Spatial Organization of Data and Data Extraction from Maptitude

N. P. Taliceo

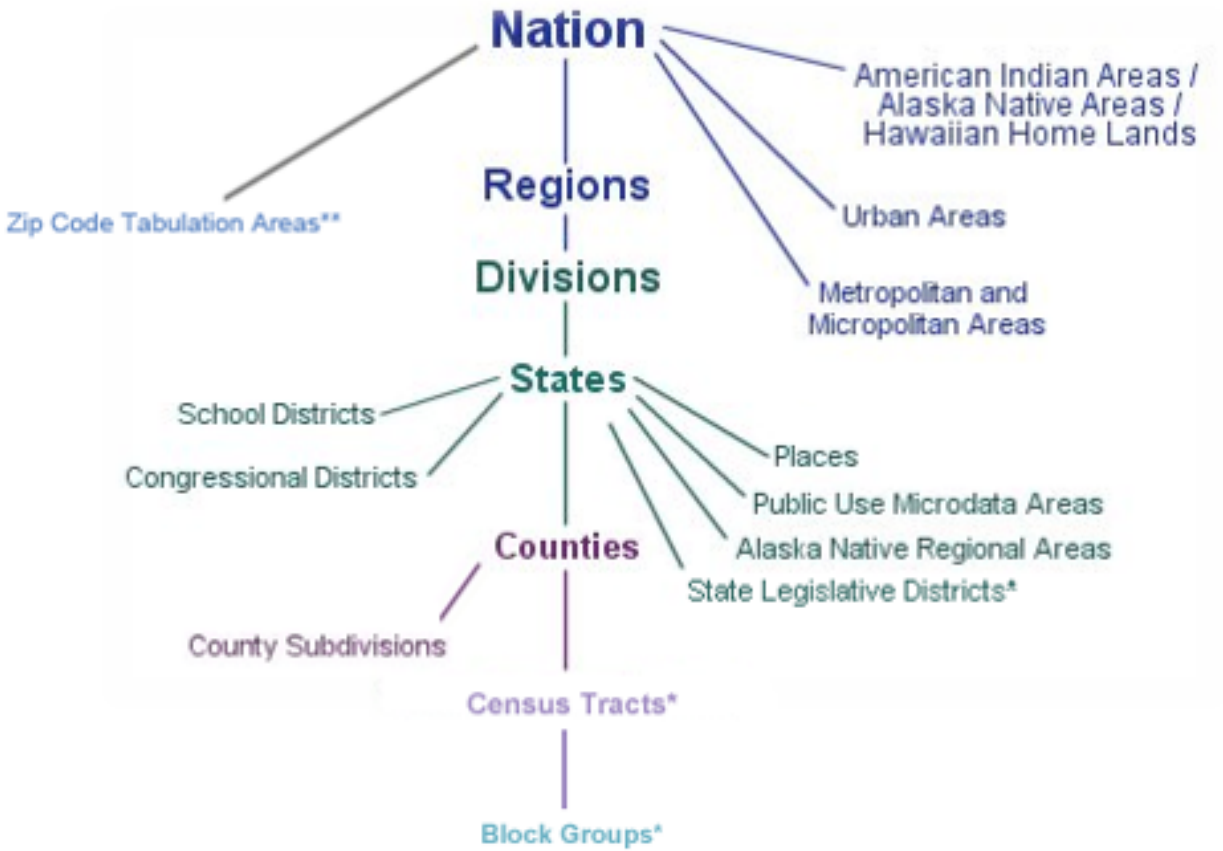
Geospatial Information Sciences
The University of Texas at Dallas

UT Dallas GIS Workshop
Richardson, TX | March 30–31, 2018

Outline

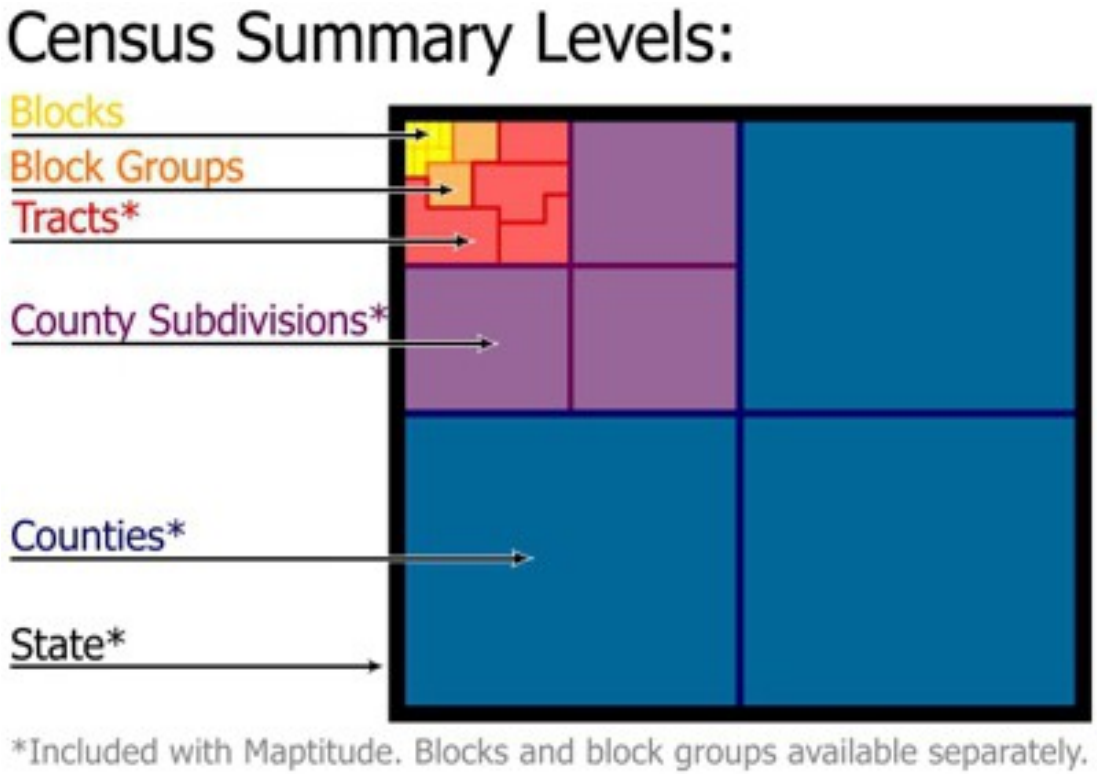
Outline

Census Summary Levels – U.S. Census Bureau



- American FactFinder
- TIGER/Line

Census Summary Levels – Maptitude



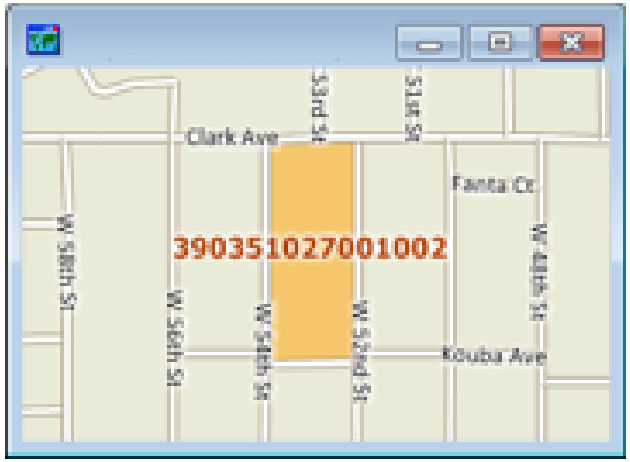
- About Census Summary Levels
- Maptitude has 2010, 2000, and 1990 Census data

Outline

Blocks & Block Groups

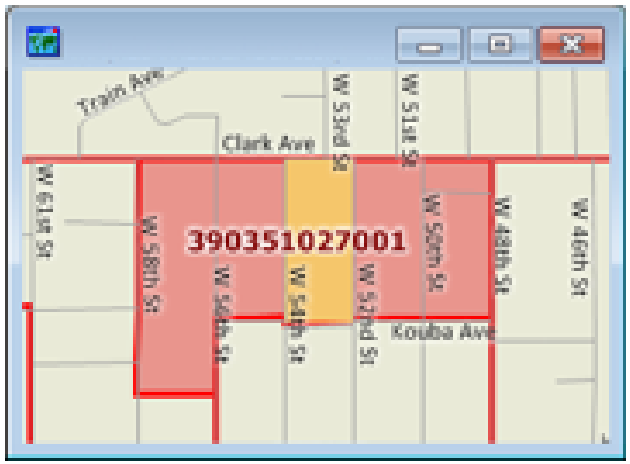
Blocks

- The smallest geographic unit from the U.S. Census
- About 11 million census blocks in the U.S. in 2010
- Typically equivalent to the city block in size (except for in rural areas)



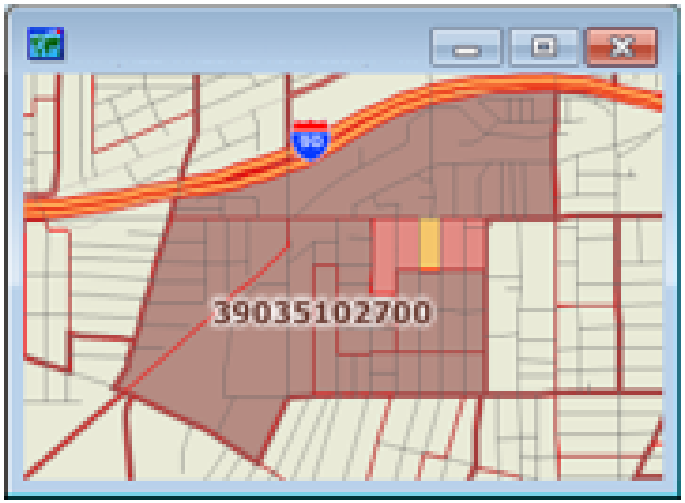
Block Groups

- A combination of census blocks; a subdivision of census tracts
- Contain 600 to 3,000 people
- Made up of 40 census blocks
- About 220,000 block groups nationwide in 2010



Census Tracts

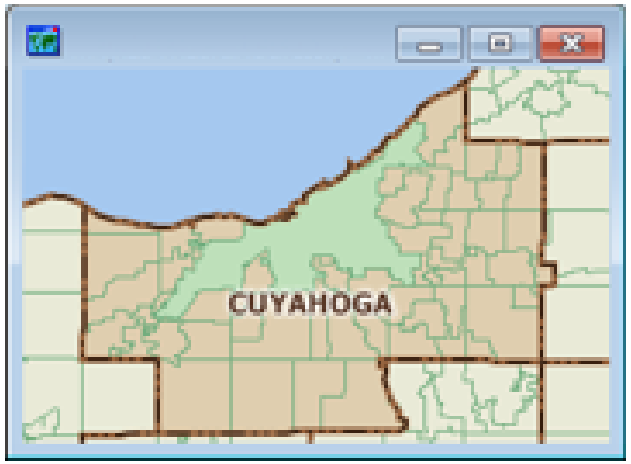
- A small subdivision of a county
- The purpose is to present census data
- Designed to have relatively homogeneous units:
  - Population characteristics
  - Economic status
  - Living conditions
- 1,500 to 8,000 people; about 4 block groups
- Maptitude tract mapping software



Counties & County Subdivisions

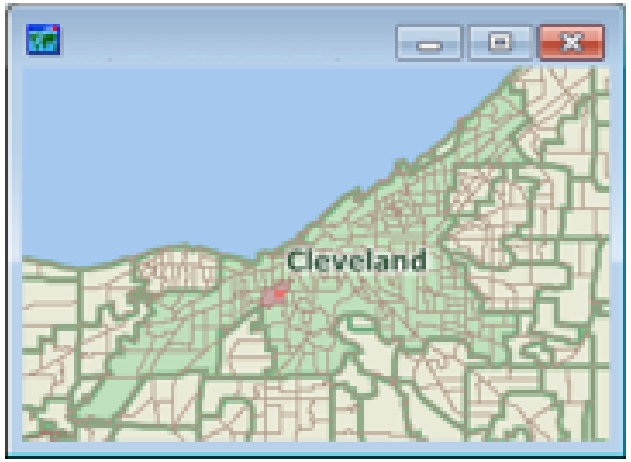
Counties

- The principal legal division in *most* states
- In Louisiana, called parishes
- There are no counties in Alaska (census areas)
- In Puerto Rico, municipios



County Subdivisions

- Subdivisions from the decennial census
- Includes civil divisions, census county divisions, census subareas, and unorganized territories

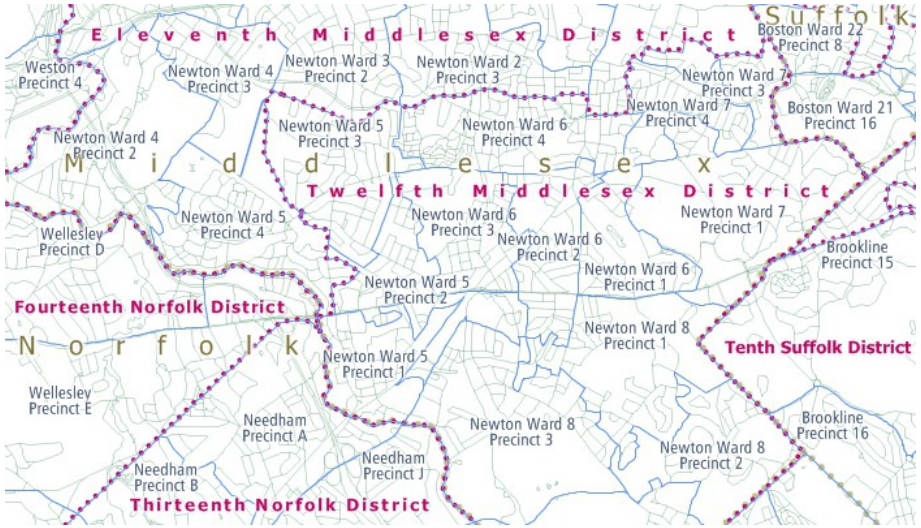


Larger Area Data

1. **States**
   - The primary governmental divisions of the United States
2. **Census Places**
   - Census designated places, consolidated cities, and incorporated places
   - A settled concentration of population identifiable by name but are not legally incorporated
   - Exceptions: towns in New England, New York, and Wisconsin and the boroughs in New York (minor civil divisions), among others
3. **Metropolitan Statistical Areas (MSAs) & Micropolitan Statistical Areas**
   - Geographically high population density in geographic regions with close economic ties
   - Micropolitan statistical areas have a population of 10,000 to 49,999

Other Available Summary Level Data

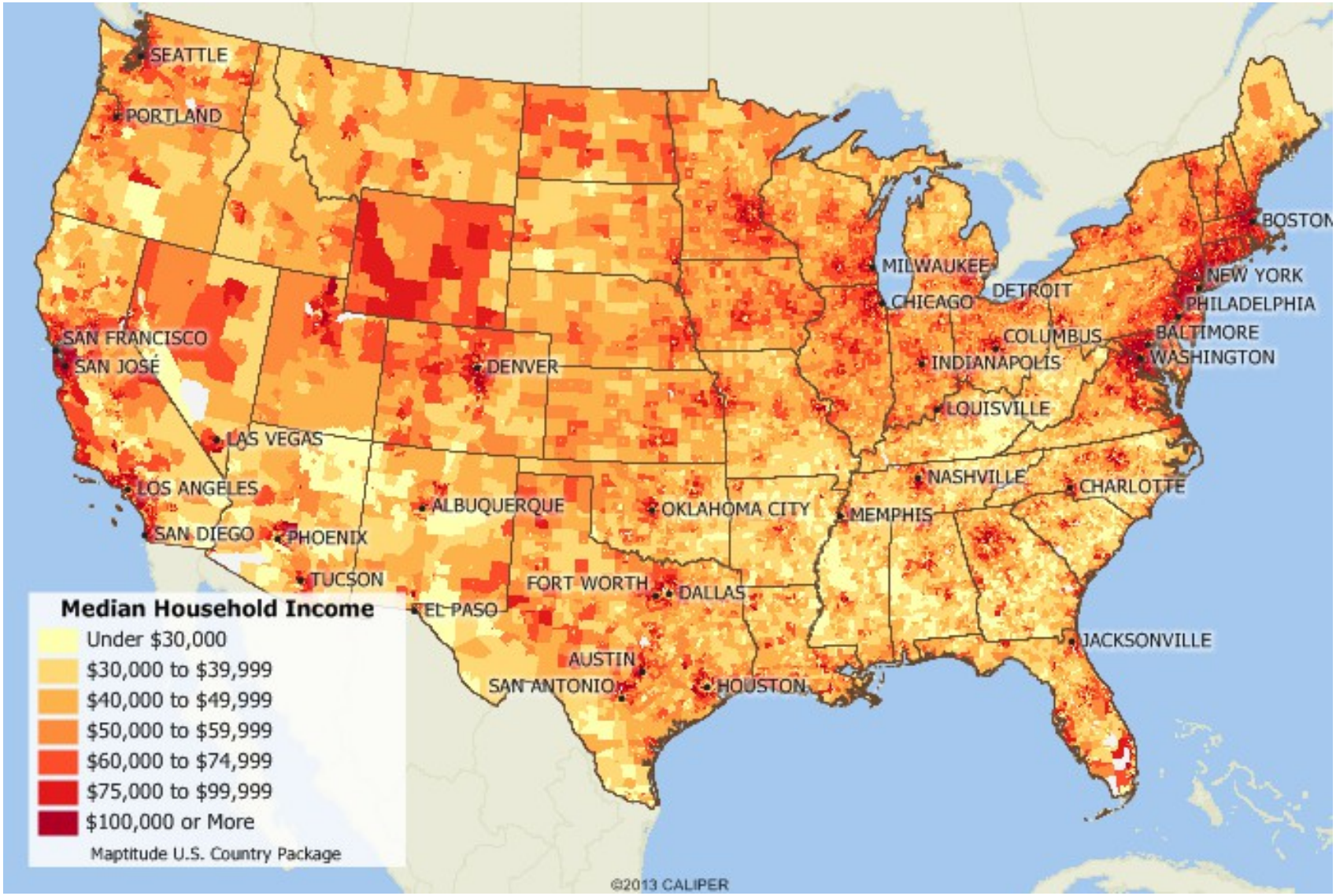
- Congressional Districts
- State Legislative Districts
- Voting Districts
- School Districts
- Urbanized Areas & Urban Clusters
- Traffic Analysis Zones (TAZs)

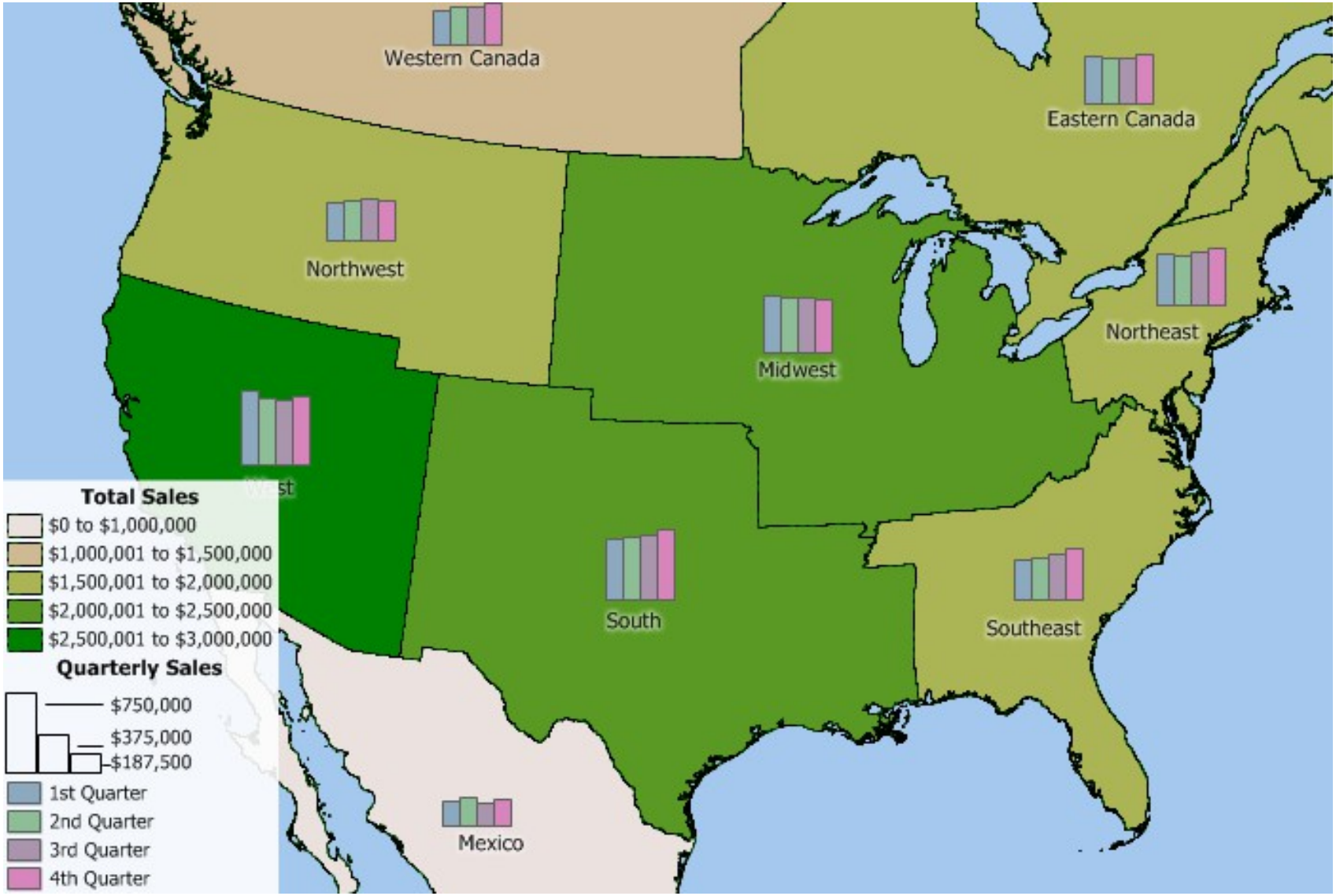


Outline

Features of Maptitude with the U.S. Country Package

- Nationwide maps with Census & ACS demographic data
  - 5-Digit ZIP Code & ZIP Code Business Pattern Data
- Demographic data: income, age, gender, households, housing units, ethnicity, ancestry, employment, buying power, population projections, health insurance coverage, etc.
- Integration with Google Earth
- Supports most data in a variety of formats (e.g., spreadsheets, databases, etc.)


SEATTLE
PORTLAND
SAN FRANCISCO
SAN JOSE
LAS VEGAS
LOS ANGELES
SAN DIEGO
PHOENIX
TUCSON
EL PASO
ALBUQUERQUE
DENVER
OKLAHOMA CITY
FORT WORTH
DALLAS
AUSTIN
SAN ANTONIO
HOUSTON
MILWAUKEE
CHICAGO
DETROIT
COLUMBUS
INDIANAPOLIS
LOUISVILLE
NASHVILLE
MEMPHIS
CHARLOTTE
JACKSONVILLE
BOSTON
NEW YORK
PHILADELPHIA
BALTIMORE
WASHINGTON
Median Household Income
Under $30,000
$30,000 to $39,999
$40,000 to $49,999
$50,000 to $59,999
$60,000 to $74,999
$75,000 to $99,999
$100,000 or More
Maptitude U.S. Country Package
©2013 CALIPER


Western Canada
Eastern Canada
Northwest
Midwest
Northeast
South
Southeast
Mexico
Total Sales
$0 to $1,000,000
$1,000,001 to $1,500,000
$1,500,001 to $2,000,000
$2,000,001 to $2,500,000
$2,500,001 to $3,000,000
Quarterly Sales
$750,000
$375,000
$187,500
1st Quarter
2nd Quarter
3rd Quarter
4th Quarter

Outline

- **Nationwide Boundary Files**
  - Census places, census tracts, CBSAs, etc.
- **Nationwide Geographic Data Layers**
  - Streets with address data for geocoding and travel time data
  - Highways (interstate, U.S., state, and county; ferry routes)
  - Building footprints (for major urban areas)
  - Area and point landmarks (e.g., commercial buildings, shops, etc.)
  - Railroads
- **Miscellaneous U.S. Data Layers**
  - Area codes, combined statistical areas, major cities and state capitals, national parks, populated places, rivers, time zones, water areas, state plane coordinate systems

Outline

Finding Available Data Fields

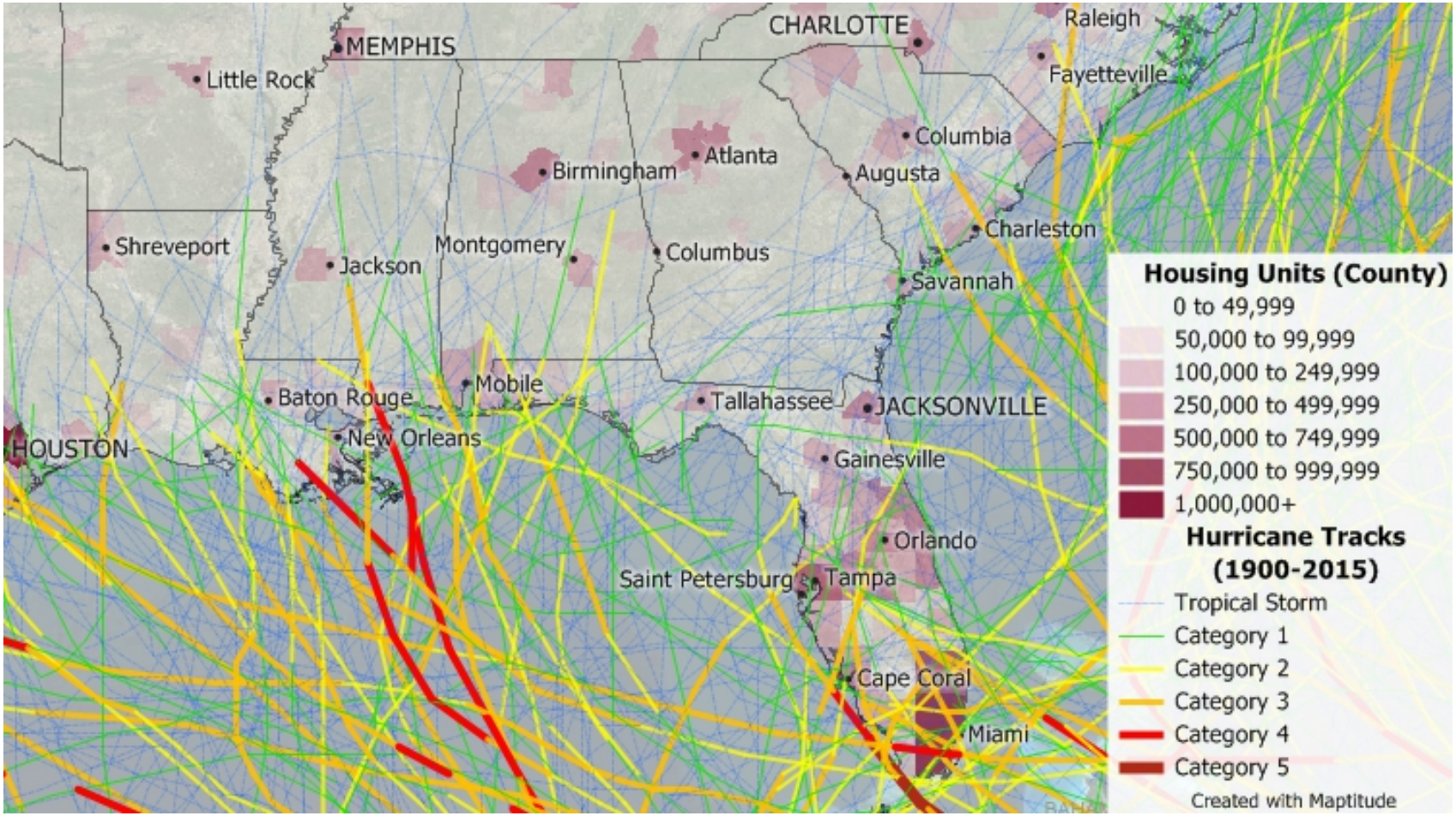
- Demographic data is included in the U.S. Country package
- By default, includes 3- and 5-digit ZIP Codes and other demographic profiles
- Blocks and Block Groups are available separately for purchase

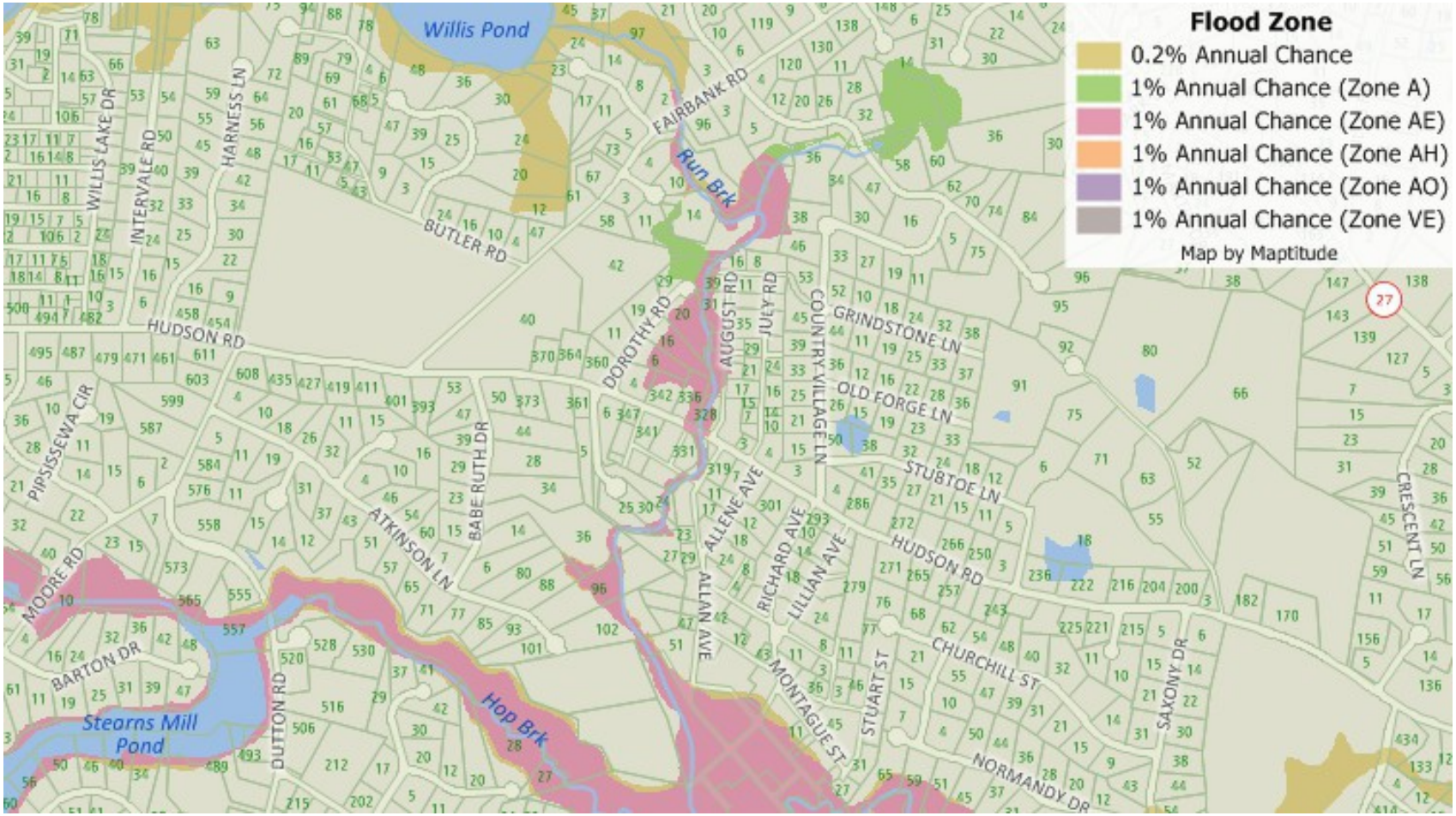
`https://www.caliper.com/maptitude/international/unitedstates.htm`

Outline

Maptitude Capabilities in Insurance

- Assess risk exposure (overexposed & underexposed areas)
- Locate customers and resources (map policy holders, risk exposure, etc.)
- Analyze insurance underwriting (flood plains, proximity to fire stations, rental properties vs. home ownership, etc.)
- Find marketing opportunities
- *Using your own data & models*
- Provide customers with a web-map app to certain information


MEMPHIS
CHARLOTTE
Raleigh
Little Rock
Fayetteville
Columbia
Atlanta
Birmingham
Augusta
Charleston
Shreveport
Jackson
Montgomery
Columbus
Savannah
Mobile
Baton Rouge
Tallahassee
JACKSONVILLE
HOUSTON
New Orleans
Gainesville
Orlando
Saint Petersburg
Tampa
Cape Coral
Miami
Housing Units (County)
0 to 49,999
50,000 to 99,999
100,000 to 249,999
250,000 to 499,999
500,000 to 749,999
750,000 to 999,999
1,000,000+
Hurricane Tracks (1900-2015)
Tropical Storm
Category 1
Category 2
Category 3
Category 4
Category 5
Created with Maptitude


Flood Zone
0.2% Annual Chance
1% Annual Chance (Zone A)
1% Annual Chance (Zone AE)
1% Annual Chance (Zone AH)
1% Annual Chance (Zone AO)
1% Annual Chance (Zone VE)
Map by Maptitude
Willis Pond
Run Brk
Hop Brk
Stearns Mill Pond
FAIRBANK RD
BUTLER RD
HUDSON RD
WILLIS LAKE DR
INTERVALE RD
HARNESS LN
DOROTHY RD
AUGUST RD
JULY RD
COUNTRY VILLAGE LN
GRINDSTONE LN
OLD FORGE LN
STUBTOE LN
CRESCENT LN
PIPSISSEWA CIR
MOORE RD
BARTON DR
DUTTON RD
ATKINSON LN
BABE RUTH DR
ALLENE AVE
ALLAN AVE
RICHARD AVE
LILLIAN AVE
MONTAGUE ST
STUART ST
CHURCHILL ST
SAXONY DR
NORMANDY DR

Outline

Task – Data Extraction Using Maptitude

Mise en scène

Image that you work for an insurance company and need to collect data to include in your *premium pricing model*.

- What data do you need to collect?
- What level of geographic detail do you need to have?

- We will use Maptitude to collect necessary data at either the ZIP Code level or census tract level.
- These data can then be used in separate software (e.g., MS Excel, `R`, `SAS`, etc.).

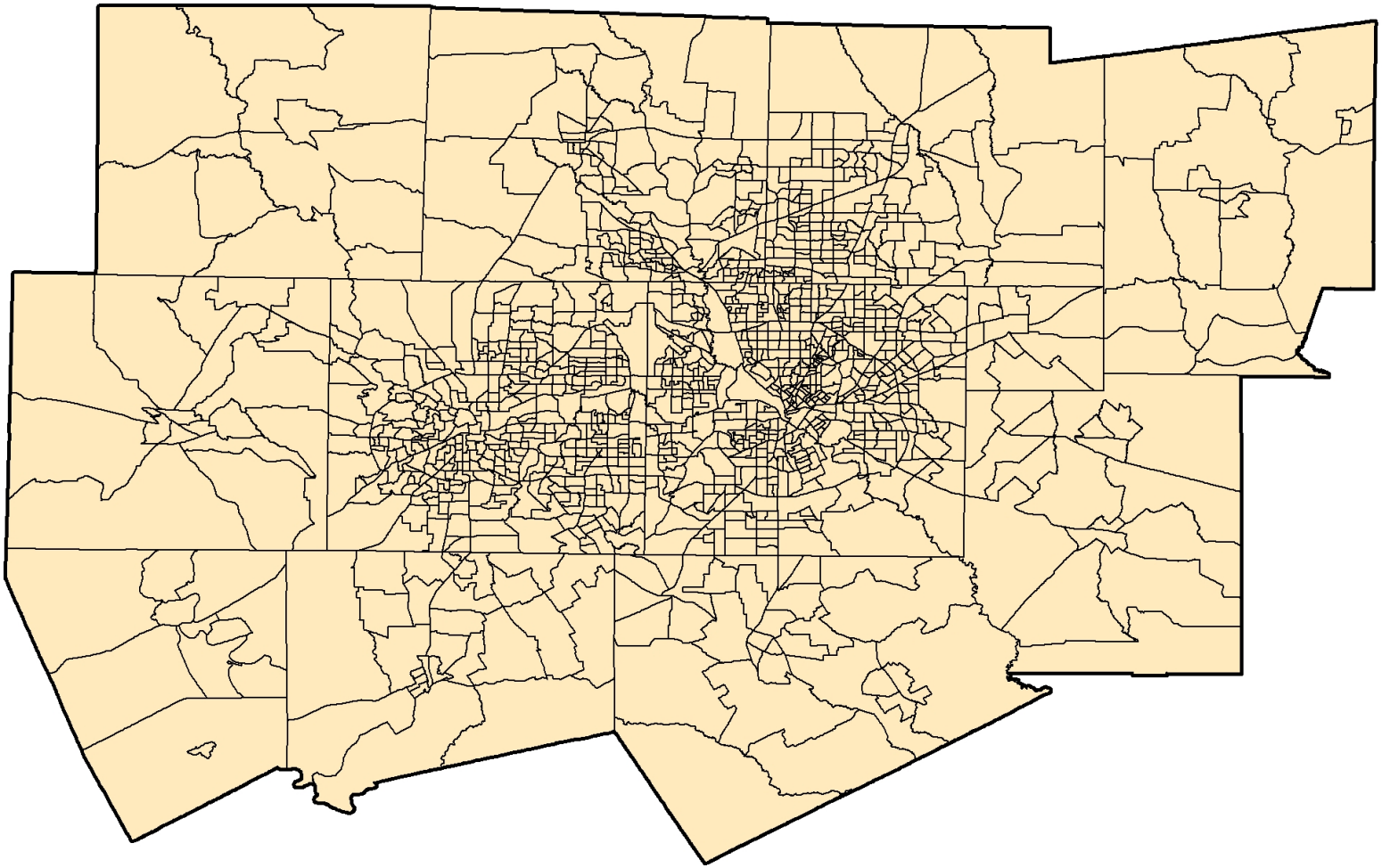
A Caveat – Reality to this Task

- Actual insurance companies pay large sums of money for "premium" data that contained detailed demographic and economic information.
- In addition, large amounts of private insurance related data are kept at the company-level. *Therefore, realistic estimates of premium pricing are nearly impossible.*
- Real models estimating base rate pricing are incredibly complicated, and are of course, proprietary.

Potential Variables

- Age distribution
- Automobile theft
- Credit score (not in Maptitude)
- Median income
- Medium home value
- Occupation
- Traffic volume estimates
- Travel demand – average commute time

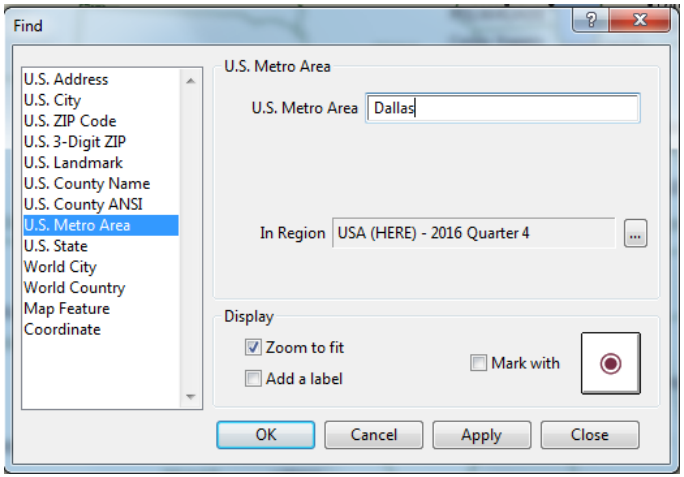
The Study Area – DFW

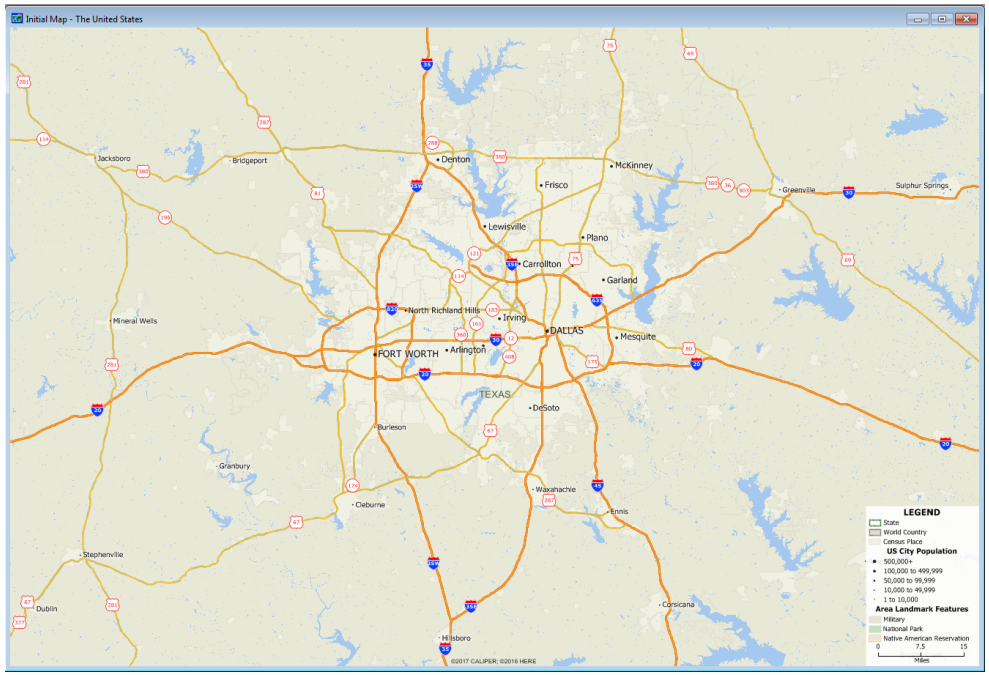


Outline

Step 1a – Location Selection: DFW

1 Open a new document by "Display the Initial Map". Zoom to the DFW area, making an attempt to capture of the DFW *metropolitan statistical area* (MSA). Alternatively, you can use the *find tool*.

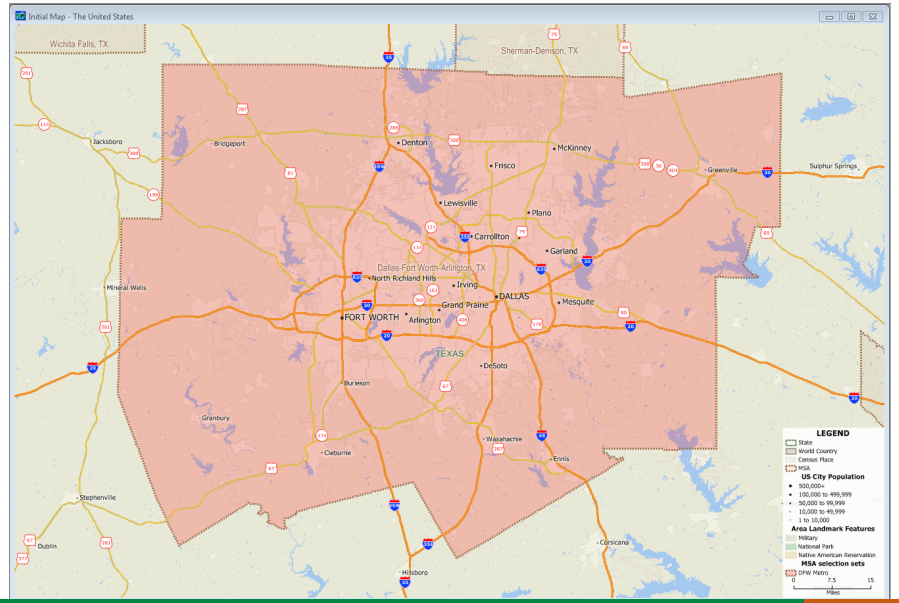


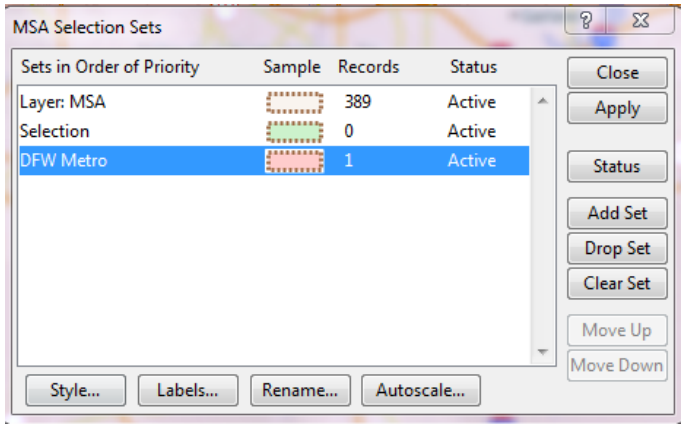


Step 1a – Location Selection: DFW

2 Enable the MSA layer and right-click the layer → Make Working Layer. In the selection window, use the "Select by Pointing" option or alternatively Selection → Select by Condition. Another option is to select within the Dataview of the MSA layer. Select the Dallas-Fort Worth-Arlington MSA.

	906900	0	17780		College Station-Bryan, TX	2,100.15	33.25	$41,067	$6
	906918	1	19100	206	Dallas-Fort Worth-Arlington, TX	9,277.78	358.29	$59,946	$8
	986450	0	26420	288	Houston-The Woodlands-Sugar Land, TX	8,258.25	1,185.57	$59,649	$8

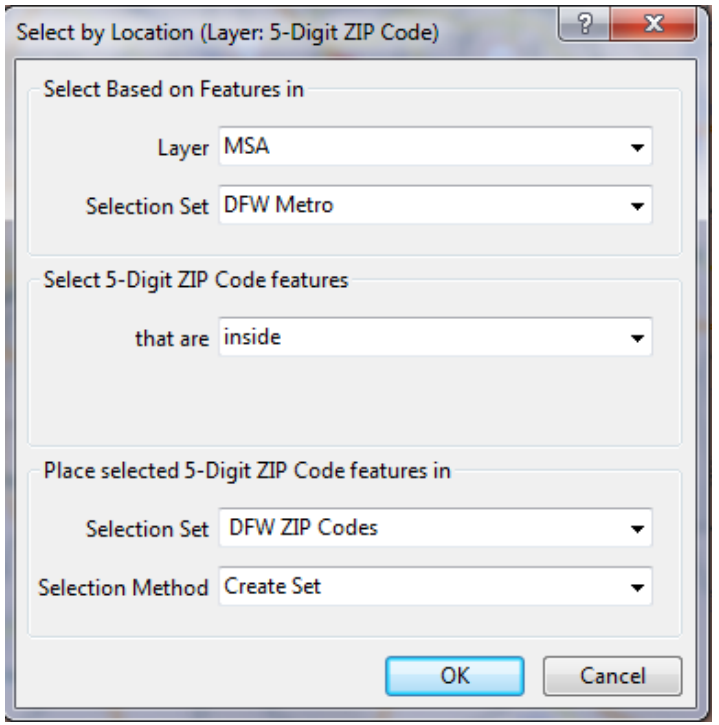




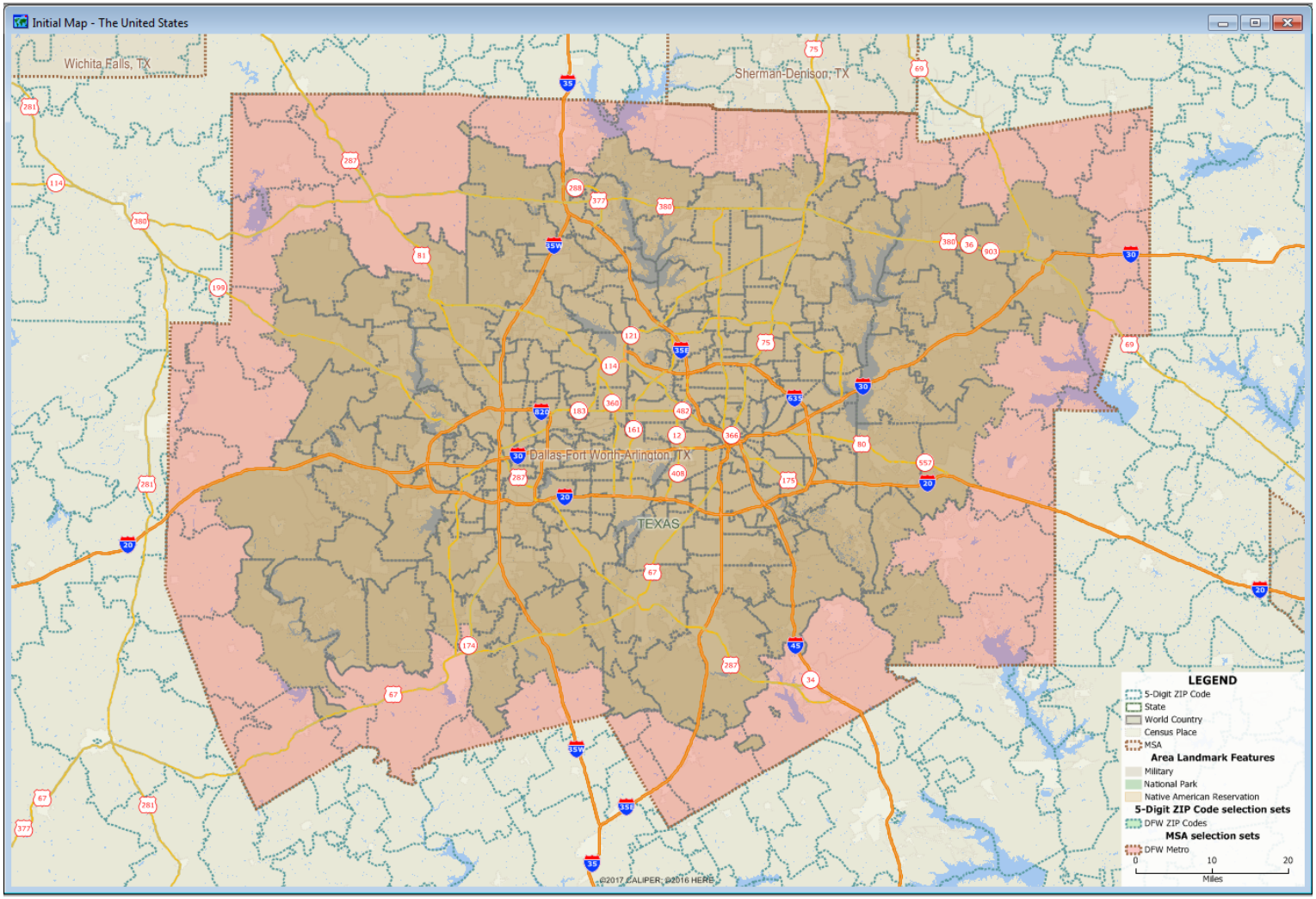
Step 1b – Location Selection: ZIP Codes

3. Make the 5-digit ZIP Codes your working layer.
4. Navigate to Selection → Select by Location
   - Layer: MSA
   - Selection Set: Selection
   - Select 5-Digit ZIP Code feature that are: touching or contained
5. From these options, make a new selection set called `DFW ZIP Codes`

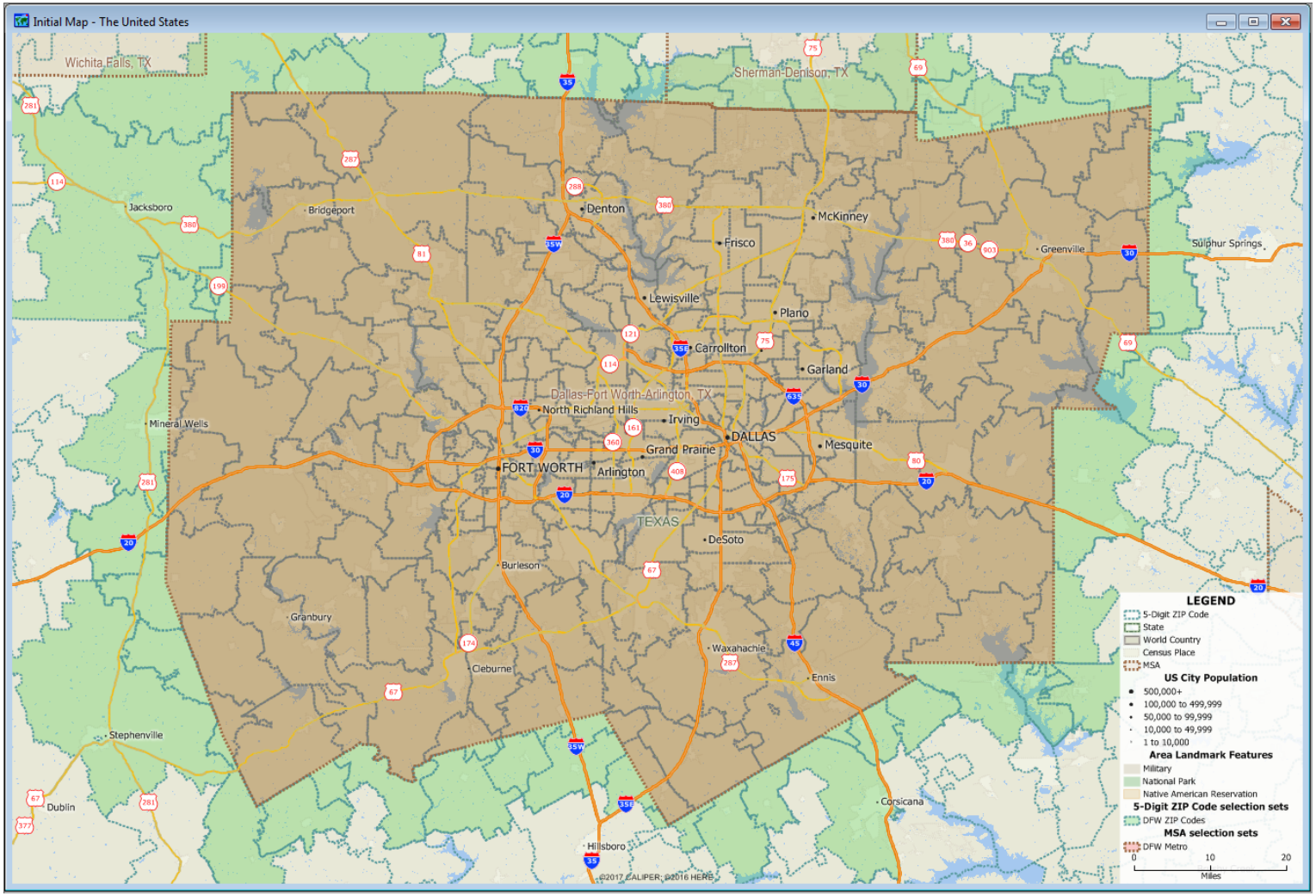
- This creates a selection issue – in one case, we are not selecting all ZIP Codes, in another case, we are selecting too many ZIP Codes.



ZIP Code Select by Location: Inside



ZIP Code Select by Location: Touching or Contained

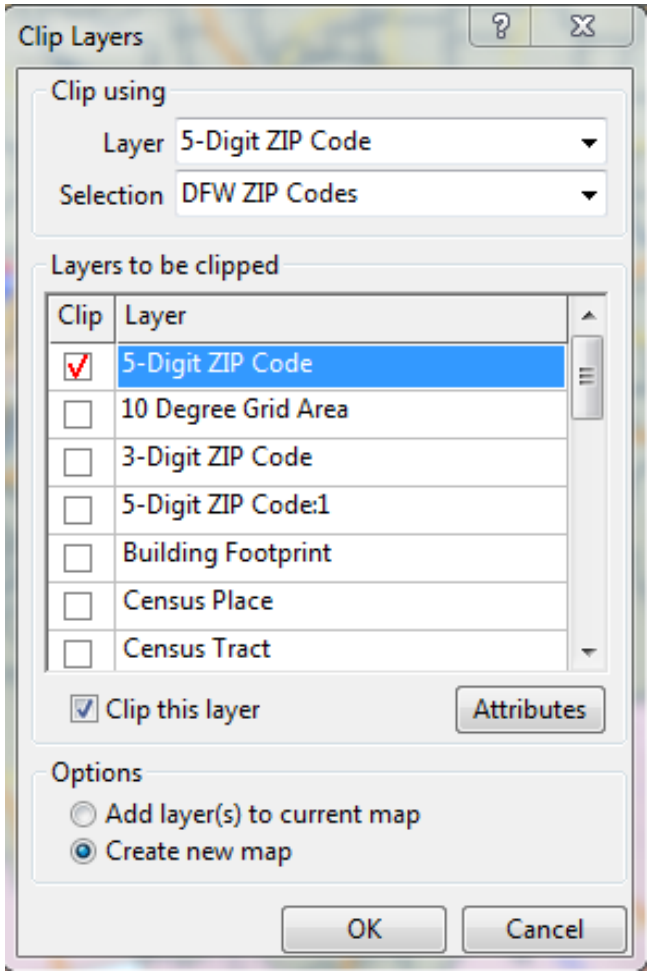


Step 1b – Location Selection: ZIP Codes

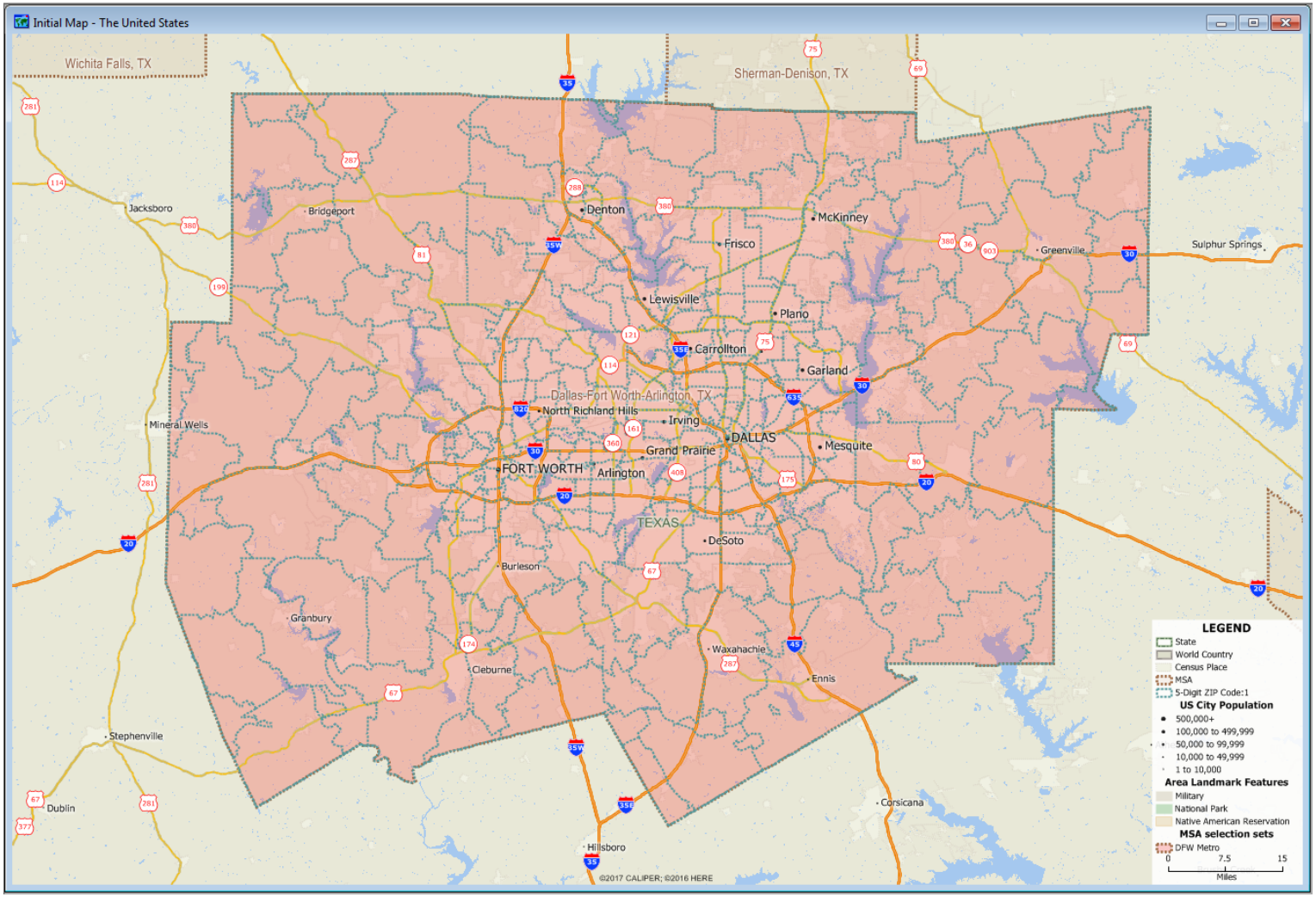
- A solution to this problem is to clip the data to the DFW MSA layer.

6 Select the ZIP Codes according to the "touching or contained" specification.

7 Navigate to Tools → Editing → Clip Layers... and input the selection parameters in the image to the right.



ZIP Code Select by Location: Clipped

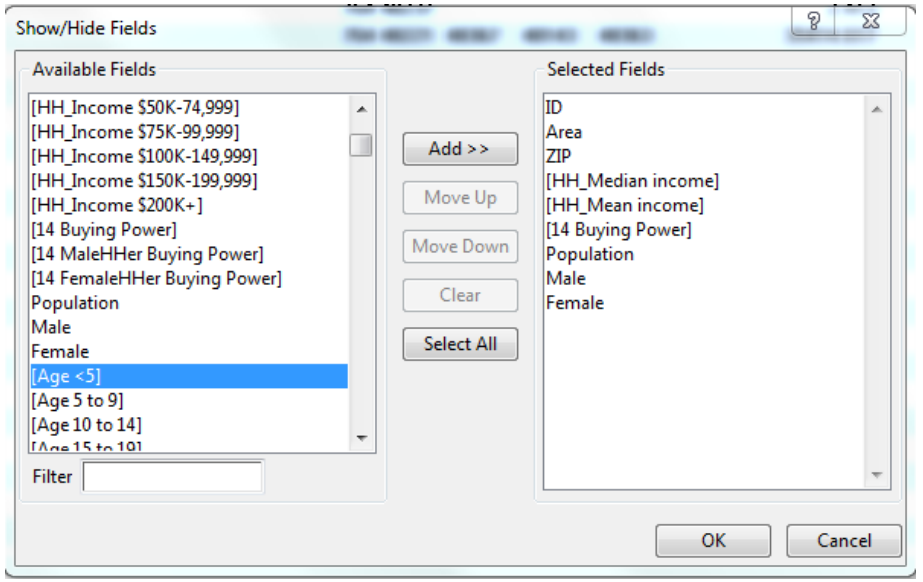


Step 1c – Location Selection: Census Tracts

3. Make the census tracts your working layer.
4. Navigate to Selection → Select by Location
   - Layer: MSA
   - Selection Set: DFW Metro (or Selection)
   - Select Census Tract features that are: inside
5. From these options, make a new selection set called `DFW CTs`

Step 2a – Selecting Relevant Fields

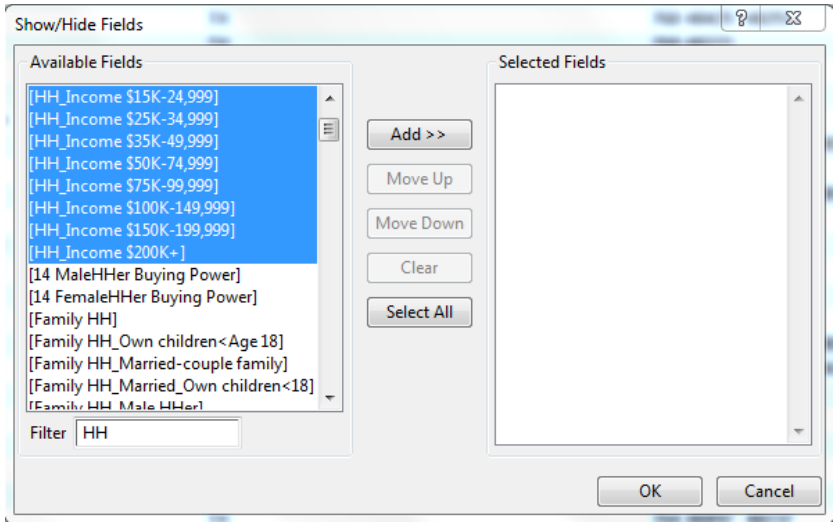
1. We first select the data that you want to choose to export. To do this, open the dataview of your current selection.
2. Click on the Show/Hide fields button 
3. In the Show/Hide Fields window, select all relevant fields that you wish to export. Click OK.

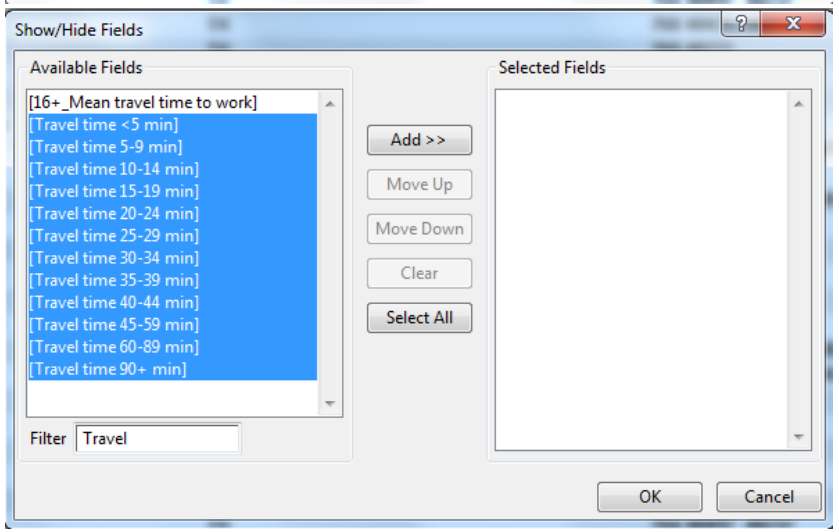


Step 2a – Selecting Relevant Fields

- You can manually select fields that are relevant by holding <CTRL> + left-click the field of interest.
- Alternatively, you can filter available fields with a common keyword in the name.
  - HH, Age, Male, or Female
  - This is useful for repetitive field names

4 Navigate to File → Save As. Save your dataview as whatever file type you wish (MS Excel .xlxs extension recommended).

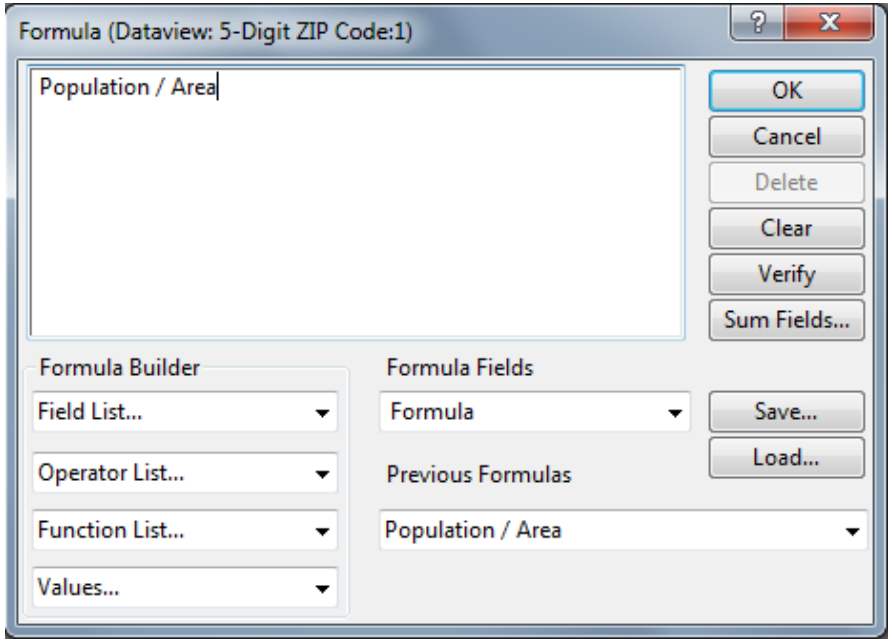




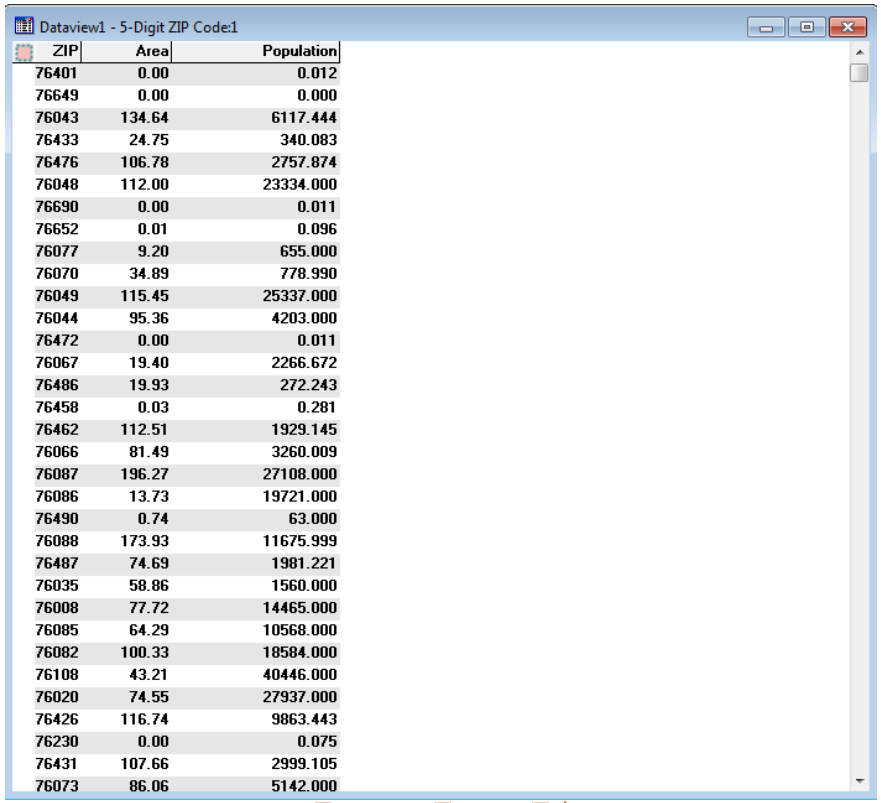
Step 2b – Calculating Fields

- You can calculate fields within dataviews. For this example, we will calculate the population density. Ensure that the area and population fields are visible.

1 Click on the formula field icon fx.



2 Create and name a formula using the drop-down windows. Click OK.



ZIP	Area	Population
76401	0.00	0.012
76649	0.00	0.000
76043	134.64	6117.444
76433	24.75	340.083
76476	106.78	2757.874
76048	112.00	23334.000
76690	0.00	0.011
76652	0.01	0.096
76077	9.20	655.000
76070	34.89	778.990
76049	115.45	25337.000
76044	95.36	4203.000
76472	0.00	0.011
76067	19.40	2266.672
76486	19.93	272.243
76458	0.03	0.281
76462	112.51	1929.145
76066	81.49	3260.009
76087	196.27	27108.000
76086	13.73	19721.000
76490	0.74	63.000
76088	173.93	11675.999
76487	74.69	1981.221
76035	58.86	1560.000
76008	77.72	14465.000
76085	64.29	10568.000
76082	100.33	18584.000
76108	43.21	40446.000
76020	74.55	27937.000
76426	116.74	9863.443
76230	0.00	0.075
76431	107.66	2999.105
76073	86.06	5142.000

Step 2b – Calculating Fields

- Notice that the new field is highlighted in green.
- This field is only a temporary calculation; if you want to open this data in Maptitude again, you must save it in a permanent field.

3 Click on the modify table icon  .

4 Click on the Add Field button and give the field an appropriate name and data type. Click OK.

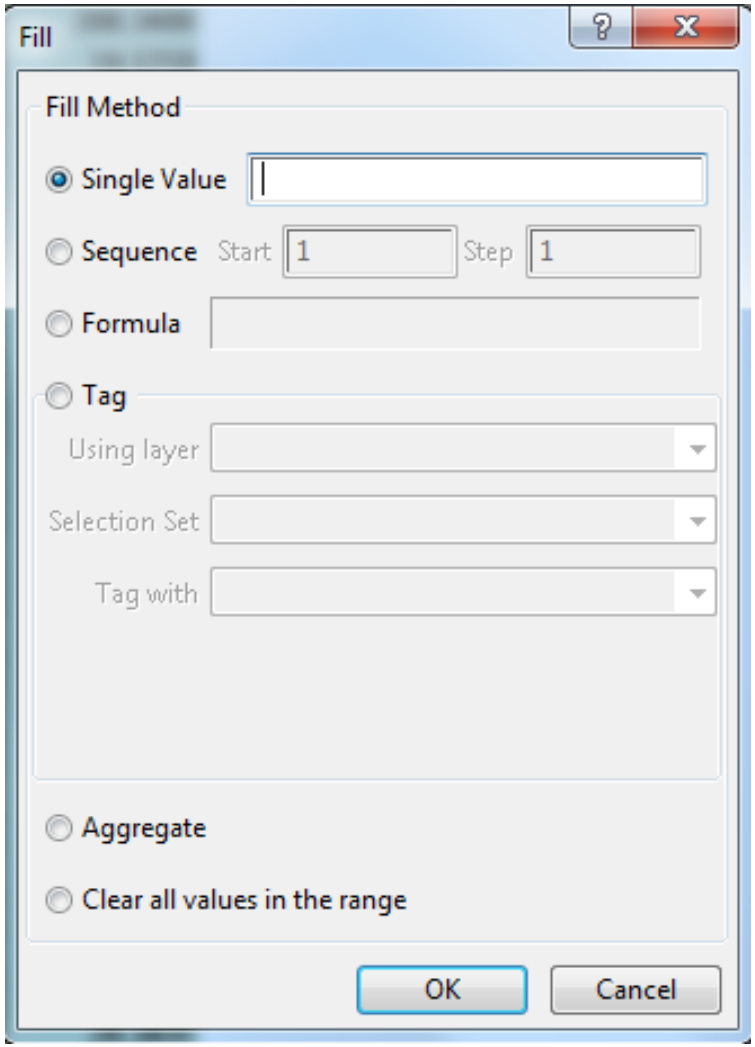
5 Right-click on the new field → Fill... → Formula.

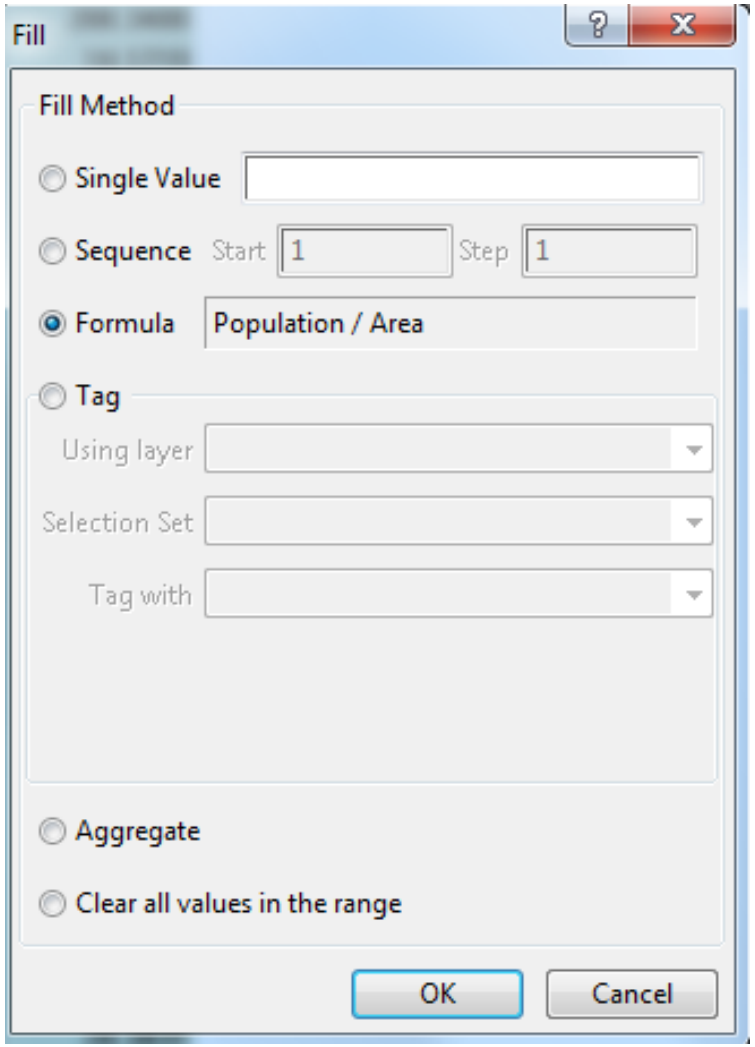
6 Use the saved formula from the temporary field. Click OK.

Dataview1 - 5-Digit ZIP Code:1

ZIP	Area	Population	POP_DEN
76401	0.00	0.012	62.1778
76649	0.00	0.000	6.6940
76043	134.64	6117.444	45.4367
76433	24.75	340.083	13.7408
76476	106.78	2757.874	25.8274
76048	112.00	23334.000	208.3400
76690	0.00	0.011	14.1719
76652	0.01	0.096	14.6713
76077	9.20	655.000	71.1940
76070	34.89	778.990	22.3270
76049	115.45	25337.000	219.4547
76044	95.36	4203.000	44.0729
76472	0.00	0.011	14.4074
76067	19.40	2266.672	116.8680
76486	19.93	272.243	13.6569
76458	0.03	0.281	10.9044
76462	112.51	1929.145	17.1467
76066	81.49	3260.009	40.0055
76087	196.27	27108.000	138.1133
76086	13.73	19721.000	1436.0177
76490	0.74	63.000	84.7155
76088	173.93	11675.999	67.1292
76487	74.69	1981.221	26.5270
76035	58.86	1560.000	26.5031
76008	77.72	14465.000	186.1086
76085	64.29	10568.000	164.3917
76082	100.33	18584.000	185.2312
76108	43.21	40446.000	936.1101
76020	74.55	27937.000	374.7305
76426	116.74	9863.443	84.4898
76230	0.00	0.075	24.3431
76431	107.66	2999.105	27.8563
76073	86.06	5142.000	59.7472

Step 2b – Calculating Fields





Step 2b – Calculating Fields

Dataview1 - 5-Digit ZIP Code:1

Area	ZIP	Population	[POP_DEN:1]	POP_DEN
0.00	76401	0.012	62.18	62.1778
0.00	76649	0.000	6.69	6.6940
134.64	76043	6117.444	45.44	45.4367
24.75	76433	340.083	13.74	13.7408
106.78	76476	2757.874	25.83	25.8274
112.00	76048	23334.000	208.34	208.3400
0.00	76690	0.011	14.17	14.1719
0.01	76652	0.096	14.67	14.6713
9.20	76077	655.000	71.19	71.1940
34.89	76070	778.990	22.33	22.3270
115.45	76049	25337.000	219.45	219.4547
95.36	76044	4203.000	44.07	44.0729
0.00	76472	0.011	14.41	14.4074
19.40	76067	2266.672	116.87	116.8680
19.93	76486	272.243	13.66	13.6569
0.03	76458	0.281	10.90	10.9044
112.51	76462	1929.145	17.15	17.1467
81.49	76066	3260.009	40.01	40.0055
196.27	76087	27108.000	138.11	138.1133
13.73	76086	19721.000	1436.02	1436.0177
0.74	76490	63.000	84.72	84.7155
173.93	76088	11675.999	67.13	67.1292
74.69	76487	1981.221	26.53	26.5270
58.86	76035	1560.000	26.50	26.5031
77.72	76008	14465.000	186.11	186.1086
64.29	76085	10568.000	164.39	164.3917
100.33	76082	18584.000	185.23	185.2312
43.21	76108	40446.000	936.11	936.1101
74.55	76020	27937.000	374.73	374.7305
116.74	76426	9863.443	84.49	84.4898
0.00	76230	0.075	24.34	24.3431
107.66	76431	2999.105	27.86	27.8563
86.06	76073	5142.000	59.75	59.7472

Step 2c – Calculating Statistics

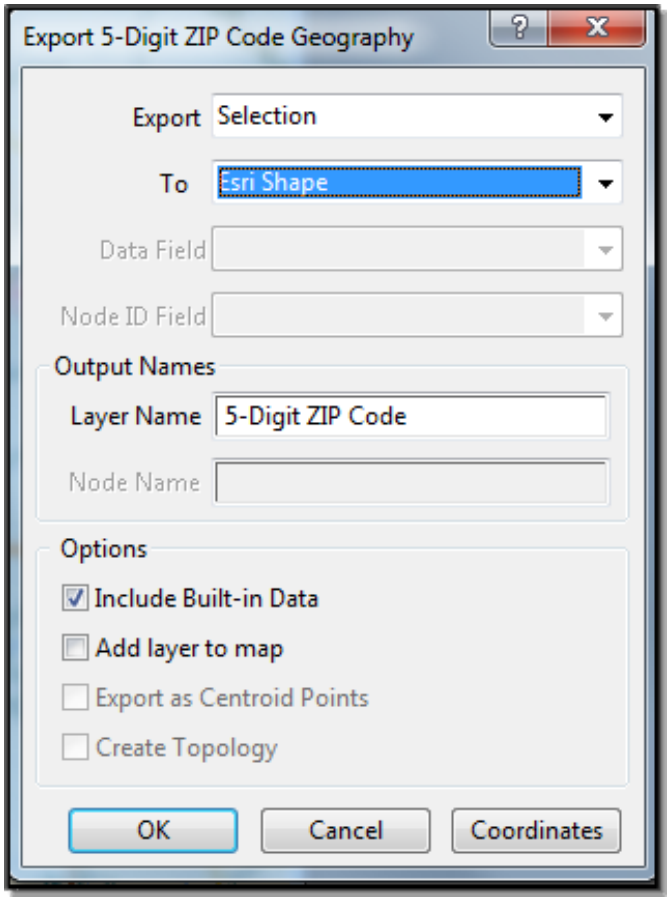
1 You can use the compute statistics tool  when the dataview is open to view summary statistics about your data.

Dataview4 - 5-Digit ZIP Code:1 Statistics

Field	Count	Sum	Minimum	Maximum	Mean	[Std. Dev.]
Area	316	9641.2030	0.0000	304.2388	30.510136	42.4200
ZIP	316	23899943.0000	75001.0000	76690.0000	75632.731013	517.2340
Population	316	6828647.0949	0.0000	92532.0000	21609.642705	19180.6727
[POP_DEN:1]	316	679667.8672	0.0000	53185.2734	2150.847681	3572.7344
POP_DEN	316	679667.8639	0.0000	53185.2725	2150.847671	3572.7343

Step 3a – Exporting Data as a Shapefile

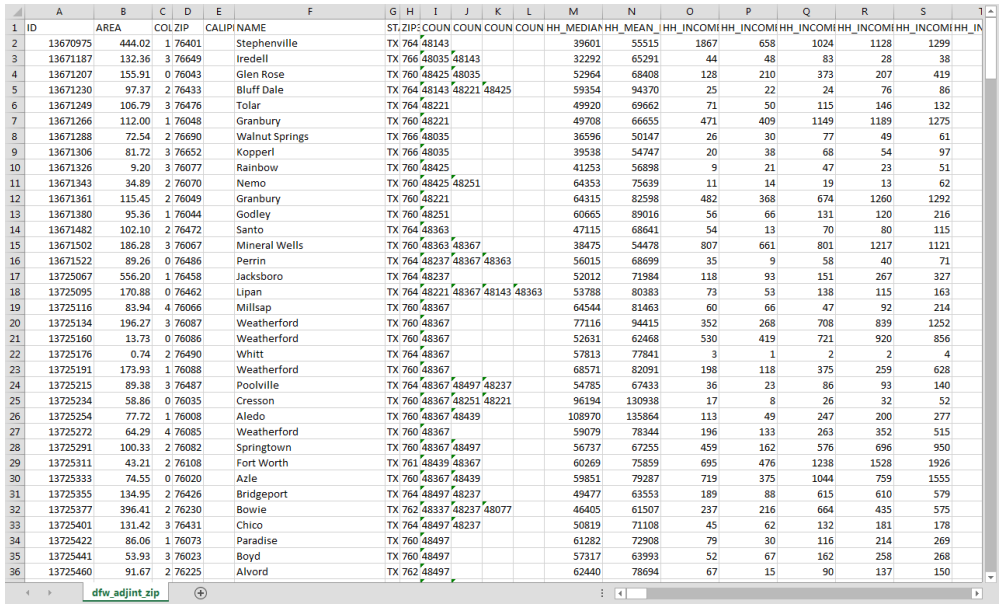
1. Once the data is selected, go to File → Export → Geography.
   - Export: Selection
   - To: ESRI Shape
2. Several files that comprise the shapefile will be exported to your local disk.
3. The DBF file will contain all relevant selected data per ZIP Code.



Step 3b – Importing Data into MS Excel

Only if you want the data directly from a shapefile database.

1. Right-click on the shapefile's DBF file → Copy → Paste.
2. Change the extension from a .DBF to .CSV. Click OK when the warning prompts.
3. Double-click on the new file to open in MS Excel.
4. Delete irrelevant columns.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	ID	AREA	COL	ZIP	CALIP	NAME	ST	ZIP	COUN	COUN	COUN	COUN	HH_MEDIAN	HH_MEAN_	HH_INCOM	HH_INCOM	HH_INCOM	HH_INCOM	HH_INCOM	HH_IN
2	13670975	444.02	1	76401		Stephenville	TX	764	48143				39601	55515	1867	658	1024	1128	1299	
3	13671187	132.36	3	76649		Iredell	TX	766	48035	48143			32292	65291	44	48	83	28	38	
4	13671207	155.91	0	76043		Glen Rose	TX	760	48425	48035			52964	68408	128	210	373	207	419	
5	13671230	97.37	2	76433		Bluff Dale	TX	764	48143	48221	48425		59354	94370	25	22	24	76	86	
6	13671249	106.79	3	76476		Tolar	TX	764	48221				49920	69662	71	50	115	146	132	
7	13671266	112.00	1	76048		Granbury	TX	760	48221				49708	66655	471	409	1149	1189	1275	
8	13671288	72.54	2	76690		Walnut Springs	TX	766	48035				36596	50147	26	30	77	49	61	
9	13671306	81.72	3	76652		Kopperl	TX	766	48035				39538	54747	20	38	68	54	97	
10	13671326	9.20	3	76077		Rainbow	TX	760	48425				41253	56898	9	21	47	23	51	
11	13671343	34.89	2	76070		Nemo	TX	760	48425	48251			64353	75639	11	14	19	13	62	
12	13671361	115.45	2	76049		Granbury	TX	760	48221				64315	82598	482	368	674	1260	1292	
13	13671380	95.36	1	76044		Godley	TX	760	48251				60665	89016	56	66	131	120	216	
14	13671482	102.10	2	76472		Santo	TX	764	48363				47115	68641	54	13	70	80	115	
15	13671502	186.28	3	76067		Mineral Wells	TX	760	48363	48367			38475	54478	807	661	801	1217	1121	
16	13671522	89.26	0	76486		Perrin	TX	764	48237	48367	48363		56015	68699	35	9	58	40	71	
17	13725067	556.20	1	76458		Jacksboro	TX	764	48237				52012	71984	118	93	151	267	327	
18	13725095	170.88	0	76462		Lipan	TX	764	48221	48367	48143	48363	53788	80383	73	53	138	115	163	
19	13725116	83.94	4	76066		Millsap	TX	760	48367				64544	81463	60	66	47	92	214	
20	13725134	196.27	3	76087		Weatherford	TX	760	48367				77116	94415	352	268	708	839	1252	
21	13725160	13.73	0	76086		Weatherford	TX	760	48367				52631	62468	530	419	721	920	856	
22	13725176	0.74	2	76490		Whitt	TX	764	48367				57813	77841	3	1	2	2	4	
23	13725191	173.93	1	76088		Weatherford	TX	760	48367				68571	82091	198	118	375	259	628	
24	13725215	89.38	3	76487		Poolville	TX	764	48367	48497	48237		54785	67433	36	23	86	93	140	
25	13725234	58.86	0	76035		Cresson	TX	760	48367	48251	48221		96194	130938	17	8	26	32	52	
26	13725254	77.72	1	76008		Aledo	TX	760	48367	48439			108970	135864	113	49	247	200	277	
27	13725272	64.29	4	76085		Weatherford	TX	760	48367				59079	78344	196	133	263	352	515	
28	13725291	100.33	2	76082		Springtown	TX	760	48367	48497			56737	67255	459	162	576	696	950	
29	13725311	43.21	2	76108		Fort Worth	TX	761	48439	48367			60269	75859	695	476	1238	1528	1926	
30	13725333	74.55	0	76020		Azle	TX	760	48367	48439			59851	79287	719	375	1044	759	1555	
31	13725355	134.95	2	76426		Bridgeport	TX	764	48497	48237			49477	63553	189	88	615	610	579	
32	13725377	396.41	2	76230		Bowie	TX	762	48337	48237	48077		46405	61507	237	216	664	435	575	
33	13725401	131.42	3	76431		Chico	TX	764	48497	48237			50819	71108	45	62	132	181	178	
34	13725422	86.06	1	76073		Paradise	TX	760	48497				61282	72908	79	30	116	214	269	
35	13725441	53.93	3	76023		Boyd	TX	760	48497				57317	63993	52	67	162	258	268	
36	13725460	91.67	2	76225		Alvord	TX	762	48497				62440	78694	67	15	90	137	150	

dfw_adjint_zip

Summary and Conclusions

- Much of the geographic hierarchy data is available pre-installed or for purchase in the Maptitude software.
- Up-to-date country-level geographic and demographic data are available for download, along with a variety of nationwide data layers.
- It is easy to find geographic and demographic data using Maptitude, and to then export these data for further analysis in third party software.

Additional Resources

- Additional data available for download: `https://www2.caliper.com/store/product-category/download-category/`
- To access information about data packages, in Maptitude, navigate to Help → Data Package Help → U.S. Region Help (HERE).
- For additional information on census summary levels, visit: `https://www.caliper.com/maptitude/census2000data/summarylevels.htm`
- For information on available United States data, visit: `https://www.caliper.com/maptitude/international/unitedstates.htm`

Questions?