USP/PLSI 492: Research Methods Instructor: Pietro Calogero

Getting and tweaking TIGER files in QGIS

The purpose of this exercise is to get the geographic data for your study area from the U.S. census. NOTE: for some jurisdictions, like the City of San Francisco, you can get a lot of ready-made GIS data. However we are going to learn how to acquire "raw" data from the US census, in a method that can be applied for any jurisdiction across the U.S.: from Winemucca to Kalamazoo.

1. Get the TIGER shapefile for your geographic area.

IN THIS TUTORIAL, I GET THE TIGER DATA A DIFFERENT WAY FROM THE ARCMAP TUTORIAL

Rather than do a separate search down through the generic TIGER database, I get the TIGER map straight from American Factinder2. The advantage of this method is that you have already set the FactFinder to seek data from a specific geographic level. For this course I have been using Census Tracts as the geographic unit, but I checked and this also works for Oakland Unified School district.

Here's how it works: When you are in FactFinder, set your geographic filters (in my example: Census tracts [140]; California; County of San Francisco; all tracts) and then click on the MAP tab:

	FactFinder
MAIN COMMUNITY FACTS	GUIDED SEARCH ADVANCED SEARCH DOWNLOAD CENTER
Search - Use the options	on the left (topics, geographies,) to narrow your search results
Your Selections	Search Results: 1-25 of 5,641 tables and other products match 'Your Selections'
Search using Census Tract All Census Tracts within San Francisco County, California 🕄	Select Geographies
clear all selections and	List Name Address Map
Search using the options below: Topics (age, income, year, dataset,)	Select geographies to add to Your Selections 🕜 🎽 Select from: • most requested geographic types — all geographic types
Geographies (states, counties, places,)	Cenetra geographic type.
Race and Ethnic Groups (race, ancestry, tribe)	California V
Industry Codes (NAICS industry,)	Select a county:
EEO Occupation Codes (executives, analysts,)	Select one or more geographic areas and click Add to Your Selections:
	All Census Tracts within San Francisco County, California Census Tract 101, San Francisco County, California Census Tract 102, San Francisco County, California

This will generate a preview which should look like the geography you are working with:



NOTE! The geography looks projected! It is! In fact, it uses the "web-mercator" projection, the same projection used by Google Earth. More on that below.

If it looks like the data you seek, then click the download button (notice my obnoxiously-recolored cursor). That invokes the download dialog:

elect a download format and click	ок. 🕜
Comma delimited (.csv) format ((.csv is compatible with spreadsheet Data and annotations in a Data and annotations in se Include descriptive data ele	data rows only) programs such as Microsoft Excel) single file parate files rment names
Presentation-ready formats PDF JPEG image (,jpg) Microsoft Excel (.xls) Rich Text Format (.rtf)	- Orientation Portrait Landscape - Paper size 8 1/2" x 11" 8 1/2" x 14"
Map Title: [
Spatial Data formats Shapefiles (.zip) cospatial data for the geograph PDFJ or D indicate a document in Adob	iles in 'Your Selections') e's Portable Document Format . To view the file you * available free from Adobe.

When you download it, remember to save the file with a name, and at a path that you can remember. Unzip the file, and it will generate a folder called "reference_map_shape" with a shapefile called "140_00". This shapefile name is the geographic code-number for census tract data.

2. Open this TIGER file in QGIS.

a) Start QGIS. By default it opens a blank data set.

b) Click LAYER > ADD VECTOR LAYER... and then Browse.. the Path to your TIGER file, and open it. c) QGIS will ask you to specify the Coordinate Reference System for this shapefile. Why? Because (at least for the moment) QGIS does not understand "web mercator" as a projection system. So I looked it up for you. It is EPSG 3857, which you can type up in the "Filter" field.(I am showing this in the right-hand image; the Filter field is highlighted in yellow) QGIS calls this "WGS84/Pseudo Mercator". I don't know why; maybe they look down on it. I don't care. That is the correct projection, because the EPSG code matches. (EPSG stands for European Petroleum Survey Group)

Coordinate Reference System	Selector	$\odot \odot \odot \otimes$	× 🔾	Coordinate Refe	rence System Selector	$\odot \odot \odot \odot$
Specify CRS for layer 140_00			Specify CRS for layer	140_00		
Filter			Filter epsg 3857	lineto offerance curbance		×
Recencty used coordinate reference systems			Recencty used coord	inace rererence systems		
Coordinate Reference System * Generated CRS (+proj=lcc +lat_1=37.0666666666666667 +lat_2=3. * Generated CRS (+proj=lcc +lat_1=35.25 +lat_2=36.4166666666 NAD83 / UTM zone 10N	Authority ID USER:100001 USER:100000 EPSG:26910		Coordinate Referen WGS 84 / Pseudo M	ice System ercator	: Authority ID EPSG:3857	
NAD27 NAD83 WGS 84	EPSG:4267 EPSG:4269 EPSG:4326					
Coordinate reference systems of the world		Hide deprecated CRSs	Coordinate reference	e systems of the world	Authority ID	Hide deprecated CRSs
Viti Levu 1916	EPSG:4731		v- III Projected	Coordinate Systems		
- Voirol 1875	EPSG:4304		✓- Mercator			
— Voirol 1875 (Paris)	EPSG:4811		WGS 84	1 / Pseudo Mercator	EP5G:3857	
Voirol 1879	EPSG:4671			13		
Voirol 1879 (Paris)	EPSG:48					
WG 5 66	EPSG:4760					
- WGS 72BE	EPSG-4324					
	EPSG:4326	Ŷ				
× C			<			
			calastad oper juice	04 / Decude Meanshee		
Selected CRS: WG5 84			Selected CRS: WG	5 64 / Pseudo Mercacor		
+proj=longlat +datum=WGS84 +no_defs			+proj=merc +a=637 +nadgrids=@null +v	8137 +b=6378137 +lat_ts=0.0 +lo vktext +no_defs	n_0=0.0 +x_0=0.0 +y_0=0 +k=1.0 +u	nits=m
W Help	×	OK 🥝 Cancel	K Help		_	V OK 💋 Cancel

When you load the shapefile it should look something like this:



0

Now, right-click on the shapefile name over in the left-hand "Layers" pane. A drop-menu will open; choose "Open Attribute Table".

) 🖭 💌 😕						
gle editing mode	(Ctrl+E) TATE	: COUNT	Y TRACT	i NAME	: LSAD	SHAPE_AREA	SHAPE_LEN	
400000US0607	06	075	010300	103	Tract	433175.1209500	2904.62864510000	
400000US0607	06	075	010500	105	Tract	1664779.053169	5827.41876756000	
400000US0607	06	075	011902	119.02	Tract	149323.1021729	1648.69404694000	
400000US0607	06	075	012901	129.01	Tract	296110.9923759	2176.86222679000	
400000US0607	06	075	013101	131.01	Tract	285471.5848259	2282.32548072000	
400000US0607	06	075	013500	135	Tract	377268.8666830	2500.82628697000	
400000US0607	06	075	015801	158.01	Tract	544084.1575459	3396.78580055000	
400000US0607	06	075	015900	159	Tract	460483.3693329	3050.18183773000	
400000US0607	06	075	017101	171.01	Tract	492519.3560960	3370.00624026000	
400000US0607	06	075	020600	206	Tract	904935.0732739	3829.56645776000	
400000US0607	06	075	022704	227.04	Tract	865620.4988130	3632.61997739000	
1000001150607	0.6	075	023300	223	Tract	1028430 171770	4498 98274774000	

Oh no! What is missing is GEOID10, the field that we will use to join in the AFF data we have prepared. Fortunately, we can create this column by concatenating together three columns we already have: STATE, COUNTY, and TRACT. So! We will use this as an opportunity to generate a new attribute field!

3. Use the "Column Calculator" to create a new field.

1.) Click the "Toggle editing mode" button in the upper left corner of the "Attribute table" window.

2.) Now click the right-most button in that same row, which looks like an abacus. That will bring up the "Field Calculator". (CONTINUE TO NEXT PAGE)



Output field name: name it something that makes sense to you, but use the "dehydrated" syntax: only letters, numbers, or the underscore (_) character. No spaces. No other punctuation.

Output field type: this should be Text (string), for the same reasons we discussed in the previous exercise with AFF data and Excel: you don't want any software stripping off leading zeroes.

Output field width: this will need to be 11 characters.

Now it is time to build the command to create this new column. In the **Function List** pane, I chose: String > concat

The syntax of this function is concat ("a", "b", "c") which produces: abc

After **concat** I typed a parenthesis mark, then in the **Function List** I opened "Fields and Values" and double-clicked STATE, COUNTY, and TRACT. Then I added a close-parenthesis mark. This gives me a valid Output preview at the bottom of the window (just above the **Help** button): 06075010300.

Since it looks correct, I hit OK. Now the Attribute Table has one more column on the right-hand end:

Save Edits (Ctrl+S)	i state 🗸	COUNTY	i TRACT	i NAME	i LSAD	SHAPE_AREA SHAPE_LEN GEO_id2	
0000000506075010300	06	075	010300	103	Tract	433175.1209500 2904.62864510000 06075010300	
100000US06075010500	06	075	010500	105	Tract	1664779.053169 5827.41876756000 06075010500	
100000US06075011902	06	075	011902	119.02	Tract	149323.1021729 1648.69404694000 06075011902	
00000US06075012901	06	075	012901	129.01	Tract	296110.9923759 2176.86222679000 06075012901	
400000US06075013101	06	075	013101	131.01	Tract	285471.5848259 2282.32548072000 06075013101	
00000US06075013500	06	075	013500	135	Tract	377268.8666830 2500.82628697000 06075013500	
100000US06075015801	06	075	015801	158.01	Tract	544084.1575459 3396.78580055000 06075015801	
00000US06075015900	06	075	015900	159	Tract	460483.3693329 3050.18183773000 06075015900	-
00000US06075017101	06	075	017101	171.01	Tract	492519.3560960 3370.00624026000 06075017101	_
00000US06075020600	06	075	020600	206	Tract	904935.0732739 3829.56645776000 06075020600	
00000US06075022704	06	075	022704	227.04	Tract	865620.4988130 3632.61997739000 06075022704	2
10000011506075023300	0.6	075	023300	222	Tract	1028430 171770 4408 08274774000 06075023300	-

Now Save Edits and you will have a GEO_id2 column in the attribute table of this shapefile that corresponds to the GEO.id2 field in Factinder CSV downloads.

At this point you could move on to the next exercise of joining your FactFinder data to this shapefile. However, I am going to show a few more steps to make this file look even better. You don't have to do the steps below for the purposes of full credit on this exercise, but you may want to know the following techniques for future use.

4. OPTIONAL: Make the shorelines crispy-clean.

What we will do here is get a detailed shoreline shapefile and use it to "erase back" the edges of the census tracts to show precise shorelines. The TIGER file we will get is called "areawater". And again, the advantage of getting this particular file is that it is available for all regions across the country.
1) Go to the main Census.gov page; > Geography link; > TIGER > **Product**: TIGER/Line shapefiles > Tab: 2010 > **Select Layer Type**: Water > **Area Hydrography**: California > San Francisco County > Download
2) Unzip the archive. Remember: name it and locate it at a sensible path.

3) Back in QGIS, add it to the working file. Even though "areawater" is an unprojected file, when you add it to this data-frame, it will reproject to align with your projected "tracts" file in this data-frame. That is because QGIS uses the first file loaded into the data-frame to set the projection. QGIS will then try to reproject every subsequent file to match that dataframe.



03

Looks nice and crispy, yes? Yes. But I want to use that "areawater" file to trim back my census tract file. That way, census tracts along the shore will have the correct land area, and the file will look nicer. The problem is that QGIS will not use a shapefile in one datum/projection to trim back another shapefile in a different datum/projection. So the next step is to reproject the "areawater" shapefile into WGS84/Web Mercator.

In the "Layers" pane on the left, put your cursor over the "areawater" shapefile and right-click. This drops down a menu with various options. At this point, choose "Save As..." This will generate a new shapefile, with a new name, and a new datum/projection. At the "Save As..." field, click "Browse" to select a path and sensible filename. At the "CRS" field, also click "Browse" to select the coordinate reference system you want. In this case, "WGS 84 / Pseudo Mercator".

× 💿	Save vector layer as	
Format	ESRI Shapefile	~
Save as	GER/tl_2010_06075_areawater/tl_2010_06075_areawater_webmercator3857.shp	Browse
Encoding	System	~
605	Selected CRS	~
CRS	WGS 84 / Pseudo Mercator	Cowse
Symbology export	No symbology	
Scale	1:50000	٥)
	OGR creation options	

QGIS will automatically add this new shapefile to the workspace. If not, go and add it manually: LAYER > ADD VECTOR LAYER... It should look exactly like the original "areawater" file in this view, but since it now has the same datum/projection as the Census Tract file (140_00), you can now use this reprojected "areawater" file to trim back the tract polygons to a more precise representation of the shoreline. Do this by going to VECTOR > GEOPROCESSING TOOLS > DIFFERENCE:



The "input layer" is the census tract layer I want to be trimmed. The "Difference Layer" is the "areawater" file, reprojected to match the census-tract file. Then I use the "Browse" button to set the path and filename of the new shapefile. The resulting file looks like the previous overlays; but now I can remove all the other shapefiles and this is what remains:



à

Okay. NOW our TIGER file is properly tweaked for further analysis. For now, let us keep the weird outlying census tracts of the Farallons, the Marin Headlands, and the western tip of Alameda island. That way when we add AFF data, there won't be any grief about missing polygons.

Appendix: State Plane and UTM zones for reprojection of various California counties

For your reference. When reprojecting into UTM or State Plane, you need to use the correct "zone" for the reprojection. Each of these projections turns the earth into a faceted ball, kind of like a mirror-ball. So you need to pick the correct "facet" for your county.

http://home.comcast.net/~rickking04/gis/spc.htm

CALIFORNIA ZONE I FIPSZONE: 0401 ADSZONE: 3276 UTM ZONE: 10 DEL NORTE, HUMBOLDT, LASSEN, MODOC, PLUMAS, SHASTA, SISKIYOU, TEHAMA, TRINITY

CALIFORNIA ZONE II FIPSZONE: 0402 ADSZONE: 3301 UTM ZONES: 10 & 11 ALPINE, AMADOR, BUTTE, COLUSA, EL DORADO, GLENN, LAKE, MENDOCINO, NAPA NEVADA, PLACER, SACRAMENTO, SIERRA, SOLANO, SONOMA, SUTTER, YOLO, YUBA

CALIFORNIA ZONE III FIPSZONE: 0403 ADSZONE: 3326 UTM ZONES: 10 & 11 ALAMEDA, CALAVERAS, CONTRA COSTA, MADERA, MARIN, MARIPOSA, MERCED, MONO, SAN FRANCISCO SAN JOAQUIN, SAN MATEO, SANTA CLARA, SANTA CRUZ, STANISLAUS, TUOLUMNE

CALIFORNIA ZONE IV FIPSZONE: 0404 ADSZONE: 3351 UTM ZONES: 10 & 11 FRESNO, INYO, KINGS, MONTEREY, SAN BENITO, TULARE

CALIFORNIA ZONE V FIPSZONE: 0405 ADSZONE: 3376 UTM ZONES: 10 & 11 KERN, LOS ANGELES, SAN BERNARDINO, SAN LUIS OBISPO, SANTA BARBARA, VENTURA

CALIFORNIA ZONE VI FIPSZONE: 0406 ADSZONE: 3401 UTM ZONE: 11 IMPERIAL, ORANGE, RIVERSIDE, SAN DIEGO

Note: Los Angeles County was in CALIFORNIA ZONE VII FIPSZONE: 0407 for NAD27. ZONE VII was eliminated for NAD23 with Los Angeles County being included with ZONE V

for NAD83 with Los Angeles County being included with ZONE V.