



# Tecniche di analisi statistica spaziale su dati georeferenziati

**Massimo Mucciardi**

Dipartimento di Scienze Economiche,  
Aziendali, Ambientali e Metodologie  
Quantitative  
Sezione di Scienze Statistiche e Matematiche  
Università degli Studi di Messina

**Edoardo Otranto**

Scienze Cognitive della Formazione e degli Studi  
Culturali  
Università degli Studi di Messina

**Centro Statistico di Ateneo**  
Università degli Studi di Messina

Giornata Italiana della Statistica  
Indicatori per la gestione del territorio  
Dai censimenti alla policy  
Messina, 24 ottobre 2014

# Outlook

---

Tecniche di analisi statistica spaziale

(Indicatori spaziali: globali e locali)

La georeferenziazione dei dati

Applicazione su dati ISTAT - Sicilia

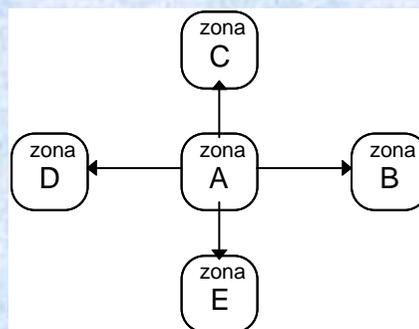
(censimento popolazione – censimento industria)

# AUTOCORRELAZIONE SPAZIALE (1)



- Nelle analisi statistiche territoriali risulta fondamentale lo studio di problemi che implicano il concetto di “spazio”. (Infatti non è possibile pensare che le manifestazioni di un fenomeno economico, sociale, ambientale et.. siano indipendenti dalla configurazione territoriale di una determinata zona oggetto di studio)
- L’esistenza di una relazione tra quello che accade presso una unità territoriale e ciò che accade altrove nello spazio, è sintomo di quella caratteristica che abitualmente si definisce come “dipendenza spaziale”.
- L’elemento più importante è costituito dalla parola “correlazione”. In generale, la correlazione misura la relazione prevalente fra una coppia di variabili. I coefficienti utilizzati normalmente in statistica forniscono informazioni sulla sua natura e il suo grado.
- Nel termine autocorrelazione spaziale, è inoltre contenuto il suffisso “auto” (si considera la stessa variabile)
- Dipendenza unidirezionale (tempo) Dipendenza multidirezionale (territorio)

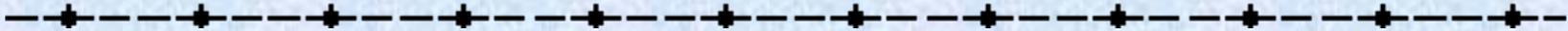
## AUTOCORRELAZIONE SPAZIALE (2)



➤ In generale esiste autocorrelazione spaziale positiva se l'intensità del fenomeno in una zona è simile all'intensità del fenomeno in zone contigue, mentre esiste autocorrelazione negativa nel caso in cui due zone contigue presentano intensità diverse.

➤ I dati spaziali non sono indipendenti: accade infatti che osservazioni che si riferiscono ad aree più vicine mostrino meno variabilità di dati relativi ad aree che risultano più lontane tra loro.

# Le matrici di contiguità (1)



- La procedura in grado di ordinare territorialmente più unità prende il nome di sistema di interconnessione o sistema di contiguità
- La contiguità spaziale è spesso un fattore che non è possibile tenere sotto controllo in maniera adeguata; ciò accade sia per le difficoltà connesse alla definizione (forma e dimensione) delle unità territoriali di rilevazione, sia perché i fenomeni socio-demografici sono rilevati su unità amministrative le quali hanno le più diverse forme ed estensioni.
- Possono inoltre sussistere, soprattutto nel campo socio-demografico, altri elementi di contatto tra le unità territoriali che giocano anch'essi il ruolo di fattori di interazione e che possono interagire con la contiguità spaziale. Ad esempio questi fattori possono essere identificati dalla lunghezza del confine tra due unità territoriali, da collegamenti stradali, ferroviari ed aerei esistenti tra le unità

## Le matrici di contiguità (2)

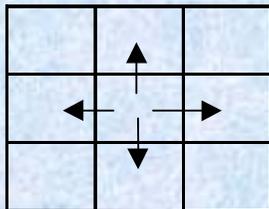


➤ La matrice presenta elementi ( $w_{ij}$ ) che assumono valore “1”, nel caso in cui due generiche unità “i” e “j” siano confinanti, “0” nel caso in cui le due generiche unità non siano confinanti (con  $w_{ii}=0$  e  $w_{ij}=w_{ji}$ ).

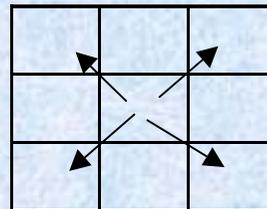
➤ In questo caso si dice che vi è contiguità di primo ordine oppure che le due unità sono connesse a ritardo “1”.

➤ Schemi di contiguità:

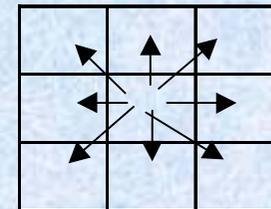
- 1) contiguità a bordo comune ;
- 2) contiguità a vertice comune ;
- 3) contiguità mista (vertice e bordo comune).



Rooks Case

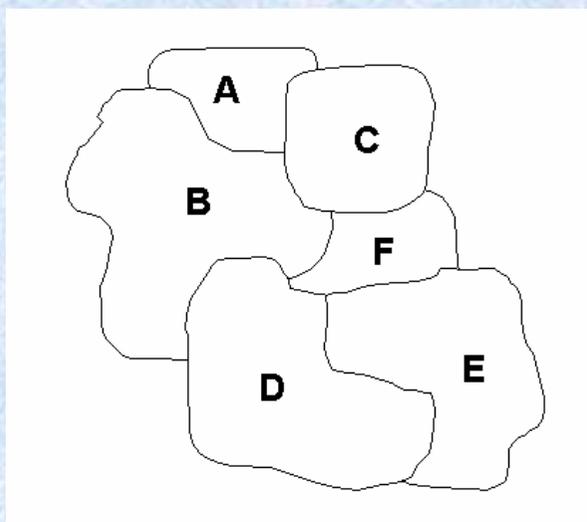
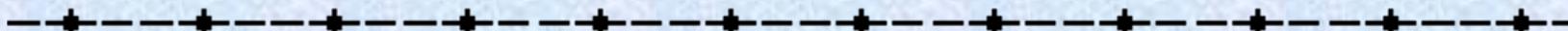


Bishops Case



Queen Case

## Le matrici di contiguità (2)



	A	B	C	D	E	F
A		1	1	0	0	0
B	1		1	1	0	1
C	1	1		0	0	1
D	0	1	0		1	1
E	0	0	0	1		1
F	0	1	1	1	1	

## Esempio di matrice spaziale (normalizzata)

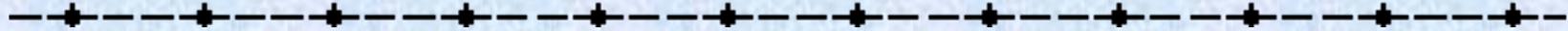


---

Adj	Torino	Vercelli	Novara	Cuneo	Asti	Alessandri	Biella	Verb-Cus-	Aosta
Torino		0.17		0.17	0.17	0.17	0.17		0.17
Vercelli	0.14		0.14			0.14	0.14	0.14	0.14
Novara		0.20						0.20	
Cuneo	0.25				0.25				
Asti	0.25			0.25		0.25			
Alessandri	0.14	0.14			0.14				
Biella	0.33	0.33							0.33
Verb-Cus-Ossola		0.33	0.33						
Aosta	0.33	0.33					0.33		
Varese			0.25					0.25	
Como									
Sondrio									
Milano			0.10						
Bergamo									
Brescia									
Pavia		0.14	0.14			0.14			
Cremona									
Mantova									

➤ Altri sistemi di contiguità (riferimento distanza, combinazione di distanza e lunghezza, funzioni Kernel etc..)

# Indici di Autocorrelazione Spaziale



$$I_m = \frac{n \sum_{i=1}^n \sum_{j=1}^n (x_i - M_x)(x_j - M_x) w_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - M_x)^2}$$

Indice di Moran

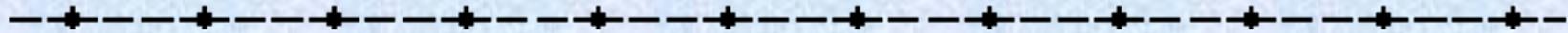
$I_m > 0 \rightarrow$  autocorrelazione positiva  
 $I_m = 0 \rightarrow$  autocorrelazione nulla  
 $I_m < 0 \rightarrow$  autocorrelazione negativa

$$I_g = \frac{n-1}{2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 w_{ij}}{\sum_{i=1}^n (x_i - M_x)^2}$$

Indice di Geary

$I_g > 1 \rightarrow$  autocorrelazione negativa  
 $I_g = 1 \rightarrow$  autocorrelazione nulla  
 $I_g < 1 \rightarrow$  autocorrelazione positiva  
 $I'_g = 1 - I_g$

## Significatività indici di Autocorrelazione Spaziale



$$E_N(I) = \frac{-1}{n-1}$$

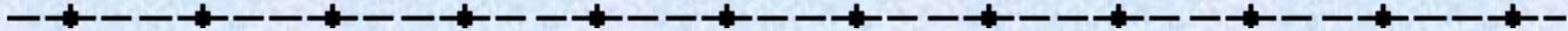
$$\text{Var}_N(I) = \left[ \frac{1}{S_0^2(n^2-1)} (n^2 S_1 - n S_2 + 3 S_0^2) \right] - \{E_N(I)\}^2$$

$$z = \frac{I - E_N(I)}{\sqrt{\text{Var}_N(I)}}$$

## L'autocorrelazione spaziale locale (LISA)

- 
- Il concetto di autocorrelazione spaziale *locale* trova invece il suo fondamento nell'idea di "outlier spaziale" (punto di instabilità del processo spaziale sottostante) e di "cluster spaziale". Si tratta di una tecnica che descrive e visualizza la distribuzione spaziale del fenomeno indagato.
  - L'analisi della correlazione locale si estrinseca nell'individuazione di quelle zone/unità amministrative caratterizzate da valori estremi della variabile; valori estremi che vengono individuati attraverso il raffronto tra il valore che la variabile assume nella *i*-esima unità ed i valori che il processo assume nelle zone/unità amministrative contigue.
  - Es. da un punto di vista economico, relazione con l'idea di distretto come cluster di concentrazione produttiva che "rompe" in qualche modo la regolarità dello spazio circostante.

# Indice di autocorrelazione locale



$$I_i = z_i \sum_{\substack{j=1 \\ j \neq i}}^n w_{ij} z_j$$

$$I_i > 0 \quad I_i < 0 \quad I_i = 0$$

Valori positivi di  $I_i$  indicano fenomeni di clusterizzazione mentre valori negativi indicano presenza di outliers spaziale

# Indice di autocorrelazione locale (inferenza)

---

$$E(I_i) = -w_i / (n - 1)$$

$$\text{Var}(I_i) = w_{i(2)}(n - b_2) / (n - 1) + 2w_{i(kh)}(2b_2 - n) / (n - 1)(n - 2) - w_i^2 / (n - 1)^2$$

$$\text{with } b_2 = m_4 / m_2^2 \quad w_{i(2)} = \sum_{j \neq i} w_{ij}^2 \quad 2w_{i(kh)} = \sum_{k \neq i} \sum_{h \neq i} w_{ik} w_{ih}$$

$$z(I_i) = \frac{I_i - E(I_i)}{\sqrt{\text{var}(I_i)}}$$



# Mapping Clusters/Outliers

## Legend

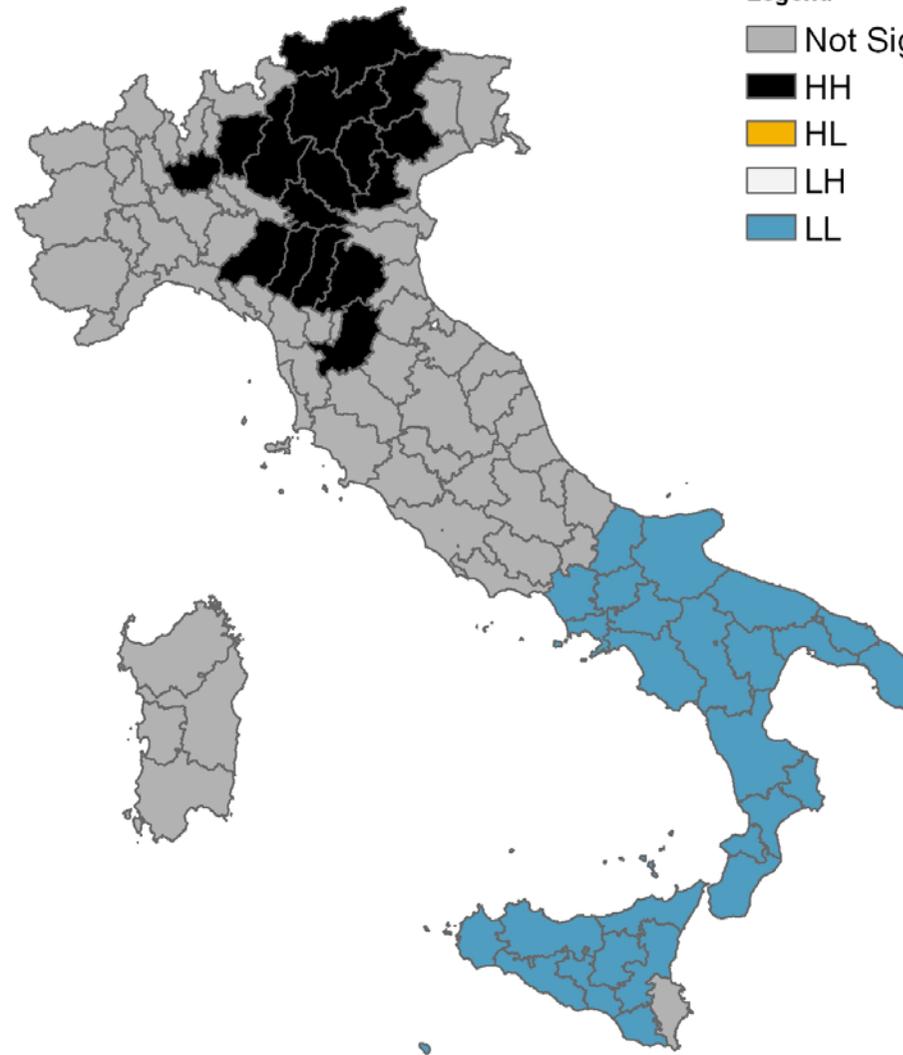
Not Significant

HH

HL

LH

LL



# Georeferenziazione dei dati ISTAT

- 
- La georeferenziazione di un evento consiste nell'associare ad esso una coppia di numeri che rappresentano le sue coordinate piane o geografiche (X ed Y o Lat. e Long.) tali da individuarlo univocamente sul territorio reale o su una mappa "georiferita" gestita mediante un GIS (Geographic Information System –ArcGis, GeoMedia, Grass, Qgis, ecc...) o comunque un software capace di rappresentare oggetti dotati di coordinate assolute
  - La georeferenziazione è essenzialmente un *legare al territorio* i diversi accadimenti o eventi.
  - Dati utilizzati: merge tra cartografia Istat – Sicilia (shape) e dati (dbf) provenienti dal Censimento della Popolazione (2011) e Censimento dell'Industria e dei Servizi (2011)  
<http://dati-censimentopopolazione.istat.it/>  
<http://dati-censimentoindustriaeservizi.istat.it/>.
  - Livello di aggregazione spaziale: "comune"

# Georeferenziazione dei dati ISTAT

---

## Variabili analizzate

- **indice di vecchiaia**
- **indice di dipendenza strutturale.**
- **indice di ricambio della popolazione attiva**
- **indice di struttura della popolazione attiva**
- **n° stranieri/popolazione residente**
- **n° addetti impresa/popolazione residente**
- **n° addetti istituzione pubblica/popolazione residente**
- **n° addetti istituzione no profit/popolazione residente**
- **n° addetti attività dei servizi di alloggio e di ristorazione /popolazione residente**
- **n° addetti attività costruzioni/popolazione residente**
- **n° addetti attività manifatturiere/popolazione residente**

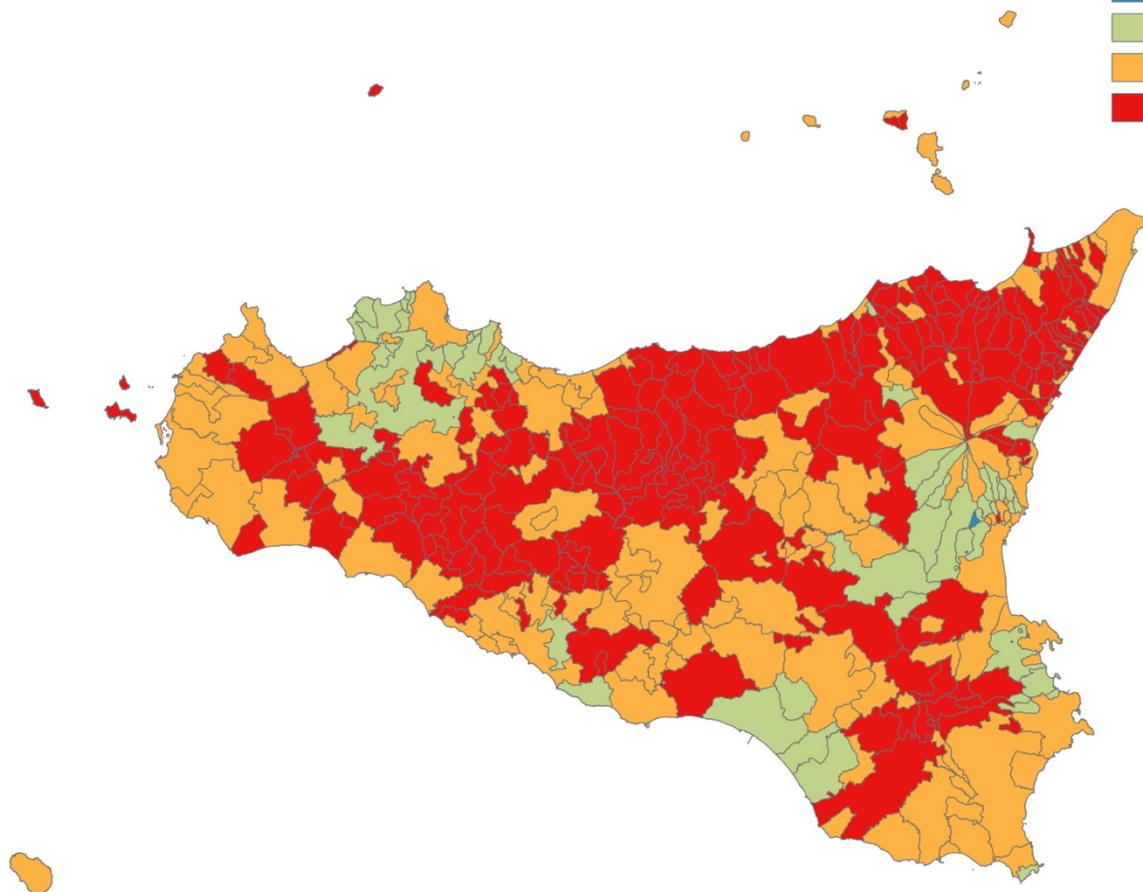
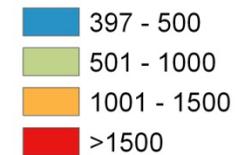
- Tutti gli indicatori sono moltiplicati per 1000 abitanti
- Matrice utilizzata = contiguità classica (standardizzata)

# Indice di vecchiaia

Analisi spaziale globale

I\_Vecchiaia  
Moran's Index = 0,520  
Mean = 1641,32

I\_VECCHIAI



Rapporto percentuale tra il numero degli ultrasessantacinquenni ed il numero dei giovani fino ai 14 anni.

# Indice di vecchiaia

Analisi spaziale locale

Moran's Index = 0,520

Mean = 1641,32

SD = 623,31

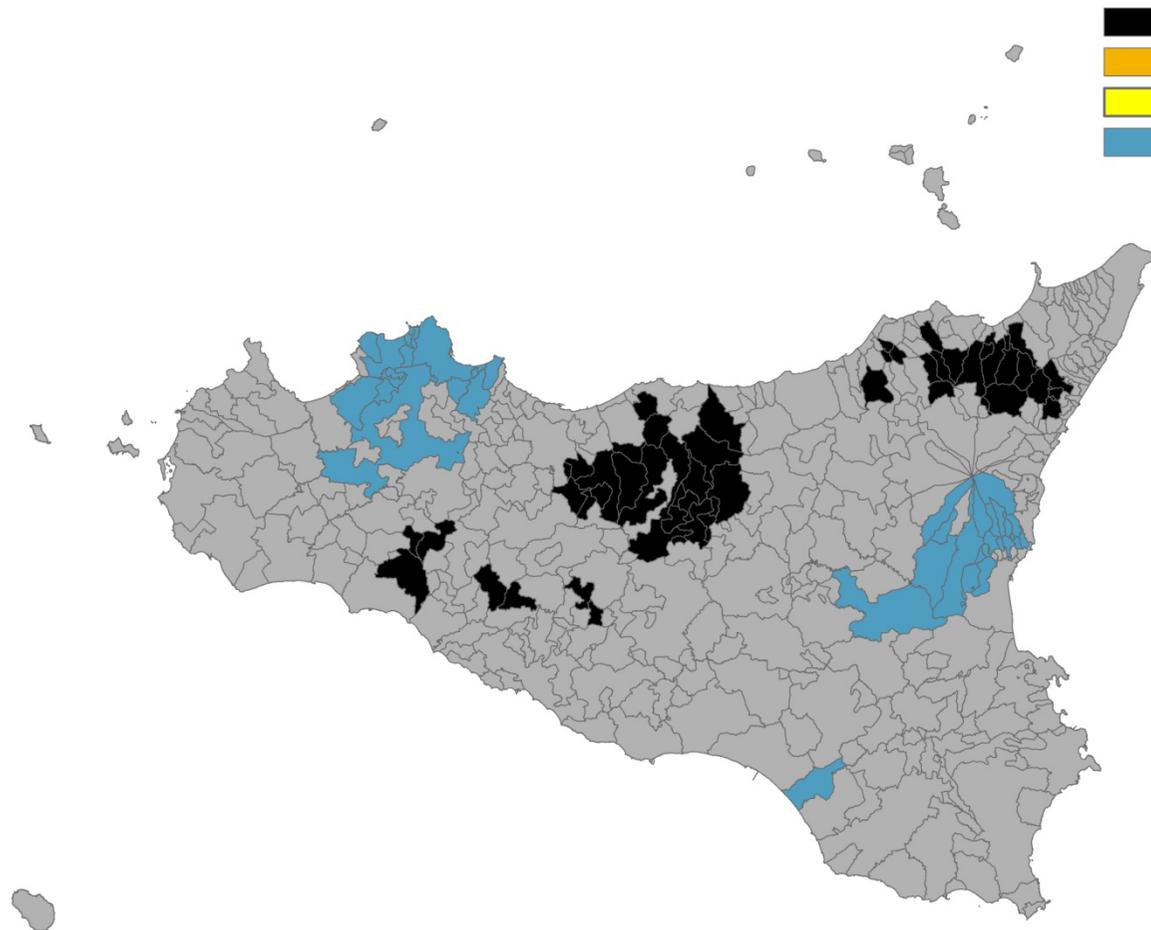
Not Significant

HH

HL

LH

LL



Comune	I_VECCHIAI	L_Moran	Tip_ass
NOVARA DI SICILIA	4219.30	5.75	HH
BASICO'	3400.00	5.62	HH
RACCUJA	3473.68	4.95	HH
BLUFI	3356.52	4.00	HH
UCRIA	3225.49	3.80	HH
TRIPÌ	3093.02	3.52	HH
FLORESTA	3604.17	3.47	HH
LIMINA	3075.27	3.27	HH
FONDACHELLI-FANTINA	2857.14	3.21	HH
BOMPIETRO	3136.65	2.47	HH
RESUTTANO	2498.01	2.43	HH
SAN SALVATORE DI FITALIA	3555.56	2.38	HH
SCLAFANI BAGNI	3794.87	2.37	HH
SCILLATO	2816.67	2.34	HH
SAN PIERO PATTI	2438.81	2.30	HH
MONTALBANO ELICONA	2672.54	2.23	HH
MALVAGNA	3038.96	2.21	HH
SUTERA	3462.12	2.11	HH
ALIMENA	2685.34	2.10	HH
ISNELLO	2815.48	2.06	HH
POLIZZI GENEROSA	2483.71	1.76	HH
GIULIANA	2929.65	1.62	HH
FRANCAVILLA DI SICILIA	2191.87	1.62	HH
CALTAVUTURO	2090.37	1.61	HH
ROCCAFIORITA	2173.91	1.52	HH
PETRALIA SOTTANA	2544.59	1.43	HH
PETRALIA SOPRANA	2141.59	1.37	HH
MONTEMAGGIORE BELSITO	2347.32	1.35	HH
SANT'ANGELO DI BROLO	2704.48	1.34	HH
CASALVECCHIO SICULO	3616.44	1.33	HH
MAZZARRA' SANT'ANDREA	2247.06	1.32	HH
FRAZZANO'	2935.06	1.28	HH
VALLEDOLMO	2184.87	1.27	HH
RODI' MILICI	2152.61	1.15	HH
ALCARA LI FUSI	3200.00	1.12	HH
ALESSANDRIA DELLA ROCCA	2318.68	1.09	HH
ANTILLO	2106.06	1.08	HH
MONGIUFFI MELIA	3070.18	1.07	HH
GRATTERI	2807.34	1.07	HH
GANGI	2273.78	1.03	HH
SAN BIAGIO PLATANI	2311.39	1.02	HH
MIRTO	2350.43	0.99	HH
SAN MAURO CASTELVERDE	2784.69	0.97	HH
CHIUSA SCLAFANI	2340.66	0.94	HH
ALIA	2202.11	0.94	HH
BOMPENSIERE	2468.75	0.93	HH
GERACI SICULO	2250.00	0.75	HH
CALTABELLOTTA	2600.00	0.72	HH

Comune	I_VECCHIAI	L_Moran	Tip_ass
CAMPOROTONDO ETNEO	396.72	3.00	LL
SAN PIETRO CLARENZA	571.64	2.31	LL
MISTERBIANCO	726.50	2.01	LL
VILLABATE	617.78	1.93	LL
FICARAZZI	649.93	1.84	LL
MOTTA SANT'ANASTASIA	789.13	1.82	LL
ACI SANT'ANTONIO	665.39	1.74	LL
CAPACI	874.87	1.72	LL
CARINI	715.84	1.71	LL
TORRETTA	810.60	1.62	LL
BELPASSO	729.65	1.60	LL
ACI CATENA	615.79	1.57	LL
CINISI	896.99	1.51	LL
ISOLA DELLE FEMMINE	787.53	1.50	LL
MASCALUCIA	769.28	1.49	LL
GIARDINELLO	869.27	1.44	LL
TRECASTAGNI	832.04	1.44	LL
PEDARA	848.75	1.40	LL
VIAGRANDE	893.38	1.39	LL
BELMONTE MEZZAGNO	698.31	1.36	LL
VALVERDE	843.82	1.36	LL
BAGHERIA	980.05	1.34	LL
MONTELEPRE	1027.40	1.32	LL
BORGETTO	923.73	1.32	LL
MISILMERI	882.46	1.28	LL
ACI BONACCORSI	1030.95	1.22	LL
NICOLOSI	1066.24	1.18	LL
SANTA MARIA DI LICODIA	945.21	1.15	LL
PATERNO'	893.91	1.14	LL
BIANCAVILLA	899.91	1.14	LL
ACATE	830.34	1.03	LL
PALERMO	1193.09	1.00	LL
SAN GIOVANNI LA PUNTA	1045.58	0.96	LL
RAMACCA	800.52	0.95	LL
TREMESTIERI ETNEO	992.95	0.84	LL
ZAFFERANA ETNEA	1017.48	0.76	LL
PARTINICO	1081.78	0.74	LL
MONREALE	836.50	0.57	LL

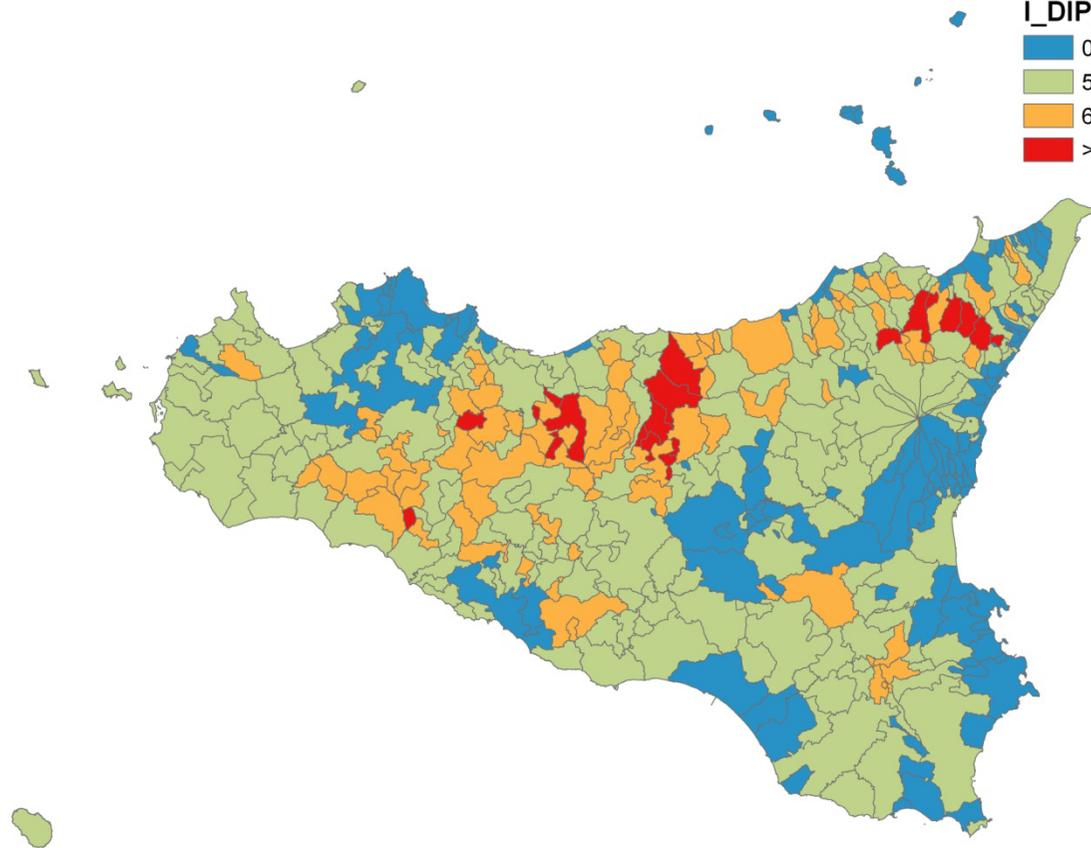
# Indice di dipendenza strutturale

Analisi spaziale globale

Moran's Index: = 0,52  
Mean = 550,73  
SD = 76,55

I\_DIP\_STRU

- 0 - 500
- 501 - 600
- 601 - 700
- >700

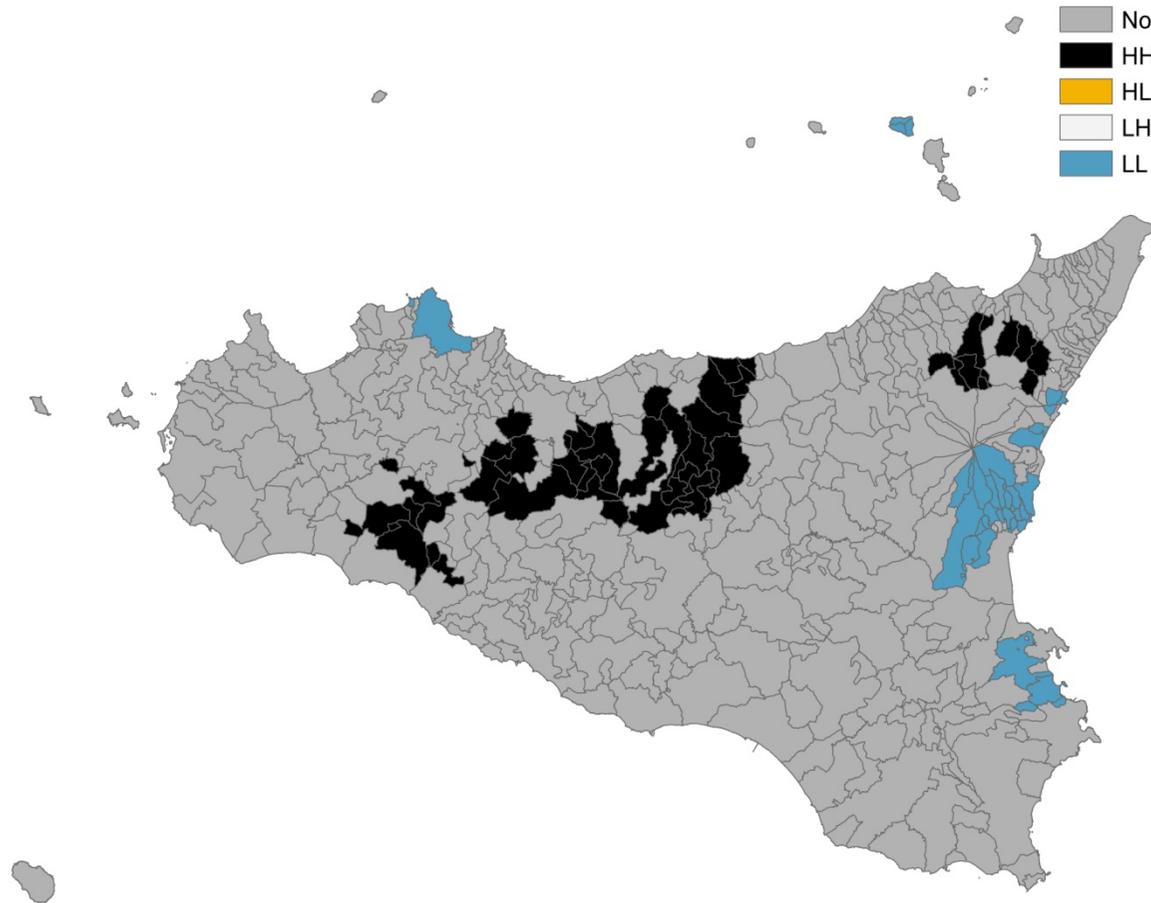


Popolazione non attiva (0-14 anni e 65 anni ed oltre) su quella attiva (15-64 anni)

# Indice di dipendenza strutturale

Analisi spaziale locale

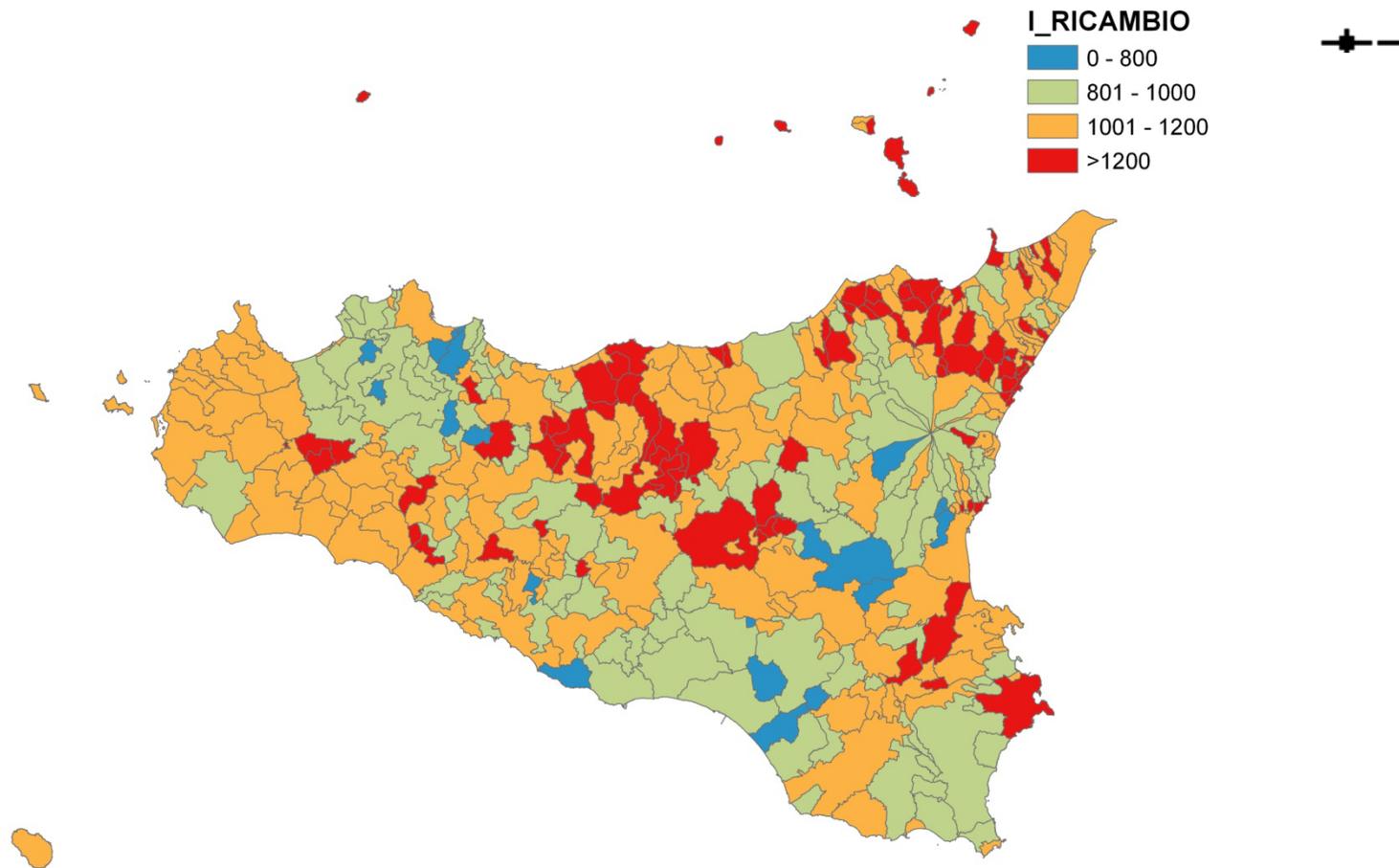
Moran's Index: = 0,52  
Mean = 550,73  
SD = 76,55



# Indice di ricambio della popolazione attiva

Analisi spaziale globale

Moran's Index: = 0,34  
Mean = 1073,50  
SD = 193,26

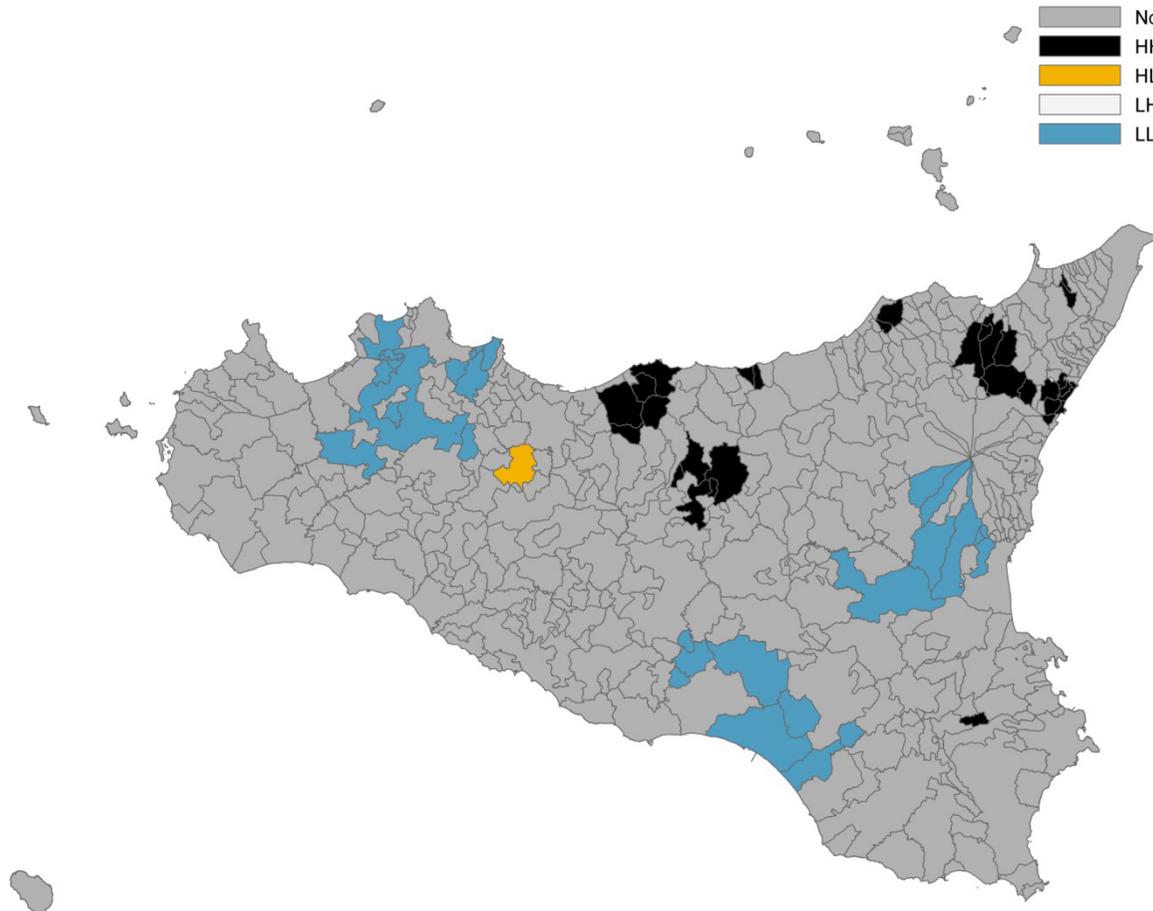
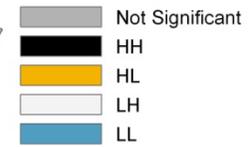


Rapporto percentuale tra la fascia di popolazione che sta per andare in pensione (55-64 anni) e quella che sta per entrare nel mondo del lavoro (15-24 anni)

# Indice di ricambio della popolazione attiva

Analisi spaziale locale

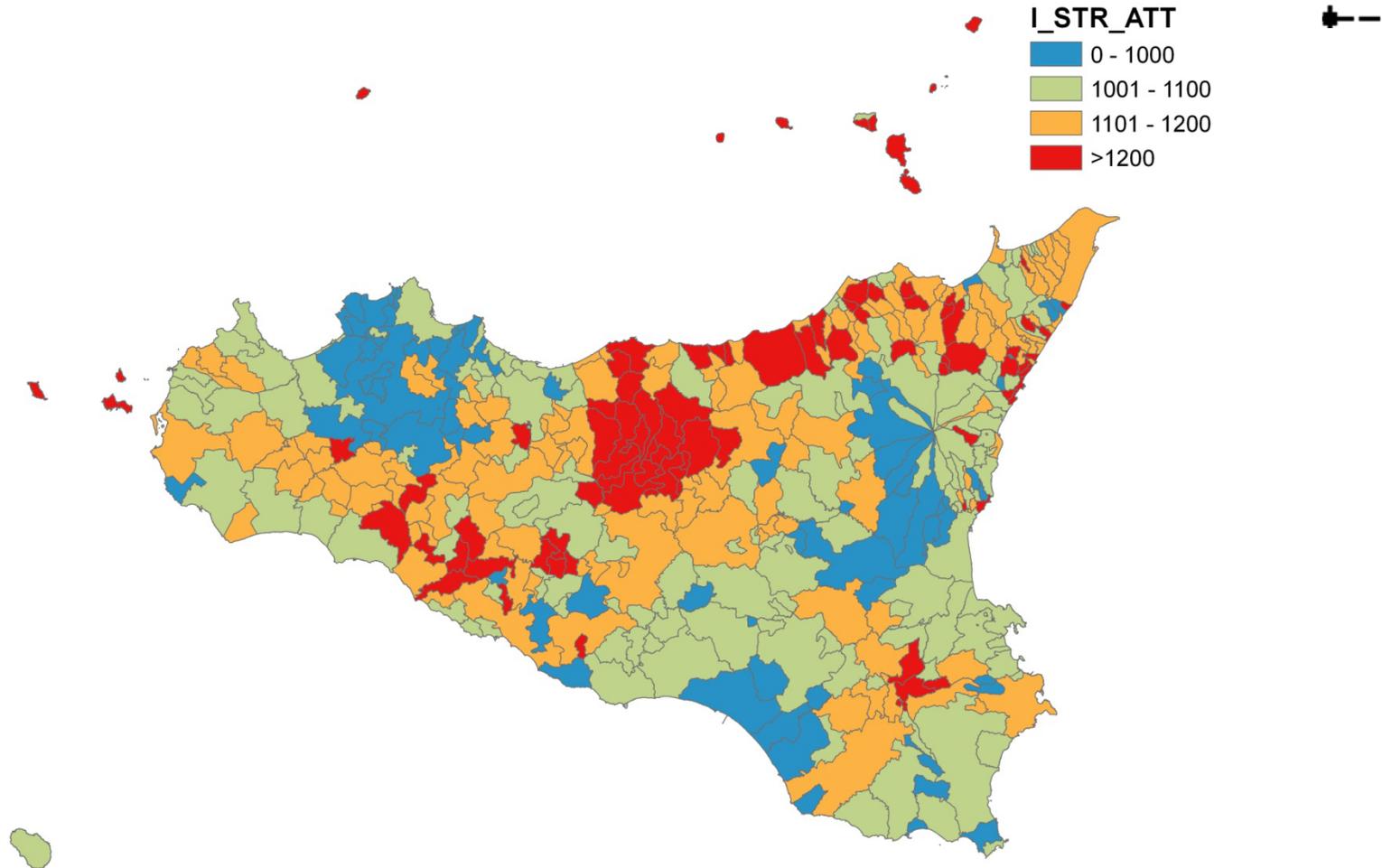
Moran's Index: = 0,34  
Mean = 1073,50  
SD =193,26



# Indice di struttura della popolazione attiva

Analisi spaziale globale

Moran's Index: = 0,41  
Mean = 1104,37  
SD = 114,53



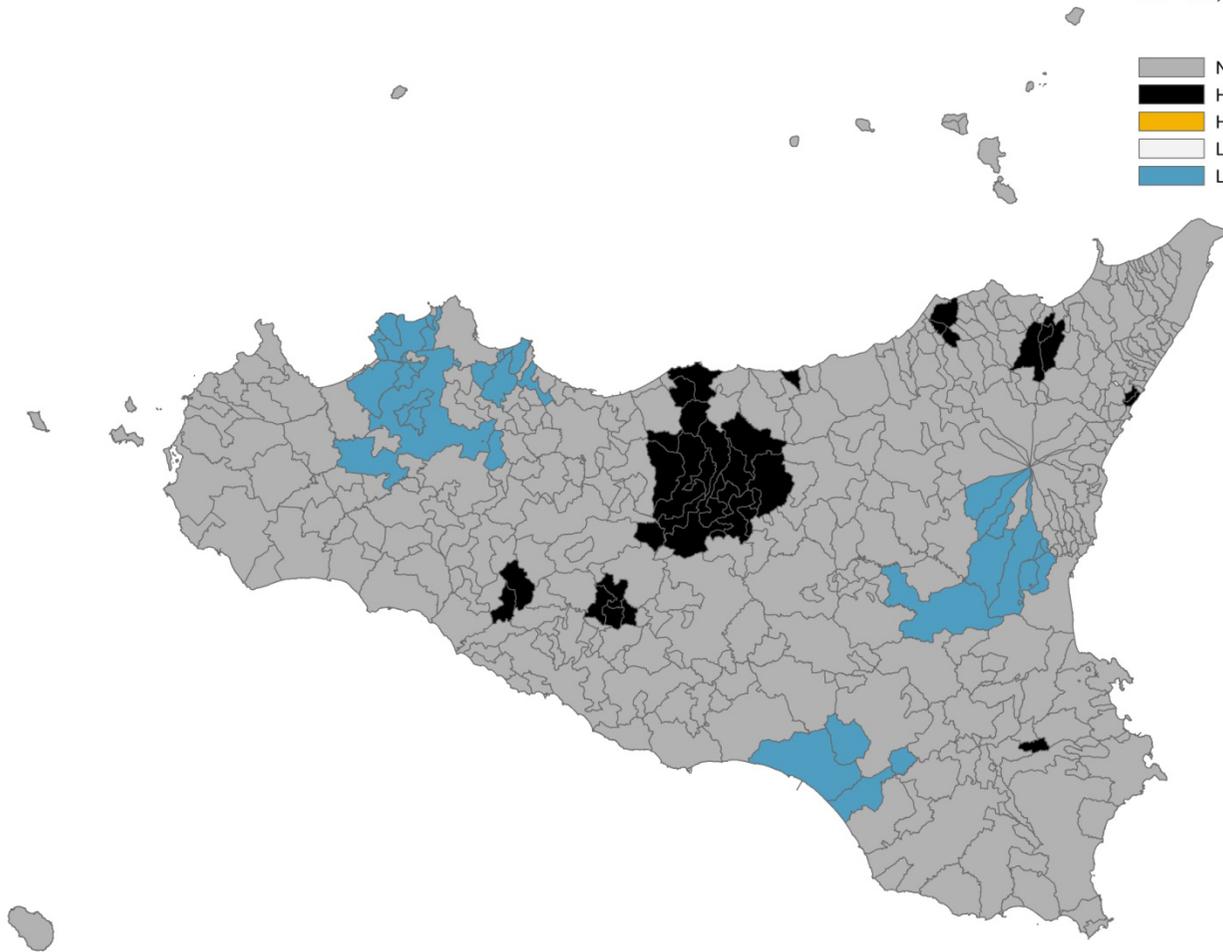
Rapporto percentuale tra la parte di popolazione in età lavorativa più anziana (40-64 anni) e quella più giovane (15-39 anni)

# Indice di struttura della popolazione attiva

Analisi spaziale locale

Moran's Index: = 0,41  
Mean = 1104,37  
SD = 114,53

- Not Significant
- HH
- HL
- LH
- LL



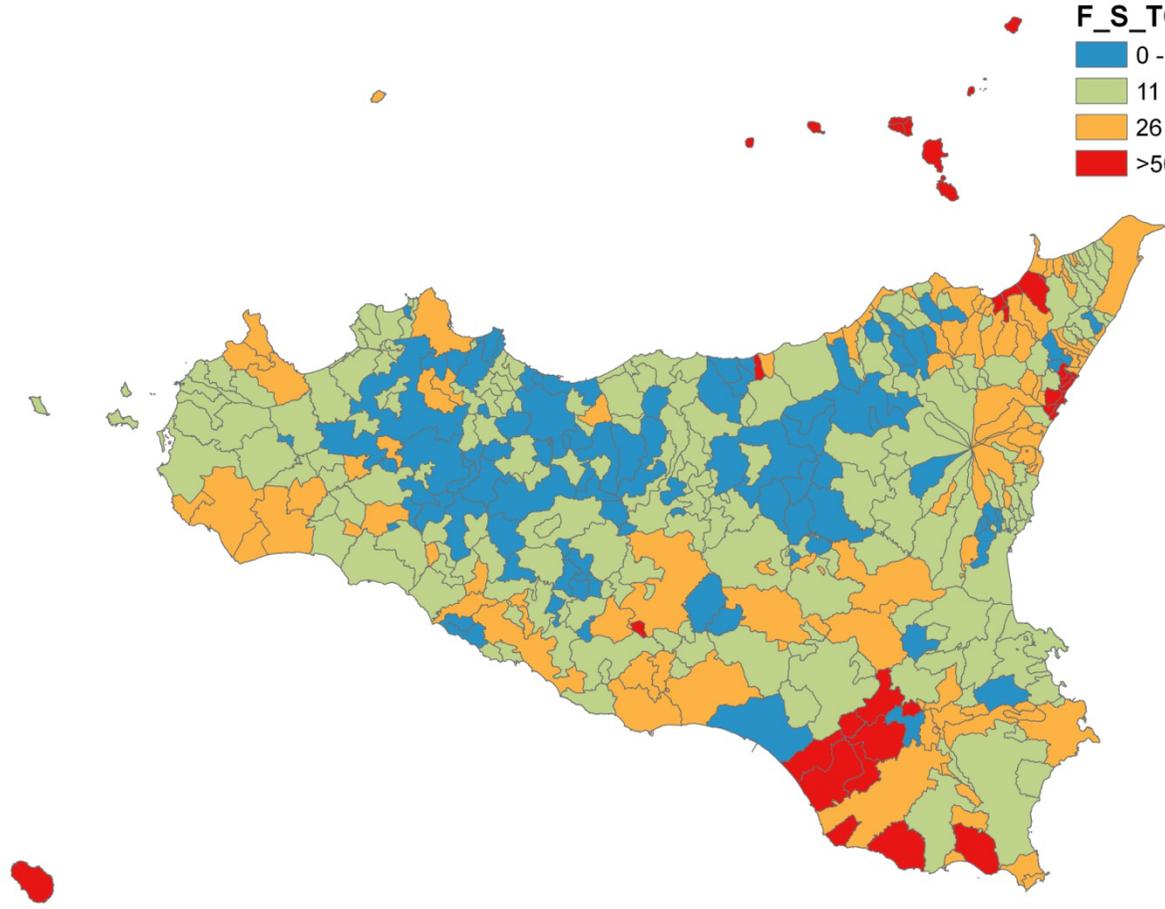
# N° di stranieri /Pop.

Analisi spaziale globale

Moran's Index: = 0,47  
Mean = 22,34  
SD = 19,57

F\_S\_TOTALE

- 0 - 10
- 11 - 25
- 26 - 50
- >50



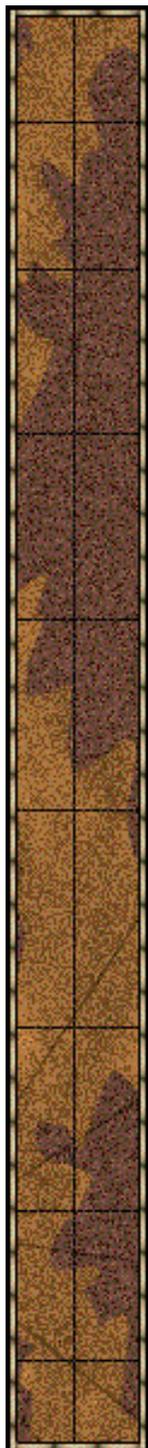
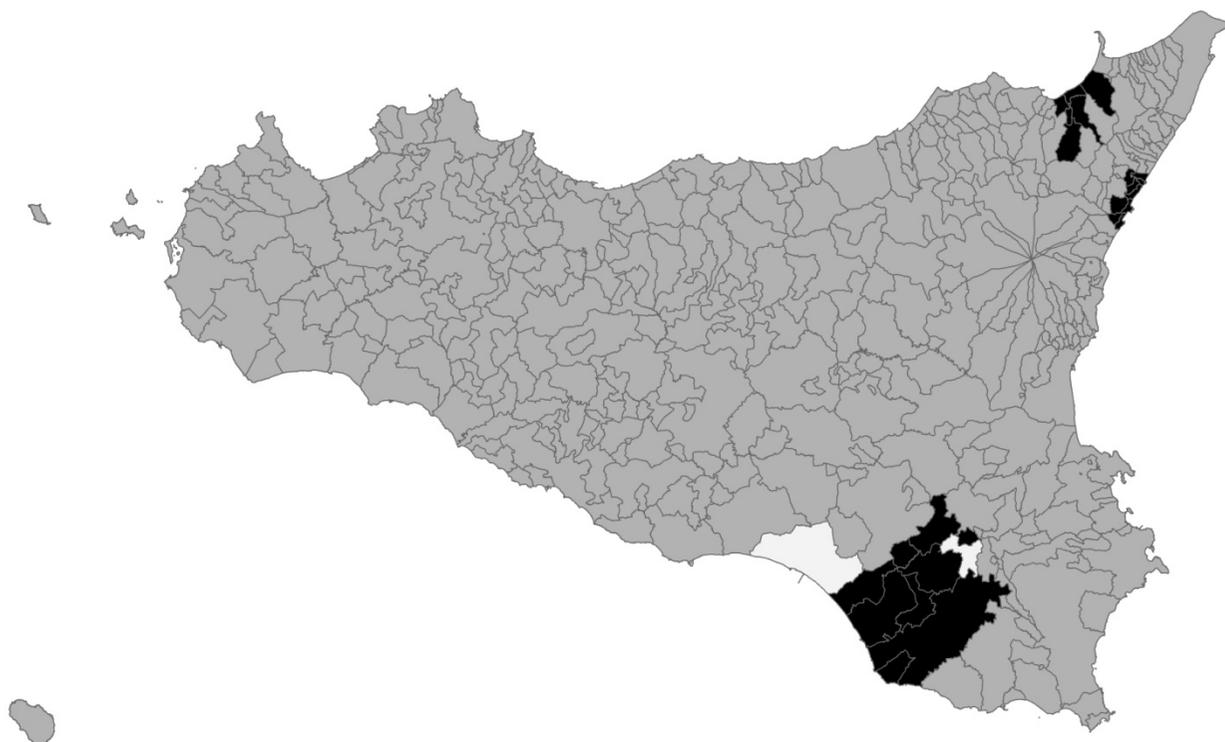
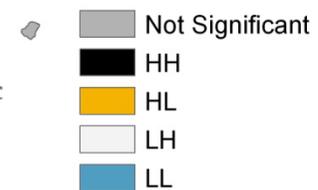
# N° di stranieri /Pop.

Analisi spaziale locale

Moran's Index: = 0,47

Mean = 22,34

SD = 19,57



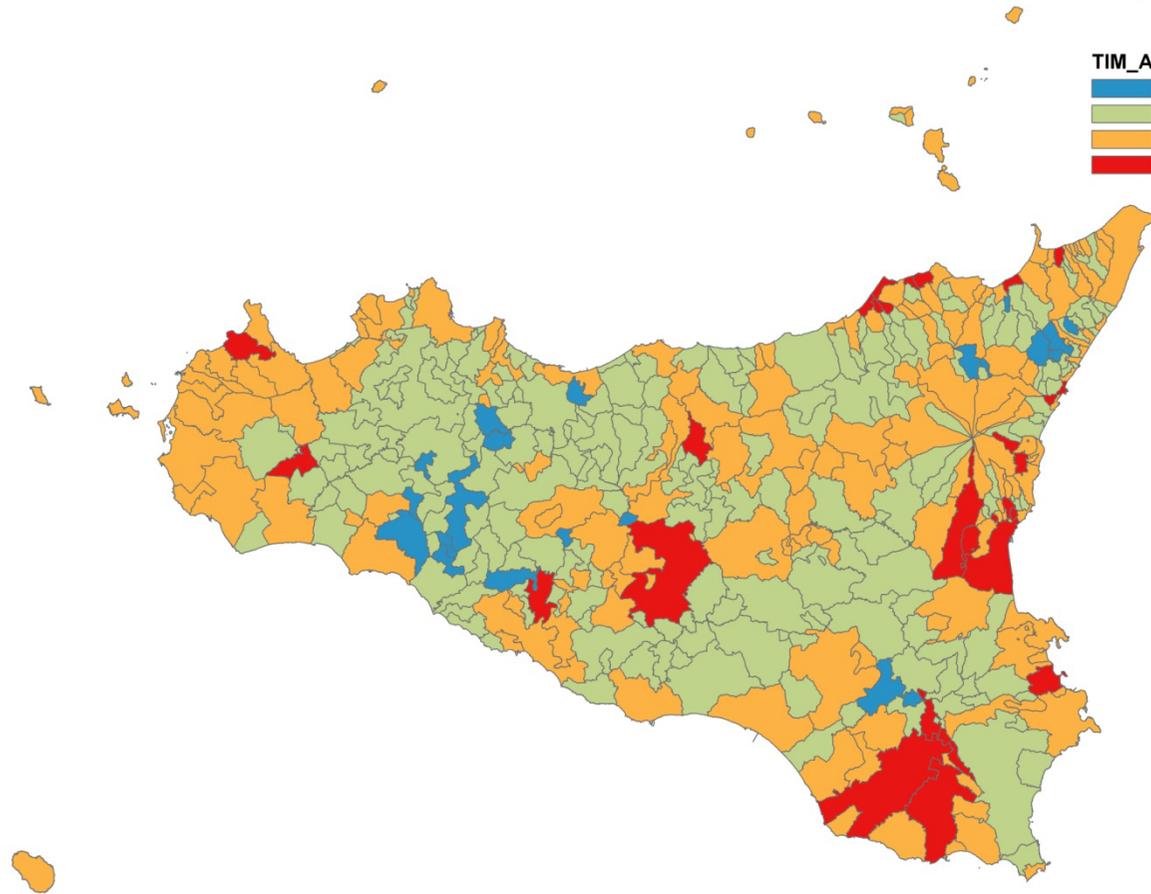
# N° di addetti Impresa /Pop.

Analisi spaziale globale

Moran's Index: = 0,24  
Mean = 109,66  
SD = 53,00

TIM\_ADDET1

- 0 - 50
- 51 - 100
- 101 - 200
- >200

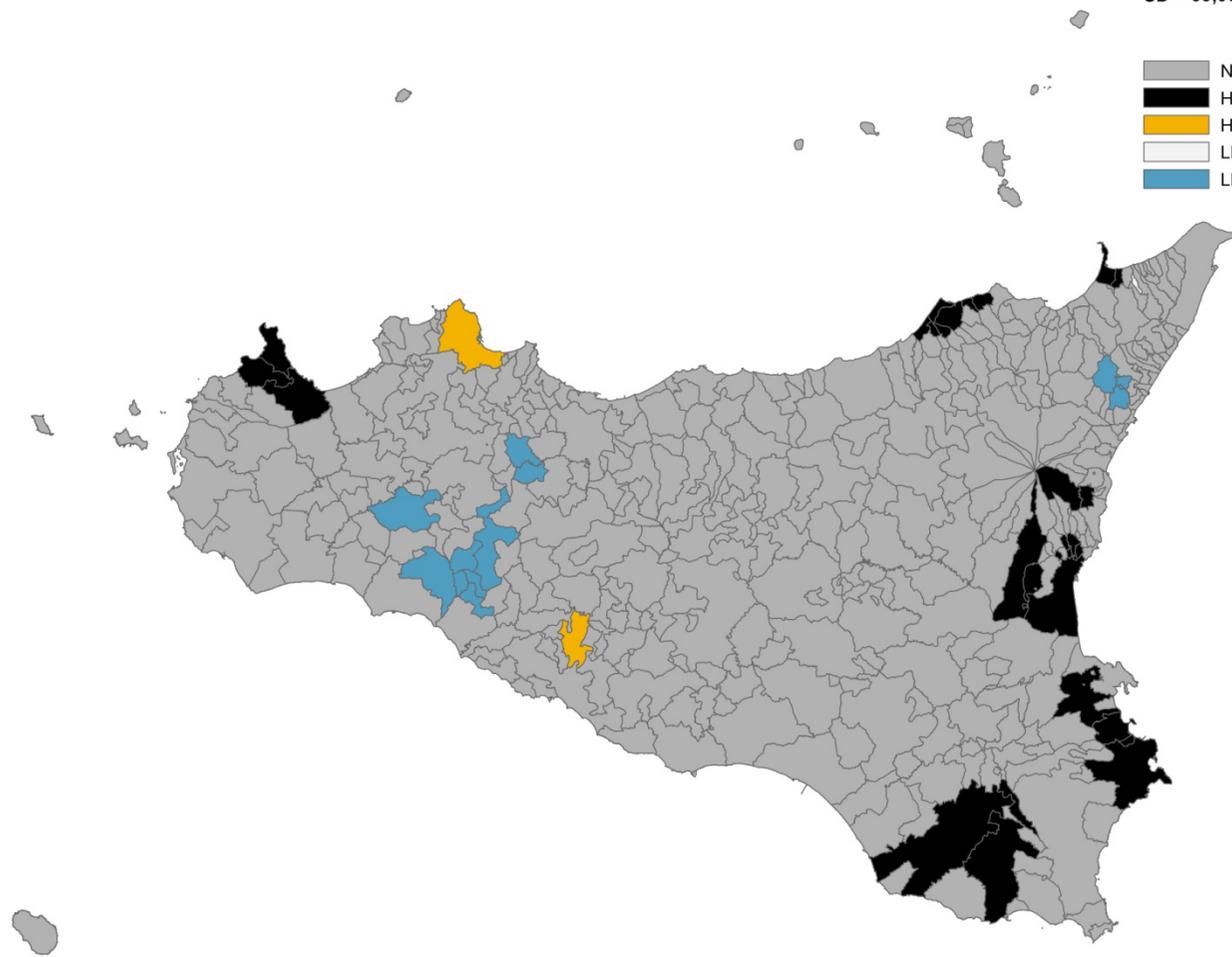


# N° di addetti Impresa /Pop.

Analisi spaziale locale

Moran's Index: = 0,24  
Mean = 109,66  
SD = 53,00

Not Significant  
HH  
HL  
LH  
LL

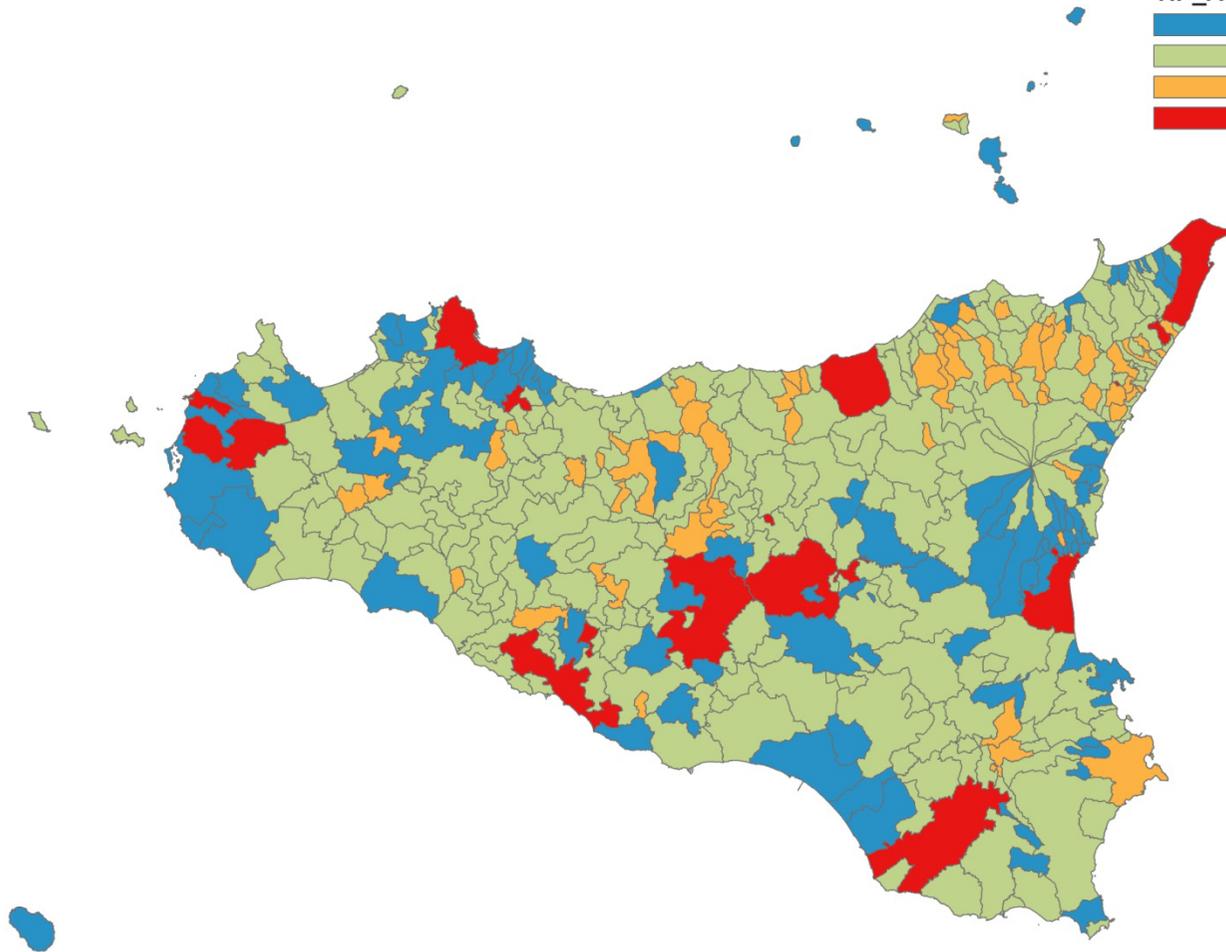
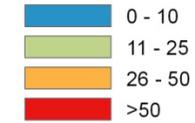


# N° di addetti Ist. Pubblica /Pop.

Analisi spaziale globale

Moran's Index: = -0,02\*  
Mean = 18,54  
SD =13,99  
\* =n.s.

TIP\_ADDET1

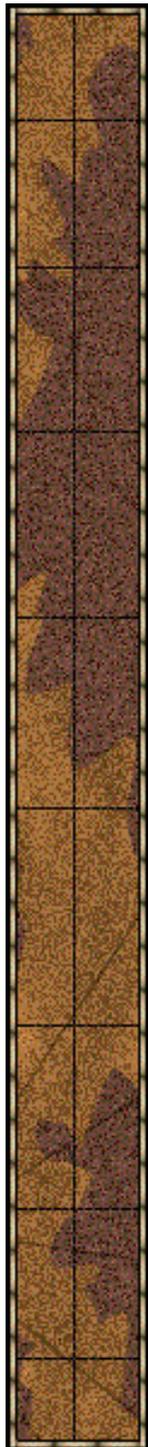
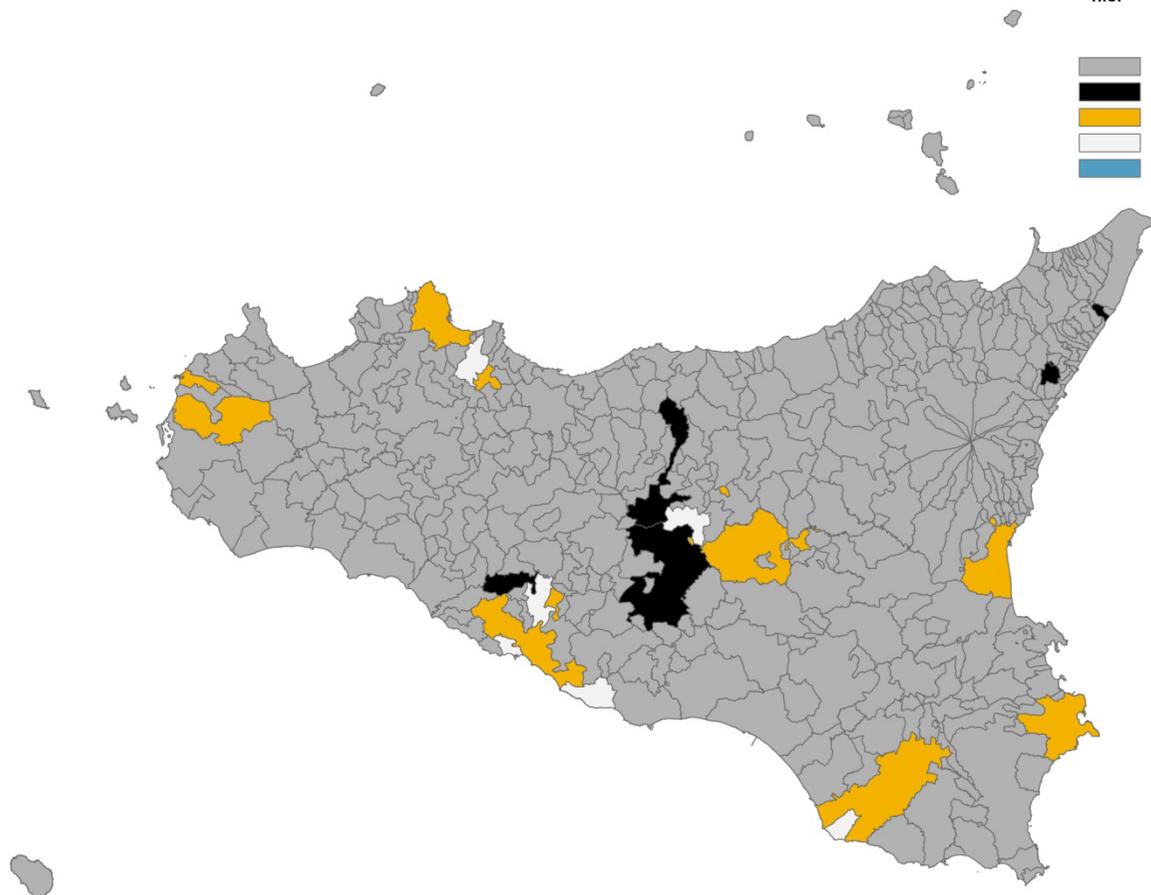


# N° di addetti Ist. Pubblica /Pop.

Analisi spaziale locale

Moran's Index: = -0,02\*  
Mean = 18,54  
SD = 13,99  
\*=n.s.

Not Significant  
HH  
HL  
LH  
LL

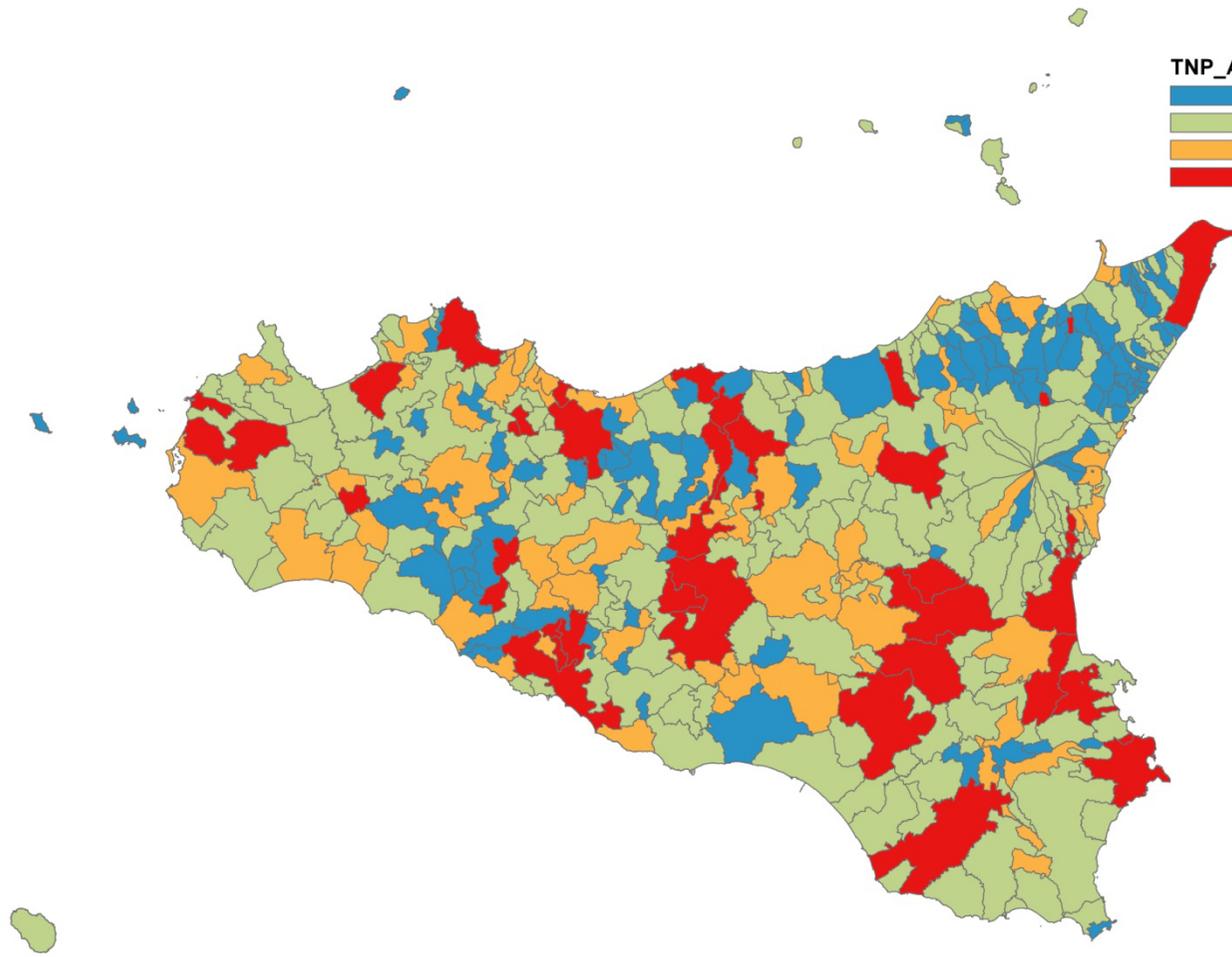
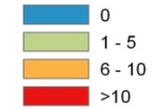


# N° di addetti Ist. No profit /Pop.

Analisi spaziale globale

Moran's Index: = 0,02\*  
Mean = 4,09  
SD = 7,44  
\*=n.s.

TNP\_ADDET1

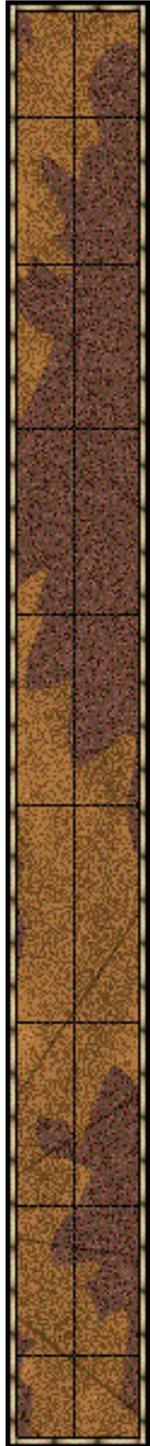
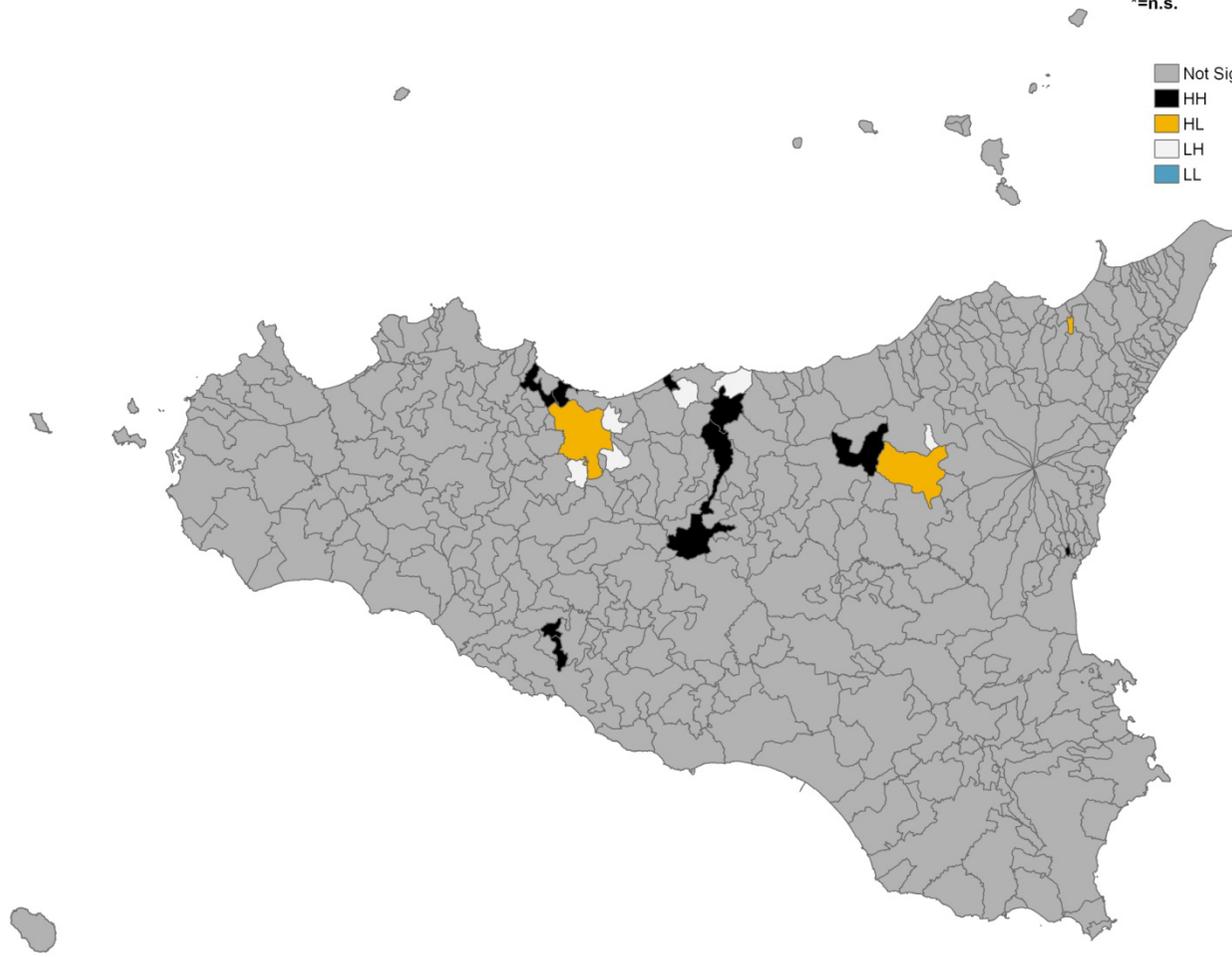


# N° di addetti Ist. No profit /Pop.

Analisi spaziale locale

Moran's Index: = 0,02\*  
Mean = 4,09  
SD = 7,44  
\*n.s.

Not Significant  
HH  
HL  
LH  
LL



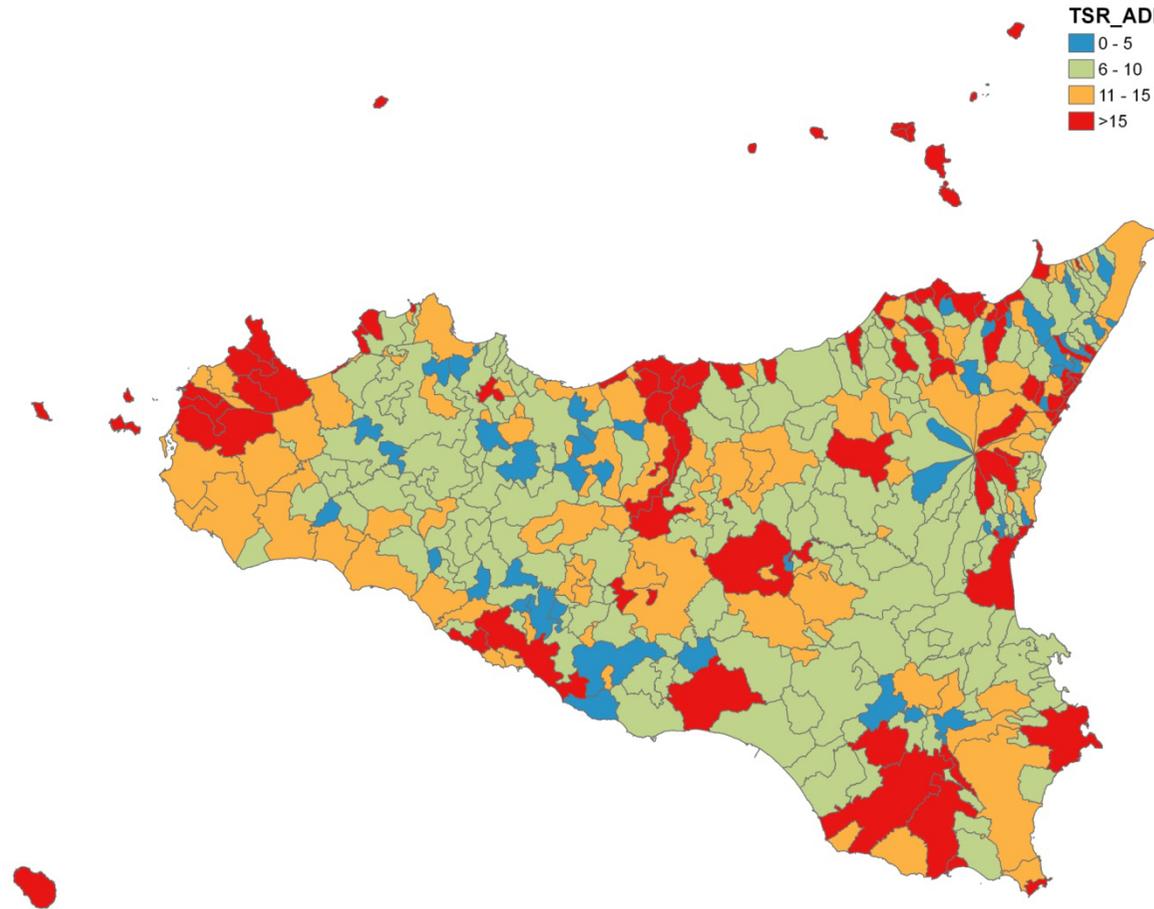
# N° di addetti Attività dei servizi di alloggio e di ristorazione /Pop.

## Analisi spaziale globale

Moran's Index: = 0,27  
Mean = 10,95  
SD = 8,05

TSR\_ADDET1

- 0 - 5
- 6 - 10
- 11 - 15
- >15

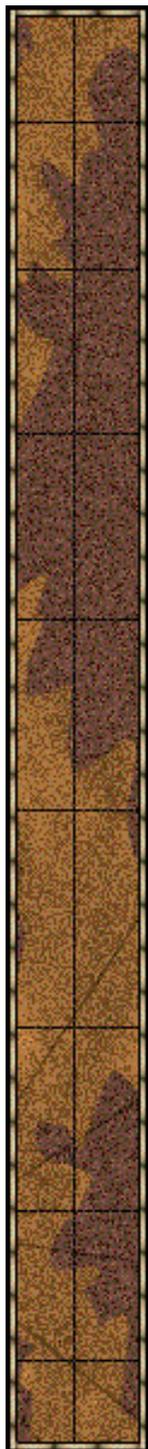
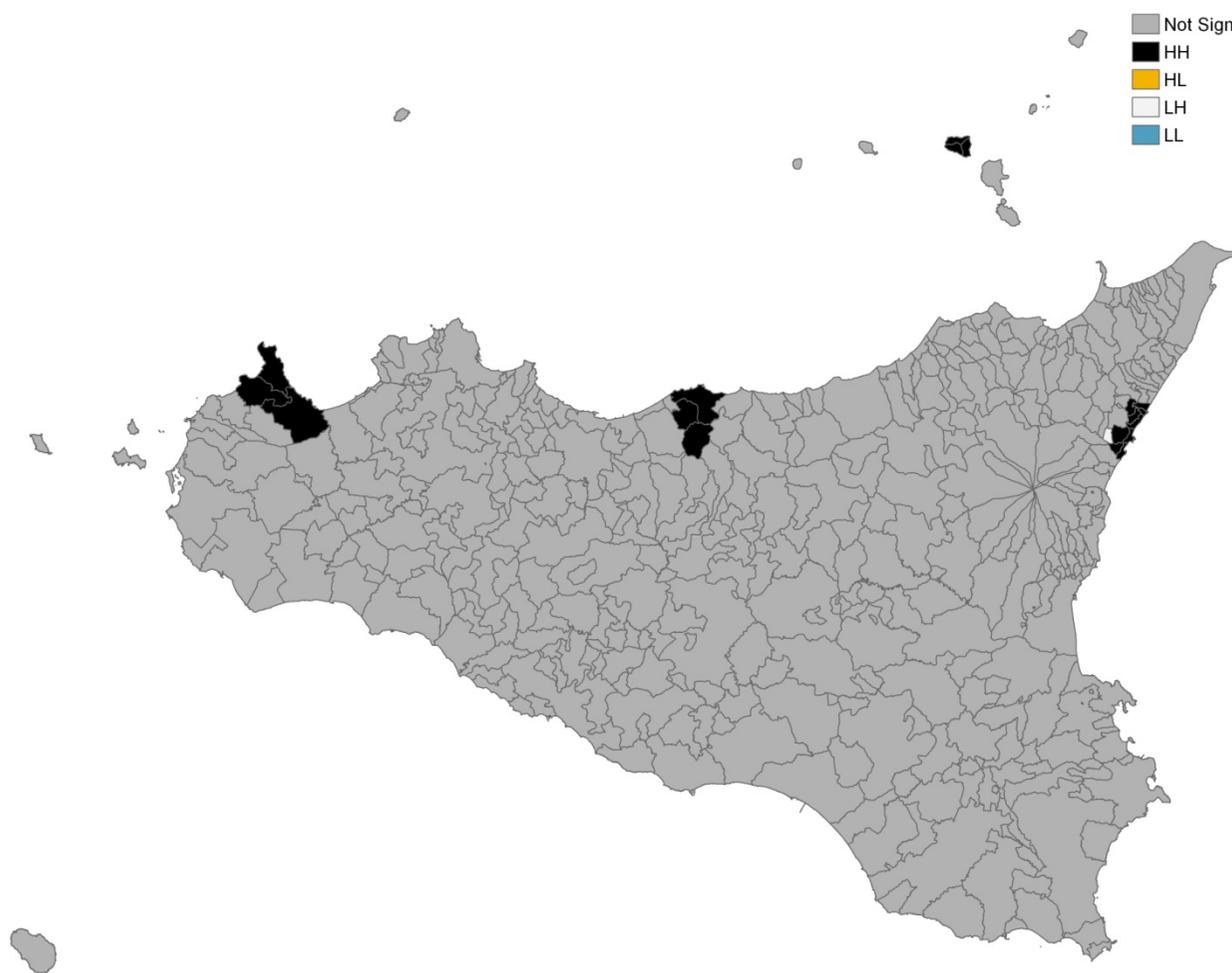


# N° di addetti Attività dei servizi di alloggio e di ristorazione /Pop.

## Analisi spaziale locale

Moran's Index: = 0,27  
Mean = 10,95  
SD = 8,05

- Not Significant
- HH
- HL
- LH
- LL

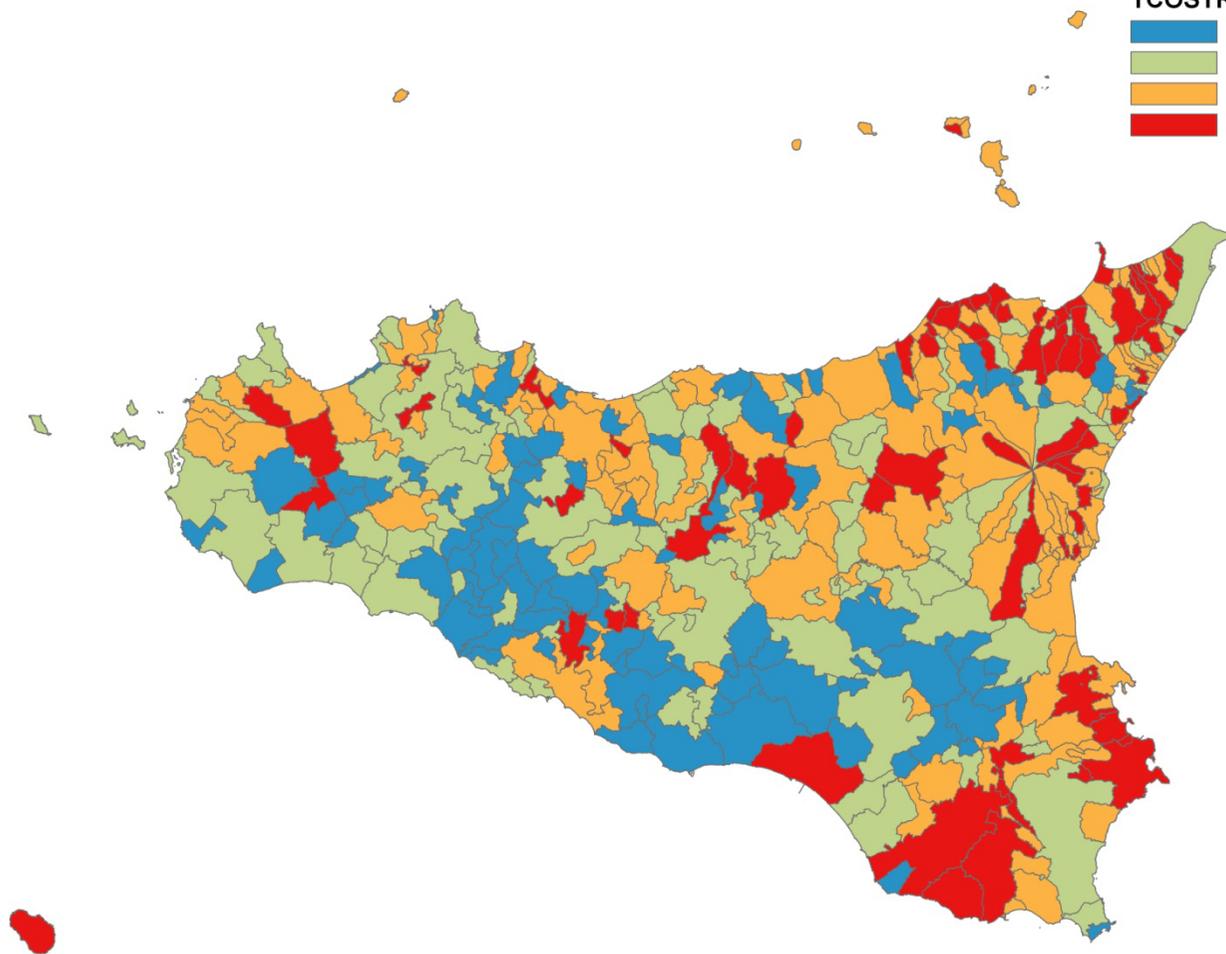
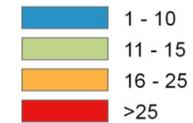


# N° di addetti Costruzioni /Pop.

Analisi spaziale globale

Moran's Index: = 0,17  
Mean = 18,48  
SD =14,02

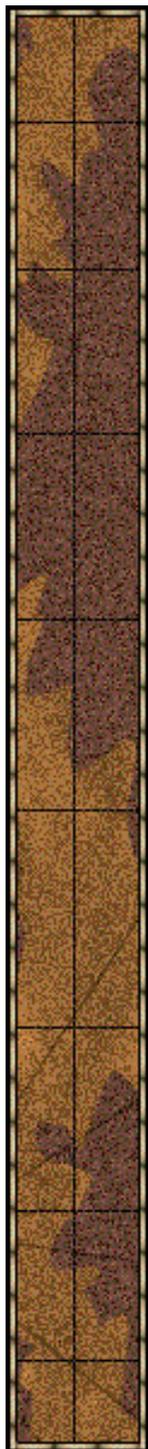
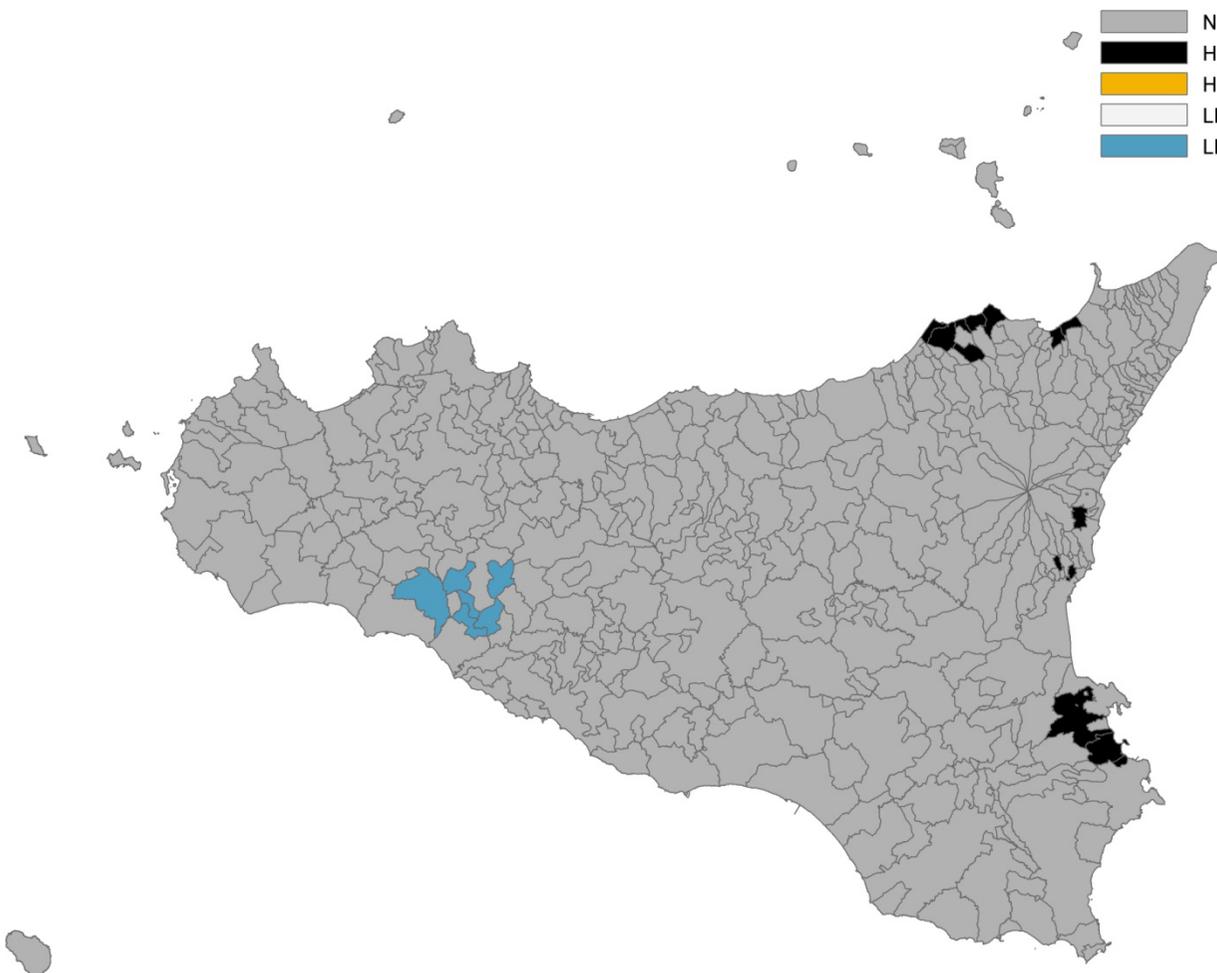
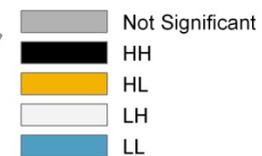
TCOSTR\_AD1



# N° di addetti Costruzioni /Pop.

Analisi spaziale locale

Moran's Index: = 0,17  
Mean = 18,48  
SD =14,02

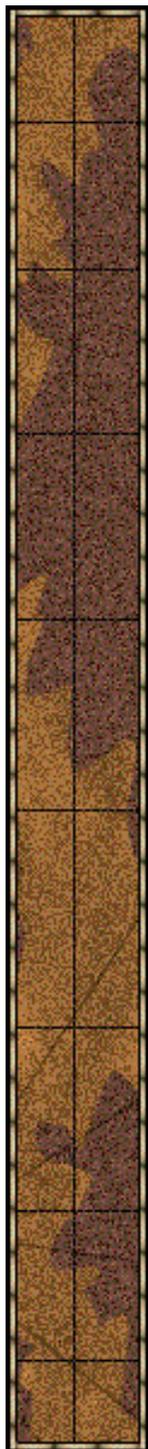
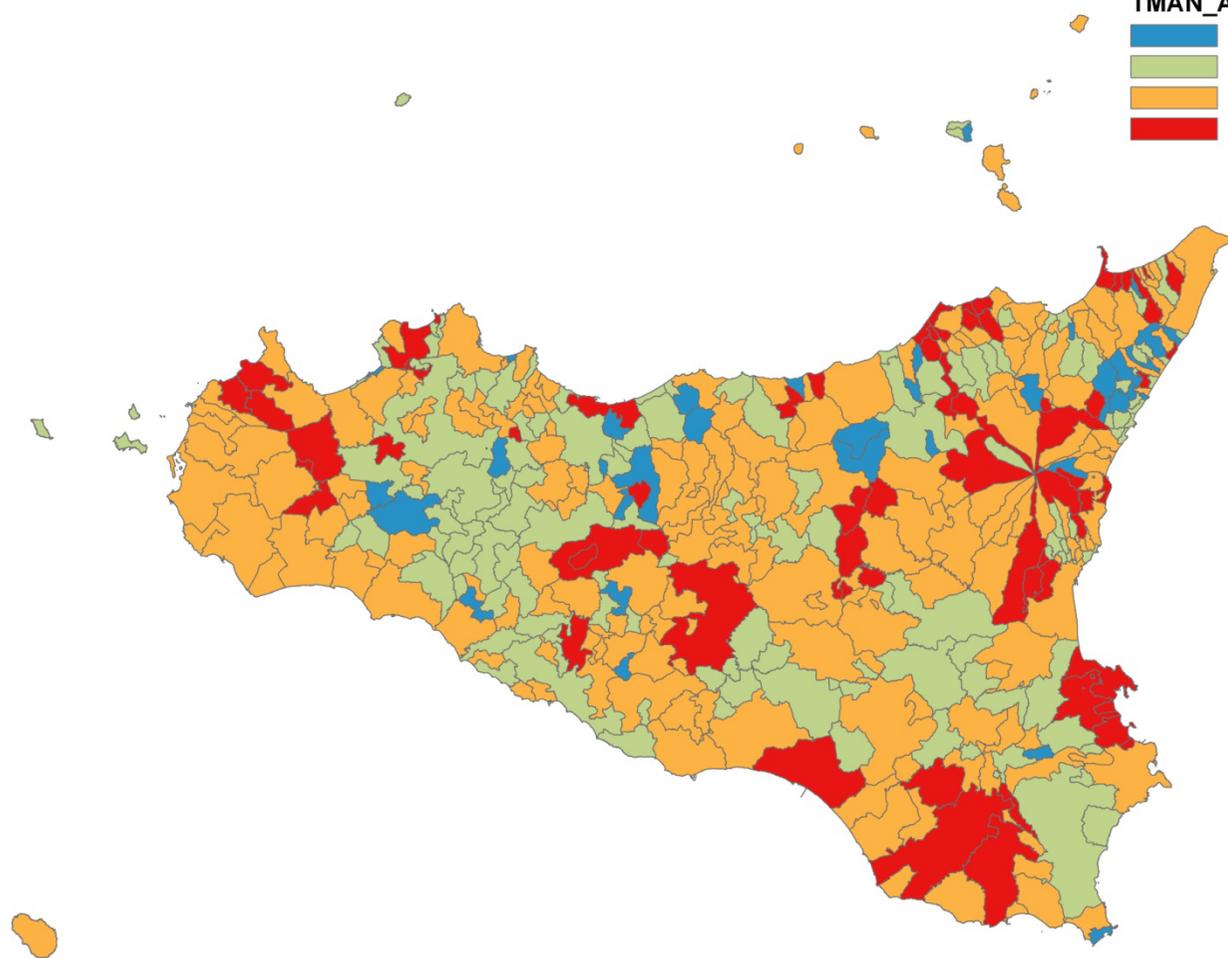
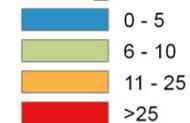


# N° di addetti attività manifatturiere /Pop.

Analisi spaziale globale

Moran's Index: = 0,13  
Mean = 17,09  
SD =19,56

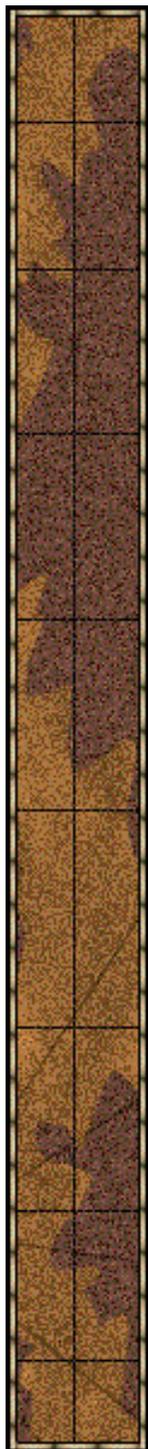
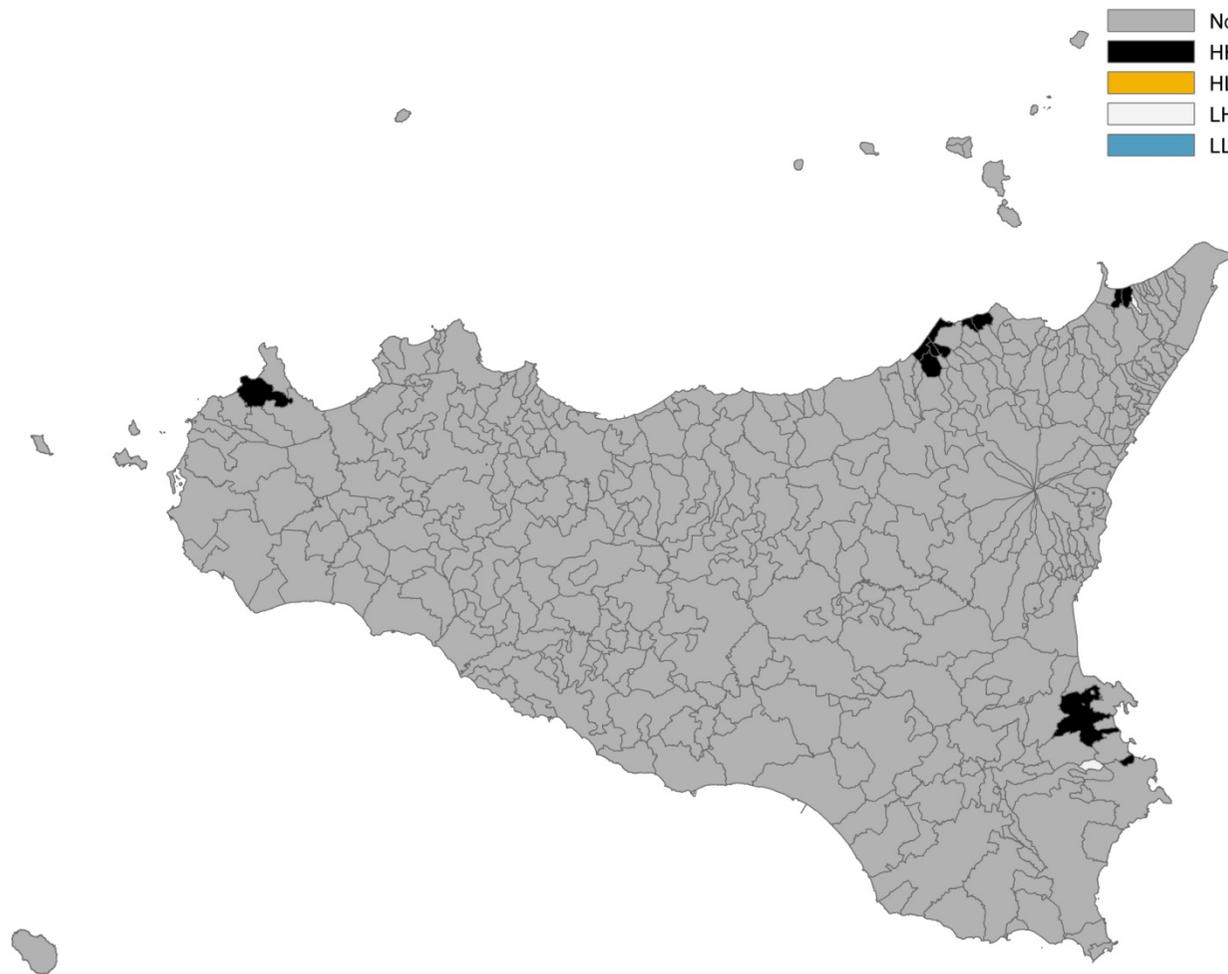
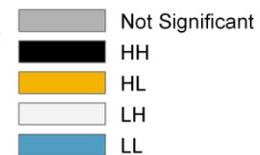
TMAN\_ADDE1



# N° di addetti attività manifatturiere /Pop.

Analisi spaziale locale

Moran's Index: = 0,13  
Mean = 17,09  
SD =19,56



# Principali riferimenti scientifici

- L. Anselin, "Local indicators of spatial association-LISA", *Geographical Analysis*, No. 27, 1995, pp. 93–115.
- L. Anselin, "The Moran scatterplot as an ESDA tool to assess local instability in spatial association", Fisher M, Scholten HJ, Unwin D (eds) *Spatial analytical perspectives on GIS*, Taylor&Francis, London, 1996.
- F. Chen, C.Tien Lu, A.P.Boedihardjo, "GLS-SOD: a generalized local statistical approach for spatial outlier detection", in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, USA, 2010
- J. Cliff, K. Ord, "Spatial Autocorrelation", Pion Limited, London, 1973.
- E. Fix, J.L. Hodges, "Discriminatory analysis. Nonparametric discrimination: Consistency properties", Technical Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
- A. S. Fotheringham, C. Brunson, M. Charlton, "Geographically Weighted Regression", J Wiley & Son, 2002.
- M. Mucciardi, "Use of a flexible weight matrix in a local spatial statistic", *Classification and Multivariate Analysis for Complex Data Structures*, Series: Studies in Classification, Data Analysis, and Knowledge Organization, 1st Edition, Springer, 2011, Vol. IX, pp. 427-434.
- M. Mucciardi P. Bertuccelli (2007) S-Joint: a new software for the analysis of spatial data, in: *Atti della Riunione Scientifica Intermedia della Società Italiana di Statistica*, Venezia.
- E. Otranto M. Mucciardi P. Bertuccelli (2014), *Spatial Effects in Dynamic Conditional Correlations*, Working Paper CRENoS;
- J. K. Ord, A. Getis, "Local spatial autocorrelation statistics: distributional issues and an application", *Geographical Analysis*, No. 27, 1995, pp. 286–305.
- T. Plümper, E. Neumayer, "Model Specification in the Analysis of Spatial Dependence" *European Journal of Political Research*, Vol. 49, Issue 3, 2010, pp 418–442.
- A. Unwin, D. Unwin, "Exploratory Spatial Data Analysis With Local Statistics", *The Statistician*, Vol. 47, No. 3, 1998, pp. 415-421.

