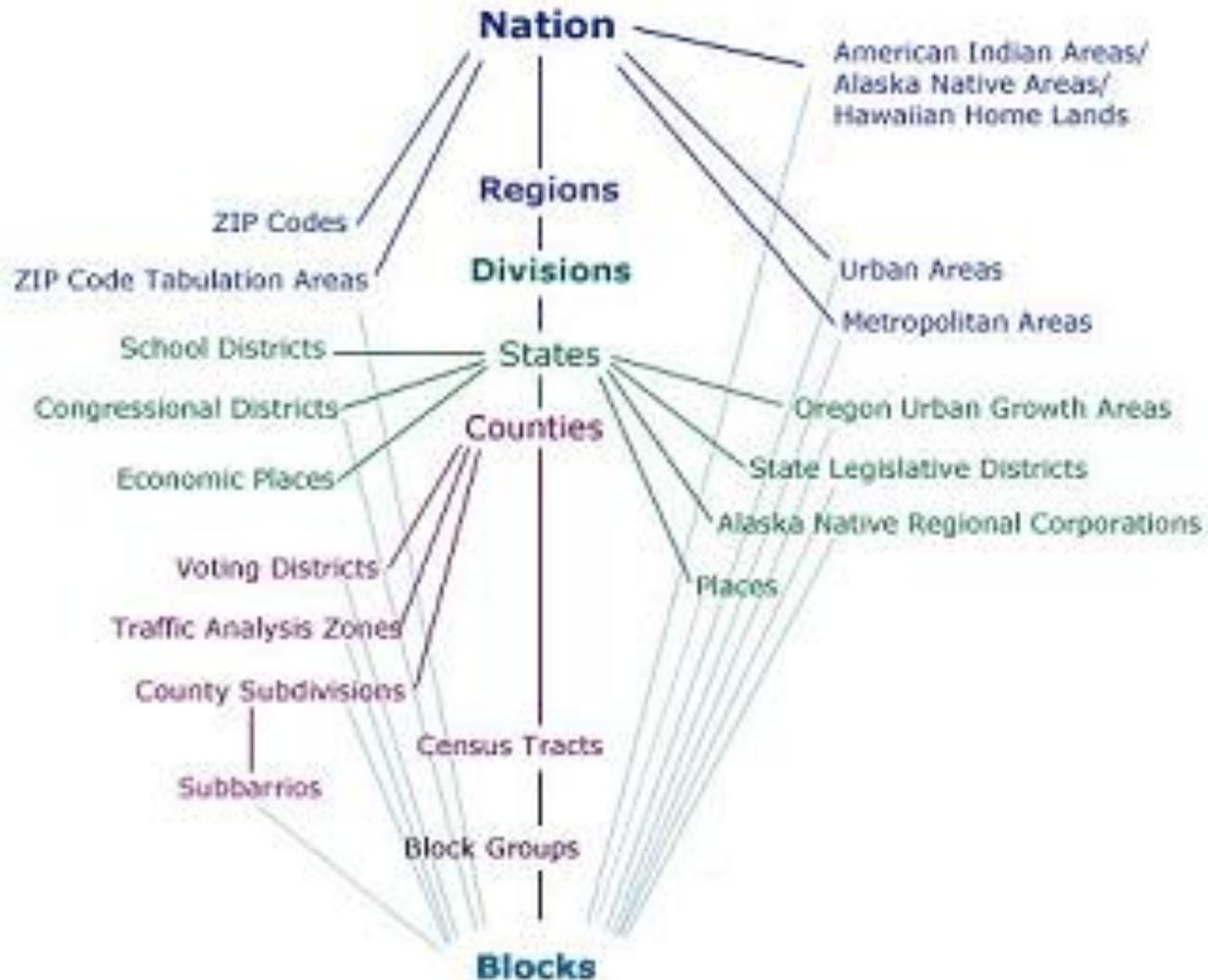
View the American FactFinder [Data Release Schedule](#)

Download large volume data from the [Census Bureau's FTP Site](#)

A how-to guide for [Building Deep Links into American Factfinder](#) (PDF) is available

Reference Maps  
Reference Maps show selected geographic boundaries for an area along with orienting features, such as roads.

# Census Geographies



# US Census historic data

## NHGIS National Historical Geographic Information System

[Home](#) | [Select Data](#) | [FAQ](#) | [Login](#)

### Data

[Select Data](#)  
[Download/Revise Extracts](#)

### User Resources

[Project Description](#)  
[User's Guide](#)  
[FAQ](#)  
[Data Availability](#)  
[Mapping Options](#)

### Data Documentation

[Tabular Data Documentation](#)  
[GIS Data Documentation](#)  
[Release Log](#)

### Research

[Citation and Use](#)  
[Related Sites](#)  
[MPC Bibliography](#)

### Contact Us

[Technical Assistance](#)  
[NHGIS Staff](#)

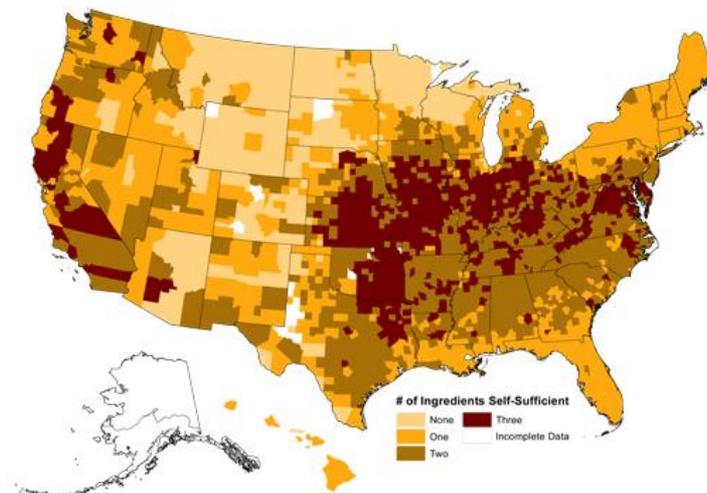
**Select  
Data**  
Tables & GIS Files

## Welcome to NHGIS

The National Historical Geographic Information System (NHGIS) provides, free of charge, aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2010.

### Thanksgiving Meal Self-Sufficiency Index - 1900 Census

Data on turkey, sweet potato and cranberry production at county level from NHGIS.



\*Approximate ability of county to produce the necessary turkey, sweet potatoes and cranberries needed for 1 Thanksgiving meal for each county resident.

### NHGIS News

[NHGIS announces conference exhibitor schedule!](#)

[NHGIS has a new website!](#)

[<read more>](#)

[IPUMS-International](#)

[IPUMS-USA](#)

[IPUMS-CPS](#)

[NAPP](#)

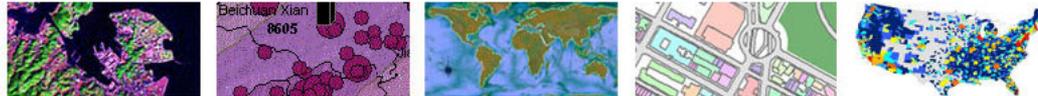
[IHIS](#)

[ATUS-X](#)



# Other resources

<http://gis.harvard.edu/icb/icb.do?keyword=k235&pageid=icb.page190080>



- Home
- GIS Help & Services
- Learn GIS
- Software
- Data Resources**
- Harvard Data Resources
- Esri Data
- Public Data Resources
- Conferences & Events
- Campus Resources
- Awards & Funding
- Employment
- News & Blogs
- Site Map / Search
- UCGIS Portfolio

## Harvard Data Resources

### CGA's Dynamic Map Service Layers



Maps of U.S. Counties, Metropolitan Statistical Areas, and Congressional Districts are published by the CGA as consumable map layers.

Rendering polygon labels dynamically on a web map can be achieved by using an ArcMap .mxd document and ArcGIS Server 10 software. The .mxd is set up to display the symbology and labels, and is published as an Esri Dynamic Map Service Layer using ArcGIS Server 10.

The 3 maps linked to below will render polygon labels consistently at all zoom scales in 8 pt. Arial bold font with a 1 point white halo mask. Labels will not overlap polygon boundaries, and one label per polygon will be rendered. The map layers are available for anyone to add to their web map application (just view and copy the well-documented page source).

[U.S. County Boundaries \(2010\)](#)

[U.S. Congressional District Boundaries \(2010\)](#)

[U.S. Metropolitan Statistical Area Boundaries](#)

### Where to Find Census Data



### GIS Data @ Harvard



#### Harvard Geospatial Library

The [Harvard Geospatial Library](#) is the University's catalog and repository for geospatial data. It houses thousands of layers of digital geospatial data, in both vector and raster (scanned maps) forms. HGL uses traditional text searching combined with map/coordinate based searches. Data can be viewed on-line, or downloaded for use in a desktop GIS. [See a powerpoint presentation explaining HGL and its capabilities.](#)

#### The Harvard Map Collection

The [Harvard Map Collection](#) maintains a large collection of geospatial data sets for use in Geographic Information Systems (GIS), Cartography, and Remote Sensing. The Map Collection's geospatial holdings include U.S. Census Bureau, Boston metropolitan area, and many worldwide and foreign data sets. These data can also be found in [HOLLIS](#).

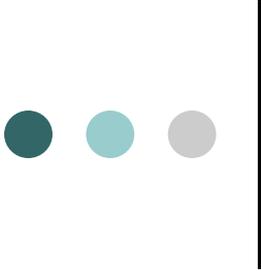
#### Graduate School of Design LAN

The GSD Frances Loeb Library and the GIS Specialist have gathered together a large amount of [GIS data stored on the GSD Local Area Network](#) for student access.

#### Public GIS Data Resource Listing

CGA maintains a list of [organizations that provide geographic data](#), including government, educational or non-profit organizations that provide free data.

Collect your own data with CGA's [GPS loaner systems](#).



# At the end of this lab you should be able to

- Join a table to a shapefile using a primary key
- Do a spatial join
- Query
  - By attribute
  - By location
- Calculate statistics for an attribute
- Save a selection as a new spatial dataset