

18 giugno 2024

*ELEMENTI DI GEO-STATISTICA E
MICROSIMULAZIONE SPAZIALE*

Analisi subcomunale e georeferenziazione dei comportamenti di risposta al censimento permanente della popolazione

Premessa

- Lavoro nato da una richiesta di supporto dell'amministrazione locale (Genova)
- Obiettivo: analisi territoriale degli esiti di risposta, in particolare per verificare caratteristiche e posizionamento sul territorio delle famiglie non rispondenti ed organizzare in modo più efficiente la fase di recupero delle mancate risposte per le prossime edizioni
- Risultati preliminari (con R.Succi e A.Marino)
- In questa sessione focus ed esempi sono sulla rappresentazione grafica e sulla geocodifica dei dati

Due parole sul censimento

RILEVAZIONE DA LISTA (2022/23)

- campione di comuni
- campione di famiglie
- **restituzione multi-canale:** nella prima fase la famiglia **compila il questionario in autonomia (CAWI)** o, recandosi presso il Centro Comunale di Rilevazione (CCR), con il supporto di operatori comunali; il **rilevatore e gli operatori di back office** intervengono nella fase di recupero delle mancate risposte

Oggi però vedremo non i risultati del censimento, ma i **comportamenti di risposta al censimento**

RESTITUZIONE SPONTANEA
Compilazione via web in autonomia/con il supporto del CCR

RECUPERO MANCATE RISPOSTE (CATI/CAPI)

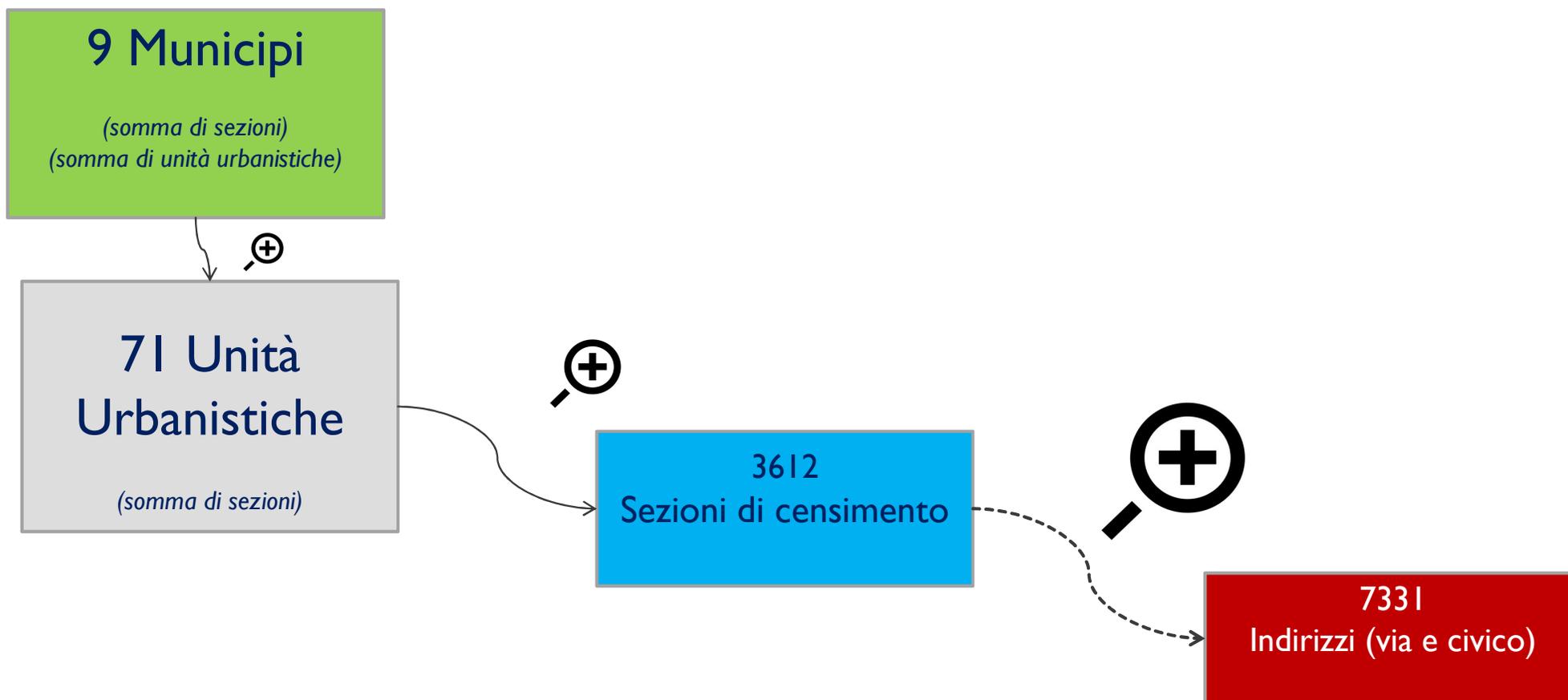
Aree subcomunali

Esempi di zonizzazione del territorio al di sotto del livello comunale:

- **Sezioni di censimento** (versione 2011 e 2021)
- **Basi Territoriali** (Località: centri abitati, nuclei abitati, case sparse)
- **Aree Subcomunali di I livello:** Municipi (Milano, Genova, Roma, Bari), Circoscrizioni (Torino, Catania), Municipalità (Napoli), Quartieri (Firenze)...
- **Altre:** Nuclei di Identità Locale (NIL), Unità urbanistiche, Zone urbanistiche, ACE...

Nota: le aree più grandi sono ricostruibili dal basso partendo dalle sezioni di censimento (aree minime)

Diversi livelli di dettaglio o 'zoom'

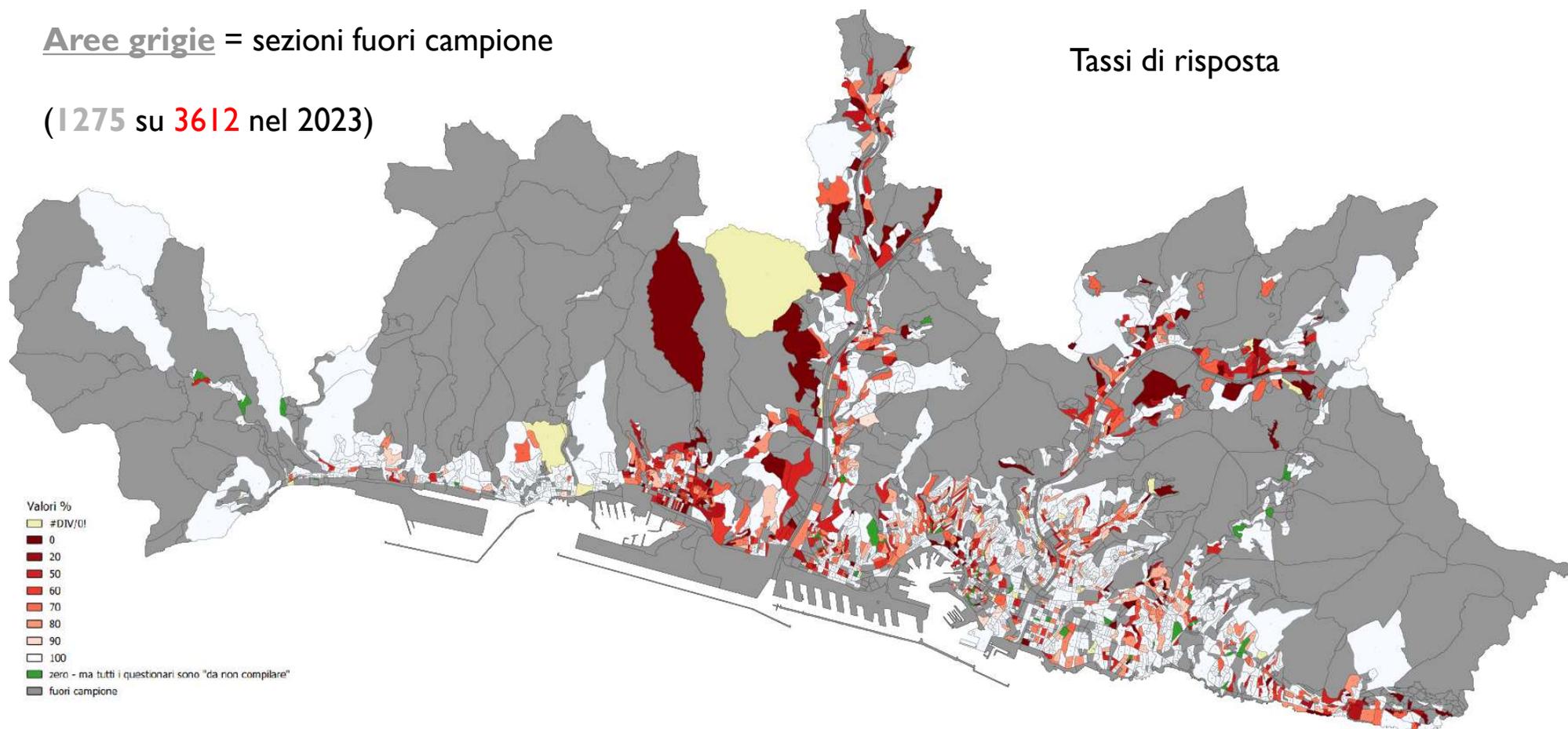


Si potrebbe partire dalle sezioni di censimento ma...

Aree grigie = sezioni fuori campione

(1275 su 3612 nel 2023)

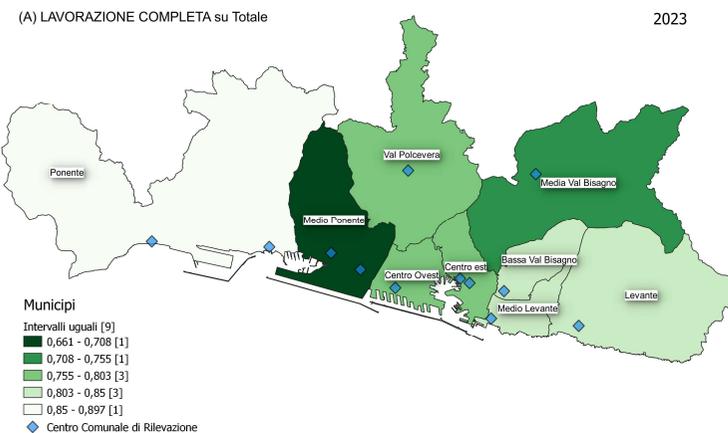
Tassi di risposta



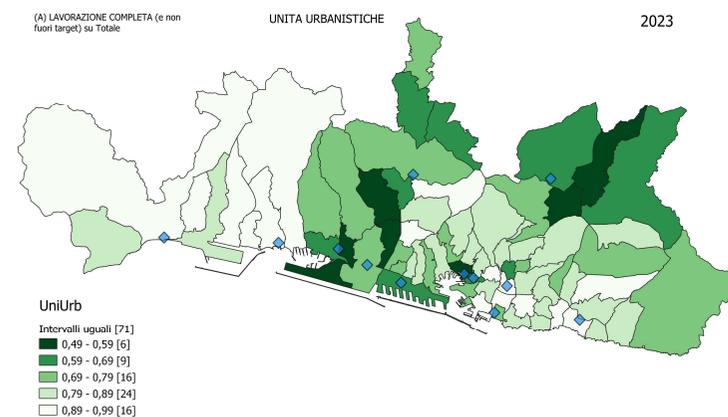
... meglio zoomare a livelli diversi /1

per rappresentare i tassi di risposta e gli esiti...

9 Municipi



71 Unità Urbanistiche



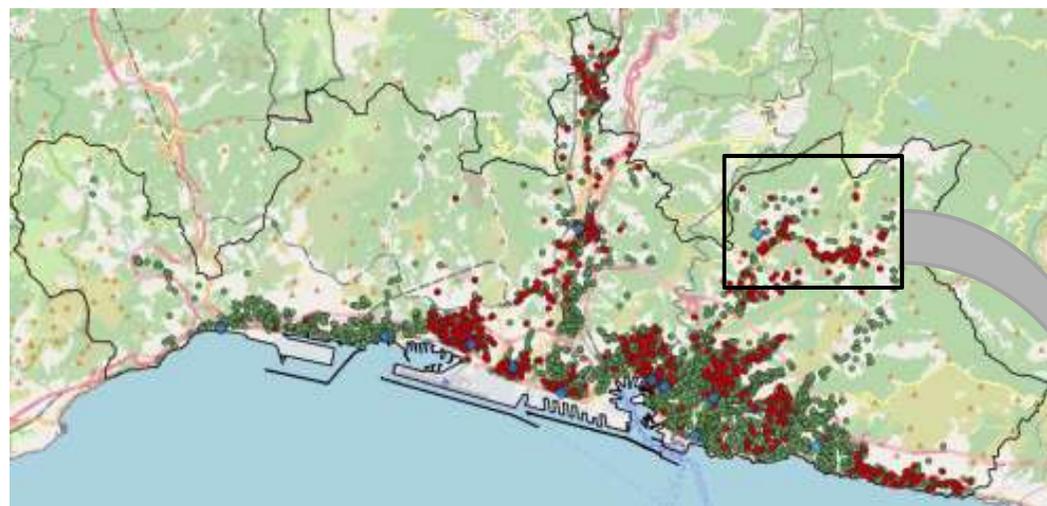
... meglio zoomare a livelli diversi /2

...e per cercare cluster di risposta sul territorio

(mappa con distribuzione di punti tramite geocodifica degli indirizzi)

7331

Questionari



Georeferenziazione / Geocodifica

In italiano sono usati sostanzialmente in modo intercambiabile, in inglese tendono a distinguerli

Georeferenziazione (caso più generale): prendo un qualunque dato (es. un record di un database, un set di immagini, una mappa digitalizzata) e gli associo un'informazione relativa alla posizione geografica, espressa in un sistema di riferimento

Geocodifica: prendo un indirizzo stradale (dato testuale) e gli associo coordinate geografiche (Latitudine e Longitudine)

Invece

Geolocalizzazione: è termine specifico per l'azione 'sul campo'. Con l'aiuto di strumenti elettronici e informatici (GPS, reti cellulari, satelliti, wi-fi) vado a determinare la posizione geografica di un oggetto, una persona, un edificio.

Geocodifica: procedura (workflow)

Ovvero: come mettiamo i punti sulla mappa ?



E' un processo iterativo, soprattutto se si tratta di una geocodifica massiva (migliaia di richieste al server o più)

Geocodifica 1: normalizzazione e formattazione

- Elementi base di un indirizzo: DUG (via, piazza) | toponimo (nome della strada) | numero civico | esponente
- Normalizzazione: correzione di eventuali inesattezze, incoerenze e ambiguità nella grafia degli indirizzi (es. apostrofi, accenti, caratteri spurii, abbreviazioni, numeri romani, spaziature...)
- Formattazione deve seguire regole del software o sistema di geocodifica e dunque prevale su tutto: quanti campi, separati da quale separatore, presenza obbligatoria di città/regione/stato o meno, civico prima o dopo il toponimo (es. via Roma 15 A o 15A, *via Roma*) e così via
- Interventi correttivi: 'a mano' (es. Excel) o con software ad hoc (es. Egon)

Geocodifica 2: gli strumenti

- Risorse interne: enti pubblici possono già disporre di database di luoghi e numeri civici geolocalizzati in proprio, o tramite società esterne ma avendo la titolarità del database prodotto
- Risorse esterne: due approcci e due 'ecosistemi'

Google Maps, Tele Atlas, Navtech, Microsoft e altri

- Precisione nei risultati, prestazioni
- Accessi gratuiti ma fino a determinate soglie: l'utilizzo sopra soglia (es. 40.000 accessi al mese; max 2500 al giorno per 30 giorni consecutivi) comporta dei costi vivi (abbonamenti e tariffe)
- Uso prevalentemente o esclusivamente online e nel cloud (-> privacy e GDPR; effetti run-time sulla riproducibilità dei risultati)
- Supporto commerciale e da community utenti

Openstreetmap/Nominatim

- Precisione dei risultati varia in funzione della copertura
- Sempre gratuiti, ma con accessi online limitati (es. 1 richiesta al secondo): oltre soglia può scattare il blocco (blacklist) dell'indirizzo IP
- Usabile anche offline, senza limitazioni: tutto il database e il software sono istanziabili localmente (-> privacy e GDPR; controllo più granulare del processo e riproducibilità dei risultati)
- Supporto da community utenti o misto
- Open-source

OpenStreetMap e Nominatim

www.openstreetmap.org

Progetto collaborativo (*crowd sourced*) finalizzato a creare una banca dati geografica di tutto il pianeta

- convoglia dati, contributi e verifiche da più fonti (cittadini, pubbliche amministrazioni, aziende private)
- open source, open data
- archivi scaricabili e riusabili localmente

www.nominatim.org

Motore di ricerca di OpenStreetMap

- Parte dall'indirizzo per arrivare alle coordinate geografiche
- (ma anche viceversa)
- software installabile localmente

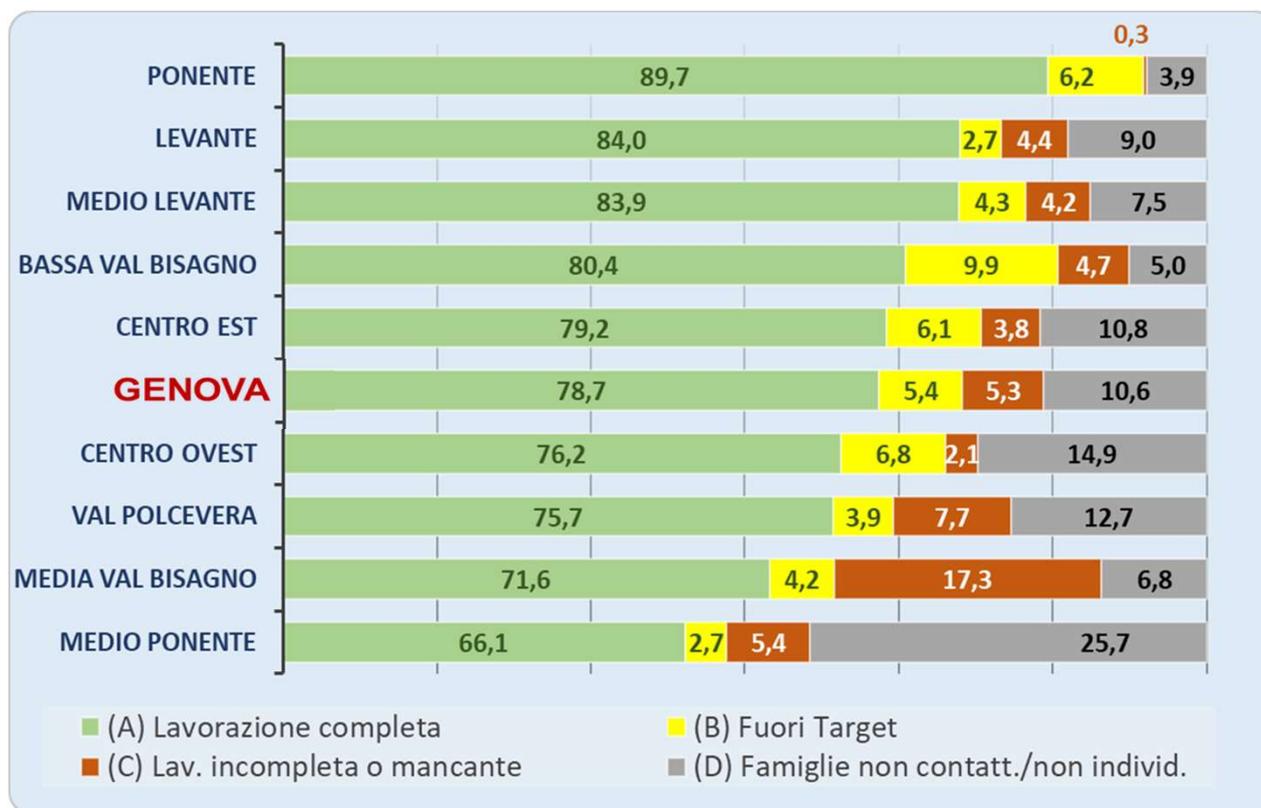
Geocodifica 3: controllo qualità

- Quali e quanti indirizzi non sono stati riconosciuti o sono stati respinti?
 - Se il sistema segnala un errore specifico tutto bene, si corregge e si recuperano i respinti
- Quali e quanti indirizzi sono stati posizionati, sì, ma nel posto sbagliato?
 - si possono applicare filtri e vincoli sulla mappa risultante (tipicamente usando software come ArcGIS/QGIS) per vedere se ci sono punti che cadono fuori dai confini amministrativi previsti - capita ad esempio tra comuni limitrofi che sono attraversati dalla stessa strada, o che hanno strade omonime
 - un buon controllo è vedere quali/quantità punti hanno ottenuto coordinate identiche: in tal caso può voler dire che il civico esatto non era presente ed è stato approssimato; a seconda del grado di precisione cercato e degli obiettivi, è possibile 'deduplicarlo' ad esempio distribuendo i punti in un intorno, sulla stessa via
 - confronto dei risultati con altri sistemi di geocodifica, a campione o sistematici
 - ricerca di outliers con metodi statistici

*...e tutto questo solo per
posizionare dei puntini su una
mappa!*



Alcuni risultati preliminari



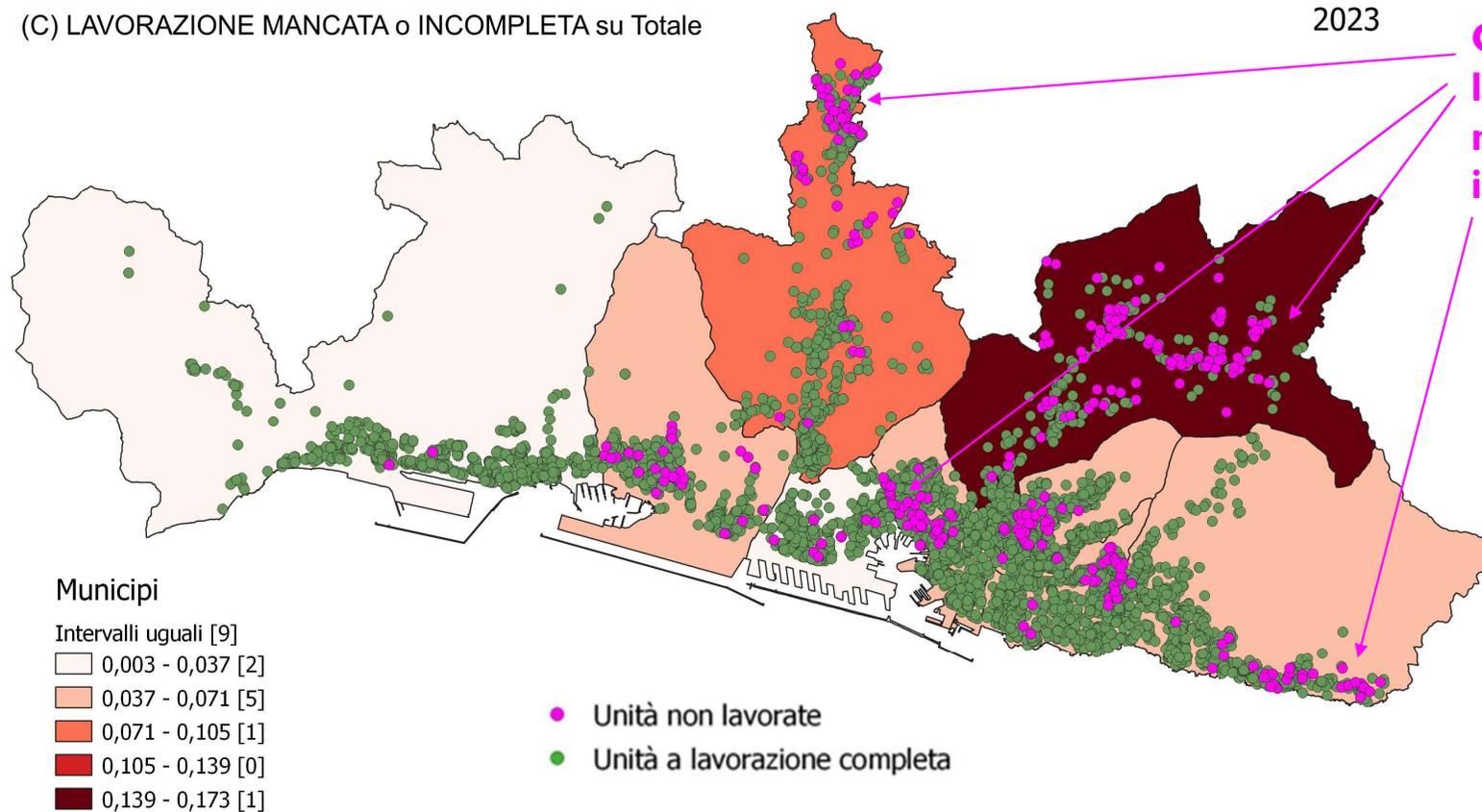
Unità di rilevazione per macro-esiti e municipi del comune di Genova (valori percentuali) - Indagine campionaria da Lista, anno 2023 (n.oss.=7331)

Alcuni risultati preliminari

(C) LAVORAZIONE MANCATA o INCOMPLETA su Totale

2023

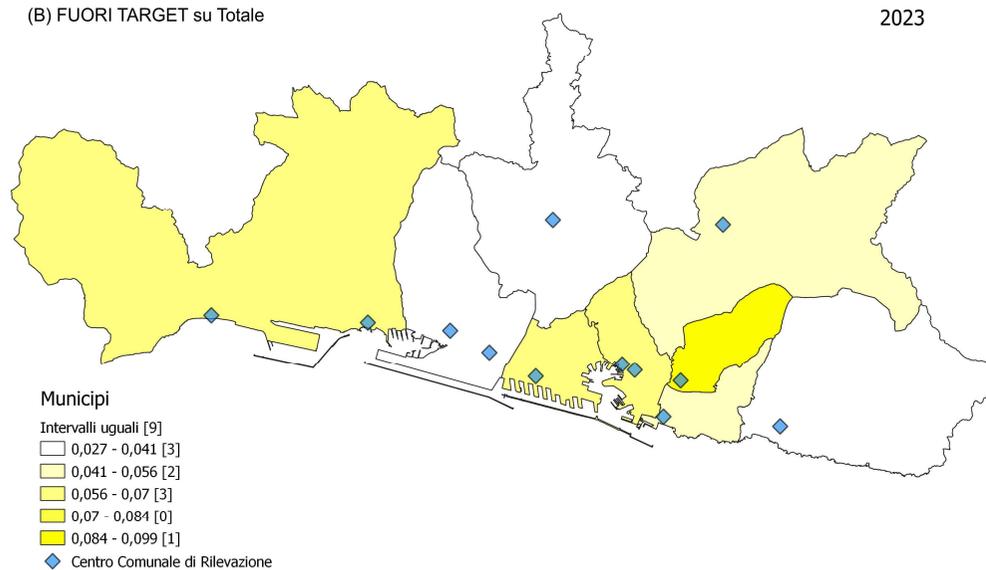
Cluster di unità a lavorazione mancante/incompleta



Alcuni risultati preliminari

(B) FUORI TARGET su Totale

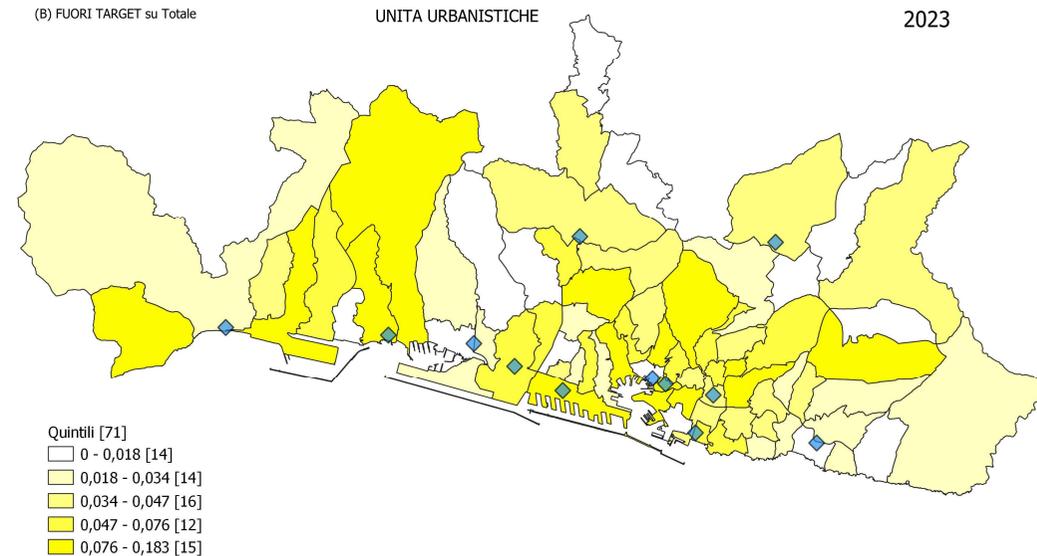
2023



(B) FUORI TARGET su Totale

UNITA URBANISTICHE

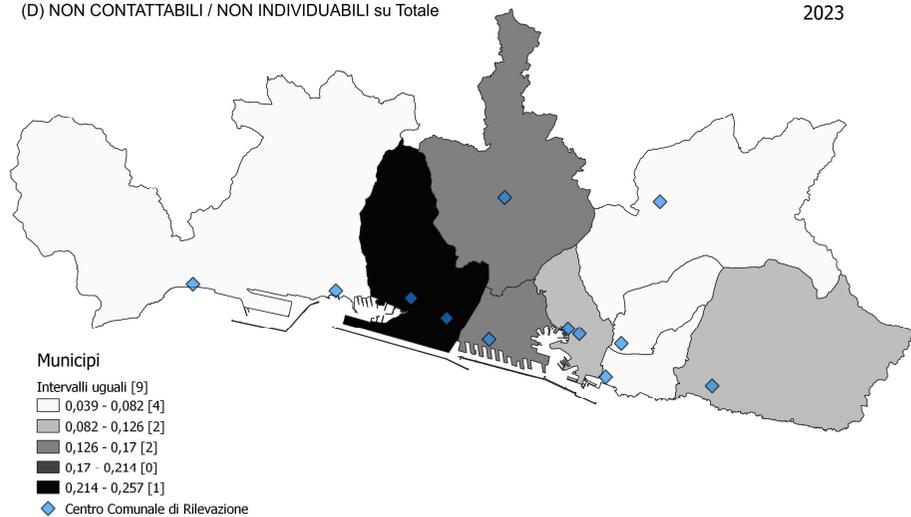
2023



Alcuni risultati preliminari

(D) NON CONTATTABILI / NON INDIVIDUABILI su Totale

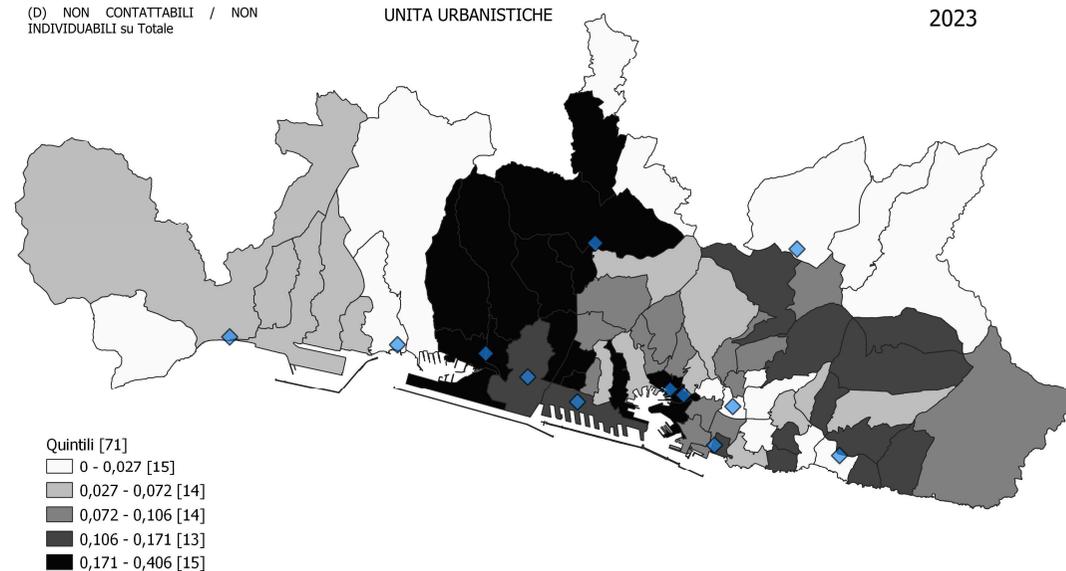
2023



(D) NON CONTATTABILI / NON INDIVIDUABILI su Totale

UNITA URBANISTICHE

2023

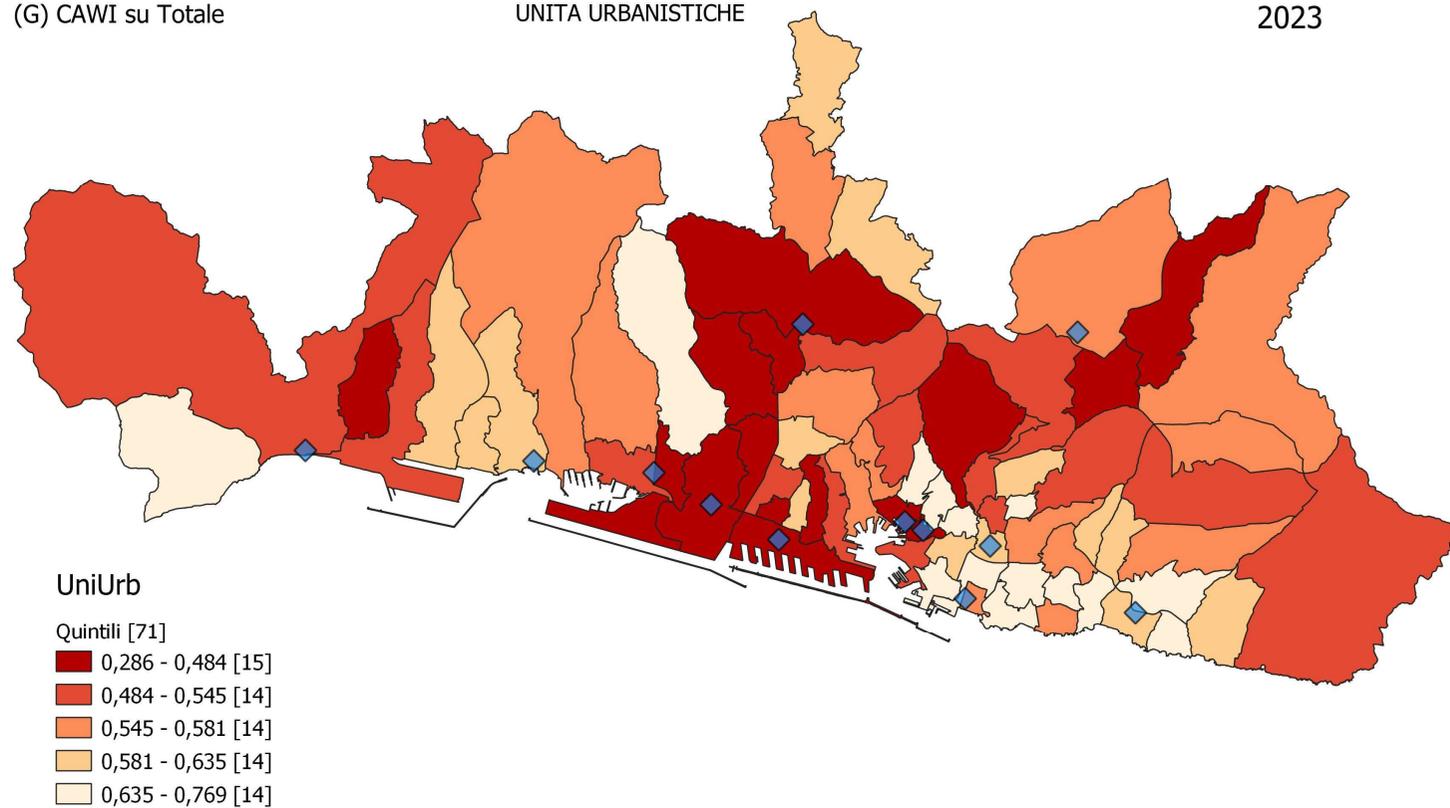


Alcuni risultati preliminari

(G) CAWI su Totale

UNITA URBANISTICHE

2023

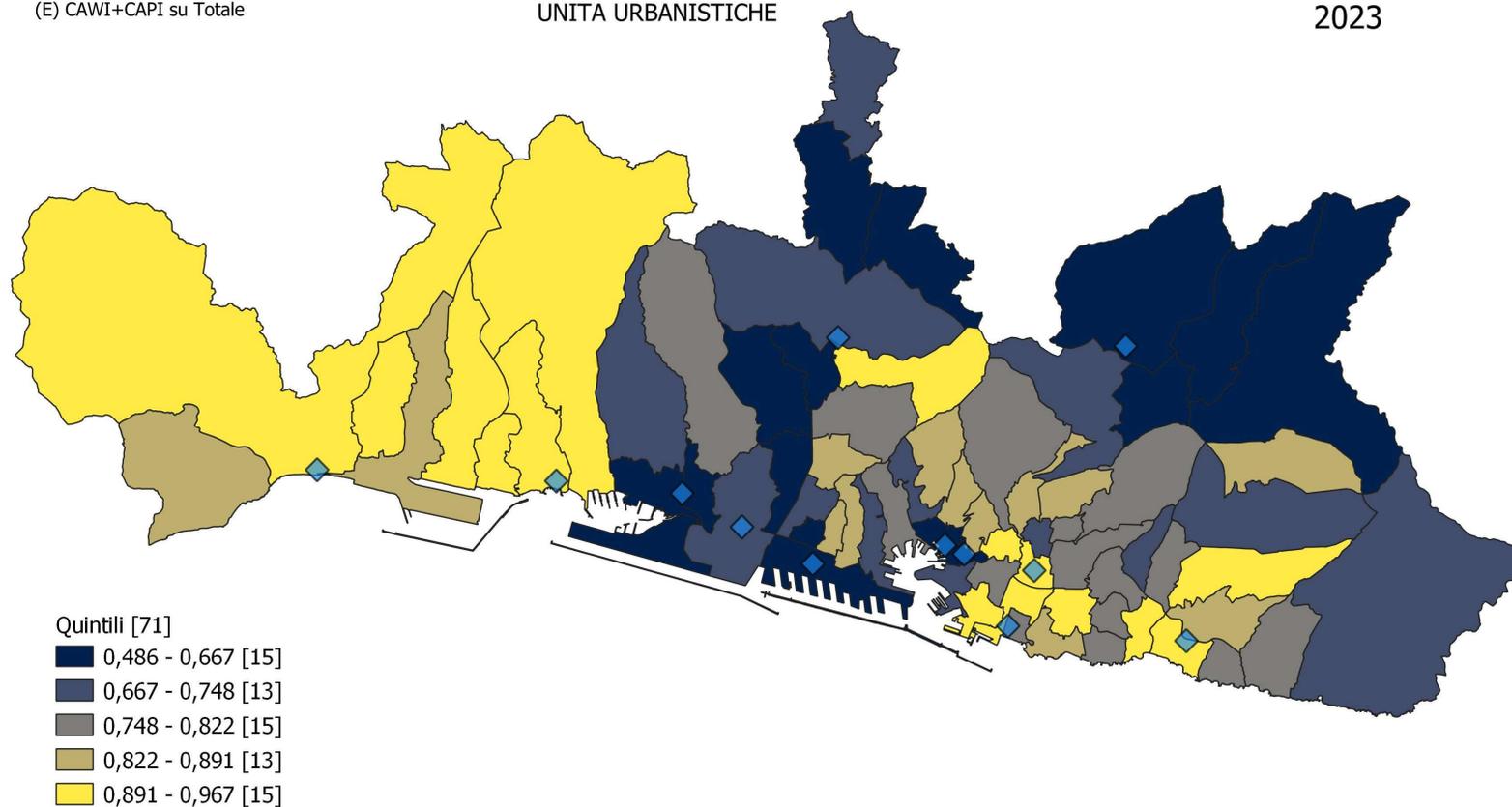


Alcuni risultati preliminari

(E) CAWI+CAPI su Totale

UNITA URBANISTICHE

2023



Alcuni risultati preliminari

Caratteristiche demo-sociali, macro-esiti e risposte CAWI all'indagine da Lista nel Comune di Genova (anni 2022-23) : risultati da stime multinomiali e logit (n.oss. = 14.639)

	MODELLO MULTINOMIALE per macro-esiti			MODELLO LOGIT per risp. CAWI (vs. altri esiti)	
	Fuori target	Non contattabili (cat. rif. = Complete)	Incomplete		
Profilo intestatario					
Età	-0.25 (-1.60)	-0.46*** (-5.16)	-0.13 (-1.13)	-0.60*** (-10.15)	
Nato all'estero	1.37*** (14.12)	0.95*** (12.58)	1.09*** (1.37)	-1.33*** (-21.80)	
Profilo sez. cens.					
Pop. 0-14	-3.24 (-1.61)	-0.34 (-0.26)	-2.39 (-1.46)	0.47 (-0.59)	
Pop. età 65+	-1.79 (-1.63)	-2.30*** (-2.97)	-2.71*** (-3.03)	2.24*** (4.80)	
Pop. bassa istruz.	0.61 (0.85)	2.11*** (3.58)	0.87 (1.13)	-2.22*** (6.59)	
Tasso occupazione	-0.17 (-0.13)	-0.23 (-0.27)	-1.84 (-1.64)	1.13** (2.05)	
Pop. extra UE	1.50** (2.13)	0.01 (0.02)	-1.73 (-1.16)	-0.65* (-1.98)	
Fam. >= 5 comp.	0.26 (2.11)	-0.48 (-0.26)	-0.47 (-0.22)	-0.61 (-0.55)	
Fam. <=2 comp.	2.82 (1.61)	3.23*** (3.41)	0.79 (0.68)	-2.85*** (-4.67)	
Fam. in affitto	-1.36*** (-2.88)	1.23*** (3.96)	0.53 (1.22)	-0.42** (-2.05)	
Case sparse	0.40 (0.72)	0.54 (1.63)	1.24*** (3.87)	-0.60*** (-2.63)	
Municipi (rif. Centro Est)					
Centro Ovest	-0.12 (-0.71)	-0.02 (-0.13)	-0.46** (-2.29)	-0.04 (-0.49)	
Bassa Val Bisagno	0.28** (1.97)	-0.83*** (-6.48)	-0.004 (-0.02)	0.08 (1.08)	
Media Val Bisagno	-0.13 (-0.66)	-0.23* (-1.70)	1.22*** (7.62)	-0.08 (-1.02)	
Val Polcevera	-0.37* (-1.89)	-0.08 (-0.63)	0.58*** (3.34)	0.05 (0.66)	
Medio Ponente	-0.49** (2.41)	0.30** (2.51)	1.68*** (10.35)	-0.04 (-0.44)	
Ponente	0.07 (0.039)	-0.73*** (-5.10)	-1.16*** (-4.64)	0.02 (0.22)	
Medio Levante	0.24 (1.51)	0.09 (0.77)	-0.36* (-1.74)	-0.12* (-1.70)	
Levante	0.10 (0.63)	0.10 (0.08)	-0.02 (-0.10)	-0.03 (-0.42)	

Nb Non riportate variabili di periodo e costante. Z-scores in parentesi. *: pval.<0.1; **: pval.<0.05; ***: pval.<0.01

Fonte: ns elab. su dati del Sistema di gestione dell'indagine anni 2022 e 2023 e variabili censuarie degli anni 2011 e 2021 per sez di cens. 2011.

Il territorio di Genova si caratterizza per un'elevata eterogeneità socio-demografica e ambientale (cfr. Istat 2017). Utilizzando strumenti di statistica spaziale e un modello logistico multinomiale, lo studio analizza per aree sub-comunali il modo in cui tali fattori hanno influenzato i profili di risposta (CAWI/CAPI) e gli esiti della mancata risposta nelle edizioni 2022 e 2023 dell'indagine campionaria da Lista del Censimento Permanente della Popolazione.

Fonti

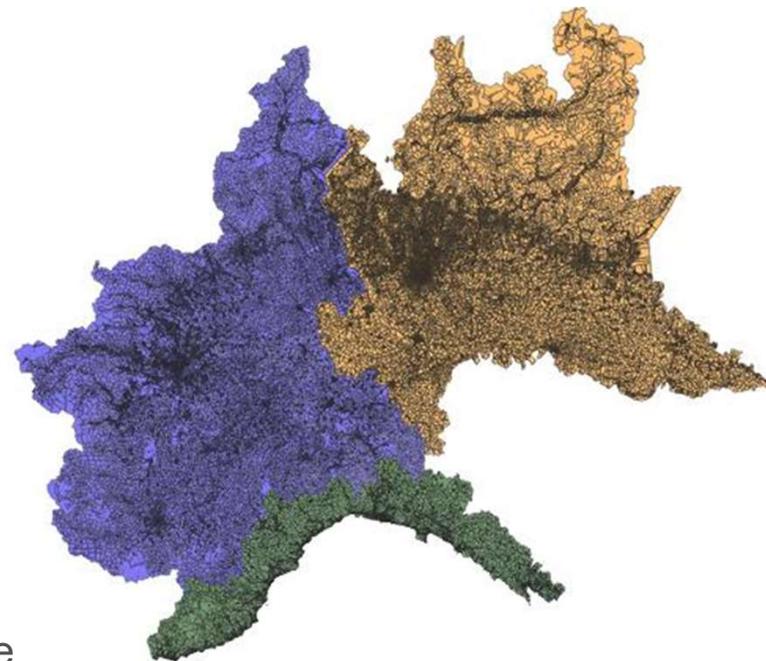
Dati geografici pubblici nel sito Istat

- Basi territoriali e sezioni
- Confini delle unità amministrative a fini statistici
- File in formato shp e/o kmz

<https://www.istat.it/it/informazioni-territoriali-e-cartografiche>

oppure dal motore di ricerca su *www.istat.it*
(chiavi: basi territoriali, cartografia, confini)

- Siti e open data dei singoli comuni per shapefile di altre aree subcomunali



grazie