



Ufficio Statistico  
Regione Puglia

# FOCUS *Aprile 2014*

## “L’uso dei GIS nella statistica descrittiva e esplorativa: un esempio di caratterizzazione economica e socio-demografica dei territori comunali”

### Introduzione

Le analisi geografiche rivestono un ruolo importante nella descrizione e nell’interpretazione di fenomeni socio-economici che, per loro natura, si realizzano a scala territoriale. L’Ufficio Statistico della Regione Puglia utilizza, tra le sue dotazioni tecnologiche, uno strumento di analisi territoriale di natura *open source*. Si tratta di *QGIS 2.2.0* che richiama, nel nome del programma, il termine inglese con cui s’identifica questo tipo di tecnologia, GIS (*Geographic Information System*). Il *focus* di questo mese ha l’obiettivo di presentare l’uso del GIS, prendendo come riferimento alcuni indicatori economici e socio-demografici, recentemente pubblicati, e descrivendoli a scala comunale. I risultati di alcune analisi esplorative di tipo multivariato si basano sui GIS. La classificazione ascendente gerarchica degli indicatori permette, ad esempio, di raggruppare i comuni sulla base della maggiore o minore vicinanza dell’insieme di valori, secondo una prescelta misura di distanza multidimensionale. Questo tipo di analisi trova il supporto dei GIS, così come si mostra nel corso di questo lavoro.

### Fonti:

Osservatorio regionale Banche-Imprese; Ministero Economia e Finanze – Dipartimento delle Finanze; Automobile Club d’Italia (ACI); Istat, datawarehouse *I.Stat*.

### Cosa sono i GIS

I GIS permettono di gestire i dati geografici, cioè i valori e gli attributi associati a punti localizzati geograficamente sul territorio. Tali dati si possono presentare in diversa

forma (mappe, tabelle, immagini) e scala. All’interno di una mappa si possono riconoscere delle forme geometriche che assumono un significato, sulla base del contenuto della mappa stessa. Ad esempio, poligoni chiusi possono rappresentare confini amministrativi comunali, aree industriali; linee, tipologie di strade (provinciali, comunali, autostrade). Le informazioni geografiche, oltre che mappe, possono essere delle tabelle di misure effettuate da strumenti localizzati su un territorio; possono essere anche delle immagini di un territorio, riprese da sensori montati su aereo o su satellite.

In un GIS, i dati devono essere referenziati all’interno di un sistema di riferimento geografico (es. *Gauss-Boaga*, UTM), attraverso procedure *ad hoc* che dipendono dal formato d’origine dei dati (mappa, tabella, immagine). Dopo la georeferenziazione, i dati sono memorizzati nei GIS attraverso due possibili forme: vettoriale e *raster*. La prima descrive principalmente le mappe e memorizza l’insieme di coordinate della spezzata che meglio approssima le poligonali, aperte o chiuse, in essa contenute: più fitta sarà la spezzata, più accurata sarà l’approssimazione. Per dati *raster* s’intende l’insieme di dati organizzati in forma di griglia regolare, in cui ad ogni elemento o cella (assimilabile ad un *pixel*) viene assegnato un valore. I dati tabellari possono diventare di tipo *raster*, attraverso procedure di interpolazione; quelli di tipo *raster* possono diventare *vettoriali* attraverso procedure di *contouring lines*. Le immagini, per loro natura, sono di tipo *raster*. La forma vettoriale, all’interno di un GIS, è strutturata, in modo differente, da sistema a sistema. Esiste uno standard *de facto*, utilizzato anche dall’ISTAT. E’ lo *shapefile*, costituito da tre distinti file: 1) quello principale, con estensione SHP, che contiene i record di lunghezza variabile, ognuno dei quali descrive una geometria, con una lista di vertici; 2) una tabella DBF associata al file SHP, che contiene i record di attributi per ogni geometria; 3) un *index* file, con estensione SHX, costituito da record che puntano a quelli corrispondenti del file SHP, per velocizzare le operazioni sulle geometrie.

I sistemi GIS offrono la possibilità di realizzare diverse tipologie di analisi spaziale, fra le quali, ad esempio, la sovrapposizione (*overlay*) di mappe (*layer*), che geometricamente significa la determinazione di a) punti su poligoni, b) linee su poligoni, c) poligoni su poligoni. Se, ad esempio, si pone come primo livello poligonale, la mappa dei confini amministrativi comunali, mediante le funzioni di *overlay* con un’altra mappa si può attribuire, a ciascun comune, uno o più tralicci elettrici (punto su

poligono), un tratto di strada (linea su poligono), un edificio o un'area industriale (poligono su poligono).

### **Cosa è il software open source**

Il *software* Open Source (Sorgente Aperto) identifica la categoria di programmi che sono distribuiti con una particolare licenza d'uso. Quest'ultima prevede una condizione molto forte: il *software* deve includere il codice sorgente. E' resa, così, la possibilità di modificare il programma, consentendone, poi, la distribuzione, sotto gli stessi termini di licenza del *software* originale. Ciascun *software* Open Source fa riferimento, quindi, a un progetto, al quale chiunque ne fosse in grado, può fornire contributi.

### **Gli indicatori per la caratterizzazione del territorio**

A titolo esemplificativo dell'operatività dei GIS nell'interpretazione dei fenomeni, principalmente per l'aspetto legato alla rappresentazione dei dati geografici, si analizzano, a scala comunale, sette variabili che forniscono un contributo per caratterizzare i territori, dal punto di vista economico e socio-demografico, in particolare: 1) Valore Aggiunto (VA) Pro-Capite; 2) Reddito Imponibile IRPEF medio, per comune; 3) Densità demografica; 4) Quoziente di natalità; 5) Quoziente di mortalità; 6) Popolazione straniera residente; 7) Diffusione delle auto ecologiche.

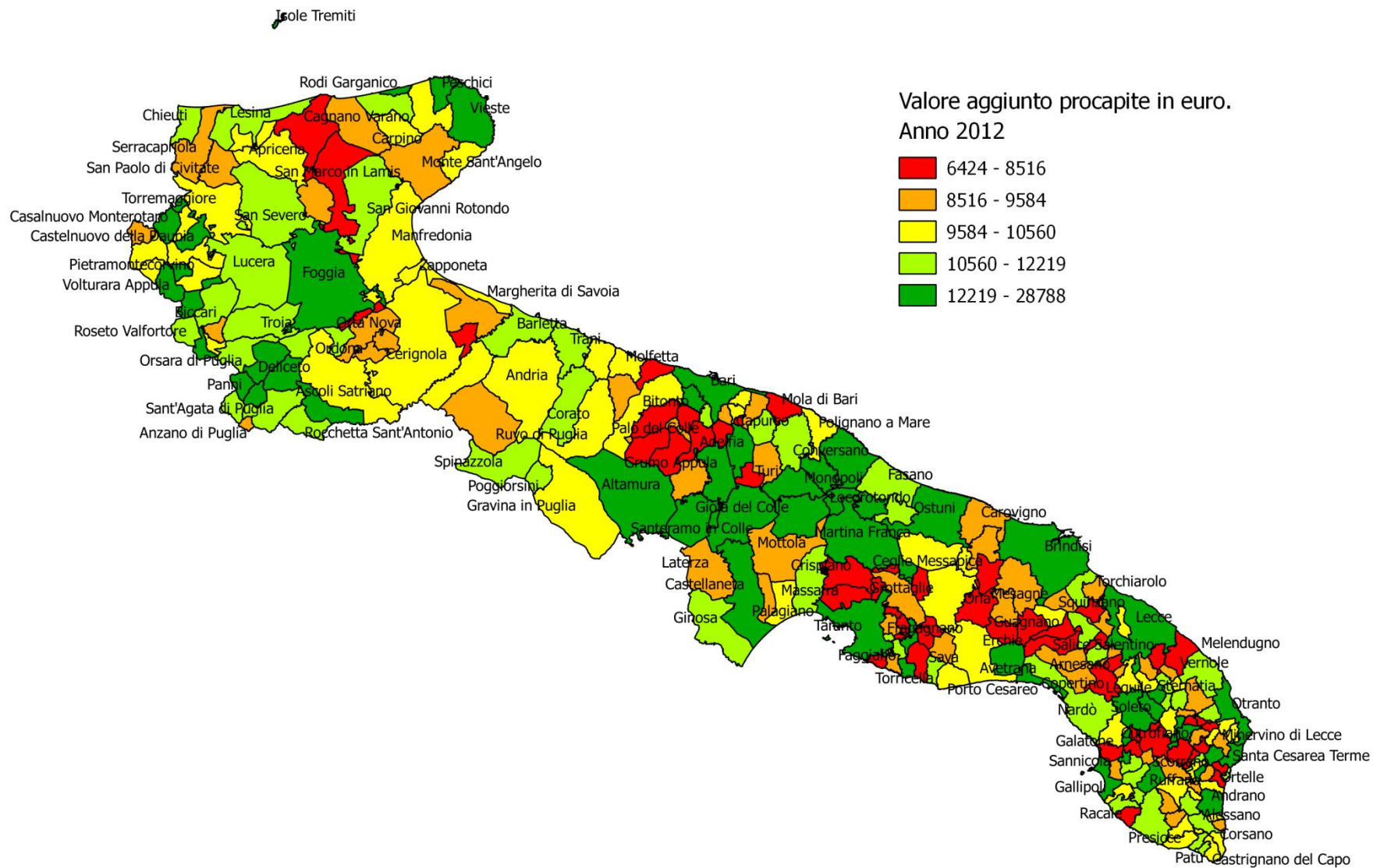
#### *Indicatori economici*

##### **Valore Aggiunto Pro-Capite**

La ricchezza prodotta all'interno dei territori comunali della Puglia, può essere descritta attraverso il valore aggiunto procapite, cioè *"la misura dell'incremento di valore che si verifica nell'ambito della produzione e distribuzione di beni e servizi finali, grazie all'intervento dei fattori produttivi (capitale e lavoro), a partire da beni e risorse primarie iniziali"*. La figura 1 rappresenta la distribuzione geografica di una stima del valore aggiunto pro-capite, per l'anno 2012. L'analisi della mappa evidenzia alcuni comuni, appartenenti alla classe più alta, di colore verde, con valori maggiori di 20.000 euro: si tratta delle Isole Tremiti (28.788 euro), Modugno (25.985 euro), Lecce (25.460 euro), mentre Bari ne è un po' al disotto (19.645 euro). Di contro, nella fascia più bassa, di colore rosso, con valori al di sotto dei 7.000

euro, si trovano: Carapelle (6.972 euro), Palmariggi (6.968 euro), Bitetto (6.747 euro), Lizzanello (6.717 euro), Carosino (6.709 euro), Statte (6.605 euro), Grumo Appula (6.424 euro). In alcuni territori si passa, in modo brusco, dalla classe a VA pro-capite più alto (verde) a quello più basso (rosso). E' il caso della Terra di Bari, in direzione Gravina di Puglia, di parte del Salento, del Tavoliere e dell'arco ionico tarantino. Il subappennino dauno, nella sua totalità, mostra una variazione più smussata, più attenuata. Lungo la costa a sud di Bari, fatta eccezione per Mola, c'è una significativa concentrazione di comuni appartenenti a classi medio alte.

Fig. 1 – Distribuzione comunale del valore aggiunto procapite, in euro. Anno 2012. Puglia.

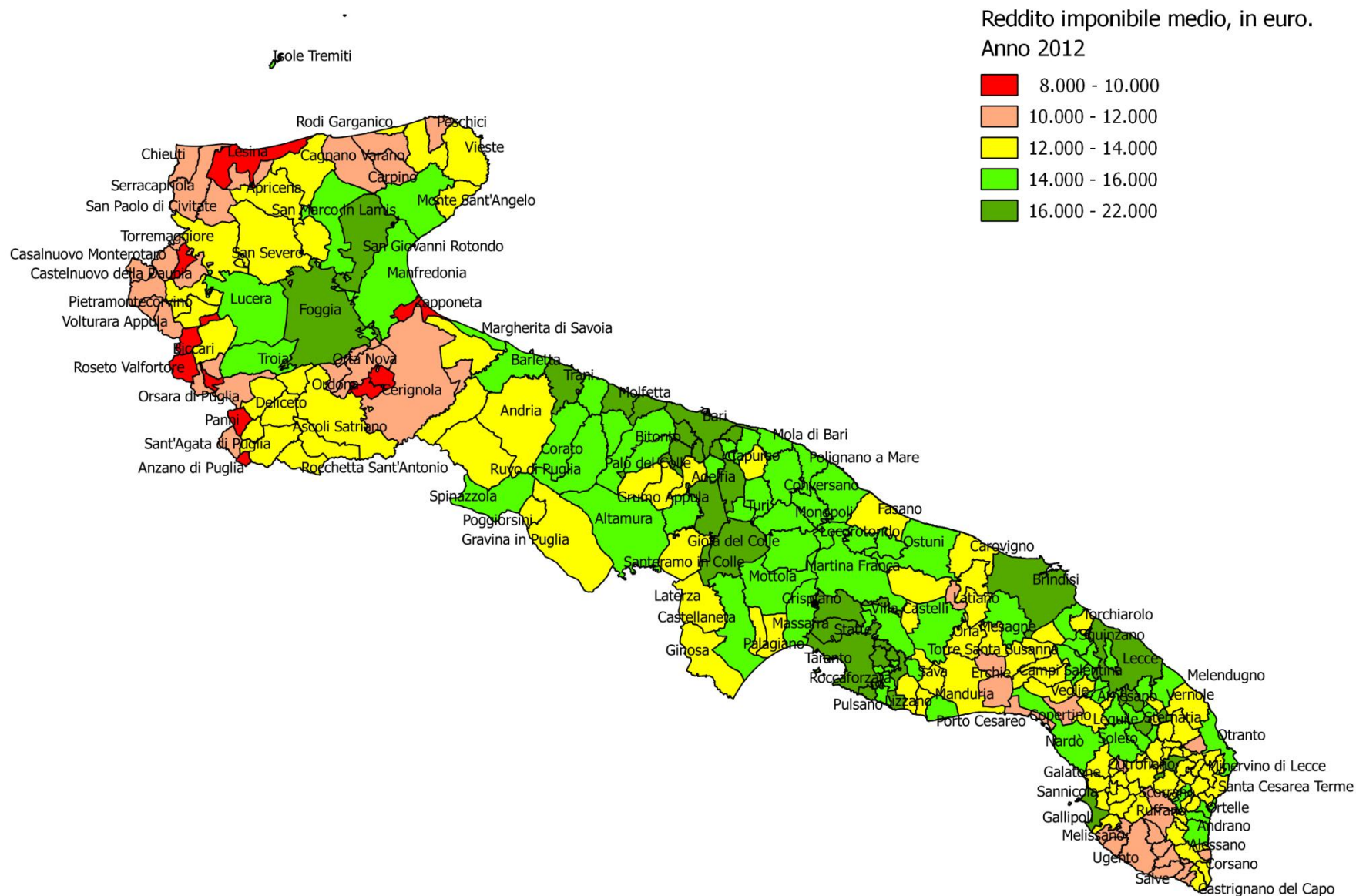


### **Reddito Imponibile IRPEF medio per comune**

Nella fig. 2 si descrive la distribuzione media comunale del reddito imponibile dei contribuenti residenti, cioè l'IRPEF delle persone fisiche per l'anno 2012. I valori più elevati si registrano per i comuni capoluogo di provincia e per i comuni ad essi più prossimi. La mappa si presenta con un'evidente prevalenza di comuni, della costa adriatica, con un alto reddito imponibile IRPEF medio. Di contro, tutto il SubAppennino dauno, mostra classi basse. Il comune con IRPEF media più alta è Lecce (21.430 euro), seguito da Bari di poco sotto (21.096 euro), infine Taranto e Leporano con valori di poco superiori ai 19.000 euro. Stornara e Roseto Valfortore sono il fanalino di coda, con IRPEF media sotto i 9.000 euro. Il passaggio più brusco da una classe a imponibile alto a quella a imponibile basso si ha solo fra il comune di Foggia (verde scuro) e quelli di Ortanova e di Ortona; è più "continuo" negli altri.

Il reddito imponibile medio IRPEF, rappresenta un valore medio relativo a una variabile che può essere più o meno "concentrata", su ciascun territorio. Volendo approfondire questo aspetto si tratta di utilizzare statistiche sulla variabilità e sulla concentrazione.

Fig. 2 – Distribuzione comunale del reddito imponibile medio per contribuente, in euro. Anno 2012. Puglia.



## *Indicatori socio-demografici*

### **Densità Demografica (ab/kmq)**

La densità della popolazione è principalmente utilizzata per valutare l'impatto antropico sull'ambiente. Dipende dalle caratteristiche geofisiche della zona, ad esempio quella di alta montagna, che può includere o meno aree non abitabili, e dalla pressione antropica (aree urbane e rurali). La densità demografica è calcolata come rapporto fra popolazione residente nel comune e relativa superficie comunale (fig. 3). Un'area ad alta densità si ha sulla costa a nord di Bari (Barletta, Trani, Bisceglie, Molfetta); il comune di Bari è quello a più alta densità con 2.714 abitanti per chilometro quadro, segue Triggiano (1.350 ab/kmq). Zone a bassa densità sono concentrate particolarmente nel subappennino dauno. I due comuni che sono al di sotto dei 10 abitanti per chilometro quadrato sono Volturara Appula e Celle di San Vito.





### **Quoziente natalità**

Il tasso di natalità è definito come il rapporto tra numero delle nascite, in un determinato territorio e periodo di tempo, e la popolazione media residente dello stesso periodo e dello stesso territorio. Su di esso incidono diversi fattori, fra i quali, lo sviluppo economico, il grado di modernizzazione del territorio ma soprattutto la struttura per età e la composizione per sesso di una popolazione. A tassi di natalità più elevati corrisponde una popolazione strutturalmente più giovane.

La distribuzione comunale di tale quoziente è in fig. 4. E' basso nel subappennino dauno, passando repentinamente dalla classe più bassa a quella più alta, in comuni dell'area del Tavoliere. Hanno quoziente di natalità molto vicino a zero i seguenti comuni: Volturara Appula, Celle di San Vito, Panni, Orsara di Puglia, Motta Montecorvino, Nociglia, Castro. Natalità alta e superiore a 4 si registra a Stornara (4), Faggiano (4,2), Arnesano (4,3) e Isole Tremiti (4,4).





### **Quoziente mortalità**

E' il rapporto tra il numero di morti registrati in un comune e l'ammontare della popolazione, ivi residente (per 1.000). Risulta più elevato nei territori con popolazioni più anziane e a bassa natalità. E' influenzato da fattori ambientali e dalle condizioni di vita. La fig. 5 descrive la distribuzione comunale di tale quoziente. E' alta in alcuni comuni del subappennino dauno, si mantiene in classi medio-basse lungo la costa adriatica. Il coefficiente di correlazione di *Pearson*, fra il quoziente di natalità e quello di mortalità, rileva un'assenza di relazione lineare fra i due quozienti. Si esclude l'ipotesi per cui i luoghi in cui si nasce in misura minore sono, verosimilmente, gli stessi in cui si muore in misura minore o maggiore. I comuni di Binetto e le Isole Tremiti sono i comuni col quoziente più basso, inferiore a 0,01, quelli con popolazione presumibilmente più anziana e, quindi, con quozienti superiori a 5 sono Pietramontecorvino (5,8), Casalnuovo Monterotaro (6,6), Castelnuovo della Daunia (7,8), Volturino (8,4), Panni (9,4), Celle di San Vito (11,6). Buona parte dei comuni in Terra di Bari appartengono a classi medio-basse di mortalità. La zona sud del Subappennino dauno e del Salento hanno significative concentrazioni di classi più elevate di mortalità.



### **Popolazione straniera residente**

L'indicatore è calcolato come rapporto fra il numero di stranieri residenti per 1000 abitanti. Per un determinato comune, la maggiore o minore presenza di popolazione straniera residente può indicare la maggiore o minore attrattività del territorio rispetto alle opportunità lavorative. Valori al di sotto di 5 stranieri residenti ogni 1000, sono a Seclì (1), Statte (3,6), Bagnolo del Salento (4,8); valori maggiore di 90 stranieri residenti ogni 1000 sono a Poggio Imperiale (91,1), Rignano Garganico (97,3), Stornarella (99,3), Ortona (103,8), Carapelle (104,1), Stornara (147,7). Ad esclusione di pochissimi comuni, tutti quelli della zona del subappennino dauno, Tavoliere delle Puglie e Gargano appartengono a classi elevate di popolazione straniera residente. Un comportamento spaziale opposto si ha nel Salento che presenta una maggiore concentrazione di classi medio-basse (fig. 6).



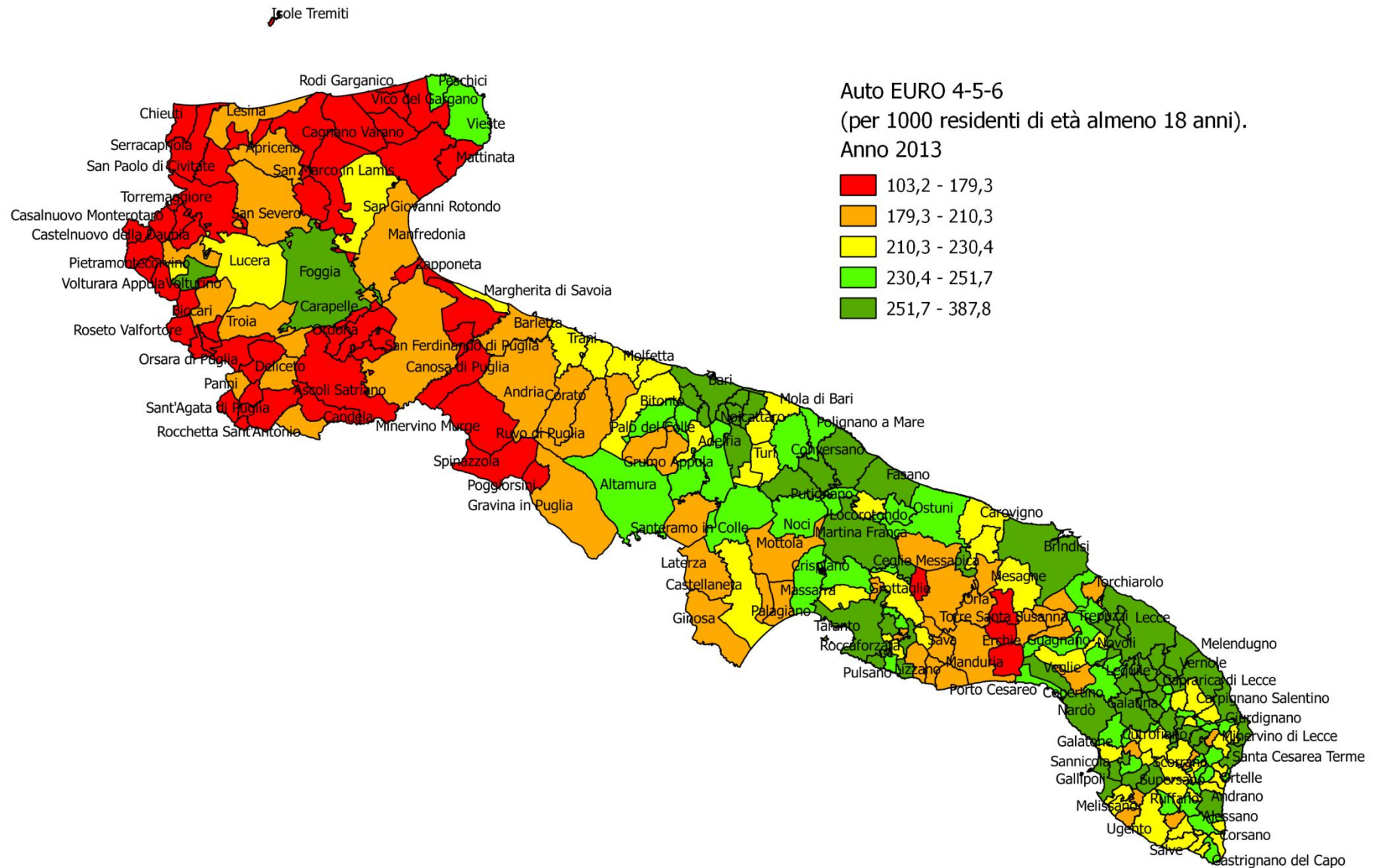
### **Auto ecologiche**

Gli inquinanti tipici di un'auto sono il monossido di carbonio, gli ossidi di azoto, gli idrocarburi incombusti e il particolato. I parametri di questi inquinanti sono regolati dalle normative internazionali che in Europa si misurano come Euro 0 (le più inquinanti) fino a Euro 6 (le più ecologiche).

In base al numero di auto ecologiche in circolazione, che siano almeno Euro 4, si può evidenziare la maggiore o minore propensione alla sensibilità ambientale, della popolazione di un territorio. La più elevata si registra a Volturino con 387,8 auto ecologiche per 1000 abitanti di età maggiore di 18 anni, seguito da Lecce (356,7) e Cavallino (355,6). La più bassa è ad Anzano di Puglia (103,2) e Casavecchio di Puglia (115,8). L'area del Salento ha la più alta concentrazione di comuni a classi di valori medio alti. La zona del Tavoliere delle Puglie, del Gargano e del Subappennino dauno hanno, di contro, quella di comuni a classe più bassa.



Fig.7 – Numero di auto di tipo Euro 4, Euro 5, Euro 6 per comuni. Anno 2013. Puglia (per 1000 residenti nello stesso comune, di età almeno 18 anni)



## Classificazione Ascendente Gerarchica

L'analisi esplorativa dei dati dimensionali trova supporto nei GIS. Alcune tecniche, come quelle di classificazione automatica, hanno l'obiettivo di formare dei gruppi omogenei, detti anche *cluster*, secondo un criterio di somiglianza o di differenza tra le unità, e secondo una particolare strategia di raggruppamento (gerarchica, non gerarchica, divisiva, agglomerativa). Ognuna di queste conduce a dei risultati differenti, quindi a *cluster* di diversa numerosità e caratteristiche statistiche. La decisione su quale soluzione adottare dipende dall'ipotesi che lo sperimentatore intende verificare, con tale classificazione. A titolo esemplificativo, si esaminano le sette variabili descritte in precedenza, al fine di caratterizzare, dal punto di vista economico e socio-demografico, i comuni. Un risultato accettabile è ottenuto attraverso una classificazione ascendente gerarchica, con il criterio della dissimilarità, misurata dalla distanza euclidea, e dal metodo di agglomerazione di *Ward*, notoriamente il più efficace quando le unità sono rappresentabili come punti di uno spazio euclideo multidimensionale. L'algoritmo di *Ward* parte da  $n$  gruppi contenenti una sola unità e, ad ogni passo, riunisce tra loro quelli che rendono minimo l'incremento, nella somma totale delle distanze, al quadrato, di ciascun comune dai rispettivi baricentri. Il risultato è mostrato in figura 8.

La mappa rappresenta i gruppi ottenuti classificando i comuni pugliesi in base al grado di "dissimilarità", misurato attraverso le variabili di cui sopra.

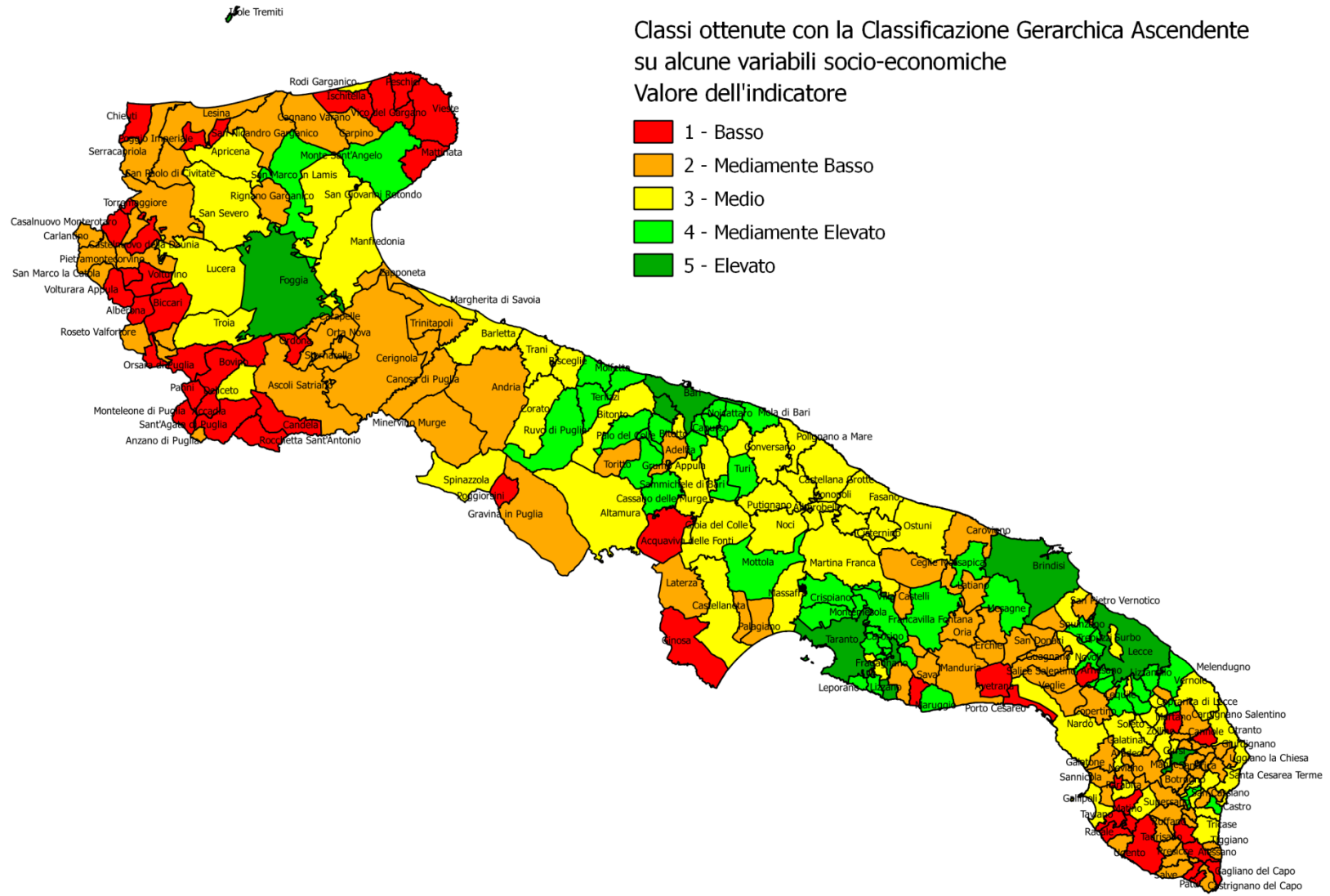
Con l'intento di costruire una sorta di "graduatoria" e quindi una classificazione dei comuni pugliesi per caratteristiche di tipo socio-demo-economico, si pone l'attenzione sulle tabelle 1 e 2. La tabella 2 definisce la "distanza euclidea" (separazione) tra le 5 classi di comuni ottenute dall'analisi di *clustering*, riscontrando come le classi più nettamente separate sono quelle con valori più elevati, viceversa la classi con valori più bassi sono quelle con profilo di separazione meno accentuato. La tabella 1 rappresenta sulle righe i baricentri delle classi dei comuni e sulle colonne le sette variabili, utilizzate per la costruzione della graduatoria di tipo socio-demo-economico. A ciascuno dei valori delle sette colonne (variabili socio-demo-economiche) sono stati attribuiti dei ranghi con la finalità di definire una "graduatoria" – dal migliore al peggiore – ai comuni, rappresentati da quelli al centro di ciascuna classe.

Ad esempio la classe 5, il cui centro è riferito al comune di Maglie, primeggia nelle variabili socio-economiche per cui possiamo dire che questa classe sia rappresentativa di un "elevato" tenore socio-demo-economico.

Viceversa, la classe 1, il cui centro è riferito al comune di Torricella, non primeggia nelle variabili potendo così affermare che questa classe sia rappresentativa di un "basso" profilo socio-demo-economico.

La classe più numerosa è la 2, con 93 comuni. Contiene, in media, come si evince dalla tabella 1, valori di rango 3 per Quoziente Natalità e Popolazione straniera residente; di rango 4 per Reddito imponibile IRPEF, Densità Demografica e auto EURO 4-5-6; di rango 2 per Quoziente Mortalità e di rango 5 per Valore Aggiunto Pro-Capite. Si tratta, quindi, di valori medio bassi di tutte le variabili considerate. Osservando la tabella 2 si riscontra che la classe 2 risulta molto distante dalla classe 5 (12040,6) e più distante dalla classe 3 (4003,1) che dalla classe 4 (3096,1). Accade che la classe 3, a sua volta, è più distante dalla classe 2 (4003,1) che non dalla 1 (2490,1) e dalla 4 (3351,8). Pertanto, non è evidente la separazione geografica dei comuni che appartengono alle classi 2, 3 e 4, anche se lo è rispetto ai criteri con cui l'algoritmo ha elaborato le sette variabili. Di contro, la classe 5 localizza geograficamente principalmente comuni o aggregati di comuni affacciati alle coste, tranne Foggia. Lo stesso accade per la classe 1; in questo caso la localizzazione dei comuni è principalmente quella al confine con le altre regioni (Molise, Campania, Basilicata). La classe 5, sulla base dei criteri imposti, è quella che in modo più accentuato si separa, nei valori e geograficamente, dalle altre quattro. La classe 5, infatti, per tutte le variabili considerate presenta valori di rango 1. Per il quoziente di mortalità di rango 5, l'interpretazione è, comunque, quella più favorevole. In fig. 8 si identificano i comuni appartenenti alle classi. In classe 5 (verde) si trovano Taranto, Brindisi, Foggia, Isole Tremiti, Bari, Modugno, Lecce, Maglie. In classe 4 (verde chiaro), relativamente distante dalla classe 5 (11116,9), si trovano 50 comuni.

Fig.8 – Le 5 classi ottenute dalla classificazione gerarchica ascendente per le sette variabili economiche e socio-demografiche considerate.



**Tab 1 – Classificazione ascendente gerarchica. Baricentri della variabili per classi e comuni al centro delle classi, in parentesi il rango all'interno della singola colonna.**

Classe – Comuni al centro	BARICENTRI, fra parentesi i ranghi all'interno della singola colonna						
	Valore Aggiunto Pro-Capite (euro)	Reddito imponibile IRPEF – medio (euro)	Densità Demografica (ab/kmq)	Quoziente Natalità	Quoziente Mortalità	Popolazione straniera residente (per 1000 ab)	auto EURO 4-5-6 (per 1000 ab maggiorenni)
1 (Torricella)	11752,6 (3)	12017,1 (5)	152,3 (5)	1,7 (5)	3,0 (1)	29,9 (1)	197,1 (5)
2 (Torre Santa Susanna)	8979,8 (5)	12190,4 (4)	222,3 (4)	2,0 (3)	2,4 (2)	25,3 (3)	198,6 (4)
3 (Caprarica di Lecce)	12317,3 (2)	14720,2 (3)	279,2 (3)	1,9 (4)	2,2 (3)	23,5 (4)	242,1 (2)
4 (Arnesano)	8622,7 (4)	15342,9 (2)	423,0 (2)	2,2 (2)	2,1 (4)	16,3 (5)	238,3 (3)
5 (Maglie)	21605,4 (1)	18662,8 (1)	823,8 (1)	2,3 (1)	1,8 (5)	29,9 (1)	288,0 (1)

**Tab 2 – Distanza euclidea fra le classi ottenute**

Classe	1 (Torricella)	2 (Torre Santa Susanna)	3 (Caprarica di Lecce)	4 (Arnesano)	5 (Maglie)
1 (Torricella)	0,0	2737,3	2490,1	4215,7	9681,2
2 (Torre Santa Susanna)	2737,3	0,0	4003,1	3096,1	12040,6
3 (Caprarica di Lecce)	2490,1	4003,1	0,0	3351,8	8148,5
4 (Arnesano)	4215,7	3096,1	3351,8	0,0	11116,9
5 (Maglie)	9681,2	12040,6	8148,5	11116,9	0,0

**Tab. 3 – Numero di comuni per classi ottenute**

Classe	1	2	3	4	5
Numero Comuni	45	93	62	50	8

**Per Eventuali Contatti**

UFFICIO STATISTICO, Via Gentile 52 - 70126 Bari

email: [ufficio.statistico@regione.puglia.it](mailto:ufficio.statistico@regione.puglia.it)