



SAPIENZA
UNIVERSITÀ DI ROMA

Introduction to Spatial Data Analysis

1) Introduction to GIS and geodata

Economic Geography | EPOS | 2022

Filippo Celata (filippo.celata@uniroma1.it)

Georeferencing and geocoding

Georeferencing and geocoding

Available information /spatial reference:

X/Y Coordinates

Addresses

Place names

Administrative entities,
area codes

Maps, images

Georeferencing tools:

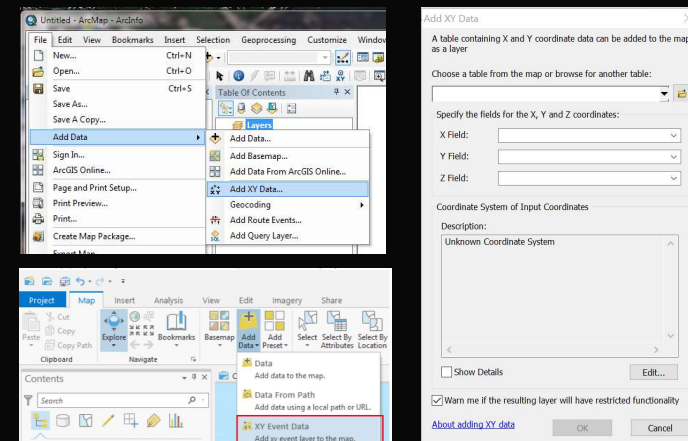
Add X/Y data (ArcGIS)

Geocoding

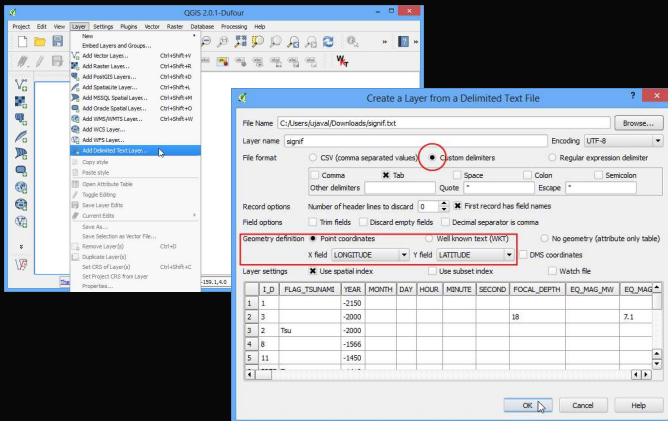
Table join

Manual georeferencing

1. Georeferencing X/Y tables: ArcGIS



2. Georeferencing X/Y tables: QGIS



X/Y coordinates formats

DMS:

75 59 32.483 W

75 59.32483' West

75° 59' 32.483'' W

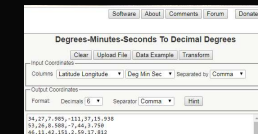
Decimal (DMS):

-7599.2356

-75.992356

-75.992356

Tools and formulas for the conversion of coordinates:



Geographic (DMS) vs. Projected coordinate systems (distance from the equator and the prime* meridian)

Xutm	Yutm
791306.7228	4643740.8315

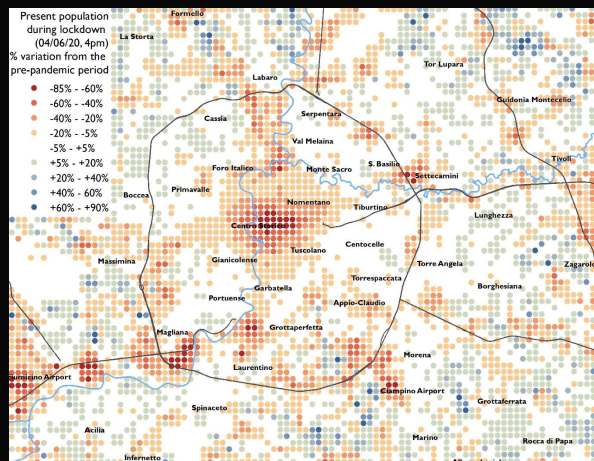
Xwgs84	Ywgs84
12.511271	41.891923

Welcome to Null Island, where lost data goes to die

● The equator and the prime meridian meet at a place denoted as 0°N, 0°E. ● This location, in the Gulf of Guinea, is where non-geocoded data goes to die. ● Recently renamed "Null Island," it has also captured the imagination — and acquired a map and several flags.



Lab: variation of the 'present' population during lockdown..



1) Click add data and add desktop/spatial22/data/fb_pop_26mar20.dbf

2) Right click on the dbf table in the table of contents to open the table

Facebook Crisis Map | Italy, tiles | March 26 2021

Xutm, Yutm: X/Y in projected coordinated system WGS_1984_UTM_Zone_32N (vs. Lon, Lat, in WGS84)

Date Time: gg, 16:00.

N_Baseline: The average number of people we expect in the area during the specified time based on pre-disaster estimates.

N_Crisis: The number of people observed in the place during the time period.

N_Difference: Crisis population - baseline population.

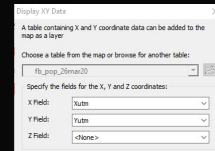
Percent Change: variation between crisis population and baseline population.

Standard (Z) Score: number of standard deviations by which the crisis population differs from the baseline population.

Convert the table using X/Y

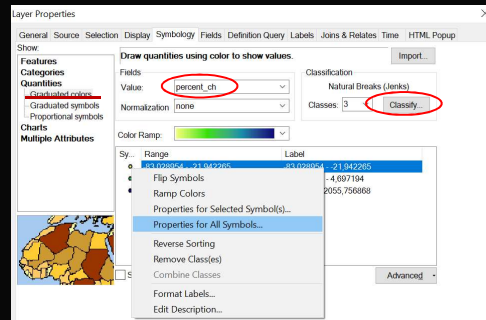
3) Right click on the table and click 'Display XY data': X Field = Xutm and Y Field = Yutm

4) Open the output layer's attribute table and check



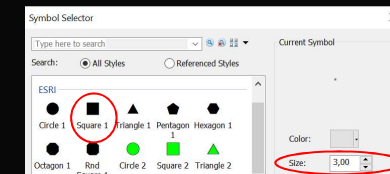
Map the present population variation <-5% and >+5%

5) Right click the 'events' layer and open its layer properties

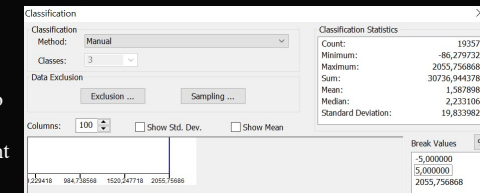


Map the present population variation <-5% and >+5%

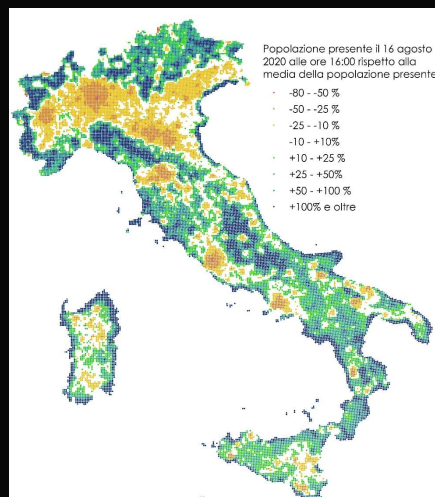
5) Right click on the default symbols and click 'properties for all symbols'



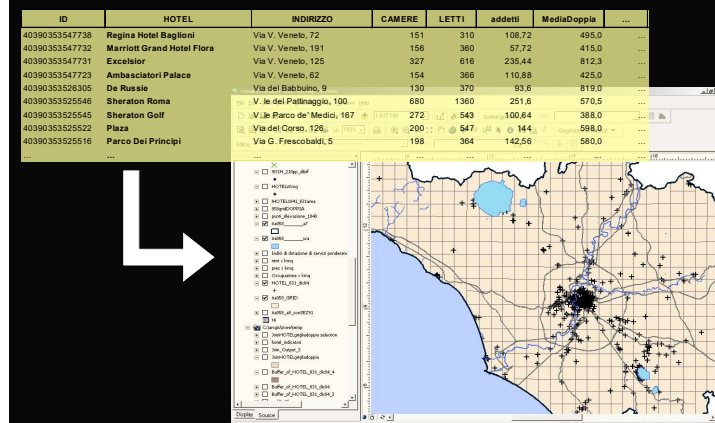
6) Enter the 'classify' menu and change the points' symbology into red (<-5%), green (>+5%) and transparent (-5% - +5%)



Present population variation during the summer



2. Geocoding services : georeferencing point data with addresses (or place names)



Geocoding services

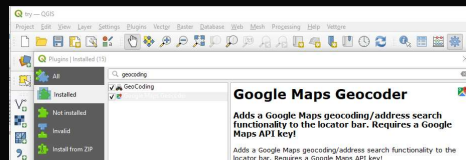
Google(maps) Geocoding API

Google Maps APIs		(Price per REQUEST)	
Web Services > Geocoding API		0-100,000	100,001-500,000
		0.005 USD per each (0.00 USD per 1000)	0.004 USD per each (4.00 USD per 1000)

The Google Maps Geocoding API is a service that provides geocoding and reverse geocoding of addresses.

Geocoding is the process of converting addresses (like a street address) into geographic coordinates (like latitude and longitude), which you can use to place markers on a map, or position the map.

Reverse geocoding is the process of converting geographic coordinates into a human-readable address.



Online geocoding services

- Eg. Batchgeo

Convalida e configura opzioni

Opzioni base

Regione:

Provincia:

Città:

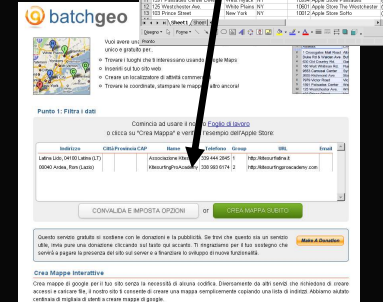
Stato:

Cap:

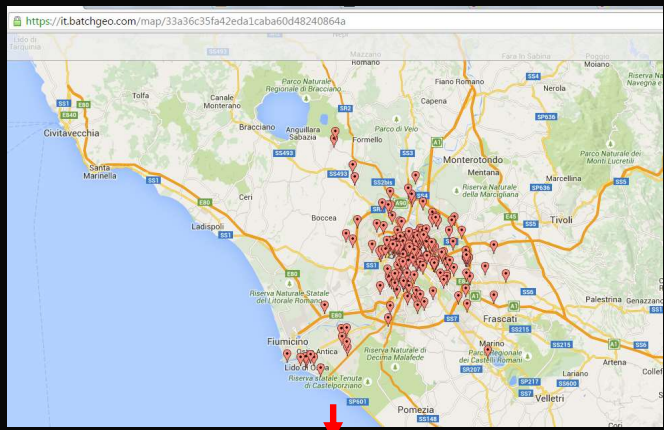
Categoria per (seleziona formato):

Anteprima casella indicatori

Convalida del file:
Nome:
Numero di telefono:
Gruppo:

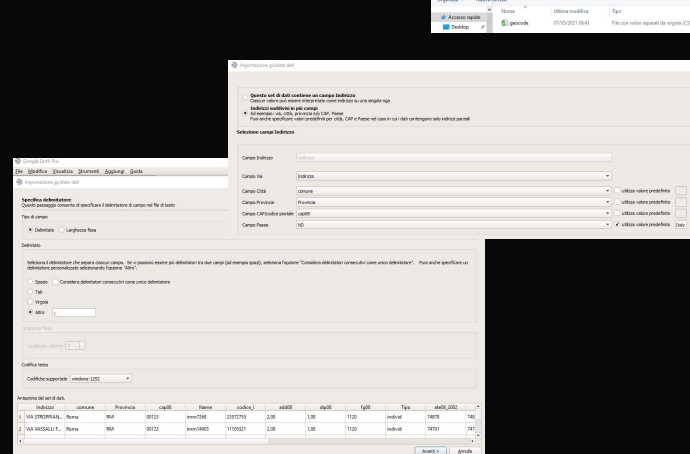


- Many others..

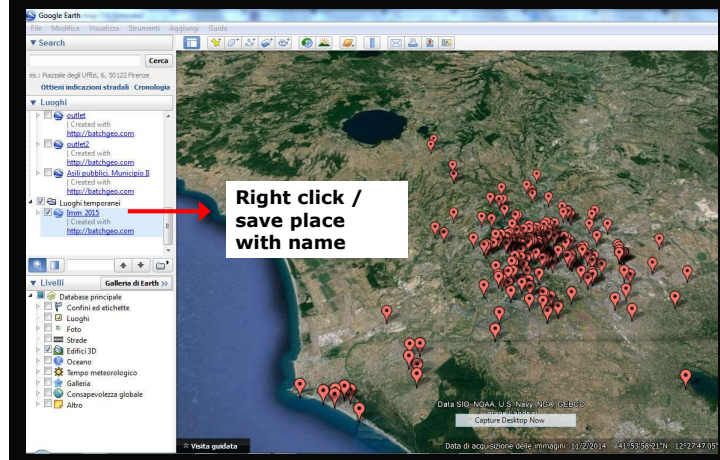


Output: map and kml files

Geocoding with Google Earth Pro

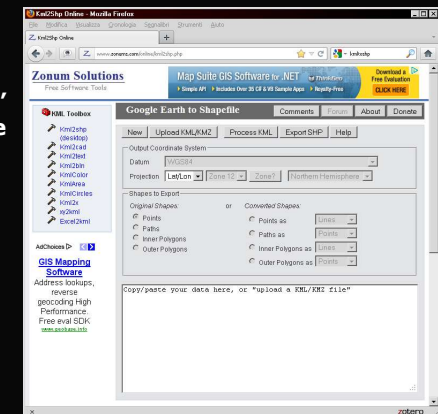


Export from google earth (save as .kmz or .kml)



Kml to shapefile conversion

Eg. Zonums' Kml2Shp online



Geocoding services :

Ex. <http://batchgeo.com>
("make maps" menu)

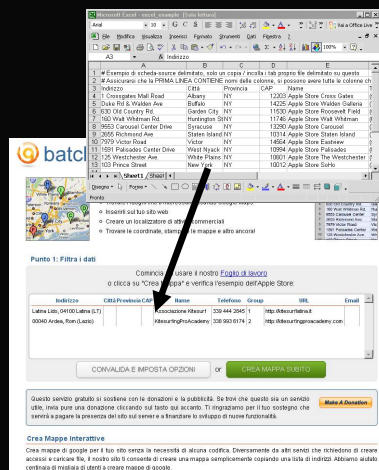
- Open in excel:
[spatial18/vv/geocode/
Roma_ricettivo_mar18.xls](http://spatial18/vv/geocode/Roma_ricettivo_mar18.xls)

- Copy and past in
batchgeo.com/make-maps

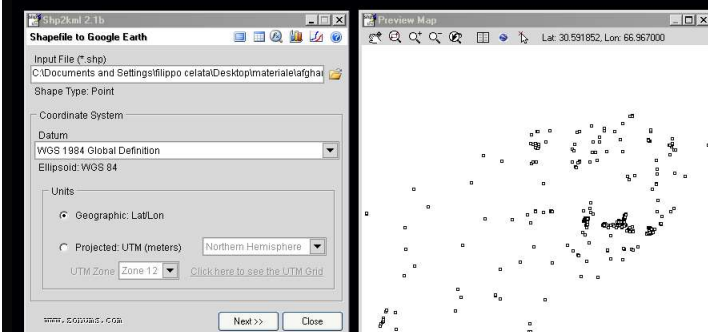
- Set the settings... (address,
city, province, postal code,
NAME) and run

- Save map (+data)

* Comune di Roma, hotels with
more than 200 accommodations,
March 2018



[From ArcGIS to google (shapefile to .kml)]



Eg. Zonums' "shp2kml" freeware

[Layers in geographic coordinate system WGS1984 (!)]

Geoextraction

Testi



Archivi



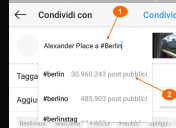
IP locator

Geolocation data from IP2Location (Product: DB6, updated on 2017-11-1)

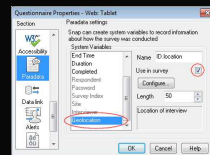
IP Address	Country	Region	City
193.100.78.200	Italy	Latvia	Rome

ISP	Organisation	Latitude	Longitude
Universita' degli Studi di Roma La Sapienza	Not Available	41.8847	12.4839

Geotag (X/Y o toponimi)



Metadata



Survey con geolocation

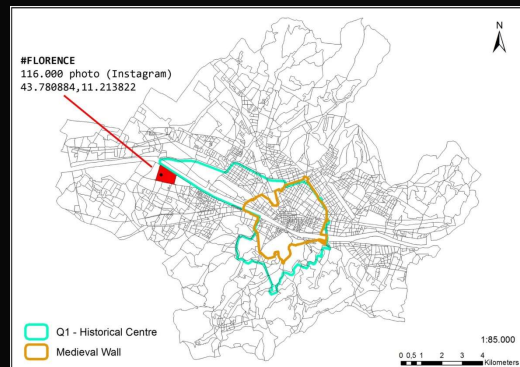
Geocoding methods accuracy

Geocoding methods (from the most to the least accurate):

- 1) ROOFTOP: precise geocode for which we have location information accurate down to street address precision.
- 2) RANGE_INTERPOLATED: approximation (usually on a road) interpolated between two precise points (such as intersections).
- 3) GEOMETRIC_CENTER: geometric center of a result such as a polyline (eg. a street) or polygon (region).
- 4) APPROXIMATE: the returned result is approximate.

Fails: "unmatched" o "unmapped"

Georeferencing: accuracy and checks



3. Table join: to associate an external table to a layer's attribute table, based on an identical field

FID #	Name	ID
190	Carson County	16027
256	Carson County	16029
268	Carson County	16031
225	Clerk County	16033
226	Clerk County	16035
227	Custer County	16037
228	Elmore County	16039

One to One

ID	% Population Change
16027	+3.9
16029	+0.5
16031	+0.9
16033	+3
16035	+0.5
16037	+0.5
16039	+3.2

ArcGIS

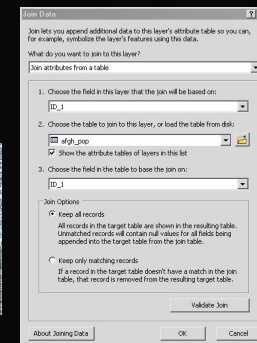
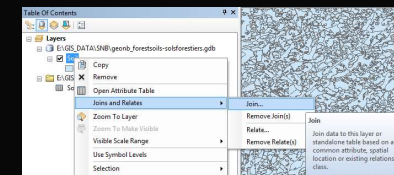


Table join: to associate an external table to a layer's attribute table, based on an identical field

FID	Name	ID
190	Canyon County	16027
256	Caribou County	16029
208	Carson County	16031
235	Clark County	16033
226	Clearwater County	16035
227	Custer County	16037
228	Doune County	16039

ID	% Population Change
16027	+3.3
16029	+0.5
16031	+0.9
16033	+0.3
16035	+0.5
16037	+0.5
16039	+3.2

QGIS

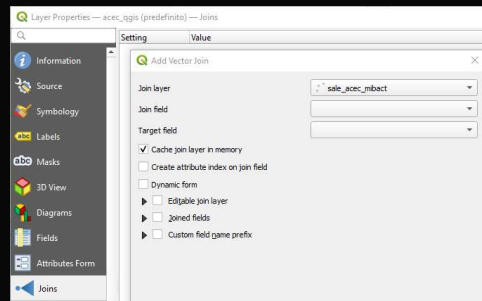


Table formats: CSV; DBF IV, .dbf; (Excel)

Spreadsheets vs. database files

Microsoft Excel - stranieri_cittadini_gombr_2015																	
File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ? Adobe PDF																	
[Icone di Excel] 100% Arial																	
A1			B C D E F G H I J K L M N O P														
A			B														
1			Popolazione straniera maschile e femminile														
2			Zona urbanistica - Maschi e Femmine														
3			Municipio I														
4			1A	1B	1C	1D	1E	1F	1G	1H	1I	1J	1K	1L	1M	1N	1O
5	ROMANIA	EUROPA COMUNITARIA	329	1.262	122	84	606	137	54	22	202	296	260	7	63	316	
6	PERUPNE	ASIA	897	293	273	70	588	190	68	24	429	506	314	19	16	1.369	
7	BANGLADESH	ASIA	68	3.181	4	21	1.660	205	20	2	65	183	362	4	10		
8	REP. POP. CINESE	ASIA	61	290	35	22	2.232	78	10	0	62	43	158	0	0		
9	UCRAINA	EUROPA NON COMUNITARIA	165	536	59	43	236	46	39	4	112	199	134	10	24		
10	PERU	AMERICA DEL SUD	209	157	41	22	186	49	21	15	84	132	115	7	9	282	
11	POLONIA	EUROPA COMUNITARIA	212	153	123	22	215	66	15	14	129	61	75	3	8	81	
12	EGITTO	AFRICA	207	217	12	9	82	13	0	2	32	27	98	1	4	23	
13	INDIA	ASIA	151	270	77	34	305	42	11	48	108	96	49	2	5	107	
14	ERI LANKA	ASIA	182	107	36	5	149	32	16	17	98	84	26	4	3	338	
15	MOLDOVA	EUROPA NON COMUNITARIA	66	140	12	9	43	28	7	1	20	54	33	1	7	57	
16	EQUADOR	AMERICA DEL SUD	109	72	46	11	110	17	6	4	36	81	54	1	7	147	
17	ALBANIA	EUROPA NON COMUNITARIA	35	42	0	3	72	22	5	1	20	16	26	0	1	16	
18	MAROCCO	AFRICA	48	83	4	2	58	2	1	4	10	7	5	3	4		
19	SPAGNA	EUROPA COMUNITARIA	192	92	41	22	120	39	15	10	86	31	27	25	4	206	
20	NGERIA	AFRICA	552	111	22	3	212	4	6	1	13	40	8	2	1	0	
21	FRANCIA	EUROPA COMUNITARIA	309	96	71	33	229	47	35	7	108	53	44	3	3	141	
22	BRASILE	AMERICA DEL SUD	98	75	23	14	123	30	12	3	43	40	35	1	3	47	
23	ERITREA	AFRICA	1.119	148	6	4	160	20	3	0	3	7	0	0	2	31	
24	COLOMBIA	AMERICA DEL SUD	128	40	11	6	87	13	10	4	27	25	10	1	1	19	
25	REGNO UNITO	EUROPA COMUNITARIA	152	90	51	26	110	25	14	6	46	31	35	6	2	48	
26	PAKISTAN	ASIA	646	59	11	0	56	5	3	0	5	6	12	0	1	5	
27	GERMANIA	EUROPA COMUNITARIA	146	91	48	15	95	43	19	4	45	35	26	2	0	61	
28	ETIOPIA	AFRICA	247	72	12	2	122	6	3	3	7	7	2	3	1	34	

Spreadsheets vs. database files

	A	B	C	D	E	F	G	H
1	CODE_ZURB	TOT_STATIS ROMANIA	FLIPINNE	BANGH	CHREF_POP	ROMANIA	PERUT	PERUT
2	1A	13567	329	897	51	165	209	
3	1B	9345	1282	293	3181	4	536	157
4	1C	1565	122	273	4	35	59	41
5	1D	694	84	70	21	22	43	22
6	1E	10250	608	568	1080	2232	236	188
7	1F	1618	157	190	205	78	46	49
8	1G	592	54	66	20	30	39	21
9	1H	239	22	24	2	0	4	15
10	1I	2266	202	429	65	62	112	84
11	1J	2402	288	506	183	43	199	132
12	1K	2227	260	314	362	68	134	135
13	1L	226	63	16	4	0	24	9
14	1M	4275	216	1366	10	12	170	262
15	1N	1158	128	264	32	13	79	60
16	1O	3348	940	728	27	84	179	181
17	1P	5850	795	1430	205	178	331	388
18	1Q	180	33	41	0	0	5	8
19	1R	138	12	45	2	0	6	3
20	1S	3657	458	686	85	86	216	193
21	1T	1214	114	138	288	73	34	18
22	1U	207	15	10	1	0	9	20
23	1V	15	2	0	0	1	1	0
24	1W	1866	284	407	128	46	97	109
25	1X	2618	380	511	80	140	177	143
26	1Y	2447	470	552	162	51	142	132
27	1Z	1988	395	638	206	29	56	74
28	1A	1272	246	145	32	41	89	60
29	1B	918	242	133	25	15	42	42
30	1C	1850	206	374	317	108	105	85
31	1D	1417	187	280	221	48	101	117
32	1E	1145	207	157	64	83	87	70
33	1F	303	107	57	27	2	9	7
34	1G	591	164	51	11	0	20	32

Field's headings:

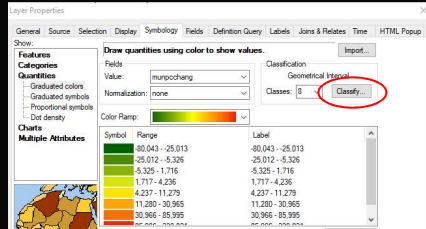
Max 10 digits

Only alphanumeric:
az, AZ, a1, _

No:
., ! ' " / & % etc.

Produce a choropleth map of the % variation and/or absolute difference of the population

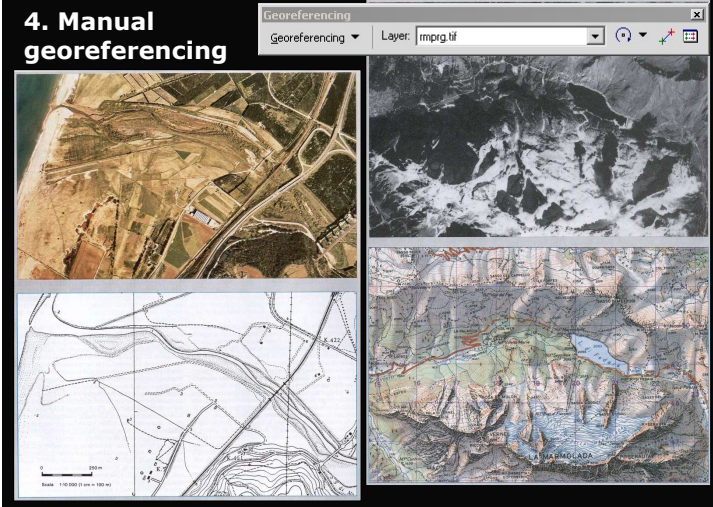
8) Right click on the layer, click on (the layer's) 'Properties and have fun with symbols, color ramps, classification methods, classes, ranges, etc..



9) In the main upper bar, click 'VIEW', **layout view** and prepare the layout + 'Insert' a **legend** (ignore the wizard) and customize it (right click/legend properties).

10) via the 'File' menu: export the map as jpg (**300 dpi**)

4. Manual georeferencing



Lab: **Produce a map of GDP growth per EU Region** (your turn...)

1. At <https://ec.europa.eu/eurostat/web/regions/data/database> (or search: "Eurostat nuts data"), download the Regional Economic indicators/"Real growth rate of regional gross value added (GVA) at basic prices by NUTS 2 regions - percentage change on previous year (nama_10r_2gvagr)", as a .xlsx spreadsheet, **with CODES**, instead of **REGIONS' NAMES** (in the **FORMAT** menu).

2. Open the file in excel, paste the GDP data in a new spreadsheet (using 'paste as values'), edit the table's heading and contents, save it as a .csv file, and add the .csv to ArcMap.

3. At <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts> (or search: "Eurostat nuts shapefile"), download the NUTS shapefile, and add it to ArcMap.

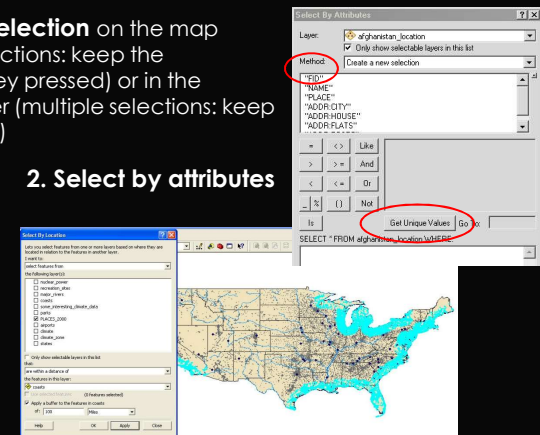
4. Associate GDP data to the NUTS layer attribute table, using a Table Join, keeping only matching values.

5. Produce a map of GDP growth per Region, 2019 data.

1. Manual selection on the map (multiple selections: keep the uppercase key pressed) or in the attribute layer (multiple selections: keep CTRL pressed)

2. Select by attributes

3. Select by location



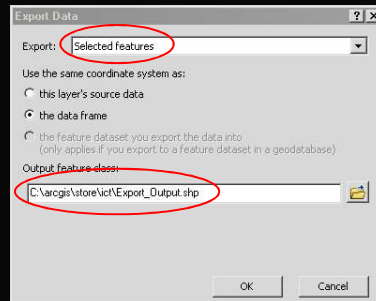
Deselect: right click on the layer, or in the Selection menu: "clear selected features"

-To create a layer or shapefile including only the selected feature

a. **Create new layer from selected features:** right click the selected layer / selection / create layer from selected features

b. Create a shapefile of the selected features: right click on the layer of selected features (1) / data / export data (as shapefile)

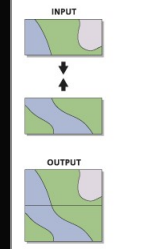
[c. To export as an autonomous layer, with symbology: right click / save as layer file]



Working with geometries: **geoprocessing**

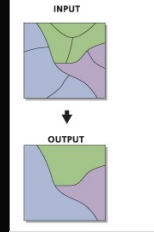
Merge

Combines input features from multiple input sources (of the same data type) into a single, new, output feature class. The input data sources may be point, line, or polygon feature classes or tables.



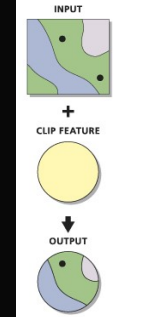
Dissolve

Aggregates features based on specified attributes.



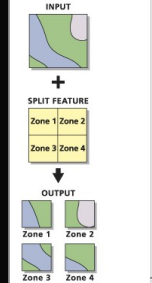
Clip

Extracts input features that overlay the clip features.



Split

Breaks the input Features into multiple output feature classes. The boundary of each unique value in the Split Field is used to split the input Features. The name of the output feature classes will be the same as the Split Field's unique values. Output feature classes are maintained in the target workspace.



Lab: selection and geoprocessing

a. In the layer with present population change per municipality, select all features where the % variation ('munpccchang') is <-25%, using Selection/**Select by attributes**

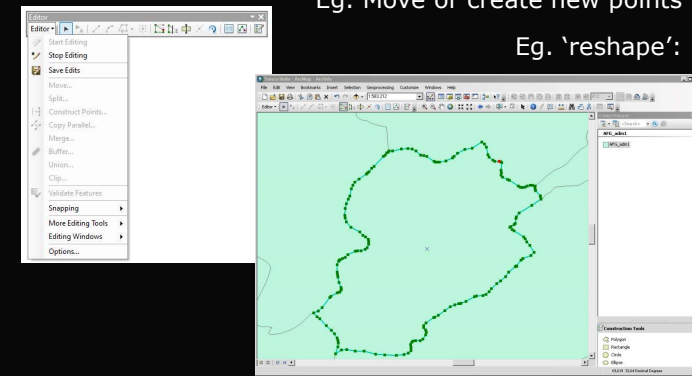
- Create a layer including only the selected municipalities (right click on the layer/selection/**create layer from selected features**, and use the layer symbology to highlight those municipalities in the map.

b. Using the layer with present population change per municipality, produce a new layer/shapefile, with the same data per Italian province: 'Geoprocessing' menu/Dissolve = set the input feature / select the field the dissolve will be based on (provinces' codes = 'COD_PROV') / select the statistic fields to be summarized (n_baseline, n_crisis, munpccchang) and the related statistic (SUM, AVERAGE)
- produce a map of present population change per province.

Working with geometries: **Editing**

Eg. Move or create new points

Eg. 'reshape':

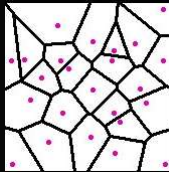


Geodata geometry conversion (points <-> polygons <-> lines)

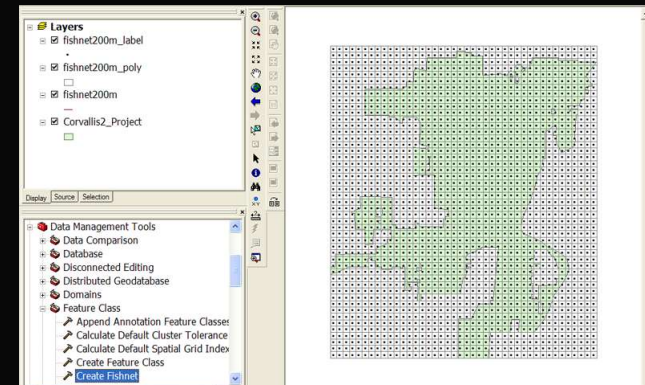
- To convert **polygons into points** (weighted/marked):
Calculate the X/Y centroids in the polygon layer's attribute table. Export the table and add it as a X/Y table

Lab: 1) open the municipalities layer attribute table. **2)** In the table properties, click 'add field', and create **two empty fields**: X and Y, format: double, precision: 20, scale: 10. **3)** Right click on the X field heading, select 'calculate geometry', and calculate the X centroid. Do the same for Y. **4)** via the table properties, export the table as **DBF**, and add it to the workspace using **add X/Y data**.

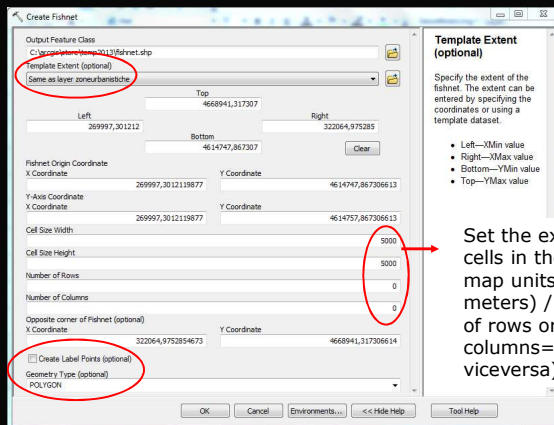
- To convert **points into polygons**: create **Voronoi Map** or **THIESSEN POLYGONS** [triangulated irregular network (TIN) that meets the Delaunay criterion]



FISHNET: to create a regular polygonal gridded shapefile



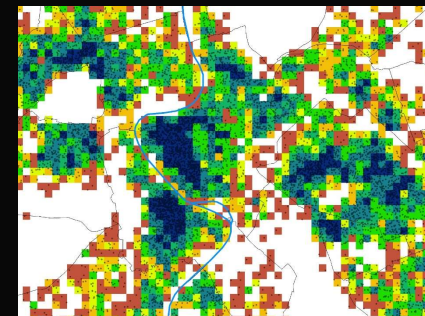
FISHNET (ArcGIS)



Set the extent of cells in the grid (in map units, eg. meters) / number of rows or columns= 0 (or viceversa)

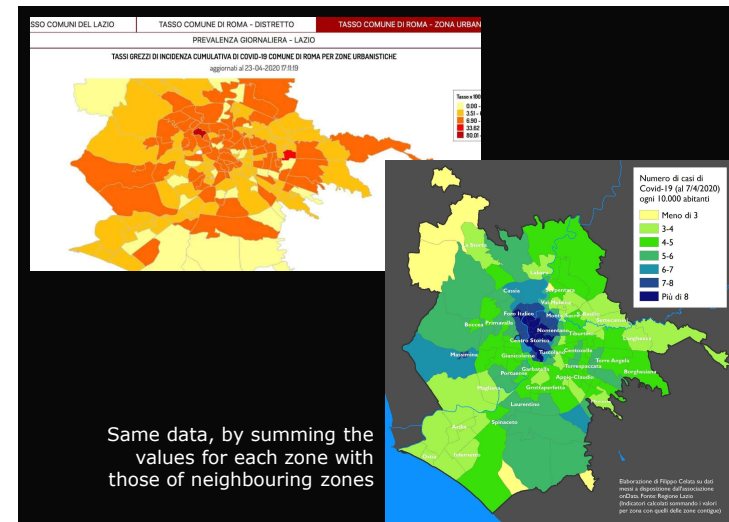
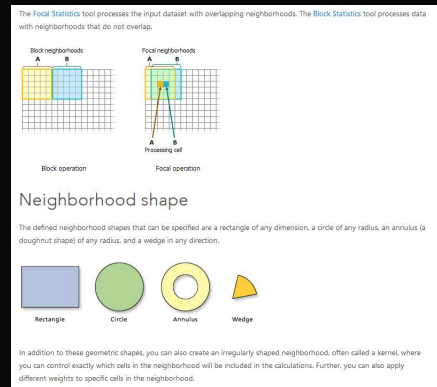
Working with rasters

Shapefile <-> Raster conversion tools: Arc toolbox/
Conversion tools/Point to raster (or raster to point)



Working with rasters

Neighbourhood statistics



Same data, by summing the values for each zone with those of neighbouring zones

1) In the layer with present population change per municipality, select all features where the % variation ('munpcchang') is <-25%, using Selection/Select by attributes. 1b) Create a layer including only the selected municipalities (right click on the layer/selection/create layer from selected features), and highlight those in the map using the layer symbology.

2) Using the layer with present population change per municipality, produce a new layer/shapefile, with the same data per Italian province: 'Geoprocessing' menu/Dissolve = set the input feature / select the field the dissolve will be based on (provinces' codes = 'COD_PROV') / select the statistic fields to be summarized (n_baseline, n_crisis, munpcchang) and the related statistic (SUM, AVERAGE). 2b) produce a map of present population change per province.

3) open the municipalities layer attribute table. 3b) In the table properties, click 'add field', and create two empty fields: X and Y, format: double, precision: 20, scale: 10. 3c) Right click on the X field heading, select 'calculate geometry', and calculate the X centroid. Do the same for Y. 3d) via the table properties, export the table as DBF, and add it to the workspace using add X/Y data. 3e) produce a map using the layer symbology

Map unemployment rates per Italian travel-to-work areas (SLL)

1) download Italian travel-to-work areas (SLL) from: https://www.istat.it/it/files/2019/12/SLL_2011_2018.zip, and add the .shp to ArcMap

2) download 2019 unemployment data from: <https://www.istat.it/it/archivio/248606> (click on 'TAVOLE').

3) open unemployment data in excel, edit the 'SLL 2019' spreadsheet in a new spreadsheet (paste as values), save it as a .csv table and add it to ArcMap (we only need the column "Tasso di disoccupazione").

4) join the csv table to the SLL shapefile, export the layer as 'SLL', and map 2019 unemployment rates.

5) repeat 3) for 2009 unemployment rates (spreadsheet 'SLL 2009'), join it to the 'SLL' shapefile, and export the layer as 'SLL_2009_2019'.

6) open the 'SLL_2009_2019' attribute table, create a new (double) field, and calculate the variation between unemployment rates of 2019 and 2009 using the 'field calculator'.

7) map the 2019-2009 difference in unemployment rates and compare it to the map produced in 4)