



**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

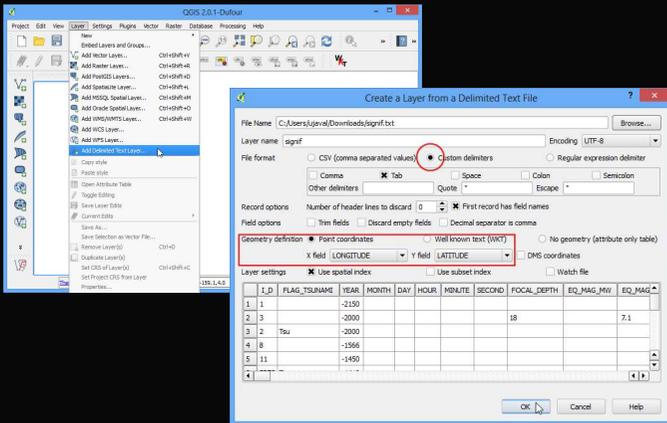
Manual georeferencing

## 1. Georeferencing X/Y tables: ArcGIS

The image shows two screenshots from the ArcGIS interface. The top screenshot displays the 'Add Data' menu with 'Add XY Data...' selected. The bottom screenshot shows the 'Add XY Data' dialog box with the following fields and options:

- Specify the fields for the X, Y and Z coordinates:**
  - X Field: [Dropdown]
  - Y Field: [Dropdown]
  - Z Field: [Dropdown]
- Coordinate System of Input Coordinates:**
  - Description: Unknown Coordinate System
- Show Details
- Warn me if the resulting layer will have restricted functionality
- Buttons: [About adding XY data](#), OK, Cancel

## 2. Georeferencing X/Y tables: QGIS



## X/Y coordinates formats

**DMS:** 75 59 32.483 W

**Decimal (DMS):** -7599.2356

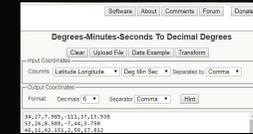
75 59.32483' West

-75.992356

75° 59' 32.483'' W

**-75.992356**

Tools and formulas for the conversion of coordinates:



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

Xwgs84	Ywgs84
12.5112771	41.891923

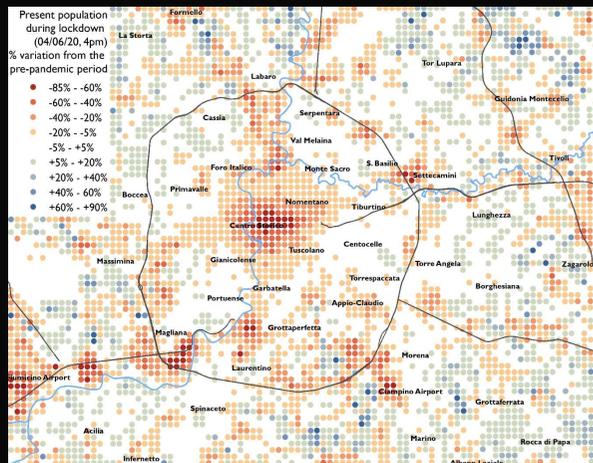
Xutm	Yutm
791306.7228	4643740.8315

**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

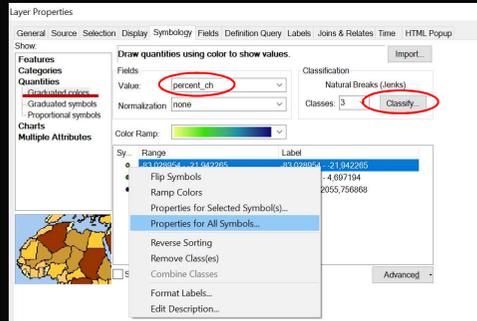
3) Right click on the table and click 'Display XY data': X Field = Xutm and Y Field = Yutm

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



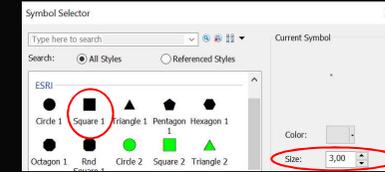
### Map the present population variation <-5% and >+5%

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

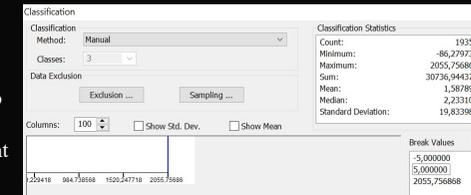


### Map the present population variation <-5% and >+5%

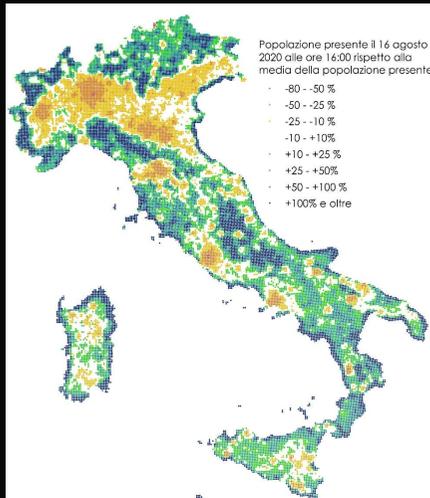
5) Right click on the default symbols and click 'properties for all symbols'



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

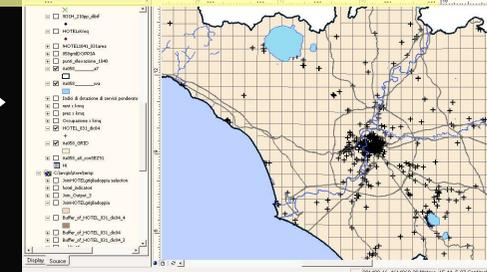


### Present population variation during the summer



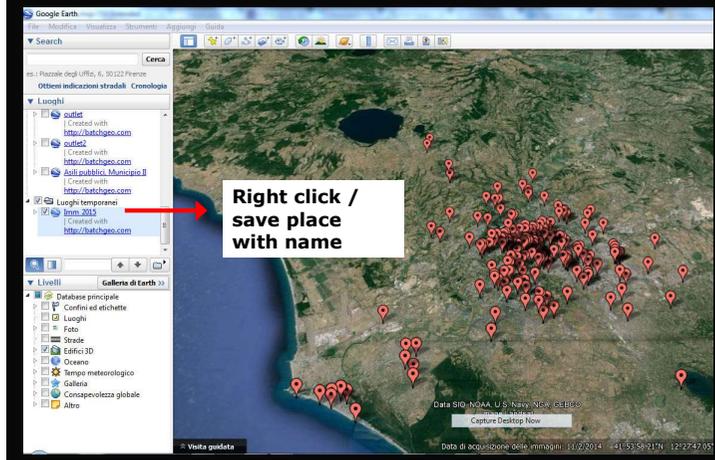
### 2. Geocoding services : georeferencing point data with addresses (or place names)

ID	HOTEL	INDIRIZZO	CAMERE	LETTI	addetti	MediaDoppia	...
40390353547738	Regina Hotel Baglioni	Via V. Veneto, 72	151	310	108,72	495,0	...
40390353547732	Marriott Grand Hotel Flora	Via V. Veneto, 191	156	360	57,72	415,0	...
40390353547731	Excelsior	Via V. Veneto, 125	327	616	235,44	812,3	...
40390353547723	Ambasciatori Palace	Via V. Veneto, 62	154	368	110,68	425,0	...
40390353526305	De Russie	Via del Babuino, 9	130	370	93,6	819,0	...
40390353525546	Sheraton Roma	V. le del Pattinaggio, 100	680	1360	251,6	570,5	...
40390353525545	Sheraton Golf	V.ile Parco de' Medici, 167	543	1086	100,64	388,0	...
40390353525522	Plaza	V. del Corso, 128	200	400	54,7	598,0	...
40390353525516	Parco Dei Principi	Via G. Frescobaldi, 5	198	364	142,56	580,0	...



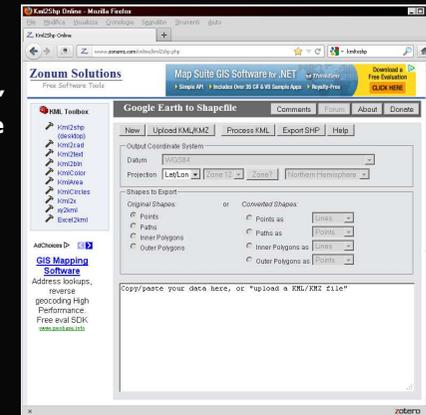


## Export from google earth (save as .kmz or .kml)



## Kml to shapefile conversion

Eg. Zonums' Kml2Shp online



## Geocoding services :

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

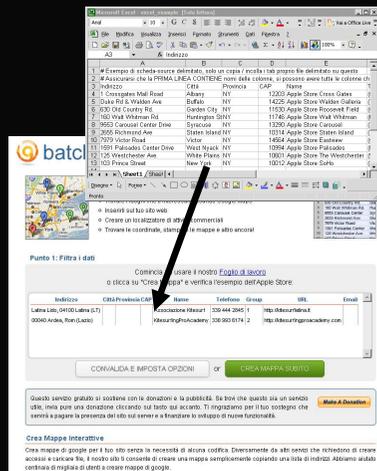
- Open in excel: [spatial18/vv/geocode/Roma\\_ricettivo\\_mar18.xls](http://spatial18/vv/geocode/Roma_ricettivo_mar18.xls)

- Copy and past in [batchgeo.com/make-maps](http://batchgeo.com/make-maps)

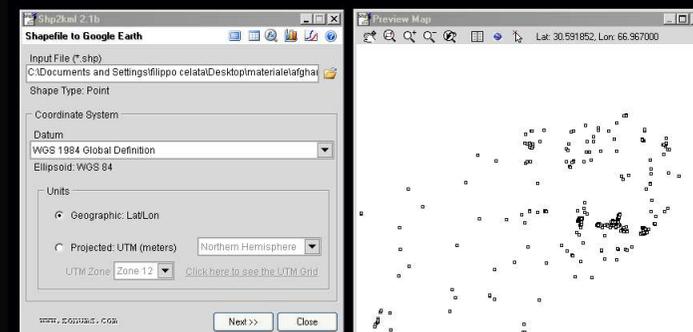
- Set the settings... (address, city, province, postal code, NAME) and run

- 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**

Die besten Clubs gibt es in **Berlin**, in anderen Gegenden muss man nochmal eine Hotline anrufen, z.B. 069-669257. Und in den ruhigen Städten **München** und **Freising** geht man einfach in einen Biergarten.

**Archivi**

3 CARATTERI ALFABETICI DEL COGNOME  
3 CARATTERI ALFABETICI DEL NOME  
2 CARATTERI NUMERICI ANNO NASCITA

4 CARATTERI UNO ALFABETICO E 3 NUMERICI PER COMUNE NASCITA  
3 CARATTERI ALFABETICI DI CONTROLO  
DUE CARATTERI NUMERICI PER GIOCO NASCITA E SESSO  
UN CARATTERE ALFABETICO PER MESE NASCITA

**IP locator**

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

IP Address	Country	Region	City
155.100.78.200	Italy	Lazio	Rome

**IP**

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

**Geotag (X/Y o toponimi)**

Condividi con Alexander Piacca a #Berlin

Tagga #Berlin 30.160.243 post pubblici

Aggiungi #Berlin 485.903 post pubblici

**Metadata**

QuantumGIS Properties - Web - Tablet

Section: Pirella Götting

Use an aerial picture visible to record information about how the survey was conducted.

System variables:

End Time: 08:00:00  
Duration: 00:00:00  
Planned: 00:00:00  
Planned: 00:00:00  
Survey Index: 00:00:00  
Date: 2017-11-11 11:21:53  
Location of receiver: 41.8847, 12.4839

**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

#FLORENCE  
116.000 photo (Instagram)  
43.780884, 11.213822

Q1 - Historical Centre  
Medieval Wall

0 0.5 1 2 3 4 Kilometers

1:85,000

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

US Counties

FIPS *	Name	ID	% Population Change
190	Calvin County	16027	+3.9
206	Carroll County	16029	+6.5
208	Cassie County	16031	+0.9
225	Clerk County	16033	+3
228	Cherokee County	16035	+0.5
227	Custer County	16037	+0.5
228	Elmore County	16039	+3.2

Population Data

ID	% Population Change
16027	+3.9
16029	+6.5
16031	+0.9
16033	+3
16035	+0.5
16037	+0.5
16039	+3.2

One to One

ArcGIS

Table of Contents

Layers

ESRI\_GIS\_DATA\SNH\geonir\_forestools-solforestiers.gdb

Copy

Remove

Open Attribute Table

Join and Relates

Join...

Remove Join(s)

Relate...

Remove Relate(s)

Join data to this layer or standalone table based on a common attribute, spatial location or existing relationship files.

Join Options

Keep all records

All records in the target table are shown in the resulting table. Unmatched records will contain null values for all fields being appended into the target table from the join table.

Keep only matching records

If a record in the target table doesn't have a match in the join table, that record is removed from the resulting target table.

Validate Join

About Joining Data

OK

Cancel

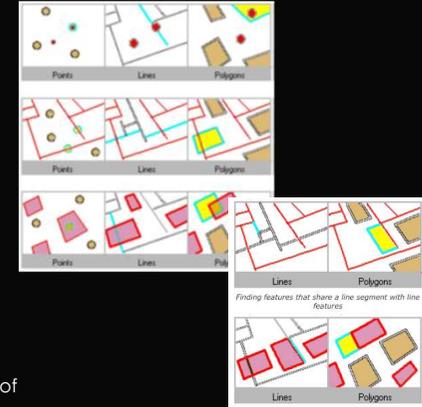


**Lab:** variation of the **'present' population** during lockdown, per municipality

1. Attribute to each facebook tile the name and code of its municipality, using the **'spatial join'** tool\*
2. Summarize the output table, per municipality, using table's **'summarize'** tool
3. Re-calculate the variation of the present population in a new table's field, using the **'field calculator'**
4. Join the table to the layer of Italian municipalities, using the **'table join'** tool
5. Produce a choropleth map of the variation of the present population, using the layer's **symbology**

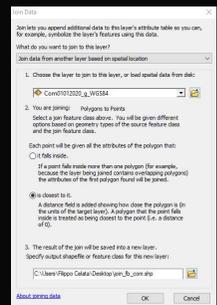
**Spatial join:** to associate to the attribute table of a layer, the content of another layer's attribute table, based on a **spatial relationship**:

- Intersect
- Are within a distance of
- Are within
- Are completely within
- Contain
- Completely contain
- Have their centroid in
- Share a line segment with
- Touch the boundary of
- Are identical to
- Are crossed by the outline of



### Convert facebook files data into municipal data

- 1) To attribute to each facebook tile the corresponding municipality: Right click on 'fb\_pop\_23mar20 events' and click on Join and relates/Join and .....
- 2) Open the output layer's attribute table and right click on the 'PROCOM' field to «SUMMARIZE» the SUM of 'n\_baseline', 'n\_crisis' fields, per municipality. Add the table to the workspace.



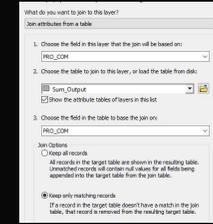
FIG_1	COD_RIP	COD_REG	COD_PROV	COD_CM	COD_UTI	PROCOM	COMUNE	COMUNE_A	CO_UTI	SHAPE_AREA	SHAPE_LEN	per	Distance
33	1	1	2	0	3	1	25027013201	25027013201	0	0	0	0	0
39	3	10	55	0	55	1	2107245605607	24492737069	0	0	0	0	0
59	1	1	2	0	2	1	404846405801	07195302440	0	0	0	0	0
63	1	1	2	0	2	1	2019070504907	26016088002	0	0	0	0	0
65	1	1	2	0	2	1	2380471779011	24272027212	0	0	0	0	0
66	5	10	85	0	85	1	2102010820303	32121266845	0	0	0	0	0
67	1	10	85	0	85	1	2005000727005	46270101860	0	0	0	0	0
68	1	10	85	0	85	1	2102010820303	32121266845	0	0	0	0	0
69	3	10	55	0	55	1	2106020512506	23897252322	0	0	0	0	0
100	3	10	56	0	56	1	2106020512506	23897252322	0	0	0	0	0
111	3	10	56	0	56	1	2106020512506	23897252322	0	0	0	0	0
112	3	10	56	0	56	1	2106020512506	23897252322	0	0	0	0	0
159	1	1	3	0	3	1	1032038702711	14054415227	0	0	0	0	0
162	1	1	3	0	3	1	1032038702711	14054415227	0	0	0	0	0
210	5	10	86	0	86	1	1620106100909	38216090605	0	0	0	0	0
213	5	10	86	0	86	1	1620106100909	38216090605	0	0	0	0	0
260	1	1	1	201	201	1	4000010100001	01664832015	0	0	0	0	0
261	1	1	1	201	201	1	4000010100001	01664832015	0	0	0	0	0

### Join the municipal summary table to the shapefile of Italian municipalities

- 3) Right click on the 'Com01012020\_g\_WGS84' layer, click Join and Relates/Join and join the 2 tables based on the 'PROCOM' field
- 4) Export the 'Com...' layer to consolidate: right click on the layer and click 'export data'

### Calculate the percentage variation of the population in each municipality

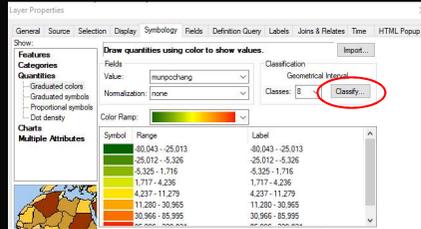
- 5) Open the last output layer's attribute table; click on the 'table options' tab and click 'Add field'
- 6) Through the 'Add field' menu, create a 'Double' field named 'munpccchang' with 'precision' = 10, 'scale' = 3
- 7) Right click on the new field, 'Field calculator' and calculate the % change:  $([Sum\_n\_cris] - [Sum\_n\_base]) / [Sum\_n\_base] * 100$



FIG_1	Shape*	COD_RIP	COD_REG	C
1	Polygon	1	1	1
2	Polygon	1	1	1
3	Polygon	1	1	1
4	Polygon	1	1	1
5	Polygon	1	1	1
6	Polygon	1	1	1

## 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.

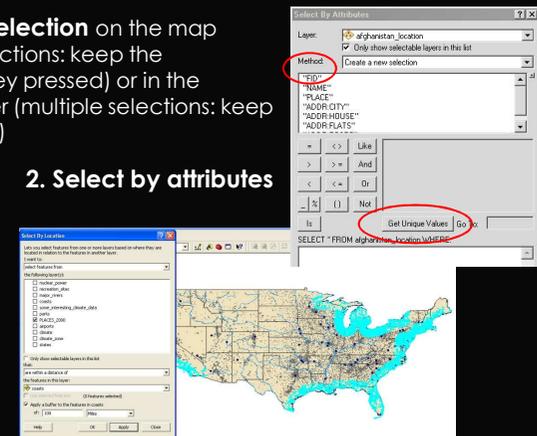
5. Produce a map of GDP growth per Region, 2019 data.

**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"

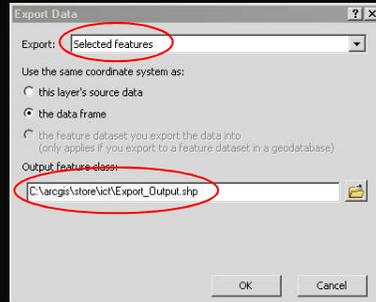


-To create a layer or shapefile including only the selected feature

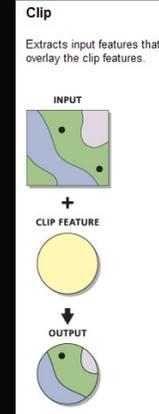
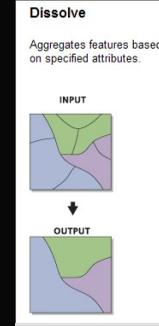
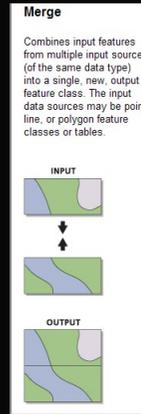
a. **Create new layer from selected features:** right click the selected layer / selection / create layer from selected features

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: **geprocessing**



## Lab: selection and geoprocessing

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

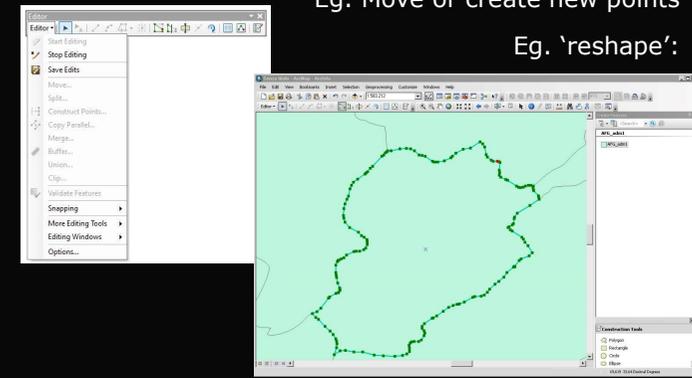
- Create a layer including only the selected municipalities (right click on the layer/selection/create layer from selected features, and use the layer symbology to highlight those municipalities in the map.

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, munpchang) 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':

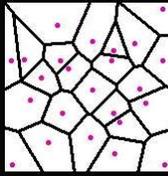


## Geodata geometry conversion (points <-> polygons <-> lines)

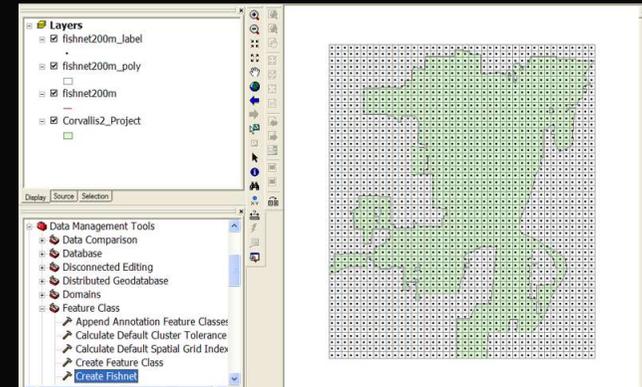
- To convert **polygons into points** (weighted/marked):  
Calculate the X/Y centroids in the polygon layer's attribute table. Export the table and add it as a X/Y table

**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**.

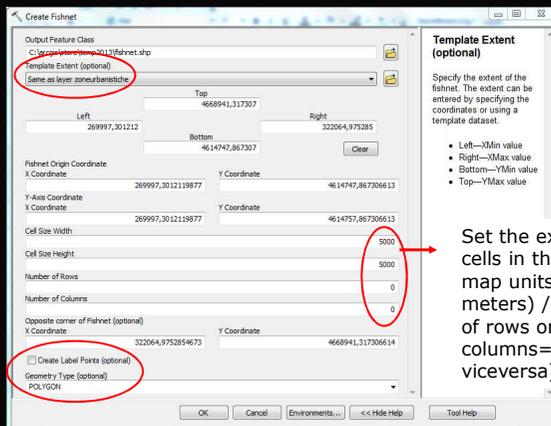
- To convert **points into polygons**: create **Voronoi Map** or **THIESSEN POLYGONS** [triangulated irregular network (TIN) that meets the Delaunay criterion]



FISHNET: to create a regular polygonal gridded shapefile



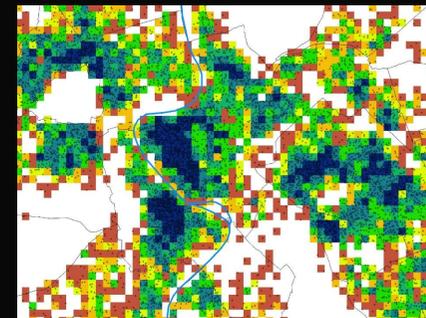
## FISHNET (ArcGIS)



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

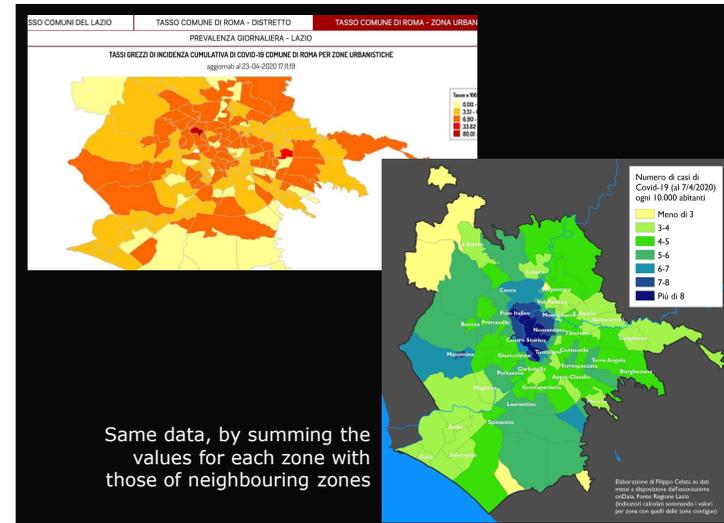
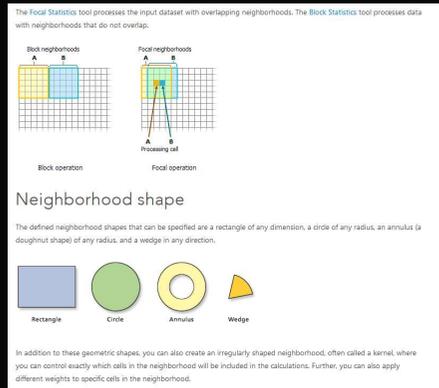
## Working with rasters

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



## Working with rasters

### Neighbourhood statistics



1) In the layer with present population change per municipality, select all features where the % variation ('munpchang') 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/shapfile, 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, munpchang) 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](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 shapfile, 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' shapfile, 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)